

Ποσοτικές Μέθοδοι στην Χρηματοοικονομική

5^ο Εξάμηνο
Καθηγητής Χρηματοοικονομικής τμήματος
Οργάνωσης και Διοίκησης Επιχειρήσεων
Νικόλαος Φίλιππας



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

ΚΕΦΑΛΑΙΟ 5

Το Κλασικό Γραμμικό Υπόδειγμα

5.1 Μέθοδος Ελαχίστων Τετραγώνων

- Σε αυτό το σημείο θα εφαρμόσουμε την μέθοδο της ανάλυσης παλινδρόμησης σε μια απλή γραμμική σχέση μεταξύ μιας εξαρτημένης μεταβλητής Y και μιας ανεξάρτητης μεταβλητής X .
- Με τον όρο γραμμική σχέση εννοούμε ότι η μεταβλητή Y επηρεάζεται από την X μέσω της παρακάτω σχέσης:

$$Y = a + \beta X + u \quad (72)$$

- Ο συντελεστής της σειράς που καλείται a , είναι ο σταθερός όρος (intercept) της παλινδρόμησης, ερμηνεύεται δε ως το επίπεδο βάσης της πρόβλεψης αυτής όταν η ανεξάρτητη μεταβλητή είναι μηδέν.
- Ο συντελεστής β (slope coefficient) ερμηνεύεται ως η γωνία της σχέσης μεταξύ της αντίστοιχης εξαρτημένης μεταβλητής και της ανεξάρτητης μεταβλητής.
- Ο συντελεστής β εκφράζει με άλλα λόγια την επίδραση που θα επιφέρει μιας μεταβολή της ανεξάρτητης μεταβλητής κατά μια μονάδα στην εξαρτημένη μεταβλητή.
- Ο όρος u ονομάζεται διαταρακτικός ή στοχαστικός όρος και αντανακλά το γεγονός ότι η μεταβολή της μεταβλητής Y δεν εξηγείται πλήρως από την μεταβολή της ανεξάρτητης μεταβλητής. Υπάρχουν και άλλοι παράγοντες οι οποίοι δεν ερμηνεύονται από το υπόδειγμα. Πρόκειται με άλλα λόγια για ένα μέτρο της ακρίβειας του υποδείγματός μας το οποίο καθορίζει την ακρίβεια της προσέγγισης της πραγματικότητας μέσω των υποθέσεων του υποδείγματός μας. Στην περίπτωση που οι υποθέσεις του υποδείγματός μας προσεγγίζουν την πραγματικότητα σε μεγάλο βαθμό η τιμή του στοχαστικού όρου θα είναι ασήμαντη και κοντά στο μηδέν.

- Προκειμένου να εξετάσει κάποιος τη σχέση μεταξύ της εξαρτημένης και ανεξάρτητης μεταβλητής είναι απαραίτητο να προσδιοριστούν οι τιμές των α , β και u στη σχέση (72).
 - Η μέθοδος εκτίμησης θα πρέπει να είναι τέτοια έτσι ώστε ο εκτιμητής που θα χρησιμοποιήσουμε να είναι ο καλύτερος, γραμμικός και αμερόληπτος από το σύνολο των διαθέσιμων εκτιμητών (Best Linear Unbiased Estimator, BLUE).
 - Ο όρος καλύτερος αναφέρεται στην επιθυμία μας οι εκτιμητές των παραμέτρων να είναι οι πιο αποτελεσματικοί.
 - Δηλαδή, οι διακυμάνσεις γύρω από τους εκτιμητές των παραμέτρων να είναι οι μικρότερες δυνατές και ως εκ τούτου πιο αξιόπιστες.
 - Αυτό επιτυγχάνεται επιλέγοντας εκείνες τις εκτιμήσεις των α και β που ελαχιστοποιούν το άθροισμα των τετραγώνων των καταλοίπων.
-
- Ο όρος γραμμικός υποδηλώνει απλά ότι η σχέση είναι γραμμική.
 - Ο όρος αμερόληπτος ορίζει ότι οι αναμενόμενες τιμές των εκτιμητών των συντελεστών της παλινδρόμησης θα είναι ίσες με τις πραγματικές τιμές των συντελεστών.
 - Εάν λοιπόν τα δεδομένα μας πληρούν τις υποθέσεις που αναφέρονται παρακάτω η πιο διαδεδομένη μέθοδος οι οποίοι μας παρέχει εκτιμητές BLUE είναι η Μέθοδος Ελαχίστων Τετραγώνων (OLS).
 - Το όνομα της προκύπτει από την προσαρμογή μιας ευθείας γραμμής η οποία περνάει όσο το δυνατόν εγγύτερα από τα διαθέσιμα ζεύγη παρατηρήσεων.
 - Ο υπολογισμός της ευθείας γραμμής προκύπτει μέσω της ελαχιστοποίησης των σφαλμάτων (u_i) τα οποία προκύπτουν από τις θεωρητικές τιμές της μεταβλητής Y , $\hat{Y}$ μέσω της ευθείας γραμμής και των πραγματικών τιμών της Y .

5.2 Στατιστικές υποθέσεις Μεθόδου Ελαχίστων Τετραγώνων

- Προκειμένου να εφαρμόσουμε την μέθοδο ελαχίστων τετραγώνων σε ένα σύνολο παρατηρήσεων και να είμαστε σε θέση να ελέγξουμε την σχέση μεταξύ των μεταβλητών πρέπει τα δεδομένα μας να ικανοποιούν τις υποθέσεις του συγκεκριμένου υποδείγματος. Οι υποθέσεις αναφέρονται κυρίως στην συμπεριφορά του στοχαστικού όρου u_i και είναι οι εξής:

1. Η μαθηματική μορφή της σχέσης μεταξύ της πραγματικής εξαρτημένης μεταβλητής Y και της ανεξάρτητης μεταβλητής X , είναι

$$Y = a + \beta X + u \quad (72.1)$$

- Σε αυτό το σημείο είναι χρήσιμο να διακρίνουμε μεταξύ ενός ντετερμινιστικού υποδείγματος και ενός πιθανοθεωρητικού.
- Τα ντετερμινιστικά υποδείγματα είναι εκείνα στα οποία η γνώση των τιμών της X προσδιορίζει την ακριβή γνώση των τιμών της εξαρτημένης μεταβλητής. Για παράδειγμα, το εισόδημα ενός πωλητή καθορίζεται από ένα σταθερό μισθό ίσο με την σταθερά a και ένα επιπλέον bonus το οποίο εκφράζεται ως $\beta\%$ της αξίας των πωλήσεων που πραγματοποιεί. Εάν γνωρίζουμε την αξία των πωλήσεων μπορούμε να καθορίσουμε το ύψος του εισοδήματος του ατόμου. Δεν χρειάζεται σε αυτή την περίπτωση να εφαρμόσουμε ανάλυση παλινδρόμησης.
- Η πιο συνηθισμένη περίπτωση ωστόσο είναι να θεωρούμε την Y ως τυχαία μεταβλητή. Εάν γνωρίζουμε το σύνολο των τιμών της X τότε έχουμε μια συνάρτηση πιθανότητας για την Y . Σε αυτή την περίπτωση το υπόδειγμά μας είναι στοχαστικό και δίνεται από την σχέση (72.1).

2. Ο στοχαστικός όρος u_i ακολουθεί την κανονική κατανομή με μέση τιμή μηδέν και σταθερή διακύμανση.

$$\text{Δηλαδή } E(u_i) = 0 \text{ και } \text{var}(u_i) = \sigma_u^2 \quad (72.2)$$

3. Τα κατάλοιπα είναι μεταξύ τους ασυσχέτιστα μεταξύ διαδοχικών χρονικών περιόδων δηλαδή

$$\text{cov}(u_i, u_j) = 0 \quad \forall i \neq j \quad (72.3)$$

4. Η ανεξάρτητη μεταβλητή είναι ασυσχέτιστη με τον στοχαστικό όρο δηλαδή

$$\text{cov}(u_i, X) = 0 \quad (72.4)$$

- Η δεύτερη υπόθεση υπαινίσσεται ότι παρά το γεγονός ότι υπάρχει ένας κύριος παράγοντας που καθορίζει την τιμή της Y υπάρχουν και άλλοι παράγοντες μικρότερης σημασίας κάποιοι από τους οποίους θα έχουν θετική επίδραση και κάποιοι άλλοι αρνητική.
- Η υπόθεση της σταθερής διακύμανσης του στοχαστικού όρου σημαίνει ότι παρά την μεγάλη ή μικρή τιμή της ανεξάρτητης μεταβλητής X η απόκλιση των τιμών του στοχαστικού όρου είναι σταθερή και σε αυτή την περίπτωση ο στοχαστικός όρος αποκαλείται ομοσκεδαστικός. Σε διαφορετική περίπτωση όταν η απόκλιση των τιμών του στοχαστικού όρου δεν είναι διαχρονικά σταθερή τότε ο στοχαστικός όρος θα αποκαλείται ετεροσκεδαστικός.
- Η τρίτη υπόθεση ότι οι τιμές του στοχαστικού όρου είναι μεταξύ τους ανεξάρτητες σημαίνει ότι δεν μπορώ να χρησιμοποιήσω την τιμή του στοχαστικού όρου τη χρονική στιγμή t για να προβλέψω ποια θα είναι η τιμή του στοχαστικού όρου την επόμενη χρονική περίοδο. Εάν οι τιμές του στοχαστικού όρου δεν είναι ασυσχέτιστες τότε λέμε ότι τα δεδομένα επιδεικνύουν αυτοσυσχέτιση (autocorrelation).

Καθώς η μεταβλητή Y συνδέεται γραμμικά με τον στοχαστικό όρο e συνεπάγεται ότι η εξαρτημένη μεταβλητή Y θα είναι τυχαία μεταβλητή. Για όλες τις τιμές της X η μεταβλητή Y θα ακολουθεί την κανονική κατανομή και η στατιστική κατανομή της Y θα περιγράφεται πλήρως από τον μέσο και την διακύμανση, δηλαδή:

$$E(Y_i) = E(\alpha + \beta_i X_i + u_i) \quad (73)$$

Επειδή τα α και β είναι σταθερές και η μεταβλητή X είναι μη στοχαστική τότε

$$E(Y_i) = \alpha + \beta_i X_i + E(u_i) \quad (74)$$

Επειδή η μέση τιμή του στοχαστικού όρου είναι μηδέν τότε η παραπάνω σχέση θα είναι ίση με:

$$E(Y_i) = \alpha + \beta_i X_i \quad (75)$$

Αναφορικά με την διακύμανση της Y έχουμε να πούμε ότι θα είναι ίση με την διακύμανση του στοχαστικού όρου u η οποία θα είναι ίση με $E(u_i^2)$ καθώς η αναμενόμενη τιμή του στοχαστικού όρου είναι ως γνωστό μηδέν.

$$\text{Τελικά ισχύει } Y_i \approx N(\alpha + \beta_i X_i, \sigma^2) \quad (76)$$

5.3 Προσαρμογή της Γραμμής Παλινδρόμησης στα δεδομένα

- Η μέθοδος ελαχίστων τετραγώνων είναι μια τεχνική η οποία συνίσταται στην προσαρμογή μιας ευθείας γραμμής που προσεγγίζει καλύτερα τα ζεύγη παρατηρήσεων των μεταβλητών X και Y .
- Η συγκεκριμένη μέθοδος περιλαμβάνει την ελαχιστοποίηση του αθροίσματος των τετραγώνων των αποκλίσεων των πραγματικών τιμών της εξαρτημένης μεταβλητής από την προσαρμοσμένη γραμμή. Δηλαδή:

$$\min \sum (Y_i - \hat{Y}_i)^2 \quad (77)$$

- Οι τιμές των συντελεστών α και β οι οποίες ελαχιστοποιούν το άθροισμα των τετραγώνων των καταλοίπων το οποίο αναφέρθηκε παραπάνω προκύπτουν μέσω των παρακάτω σχέσεων:

$$\hat{\beta} = \frac{\text{cov}(X, Y)}{\text{var}_X} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \quad (78) \quad \text{όπου } \text{cov}(X, Y) \text{ είναι η συνδιακύμανση μεταξύ των μεταβλητών.}$$

- Ισοδύναμα ο εκτιμητής του συντελεστή κλίσεως δίνεται από την σχέση:

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

$$\text{και } \hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X} \quad (80)$$

$$\hat{\beta} = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2} \sqrt{\sum y_i^2}} \quad (79) \quad \text{όπου } x_i = X_i - \bar{X} \text{ και } y_i = Y_i - \bar{Y}$$

- Ας δούμε αναλυτικότερα πως καταλήγουμε στις παραπάνω σχέσεις για τους συντελεστές του υποδείγματος.
- Όπως ήδη έχουμε αναφέρει, οι εκτιμητές των παραμέτρων προκύπτουν μέσω της ελαχιστοποίησης του αθροίσματος των τετραγώνων του στοχαστικού όρου u_i .
- Ο στοχαστικός όρος ο οποίος εκφράζει την άγνοιά μας για τους παράγοντες που διαμορφώνουν την εξαρτημένη μεταβλητή ορίζεται ως εξής:

$$u_i = (Y_i - \hat{Y}_i) \quad (81)$$

- όπου Y_i είναι η πραγματική τιμή της εξαρτημένης μεταβλητής $\hat{Y}_i$ και η θεωρητική τιμή που προκύπτει μέσω του υποδείγματος ή με άλλα λόγια η γνώση μας για τους παράγοντες που διαμορφώνουν την εξαρτημένη μεταβλητή του φαινομένου που εξετάζουμε.
- Το άθροισμα τετραγώνων των καταλοίπων υπολογίζεται με την βοήθεια της παρακάτω σχέσης:

$$SS = \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - \hat{\alpha} - \hat{\beta}X_i)^2 \quad (82) \quad \text{όπου } \hat{\alpha} \text{ και } \hat{\beta} \text{ είναι οι παράμετροι προς εκτίμηση.}$$

- Στην συνέχεια, παίρνουμε την πρώτη παράγωγο του αθροίσματος τετραγώνων διαδοχικά ως προς κάθε παράμετρο προς εκτίμηση για να προκύψουν τελικά οι κανονικές εξισώσεις, δηλαδή:

$$\frac{\partial SS}{\partial \hat{\alpha}} = \sum (-2(Y_i - \hat{\alpha} - \hat{\beta}X_i)) \text{ και} \quad (83)$$

$$\frac{\partial SS}{\partial \hat{\beta}} = \sum (-2X_i(Y_i - \hat{\alpha} - \hat{\beta}X_i)) \quad (84)$$

- Το άθροισμα των τετραγώνων ελαχιστοποιείται όταν οι παραπάνω μερικές παράγωγοι ισούται με το μηδέν, όταν δηλαδή ισχύει:

$$\sum (Y_i - \hat{\alpha} - \hat{\beta}X_i) = 0$$

$$\sum X_i(Y_i - \hat{\alpha} - \hat{\beta}X_i) = 0$$

Αυτό επιτυγχάνεται όταν

$$\sum Y_i = n\hat{\alpha} + \hat{\beta}\sum X_i \quad (85)$$

$$\sum X_i Y_i = \hat{\alpha}\sum X_i + \hat{\beta}\sum X_i^2 \quad (86)$$

Τώρα πρέπει να λύσουμε ένα σύστημα εξισώσεων με δυο αγνώστους τα $\hat{\alpha}$ και $\hat{\beta}$. Αρχικά πολλαπλασιάζουμε την πρώτη εξίσωση με $\sum X_i$ και την δεύτερη με n :

$$\sum X_i Y_i = n\hat{\alpha}\sum X_i + \hat{\beta}(\sum X_i)^2$$

$$n\sum X_i Y_i = n\hat{\alpha}\sum X_i + n\hat{\beta}\sum X_i^2$$

Αφαιρώντας την πρώτη εξίσωση από την δεύτερη προκύπτει:

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

$$\text{Επομένως } \hat{\beta} = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2} \quad (87)$$

Η σχέση που μας παρέχει την εκτίμηση του συντελεστή $\hat{\alpha}$ προκύπτει εάν διαιρέσουμε την εξίσωση (85) με n θα έχουμε δηλαδή

$$\frac{\sum Y_i}{n} = \frac{n \hat{\alpha} + \hat{\beta} \sum X_i}{n} \text{ το οποίο είναι ισοδύναμο με}$$

$$\bar{Y} = \hat{\alpha} + \hat{\beta} \bar{X}$$

και τελικά λύνοντας την τελευταία εξίσωση ως προς το $\hat{\alpha}$ λαμβάνουμε

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X} \quad (88)$$

Εφαρμογή: Υπολογισμός συντελεστών παλινδρόμησης

Α. Με βάση τα παρακάτω στοιχεία για την συνολική κατανάλωση και το διαθέσιμο εισόδημα να υπολογίσετε τους συντελεστές α και β της παλινδρόμησης της κατανάλωσης ως προς το διαθέσιμο εισόδημα.

Πίνακας 21
Συνολική Κατανάλωση και Διαθέσιμο Εισόδημα

n	Y_i	X_i	$X_i Y_i$	X_i^2
1	102	114	11628	12996
2	106	118	12508	13924
3	108	126	13608	15876
4	110	130	14300	16900
5	122	136	16592	18496
6	124	140	17360	19600
7	128	148	18944	21904
8	130	156	20280	24336
9	142	160	22720	25600
10	148	164	24272	26896
11	150	170	25500	28900
12	154	178	27412	31684
n=12	$\Sigma Y_i=1524$ $\bar{Y}=127$	$\Sigma X_i=1740$ $\bar{X}=145$	$\Sigma X_i Y_i=225124$	$\Sigma X_i^2=257112$

Απάντηση

Αντικαθιστώντας τις τιμές των μεταβλητών που έχουν υπολογιστεί στον παραπάνω πίνακα στους γνωστούς τύπους προκύπτουν τα εξής αποτελέσματα:

$$\beta = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2} = \frac{(12)(225124) - (1740)(1524)}{(12)(257112) - (1740)^2} = \frac{49728}{57744} = 0,86$$

$$\alpha = \bar{Y} - \beta \bar{X} = 127 - 0,86(145) = 127 - 124,7 = 2,3$$

Επομένως η εξίσωση της γραμμής παλινδρόμησης θα είναι η παρακάτω:

$$\hat{Y}_i = 2,3 + 0,86 X_i$$

Εφαρμογή: Υπολογισμός συντελεστών παλινδρόμησης

Β. Στην συνέχεια, με βάση τα στοιχεία του παραπάνω πίνακα υπολογίστε τον συντελεστή β της γραμμής παλινδρόμησης όταν οι μεταβλητές μετρώνται με τις αποκλίσεις από τον αντίστοιχο μέσο τους.

Πίνακας 22

Συνολική Κατανάλωση και Διαθέσιμο Εισόδημα: Εναλλακτικοί υπολογισμοί

n	Y_i	X_i	y_i	x_i	$x_i y_i$	x_i^2
1	102	114	-25	-31	775	961
2	106	118	-21	-27	567	729
3	108	126	-19	-19	361	361
4	110	130	-17	-15	255	225
5	122	136	-5	-9	45	81
6	124	140	-3	-5	15	25
7	128	148	1	3	3	9
8	130	156	3	11	33	121
9	142	160	15	15	225	225
10	148	164	21	19	399	361
11	150	170	23	25	575	625
12	154	178	27	33	891	1089
			$\sum y_i = 0$	$\sum x_i = 0$	$\sum x_i y_i = 4144$	$\sum x_i^2 = 4812$

Παρατηρούμε ότι στην περίπτωση που οι μεταβλητές εκφράζονται ως αποκλίσεις από τον αντίστοιχο μέσο ισχύει $\sum y_i = \sum x_i = 0$.

$$\beta = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{4144}{4812} = 0,86$$

5.4 Ερμηνεία της εξίσωσης παλινδρόμησης

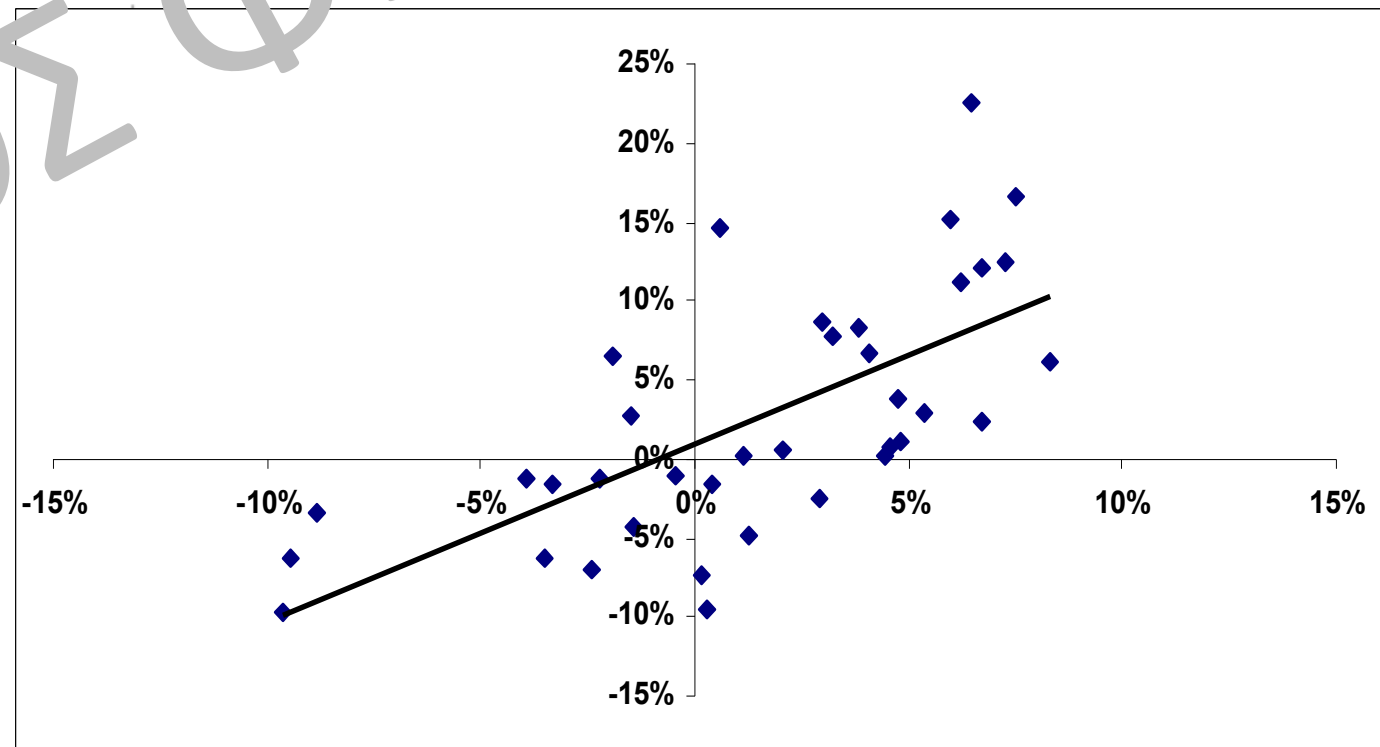
- Ας υποθέσουμε ότι η εφαρμογή της γραμμής παλινδρόμησης για ένα δείγμα παρατηρήσεων των μεταβλητών X και Y μας έδωσε την παρακάτω εξίσωση:

$$\hat{Y} = 0,87 + 1,118X \quad (89)$$

- Το παραπάνω αποτέλεσμα ερμηνεύεται λέγοντας ότι η προβλεφθείσα τιμή της εξαρτημένης μεταβλητής $\hat{Y}$ ισούται με ένα σταθερό ποσό ίσο με 0,87 συν 1,118 για κάθε μονάδα μεταβολής της ανεξάρτητης μεταβλητής X .
- Η σταθερά αντιπροσωπεύει την τιμή της εξαρτημένης μεταβλητής όταν η ανεξάρτητη μεταβλητή λαμβάνει την τιμή 0, παραμένει δηλαδή αμετάβλητη.
- Διαγραμματικά η σταθερά a απεικονίζεται ως η κάθετη απόσταση μεταξύ της αρχής των αξόνων και του σημείου όπου η ευθεία τέμνει τον κάθετο άξονα.

Διάγραμμα 20

Προσαρμογή ευθείας γραμμής στα δεδομένα



5.5 Κατάλοιπα

- Πρωταρχικό ρόλο στην κατασκευή, εκτίμηση και έλεγχο κάθε υποδείγματος διαδραματίζουν τα κατάλοιπα ή αλλιώς σφάλματα τα οποία προσεγγίζουν την αβεβαιότητα μας σχετικά με ένα τμήμα της συμπεριφοράς των εξεταζόμενων μεταβλητών.
- Για παράδειγμα, στο υπόδειγμα της αγοράς, υποθέτουμε ότι ο όρος του σφάλματος λαμβάνει υπόψη την αβεβαιότητα που σχετίζεται με την εξαρτημένη μεταβλητή, δηλαδή τις αποδόσεις της μετοχής, ενώ η ανεξάρτητη μεταβλητή, δηλαδή οι αποδόσεις της αγοράς, θα θεωρούνται εξωγενείς – οι αποδόσεις της αγοράς θεωρούνται μη-τυχαίες ή δεδομένες.
- Το σφάλμα προκύπτει ως η διαφορά μεταξύ της παρατηρούμενης και της προβλεπόμενης (fitted) τιμής της εξαρτημένης μεταβλητής και δίνεται, όπως έχουμε ήδη αναφέρει από την παρακάτω σχέση

$$u_i = (Y_i - \hat{Y}_i) \quad (90)$$

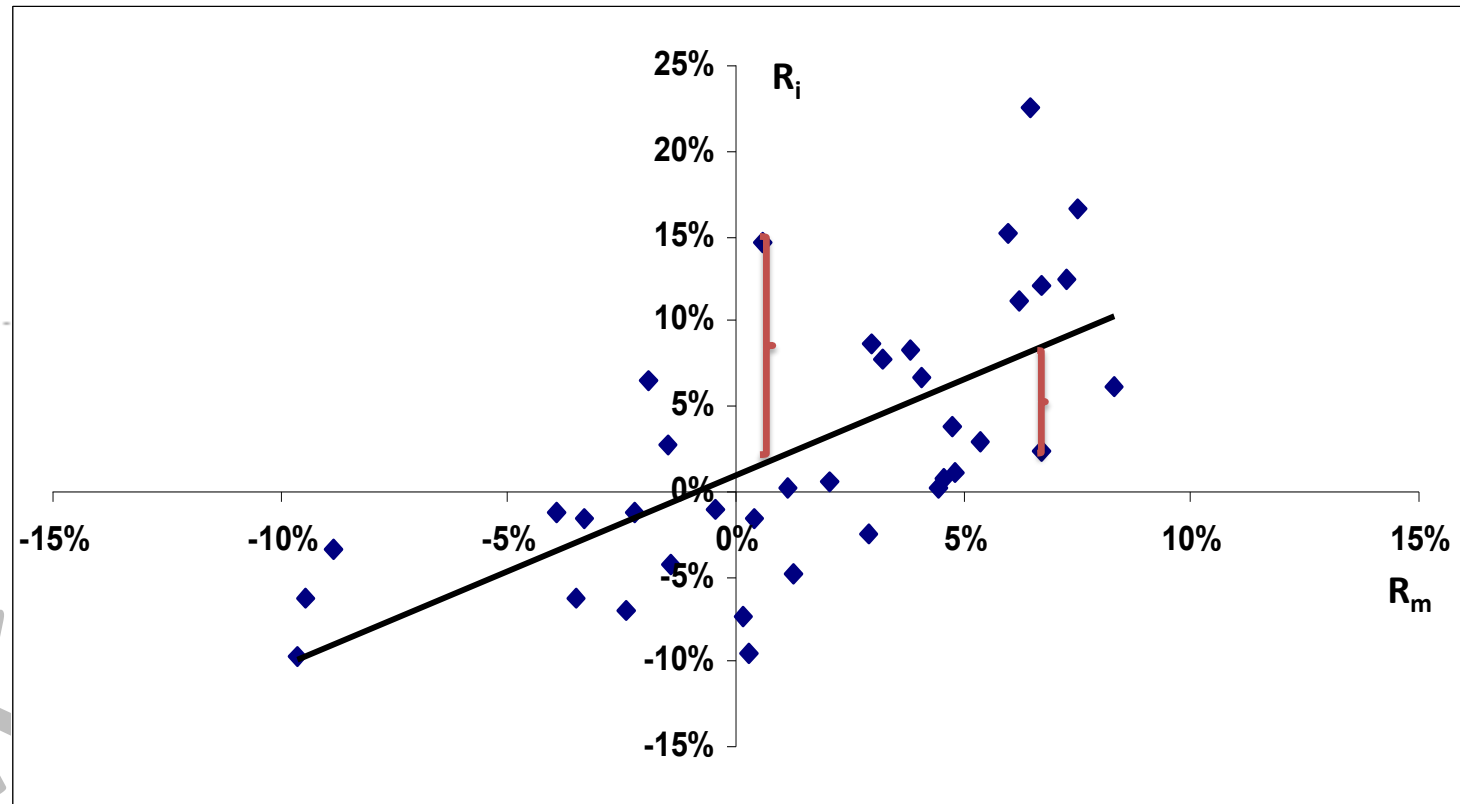
όπου Y_i είναι η πραγματική τιμή της εξαρτημένης μεταβλητής και $\hat{Y}_i$ η θεωρητική τιμή που προκύπτει μέσω του υποδείγματος.

Σε αντιπαράβολή με τη υποτιθέμενο ως πραγματικό υπόδειγμα όπως αποτυπώνεται από τη θεωρητική σχέση π.χ $R_{it} = \alpha + \beta R_{mt} + u_{it}$, το εκτιμημένο υπόδειγμα δηλαδή το $\hat{R}_{it} = \hat{\alpha} + \hat{\beta} R_{mt} + u_{it}$, φέρει καπέλα (^).

- Διαγραμματικά, το σφάλμα απεικονίζεται από την κάθετη απόσταση της παρατηρούμενης τιμής από την προσαρμοσμένη χαρακτηριστική γραμμή όπως φαίνεται στο παρακάτω διάγραμμα.

Διάγραμμα 21

Γραφική απεικόνιση καταλοίπων



- Ο όρος σφάλματος u_i μπορεί να είναι θετικός ή αρνητικός, εξαρτώμενος από το εάν η πραγματοποίηση της απόδοσης της μετοχής βρίσκεται αντίστοιχα πάνω ή κάτω από τη χαρακτηριστική γραμμή του διαγράμματος 21.

- Τώρα το λογικό ερώτημα που προκύπτει είναι: «πως λάβαμε τη χαρακτηριστική γραμμή;».
- Η απάντηση είναι με τη «Μέθοδο των Ελαχίστων Τετραγώνων – Ordinary Least Squares (OLS)».
- Με τη μέθοδο αυτή η χαρακτηριστική γραμμή επιλέγεται με τέτοιο τρόπο ώστε να ελαχιστοποιείται το άθροισμα των τετραγώνων των καταλοίπων u_{it} (Sum of Squares for Error - SSE).

$$\min \{ \text{Το Άθροισμα των Τετραγώνων των Καταλοίπων - SSE} \} = \sum u_{it}^2 = \sum (R_{it} - \hat{R}_{it})^2 \quad (91)$$

- Στους δυο πίνακες που ακολουθούν παρουσιάζονται κατά σειρά οι παρατηρούμενες (actual) αποδόσεις δυο μετοχών X και Y, οι προβλεφθείσες με βάση το υπόδειγμα της Αγοράς και στην τελευταία στήλη απεικονίζονται τα σφάλματα τα οποία προκύπτουν ως η διαφορά των δυο προαναφερθέντων τιμών.
- Από τη γραφική απεικόνιση των καταλοίπων παρατηρούμε ότι τα κατάλοιπα της μετοχής Y παρουσιάζουν υψηλότερη μεταβλητότητα σε σχέση με τα κατάλοιπα της μετοχής X.
- Το γεγονός αυτό θα έχει επιπτώσεις, όπως είναι αναμενόμενο στην προσαρμογή του γραμμικού υποδείγματος στα δεδομένα της μετοχής Y το οποίο με τη σειρά θα εκφραστεί με ένα χαμηλό συντελεστή προσδιορισμού R^2

Actual	Fitted	Residual	Residual Plot
6.36000	9.76752	-3.40752	* .
-0.39000	1.48174	-1.87174	. * .
-1.45000	-3.56752	2.11752	. * .
14.6000	8.19853	6.40147	. . *
-0.53000	-4.10237	3.57237	. *
-5.82000	-3.29284	-2.52716	. * .
-0.94000	-1.05650	0.11650	. * .
1.45000	0.24898	1.20102	. * .
8.94000	1.26656	7.67344	. . *
11.7400	9.08231	2.65769	. * .
6.22000	8.76098	-2.54098	. * .
4.46000	6.64850	-2.18850	. * .
6.53000	6.46527	0.06473	. * .
11.4300	10.0741	1.35587	. * .
-9.45000	-10.9145	1.46451	. * .
-0.09000	1.10574	-1.19574	. * .
5.22000	4.45184	0.76816	. * .
2.36000	4.75440	-2.39440	. * .
7.55000	9.06059	-1.51059	. * .
-0.26000	-1.00776	0.74776	. * .

[illegible]

Πίνακας των Πραγματικών, των Προσαρμοσμένων Αποδόσεων της μετοχής Υ με το Υπόδειγμα της Αγοράς και των Τυπικών Σφαλμάτων του Υποδείγματος

Actual	Fitted	Residual	Residual Plot
-2.06061	3.55657	-5.61718	. * .
-3.09406	0.00194	-3.09600	. * .
-0.12771	-2.16421	2.03650	. * .
-1.66240	2.88347	-4.54588	. * .
6.63199	-2.39366	9.02565	. *
-2.07317	-2.04637	-0.02680	. * .
3.36239	-1.08698	4.44937	. * .
-0.84337	-0.52692	-0.31645	. * .
-12.7582	-0.09037	-12.6678	* . .
-19.2201	3.26261	-22.4827	* . .
-1.55172	3.12476	-4.67649	. * .
2.45184	2.21850	0.23334	. * .
-1.88034	2.13990	-4.02024	. * .
9.75610	3.68811	6.06799	. * .
7.30159	-5.31610	12.6177	. . *
-0.14793	-0.15936	0.01144	. * .
1.92593	1.27612	0.64980	. * .
-8.43023	1.40592	-9.83616	* .
-0.15873	3.25330	-3.41203	. * .

5.6 Έλεγχος στατιστικής σημαντικότητας συντελεστών υποδείγματος

- Για τον έλεγχο των θεωριών και την αξιολόγηση της εμπειρικής σημαντικότητας και της ακρίβειας των επιμέρους παραμέτρων, με σκοπό τη χρησιμοποίησή τους στην άσκηση πολιτικής, είναι αναγκαίο να γίνει έλεγχος της σημαντικότητας των συντελεστών.
- Ο έλεγχος της συνολικής ερμηνευτικής ικανότητας ενός υποδείγματος αποτελεί το πρώτο μόνο βήμα.
- Πριν προχωρήσουμε στον έλεγχο της στατιστικής σημαντικότητας των εκτιμηθέντων παραμέτρων χρειάζεται να υπολογίσουμε τα τυπικά σφάλματα των εκτιμητών.
- Τα τυπικά σφάλματα είναι η τυπική απόκλιση των εκτιμηθέντων συντελεστών και μετρούν την στατιστική αξιοπιστία των συντελεστών παλινδρόμησης (όσο μικρότερα τα τυπικά σφάλματα, τόσο πιο αξιόπιστοι είναι οι συντελεστές παλινδρόμησης) .
- Το τυπικό σφάλμα της σταθεράς της παλινδρόμησης είναι ίσο με

$$SE_{\hat{\alpha}} = \sqrt{\frac{\sum (Y_i - \hat{Y})^2}{n-2} \frac{\sum X_i^2}{n \sum (X_i - \bar{X})^2}} \quad (92)$$

- ενώ το τυπικό σφάλμα της εκτίμησης του συντελεστή κλίσεως υπολογίζεται μέσω της σχέσης

$$SE_{\hat{\beta}} = \sqrt{\frac{\sum (Y_i - \hat{Y})^2}{n-2} \frac{1}{\sum (X_i - \bar{X})^2}} \quad (93)$$

όπου $\frac{\sum (Y_i - \hat{Y})^2}{n-2}$ είναι το τυπικό σφάλμα της παλινδρόμησης και το οποίο είναι ένα συνοπτικό μέτρο του μεγέθους των σφαλμάτων πρόβλεψης. Το τυπικό σφάλμα της παλινδρόμησης λοιπόν υπολογίζεται με την βοήθεια της παρακάτω σχέσης;

$$SE = \sqrt{\frac{\sum_{i=1}^n u_i^2}{n-2}} \quad (94)$$

όπου $\sum_{i=1}^n u_i^2$ είναι το άθροισμα των τετραγώνων των καταλοίπων και
n ο αριθμός των παρατηρήσεων

- Αξίζει να σημειωθεί ότι στον παραπάνω τύπο ο παρανομαστής είναι $n-2$ και εκφράζει τους βαθμούς ελευθερίας από την εκτίμηση του υποδείγματος καθώς απαιτείται η εκτίμηση δυο παραμέτρων ($\hat{\alpha}, \hat{\beta}$) για τον υπολογισμό των καταλοίπων.
- Όταν πρόκειται για ένα χαρτοφυλάκιο ή μετοχή το τυπικό σφάλμα της παλινδρόμησης εκφράζει τον μη συστηματικό κίνδυνο του περιουσιακού στοιχείου και επιθυμούμε να είναι όσο το δυνατόν μικρότερο.
- Γνωρίζουμε ότι για δεδομένα τα οποία προέρχονται από μια κανονική κατανομή η διαφορά μεταξύ μιας τυχαίας μεταβλητής και του μέσου της διαιρεμένη με μια εκτίμηση της τυπικής απόκλισης ακολουθεί μια κατανομή t-student.
- Έτσι εάν διαιρέσουμε τις εκτιμήσεις των συντελεστών της παλινδρόμησης με τα αντίστοιχα τυπικά τους σφάλματα θα προκύψει μια ποσότητα η οποία ακολουθεί την κατανομή t-student με $n-2$ βαθμούς ελευθερίας.
- Με αυτή την πληροφορία είμαστε σε θέση να κατασκευάσουμε διαστήματα εμπιστοσύνης γύρω από τις σημειακές εκτιμήσεις των συντελεστών.

Έστω ότι $1-c$ συμβολίζει το επίπεδο εμπιστοσύνης όπου c είναι η πιθανότητα να βρίσκεται η τυχαία μεταβλητή στα άκρα της κατανομής. Τότε θα ισχύει:

$$P(-t_{n-2,c/2} \leq \frac{\hat{\alpha} - \alpha}{SE_{\alpha}} \leq t_{n-2,c/2}) = 1 - c \quad (95)$$

$$P(-t_{n-2,c/2} \leq \frac{\hat{\beta} - \beta}{SE_{\beta}} \leq t_{n-2,c/2}) = 1 - c \quad (96)$$

Από τις παραπάνω σχέσεις προκύπτουν τα διαστήματα εμπιστοσύνης των συντελεστών της παλινδρόμησης ως εξής:

$$P(\hat{\alpha} - SE_{\alpha} t_{n-2,c/2} \leq \alpha \leq \hat{\alpha} + SE_{\alpha} t_{n-2,c/2}) = 1 - c \quad (97)$$

$$P(\hat{\beta} - SE_{\beta} t_{n-2,c/2} \leq \beta \leq \hat{\beta} + SE_{\beta} t_{n-2,c/2}) = 1 - c \quad (98)$$

- Έτσι υπάρχει πιθανότητα ίση με $1-c$ ότι η πραγματική τιμή των συντελεστών περιέχεται στο προκαθορισμένο διάστημα. Εάν το συγκεκριμένο διάστημα περιλαμβάνει την τιμή μηδέν τότε οι συντελεστές δεν είναι στατιστικά διαφορετικοί του μηδενός.

- Σύμφωνα με τη θεωρία της παλινδρόμησης λοιπόν με πιθανότητα 66% οι πραγματικοί συντελεστές παλινδρόμησης θα βρίσκονται σ'ένα εύρος ενός τυπικού σφάλματος των αναφερόμενων συντελεστών

$$(\hat{\beta} - 1 * SE_{\beta} < \beta < \hat{\beta} + 1 * SE_{\beta})$$

- και με 95% πιθανότητα να βρίσκονται σ'ένα εύρος δυο τυπικών σφαλμάτων

$$(\hat{\beta} - 2 * SE_{\beta} < \beta < \hat{\beta} + 2 * SE_{\beta}).$$

- Στην όλη διαδικασία ελέγχου της σημαντικότητας ενός συντελεστή, αποφασιστικό ρόλο διαδραματίζει ο τρόπος που θα διατυπωθούν η ελεγχόμενη και η εναλλακτική υπόθεση.
- Αναφορικά με τη σημαντικότητα ενός συντελεστή, δυο είναι τα ενδεχόμενα:
 - (i) να μην είναι στατιστικά σημαντικός και
 - (ii) να είναι στατιστικά σημαντικός.
- Ο όρος «στατιστικά» υποδηλώνει ότι η σχετική απόφαση θα ληφθεί σε όρους πιθανότητας, όπως είναι σωστό να γίνεται στις περιπτώσεις των στοχαστικών υποδειγμάτων.
- Τα παραπάνω θα μπορούσαν να διατυπωθούν συμβολικά ως εξής:

$$H_0 : \hat{\beta}_t = 0$$

$$H_1 : \hat{\beta}_t \neq 0$$

(99)

- Όπου $\hat{\beta}_j$ η παράμετρος της οποίας επιθυμούμε να ελέγξουμε τη σημαντικότητα. Για τον έλεγχο αυτό χρησιμοποιείται είτε το κριτήριο της τυποποιημένης κανονικής κατανομής, είτε το κριτήριο t του student.
- Όμως, επειδή το κριτήριο της τυποποιημένης κανονικής κατανομής προϋποθέτει την γνώση της τυπικής απόκλισης που ως γνωστό υπολογίζεται ως η θετική τετραγωνική ρίζα της διακύμανσης η πρακτική εφαρμογή του κριτηρίου αυτού αντιμετωπίζει προβλήματα όταν το δείγμα είναι μικρό.
- Γι αυτό στην πράξη, κατά κανόνα, χρησιμοποιείται το κριτήριο t του student.
- Ειδικότερα η τιμή του t υπολογίζεται από τον τύπο:

$$t_j = \frac{\hat{\beta}_j}{S_{\hat{\beta}_j}} \quad (100)$$

$\hat{\beta}_j$ η εκτίμηση της παραμέτρου β_j

$S_{\hat{\beta}_j}$ = τυπικό σφάλμα του συντελεστή $\hat{\beta}_j$ και

m = οι βαθμοί ελευθερίας

- Σημειώνεται ότι το τυπικό σφάλμα αποτελεί εκτίμηση της τυπικής απόκλισης, η οποία είναι θεωρητική έννοια.
- Αν η τιμή του t_j είναι μεγαλύτερη από την αντίστοιχη κριτική τιμή που προκύπτει από τους πίνακες για ορισμένο επίπεδο σημαντικότητας, τότε λέμε ότι η παράμετρος β_j είναι στατιστικά διαφορετική από το μηδέν. Στην αντίθετη περίπτωση θα συναχθεί αντίθετο συμπέρασμα.

- Αξίζει να τονισθεί ότι κατά την διατύπωση της μηδενικής υπόθεσης δεν είναι απαραίτητο το να τίθεται πάντοτε ίσο με το μηδέν.
- Η παράμετρος είναι δυνατό να λάβει οποιαδήποτε τιμή.
- Όμως, είναι αναγκαίο η τιμή αυτή να είναι συγκεκριμένη. Για παράδειγμα, μπορεί να τεθεί:

$$H_0 : \hat{\beta}_j = 1$$

$$H_1 : \hat{\beta}_j \neq 1 \quad (101)$$

- Στην περίπτωση αυτή ο τύπος υπολογισμού του κριτηρίου t statistic γράφεται:

$$t_m = \frac{\hat{\beta}_j - 1}{S_{\hat{\beta}_j}} \quad (102)$$

- Εξάλλου η διατύπωση της (101) υποδηλώνει ότι το β μπορεί να λάβει είτε θετική είτε αρνητική τιμή. Ωστόσο, σε πάρα πολλά προβλήματα της πραγματικής ζωής το πρόσημο του β_j είναι προσδιορισμένο (θετικό ή αρνητικό).
- Συνήθως ο έλεγχος στατιστικής σημαντικότητας των συντελεστών πραγματοποιείται σε επίπεδο σημαντικότητας 95 ή 99%. Αυτό σημαίνει ότι υπάρχει 95 ή 99% πιθανότητα αντίστοιχα οι πραγματικές τιμές των συντελεστών α και β δεν οφείλονται στην τύχη.

5.7 Έλεγχοι καταλληλότητας του υποδείγματος

- Τα κριτήρια που χρησιμοποιούνται συνήθως για τον έλεγχο της συνολικής ερμηνευτικής ικανότητας ενός υποδείγματος, μπορούν να διακριθούν σε δυο κατηγορίες:
 - α) στα περιγραφικά και
 - β) στα αναλυτικά ή στατιστικά.
- Στην πρώτη κατηγορία ανήκουν η ανερμήνευτη διακύμανση (residual variance) S^2 , και ο συντελεστής προσδιορισμού (coefficient of determination) R^2 . Το πρώτο αποτελεί μέτρο του βαθμού της συνολικής ερμηνευτικής ικανότητας του υποδείγματος σε απόλυτους όρους, ενώ το δεύτερο μετράει το ίδιο πράγμα σε σχετικούς όρους. Έτσι, η τιμή της ανερμήνευτης διακύμανσης εξαρτάται από τις μονάδες μέτρησης των μεταβλητών, γεγονός που εμποδίζει τις απευθείας συγκρίσεις μεταξύ εναλλακτικών υποδειγμάτων. Οι συγκρίσεις αυτές θα έχουν νόημα μόνον εφόσον οι μεταβλητές στα υπό σύγκριση υποδείγματα έχουν εκφραστεί στις ίδιες μονάδες. Σε αντίθεση με το S^2 ο συντελεστής προσδιορισμού είναι απαλλαγμένος από τις μονάδες στις οποίες μετρώνται οι μεταβλητές.
- Τα κριτήρια S^2 και R^2 χαρακτηρίζονται ως περιγραφικά, διότι οι αποφάσεις περί της ερμηνευτικής ικανότητας του υποδείγματος λαμβάνονται ανάλογα με το μέγεθος της αριθμητικής τους τιμής. Όσο πιο μικρή είναι η τιμή του S^2 ή όσο πιο μεγάλη είναι η τιμή του R^2 , τόσο καλύτερη θεωρείται η προσαρμοστική ικανότητα του υποδείγματος στα δεδομένα. Με άλλα λόγια, οι σχετικές αποφάσεις δεν λαμβάνονται σε όρους πιθανότητας όπως απαιτεί η στοχαστική φύση των οικονομετρικών υποδειγμάτων. Αποφάσεις σε όρους πιθανότητας λαμβάνονται μόνο με τα αναλυτικά κριτήρια, όπως είναι το κριτήριο F και το κριτήριο ελέγχου της προβλεπτικής ικανότητας του υποδείγματος.

5.7.1 Η ανερμήνευτη διακύμανση

- Η ανερμήνευτη διακύμανση αποτελεί μέτρο της διασποράς των παρατηρούμενων τιμών της εξαρτημένης μεταβλητής γύρω από την εκτιμώμενη γραμμή παλινδρόμησης και υπολογίζεται από τον τύπο:

$$S_u^2 = \frac{\sum u_i^2}{n - \lambda} = \frac{\sum (Y_i - \hat{Y}_i)^2}{n - \lambda} \quad (103)$$

- όπου τα u_i παριστάνουν τα ανερμήνευτα κατάλοιπα της μεθόδου ελαχίστων τετραγώνων, n = το μέγεθος του δείγματος, λ = τον αριθμό των προς εκτίμηση παραμέτρων και $n - \lambda$ = τους βαθμούς ελευθερίας.
- Εξάλλου τα Y_i ΚΑΙ $\hat{Y}_i$ παριστάνουν τις πραγματικές και τις θεωρητικές (εκτιμημένες) τιμές της εξαρτημένης μεταβλητής, αντίστοιχα.
- Η προσαρμοστικότητα του υποδείγματος είναι αντιστρόφως ανάλογη με το μέγεθος της τιμής του S_u^2 .
- Και τούτο διότι, όπως συνάγεται από την σχέση (103) η τιμή τούτου εξαρτάται από το ανερμήνευτο μέρος (τα κατάλοιπα u_i) της συνολικής μεταβλητικότητας των τιμών της εξαρτημένης μεταβλητής.
- Όσο μικρότερο είναι το μέρος αυτό, τόσο ικανοποιητικότερη είναι η ερμηνευτική ικανότητα του υποδείγματος.
- Αξίζει να σημειωθεί ότι η διαίρεση του αθροίσματος των καταλοίπων με τους βαθμούς ελευθερίας και όχι με το μέγεθος του δείγματος, είναι αναγκαία στα μικρά δείγματα (δηλαδή $n < 30$).
- Οι βαθμοί ελευθερίας είναι $(n - \lambda)$ διότι από το σύνολο των n καταλοίπων (u_i) μόνο τα $(n - \lambda)$ παίρνουν ανεξάρτητες τιμές.

5.7.2 Ο συντελεστής προσδιορισμού

- Όπως είναι γνωστό, ο συντελεστής προσδιορισμού μετράει το ποσοστό της συνολικής μεταβλητικότητας στις τιμές της εξαρτημένης μεταβλητής που ερμηνεύεται από το υπόδειγμα και υπολογίζεται από τον τύπο:

$$R^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} = \frac{\text{Ερμηνευόμενο μέρος}}{\text{Συνολική μεταβλητότητα τη } Y} \quad (104)$$

- Αξίζει να τονιστεί ότι ο υπολογισμός του R^2 από τον παραπάνω τύπο είναι σωστός μόνον εφόσον στο υπόδειγμα υπάρχει σταθερός όρος.
- Η συνολική διακύμανση της εξαρτημένης μεταβλητής εκφράζεται με το συνολικό άθροισμα τετραγώνων (sum of squares total):

$$SST = \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n y_i^2 \quad (105)$$

- Το άθροισμα τετραγώνων της παλινδρόμησης (sum of squares of regression) ορίζεται ως:

$$SSR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 = \sum_{i=1}^n \hat{y}_i^2 \quad (106)$$

το οποίο εκφράζει το τμήμα της ερμηνευθείσας διακύμανσης της εξαρτημένης μεταβλητής μέσω της παλινδρόμησης.

- Το άθροισμα τετραγώνων των καταλοίπων (sum of squares of errors) προκύπτει ως η απόκλιση των πραγματικών τιμών της εξαρτημένης μεταβλητής και των αντίστοιχων θεωρητικών $\hat{Y}$.
- Μια μικρή τιμή του συγκεκριμένου αθροίσματος υποδηλώνει πως η προσαρμογή του υποδείγματος στην πραγματικότητα είναι ικανοποιητική.

$$SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n u_i^2 \quad (107)$$

- Ξεκινώντας από την βασική εξίσωση των επιμέρους αθροισμάτων η οποία αποδεικνύεται σχετικά εύκολα μπορούμε να καταλήξουμε στον τύπο υπολογισμού του συντελεστή προσδιορισμού.

$$SST = SSR + SSE \quad (108)$$

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (109)$$

$$\sum_{i=1}^n y_i^2 = \sum_{i=1}^n \hat{y}_i^2 + \sum_{i=1}^n u_i^2 \quad (110)$$

Διαιρώντας και τα δυο μέλη της εξίσωσης (110) με $\sum_{i=1}^n y_i^2$ προκύπτει

$$\frac{\sum_{i=1}^n y_i^2}{\sum_{i=1}^n y_i^2} = \frac{\sum_{i=1}^n \hat{y}_i^2}{\sum_{i=1}^n y_i^2} + \frac{\sum_{i=1}^n u_i^2}{\sum_{i=1}^n y_i^2} \quad (111)$$

$$1 = \frac{\sum_{i=1}^n \hat{y}_i^2}{\sum_{i=1}^n y_i^2} + \frac{\sum_{i=1}^n u_i^2}{\sum_{i=1}^n y_i^2} \quad (112)$$

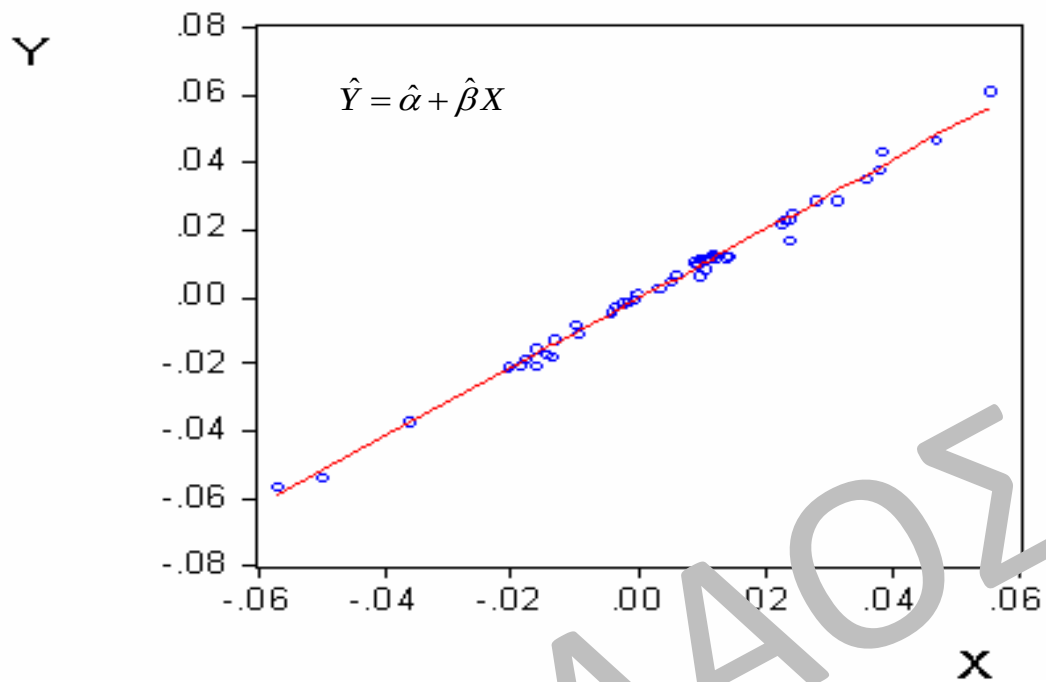
Επομένως $R^2 = \frac{\sum_{i=1}^n \hat{y}_i^2}{\sum_{i=1}^n y_i^2} = 1 - \frac{\sum_{i=1}^n u_i^2}{\sum_{i=1}^n y_i^2} \quad (113)$

- Το R^2 λοιπόν είναι ένα μέτρο καλής προσαρμογής του υποδείγματος και τα όρια τιμών του βρίσκονται στο διάστημα από 0 έως 1 .
- Εάν η παλινδρόμηση αναπαραγάγει ικανοποιητικά την πραγματικότητα τότε όπως φαίνεται από την σχέση (113) η τιμή του στοχαστικού όρου u_i θα τείνει στο μηδέν και η τιμή του R^2 τείνει στο 1 .
- Στην αντίθετη περίπτωση η τιμή του R^2 τείνει στο μηδέν .
- Πρέπει να σημειωθεί ότι είναι δυνατόν το R^2 να είναι αρνητικό εάν η παλινδρόμηση δεν περιλαμβάνει έναν σταθερό όρο ή εάν χρησιμοποιείται η μέθοδος ελαχίστων τετραγώνων σε δυο στάδια (2SLS).
- Ο συντελεστής προσδιορισμού στην χρηματοοικονομική επιστήμη έχει μια πολύ σημαντική ερμηνεία.
- Εκφράζει το ποσοστό των μεταβολών της εξαρτημένης μεταβλητής που οφείλονται σε μεταβολές της ανεξάρτητης μεταβλητής.
- Με άλλα λόγια, εκφράζει το ποσοστό του συνολικού κινδύνου μιας μετοχής ή χαρτοφυλακίου που είναι συστηματικός ενώ το υπόλοιπο $1-R^2$ εκφράζει το μη συστηματικό κίνδυνο του περιουσιακού στοιχείου.
- Για παράδειγμα, εάν παλινδρομήσουμε τις αποδόσεις ενός μετοχικού χαρτοφυλακίου με τις αποδόσεις ενός κατάλληλου δείκτη αναφοράς και προκύψει ένας συντελεστής προσδιορισμού 85% σημαίνει ότι το 85% του κινδύνου είναι συστηματικός, οφείλεται δηλαδή σε όλους εκείνους τους παράγοντες που επηρεάζουν συνολικά τις αποδόσεις των περιουσιακών στοιχείων και το υπόλοιπο 15% οφείλεται σε μη συστηματικούς παράγοντες.
- Στην συνέχεια, παρουσιάζονται δυο διαγράμματα τα οποία απεικονίζουν μια περίπτωση καλής προσαρμογής της ευθείας παλινδρόμησης στα δεδομένα και μια περίπτωση όπου η προσαρμογή της ευθείας γραμμής δεν εξηγεί επαρκώς της συμπεριφορά της εξαρτημένη μεταβλητή.

Διάγραμμα 22

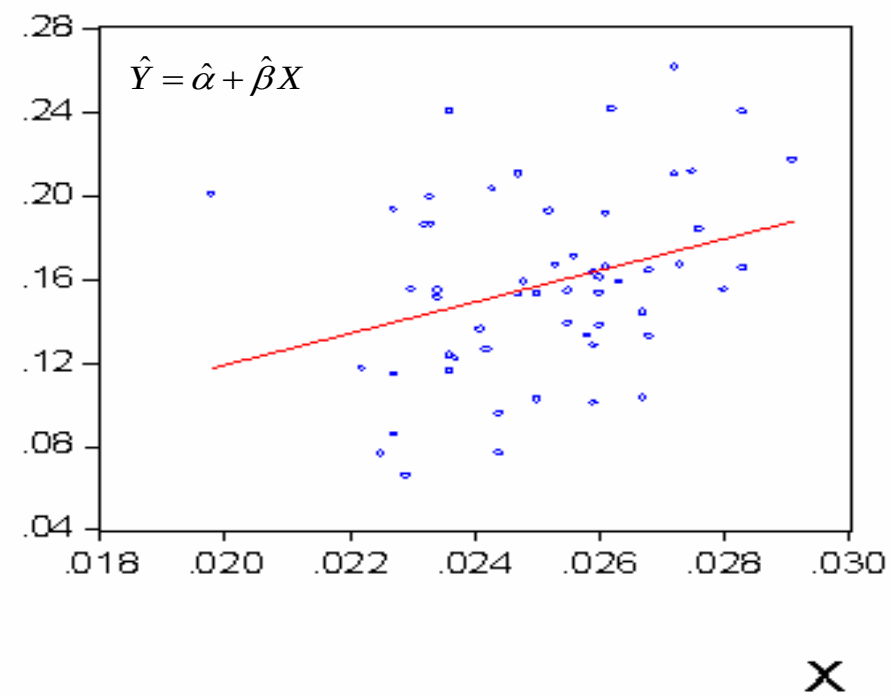
Προσαρμογή υποδείγματος στα δεδομένα

A.



3.

Y



5.7.3 Το κριτήριο F

- Το κριτήριο F δίνει απάντηση στο ερώτημα σχετικά με την συνολική ερμηνευτική ικανότητα του υποδείγματος σε όρους πιθανότητας.
- Αυτό προκύπτει ως ο λόγος της ερμηνευόμενης προς την ανερμήνευτη διακύμανση της εξαρτημένης μεταβλητής διαιρούμενου με τους αντίστοιχους βαθμούς ελευθερίας.
- Στην περίπτωση της παλινδρόμησης ελαχίστων τετραγώνων το στατιστικό κριτήριο δίνεται από την παρακάτω σχέση:

$$F = \frac{\frac{R^2}{k-1}}{\frac{1-R^2}{n-k}} \quad (114)$$

όπου

R^2 είναι ο συντελεστής προσδιορισμού του υποδείγματος

k είναι ο αριθμός των παραμέτρων προς εκτίμηση και

n είναι ο αριθμός των παρατηρήσεων

- Κάτω από την μηδενική υπόθεση της κανονικής κατανομής των καταλοίπων το στατιστικό κριτήριο ακολουθεί την κατανομή F με $k-1$ βαθμούς ελευθερίας του αριθμητή και $n-k$ βαθμούς ελευθερίας του παρανομαστή.
 - Για την εξαγωγή αξιόπιστων συμπερασμάτων, προϋπόθεση είναι η σωστή διατύπωση της υπόθεσης που πρόκειται να ελεγχθεί και η ισχύ των προϋποθέσεων κάτω από τις οποίες ισχύει το κριτήριο.
 - Σε κάθε ελεγχόμενη υπόθεση H_0 υπάρχει και η εναλλακτική H_1 .
 - Πρόκειται για την πρόταση η οποία θα γίνει αποδεκτή σε περίπτωση που απορριφθεί η ελεγχόμενη υπόθεση.
 - Η εναλλακτική υπόθεση εκφράζει το αποτέλεσμα που αναμένεται να προκύψει από τον σχετικό έλεγχο.
 - Η απόφαση για την εκτίμηση ενός υποδείγματος λαμβάνεται κάτω από την προσδοκία ότι τούτο ερμηνεύει το φαινόμενο το οποίο ερευνούμε.
 - Στην αντίθετη περίπτωση, η όλη προσπάθεια δεν θα είχε κανένα νόημα.
-
- Σύμφωνα με τα παραπάνω, η διατύπωση της ελεγχόμενης υπόθεσης H_0 πρέπει να γίνει κατά τρόπο που να υποδηλώνει ότι οι ερμηνευτικές μεταβλητές δεν ασκούν καμία επίδραση στην διαμόρφωση της συμπεριφοράς της εξαρτημένης μεταβλητής.
 - Έτσι, η απόρριψή της θα οδηγούσε στην παραδοχή της εναλλακτικής υπόθεσης H_1 , γεγονός που θα σήμαινε ότι οι διακυμάνσεις της εξαρτημένης μεταβλητής ερμηνεύονται πράγματι από αντίστοιχες διακυμάνσεις των ανεξάρτητων μεταβλητών.
 - Αυτό, με τη σειρά του, θα σήμαινε ότι το υπόδειγμα ερμηνεύει ικανοποιητικά το υπό διερεύνηση φαινόμενο.

5.8 Προβλέψεις μέσω του κλασικού γραμμικού υποδείγματος

- Έχει ήδη αναφερθεί ότι ένας από τους βασικούς σκοπούς εκτίμησης των οικονομετρικών υποδειγμάτων είναι η πρόβλεψη των τιμών της εξαρτημένης μεταβλητής σε περιοχές εντός και εκτός της περιοχής αναφοράς του δείγματος.
- Στην περίπτωση των χρονολογικών σειρών, αυτό σημαίνει ότι η πρόβλεψη αναφέρεται σε μελλοντικές περιόδους.
- Βεβαίως, οι προβλέψεις είναι δυνατόν να αφορούν και το παρελθόν.
- Εξάλλου, στην περίπτωση των διακλαδικών ή διαστρωματικών δεδομένων η πρόβλεψη αφορά άλλους κλάδους, άλλα στρώματα ή άλλες περιοχές.
- Για να είναι χρήσιμη στην λήψη αποφάσεων και την άσκηση πολιτικής, η πρόβλεψη πρέπει να είναι ποσοτικής φύσεως και επαληθεύσιμη.
- Αυτό σημαίνει ότι η ακρίβειά της θα είναι δυνατό να επιβεβαιωθεί ή να διαψευσθεί από τα πραγματικά γεγονότα.
- Οι προβλέψεις, όπως φαίνεται και στο παρακάτω διάγραμμα, διακρίνονται σε εντός του δείγματος προβλέψεις (in-sample) και εκτός του δείγματος (out-of-sample).
- Έτσι, εάν το δείγμα μας αναφέρεται στην περίοδο π.χ. 1990-2010 και χρησιμοποιήσουμε το διαθέσιμο δείγμα και την εξίσωση που προκύπτει από την εκτίμηση του υποδείγματος για να υπολογίσουμε τις θεωρητικές τιμές της εξαρτημένης μεταβλητής $\hat{Y}_T$ όπου $T=1990-2010$ τότε εκτελούμε προβλέψεις εντός του δείγματος. Από την άλλη πλευρά, μπορούμε να χρησιμοποιήσουμε την εκτιμηθείσα σχέση και να προβλέψουμε την συμπεριφορά της εξαρτημένης μεταβλητής $\hat{Y}_T$ για τιμές $T+1, T+2, \text{ κ.λπ.}$ πραγματοποιώντας προβλέψεις εκτός του δείγματος.

Διάγραμμα 23

Εντός και εκτός του δείγματος προβλέψεις

Χρόνος	Περίοδος ανάλυσης	
	Προβλέψεις εντός δείγματος	Προβλέψεις εκτός δείγματος

- Όταν το υπόδειγμα πρόκειται να χρησιμοποιηθεί για προβλέψεις, η αξιολόγηση της καταλληλότητας του γίνεται όχι με βάση το βαθμό της προσαρμοστικής του ικανότητας, αλλά με κριτήριο την προβλεπτική του ικανότητα.
- Βεβαίως, ένα υπόδειγμα που εμφανίζει υψηλό βαθμό προσαρμογής στα δεδομένα, αναμένεται να οδηγήσει σε ικανοποιητικές προβλέψεις.
- Όμως αυτό δεν συμβαίνει πάντοτε, και τούτο διότι κατά την διενέργεια των προβλέψεων υπεισέρχονται σφάλματα, πέρα από εκείνα που αφορούν την περίοδο εκτίμησης του υποδείγματος.
- Τα σημαντικότερα από αυτά προέρχονται:
 - (i) από τυχόν μεροληψίες που υπάρχουν στους εκτιμηθέντες συντελεστές,
 - (ii) από διαρθρωτικές μεταβολές που σημειώνονται στο υπόδειγμα κατά την περίοδο της πρόβλεψης,
 - (iii) από σφάλματα που γίνονται κατά την πρόβλεψη των εξωγενών (ερμηνευτικών) μεταβλητών και
 - (iv) από σφάλματα που διαπράττονται κατά την εκτίμηση των μελλοντικών τιμών, που αναμένεται να πάρει ο στοχαστικός όρος.

- Ο έλεγχος της προβλεπτικής ικανότητας του υποδείγματος μπορεί να αναφέρεται:
 - a) στον έλεγχο συγκεκριμένης τιμής της εξαρτημένης μεταβλητής και
 - b) στον συνολικό έλεγχο μιας σειράς τιμών της εξαρτημένης μεταβλητής.
- Στην πρώτη κατηγορία διακρίνουμε την περίπτωση κατά την οποία η πραγματική μελλοντική τιμή της εξαρτημένης μεταβλητής είναι γνωστή και στην περίπτωση που αυτή είναι άγνωστη. Με σκοπό την καλύτερη κατανόηση, χρησιμοποιούμε το ακόλουθο απλό υπόδειγμα:

$$Y_t = \alpha + \beta X_t + u_t \quad (115)$$

- Y_t = συνολικές εισαγωγές
- X_t = εθνικό εισόδημα
- Έστω ότι το υπόδειγμα εκτιμάται για την περίοδο 1980-2005. Οι προβλέψεις των τιμών της Y μετά το 2005 θα γίνουν με βάση το εκτιμημένο υπόδειγμα:

$$\hat{Y}_{T+1} = \hat{\alpha} + \hat{\beta} X_{T+1} + \hat{u}_{T+1} \quad (116)$$

- Εξάλλου η πραγματική τιμή της Y το έτος $T+1$ θα είναι Y_{T+1} .
- Υποθέτουμε ότι το Y_{T+1} αναφέρεται στο έτος μετά την λήξη της περιόδου ανάλυσης για το οποίο ήδη οι συνολικές εισαγωγές και το εθνικό εισόδημα είναι γνωστά, δηλ. τα Y_{T+1} και X_{T+1} .
- Στην περίπτωση αυτή ο έλεγχος της προβλεπτικής ικανότητας του υποδείγματος γίνεται με βάση τον τύπο:

$$t = \frac{Y_{T+1} - \hat{Y}_{T+1}}{S_F} \approx t_{n-2} \quad (117)$$

όπου $S_F = \sqrt{S_F^2}$ και $S_F^2 =$ αμερόληπτη εκτίμηση του σφάλματος πρόβλεψης η οποία υπολογίζεται μέσω της σχέσης:

$$S_F^2 = S^2 \left[1 + \frac{1}{T} + \frac{(X_{T+1} - \bar{X})^2}{\sum (X_t - \bar{X})^2} \right] \quad (118)$$

όπου $T =$ μέγεθος του δείγματος και $S^2 =$ αμερόληπτη εκτίμηση της σ_u^2

- Με άλλα λόγια, ο σχετικός έλεγχος γίνεται με βάση τις τιμές του κριτηρίου t του student.
- Εάν η τιμή που θα προκύψει από τον παραπάνω τύπο είναι μικρότερη από την κριτική τιμή του t σε ορισμένο επίπεδο στατιστικής σημαντικότητας, τότε συμπεραίνουμε ότι η παρατηρούμενη διαφορά μεταξύ της πραγματικής τιμής της Y_{T+1} και εκείνης που προβλέπεται με βάση το υπόδειγμα, δηλ. της $\hat{Y}_{T+1}$ δεν είναι στατιστικά σημαντική.
- Αυτό σημαίνει ότι η προβλεπτική ικανότητα του υποδείγματος είναι ικανοποιητική. Το αντίθετο συμπέρασμα θα συναχθεί όταν η τιμή που θα προκύψει από τον παραπάνω τύπο είναι μεγαλύτερη της κριτικής τιμής.

Εάν το Y_{T+1} αφορούσε το 2007, η πραγματική του τιμή θα ήταν άγνωστη.

Ωστόσο, αν έχουμε κάποια πρόβλεψη για το $\hat{Y}_{T+1}$, τότε μπορούμε να κατασκευάσουμε όρια εμπιστοσύνης μέσα στα οποία αναμένεται να βρίσκεται

η τιμή της Y κατά την περίοδο $T+1$. Ο σχετικός τύπος είναι:

$$\hat{Y}_{T+1} - S_F t_{n-2}, \alpha / 2 \leq Y_{T+1} \leq \hat{Y}_{T+1} + S_F t_{n-2}, \alpha / 2 \quad (119)$$

Για τον συνολικό έλεγχο της ακρίβειας πολλών προβλέψεων, χρησιμοποιείται ο συντελεστής ανισότητας του Theil. Αυτός υπολογίζεται από τον τύπο:

$$U_T = \frac{\sqrt{\frac{1}{m} \sum (A_{Ti} - F_{Ti})^2}}{\sqrt{\frac{1}{m} \sum A_{Ti}^2}} \quad (120)$$

όπου

U_T = ο συντελεστής ανισότητας

A_{Ti} = η ποσοστιαία μεταβολή στις πραγματικές τιμές της Y

F_{Ti} = η ποσοστιαία μεταβολή στις προβλεπόμενες τιμές της Y

m = ο αριθμός περιόδων για τις οποίες έχουν γίνει προβλέψεις

- Όταν $A_{Ti} = F_{Ti}$ για όλα τα i , έχουμε την περίπτωση των ιδανικών προβλέψεων.
- Σε αυτή την περίπτωση το U_T ισούται με το μηδέν. Εξάλλου, όταν $F_{Ti}=0$ για όλα τα i , τότε $U_T = 1$.
- Είναι φανερό ότι όσο περισσότερο οι τιμές του U_T πλησιάζουν το μηδέν, τόσο ικανοποιητικότερη είναι η προβλεπτική ικανότητα του υποδείγματος.
- Τέλος, η τιμή του U_T είναι πιθανό να υπερβαίνει την μονάδα. Σε αυτή την περίπτωση οι προβλέψεις είναι ιδιαίτερα κακές.

5.9 Παραβίαση Υποθέσεων Γραμμικού Υποδείγματος

- Η μέθοδος ελαχίστων τετραγώνων προϋποθέτει την υιοθέτηση αρκετά περιοριστικών υποθέσεων οι οποίες πρέπει να ελέγχονται εμπειρικά. Οι υποθέσεις αυτές αφορούν κυρίως αλλά όχι αποκλειστικά την συμπεριφορά του στοχαστικού όρου.
- Εάν παραβιάζεται η υπόθεση της ανεξαρτησίας των καταλοίπων (ύπαρξη αυτοσυσχέτισης) και εκτιμήσουμε το υπόδειγμα της αγοράς με την απλή μέθοδο ελαχίστων τετραγώνων οι συντελεστές του υποδείγματος θα είναι αμερόληπτοι μεν πλην όμως δεν θα είναι αποτελεσματικοί.
- Επιπρόσθετα, και ίσως το πιο σημαντικό, οι δειγματικές διακυμάνσεις των εκτιμηθέντων συντελεστών θα είναι μεροληπτικές με αποτέλεσμα οι συνήθεις έλεγχοι στατιστικής σημαντικότητας και η κατασκευή διαστημάτων εμπιστοσύνης να μην είναι έγκυρη.
- Τις ίδιες ακριβώς επιπτώσεις θα έχουμε εάν παραβιάζεται η υπόθεση της ομοσκεδαστικότητας και εμείς εκτιμήσουμε το υπόδειγμα της αγοράς με την απλή μέθοδο ελαχίστων τετραγώνων.
- Η παραβίαση της υπόθεσης της ανεξαρτησίας της ερμηνευτικής μεταβλητής και του στοχαστικού όρου δημιουργεί σοβαρά οικονομετρικά προβλήματα στον βαθμό που οι εκτιμηθέντες συντελεστές με τη μέθοδο ελαχίστων τετραγώνων είναι μεροληπτικοί και ασυνεπείς.
- Η παραβίαση της υπόθεσης του μηδενικού μέσου του στοχαστικού όρου όπως και η παραβίαση της υπόθεσης της κανονικότητας του στοχαστικού όρου δεν έχουν σημαντικές επιπτώσεις στην συνέπεια και στην ασυμπτωτική αποτελεσματικότητα των εκτιμητών των ελαχίστων τετραγώνων. Αυτό ισχύει ιδιαίτερα όταν το μέγεθος του δείγματος είναι μεγάλο.

- Πρέπει όμως να σημειώσουμε ότι η παραβίαση της υπόθεσης περί κανονικής κατανομής των κατάλοιπων έχει μεγάλη σημασία κατά την διατύπωση συμπερασμάτων αναφορικά με τον βαθμό αξιοπιστίας των εκτιμηθέντων συντελεστών του υποδείγματος.
- Όπως είναι γνωστό τα κλασσικά κριτήρια στατιστικού ελέγχου της σημαντικότητας των διαφορών υποθέσεων βασίζονται στην υπόθεση της κανονικότητας.
- Κατά συνέπεια αυτά δεν μπορούν να χρησιμοποιηθούν για τον έλεγχο των εκτιμηθέντων συντελεστών και την κατασκευή διαστημάτων εμπιστοσύνης όταν τα u_i δεν ακολουθούν την κανονική κατανομή.
- Η προηγούμενη ανάλυση καθιστά φανερό ότι η εκτίμηση του υποδείγματος της αγοράς πρέπει να γίνεται προσεκτικά εξετάζοντας αναλυτικά την παραβίαση των περιοριστικών υποθέσεων οι οποίες γίνονται παραδεκτές.
- Πιο αναλυτικά το συγκεκριμένο θέμα θα αναλυθεί στο επόμενο κεφάλαιο.

5.10 Σχέση μεταξύ συντελεστή προσδιορισμού και του συντελεστή συσχέτισης του Pearson

- Ο συντελεστής προσδιορισμού (R^2) και ο συντελεστής συσχέτισης του Pearson (r) συνδέονται στα πλαίσια του κλασικού μονοπαραγοντικού υποδείγματος.
- Η γραμμική εξίσωση που περιγράφει καλύτερα την σχέση μεταξύ των μεταβλητών X και Y υπολογίζεται μέσω της εξίσωσης της γραμμής παλινδρόμησης.
- Όπως είναι γνωστό, η συγκεκριμένη εξίσωση μπορεί να χρησιμεύσει για την πρόβλεψη των τιμών της μιας μεταβλητής (Y) μέσω της γνώσης της τιμής της άλλης (X).
- Έτσι, για κάθε τιμή της μεταβλητής X η συγκεκριμένη εξίσωση μας παρέχει μια τιμή για την Y η οποία αποτελεί την καλύτερη εκτίμηση της Y με βάση τα διαθέσιμα στοιχεία.
- Η προβλεπόμενη τιμή της μεταβλητής Y συμβολίζεται με $\hat{Y}$
- Οποιαδήποτε τιμή της Y μπορεί να οριστεί λοιπόν ως το άθροισμα της προβλεπόμενης τιμής $\hat{Y}$ και της διαφοράς της πραγματικής από την προβλεπόμενη τιμή όπως φαίνεται από την παρακάτω σχέση:

$$Y = \hat{Y} + (Y - \hat{Y}) \quad (121)$$

- Ο δεύτερος όρος στο δεξί τμήμα της παραπάνω σχέσης αντιπροσωπεύει όπως είναι γνωστό το κατάλοιπο (residual) της παλινδρόμησης το οποίο συμβολίζεται συνήθως με u . Η διακύμανση της μεταβλητής Y θα ισούται με το άθροισμα της διακύμανσης των δυο επιμέρους συστατικών, δηλαδή:

$$s_y^2 = s_{\hat{y}}^2 + s_u^2 \quad (122)$$

- Γνωρίζουμε ήδη ότι ο συντελεστής προσδιορισμού R^2 υπολογίζεται με την βοήθεια της παρακάτω σχέσης:

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{s_u^2}{s_y^2} \quad (123) \text{ από όπου προκύπτει η παρακάτω σχέση}$$

$$s_u^2 = (1 - R^2) s_y^2 \quad (124) \text{ και συνδυάζοντας τις παραπάνω σχέσεις θα έχουμε:}$$

$$R^2 = r^2 = \frac{s_{\hat{y}}^2}{s_y^2} \quad (125)$$