Ανάλυση Δεδομένων στη Λογιστική και Χρηματοοικονομική

Ενότητα 6η

Περαιτέρω συμπεράσματα για το μοντέλο πολλαπλής παλινδρόμησης

Περιγραφή Ενότητας

1. Οι βασικές υποθέσεις του γραμμικού υποδείγματος παλινδρόμησης
2. Έλεγχος για Ετεροσκεδαστικότητα
3. Έλεγχος για Αυτοσυσχέτιση
4. Έλεγχος για Κανονική Κατανομή
5. Έλεγχος για Πολυσυγγραμικότητα

1. Οι βασικές υποθέσεις του γραμμικού υποδείγματος παλινδρόμησης

Ας θυμηθούμε ποιες είναι συνοπτικά οι βασικές υπόθεσειςτου γραμμικού υποδείγματος παλινδρόμησης

1. $E(u_t) = 0$ — Ο διαταρακτικός όρος έχει μέσο όρο μηδέν
2. $Var(u_t) = \sigma^2$ — Η διακύμανση του διαταρακτικού όρου είναι σταθερή και συγκεκριμένη για όλες τις τιμές της ανεξάρτητης μεταβλητής Χ
3. $Cov(u_i, u_j) = 0$ — Οι διαταρακτικοί όροι είναι στατιστικά ανεξάρτητοι μεταξύ τους (έχουν συνδιακύμανση μηδέν)
4. $Cov(u_t, x_t) = 0$ — Η συνδιακύμανση μεταξύ του διαταρακτικού όρου και ανεξάρτητης μεταβλητής είναι μηδέν
5. $u_t \sim N(0, \sigma^2)$ — Ο διαταρακτικός όρος κατανέμεται σύμφωνα με την κανονική κατανομή

Στην ενότητα αυτή θα αναλύσουμε περαιτέρω τις παραπάνω υποθέσεις και θα απαντήσουμε στα ακόλουθα ερωτήματα
Α. Πώς οι παραβάσεις (violations) στις παραπάνω υποθέσεις μπορούν να διαγνωστούν;
Β. Ποιες είναι οι κύριες αιτίες των παραβάσεων των υποθέσεων ;
Γ. Ποιες είναι οι συνέπειες στο γραμμικό υποδείγμα παλινδρόμησης σε περίπτωση που μία ή περισσότερες υποθέσεις παραβιάζονται, αλλά το γεγονός αυτό αγνοείται;

- **Οφείλουμε να γνωρίζουμε**

1. Τους διαγνωστικούς ελέγχους για παραβάσεις
2. Τις συνέπειες των παραβάσεων των υποθέσεων στο μοντέλο
   - οι εκτιμήσεις των συντελεστών είναι λανθασμένες.
   - τα αντίστοιχα τυπικά σφάλματα (standard errors) είναι λάθος.
   - η κατανομή που υποθέσαμε για τις στατιστικούς ελέγχους θα είναι ακατάλληλη.
3. Τις απαραίτητες διορθώσεις που πρέπει να κάνουμε.

- **Μια σημαντική παρατήρηση**

Οι Στατιστικές Κατανομές για τους διαγνωστικούς ελέγχους αποτελούν οι F και χ^2

- Η F-statistic περιλαμβάνει δύο υποδείγματα παλινδρόμησης, ένα «χωρίς περιορισμούς» υπόδειγμα (unrestricted regression) και ένα «με περιορισμούς» υπόδειγμα (restricted regression) και συγκρίνουμε τα *RSS*.
- *Η* χ^2-statistic μερικές φορές ονομάζεται και έλεγχος Lagrange Multiplier (LM) και έχει μόνο έναν βαθμό ελευθερίας: τον αριθμό των περιορισμών που ελέγχονται, *m*.
- Ασυμπτωτικά, οι 2 στατιστικές είναι ισοδύναμες αφού η χ^2 είναι ειδική περίπτωση της κατανομής F: $$\frac{\chi^2(m)}{m} \rightarrow F(m, T-k) \text{ as } T-k \rightarrow \infty$$
- Για μικρά δείγματα, προτιμάται η στατιστική F.

Υπόθεση 1: Ο διαταρακτικός όρος έχει μέσο όρο μηδέν

- Υποθέτουμε ότι ο μέσος όρος του διαταρακτικού όρου είναι μηδενικός.

- Για όλους τους διαγνωστικούς ελέγχους, δεν μπορούμε να παρατηρήσουμε τους διαταρακτικούς όρους. Για τον λόγο αυτό πραγματοποιήσουμε τους διαγνωστικούς ελέγχους στα κατάλοιπα.

- Ο μέσος όρος των καταλοίπων θα είναι πάντα μηδενικός υπό τον όρο ότι υπάρχει σταθερός όρος στην παλινδρόμηση.

Υπόθεση 2: Ομοσκεδαστικότητα

$$\mathbf{Var}(u_t) = \sigma^2 < \infty$$

Μέχρι στιγμής υποθέσαμε ότι η διακύμανση του διαταρακτικού όρου είναι σταθερή, σ^2 ·Στην περίπτωση αυτή το υπόδειγμα χαρακτηρίζεται από ομοσκεδαστικότητα (homoskedasticity). Σε αντίθετη περίπτωση, όταν η διακύμανση του διαταρακτικού όρου δεν είναι σταθερή, τότε το υπόδειγμα χαρακτηρίζεται από ετεροσκεδαστικότητα (heteroskedasticity).

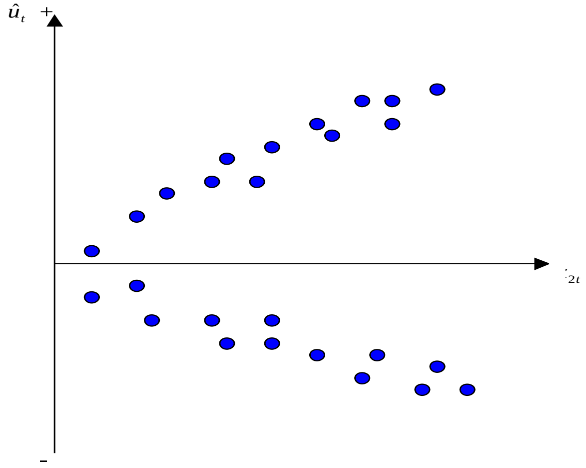
Διάγνωση της Ετεροσκεδαστικότητας

Α. Γραφικός Έλεγχος

Ένας τρόπος διερεύνησης της ύπαρξης ετεροσκεδαστικότητας είναι να εκτιμήσετε το μοντέλο σας χρησιμοποιώντας τη μέθοδο ελαχίστων τετραγώνων και να σχεδιάσετε τα κατάλοιπα.

- ❖ Εάν τα κατάλοιπα είναι ομοσκεδαστικά, δεν πρέπει να υπάρχουν μοτίβα οποιουδήποτε είδους.
- ❖ Εάν τα κατάλοιπα είναι ετεροσκεδαστικά, μπορεί να τείνουν να παρουσιάζουν μεγαλύτερη ή μικρότερη διακύμανση με κάποιο συστηματικό τρόπο.

- Έστω ότι κάνουμε εκτίμηση ενός υποδείγματος παλινδρόμησης και υπολογίζουμε τα κατάλοιπα $\hat{u}_t$



Από το διάγραμμα μπορούμε να συμπεράνουμε ότι τα σφάλματα είναι ετεροσκεδαστικά, επειδή η διακύμανση αυξάνεται συστηματικά με την ανεξάρτητη μεταβλητή.

Β. Ο έλεγχος των Goldfeld-Quandt

- Χωρίστε το συνολικό δείγμα μήκους T σε δύο υπο-δείγματα μήκους T_1 και T_2. Το μοντέλο παλινδρόμησης υπολογίζεται σε κάθε υπόδειγμα και υπολογίζονται οι δύο διακυμάνσεις των καταλοίπων.

- Η μηδενική υπόθεση είναι ότι οι δύο διακυμάνσεις των καταλοίπων είναι ίσες $\sigma_1^2 = \sigma_2^2$

- Η στατιστική είναι ο λόγος των δύο διακυμάνσεων των καταλοίπων, όπου η μεγαλύτερη από τις δύο διακυμάνσεις πρέπει να τοποθετηθεί στον αριθμητή:

$$GQ = \frac{s_1^2}{s_2^2}$$

- Η GQ στατιστική ακολουθεί την F (T_1-k, T_2-k) κατανομή κάτω από τη μηδενική υπόθεση της ομοσκεδαστικότητας.

- Ένα πρόβλημα με αυτόν τον έλεγχο είναι ότι η επιλογή του σημείο διαχωρισμού του δείγματος είναι συνήθως αυθαίρετη και μπορεί να επηρεάσει σημαντικά το αποτέλεσμα του ελέγχου.

Γ. Ο έλεγχος του White

- Ο έλεγχος αυτός βασίζεται στα κατάλοιπα, αλλά δεν έχει τους περιορισμούς του ελέγχου Goldfeld-Quandt. Επίσης, δεν προϋποθέτει τον καθορισμό των μεταβλητών που ευθύνονται για την ετεροσκεδαστικότητα. Γενικά, ο έλεγχος του White αποτελέι έναν από τους πιο δημοφιλείς διαγνωστικούς ελέγχους για την ετεροσκεδαστικότητα επειδή κάνει λίγες υποθέσεις σχετικά με τη μορφή της ετεροσκεδαστικότητας.
- <u>Βήματα:</u>

1. Εκτιμούμε το πολλαπλό γραμμικό υπόδειγμα $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$ με την μέθοδο των ελαχίστων τετραγώνων και σώζουμε τα κατάλοιπα $\hat{u}_t$.
2. Στη συνέχεια, εκτιμούμε την βοηθητική παλινδρόμηση

$$\hat{u}_t^2 = \alpha_1 + \alpha_2 x_{2t} + \alpha_3 x_{3t} + \alpha_4 x_{2t}^2 + \alpha_5 x_{3t}^2 + \alpha_6 x_{2t} x_{3t} + v_t$$

3. Παίρνουμε τον συντελεστή προσδιορισμού R^2.

4. Υπολογίζουμε την στατιστική του White ως εξής:

$W= T R^2 \sim \chi^2 (m)$ όπου T ο αριθμός των παρατηρήσεων (μέγεθος του δείγματος) και m είναι ο αριθμός των ανεξάρτητων μεταβλητών της βοηθητικής παλινδρόμησης χωρίς τον σταθερό όρο.

5. Εάν η στατιστική χ^2 είναι μεγαλύτερη από την αντίστοιχη τιμή από τον στατιστικό πίνακα, τότε απορρίψτε την μηδενική υπόθεση ότι τα κατάλοιπα χαρακτηρίζονται από ομοσκεδαστικότητα.

Οι συνέπειες της Ετεροσκεδαστικότητας

- Οι εκτιμητές των συντελεστών που προκύπτουν από την μέθοδο ελαχίστων τετραγώνων, όταν ο διαταρακτικός όρος είναι ετεροσκεδαστικός, εξακολουθούν να είναι γραμμικοί και αμερόληπτοι.

- Το πρόβλημα που δημιουργείται αναφέρεται κυρίως στις εκτιμήσεις των διακυμάνσεων τους και την αποτελεσματικότητα τους.

- Αυτό συνεπάγεται ότι εάν εξακολουθούμε να χρησιμοποιούμε την εκτίμηση ελαχίστων τετραγώνων παρουσία ετεροσκεδαστικότητας, τα τυπικά μας σφάλματα (standard errors) θα μπορούσαν να είναι ακατάλληλα και συνεπώς τυχόν συμπεράσματα που κάνουμε θα μπορούσαν να είναι παραπλανητικά.

- Το αν τα τυπικά σφάλματα που υπολογίζονται είναι πολύ μεγάλα ή πολύ μικρά, εξαρτάται από τη μορφή της ετεροσεδαστικότητας.

- Συνήθως γίνεται υποεκτίμηση των τυπικών σφαλμάτων και διακυμάνσεων

  - Η υποεκτίμηση των τυπικών σφαλμάτων και διακυμάνσεων οδηγεί σε υψηλότερες τιμές των στατιστικών $\boldsymbol{t}$ και $\boldsymbol{F}$.

  - Τα διαστήματα εμπιστοσύνης είναι αναξιόπιστα και οι προβλέψεις του υποδείγματος είναι μη αποτελεσματικές.

  - Τα συμπεράσματα μας για τις παραμέτρους του πληθυσμού θα είναι αναξιόπιστα.

  - Οι εκτιμητές όμως εξακολουθούν να είναι αμερόληπτοι και συνεπείς, γιατί καμία από τις ανεξάρτητες μεταβλητές δε συσχετίζεται με τον διαταρακτικό όρο. Άρα, οι τιμές των εκτιμημένων συντελεστών θα είναι πολύ κοντά στις πραγματικές παραμέτρους.

Αντιμετώπιση της ετεροσκεδαστικότητας

Η αποτελεσματική αντιμετώπιση της ετεροσκεδαστικότητας εξαρτάται από το κατά πόσο είναι γνωρίζουμε τη μορφή της ετεροσκεδαστικότητας. Στη συνέχεια θα δείξουμε προσεγγίσεις αντιμετώπισης της ετεροσκεδαστικότητας.

Α. Εκτίμηση του υποδείγματος όταν υποθέτουμε τη μορφή της ετεροσκεδαστικότητας

Εάν είναι γνωστή η μορφή (δηλ. η αιτία) της ετεροσκεδαστικότητας, τότε μπορούμε να χρησιμοποιήσουμε μια μέθοδο εκτίμησης που το λαμβάνει υπόψη, την **γενικευμένη μέθοδο ελαχίστων τετραγώνων (generalised least squares, GLS).**

- Έστω το γραμμικό υπόδειγμα $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$ και έστω ότι η διακύμανση σφάλματος σχετίζεται με μια άλλη μεταβλητή z_t : $\operatorname{var}(u_t) = \sigma^2 z_t^2$
- Για να αφαιρέσετε την ετεροσεδαστικότητα, διαιρέστε την εξίσωση παλινδρόμησης με z_t

$$\frac{y_t}{z_t} = \beta_1 \frac{1}{z_t} + \beta_2 \frac{x_{2t}}{z_t} + \beta_3 \frac{x_{3t}}{z_t} + v_t \quad \text{όπου} \quad v_t = \frac{u_t}{z_t} \quad \text{και}$$

$$\operatorname{var}(v_t) = \operatorname{var}\left(\frac{u_t}{z_t}\right) = \frac{\operatorname{var}(u_t)}{z_t^2} = \frac{\sigma^2 z_t^2}{z_t^2} = \sigma^2 \quad \text{για γνωστό } z_t$$

Έτσι, ο διαταρακτικός όρος στο νέο υπόδειγμα παλινδρόμησης θα είναι ομοσκεδαστικός.

Η γενικευμένη μέθοδος ελαχίστων τετραγώνων είναι επίση γνωστή ωσσταθμική μέθοδο των ελαχίστων τετραγώνων.

Β. Πιο «Πρακτικές» Προσεγγίσεις στην Αντιμετώπιση της Ετεροσκεδαστικότητας

Συνήθως είναι πολύ δύσκολο για τους ερευνητές να γνωρίζουν την ακριβή αιτία και κατά συνέπεια μορφή της ετεροσκεδαστικότητας. Έτσι είναι πολύ δύσκολο να εφαρμοστεί στην πράξη η προσέγγιση που είδαμε στην προηγούμενη διαφάνεια. Δύο προσεγγίσεις που ακολουθούνται στην πράξη είναι οι ακόλουθες:

- Μετατροπή των μεταβλητών σε λογαριθμική μορφή ή με κατά κάποιο άλλο τρόπο μετασχηματισμό των μεταβλητών (π.χ. rescaling).
- Χρησιμοποιήστε τις συνεπή σε ετεροσκεδαστικότητα τυπικές εκτιμήσεις σφαλμάτων του White.

  - ❖ Το αποτέλεσμα της χρήσης της διόρθωσης του White είναι ότι γενικά τα τυπικά σφάλματα (standard errors) για τους συντελεστές (slope coefficients) αυξάνονται σε σχέση με τα συνήθη τυπικά σφάλματα των ελαχιστών τετραγώνων. Αυτό μας καθιστά πιο «συντηρητικούς» στον έλεγχο υποθέσεων, έτσι ώστε να χρειαζόμαστε περισσότερα στοιχεία ενάντια στην μηδενική υπόθεση πριν την απορρίψουμε.

Υπόθεση 3: Μη αυτοσυσχέτιση μεταξύ των διαταρακτικών όρων

$$Cov\,(u_i, u_j) = 0$$

- Στις βασικές υποθέσεις των απλών και πολλαπλών γραμμικών υποδειγμάτων, υποθέσαμε ότι δεν υπάρχει αυτοσυσχέτιση μεταξύ των σφαλμάτων (διαταρακτικών όρων) u_t. Αυτή η υπόθεση δηλώνει ότι οι τιμές των διαταρακτικών όρων είναι ανεξάρτητες μεταξύ τους (σειριακή ανεξαρτησία). Αυτό είναι ουσιαστικά το ίδιο με το να πούμε ότι δεν υπάρχει κάποιο συγκεκριμένο μοτίβο στα σφάλματα.

- Δηλαδή για δύο διαφορετικές παρατηρήσεις του σφάλματος (διαταρακτικού όρου) η συνδιακύμανση τους είναι μηδέν: $cov(u_i,u_j)=0$, για $i\neq j$. Αν αυτή η υπόθεση δεν ισχύει, τότε οι διαταρακτικοί όροι δεν είναι ανεξάρτητοι κατά ζεύγη, αλλά αυτοσυσχετιζόμενοι κατά ζεύγη (ή σειριακά συσχετιζόμενοι).

- Η αυτοσυσχέτιση αναφέρεται στη συσχέτιση δύο διαδοχικών τιμών της ίδιας μεταβλητής.

- Προφανώς δεν μπορούμε να έχουμε τα πραγματικά u_t, οπότε χρησιμοποιούμε τα κατάλοιπά τους.

Η έννοια της χρονικής υστέρησης

Προτού προχωρήσουμε στην ανάλυση της αυτοσυσχέτισης, είναι απαραίτητο να κατανοήσουμε την έννοια της χρονικής υστέρησης.

t	y_t	y_{t-1}	Δy_t
1989M09	0.8	-	-
1989M10	1.3	0.8	1.3-0.8=0.5
1989M11	-0.9	1.3	-0.9-1.3=-2.2
1989M12	0.2	-0.9	0.2--0.9=1.1
1990M01	-1.7	0.2	-1.7-0.2=-1.9
1990M02	2.3	-1.7	2.3--1.7=4.0
1990M03	0.1	2.3	0.1-2.3=-2.2
1990M04	0.0	0.1	0.0-0.1=-0.1
.	.	.	.
.	.	.	.
.	.	.	.

Διάγνωση της Αυτοσυσχέτισης

Α. Γραφικός Έλεγχος

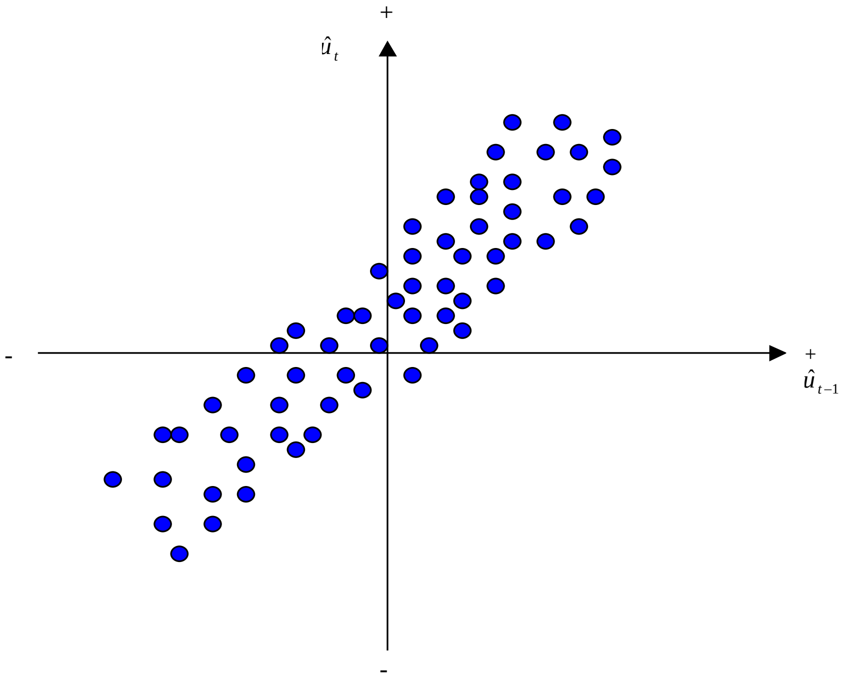
Μπορούμε να διαπιστώσουμε την ύπαρξη αυτοσυσχέτισης (πρώτης τάξης)

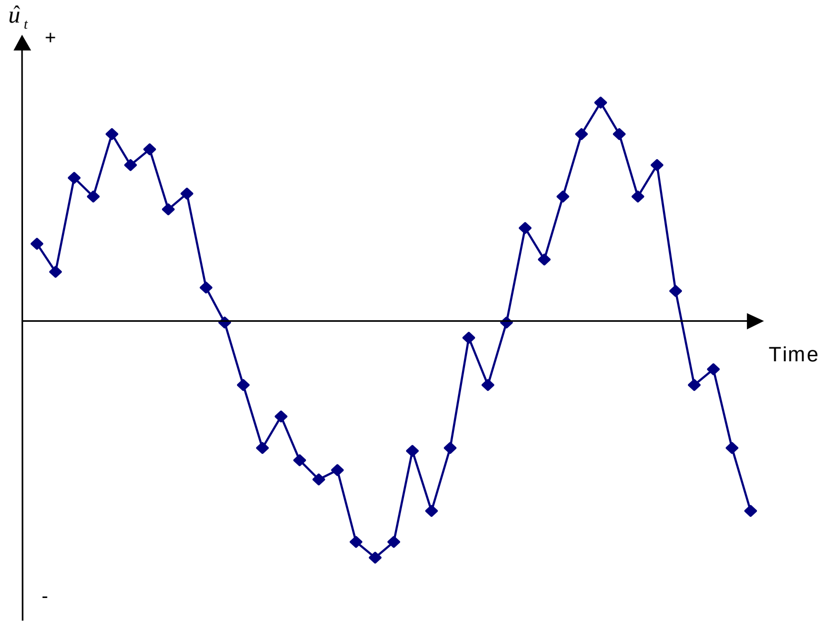
Α. με βάση το διάγραμμα διασποράς των καταλοίπων. Συγκεκριμένα παρατηρούμε τη γραφική παράσταση των καταλοίπων $\boldsymbol{u_{t-1}}$ και $\boldsymbol{u_t}$.

Β. με βάση ένα διάγραμμα χρονοσειράς, όπου στον οριζόντιο άξονα είναι ο χρόνος και στον κατακόρυφο άξονα οι τιμές των καταλοίπων

Στις επόμενες διαφάνειες θα δείτε μοτίβα αυτοσυσχέτισης με βάση τα διαγράμματα Α και Β.

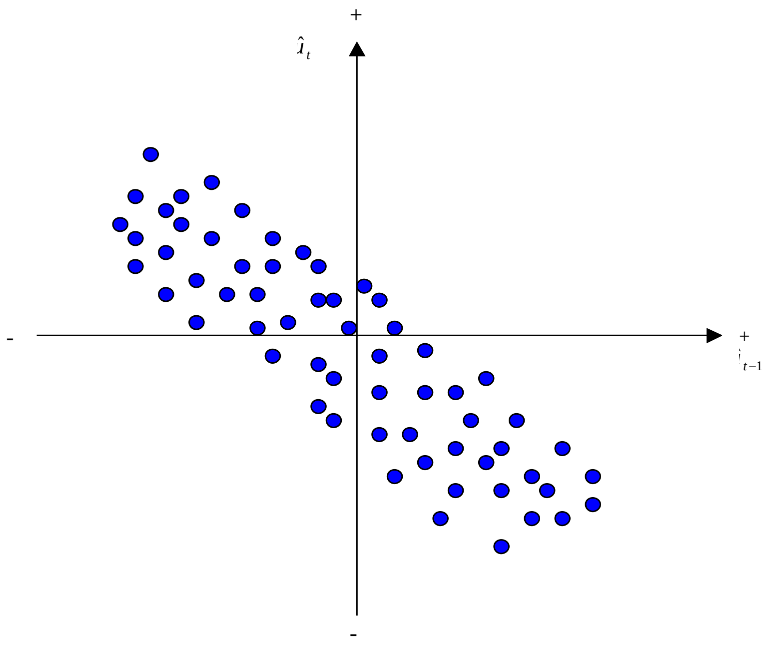
Θετική Αυτοσυσχέτιση

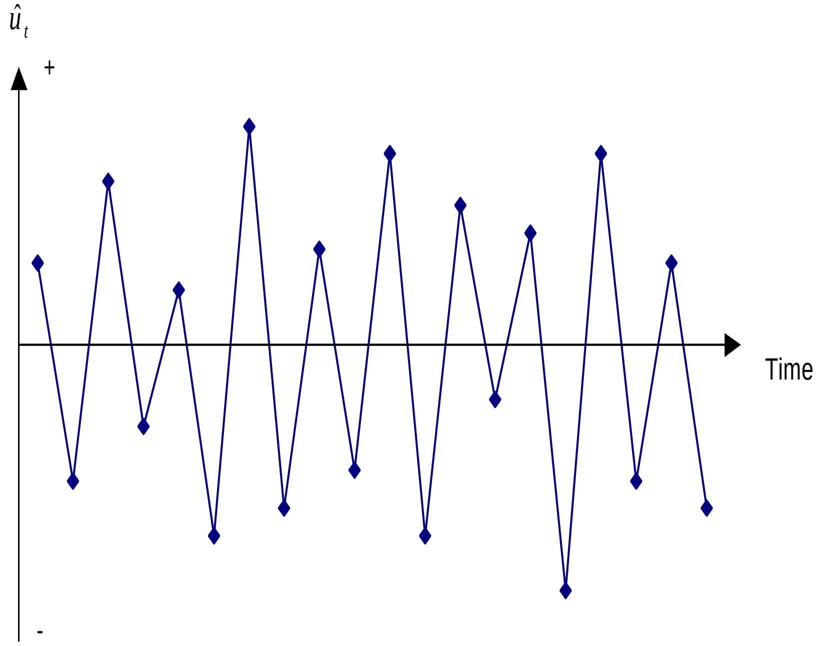




Η θετική αυτοσυσχέτιση υποδεικνύεται από ένα κυκλικό διάγραμμα καταλοίπων με την πάροδο του χρόνου.

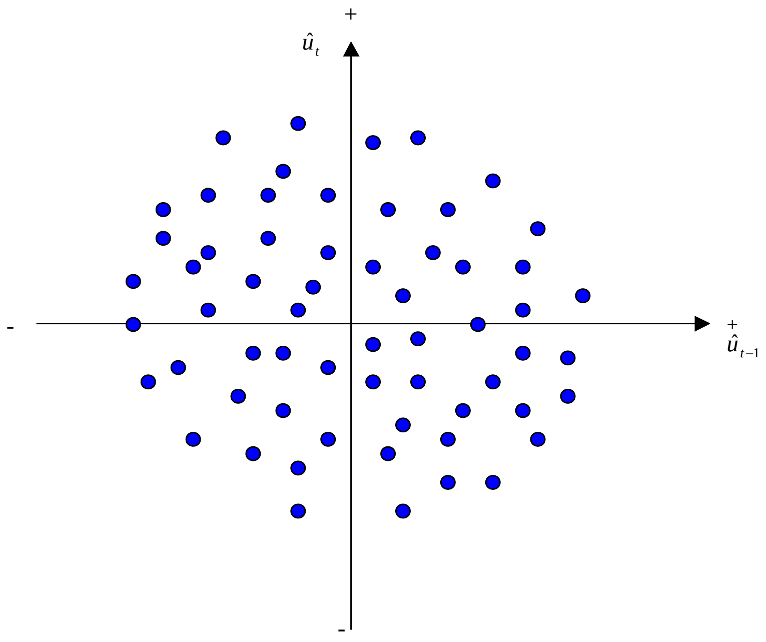
Αρνητική Αυτοσυσχέτιση

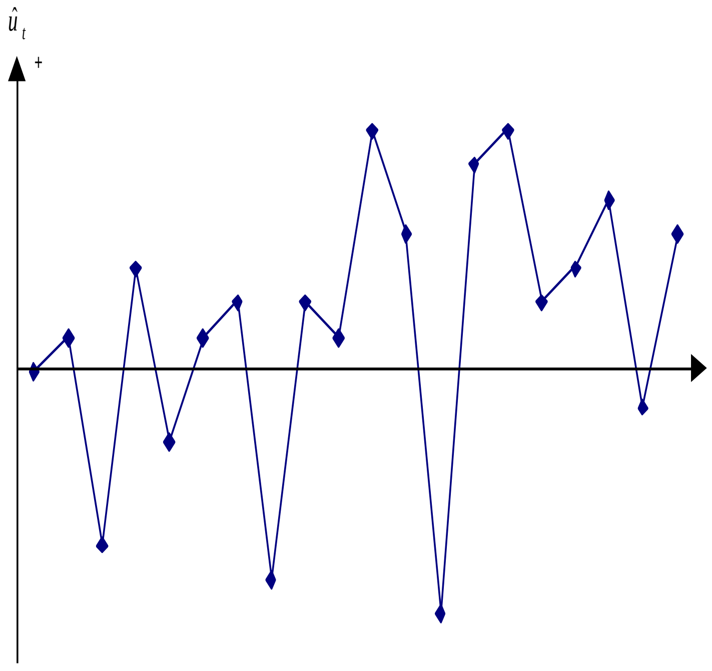




Η αρνητική αυτοσυσχέτιση υποδεικνύεται από ένα εναλλασσόμενο μοτίβο όπου τα κατάλοιπα διασχίζουν τον οριζόντιο άξονα (άξονα του χρόνου) συχνότερα από ό, τι αν κατανέμονταν τυχαία.

Χωρίς μοτίβο στα κατάλοιπα - Χωρίς αυτοσυσχέτιση





Δεν υπάρχει καθόλου μοτίβο στα κατάλοιπα.

Β. Ο έλεγχος Durbin-Watson (DW)

- Ο έλεγχος των Durbin-Watson (DW) αποτελεί έναν έλεγχο για **αυτοσυσχέτιση πρώτης τάξης** - δηλαδή υποθέτει ότι η σχέση είναι μεταξύ ενός σφάλματος και του προηγούμενου

$$u_t = \rho u_{t-1} + v_{t,}\text{, όπου } v_t \sim N(0, \sigma_v^2)$$

- **Προσοχή:** Προϋποθέσεις που πρέπει να πληρούνται για να είναι ο έλεγχος των Durbin-Watson (DW) έγκυρος

1. Σταθερός όρος στην παλινδρόμηση.

2. Οι ανεξάρτητες μεταβλητές είναι μη στοχαστικές.

3. Να μην υπάρχει ως ανεξάρτητη μεταβλητή η εξαρτημένη με χρονική υστέρηση.

- Η στατιστική του DW εξετάζει τις ακόλουθες υποθέσεις:

  $H_0 : \rho = 0$ Δεν υπάρχει αυτοσυσχέτιση πρώτης τάξης στα κατάλοιπα

  $H_1 : \rho \neq 0$ Υπάρχει αυτοσυσχέτιση πρώτης τάξης στα κατάλοιπα

- Βήματα

1. Εκτιμούμε το υπόδειγμα παλινδρόμησης με την μέθοδο των ελαχίστων τετραγώνων και παίρνουμε τα κατάλοιπα $\hat{u}_t$

2. Υπολογίζουμε τη στατιστική του DW υπολογίζεται ως εξής

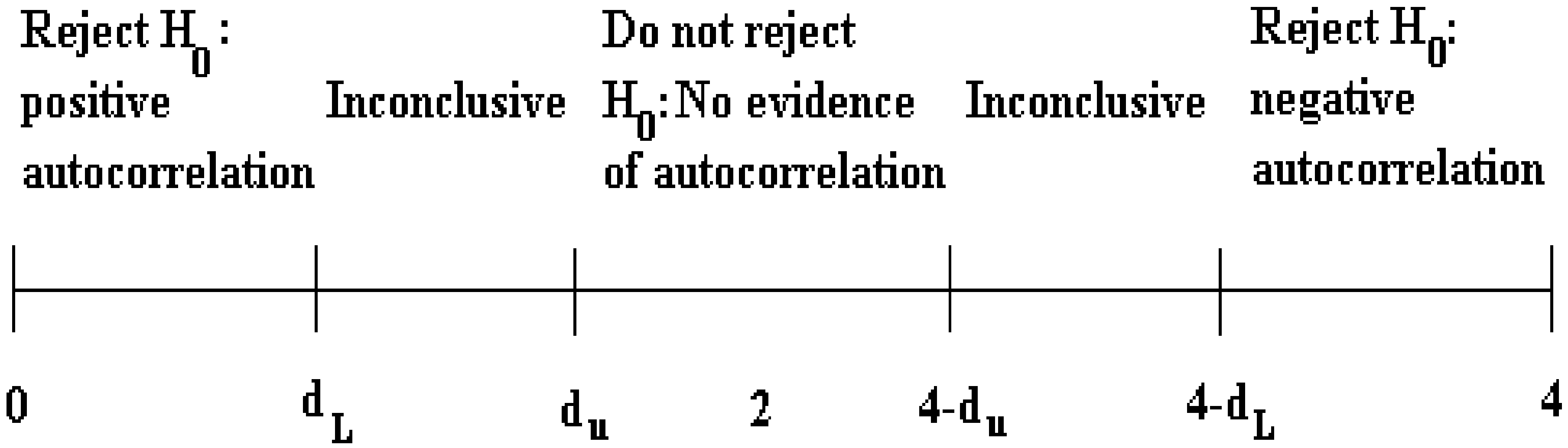
$$DW = \frac{\sum_{t=2}^{T} \left(\hat{u}_t - \hat{u}_{t-1}\right)^2}{\sum_{t=2}^{T} \hat{u}_t^2}$$

- Μπορούμε επίσης να γράψουμε $DW \approx 2(1-\hat{\rho})$

  όπου ο $\hat{\rho}$ είναι ο εκτιμώμενος συντελεστής συσχέτισης. Δεδομένου ότι ο $\hat{\rho}$ αποτελεί συντελεστή συσχέτισης, υπονοεί ότι $-1 \leq \hat{p} \leq 1$
- Άρα θα μπορούσαμε επίσης να γράψουμε ότι $0 \leq DW \leq 4$.

- Εάν $\hat{\rho}$ = 0, *DW* = 2. Σε γενικές γραμμές, δεν απορρίπτουμε την μηδενική υπόθεση εάν *DW* είναι κοντά στο 2 δηλαδή υπάρχουν μικρές ενδείξεις ύπαρξης αυτοσυσχέτισης

3: Συγκρίνουμε την τιμή του DW με δύο κριτικές τιμές. Το DW έχει 2 κριτικές τιμές, μια ανώτερη κριτική τιμή (d_u) και μια χαμηλότερη κριτική τιμή (d_L), και υπάρχει επίσης μια ενδιάμεση περιοχή όπου δεν μπορούμε ούτε να απορρίψουμε ούτε να μην απορρίψουμε την μηδενική υπόθεση H_0. Δείτε συγκεκριμένα την επόμενη διαφάνεια:


Reject H_0 : positive autocorrelation
Inconclusive
Do not reject H_0: No evidence of autocorrelation
Inconclusive
Reject H_0: negative autocorrelation
0
d_L
d_u
2
$4\text{-}d_u$
$4\text{-}d_L$
4

Γ. Ο έλεγχος Breusch-Godfrey(BG)

- Αποτελεί έναν γενικότερο έλεγχο όπου εξετάζει αυτοσυσχέτιση **μεγαλύτερης τάξης**

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \ldots + \rho_r u_{t-r} + v_t \quad , \quad v_t \sim \mathrm{N}(0, \sigma_v^2)$$

- Η μηδενική και εναλλακτική υπόθεση είναι οι ακόλουθες:

$\mathrm{H}_0 : \rho_1 = 0$ και $\rho_2 = 0$ και ... και $\rho_r = 0$

$\mathrm{H}_1 : \rho_1 \neq 0$ ή $\rho_2 \neq 0$ ή ... ή $\rho_r \neq 0$

- Τα βήματα είναι τα εξής:

1. Εκτιμούμε το υπόδειγμα παλινδρόμησης με την μέθοδο των ελαχίστων τετραγώνων και σώζουμε τα κατάλοιπα $\hat{u}_t$.

2. Εκτιμούμε την παλινδρόμηση με εξαρτημένη μεταβλητή $\hat{u}_t$ και ανεξάρτητες τις X και $\hat{u}_{t-1}, \hat{u}_{t-2}, \ldots, \hat{u}_{t-r}$ και παίρνουμε τον συντελεστή προσδιορισμού R^2

3. Εκτιμούμε τη στατιστική του ελέγχου BG = $(T\text{-}r)R^2 \sim \chi^2(r)$ (r βαθμοί ελευθερίας)

- Εάν η στατιστική BG υπερβεί την κριτική τιμή από τους στατιστικούς πίνακες, τότε απορρίπτουμε την μηδενική υπόθεση ότι δεν υπάρχει αυτοσυσχέτιση.

Οι συνέπειες της Αυτοσυσχέτισης

- Οι εκτιμήσεις συντελεστών που προκύπτουν από την εκτίμηση του υποδείγματος με τη μέθοδο ελαχίστων τετραγώνων παραμένουν αμερόληπτες (unbiased) και συνεπείς, αλλά είναι αναποτελεσματικές.

- Η μέθοδος των ελαχίστων τετραγώνων δεν δίνει πλέον τους πιο αποτελεσματικούς εκτιμητές και προβλέψεις. Δεν είναι πλέον **BLUE (Best Linear Unbiased Estimator).**

- Οι εκτιμητές των διακυμάνσεων των εκτιμητών είναι μεροληπτικοί και ασυνεπείς. Οι στατιστικοί έλεγχοι που βασίζονται στις διακυμάνσεις αυτές οδηγούν σε λάθος συμπεράσματα.

- Αν μεταξύ των ανεξάρτητων μεταβλητών περιλαμβάνεται και η εξαρτημένη μεταβλητή με χρονική υστέρηση, τότε οι εκτιμητές δεν είναι ούτε συνεπείς.

- Ο συντελεστής προσδιορισμού R^2 είναι πιθανό να διογκωθεί σε σχέση με τη "σωστή" τιμή του για τα θετικά συσχετιζόμενα κατάλοιπα.

Αντιμετώπιση της αυτοσυσχέτισης

- Εάν η μορφή της αυτοσυσχέτισης είναι γνωστή, θα μπορούσαμε να χρησιμοποιήσουμε τη **γενικευμένη μέθοδο ελαχίστων τετραγώνων (Generalised Least Square (GLS) -** δηλαδή μια προσέγγιση που λαμβάνει υπόψιν τα αυτοσυσχετιζόμενα κατάλοιπα, π.χ., Cochrane-Orcutt.

- Αλλά τέτοιες διαδικασίες που «διορθώνουν» την αυτοσυσχέτιση απαιτούν υποθέσεις σχετικά με τη μορφή της αυτοσυσχέτισης.

- Εάν αυτές οι υποθέσεις δεν είναι βάσιμες, η «θεραπεία» (αντιμετώπιση) θα ήταν πιο επικίνδυνη από την ασθένεια! - Hendry and Mizon (1978).

- Ωστόσο, είναι απίθανο να είναι γνωστή η μορφή της αυτοσυσχέτισης και μια πιο «σύγχρονη» άποψη είναι ότι η αυτοσυσχέτιση των καταλοίπων δίνει την ευκαιρία να τροποποιηθεί η παλινδρόμηση.

- Στις παρακάτω διαφάνειες ακολουθεί η ανάλυση όλων των παραπάνω:

Εκτίμηση του υποδείγματος όταν ο συντελεστής αυτοσυσχέτισης $\boldsymbol{\rho}$ είναι γνωστός

- **Γενικευμένη μέθοδος ελαχίστων τετραγώνων**

Εφαρμόζεται η γενικευμένη μέθοδος ελαχίστων τετραγώνων. Έστω το υπόδειγμα

$$y_t = \beta_1 + \beta_2 x_{2t} + u_t,\ t=1,2,\ldots,T$$

και $$u_t = \rho u_{t-1} + v_t,\ \text{όπου } v_t \sim N(0,\ \sigma_v^2)$$

Γράφουμε το υπόδειγμα για χρονική περίοδο *t-1* (υστερούμε την αρχική παλινδρόμηση κατά μια περίοδο):

$$y_{t-1} = \beta_1 + \beta_2 x_{2t-1} + u_{t-1}$$

Πολλαπλασιάζουμε και τα δύο μέλη με τον συντελεστή συσχέτισης ρ

$$\rho y_{t-1} = \rho\beta_1 + \rho\beta_2 x_{2t-1} + \rho u_{t-1}$$

Αφαιρούμε την παλινδρόμηση αυτή από την αρχική:

$$y_t - \rho y_{t-1} = \beta_1 - \rho\beta_1 + (\beta_2 x_{2t} - \rho\beta_2 x_{2t-1}) + (u_t - \rho u_{t-1}) \quad \text{ή}$$

$$y_t - \rho y_{t-1} = (\beta_1 - \rho\beta_1) + \beta_2(x_{2t} - \rho x_{2t-1}) + (u_t - \rho u_{t-1}) \quad \text{ή}$$

$$y_t^* = \beta_1^* + \beta_2 x_{2t}^* + u_t^*$$

Τέλος θέτουμε:

$$y_t^* = y_t\sqrt{1-\rho^2} \qquad x_{2t}^* = x_{2t}\sqrt{1-\rho^2} \qquad u_t^* = u_t\sqrt{1-\rho^2}$$

Το υπόδειγμα αυτό δεν έχει πρόβλημα αυτοσυσχέτισης.

- **Μέθοδος Cochrane-Orcutt**

Βήματα:

1. Έστω το υπόδειγμα $y_t = \beta_1 + \beta_2 x_{2t} + u_t$, $t=1,2,\ldots,T$
2. Εκτιμήστε το παραπάνω υπόδειγμα παλινδρόμησης αγνοώντας την ύπαρξη αυτοσυσχέτισης καταλοίπων.
3. Από την εκτίμηση της παλινδρόμησης, παίρνετε τα κατάλοιπα $\hat{u}$.
4. Εκτιμήστε την ακόλουθη παλινδρόμηση

$$u_t = \rho u_{t-1} + v_t,\ \text{όπου } v_t \sim N(0, \sigma_v^2)$$

5. Από την παραπάνω εκτίμηση παίρνετε το $\hat{\rho}$.
6. Με βάση το $\hat{\rho}$ κατασκευάστε τα $y_t^* = y_t\sqrt{1-\rho^2}$ και $x_{2t}^* = x_{2t}\sqrt{1-\rho^2}$
7. Εκτιμήστε το ακόλουθο υπόδειγμα

$$y_t^* = \beta_1^* + \beta_2 x_{2t}^* + v_t$$

8. Με βάση τα καινούρια β_1^* και β_2, ξανα-κατασκευάστε τα κατάλοιπα και επαναλάβετε τα βήματα 4 έως 7 έως ότου η διαφορά του $\hat{\rho}$ του βήματος 5 μεταξύ δύο διαδοχικών επαναλήψεων είναι μικ΄ροτερη από μια συγκεκριμένη τιμή (π.χ. 0.01).

- **Η μέθοδος των πρώτων διαφορών**
  - Ένας άλλος τρόπος για να αντιμετωπίσετε το πρόβλημα της αυτοσυσχέτισης είναι να μεταβείτε σε ένα μοντέλο βασισμένο στις πρώτες διαφορές. Έστω το υπόδειγμα $y_t = \beta_1 + \beta_2 x_{2t} + u_t$, $t=1,2,\dots,T$

Εάν ρ=1

- Δηλώστε την πρώτη διαφορά του y_t, δηλαδή $\Delta y_t = y_t - y_{t-1}$
- Ομοίως για τις ανεξάρτητες μεταβλητές x, δηλαδή $\Delta x_{2t} = x_{2t} - x_{2t-1}$
- Το μοντέλο μετασχηματίζεται ως εξής:

$\Delta y_t = \beta_2 \Delta x_{2t} + v_t$ Με άλλα λόγια στη θέση των αρχικών μεταβλητών έχουμε τις πρώτες διαφορές και δεν υπάρχει ο σταθερός όρος.

Εάν ρ= -1

- Το μοντέλο μετασχηματίζεται ως εξής:

$y_t + y_{t-1} = 2\beta_1 + \beta_2(x_{2t}+x_{2t-1}) + u_t + u_{t-1}$ ή

$y_t^{new} = \beta_1 + \beta_2 x_{2t}^{new} + v_t$ Όπου $y_t^{new} = \frac{y_t + y_{t-1}}{2}$, $x_{2t}^{new} = \frac{x_{2t} + x_{2t-1}}{2}$, $v_t = \frac{u_t + u_{t-1}}{2}$

Με άλλα λόγια στη θέση των αρχικών μεταβλητών έχουμε τους κινητούς μέσους δύο περιόδων

Εκτίμηση του υποδείγματος όταν ο συντελεστής αυτοσυσχέτισης $\boldsymbol{\rho}$ είναι άγνωστος

- **Η τεχνική του Durbin**

Έστω το υπόδειγμα $y_t = \beta_1 + \beta_2 x_{2t} + u_t$, *t=1,2,...,T*

Όπως και στην περίπτωση της γενικευμένης μεθόδου ελαχίστων τετραγώνων, το υπόδειγμα μετασχηματίζεται ως εξής:

$$y_t - \rho y_{t-1} = (\beta_1 - \rho\beta_1) + \beta_2 x_{2t} - \rho\beta_2 x_{2t-1} + (u_t - \rho u_{t-1})$$

$$y_t = \beta_1(1-\rho) + \rho y_{t-1} + \beta_2 x_{2t} - \rho\beta_2 x_{2t-1} + \nu_t$$

$$y_t = \delta_1 + \rho y_{t-1} + \delta_2 x_{2t} + \delta_3 x_{2t-1} + \nu_t$$

$$\delta_1 = \beta_1(1-\rho) \qquad \delta_2 = \beta_2 \quad \delta_3 = -\rho\beta_2$$

Εκτιμάμε τους συντελεστές του μετασχηματισμένου υποδείγματος με τη μέθοδο ελαχίστων τετραγώνων

$$y_t = \delta_1 + \rho y_{t-1} + \delta_2 x_{2t} + \delta_3 x_{2t-1} + \nu_t$$

και έπειτα υπολογίζουμε τους συντελεστές του αρχικού υποδείγματος $y_t = \beta_1 + \beta_2 x_{2t} + u_t$, *t=1,2,...,T*

από τις σχέσεις $\delta_1 = \beta_1(1-\rho) \qquad \delta_2 = \beta_2 \qquad \delta_3 = -\rho\beta_2$

Μια Πιο «Πρακτική» Προσέγγιση στην Αντιμετώπιση της Αυτοσυσχέτισης

Για να μπορέσουμε να χρησιμοποιήσουμε τις παραπάνω προσεγγίσεις θα πρέπει πρώτα να εξεταστεί η εγκυρότητά τους. Μια προσεγγίση που ακολουθείται στην πράξη είναι ο υπολογισμός των τυπικών σφαλμάτων που να λαμβάνουν υπόψιν την αυτοσυσχέτιση των καταλοίπων.

- Θυμηθείτε ότι μια πρακτική προσέγγιση για την ετεροσκεδαστικότητα είναι ο υπολογισμός των συνεπή σε ετεροσκεδαστικότητα τυπικών εκτιμήσεων των σφαλμάτων **του White.** **<u>Η μέθοδος White είναι κατάλληλη όταν τα κατάλοιπα είναι ετεροσκεδαστικά αλλά μη αυτοσυσχετιζόμενα.</u>**

- Οι **Newey West** αναπτύσουν έναν εκτιμητή διακυμάνσεων- συνδυακυμάνσεων των συντελεστών (δηλάδη υπολογίζουν τις τυπικές εκτιμήσεις των σφαλμάτων) που να είναι συνεπής **ΚΑΙ στην ετεροσκεδαστικότητα ΚΑΙ αυτοσυσχέτιση.**

Δυναμικά Μοντέλα

- Μια πιο «μοντέρνα» άποψη που αφορά την αυτοσυσχέτιση αναφέρει ότι η αυτοσυσχέτιση στα σφάλματα προκύπτει λόγω λανθασμένης δυναμικής (misspecified dynamics). Μπορούμε να εκφράσουμε την εξαρτημένη μεταβλητή ως το άθροισμα ενός μέρους που μπορεί να εξηγηθεί από το μοντέλο και ένα μέρος που δεν μπορεί να εξηγηθεί από το μοντέλο, δηλαδή $y_t = \widehat{y_t} + \widehat{u_t}$. Η αυτοσυσχέτιση στα κατάλοιπα οφείλεται σε μια δυναμική της y που δεν έχει μοντελοποιηθεί και άρα δεν έχει ενσωματωθεί στην $\widehat{y_t}$. Με πιο απλά λογια υπάρχει μια πιο «πλούσια» δομή στην εξαρτημένη μεταβλητή y_t και μεγαλύτερη πληροφόρηση στο δείγμα για τη δομή αυτή που δεν έχει ενσωματωθεί στα μοντέλα που έχουμε δει μέχρι τώρα, τα οποία ήταν στατικά, $y_t = \beta_1 + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t$

- Αυτό που ίσως απαιτείται είναι η δημιουργία ενός δυναμικού μοντέλου που θα ενσωματώνει αυτή την νέα δομή. Αλλά μπορούμε εύκολα να επεκτείνουμε αυτήν την ανάλυση στην περίπτωση όπου η τρέχουσα τιμή του y_t εξαρτάται από τις προηγούμενες τιμές του y ή ενός από τα x, π.χ.

$$y_t = \beta_1 + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + \gamma_1 y_{t-1} + \gamma_2 x_{2t-1} + \ldots + \gamma_k x_{kt-1} + u_t$$

- Θα μπορούσαμε να επεκτείνουμε το μοντέλο ακόμη περισσότερο προσθέτοντας επιπλέον χρονικές υστερήσεις, π.χ. x_{2t-2} , y_{t-3} .

Γιατί ίσως θέλουμε/πρέπει να συμπεριλάβουμε χρονικές υστερήσεις σε μια παλινδρόμηση;

- Αδράνεια της εξαρτημένης μεταβλητής.
- Υπερβολικές αντιδράσεις (Over-reactions).
- Μέτρηση χρονοσειρών ως επικαλυπτόμενων κινητών μέσων όρων.
- Ωστόσο, άλλα προβλήματα με την παλινδρόμηση θα μπορούσαν να προκαλέσουν την απόρριψη της μηδενικής υπόθεσης ότι δεν υπάρχει αυτοσυσχέτιση:
    - ❖ Παράλειψη σχετικών μεταβλητών, οι οποίες είναι οι ίδιες αυτοσυσχετιζόμενες.
    - ❖ Εάν έχουμε «διαπράξει» σφάλμα "εσφαλμένου προσδιορισμού" χρησιμοποιώντας μια ακατάλληλη λειτουργική φόρμα.
    - ❖ Αυτοσυσχέτιση που προκύπτει από μη παραμετροποιημένη εποχικότητα.

Προβλήματα με την προσθήκη χρονικών υστερήσεων για να «θεραπεύσουμε» την αυτοσυσχέτιση

- Η συμπερίληψη χρονικών υστερήσεων της εξαρτημένης μεταβλητής παραβιάζει την υπόθεση ότι οι μεταβλητές στη δεξία πλευρά της παλινδρόμησης είναι μη στοχαστικές.
- Τι σημαίνει στην πραγματικότητα μια παλινδρόμηση με μεγάλο αριθμό χρονικών υστερήσεων;
- Σημειώστε ότι εάν εξακολουθεί να υπάρχει αυτοσυσχέτιση στα κατάλοιπα ενός μοντέλου, συμπεριλαμβανομένων των χρονικών υστερήσεων, τότε οι εκτιμητές ελαχίστων τετραγώνων δεν θα είναι καν συνεπείς.

Σημείωση

- Όταν στο στοχαστικό διαταρακτικό όρο $\boldsymbol{u_t}$ ισχύουν οι παρακάτω τρεις υποθέσεις:
- $\boldsymbol{E(u_t) = 0}$
- $\boldsymbol{Var(u_t) = \sigma^2}$
- $\boldsymbol{Cov\ (u_t\ ,\ u_{t+s}) = 0}$ για $\mathbf{t} \neq \boldsymbol{s}$

τότε λέμε ότι ο διαταρακτικός όρος είναι λευκός θόρυβος (white noise).

Υπόθεση 4: Οι ανεξάρτητες μεταβλητές είναι μη στοχαστικές

$$\mathrm{Cov}\,(u_t, x_t) = 0$$

- Οι εκτιμητές ελαχίστων τετραγώνων είναι συνεπείς και αμερόληπτοι όταν οι ανεξάρτητες μεταβλητές είναι στοχαστικές, δεδομένου όμως ότι οι ανεξάρτητες μεταβλητές δεν είναι συσχετιζόμενες με τον διαταρακτικό όρο.
- Θυμηθείτε ότι

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = (X'X)^{-1} X' y \qquad \text{και } y = X\beta + u$$

$$\hat{\beta} = (X'X)^{-1}X'(X\beta + u)$$

$$\hat{\beta} = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u$$

$$\hat{\beta} = \beta + (X'X)^{-1}X'u$$

- Άρα $\widehat{E(\beta)} = \mathrm{E}(\beta) + \mathrm{E}((X'X)^{-1}X'u)$

$$\widehat{E(\beta)} = \beta + \mathrm{E}[(X'X)^{-1}X']\mathrm{E}(u)$$

- Επειδή $\mathrm{E}(u)$, τότε $\widehat{E(\beta)} = \beta$ που σημαίνει ότι ο εκτιμητής παραμένει αμερόληπτος ακόμη και στην περίπτωση που οι ανεξάρτητες μεταβλητές είναι στοχαστικές.

- Στην περίπτωση όμως όπου μία ή περισσότερες μεταβλητές συσχετίζονται με τον διαταρακτικό όρο, τότε ο εκτιμητής ελαχίστων τετραγώνων δε θα είναι ούτε συνεπής.
- Εξήγηση:

Έστω το υπόδειγμα $y_t = \beta_1 + \beta_2 x_{2t} + u_t$, *t=1,2,...,T* και ότι η x_{2t} έιναι θετικά συσχετιζόμεη με τον u_t.

Όταν ο u_t τυγχάνει να έχει υψηλή τιμή , τότε και η y_t θα έχει υψηλή τιμή. Αλλά εάν η x_{2t} έιναι θετικά συσχετιζόμενη με τον u_t, τότε και η x_{2t} είναι πιθανόν να έχει υψηλή τιμή.

Άρα ο εκτιμητής ελαχίστων τετραγώνων λανθασμένα θα αποδόσει την υψηλή τιμή της y_t στην υψηλή τιμή της x_{2t}, αλλά στην πραγματικότητα η y_t έχει υψηλή τιμή επείδη ο u_t έχει υψηλή τιμή.

Το παραπάνω θα έχει ως αποτέλεσμα ο εκτιμητής ελαχίστων τετραγώνων να είναι μεροληπτικός και ασυνεπής και η γραμμή προσαρμογής της παλινδρόμησης θα λάβει υπόψιν τα χαρακτηριστικά των δεδομένων πολύ καλύτερα από ότι είναι στην πραγματικότητα.

Υπόθεση 5: Ο διαταρακτικός όρος κατανέμεται κανονικά με μέσο μηδέν και σταθερή διακύμανση.

$$u_t \sim N(0, \sigma^2)$$

Μια από τις υποθέσεις του υποδείγματος της γραμμικής παλινδρόμησης είναι ότι ο διαταρακτικός όρος κατανέμεται κανονικά με μέσο μηδέν και σταθερή διακύμανση.

Άρα θα έχουμε: $u_t \sim N(0, \sigma^2)$

Οι Jarque και Bera (1987) πρότειναν μια μέθοδο που βασίζεται στα κατάλοιπα που προκύπτουν από τη μέθοδο των ελαχίστων τετραγώνων:

- Οι συντελεστές της ασυμμετρίας και της κύρτωσης ορίζονται ως εξής:

$$b_1 = \frac{E[u^3]}{\left(\sigma^2\right)^{3/2}} \qquad b_2 = \frac{E[u^4]}{\left(\sigma^2\right)^2}$$

- Η στατική των Bera Jarque είναι η ακόλουθη $JB = T\left[\frac{b_1^2}{6} + \frac{(b_2 - 3)^2}{24}\right] \sim \chi^2(2)$

Βήματα:

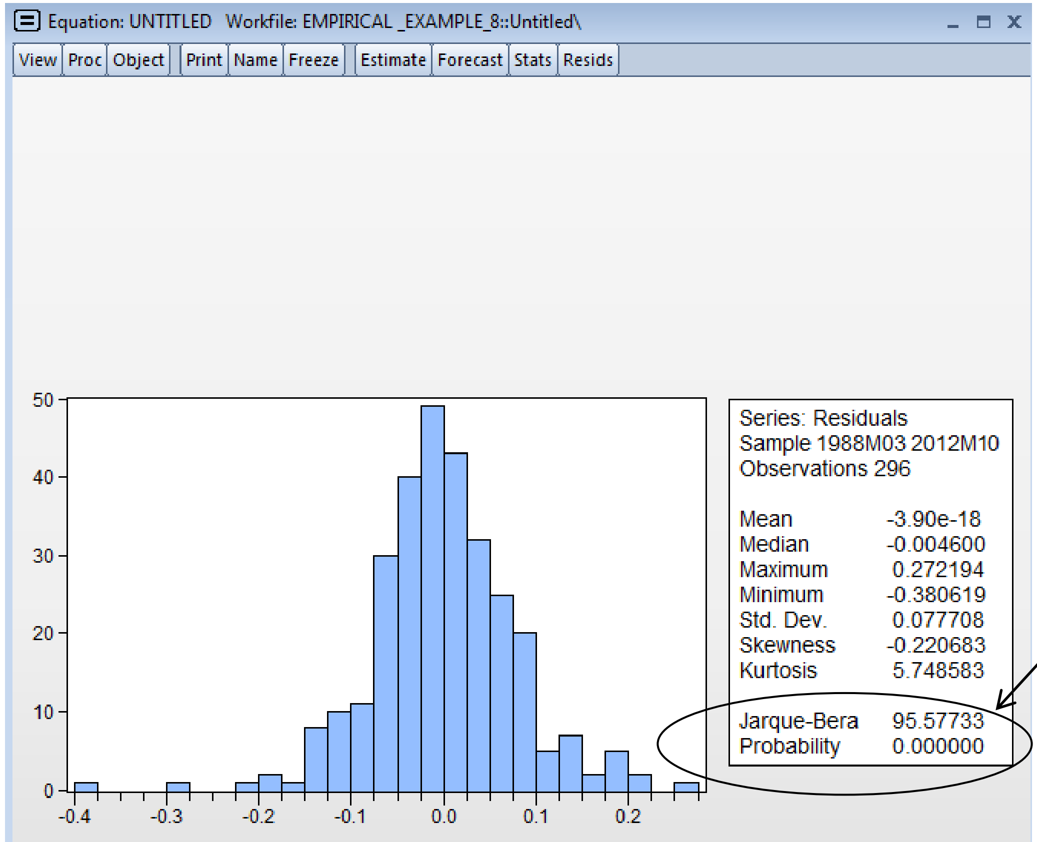
- Εκτιμούμε το υπόδειγμα με τη μέθοδο των ελαχίστων τετραγώνων και σώζουμε τα κατάλοιπα $\hat{u}$
- Χρησιμοποιώντας τα κατάλοιπα εκτιμούμε τους συντελεστές της ασυμμετρίας και της κύρτωσης.

❖ Σημειώστε ότι ότι η ασυμμετρία ορίζεται με βάση την τρίτη κεντρική ροπή ως προς το μέσο, ενώ η κύρτωση ορίζεται με βάση την τέταρτη κεντρική ροπή ως προς τον μέσο. Αν τα κατάλοιπα ακολουθούν την κανονική κατανομή τότε θα έχω $b_1 = 0$ και $b_2 = 3$.

Υποθέσεις:

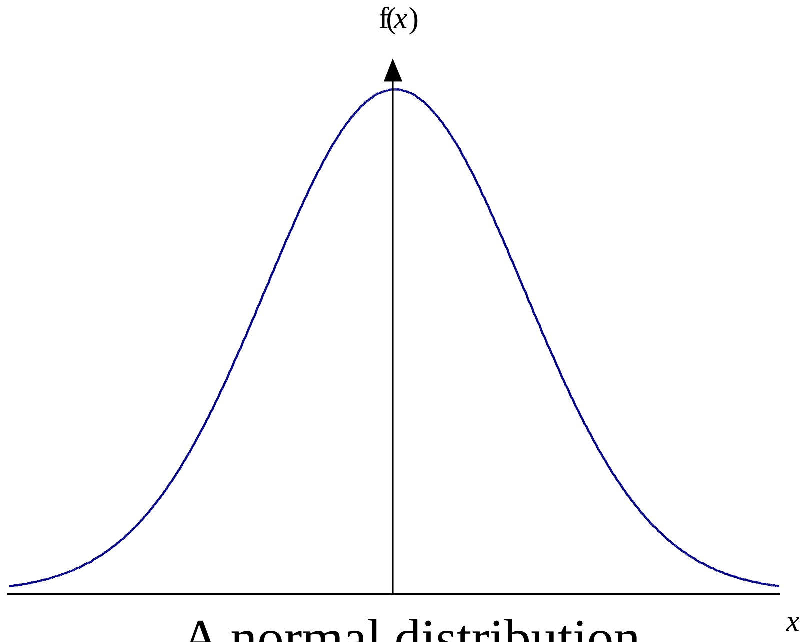
- Μηδενική υπόθεση (null hypothesis) H_0 : Τα κατάλοιπα ακολουθούν την κανονική κατανομή (κατανέμονται κανονικά).
- Εναλλακτική υπόθεση (alternative hypothesis) H_1 : Τα κατάλοιπα δεν ακολουθούν την κανονική κατανομή (δεν κατανέμονται κανονικά).

Αν η τιμή της στατιστικής των Jarque – Bera (JB) είναι μεγαλύτερη από την κριτική τιμή της χ^2κατανομής (τιμή των πινάκων) τότε απορρίπτουμε τη μηδενική υπόθεση οπότε τα κατάλοιπα δεν κατανέμονται κανονικά. Αντίθετα, αν η τιμή της στατιστικής των Jarque – Bera (JB) είναι μικρότερη από την κριτική τιμή της χ^2 κατανομής (τιμή των πινάκων) τότε η μηδενική υπόθεση δεν απορρίπτεται και συνεπώς τα κατάλοιπα κατανέμονται κανονικά.

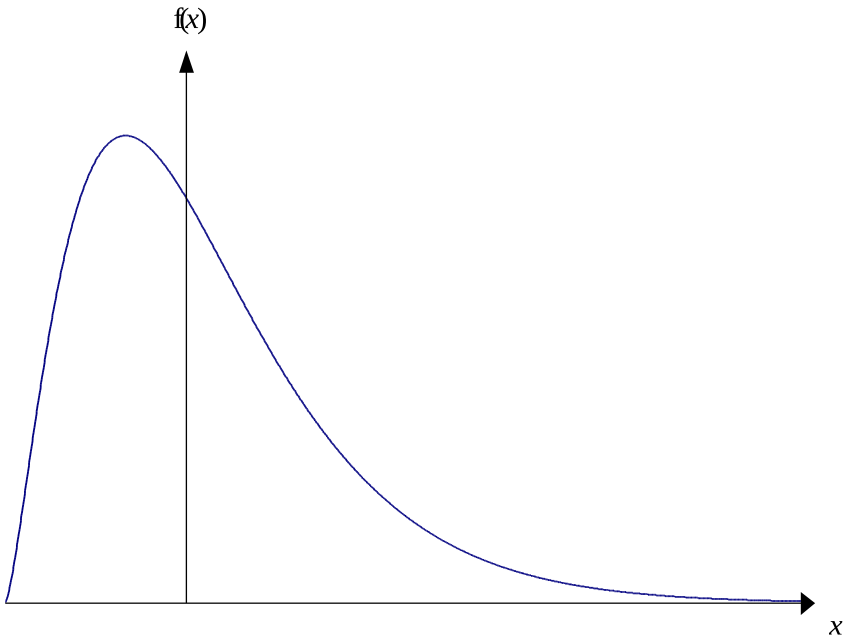


p-value is 0.000, άρα απορρίπτουμε την μηδενική υπόθεση ότι τα κατάλοιπα ακολουθούν την κανονική κατανομή

Κανονική Κατανομή vs. Κατανομή με Ασυμμετρία

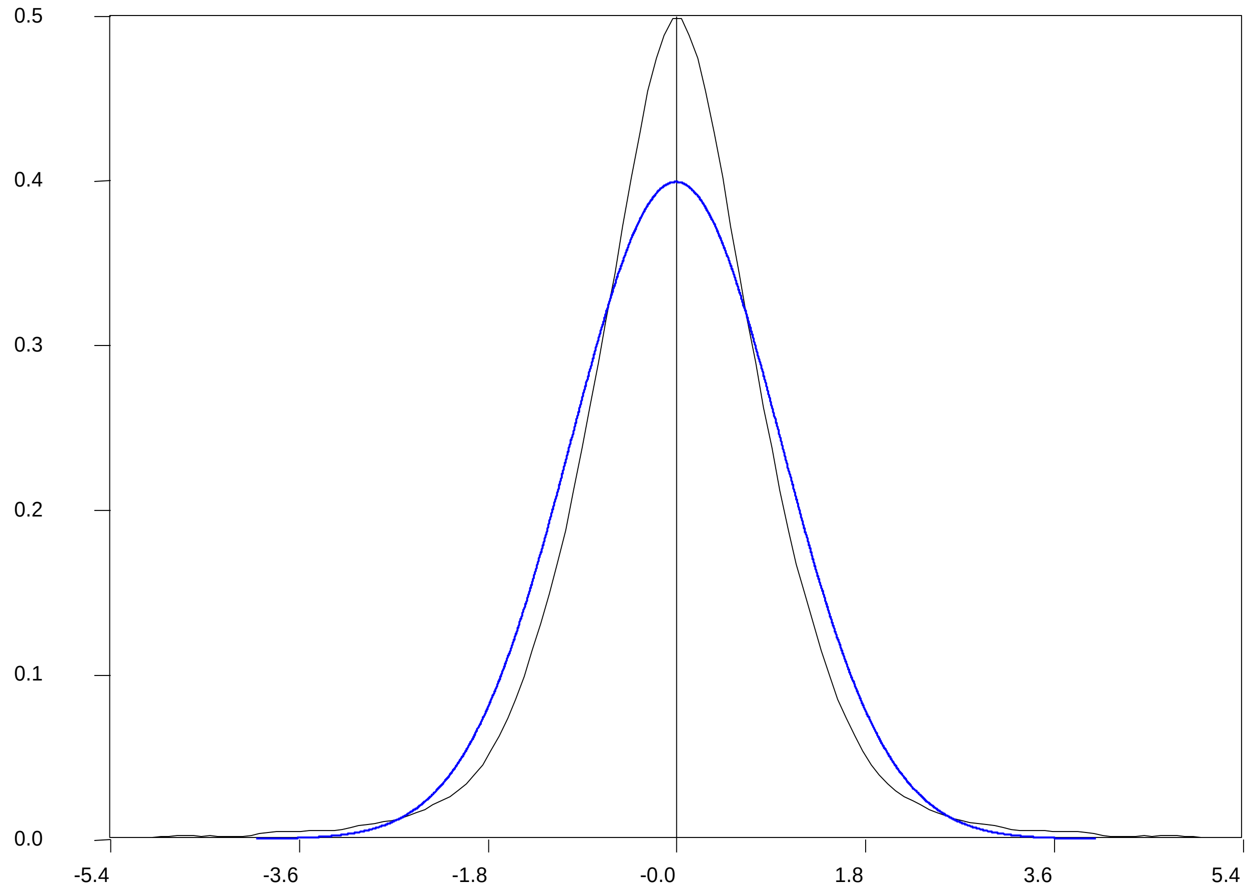


A normal distribution



A skewed distribution

Λεπτόκυρτη (Leptokurtic) vs. Κανονική Κατανομή (Normal Distribution)



Μια έμμεση (implicit) υπόθεση: Οι ανεξάρτητες μεταβλητές δε συσχετίζονται μεταξύ τους

Ξεκινώντας με τον όρο συγγραμικότητα (collinearity), αυτός αποτελεί την ύπαρξη μιας γραμμική σχέσης μεταξύ δύο ανεξάρτητων μεταβλητών. Με τον όρο πολυσυγγραμικότητα αναφερόμαστε στην περίπτωση που δύο ή περισσότερες ανεξάρτητες μεταβλητές, σχετίζονται με υψηλή γραμμική σχέση.

Με άλλα λόγια αυτό το πρόβλημα παρουσιάζεται όταν οι ανεξάρτητες μεταβλητές συσχετίζονται πολύ μεταξύ τους, το οποίο δημιουργεί σοβαρά προβλήματα στην εκτίμηση και την αξιοπιστία των συμπερασμάτων. Η πολυσυγγραμικότητα είναι ένα φαινόμενο που εμφανίζεται κυρίως στις χρονολογικές σειρές, επειδή οι μεταβλητές αυτές μεταβάλλονται διαχρονικά, αλλά και σε διαστρωματικά στοιχεία.

Είδη πολυσυγγραμικότητας

α. Πλήρης πολυσυγγραμικότητα (Perfect multicollinearity):

- Αν η συσχέτιση μεταξύ δύο ανεξάρτητων μεταβλητών είναι ίση με + 1 ή −1, τότε λέμε ότι υπάρχει πλήρης πολυσυγγραμικότητα.
- Σε πολλαπλά γραμμικά υποδείγματα παλινδρόμησης όπου υπάρχουν περισσότερες από δύο ανεξάρτητες μεταβλητές, μπορεί να υπάρχει πλήρης πολυσυγγραμικότητα χωρίς ο συντελεστής συσχέτισης να είναι απαραίτητος ίσος με τη μονάδα.
- *Συνέπειες:* Δεν μπορούν να προσδιοριστούν οι εκτιμητές του πολλαπλού γραμμικού υποδείγματος της παλινδρόμησης με την μέθοδο των ελαχίστων τετραγώνων. Υποθέστε ότι $x_3 = 2x_2$ και ότι και οι δύο μεταβλητές χρησιμοποιούνται στο μοντέλο παλινδρόμησης, δηλαδή $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t$.

Διαισθητικά και οι δύο μεταβλητές μαζί περιέχουν αρκετή πληροφορία για την εκτίμηση μιας παραμέτρου και όχι δύο παραμέτρων.

Τεχνικά (Μαθηματικά) η δυσκολία βρίσκεται στην προσπάθεια να αντιστραφεί η μήτρα $(\mathbf{X'X})$. Όταν υπάρχει πολυσυγραμμικότητα, η μήτρα $(\mathbf{X'X})$ ονομάζεται ιδιάζουσα, και δεν έχει αντίστροφη μήτρα

- Θυμηθείτε ότι για την εκτίμηση των συντελεστών του υποδείγματος αυτού με την μέθοδο των ελαχίστων τετραγώνων, πρέπει να μπορούμε να υπολογίσουμε τον πίνακα $(\mathbf{X'X})^{-1}$ (Αντίστροφος του $\mathbf{X'X}$).

β. Μερική πολυσυγγραμικότητα (Near multicollinearity) :

❖ Συναντάται στην πράξη και αφορά την κατάσταση στην οποία υπάρχει μια σχέση, όχι όμως τέλεια, μεταξύ δύο ή περισσότερων μεταβλητών.

❖ Συνέπειες εάν υπάρχει μερική πολυσυγγραμικότητα αλλά αγνοείται:

- Ο συντελεστής προσδιορισμού θα είναι υψηλός, αλλά οι συντελεστές θα έχουν high standard errors.
- Η παλινδρόμηση γίνεται πολύ ευαίσθητη σε μικρές αλλαγές στις προδιαγραφές του υποδείγματος.
- Τα διαστήματα εμπιστοσύνης για τις παραμέτρους θα είναι πολύ μεγάλα και, συνεπώς, τα significance tests θα μπορούσαν να δώσουν ακατάλληλα συμπεράσματα.

Μέτρηση πολυσυγγραμικότητας

- Ο ευκολότερος τρόπος για να μετρήσετε την *έκταση* της πολυγραμμικότητας είναι απλώς να εξετάσετε τη μήτρα συσχετίσεων μεταξύ των επιμέρους μεταβλητών. π.χ.

ρ	x_2	x_3	x_4
x_2	-	0.2	<u>0.8</u>
x_3	0.2	-	0.3
x_4	<u>0.8</u>	0.3	-

- Ένας δείκτης που χρησιμοποιείται για την διαπίστωση της πολυσυγγραμικότητας αποτελεί ο εκτιμητής διόγκωσης της διακύμανσης (Variance Inflation Factor) ο οποίος δείχνει την έκταση (ταχύτητα) αύξησης της διακύμανσης ενός εκτιμητή όταν οι ανεξάρτητες μεταβλητές συσχετίζονται και ορίζεται ως εξής:

$$VIF = \frac{Var(\hat{\beta}_j)}{Var(\hat{\beta}_j)^*} = \frac{1}{1 - R_j^2}$$

όπου $Var(\hat{\beta}_j)$ είναι η διακύμανση του εκτιμητή $\hat{\beta}_j$,$Var(\hat{\beta}_j)^*$είναι η δυνητική (ιδανική) διακύμανση του εκτιμητή $\hat{\beta}_j$ (αν ίσχυε R_j^2=0) και R_j^2,$\boldsymbol{j} = \mathbf{1}, \ldots, \boldsymbol{K}$ είναι ο συντελεστής προσδιορισμού που προέρχεται από την παλινδρόμηση με εξαρτημένη την ερμηνευτική μεταβλητή $\mathbf{X}_j$ του αρχικού μας υποδείγματος, περιλαμβανομένου της σταθεράς, στις υπόλοιπες *(k -1)* ερμηνευτικές μεταβλητές.

Για παράδειγμα, έαν VIF=4 για μια συγκεκριμένη μεταβλητή, αυτό σημαίνει ότι διακύμανση του εκτιμητή είναι τέσσερις φορές μεγαλύτερη από την διακύμανση του εκτιμητή στην περίπτωση που η μεταβλητή ήταν ανεξάρτητη από τις υπόλοιπες ερμηνευτικές μεταβλητές.

Όσο μεγαλύτερη είναι η τιμή του VIF τόσο μεγαλύτερο είναι το πρόβλημα της πολυσυγγραμικότητας.

Πρακτικός κανόνας: Αν ο VIF<5 ή 10 τότε έχουμε αμελητέα πολυσυγγραμικότητα. Αν ο $VIF >$5 ή 10 έχουμε ισχυρή ένδειξη πολυσυγγραμικότητας.

Λύσεις στο Πρόβλημα της Πολυσυγραμμικότητας

Οι ευκολότεροι τρόποι για να «θεραπεύσετε» τα προβλήματα είναι

- Αποκλείστε από το μοντέλο μία από τις γραμμικές μεταβλητές. Μπορούμε να εξαιρέσουμε από την ανάλυση κάποια ανεξάρτητη μεταβλητή δευτερευούσης σημασίας.

- Μετατρέψτε τις μεταβλητές με υψηλή συσχέτιση σε έναν λόγο (ratio), σε πρώτες διαφορές (first differences) από τα επίπεδά τους (levels) για τα διαχρονικά δεδομένα.

- Βγείτε έξω και συλλέξτε περισσότερα δεδομένα !!!

α.μεγαλύτερη διάρκεια δεδομένων

β. μετάβαση σε υψηλότερη συχνότητα: Εμπειρικές έρευνες έχουν δείξει ότι αν αλλάξουμε τη συχνότητα ενός δείγματος π.χ. από ετήσια σε τριμηνιαία ή μηνιαία, αυτό συμβάλλει στη διόρθωση του προβλήματος της πολυσυγγραμικότητας καθώς αυξάνει το πλήθος των παρατηρήσεων

- Χρησιμοποιείστε τη μέθοδο της ραχοειδούς παλινδρόμησης (ridge regression) Η μέθοδος αυτή συνίσταται στην αντικατάσταση της μήτρας (**X′X**) με μια νέα μήτρα (**X ′X** + **λI**), όπου **λ** ένας πολύ μικρός αριθμός. Δηλαδή προσθέτουμε στις διακυμάνσεις των εκτιμητών μια σταθερά, με αποτέλεσμα να μικραίνει ο συντελεστής συσχέτισης ανάμεσα στις ανεξάρτητες μεταβλητές του πολλαπλού γραμμικού υποδείγματος της παλινδρόμησης. Δοκιμάζονται διάφορες τιμές **λ** ώστε να βρεθεί το κατάλληλο που δίνει λογικά αποτελέσματα.