Α.Τ.Ε.Ι. ΠΑΤΡΑΣ

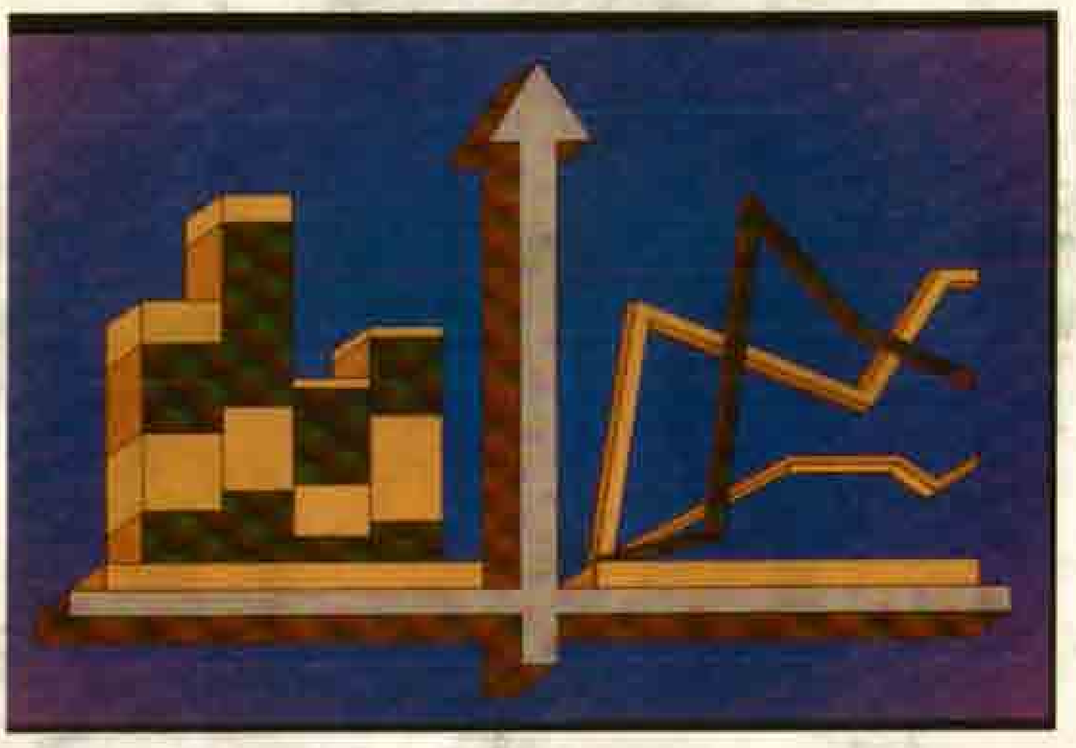
ΣΧΟΛΗ: ΟΙΚΟΝΟΜΙΑΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ

ΤΜΗΜΑ: ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΥ ΣΧΕΔΙΑΣΜΟΥ ΚΑΙ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΘΕΜΑ

ΣΥΣΧΕΤΙΣΗ: ΑΠΛΗ-ΠΟΛΛΑΠΛΗ ΠΑΛΙΝΔΡΟΜΗΣΗ



ΕΠΟΠΤΕΥΩΝ ΚΑΘΗΓΗΤΗΣ:
ΠΑΠΑΘΑΝΑΣΟΠΟΥΛΟΥ
ΧΡΥΣΑΝΘΗ

ΟΙ ΣΠΟΥΔΑΣΤΕΣ:
ΤΖΕΛΗ ΑΚΡΙΒΗ
ΣΤΑΥΡΟΠΟΥΛΟΥ ΚΩΝΣΤΑΝΤΙΝΑ

ΠΑΤΡΑ - 2006

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ 4ο

ΚΕΦΑΛΑΙΟ 5ο

ΚΕΦΑΛΑΙΟ 1ο

ΕΙΣΑΓΩΓΗ

Συχνά δύο ή και περισσότερες μεταβλητές εξετάζονται μαζί με την ελπίδα να προσδιοριστεί η σχέση που μπορεί να υπάρχει μεταξύ τους. Για παράδειγμα θα μπορούσε να εξεταστεί η επίδραση που έχει η θερμοκρασία στα αποτελέσματα ενός πειράματος. Άλλες φορές πάλι δύο ή περισσότερες μεταβλητές εξετάζονται με στόχο την πρόβλεψη της μίας από τις άλλες. Για παράδειγμα γνωρίζοντας τη σχέση μεταξύ εξόδων που δαπανώνται για διαφήμιση και εσόδων από τις πωλήσεις κάποιου προϊόντος θα μας ενδιέφερε να προβλέψουμε, κατά πιθανότητα πάντα, τα έσοδα όταν για τη διαφήμιση του προϊόντος διατεθεί κάποιο συγκεκριμένο χρηματικό ποσό.

Η στατιστική μεθοδολογία που χρησιμοποιεί τη σχέση μεταξύ δύο ή περισσοτέρων μεταβλητών έτσι ώστε η μία να μπορεί να προβλεφθεί από την άλλη ή τις άλλες καλείται **Ανάλυση της Παλινδρόμησης** (Regression Analysis). Τον όρο Παλινδρόμηση χρησιμοποίησε για πρώτη φορά ο F. Galton το 1900 όταν μελετώντας τη σχέση μεταξύ του ύψους των γονέων και αυτού των παιδιών τους, παρατήρησε ένα είδος επαναφοράς (παλινδρόμησης) του ύψους των παιδιών στο ύψος των γονέων τους.

Η Ανάλυση της Παλινδρόμησης αποτελεί έναν από τους σημαντικότερους κλάδους της στατιστικής με ευρείες εφαρμογές σε όλες τις σύγχρονες επιστήμες.

Μία εκτεταμένη όμως μελέτη της δεν είναι ο σκοπός αυτής της εργασίας. Στόχος μας είναι να αναπτύξουμε τη μεθοδολογία που εφαρμόζεται στην απλή και πολλαπλή παλινδρόμηση καθώς και η αναφορά σε άλλα είδη συσχέτισης.

Πιο συγκεκριμένα, στο δεύτερο κεφάλαιο, θα ασχοληθούμε εκτενέστερα με την απλή γραμμική παλινδρόμηση. Το απλό παλινδρομικό μοντέλο εκφράζει τη σχέση ανάμεσα σε μία μεταβλητή **Y** (εξαρτημένη) και σε μία μεταβλητή **X** (ανεξάρτητη). Όμως η ευθεία που ορίζεται από τη σχέση που προκύπτει δεν περνάει από όλα τα σημεία που δημιουργούνται από τα ζεύγη ($\mathbf{Y}_t, \mathbf{X}_t$). Οι αποκλίσεις από την ευθεία μπορούν να ληφθούν υπόψη με την προσθήκη μίας τυχαίας μεταβλητής $\mathbf{u}_t$.

Για να εφαρμοστεί η απλή παλινδρόμηση θα πρέπει πρώτα να τηρούνται κάποιες βασικές υποθέσεις έτσι ώστε να είναι έγκυρες και αποτελεσματικές οι διαδικασίες που θα ακολουθηθούν καθώς και τα αποτελέσματα που θα προκύψουν. Για να διαμορφώσουμε ένα παλινδρομικό μοντέλο το οποίο θα εκφράζει όσο γίνεται καλύτερα τη σχέση που μπορεί να υπάρχει ανάμεσα στις δύο μεταβλητές θα πρέπει να εκτιμήσουμε τις παραμέτρους του υποδείγματος και στη συνέχεια να προβούμε σε έλεγχο υποθέσεων για αυτές τις παραμέτρους έτσι ώστε να δούμε κατά πόσο είναι αποδεκτά τα αποτελέσματα. Για να είμαστε σίγουροι για τη σχέση που υπάρχει ανάμεσα στις δύο μεταβλητές είναι χρήσιμο να εξετάσουμε τον συντελεστή προσδιορισμού καθώς και το συντελεστή συσχέτισης και τέλος θα δούμε τον τρόπο με τον οποίο χρησιμοποιούμε το απλό παλινδρομικό μοντέλο για να κάνουμε πρόβλεψη.

Στη συνέχεια, στο τρίτο κεφάλαιο, θα ασχοληθούμε με τη μελέτη τριών ή περισσοτέρων μεταβλητών με αντικειμενικό σκοπό να εξακριβώσουμε, κατά πρώτο λόγο, πώς μεταβάλλονται οι τιμές της εξαρτημένης μεταβλητής **Y**, όταν γνωρίζουμε τις τιμές των άλλων μεταβλητών και κατά δεύτερο λόγο, να μελετήσουμε την ένταση με την οποία η διαμόρφωση των τιμών της **Y** σχετίζεται με τις τιμές και τις μεταβολές των άλλων μεταβλητών. Η σχέση που συνδέει μια εξαρτημένη μεταβλητή με περισσότερες από δύο ανεξάρτητες μεταβλητές ονομάζεται πολλαπλή παλινδρόμηση.

Ενώ η απλή παλινδρόμηση παριστάνεται με μια ευθεία ή μια άλλη καμπύλη, η πολλαπλή παλινδρόμηση παριστάνεται με μια επιφάνεια. Όπως είναι φυσικό και στην περίπτωση της πολλαπλής παλινδρόμησης ισχύουν οι υποθέσεις που ισχύουν και στην απλή αλλά υπάρχουν και κάποιες επιπλέον. Αφού εξετάσουμε αν αυτές οι υποθέσεις παραβιάζονται ή όχι στη συνέχεια θα προχωρήσουμε στη διατύπωση των ιδιοτήτων των παραμέτρων της πολλαπλής παλινδρόμησης, στην εκτίμηση τους καθώς και στον έλεγχο αυτών. Θα δούμε ακόμα τον συντελεστή πολλαπλού προσδιορισμού, τους συντελεστές μερικής συσχέτισης και τον συντελεστή μερικού προσδιορισμού, τη λειτουργία αυτών και το πόσο σημαντικοί είναι για τη δημιουργία ενός ορθού μοντέλου. Επίσης θα ασχοληθούμε με τους μετασχηματισμούς που μπορούμε να κάνουμε στην περίπτωση που δεν υπάρχει γραμμική σχέση ανάμεσα στις μεταβλητές καθώς και με την σωστή επιλογή των μεταβλητών που θα χρησιμοποιήσουμε σε ένα υπόδειγμα.

Ακόμα, στο ίδιο κεφάλαιο, θα μας απασχολήσουν τρεις σημαντικές έννοιες, η πολυσυγγραμμικότητα, η ετεροσκεδαστικότητα και η αυτοσυσχέτιση. Πιο συγκεκριμένα θα ορίσουμε αυτούς τους όρους και θα μιλήσουμε για τις συνέπειες τους όταν εμφανίζονται σε ένα υπόδειγμα αλλά θα δώσουμε και λύσεις στα προβλήματα που μπορεί να εμφανιστούν εξαιτίας τους.

Στη συνέχεια, στο τέταρτο κεφάλαιο, θα ασχοληθούμε με τα διάφορα άλλα είδη συσχέτισης που μπορεί να παρουσιαστούν ανάμεσα σε μεταβλητές οι οποίες παρουσιάζουν κάποια αλληλεξάρτηση. Πιο συγκεκριμένα θα δούμε μη γραμμικές εξισώσεις παλινδρόμησης και θα δώσουμε αριθμητικά παραδείγματα καθώς και διαγράμματα έτσι ώστε να καταλάβουμε καλύτερα τη χρησιμότητά τους στην οικονομική θεωρία κυρίως. Επίσης θα προχωρήσουμε και στην ανάλυση της συσχέτισης δεδομένων που είναι χωρισμένα κατά τάξεις.

Τέλος, στο πέμπτο κεφάλαιο, ασχοληθήκαμε με μία γενική εφαρμογή στην οποία καταφέραμε να εφαρμόσουμε αρκετές από τις τεχνικές που αναλύσαμε στο θεωρητικό μέρος. Στην εφαρμογή προσπαθήσαμε να δημιουργήσουμε ένα μοντέλο το οποίο να προσδιορίζει όσο το δυνατόν πιο αποτελεσματικά τη σχέση που έχει η ανταγωνιστικότητα με όλους τους άλλους οικονομικούς παράγοντες για μία επιχείρηση. Η εφαρμογή έχει υλοποιηθεί με τη βοήθεια του στατιστικού προγράμματος **SPSS** διότι η πρακτική εφαρμογή της θεωρίας χωρίς τη βοήθεια του συγκεκριμένου προγράμματος θα ήταν αδύνατο να γίνουν.

ΚΕΦΑΛΑΙΟ 2ο

ΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

2.1 ΕΙΣΑΓΩΓΗ

Σε αυτό το κεφάλαιο θα ασχοληθούμε εκτενέστερα με την απλή γραμμική παλινδρόμηση. Το απλό παλινδρομικό μοντέλο εκφράζει τη σχέση ανάμεσα σε μία μεταβλητή **Y** (εξαρτημένη) και σε μία μεταβλητή **X** (ανεξάρτητη). Έστω λοιπόν ότι θέλουμε να ερευνήσουμε τη σχέση που υπάρχει ανάμεσα στις δαπάνες κατανάλωσης και στο διαθέσιμο εισόδημα σύμφωνα με τα δεδομένα που έχουμε συλλέξει για ένα δεδομένο αριθμό οικογενειών σε ένα δεδομένο χρονικό διάστημα. Όπου $\mathbf{Y_t}$ είναι οι δαπάνες κατανάλωσης της οικογένειας **t** και $\mathbf{X_t}$ το διαθέσιμο εισόδημα της κάθε οικογένειας **t**. Υποθέτουμε ότι η μαθηματική μορφή της συναρτησιακής σχέσης αυτών των δυο μεταβλητών είναι γραμμική και εκφράζεται ως εξής :

$$\mathbf{Y_t = \beta_0 + \beta_1 X_t} \qquad (2.1)$$

Η σχέση (2.1) ονομάζεται προσδιοριστική και σημαίνει ότι όλες οι οικογένειες με το ίδιο εισόδημα έχουν και τις ίδιες δαπάνες κατανάλωσης. Όμως αυτό δεν ισχύει στην πραγματικότητα. Η ευθεία που ορίζεται από την θεωρητική σχέση (2.1) δεν περνάει από όλα τα σημεία που προκύπτουν από τα ζεύγη $(\mathbf{Y_t, X_t})$.

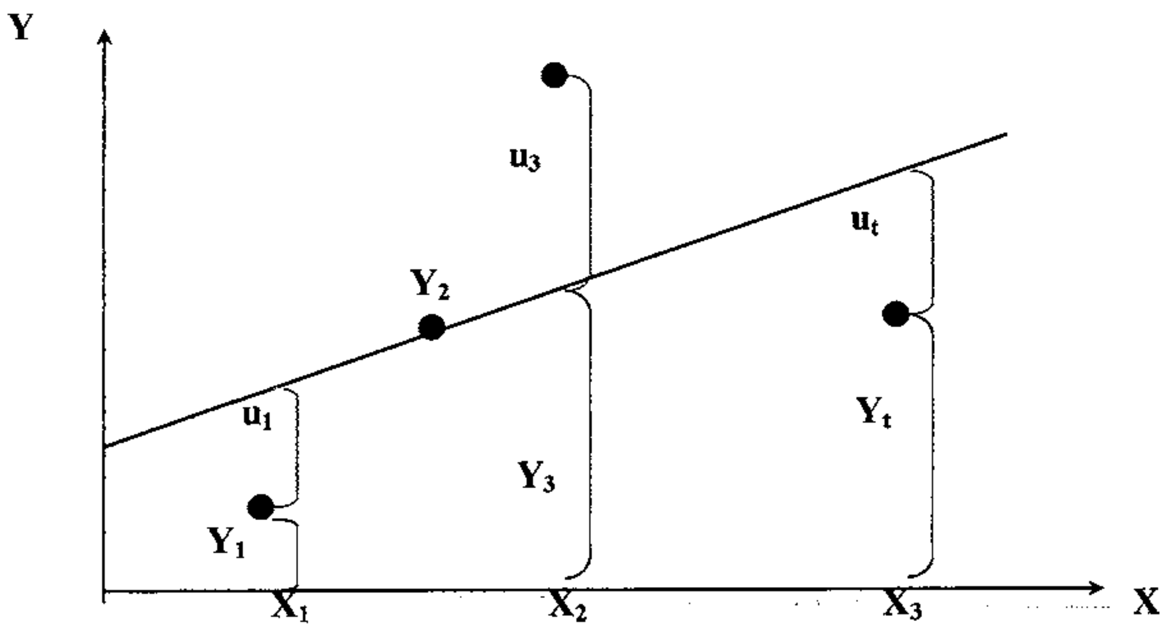
Οι αποκλίσεις από την ευθεία μπορούν να ληφθούν υπόψη με την προσθήκη μίας τυχαίας μεταβλητής $\mathbf{u}_t$ οπότε η παραπάνω σχέση μετατρέπεται σε μία στοχαστική σχέση οποία είναι η εξής :

$$\mathbf{Y}_t = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1\mathbf{X}_t + \mathbf{u}_t \qquad (2.2)$$

Γενικά μία οικονομική σχέση αποτελείται από δύο μέρη, το συστηματικό μέρος $\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1\mathbf{X}_t$ και το μη συστηματικό μέρος που παριστάνεται από τον τυχαίο όρο $\mathbf{u}_t$. Η τυχαία μεταβλητή $\mathbf{u}_t$ ονομάζεται και διαταρακτικός όρος.

Τα παραπάνω μπορούν να δοθούν γραφικά με το παρακάτω διάγραμμα.

<u>Διάγραμμα Διασποράς (2.1)</u>



Τα σημεία που ορίζονται από τις παρατηρήσεις παριστάνουν την πραγματική σχέση που υπάρχει ανάμεσα στις μεταβλητές **X**, **Y**. Όπως είναι φανερό από το διάγραμμα (2.1) είναι απαραίτητη η ύπαρξη του όρου $\mathbf{u}_t$.

2.2 ΒΑΣΙΚΕΣ ΥΠΟΘΕΣΕΙΣ ΤΗΣ ΑΠΛΗΣ ΓΡΑΜΜΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Η σχέση (2.2) αποτελεί την οικονομετρική μορφή της γραμμικής σχέσης που υποθέτουμε ότι υπάρχει ανάμεσα στις μεταβλητές **X** και **Y**. Η στοχαστική φύση της σχέσης (2.2) εκφράζει ότι για κάθε τιμή της **X** δεν υπάρχει μόνο μία τιμή της **Y** αλλά μία ολόκληρη κατανομή τιμών. Για να είναι το υπόδειγμα πλήρως εξειδικευμένο απαιτείται όχι μόνο ο μαθηματικός καθορισμός της σχέσης που συνδέει την εξαρτημένη μεταβλητή **Y** με την ανεξάρτητη μεταβλητή **X** αλλά επίσης και η εξειδίκευση της κατανομής του $\mathbf{u}_t$. Σύμφωνα με τον Jeffrey Jarrett (2002), « Μέθοδοι Προβλέψεων », Αθήνα : Εκδόσεις Gutenberg η παλινδρόμηση προϋποθέτει ότι ισχύει το σύνολο των παρακάτω υποθέσεων στον πληθυσμό στον οποίο ανήκουν οι δύο μεταβλητές, προκειμένου να χρησιμοποιηθεί γραμμικό μοντέλο παλινδρόμησης.

ΥΠΟΘΕΣΗ 1 : η σχέση η οποία υπάρχει ανάμεσα στις μεταβλητές περιγράφεται από μία ευθεία γραμμή, δηλαδή είναι γραμμική.

Από μαθηματικής άποψης το γραμμικό μοντέλο παλινδρόμησης μπορεί να εκφραστεί ως εξής :

$$\mathbf{Y=\beta_0 + \beta_1 X+e} \qquad (2.3)$$

Σε αυτή την εξίσωση το **e** είναι το τυχαίο σφάλμα που μετράει την κάθετη απόκλιση της κάθε τιμής της εξαρτημένης μεταβλητής από την πληθυσμιακή παλινδρομική γραμμή για την αντίστοιχη τιμή της ανεξάρτητης μεταβλητής. Αυτό το σφάλμα μπορεί να προέλθει από μία ή και περισσότερες αιτίες.

1. Ατελής θεωρία : Συγκεντρωτικές μεταβλητές που συσχετίζονται με τη συμπεριφορά της **Y** μπορεί να έχουν παραλειφθεί. Μία περισσότερο πλήρης θεωρία ίσως απαιτούσε να συμπεριληφθούν πληροφορίες που να αφορούν και άλλα μεγέθη.. Επομένως, θα μπορούσαν να προστεθούν περισσότερες ανεξάρτητες μεταβλητές στις παλινδρομικές σχέσεις (πολλαπλή παλινδρόμηση).
2. Ατελής προσδιορισμός : Πρόκειται για τη περίπτωση που εκτιμάται μία γραμμική σχέση, ενώ η πραγματική σχέση ίσως δεν είναι γραμμική, αλλά εκφράζεται καλύτερα με την εξίσωση μιας παραβολής.
3. Σφάλματα μέτρησης : Ακόμα και όταν η σχέση έχει προετοιμαστεί συστηματικά, μπορούν να συμβούν και υπολογιστικά σφάλματα κατά τη μέτρηση των μεταβλητών.

Όταν οι αιτίες που προκαλούν τα τυχαία σφάλματα είναι ελεγχόμενες τότε είναι δυνατό να κάνουμε προβλέψεις με ακρίβεια.

ΥΠΟΘΕΣΗ 2 : η κατανομή των τιμών της **X** που θα εμφανιστεί θα είναι ίδια σε οποιαδήποτε επανάληψη του πειράματος. Όμως η κατανομή των τιμών της **Y** για δεδομένες τιμές της **X** μπορεί να διαφέρει από πείραμα σε πείραμα εξαιτίας της επίδρασης του όρου του τυχαίου σφάλματος.

ΥΠΟΘΕΣΗ 3 : η αναμενόμενη τιμή του τυχαίου σφάλματος του τυχαίου σφάλματος είναι μηδέν και εκφράζεται ως εξής :

$$E(e)=0$$

Δηλαδή κατά μέσο όρο η τιμή του τυχαίου σφάλματος είναι ίση με το μηδέν. Για να κατανοήσουμε την παραδοχή αυτή θεωρούμε την αναμενόμενη τιμή και των δύο μερών της εξίσωσης του υποδείγματος για δεδομένη τιμή της **X**. Η αναμενόμενη τιμή του αθροίσματος :

$$\mathbf{E(Y/X)=E(\beta_0 + \beta_1 X+e)} \quad (2.4)$$

μπορεί να αναχθεί στο άθροισμα των αναμενόμενων τιμών :

$$\mathbf{E(Y/X)=E(B_0)+E(B_1X)+E(e)}$$

$\boldsymbol{\beta_0}$ και $\boldsymbol{\beta_1}$ είναι σταθερές και η **X** είναι δεδομένη, έτσι ώστε να μπορούμε να τη μεταχειριστούμε ως σταθερά. Επομένως, η εξίσωση γίνεται :

$$\mathbf{E(Y/X)=\beta_0 + \beta_1 X} \quad (2.5)$$

Η τρίτη υπόθεση συνεπάγεται ότι για μια δεδομένη τιμή της **X**, ο μέσος των τιμών **Y** βρίσκεται πάνω στη γραμμή παλινδρόμησης.

ΥΠΟΘΕΣΗ 4 : η διακύμανση του όρου του τυχαίου σφάλματος είναι ίδια για κάθε τιμή του **X**. Η υπόθεση αυτή ονομάζεται ομοσκεδαστικότητα :

$$\mathbf{VAR(e)=E[e-E(e)]^2=E(e^2)=C} \quad (2.6)$$

όπου **C** είναι ένας σταθερός όρος. Αλλά από το μοντέλο έχουμε :

$$\mathbf{e=Y-(B_0+B_1X)} \quad \mathbf{(2.7)}$$

Αυτή η υπόθεση εισάγει ότι **$E[Y-(\beta 0+\beta 1X)]^2=C$**, ή ότι η διακύμανση και η σταθερή απόκλιση είναι ίδιες για κάθε τιμή **Χ** στον πληθυσμό. Η **VAR(e)** μετράει τη μεταβλητότητα στην τιμή της εξαρτημένης μεταβλητής **Y**, περί τη γραμμή παλινδρόμησης για δεδομένες τιμές της ανεξάρτητης μεταβλητής **Χ**.

ΥΠΟΘΕΣΗ 5 : οι τιμές του e είναι ανεξάρτητες μεταξύ τους, δηλαδή **$E(e_i e_j)=0$**. Αυτό σημαίνει ότι η τιμή του σφάλματος για οποιαδήποτε δεδομένη τιμή της ανεξάρτητης μεταβλητής δεν σχετίζεται με το σφάλμα για οποιαδήποτε άλλη τιμή της **Χ.** Όταν τα σφάλματα αυτά σχετίζονται, τότε έχουμε το πρόβλημα της αυτοσυσχέτισης.

ΥΠΟΘΕΣΗ 6 : ο όρος του σφάλματος για κάθε τιμή της **Χ** κατανέμεται κανονικά. Από τη στιγμή που οι τιμές της **Y** για δεδομένες τιμές της **Χ** ποικίλλουν από πείραμα σε πείραμα μόνο εξαιτίας του όρου του τυχαίου σφάλματος (οι τιμές της **Χ** μένουν ίδιες από πείραμα σε πείραμα) , οι τιμές της **Y** πρέπει αναγκαστικά να κατανέμονται κανονικά.

2.3 ΕΚΤΙΜΗΣΗ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΤΟΥ ΑΠΛΟΥ ΠΑΛΙΝΔΡΟΜΙΚΟΥ ΜΟΝΤΕΛΟΥ

Σκοπός της εκτίμησης της συναρτησιακής μας σχέσης είναι η εκτίμηση των άγνωστων παραμέτρων άρα και η εκτίμηση του προσδιοριστικού μέρους της εξίσωσης που συμβολίζεται ως εξής :

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 X_i \qquad (2.8)$$

Όπου $\hat{b}_0$, $\hat{b}_1$ είναι οι εκτιμήτριες των παραμέτρων β_0 και β_1 αντίστοιχα. Το μη ερμηνευμένο μέρος $\hat{e}_i = Y_i - \hat{Y}_i$ ονομάζεται κατάλοιπο και αποτελεί μία εκτίμηση του άγνωστου όρου σφάλματος $\mathbf{e}_i$. Για να βρούμε όμως τις άγνωστες παραμέτρους χρειαζόμαστε κάποιο κριτήριο προσαρμογής το οποίο να δίνει εκτιμήτριες που έχουν κάποιες συγκεκριμένες ιδιότητες όσον αφορά τη σχέση τους με τις αντίστοιχες παραμέτρους. Ένα τέτοιο κριτήριο είναι η μέθοδος των ελαχίστων τετραγώνων η οποία έχει να κάνει με την ελαχιστοποίηση των τετραγώνων των καταλοίπων και είναι η πιο γνωστή και η πιο απλή μέθοδος.

$$\sum_{i=1}^{n}\hat{\mathbf{e}}_i^2 = \sum_i(\mathbf{Y}_i - \hat{\mathbf{b}}_0 - \hat{\mathbf{b}}_1\mathbf{X}_i)^2 \qquad (2.9)$$

οι συνθήκες πρώτης τάξης ως προς τα $\hat{\mathbf{b}}_0$ και $\hat{\mathbf{b}}_1$ μας δίνουν τα παρακάτω :

$$\frac{\partial \sum_i \hat{\mathbf{e}}_i^2}{\partial \hat{\mathbf{b}}_0} = \sum_i 2(\mathbf{Y}_i - \hat{\mathbf{b}}_0 - \hat{\mathbf{b}}_1\mathbf{X}_i)(-1) = 0 \qquad \text{δηλαδή} \qquad \sum_i \hat{\mathbf{e}}_i = 0$$

και

$$\frac{\partial \sum_i \hat{\mathbf{e}}_i^2}{\partial \hat{\mathbf{b}}_1} = \sum_i 2(\mathbf{Y}_i - \hat{\mathbf{b}}_0 - \hat{\mathbf{b}}_1\mathbf{X}_i)(-\mathbf{X}_i) = 0 \qquad \text{δηλαδή} \qquad \sum_i \mathbf{X}_i\hat{\mathbf{e}}_i = 0$$

Στη συνέχεια προκύπτουν οι κανονικές εξισώσεις των ελαχίστων τετραγώνων οι οποίες είναι :

$$\sum_i \mathbf{Y}_i = \mathbf{n}\hat{\mathbf{b}}_0 + (\sum_i \mathbf{X}_i)\hat{\mathbf{b}}_1 \qquad (2.10)$$

$$\sum_i \mathbf{X}_i\mathbf{Y}_i = (\sum_i \mathbf{X}_i)\hat{\mathbf{b}}_0 + (\sum_i \mathbf{X}_i^2)\hat{\mathbf{b}}_1 \qquad (2.11)$$

η λύση του παραπάνω συστήματος μας δίνει τις εκτιμήτριες ελαχίστων τετραγώνων :

$$\hat{b}_1 = \frac{\sum_i X_i Y_i - n\bar{X}\bar{Y}}{\sum_i X_i^2 - n\bar{X}^2} = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_i (Xi - \bar{X})^2} \qquad (2.12)$$

και

$$\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X} \qquad (2.13)$$

όπου **X** και **Y** είναι οι δειγματικοί μέσοι των $\bar{X}_i$ και $\bar{Y}_i$.

Χρησιμοποιούμε τη μέθοδο των ελαχίστων τετραγώνων γιατί οι παράμετροι που εκτιμώνται μέσω τις αυτής της μεθόδου έχουν τις ιδιότητες **BLUE** (Best Linear Unbiased Estimators) που αυτό σημαίνει τα ακόλουθα σύμφωνα με τον Damodar Gujarati (1999), « Essentials of Econometrics », 2nd Edition McGraw Hill. :

Best: οι εκτιμήσεις είναι βέλτιστες γιατί έχουν την ελάχιστη διακύμανση σε σχέση με άλλη μέθοδο.

Linear: οι εκτιμήσεις μας είναι γραμμικές συναρτήσεις της τυχαίας μεταβλητής Y.

Unbiased: η εκτιμώμενη τιμή είναι ίση με την πραγματική τιμή του πληθυσμού καθώς και η εκτιμώμενη διακύμανση είναι ίση με αυτή του πληθυσμό.

$$\mathbf{E(b_1) = B_1,\ E(b_2) = B_2,\ E(\hat{s}^2) = s^2}$$

Θα προχωρήσουμε σε ένα παράδειγμα με τη μέθοδο των ελαχίστων τετραγώνων για να την κατανοήσουμε καλύτερα. Ο παρακάτω πίνακας αναφέρεται στην αξία των εισαγωγών καταναλωτικών αγαθών (Y) και στο διαθέσιμο εισόδημα (X) σε σταθερές τιμές για την Ελληνική οικονομία και για την χρονική περίοδο 1988 έως 2003. Στον πίνακα αυτό δίνονται ακόμα και κάποιες περαιτέρω πληροφορίες για να γίνουν πιο εύκολα οι υπολογισμοί που θέλουμε να κάνουμε στη συνέχεια.

Παράδειγμα

Πίνακας (2.1) Αξία των εισαγωγών καταναλωτικών αγαθών και το διαθέσιμο εισόδημα

Έτος	Αξία εισαγωγών καταναλωτικών αγαθών (Y) σε δισεκ. δρχ	Διαθέσιμο εισόδημα (X) Σε δισεκ. δρχ	$\hat{Y}$	$\hat{u}$ (κατάλοιπα)
1988	5,121	105,508	5,094	0,026
1989	4,134	107,497	5,169	-1,035
1990	4,653	111,875	5,333	-0,679
1991	5,622	124,676	5,814	-0,191
1992	5,499	130,118	6,018	-0,518
1993	6,453	142,140	6,469	-0,016
1994	7,093	155,338	7,964	0,129
1995	8,907	171,456	7,568	1,338
1996	8,625	182,420	7,980	0,644
1997	9,204	192,895	8,373	0,831
1998	9,647	204,164	8,796	0,851
1999	10,167	221,908	9,461	0,705
2000	9,961	240,471	10,158	-0,197
2001	10,580	267,849	11,185	-0,605
2002	10,658	289,450	11,995	-1,337
2003	13,139	318,550	13,087	0,052

ΠΗΓΗ :Τάκης Παπαϊωάννου – Σωτήρης Λουκάς(1990), « Εισαγωγή στη Στατιστική », Ιωάννινα.

Επίσης υπολογίσαμε και τα εξής ποσά :

ΣY = 129,463

ΣX = 2966,315 $\bar{X}$ **= 185,394**

ΣXY = 26541,949

T = 16 (αριθμός παρατηρήσεων) $\bar{Y}$ **= 8,091**

ΣX2 = 617645,622 Σx^2 **= 67706,579**

ΣY2 = 1151,018 Σy^2 **= 103,476**

Σxy = 2540,196

Κάνοντας αντικατάσταση στους τύπους (2.12) και (2.13) προκύπτουν τα εξής :

$$b_1 = 0{,}0375 \quad \text{και} \quad b_0 = 1{,}136$$

Επομένως η συνάρτηση εισαγωγών καταναλωτικών αγαθών για την περίοδο 1988 έως 2003 είναι :

$$\hat{Y}_t = 1{,}136 + 0{,}0375\, X_t$$

Ο συντελεστής παλινδρόμησης b_1 παριστάνει τη μεταβολή στην προσδοκώμενη τιμή της εξαρτημένης μεταβλητής (**Y**) όταν η ερμηνευτική μεταβλητή (**X**) μεταβάλλεται κατά μία μονάδα. Είναι με άλλα λόγια η παράγωγος της $E(Y_t)$ ως προς X_t δηλαδή :

$$\frac{dE(Y_t)}{dX_t} = b_1 \qquad (2.14)$$

Στο παράδειγμα μας ο συντελεστής b_1 παριστάνει την οριακή ροπή για τις εισαγωγές των καταναλωτικών αγαθών. Για την περίοδο 1988-2003 η εκτίμηση της οριακής ροπής είναι 0,0375, αυτό σημαίνει ότι όταν το διαθέσιμο εισόδημα αυξάνεται κατά ένα δισεκατ. δρχ. οι εισαγωγές καταναλωτικών αγαθών θα αυξηθούν κατά 37,5 εκατ. δρχ.

Μπορούμε όμως να προχωρήσουμε και παραπέρα και να βρούμε την ελαστικότητα της **Y** ως προς **X** η οποία είναι :

$$\eta_{yx} = \frac{dY}{dX}\,\frac{X}{Y} \qquad (2.15)$$

ή

$$\eta_{yx} = \frac{dY}{dx}\,\frac{Y}{X} \qquad (2.16)$$

Επειδή όπως είναι γνωστό η ελαστικότητα δεν παραμένει σταθερή σε όλο το μήκος της συνάρτησης συνήθως υπολογίζουμε την ελαστικότητα όταν η **Y** και η **X** παίρνουν για τιμές τους αντίστοιχους μέσους οπότε έχουμε :

$$\eta_{yx} = b_1 \frac{\bar{X}}{\bar{Y}} \qquad (2.17)$$

Όπου $\bar{\eta}_{yx}$ = η ελαστικότητα στο σημείο των μέσων, $\bar{X}$ = ο μέσος των **X**, $\bar{Y}$ = ο μέσος των **Y**.

Στο παράδειγμά μας η εισοδηματική ελαστικότητα των εισαγωγών καταναλωτικών αγαθών στο σημείο των μέσων είναι

$$\bar{\eta}_{yx} = 0{,}0375\ \frac{185{,}394}{8{,}091} = 0{,}86$$

Το παραπάνω αποτέλεσμα ερμηνεύεται ως εξής, όταν το εισόδημα αυξάνεται κατά 10% οι εισαγωγές καταναλωτικών αγαθών αυξάνονται κατά 8,6 %.

2.4 ΈΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ ΓΙΑ ΤΙΣ ΠΑΡΑΜΕΤΡΟΥΣ b_0, b_1

Για να προχωρήσουμε στον έλεγχο υποθέσεων οποιασδήποτε παραμέτρου πρέπει να ακολουθήσουμε τα παρακάτω βήματα σύμφωνα με τον Πέτρο Α. Κιόχο (1993), « Στατιστική », Αθήνα : Εκδόσεις Interbooks.:

1. Ορίζουμε τη μηδενική ή αρχική υπόθεση που τη συμβολίζουμε με $\mathbf{H_0}$, η οποία είναι η υπόθεση που κάνουμε για μία συγκεκριμένη τιμή της παραμέτρου και είναι η πιο σοβαρή τιμή στον έλεγχο.
2. Ορίζουμε την εναλλακτική υπόθεση που τη συμβολίζουμε με $\mathbf{H_1}$.
3. Προσδιορίζουμε τον έλεγχο που θα εφαρμόσουμε για την αποδοχή ή την απόρριψη της $\mathbf{H_0}$. Ο έλεγχος αυτός πάντα θα βασίζεται σε μία στατιστική συνάρτηση.
4. Καθορίζουμε την απορριπτική περιοχή της $\mathbf{H_0}$ ή αλλιώς την κρίσιμη περιοχή του ελέγχου, δηλαδή την περιοχή του δειγματικού χώρου για την οποία απορρίπτεται η $\mathbf{H_0}$.
5. Τέλος εξάγω συμπεράσματα. Για να κάνουμε το τελευταίο βήμα, πρέπει να γίνει σύγκριση ανάμεσα στην κρίσιμη τιμή και στο **επίπεδο σημαντικότητας (α)** που έχουμε ήδη ορίσει.

Σε κάθε έλεγχο υπάρχει πιθανότητα να γίνουν δύο είδη σφαλμάτων :

1. **Σφάλμα τύπου Ι** : το οποίο αναφέρεται στην απόρριψη της $\mathbf{H_0}$ ενώ στην πραγματικότητα αυτή είναι αληθής.
2. **Σφάλμα τύπου ΙΙ** : το οποίο αναφέρεται στην αποδοχή της $\mathbf{H_0}$ ενώ στην πραγματικότητα είναι ψευδής.

Υπάρχουν επίσης και δύο είδη ελέγχων :

1. Ο αμφίπλευρος έλεγχος που εκφράζεται ως εξής :

   $$\mathbf{H_0 : \theta = \theta_0 \quad v.s \quad H_1 : \theta \neq \theta_0}$$

   σε αυτή την περίπτωση απορρίπτεται η μηδενική υπόθεση όταν το κρίσιμο σημείο είναι μεγαλύτερο από την κρίσιμη τιμή του επιπέδου σημαντικότητας της κατανομής στην οποία ανήκουν τα δεδομένα μας.

2. Ο μονόπλευρος έλεγχος ο οποίος έχει τις εξής δύο μορφές :

   $$\mathbf{H_0 : \theta = \theta_0 \quad v.s \quad H_1 : \theta > \theta_0}$$

   σε αυτή την περίπτωση απορρίπτεται η μηδενική υπόθεση όταν το κρίσιμο σημείο είναι μεγαλύτερο από την κρίσιμη τιμή του επιπέδου σημαντικότητας της κατανομής στην οποία ανήκουν τα δεδομένα μας.

   $$\mathbf{H_0 : \theta = \theta_0 \quad v.s \quad H_1 : \theta < \theta_0}$$

   σε αυτή την περίπτωση απορρίπτεται η μηδενική υπόθεση όταν το κρίσιμο σημείο είναι μικρότερο από την αρνητική τιμή του επιπέδου σημαντικότητας της κατανομής στην οποία ανήκουν τα δεδομένα μας.

Αφού μιλήσαμε πρώτα γενικά για τον έλεγχο υποθέσεων ας προχωρήσουμε τώρα στην διατύπωση των ελέγχων για τις παραμέτρους $\mathbf{b_0}$ και $\mathbf{b_1}$ καθώς και των τύπων που μας δίνουν τα κρίσιμα σημεία.

Πάντα χρησιμοποιούμε τον αμφίπλευρο έλεγχο και για τις δύο παραμέτρους και ως στατιστική ελέγχου την **t**. Έτσι λοιπόν έχουμε :

- Για την παράμετρο $\mathbf{b_0}$ με επίπεδο σημαντικότητας **α** έχουμε:

$$\mathbf{H_0 : b_0 = 0 \quad v.s \quad H_1 : b_0 \neq 0}$$

$$\mathbf{t_{n-2} = \frac{\hat{b}_0 - 0}{S_{\hat{b}0}}} \qquad (2.17)$$

όπου $\hat{S}_{b0} = S\sqrt{\dfrac{\Sigma X^2}{n\Sigma X^2}}$ (2.18)

και $S = \dfrac{\Sigma(X-\bar{X})^2\,\Sigma(Y-\bar{Y})^2 - [\Sigma(X-\bar{X})(Y-\bar{Y})]^2}{\Sigma(X-\bar{X})^2}$ (2.19)

Η μηδενική υπόθεση απορρίπτεται όταν $|\mathbf{t}| \geq \mathbf{t}_{n-2,\, a/2}$. Σε αυτή την περίπτωση η γραμμή παλινδρόμησης δεν περνάει από την αρχή των αξόνων.

- Για την παράμετρο $\mathbf{b_1}$ με επίπεδο σημαντικότητας **α** έχουμε:

$$\mathbf{H_0 : b_1 = 0 \quad v.s \quad H_1 : b_1 \neq 0}$$

$$\mathbf{t_{n-2} = \frac{\hat{b}_1 - 0}{\hat{S}_{b1}}} \qquad (2.20)$$

όπου $\hat{S}_{b1} = S\sqrt{\dfrac{1}{\Sigma X^2}}$ (2.21)

Η μηδενική υπόθεση απορρίπτεται όταν $|\mathbf{t}| \geq \mathbf{t}_{n-2,\, a/2}$. Η απόρριψη της μηδενικής υπόθεσης σημαίνει ότι υπάρχει σχέση ανάμεσα στις μεταβλητές **X** και **Y**. Δεχόμαστε δηλαδή ότι η μεταβλητή **X** είναι σημαντική στην ερμηνεία της συμπεριφοράς της **Y**.

2.5 Ο ΣΥΝΤΕΛΕΣΤΗΣ ΠΡΟΣΔΙΟΡΙΣΜΟΥ

1. Ο δείκτης ή αλλιώς ο συντελεστής προσδιορισμού χρησιμοποιείται για τον έλεγχο της καλής προσαρμογής της εκτιμήτρια της ευθείας παλινδρόμησης στα ζεύγη των δεδομένων μας. Ο δείκτης αυτός συμβολίζεται με $\mathbf{R^2}$ και σύμφωνα από το βιβλίο του Δ.Π. Ψωϊνού (1999), « Στατιστική », Θεσσαλονίκη : Εκδόσεις Ζήτη, δίνεται από τον τύπο :

$$\mathbf{R^2 = \frac{1 - \sigma^2}{\sigma_Y^{\,2}}} \qquad (2.21)$$

όπου $$\mathbf{\sigma_Y^{\,2} = \frac{\Sigma Y_i^{\,2}}{N} - \mu_y^{\,2}} \qquad (2.22)$$

Ο δείκτης $\mathbf{R^2}$ είναι καθαρός αριθμός και μπορεί να χρησιμοποιηθεί για συγκρίσεις. Παίρνει τιμές από **0 έως 1**. Όταν η τιμή του $\mathbf{R^2}$ τείνει προς τη μονάδα τόσο πιο καλή είναι η προσαρμογή της εκτιμήτριας ευθείας και αν $\mathbf{R^2 = 1}$ τότε η ευθεία περνάει από όλα τα σημεία $(\mathbf{X_i}, \mathbf{Y_i})$ του διαγράμματος διασποράς. Ο συντελεστής προσδιορισμού δείχνει το ποσοστό της εξαρτημένης μεταβλητής (**Y**) που ερμηνεύεται από τις μεταβολές της ανεξάρτητης μεταβλητής (**X**). Αν για παράδειγμα το $\mathbf{R^2}$ ήταν ίσο με 0,85 τότε αυτό θα σήμαινε ότι το 85% της συνολικής μεταβλητότητας της εξαρτημένης μεταβλητής **Y** οφείλεται στη σχέση που υπάρχει στις μεταβλητές **X** και **Y** και ότι μόνο το υπόλοιπο 25% της διακύμανσης της μεταβλητής **Y** οφείλεται σε άλλες άγνωστες αιτίες που θεωρούνται μη ελεγχόμενοι παράγοντες.

2.6 Ο ΣΥΝΤΕΛΕΣΤΗΣ ΓΡΑΜΜΙΚΗΣ ΣΥΣΧΕΤΙΣΗΣ

Ο συντελεστής συσχέτισης μας δείχνει το βαθμό συσχέτισης που υπάρχει ανάμεσα στις δύο μεταβλητές και όχι τον τρόπο εξάρτησης και συμβολίζεται με **ρ**. Με **r** συμβολίζεται ο δειγματικός συντελεστής συσχέτισης. Ο συντελεστής αυτός είναι καθαρός αριθμός όπως και ο συντελεστής προσδιορισμού που είδαμε πιο πριν. Ο συντελεστής συσχέτισης θα μας δείξει αν η σχέση είναι έντονη, μέτρια ή ανύπαρκτη και ορίζεται ως ο μέσος γεωμετρικός

όρος των δύο συντελεστών $\hat{b}_1$ και $\hat{b}'_1$ των εξισώσεων παλινδρόμησης :

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 X \quad \text{και} \quad \hat{X} = \hat{b}'_0 + \hat{b}'_1 Y$$

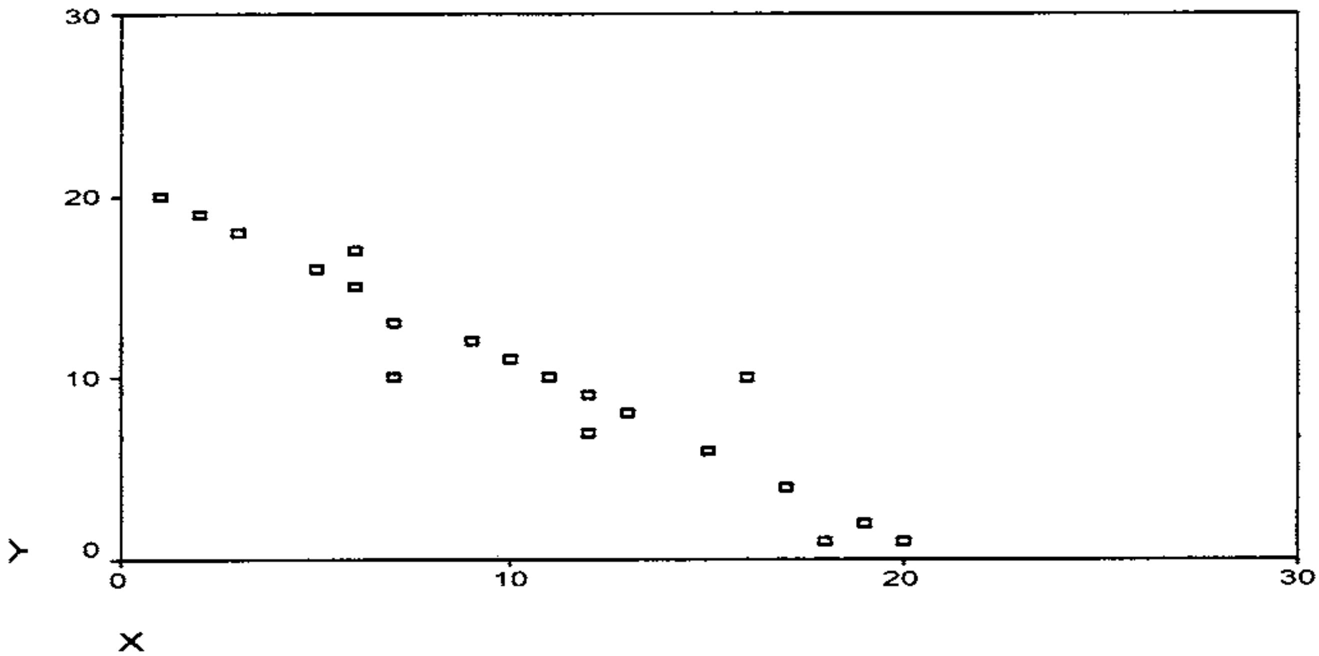
Ο συντελεστής συσχέτισης έχει το πρόσημο + όταν οι παραπάνω συντελεστές είναι ομόσημοι και έχει το πρόσημο – όταν είναι ετερόσημοι. Επίσης πρέπει να αναφερθεί ότι ο συντελεστής προσδιορισμού ισούται με τον συντελεστή γραμμικής συσχέτισης ($R^2 = \rho^2$). Ο τύπος που μας δίνει το **ρ** είναι :

$$\rho = \frac{n(\Sigma X_i Y_i) - (\Sigma X_i)(\Sigma Y_i)}{\sqrt{[n(\Sigma X_i^2) - (\Sigma X_i)^2] * [n(\Sigma Y_i^2) - (\Sigma Y_i)^2]}} \qquad (2.23)$$

Ο συντελεστής συσχέτισης έχει τις εξής ιδιότητες : α) ότι πρόσημο έχει ο συντελεστής παλινδρόμησης **β** το ίδιο πρόσημο έχει και το **ρ**, β) η τιμή του συντελεστή συσχέτισης κυμαίνεται από **-1 έως +1** και γ) ο συντελεστής συσχέτισης προσδιορίζει όπως είπαμε την ένταση της εξάρτησης όταν η σχέση εξάρτησης είναι γραμμική ενώ δεν είναι κατάλληλο στατιστικό μέτρο συσχέτισης στην περίπτωση καμπυλόγραμμης σχέσης.

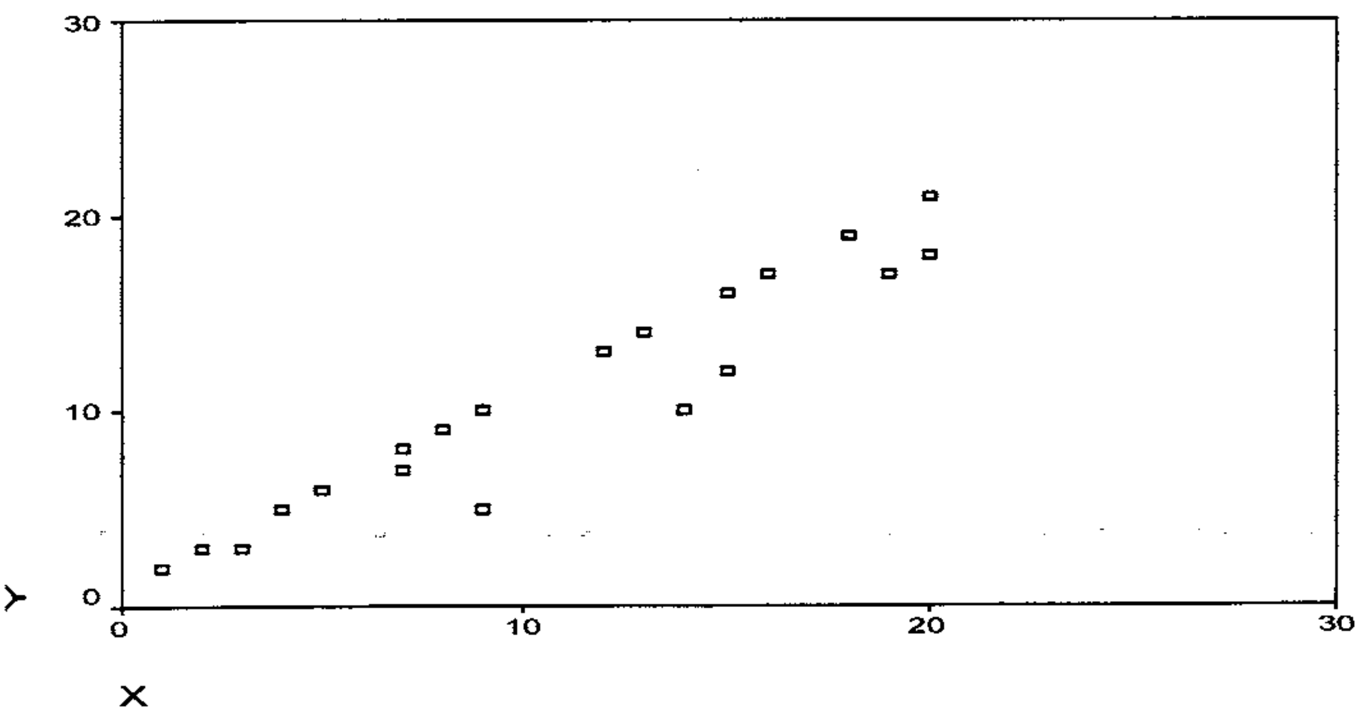
Στη συνέχεια θα δούμε μερικά διαγράμματα που μας δείχνουν τη συσχέτιση που υπάρχει ανάμεσα στις μεταβλητές **X** και **Y**.

Διάγραμμα Διασποράς (2.2)



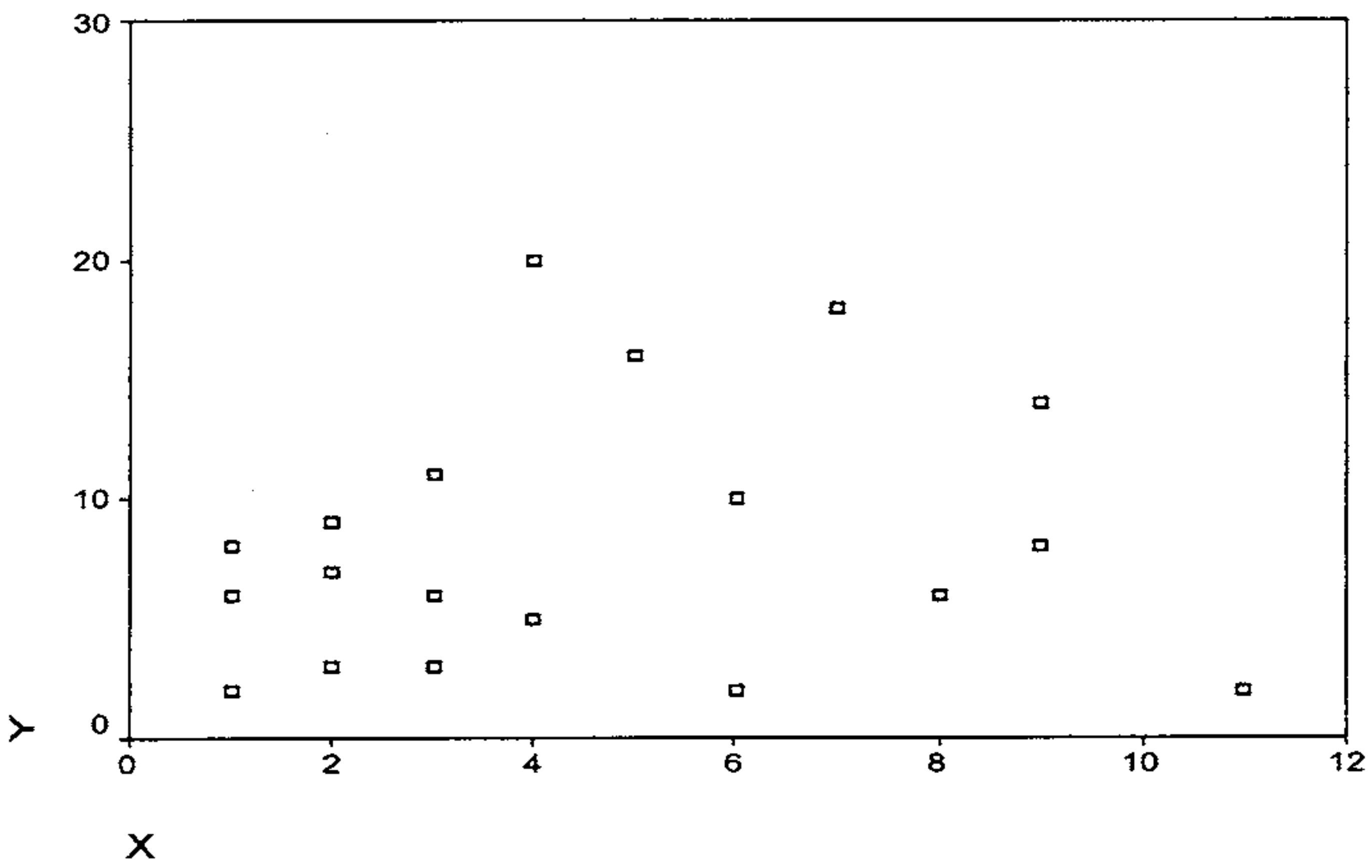
(a) **r < 0**

Διάγραμμα Διασποράς (2.3)



(b) **r > 0**

Διάγραμμα Διασποράς (2.4)



(c) **r = 0**

2.7 ΠΡΟΒΛΕΨΗ

Τώρα θα ασχοληθούμε με τους τρόπους με τους οποίους μπορούμε να προβλέψουμε τιμές της εξαρτημένης μεταβλητής. Όμως πρώτα θα πούμε κάποια πράγματα που θα πρέπει να προσέχουμε πριν προχωρήσουμε στην πρόβλεψη.

Κάθε πρόβλεψη που βασίζεται σε ένα μοντέλο παλινδρόμησης είναι μία πρόβλεψη υπό όρους αφού η πρόβλεψη της εξαρτημένης μεταβλητής **Y** βασίζεται στην τιμή της ανεξάρτητης μεταβλητής **X**. Έστω ότι έχουμε προσδιορίσει το κατάλληλο μοντέλο παλινδρόμησης για τη σχέση μεταξύ δύο μεταβλητών, για να προχωρήσουμε πρέπει να γνωρίζουμε την μελλοντική τιμή της μεταβλητής **X**.

Ο προσδιορισμός της γραμμής παλινδρόμησης γίνεται με τη βοήθεια δεδομένων που προέρχονται από το παρελθόν. Αν αλλάξει κάτι στη σχέση που υπάρχει ανάμεσα στις δύο μεταβλητές τότε η γραμμή παλινδρόμησης χάνει πια τη χρησιμότητά της. Αυτό σημαίνει ότι πρέπει πάντα να εξετάζουμε πολύ καλά και να είμαστε σίγουρη για τη σχέση που συνδέει τις δύο μεταβλητές.

Τέλος θα πρέπει να επιβεβαιώσουμε αν η σχέση που συνδέει τις δύο μεταβλητές είναι σχέση αιτίας και αποτελέσματος έτσι ώστε να μπορεί να γίνει και θεωρητική τεκμηρίωση των αποτελεσμάτων που θα προκύψουν.

Τώρα που αναφέραμε τα παραπάνω μπορούμε να προχωρήσουμε και να δούμε πως μπορεί να γίνει η πρόβλεψη των τιμών της εξαρτημένης μεταβλητής **Y**. Ας υποθέσουμε λοιπόν ότι έχουμε την εξής γραμμή παλινδρόμησης του δείγματος $\hat{Y}_t = \hat{b}_0 + \hat{b}_1 X_t$ η οποία είναι εκτιμήτρια της γραμμής παλινδρόμησης στην πληθυσμό $E(Y_t) = \beta_0 + \beta_1 X_t$ Με άλλα λόγια η $\hat{Y}_t$ είναι εκτιμητής του υπό συνθήκη μέσου της Y_t δεδομένης της τιμής της X_t. Επίσης επειδή $E(Y_t) = E(b_0 + b_1 X_t) = \beta_0 + \beta_1 X_t$, αυτό σημαίνει ότι είναι αμερόληπτος εκτιμητής του $E(Y_t)$. Η διακύμανση του Y_t είναι :

$$V(Y_t) = \sigma^2 \left[\frac{1}{n} + \frac{(X_t - \bar{X})^2}{\Sigma X^2} \right] \quad (2.24)$$

Επειδή όμως το $\hat{Y}_t$ είναι γραμμική συνάρτηση των τυχαίων μεταβλητών u_t που έχουμε υποθέσει ότι κατανέμονται κανονικά και ανεξάρτητα μεταξύ τους αυτί συνεπάγεται ότι και το $\hat{Y}_t$ θα ακολουθεί την κανονική κατανομή με μέσο $\beta_0 + \beta_1 X_t$ και διακύμανση $\sigma^2_{\hat{Y}t}$

Άρα έχουμε :

$$\frac{\hat{Y}_t - (\beta_0 + \beta_1 X_t)}{\sigma_{\hat{Y}t}} \sim N(0,1) \quad (2.25)$$

Στη θέση του της άγνωστης διακύμανσης σ^2 βάζουμε την αμερόληπτη εκτίμηση της S^2 που είναι :

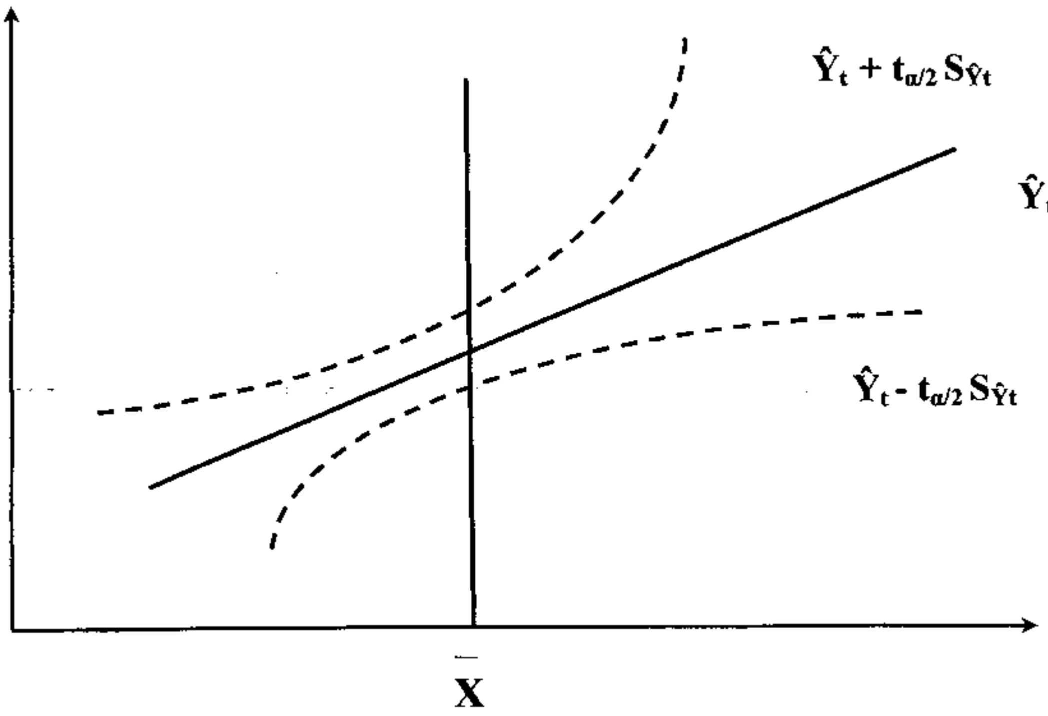
$$S_{\hat{Y}t} = S\sqrt{\frac{1}{n} + \frac{(X-\bar{X})^2}{\Sigma X^2}} \qquad (2.26)$$

Το διάστημα εμπιστοσύνης **1-α** για την προσδοκώμενη τιμή της **Y** δίνεται από την παρακάτω σχέση:

$$\hat{Y}_t - t_{\alpha/2}\, S_{\hat{Y}t} \leq E(Y_t) \leq \hat{Y}_t + t_{\alpha/2}\, S_{\hat{Y}t} \qquad (2.27)$$

Όσο απομακρύνεται η X_t από το μέσο **X**, τόσο πιο μεγάλο είναι το τυπικό σφάλμα της $\hat{Y}_t$ και αυτό σημαίνει κατ'επέκταση ότι και το πλάτος του διαστήματος εμπιστοσύνης θα είναι πιο μεγάλο. Το τυπικό σφάλμα της $\hat{Y}_t$ παίρνει την ελάχιστη τιμή του όταν το X_t ισούται με το μέσο X_t. Γραφικά αυτό μπορούμε να το δούμε στο παρακάτω διάγραμμα.

Διάγραμμα (2.5)



Μπορούμε να χρησιμοποιήσουμε τα δεδομένα από το παράδειγμα που είδαμε πιο πάνω με τις αξίες εισαγωγών καταναλωτικών αγαθών και το διαθέσιμο εισόδημα για να μπορέσουμε να βρούμε το 95% διάστημα εμπιστοσύνης της $\mathbf{E(Y_t)}$. Έτσι λοιπόν από τα δεδομένα του πίνακα (2.1) το διαθέσιμο εισόδημα για το έτος 2000 είναι 240,471 και η υπολογισμένη τιμή της των αξιών εισαγωγικών αγαθών είναι 10,158. Κάνοντας αντικατάσταση στους τύπους (2.25), (2.26), (2.27) προκύπτει ότι για **α = 5%,** η τιμή του **t** από τους στατιστικούς πίνακες είναι 2,145 και το διάστημα εμπιστοσύνης που αντιστοιχεί στην $\mathbf{E(Y_t)}$ όταν το $\mathbf{X_t}$ ισούται με 240,471 είναι το εξής :

$$\mathbf{10{,}158 - 2{,}145 * 0{,}250 \leq E(Y_t) \leq 10{,}158 + 2{,}145 * 0{,}250}$$
$$\mathbf{9{,}622 \leq E(Y_t) \leq 10{,}694}$$

Τώρα θα προχωρήσουμε ακόμα παρακάτω και θα υποθέσουμε ότι η τιμή της ανεξάρτητης μεταβλητής για μία χρονική περίοδο είναι $\mathbf{X_f}$ και ότι θέλουμε να προβλέψουμε την τιμή $\mathbf{Y_f}$. Έστω ότι θεωρούμε πως ισχύει η ίδια γραμμική σχέση που συνέδεε τις δυο μεταβλητές και κατά την περίοδο της πρόβλεψης.

$$\mathbf{Y_f = \beta_0 + \beta_1 X_f + u_f} \qquad (2.28)$$

Αν λοιπόν γνωρίζαμε την γραμμή παλινδρόμησης στον πληθυσμό καθώς και τον τυχαίο όρο θα μπορούσαμε να βρούμε την $\mathbf{Y_f}$. Επειδή η $\mathbf{Y_f}$ είναι μία τυχαία μεταβλητή ακόμα και αν γνωρίζαμε την παλινδρόμηση του πληθυσμού, η πρόβλεψη θα ήταν πρόβλεψη του μέσου της $\mathbf{Y_f}$

$$\mathbf{E(Y_f) = \beta_0 + \beta_1 X_f} \qquad (2.29)$$

Άρα όταν χρησιμοποιούμε τη γραμμή παλινδρόμησης του δείγματος η $\hat{Y}_f = \hat{\beta}_0 + \hat{\beta}_1 X_f$ είναι η καλύτερη εκτίμηση της $E(Y_f)$. Όταν όμως χρησιμοποιούμε την $\hat{Y}_f$ σαν εκτίμηση της Y_f κάνουμε σφάλμα πρόβλεψης που το συμβολίζουμε με **f**.

$$F = Y_f - \hat{Y}_f \qquad (2.30)$$

Από όλα τα παραπάνω μπορούμε να διαπιστώσουμε ότι το σφάλμα πρόβλεψης είναι συνάρτηση των εκτιμητών των β_0 και β_1 και του διαταρακτικού όρου u_f. Οι εκτιμητές όμως των β_0 και β_1 είναι κατ'επέκταση γραμμικές συναρτήσεις των τυχαίων μεταβλητών u_f. Οπότε το σφάλμα πρόβλεψης είναι γραμμικός συνδυασμός των τυχαίων μεταβλητών uf οι οποίες είναι ανεξάρτητες μεταξύ τους όπως έχουμε ήδη πει και ακολουθούν την κανονική κατανομή σύμφωνα με το κεντρικό οριακό θεώρημα. Επομένως και το **f** ακολουθεί την κανονική κατανομή με μέσο μηδέν και διακύμανση σ_f^2 και ισχύει η παρακάτω σχέση :

$$\frac{Y_f - \hat{Y}_f}{\sigma_f} \sim N(0,1) \qquad (2.31)$$

Στη συνέχεια μπορούμε να χρησιμοποιήσουμε την κατανομή **t** για να κατασκευάσουμε το **1- α** διάστημα εμπιστοσύνης για την Y_f, αφού όμως πρώτα βάλουμε στη θέση της σ_f την αμερόληπτη εκτίμησή της, την S_f. Έτσι προκύπτει η παρακάτω σχέση:

$$\hat{Y}_f - t_{\alpha/2} S_{\hat{Y}f} \leq E(Y_f) \leq \hat{Y}_f + t_{\alpha/2} S_{\hat{Y}f} \qquad (2.32)$$

όπου $S_{\hat{Y}f} = S\sqrt{\frac{1}{n} + \frac{(X_f - \bar{X})^2}{\Sigma \bar{X}^2}} \qquad (2.33)$

Χρησιμοποιώντας τα δεδομένα του ίδιου παραδείγματος θα προσπαθήσουμε να βρούμε για **α = 5%** το διάστημα εμπιστοσύνης για τις προβλεπόμενες εισαγωγές καταναλωτικών αγαθών για το έτος 2004 με X_{2004} = 296,657.

$$\hat{Y}_{2004} = 1{,}136 + 0{,}0375 * 296{,}657 = 12{,}250$$

και με αντικατάσταση του τύπου (2.33) έχουμε :

$$S_{\hat{Y}f} = 0{,}379$$

Άρα για το Y_{2004} έχουμε :

$$12{,}250 - 0{,}855 * 2{,}145 \leq Y_{2004} \leq 12{,}250 + 0{,}855 * 2{,}145$$

$$10{,}716 \leq Y_{2004} \leq 14{,}384$$

Το αντίστοιχο διάστημα για την $E(Y_{2004})$ είναι :

$$12{,}250 - 0{,}379 * 2{,}145 \leq E(Y_{2004}) \leq 12{,}250 + 0{,}379 * 2{,}145$$

$$11{,}435 \leq E(Y_{2004}) \leq 13{,}065$$

Έστω πως τώρα θέλουμε να δούμε και κάτι ακόμα. Υποθέτουμε λοιπόν ότι γνωρίζουμε την τιμή της **Y** για την περίοδο πρόβλεψης, δηλαδή θεωρούμε ότι ξέρουμε το ζεύγος (Y^*, X_f) και θέλουμε να βρούμε αν η διαφορά $Y^* - \hat{Y}_f$ είναι στατιστικά σημαντική. Για να το κάνουμε αυτό θα κάνουμε τον παρακάτω έλεγχο υποθέσεων:

$$H_0: Y^* = Y_f$$

$$H_1: Y^* \neq Y_f$$

Ο έλεγχος της παραπάνω υπόθεσης μπορεί να γίνει με την κατανομή t και το κρίσιμο σημείο μπορούμε να το βρούμε από τον εξής τύπο :

$$\mathbf{t^* = \frac{Y^* - \hat{Y}_f}{S_f}} \qquad (2.34)$$

Η μηδενική υπόθεση γίνεται αποδεκτή αν η τιμή της **t** από το δείγμα **t*** είναι μικρότερη από τη θεωρητική τιμή της **t** με **n-2** βαθμούς ελευθερίας. Αποδοχή της μηδενικής υπόθεσης σημαίνει ότι δεν υπάρχει σημαντική διαφορά ανάμεσα στην πραγματική τιμή της εξαρτημένης μεταβλητής και της προβλεπόμενης από το υπόδειγμα. Επίσης αυτό συνεπάγεται ότι η ικανότητα του υποδείγματος για πρόβλεψη είναι καλή. Ακριβώς το αντίθετο συμβαίνει στην περίπτωση που απορρίψουμε τη μηδενική υπόθεση.

2.8 ΣΥΜΠΕΡΑΣΜΑ

Ο σκοπός της απλής παλινδρόμησης είναι να αναλύσει και να εκτιμήσει κυρίως τη σχέση που μπορεί να υπάρχει ανάμεσα σε μία εξαρτημένη μεταβλητή **Y** και μία ανεξάρτητη μεταβλητή **X**. Η ανάλυση αυτή μπορεί να μας δώσει απάντηση στο ερώτημα που αφορά το βαθμό εξάρτησης μεταξύ αυτών των δύο μεταβλητών. Η ύπαρξη υψηλού βαθμού συσχέτισης δε σημαίνει πάντα ότι υπάρχει κάποιος λόγος που υπάρχει η συσχέτιση αυτή για αυτό θα πρέπει να εξετάζουμε πρώτα τη θεωρία και να εξασφαλίζουμε την θεωρητική τεκμηρίωση.

Το μοντέλο της απλής γραμμικής παλινδρόμησης στηρίζεται όπως είδαμε σε κάποιες υποθέσεις που αν παραβιαστούν όλα τα αποτελέσματα και τα τελικά συμπεράσματα δε θα έχουν καμία αξία για την έρευνά μας, ειδικά αν παραβιαστεί η υπόθεση της κανονικότητας στην οποία στηρίζεται η αξιοπιστία των εκτιμηθέντων συντελεστών του υποδείγματος. Η γραφική παράσταση των πραγματικών και των προβλεφθεισών τιμών και των καταλοίπων είναι ένα πολύ χρήσιμο εργαλείο για να διαπιστώσουμε αν υπάρχουν αποκλίσεις από τις υποθέσεις του απλού παλινδρομικού μοντέλου.

Μία γραμμή παλινδρόμησης έχει σημασία αν η κλίση της ευθείας είναι διαφορετική από το μηδέν. Επίσης πολύ σημαντικό ρόλο παίζουν ο συντελεστής προσδιορισμού καθώς και ο συντελεστής συσχέτισης που μας βοηθούν να εκτιμήσουμε καλύτερα το μοντέλο μας.

Τέλος τα διαστήματα εμπιστοσύνης για τη μέση τιμή καθώς και για την ατομική τιμή μπορούν να μας βοηθήσουν να ελέγξουμε τη διασπορά στις εκτιμήσεις και να βγάλουμε ακόμα πιο καλά συμπεράσματα για το υπόδειγμα μας, δηλαδή αν είναι ικανοποιητικό και αν οι προβλέψεις που μπορούμε να κάνουμε με τη βοήθεια του μπορούν να θεωρηθούν έγκυρες και αξιόπιστες.

ΚΕΦΑΛΑΙΟ 3ο

ΠΟΛΛΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

3.1 ΕΙΣΑΓΩΓΗ

Στο προηγούμενο κεφάλαιο ασχοληθήκαμε με την περίπτωση μελέτης των μονάδων ενός πληθυσμού από την άποψη δύο χαρακτηριστικών ιδιοτήτων (**X**, **Y**) και, ειδικότερα, πώς διαμορφώνονται οι τιμές μιας μεταβλητής (της εξαρτημένης) όταν μεταβάλλονται οι τιμές της άλλης μεταβλητής (της ανεξάρτητης).

Στη συνέχεια, θα ασχοληθούμε με τη μελέτη τριών ή περισσοτέρων μεταβλητών με αντικειμενικό σκοπό να εξακριβώσουμε, κατά πρώτο λόγο, πώς μεταβάλλονται οι τιμές της μιας μεταβλητής, που ονομάζεται εξαρτημένη, όταν γνωρίζουμε τις τιμές των άλλων μεταβλητών, που ονομάζονται ανεξάρτητες ή ερμηνευτικές, και, κατά δεύτερο λόγο, να μελετήσουμε την ένταση με την οποία η διαμόρφωση των τιμών μιας μεταβλητής σχετίζεται με τις τιμές και τις μεταβολές των άλλων μεταβλητών που ονομάζονται ανεξάρτητες.

Η σχέση που συνδέει μια εξαρτημένη μεταβλητή με περισσότερες από δύο ανεξάρτητες μεταβλητές ονομάζεται πολλαπλή παλινδρόμηση.

Ενώ η απλή παλινδρόμηση παριστάνεται με μια ευθεία ή μια άλλη καμπύλη, η πολλαπλή παλινδρόμηση παριστάνεται με μια επιφάνεια (επίπεδο) ή με μια υπερεπιφάνεια.

3.2 ΠΡΟΣΔΙΟΡΙΣΜΟΣ ΜΟΝΤΕΛΟΥ

Όπως ήδη έχει προαναφερθεί, η μέθοδος της απλής γραμμικής συσχέτισης και παλινδρόμησης αναφέρεται σε σχέσεις που περιλαμβάνουν μία μόνο ερμηνευτική μεταβλητή, δηλαδή, πρόκειται για μοντέλο δύο μεταβλητών (εξαρτημένη & ανεξάρτητη μεταβλητή) με συναρτησιακή μορφή την ακόλουθη :

$$\mathbf{Y_i = B_1 + B_2X_i} + \varepsilon_i \quad (3.1)$$

Όπου : $\mathbf{Y_i}$ = τιμή εξαρτημένης μεταβλητής

$\mathbf{B_1}$ = σταθερά

$\mathbf{B_2}$ = συντελεστής παλινδρόμησης που περιγράφει την επίδραση της ανεξάρτητης μεταβλητής πάνω στην εξαρτημένη (κλίση)

$\mathbf{X_i}$ = τιμή ανεξάρτητης μεταβλητής

$\boldsymbol{\varepsilon_i}$ = τυπικό σφάλμα, δηλαδή, η διαφορά μεταξύ της πραγματικής τιμής της Yi και της τιμής της πρόβλεψης που προκύπτει από το υπόδειγμα) .

Όμως σε πολλές πραγματικές καταστάσεις η συμπεριφορά των περισσότερων οικονομικών και όχι μόνο μεταβλητών είναι συνάρτηση όχι μιας αλλά πολλών μεταβλητών, άρα θα υπάρχουν περισσότερες από μία ανεξάρτητες μεταβλητές που θα επηρεάζουν την εξαρτημένη μεταβλητή η οποία μας ενδιαφέρει κάθε φορά. Σε αυτές τις περιπτώσεις χρειάζεται να καταφύγουμε σε μια τεχνική που ονομάζεται πολλαπλή παλινδρόμηση. Ο σκοπός της πολλαπλής παλινδρόμησης σύμφωνα με τον Κοσμά Μιλτ. Μαρκάτο , « Στατιστική Επιχειρήσεων », Εκδόσεις ''Ελληνικός Λόγος'', είναι να περιγράψει τη σχέση μεταξύ της εξαρτημένης μεταβλητής **Y** και των **k** ανεξάρτητων μεταβλητών $\mathbf{X_1}$, $\mathbf{X_2}$, $\mathbf{X_3}$.........$\mathbf{X_\kappa}$. Έστω ότι $\mathbf{Y = f(X_1, X_2, \ldots\ldots, X_k)}$, δηλαδή η Y είναι συνάρτηση των k ερμηνευτικών μεταβλητών $\mathbf{X_1, X_2, \ldots\ldots, X_k}$.

Αν υποθέσουμε ότι η παραπάνω συναρτησιακή σχέση είναι γραμμική, για ένα δείγμα από T παρατηρήσεις, μπορούμε να γράψουμε :

$$\mathbf{Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \ldots\ldots + \beta_k X_{tk} + \varepsilon_t} \qquad (3.2)$$

όπου X_{t1} είναι η t παρατήρηση της ερμηνευτικής μεταβλητής X_1, X_{t2} η t παρατήρηση της ερμηνευτικής μεταβλητής X_2, κ.ο.κ. Ο πρώτος δηλαδή δείκτης αναφέρεται στην παρατήρηση και ο δεύτερος στην ερμηνευτική μεταβλητή. Η σχέση αυτή αποτελεί το υπόδειγμα της γραμμικής πολυμεταβλητής παλινδρομήσεως που είναι επέκταση της απλής παλινδρόμησης για περισσότερες από μία ερμηνευτικές μεταβλητές.

Το υπόδειγμα της πολλαπλής γραμμική παλινδρόμησης, βασίζεται πάνω σε κάποιες υποθέσεις όσον αφορά τα τυπικά σφάλματα ε_i (οι υποθέσεις είναι ανάλογες με αυτές της απλής γραμμικής παλινδρόμησης με την προσθήκη μίας ακόμα υπόθεσης) σύμφωνα με τον Damodar Gujarati (1999), « Essentials of Econometrics », 2nd Edition McGraw Hill. :

1) Οι ερμηνευτικές μεταβλητές δεν συσχετίζονται με το τυπικό σφάλμα ε_i.
2) Η αναμενόμενη μέση τιμή του τυπικού σφάλματος είναι μηδέν ($\mathbf{E(\varepsilon_i) = 0}$).
3) Η διακύμανση κάθε ε_i είναι σταθερή ($\mathbf{Var(\varepsilon_i) = \sigma^2}$).
4) Δεν υπάρχει συσχέτιση μεταξύ δύο τυπικών σφαλμάτων ($\mathbf{i \neq j, Cov(\varepsilon_i, \varepsilon_j) = 0}$).
5) Στη συνάρτηση παλινδρόμησης πληθυσμού $\mathbf{Y = B_1 + B_2X_i} + \varepsilon_i$ το τυπικό σφάλμα ε_i ακολουθεί την κανονική κατανομή, έχει μέσο όρο 0 και σταθερή διακύμανση ($\varepsilon_i \sim \mathbf{N(0, \sigma^2)}$).
6) Δεν υπάρχει τέλεια συγγραμμικότητα μεταξύ των $\mathbf{X_1, X_2 \ldots\ldots X_k}$, δηλαδή δεν υπάρχει ακριβής γραμμική σχέση μεταξύ των μεταβλητών.

Ακολουθεί μια απλή εφαρμογή ως παράδειγμα χρήσης της πολλαπλής παλινδρόμησης. Παρακάτω θα αναπτύξουμε τις μεθόδους εκτιμήσεως και ελέγχου του υποδείγματος της πολλαπλής παλινδρόμησης

3.3 ΕΦΑΡΜΟΓΗ

Ας υποθέσουμε ότι έχουμε στη διάθεσή μας παρατηρήσεις για το πλήθος από μπλουζάκια Benetton που έχουν πωληθεί, την τιμή τους, και το κατά κεφαλή εισόδημα των κατοίκων 15 διαφορετικών πόλεων μιας χώρας σε μια συγκεκριμένη χρονική περίοδο. Θα παραστήσουμε την εξαρτημένη μεταβλητή με **Y**, η οποία και θα αντιστοιχεί στην ποσότητα των πωλούμενων μπλουζακίων, και θα έχουμε δύο ανεξάρτητες μεταβλητές: την X_1 που θα παριστάνει την τιμή και την X_2 που θα παριστάνει το εισόδημα.

Πίνακας (3.1) Ζήτηση, Τιμή και Εισόδημα

Πόλη	Y (ζήτηση για μπλουζάκια)	X_1 (τιμή)	X_2 (εισόδημα)
1	166	10	20
2	180	9	21
3	73	10	12
4	81	14	16
5	229	8	24
6	182	15	24
7	233	6	23
8	102	10	15
9	190	7	20
10	150	10	19
11	221	11	25
12	137	15	21
13	173	8	19
14	150	12	20
15	92	10	14

ΠΗΓΗ : Dawning & Clark (1998), « Στατιστική των Επιχειρήσεων », 3η Αμερικανική Έκδοση, Αθήνα : Εκδόσεις Barron's Κλειδάριθμος.

Θα υποθέσουμε ότι η σχέση μεταξύ των **Y**, $\mathbf{X_1}$, και $\mathbf{X_2}$ δίνεται από την ακόλουθη εξίσωση:

$$\mathbf{Y_1 = B_1X_{i1} + B_2X_{i2} + B_3 + e_i}$$

όπου το Y_i αντιπροσωπεύει την *i*-οστή τιμή της εξαρτημένης μεταβλητής και το X_{ij} αντιπροσωπεύει την *i*-οστή παρατήρηση της *j*-οστής ανεξάρτητης μεταβλητής. Αυτή τη φορά χρειαζόμαστε δύο δείκτες γιατί πρέπει να χρησιμοποιήσουμε τον ένα για να παρακολουθούμε τον αριθμό της παρατήρησης και τον άλλο για να παρακολουθούμε τον αριθμό της μεταβλητής.

Στην περίπτωση των παραπάνω παρατηρήσεων, το $\mathbf{X_{11}}$ ισούται με 10, το $\mathbf{X_{21}}$ με 9, το $\mathbf{X_{12}}$ με 20, το $\mathbf{X_{22}}$ με 21, κ.ο.κ.

Οι πραγματικές τιμές των $\mathbf{B_1}$, $\mathbf{B_2}$, και $\mathbf{B_3}$ είναι άγνωστες, αλλά θα προσπαθήσουμε να τις εκτιμήσουμε. Το $\mathbf{B_1}$ αντιπροσωπεύει την επίδραση που έχει στην **Y** η ανεξάρτητη μεταβλητή $\mathbf{X_1}$ με την προϋπόθεση ότι η $\mathbf{X_2}$ παραμένει σταθερή. Παρόμοια, το $\mathbf{B_2}$ αντιπροσωπεύει την επίδραση που έχει στην **Y** η ανεξάρτητη μεταβλητή $\mathbf{X_2}$ όταν η $\mathbf{X_1}$ παραμένει σταθερή. Αν η $\mathbf{X_1}$ αυξηθεί κατά 1 μονάδα και όλα τα άλλα παραμείνουν όπως ήταν, τότε η **Y** θα αυξηθεί κατά $\mathbf{B_1}$. Όταν σχεδιάζουμε το μοντέλο με αυτόν τον τρόπο, υποθέτουμε ότι οι επιδράσεις των $\mathbf{X_1}$ και $\mathbf{X_2}$ στην **Y** είναι αθροιστικές. Αυτό σημαίνει ότι η ποσότητα κατά την οποία επηρεάζει η $\mathbf{X_1}$ την **Y** δεν εξαρτάται από το μέγεθος της $\mathbf{X_2}$, και αντίστροφα. Περιμένουμε ότι το $\mathbf{B_2}$ θα είναι θετικό γιατί, λογικά, οι καταναλωτές θα αγοράζουν περισσότερα μπλουζάκια όταν αυξάνεται το εισόδημά τους, και το $\mathbf{B_1}$ θα είναι αρνητικό γιατί, και πάλι λογικά, η ζήτηση για μπλουζάκια θα είναι μικρότερη όταν ανεβαίνει η τιμή τους. Το $\mathbf{B_3}$ ονομάζεται σταθερός όρος του μοντέλου, και είναι ανάλογος με τον όρο της κατακόρυφης απόστασης του μοντέλου της απλής γραμμικής παλινδρόμησης. Το **e** είναι και πάλι μια τυχαία μεταβλητή που ονομάζεται τυπικό σφάλμα (standard error) και που αντιπροσωπεύει την επίδραση όλων των δυνατών παραμέτρων εκτός της τιμής και του εισοδήματος οι οποίοι ενδέχεται να επηρεάζουν την ζήτηση για μπλουζάκια Benetton. Κατά συνέπεια, το **e** θα πρέπει να είναι ανεξάρτητο από τα X_1 και X_2. Αυτό μπορεί να ελεγχθεί εύκολα αν παραστήσουμε γραφικά τις τιμές του **e** σε συνδυασμό με τις τιμές του X_1 και τις τιμές του **e** σε συνδυασμό με τις τιμές του X_2 για να δούμε αν υπάρχει κάποιο επαναλαμβανόμενο φαινόμενο. Η αναμενόμενη τιμή του **e** είναι **0** και η διακύμανσή του σ^2, η οποία και είναι άγνωστη. Θα υποθέσουμε επίσης ότι το **e** ακολουθεί την κανονική κατανομή.

Αν εισάγουμε τα δεδομένα του πίνακα στον υπολογιστή και με την χρήση κατάλληλου προγράμματος θα μας δοθεί ένα αποτέλεσμα σαν το παρακάτω:

$$\mathbf{Y = -7.738x_1 + 12.286x_2 - 2.765}$$

Αυτή η εξίσωση μας δίνει τις εκτιμώμενες τιμές των συντελεστών. Το πρώτο πράγμα που μπορούμε να διαπιστώσουμε από αυτό το αποτέλεσμα επιβεβαιώνει αυτό που περιμέναμε: ότι ο συντελεστής του X_1 είναι αρνητικό, πράγμα που σημαίνει ότι η υψηλότερη τιμή οδηγεί σε λιγότερες πωλήσεις, και ότι ο συντελεστής του X_2 είναι θετικός, γεγονός που σημαίνει ότι το υψηλότερο εισόδημα οδηγεί σε μεγαλύτερες πωλήσεις. Μπορούμε να χρησιμοποιήσουμε τις εκτιμώμενες τιμές των συντελεστών για να κάνουμε προβλέψεις για την τιμή του **Y**.

Για παράδειγμα, αν είχαμε υπόψη μας μια πόλη στην οποία το μέσο εισόδημα είναι 20 και η τιμή για τα μπλουζάκια Benetton 6, θα μπορούσαμε να προβλέψουμε ότι η ζήτηση για μπλουζάκια θα είναι:

$$\mathbf{(-7{,}738 * 6) + (12{,}286 * 20) - 2{,}765 = 196{,}5}$$

3.4 ΕΚΤΙΜΗΣΗ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΤΗΣ ΠΟΛΛΑΠΛΗΣ ΓΡΑΜΜΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Γενικά θα παριστάνουμε με $\mathbf{b_0}$ την σταθερά, με $\mathbf{b_1}$ την εκτίμηση για τον συντελεστή $\mathbf{B_1}$, με $\mathbf{b_2}$ την εκτίμηση για τον συντελεστή $\mathbf{B_2}$ κ.ο.κ. Έτσι λοιπόν η εκτιμώμενη εξίσωση παλινδρόμησης που θα προκύψει θα έχει την εξής μορφή σύμφωνα με τον Γ. Δρόσο – Δ. Καραπιστόλα (1994), « Στατιστική Επιχειρήσεων », Αθήνα : Εκδόσεις Ελλήν :

$$\hat{\mathbf{Y}} = \mathbf{b_0} + \mathbf{b_1X_1} + \mathbf{b_2X_2} + \ldots\ldots + \mathbf{b_kX_k} \quad (3.3)$$

Οι συντελεστές $\mathbf{b_1}$, $\mathbf{b_2}$, …., $\mathbf{b_k}$ αντιπροσωπεύουν την μερική επίδραση που έχουν οι ανεξάρτητες μεταβλητές πάνω στην εξαρτημένη. Για παράδειγμα, ο συντελεστής $\mathbf{b_1}$ υποδηλώνει την μεταβολή της $\mathbf{Y}$, που θα προκύψει εάν η μεταβλητή $\mathbf{X_1}$ μεταβληθεί κατά μία μονάδα (ανάλογα με την μονάδα μέτρησής της), και οι άλλες ανεξάρτητες μεταβλητές ($\mathbf{X_2}$, $\mathbf{X_3}$, ……, $\mathbf{X_k}$) παραμείνουν σταθερές. Επομένως γι' αυτό το λόγο οι συντελεστές $\mathbf{b_1}$, $\mathbf{b_2}$, ……., $\mathbf{b_k}$ ονομάζονται και συντελεστές μερικής παλινδρόμησης.

Η εκτίμηση της εξίσωσης της γραμμικής παλινδρόμησης θα προκύψει μέσω της μεθόδου των ελαχίστων τετραγώνων, δηλαδή θα προσπαθήσουμε να βρούμε τις τιμές εκείνες των συντελεστών παλινδρόμησης ($\mathbf{b_1}$, $\mathbf{b_2}$, ….., $\mathbf{b_k}$) που ελαχιστοποιούν το άθροισμα των τετραγώνων της απόκλισης μεταξύ των πραγματικών τιμών της Υ και των εκτιμώμενων τιμών της $\hat{\mathbf{Y}}$ που θα προκύψουν από την εξίσωση της παλινδρόμησης.

Ας εξετάσουμε θεωρητικά μια περίπτωση με δύο ανεξάρτητες μεταβλητές $\mathbf{X_1}$ και $\mathbf{X_2}$.

Έχουμε ένα εκτιμώμενο υπόδειγμα της μορφής:

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 \qquad (3.4)$$

Η $\hat{Y}$ είναι η εκτίμηση της **E(Y)** έτσι λοιπόν σύμφωνα με την παραπάνω εξίσωση οι αποκλίσεις μεταξύ των πραγματικών τιμών της **Y** και των τιμών της $\hat{Y}$ συμβολίζονται με

$$e_t = Y_t - b_0 - b_1X_{1t} - b_2X_{2t} \qquad (3.5)$$

Υψώνοντας στο τετράγωνο την παραπάνω εξίσωση και από τις δύο πλευρές και ταυτόχρονα αθροίζοντας τις παρατηρήσεις παίρνουμε το εξής αποτέλεσμα:

$$\Sigma e_t^2 = \Sigma(\,Y_t - b_0 - b_1\, X_{1t} - b_2X_{2t})^2 \qquad (3.6)$$

Παραγωγίζοντας την παραπάνω εξίσωση και εξισώνοντας τις παραγώγους με μηδέν καταλήγουμε σε ένα σύστημα τριών εξισώσεων με τρεις αγνώστους χωρίς να επεκταθούμε περισσότερο κάνοντας τους απαραίτητους υπολογισμούς καταλήγουμε στους εξής τύπους υπολογισμού των παραμέτρων.

$$b_0 = Y - b_1X_1 - b_2X_2 \qquad (3.7)$$

$$b_1 = \frac{(\Sigma y_t x_{1t})(\Sigma x^2_{2t}) - (\Sigma y_t x_{2t})(\Sigma x_{1t}x_{2t})}{(\Sigma x^2_{1t})(\Sigma x^2_{2t}) - (\Sigma x_{1t}x_{2t})^2} \qquad (3.8)$$

$$b_2 = \frac{(\Sigma y_t x_{2t})(\,\Sigma x^2_{1t}) - (\,\Sigma y_t x_{1t})(\Sigma x_{1t}x_{2t})}{(\Sigma x^2_{1t})(\,\Sigma x^2_{2t}) - (\Sigma x_{1t}x_{2t})^2} \qquad (3.9)$$

3.4.1 ΙΔΙΟΤΗΤΕΣ ΤΩΝ ΕΚΤΙΜΗΤΩΝ

Με δεδομένες τις υποθέσεις του κλασσικού γραμμικού υποδείγματος παλινδρόμησης οι εκτιμήσεις που παίρνουμε από την **OLS** έχουν τις ιδιότητες **BLUE** (Best, Linear, Unbiased Estimators) με άλλα λόγια είναι άριστοι, γραμμικοί και αμερόληπτοι όπως συμβαίνει και στην απλή παλινδρόμηση.

1) Άριστοι, σημαίνει πως οι παράμετροί μας έχουν την ελάχιστοι διακύμανση συγκρινόμενες με αυτές που θα παίρναμε από άλλη μέθοδο.
2) Γραμμικοί, σημαίνει ότι θα πρέπει να είναι γραμμικές οι συναρτήσεις της τυχαίας μεταβλητής **Y**.
3) Αμερόληπτη, σημαίνει ότι θα πρέπει να ισχύουν τα εξής: $\mathbf{E(b_1) = B_1}$, $\mathbf{E(b_2) = B_2}$, $\mathbf{E(\hat{s}^2) = s^2}$.

Ωστόσο, θα πρέπει να τονίσουμε ότι για να είναι άριστοι οι εκτιμητές θα πρέπει υποχρεωτικά να είναι αμερόληπτοι και γραμμικοί ως προς **Y**. Επίσης, αν τα e κατανέμονται κανονικά τότε αυτοί συμπίπτουν με τους εκτιμητές μέγιστης πιθανοφάνειας. Οι εκτιμητές μέγιστης πιθανοφάνειας δίνονται από τους ίδιους μαθηματικούς τύπους από τους οποίους εκτιμούνται οι εκτιμητές ελαχίστων τετραγώνων. Αξίζει να τονιστεί ότι οι ιδιότητες των εκτιμητών εξαρτώνται άμεσα από την ισχύ των υποθέσεων που ήδη έχουν αναφερθεί προηγουμένως. Η παραβίαση των υποθέσεων αυτών θα έχει ως αποτέλεσμα να μην διατηρούν οι εκτιμητές τις επιθυμητές ιδιότητές τους.

3.5 ΣΥΝΤΕΛΕΣΤΗΣ ΠΟΛΛΑΠΛΟΥ ΠΡΟΣΔΙΟΡΙΣΜΟΥ R^2

Μετά από κάποιους υπολογισμούς θα μας δοθεί επίσης και μια τιμή $\mathbf{R^2}$ για την παλινδρόμηση. Το $\mathbf{R^2}$ ονομάζεται συντελεστής πολλαπλού προσδιορισμού (coefficient of multiple determination). Η ερμηνεία της τιμής $\mathbf{R^2}$ είναι παρόμοια με την ερμηνεία της τιμής $\boldsymbol{\rho^2}$ μιας απλής γραμμικής παλινδρόμησης. Το $\mathbf{R^2}$ μετρά το ποσοστό της απόκλισης της εξαρτημένης μεταβλητής το οποίο μπορεί να εξηγηθεί από την παλινδρόμηση. Η τιμή του $\mathbf{R^2}$ κυμαίνεται πάντα από **0 έως 1**.

Για να υπολογίσουμε το $\mathbf{R^2}$ θα πρέπει πρώτα να υπολογίσουμε ορισμένα στατιστικά στοιχεία αθροίσματος τετραγώνων σύμφωνα με τον Dawning & Clark (1998), « Στατιστική των Επιχειρήσεων », 3η Αμερικανική Έκδοση, Αθήνα : Εκδόσεις Barron's Κλειδάριθμος. Το συνολικό άθροισμα τετραγώνων (total sum of squares – και για συντομία **TSS**) είναι το άθροισμα των τετραγώνων των τιμών της **y** γύρω από το $\hat{\mathbf{y}}$:

$$\mathbf{TSS = \Sigma(y_i - \bar{y})^2} \qquad (3.10)$$

Σημειώνουμε ότι το **TSS** ισοδυναμεί με την ποσότητα που στην περίπτωση της απλής γραμμικής παλινδρόμησης ονομάζαμε συνολικό τετραγωνικό σφάλμα. Θα παριστάνουμε με $\hat{\mathbf{y}}_i$ την προβλεπόμενη από την παλινδρόμηση τιμή του **y** για την ***i***- οστή παρατήρηση:

$$\mathbf{\hat{y}_i = b_0 + b_1x_1 + b_2x_2 + \ldots\ldots + b_kX_k} \quad (3.11)$$

Για κάθε παρατήρηση μπορούμε να υπολογίσουμε τη διαφορά μεταξύ της τιμής του **y** που προβλέπουμε από τη γραμμή παλινδρόμησης και τη μέση τιμή του **y**.

Στη συνέχεια μπορούμε να αθροίσουμε τα τετράγωνα όλων αυτών των αποκλίσεων και να ονομάσουμε αυτό που θα πάρουμε άθροισμα τετραγώνων παλινδρόμησης (regression sum of squares – και για συντομία **RGRSS**) :

$$\mathbf{RGRSS} = \Sigma(\hat{y}_i - \bar{y})^2$$

Κατόπιν μπορούμε να υπολογίσουμε το υπόλοιπο κάθε παρατήρησης, που δεν είναι παρά η διαφορά μεταξύ της πραγματικής τιμής του y και της προσαρμοσμένης τιμής από τη γραμμή παλινδρόμησης:

$$(\boldsymbol{\upsilon\pi\acute{o}\lambda o\iota\pi o}_i) = \mathbf{y}_i - \hat{\mathbf{y}}_i$$

Το άθροισμα των τετραγώνων όλων των υπολοίπων ονομάζεται άθροισμα τετραγωνικών σφαλμάτων (error sum of squares – και για συντομία **ERSS**):

$$\mathbf{ERSS} = \Sigma(\mathbf{y}_i - \hat{\mathbf{y}}_i)^2$$

Αυτή η ποσότητα είναι ανάλογη με την ποσότητα που στην περίπτωση της απλής γραμμικής παλινδρόμησης ονομάζαμε άθροισμα των τετραγώνων όλων των σφαλμάτων. Εκτελώντας κάποιους υπολογισμούς , μπορούμε να αποδείξουμε ότι:

$$\mathbf{TSS} = \mathbf{RGRSS} + \mathbf{ERSS} \qquad (3.12)$$

Μπορούμε να θεωρήσουμε ότι το **TSS** αντιπροσωπεύει τη συνολική διακύμανση των τιμών του **y**. Το **RGRSS** είναι η ποσότητα αυτής της διακύμανσης η οποία μπορεί να εξηγηθεί από την παλινδρόμηση, και το **ERSS** είναι η ποσότητα της διακύμανσης που απομένει και δεν μπορεί να εξηγηθεί από την παλινδρόμηση. Αν η παλινδρόμηση προσαρμόζεται στα δεδομένα πολύ καλά, η τιμή του **RGRSS** θα είναι πολύ μεγαλύτερη από την τιμή του **ERSS**.

Μπορούμε να υπολογίσουμε την τιμή του $\mathbf{R^2}$ με οποιονδήποτε από τους δύο επόμενους τύπους:

$$\mathbf{R^2 = 1 - \frac{ERSS}{TSS}} \quad \text{ή} \quad \mathbf{R^2 = \frac{RGRSS}{TSS}} \qquad (3.13)$$

Παρόλα αυτά, υπάρχει ένα σημαντικό σημείο που θα πρέπει να έχουμε υπόψη μας κατά την ερμηνεία της τιμής του $\mathbf{R^2}$ μιας πολλαπλής παλινδρόμησης. Υπάρχει πάντα η δυνατότητα να αυξήσουμε την τιμή του $\mathbf{R^2}$ με την προσθήκη στην παλινδρόμηση περισσότερων ανεξάρτητων μεταβλητών, είτε έχουν κάποια σχέση με την εξαρτημένη μεταβλητή είτε όχι. Για να είναι αξιόπιστα τα αποτελέσματα της παλινδρόμησης θα πρέπει το πλήθος των παρατηρήσεων να είναι σημαντικά μεγαλύτερο από το πλήθος των συντελεστών που προσπαθούμε να εκτιμήσουμε. Έτσι, μερικές φορές συνιστάται ο υπολογισμός του προσαρμοσμένου $\mathbf{R^2}$.

$$\mathbf{(\text{προσαρμοσμένο } R^2) = 1 - \frac{(1 - R^2)(n - 1)}{n - k}} \qquad (3.14)$$

Το προσαρμοσμένο $\mathbf{R^2}$ είναι σταθμισμένο με τους βαθμούς ελευθερίας κάθε αντίστοιχου αθροίσματος τετραγώνων. Άρα το προσαρμοσμένο $\mathbf{R^2}$ αυξάνεται (βέβαια με την προσθήκη άλλης μιας μεταβλητής δεν αυξάνεται αναγκαστικά διότι κάτι τέτοιο θα αύξανε την τιμή του **k**) ή μειώνεται ανάλογα με το αν η συμβολή της νέας ερμηνευτικής μεταβλητής υπεραντισταθμίζει ή όχι την απώλεια των βαθμών ελευθερίας. Αποδεικνύεται επίσης ότι το προσαρμοσμένο $\mathbf{R^2}$ μπορεί να πάρει και αρνητικές τιμές όταν $\mathbf{R^2 < \frac{k}{n - 1}}$.

*(ο περιορισμός $0 < R^2 < 1$ δεν ισχύει όταν δεν υπάρχει σταθερός όρος στην εξίσωση)

Παρότι υπάρχουν συγκεκριμένα θεωρητικά υποδείγματα των οποίων η οικονομετρική τους εξειδίκευση δεν περιέχει σταθερό όρο, θα πρέπει να θεωρήσουμε τη χρήση του σταθερού όρου στην παλινδρόμηση ως υποχρεωτική. Εξάλλου, ο αποκλεισμός του από το προς εκτίμηση υπόδειγμα σημαίνει και μια αυθαίρετη επιβολή του συγκεκριμένου μηδενικού περιορισμού, που κανονικά θα πρέπει να ελέγξουμε για να αποφανθούμε σχετικά με την ισχύ του.

3.6 ΟΙ ΣΥΝΤΕΛΕΣΤΕΣ ΜΕΡΙΚΗΣ ΣΥΣΧΕΤΙΣΗΣ

Μέχρι στιγμής έχουμε δει τους συντελεστές απλής συσχέτισης r_{yxk} οι οποίοι μας δίνουν το μέτρο της γραμμικής συσχέτισης μεταξύ δύο μεταβλητών. Στην πολλαπλή παλινδρόμηση μας ενδιαφέρει να γνωρίζουμε και τον βαθμό γραμμικής συσχέτισης μεταξύ δύο μεταβλητών όταν έχει ληφθεί όμως υπόψη και η επίδραση κάποιων άλλων μεταβλητών. Οι συντελεστές μερικής συσχέτισης (l – 1) τάξης μας δίνουν το βαθμό συσχέτισης μεταξύ του **Y** και του X_k μετά την εισαγωγή (δηλαδή αφού έχει ληφθεί υπόψη η επίδραση) των μεταβλητών. Άρα, για παράδειγμα, ο συντελεστής μερικής συσχέτισης πρώτης τάξης $r_{yxi \cdot xj}$ μετράει τη συσχέτιση **Y** και X_i αφού έχει εισαχθεί και η (μια μεταβλητή) X_j.

Οι συντελεστές μερικής συσχέτισης παίζουν σημαντικό ρόλο στην απόφασή μας για εισαγωγή ή όχι και άλλων ερμηνευτικών μεταβλητών. Αν, για παράδειγμα, ο συντελεστής απλής συσχέτισης r_{yx2} είναι πολύ υψηλός, ενώ ο συντελεστής μερικής συσχέτισης $r_{yx2 \cdot x1}$ είναι πολύ χαμηλός, αυτό σημαίνει ότι, ενώ μόνη της η X_2 ερμηνεύει πολύ ικανοποιητικά την **Y**, αν εισαχθεί και η X_1 τότε η X_2 δεν βοηθά πλέον σημαντικά την ερμηνεία της **Y**. Στην περίπτωση αυτή, το γεγονός ότι η X_2 ερμηνεύει μόνη της ικανοποιητικά την **Y** οφείλεται στο ότι η Y και η X_2 διαμορφώνονται από την X_1, και συνεπώς δεν υπάρχει λόγος να διατηρήσουμε την X_2 στην παλινδρόμηση, μετά την εισαγωγή της X_1.

3.6.1 Ο ΣΥΝΤΕΛΕΣΤΗΣ ΜΕΡΙΚΟΥ ΠΡΟΣΔΙΟΡΙΣΜΟΥ

Όμως, εκτός από τον συντελεστή μερικής συσχέτισης, για να μετρήσουμε και να εκτιμήσουμε την συνεισφορά κάθε ανεξάρτητης μεταβλητής στην ερμηνεία των μεταβολών της Y έχουμε και τον συντελεστή μερικού προσδιορισμού $r^2_{yxk\ .xj}$, όπου δεν είναι τίποτε άλλο από το τετράγωνο του συντελεστή μερικής συσχέτισης.

Αναλυτικότερα, εφόσον η Y εκφράζεται ως γραμμική συνάρτηση της X_i που ερμηνεύει ένα ποσοστό της συνολικής διασποράς της Y, μένει ένα υπόλοιπο ανερμήνευτο. Εάν, τώρα προσθέσουμε στο υπόδειγμά μας και μια νέα μεταβλητή, την X_j, θα έχουμε ένα ποσοστό ανερμήνευτης διασποράς που θα ερμηνεύσει η νέα μεταβλητή.

Έτσι, λοιπόν το ποσοστό αυτό της ανερμήνευτης διασποράς της μεταβλητής X_j, ονομάζεται συντελεστής μερικού προσδιορισμού. Όμως, επειδή η X_j συνήθως συσχετίζεται με την Y όσο και με την X_i, για να μετρήσουμε την πραγματική συνεισφορά της στο ανερμήνευτο από την X_i μέρος της διασποράς της Y, πρέπει πρώτα να αφαιρέσουμε την επίδραση της X_i και από την Y και από την X_j για να προκύψει η πραγματική επίδραση της X_j, χωρίς να επηρεάζεται από την επίδραση της X_i.

Ο συμβολισμός του συντελεστή μερικού προσδιορισμού όπως ήδη έχει προαναφερθεί είναι $r^2_{yxk\ .xj}$. Ο τρόπος υπολογισμού του συντελεστή μερικού προσδιορισμού δεν έχει νόημα να αναφερθεί εξαιτίας της πολυπλοκότητάς του, εξάλλου στην καθημερινή ανάλυση των δεδομένων υιοθετούμε εξειδικευμένα στατιστικά προγράμματα που κάνουν όλους τους απαραίτητους υπολογισμούς.

Έτσι, λοιπόν η χρησιμότητα του συντελεστή μερικού προσδιορισμού είναι για να βοηθήσει στην από πριν εκτίμηση του πόσο θα μειωθεί το ανερμήνευτο μέρος των μεταβολών της Y, εάν προστεθεί στο υπόδειγμα παλινδρόμησης μια νέα μεταβλητή. Έτσι, μπορούμε να επιλέξουμε, μεταξύ μιας ομάδας ανεξάρτητων μεταβλητών, εκείνες που θα συνεισφέρουν περισσότερο στη μείωση της ανερμήνευτης μεταβλητότητας της εξαρτημένης μεταβλητής Y. Αυτός ο τρόπος επιλογής των ανεξάρτητων μεταβλητών ονομάζεται πολλαπλή παλινδρόμηση με διαδοχική επιλογή των ανεξάρτητων μεταβλητών, σε αυτό όμως θα αναφερθούμε αργότερα παρακάτω.

3.7 ΈΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ

Οι έλεγχοι στατιστικής σημαντικότητας στην ανάλυση της πολλαπλής παλινδρόμησης έχουν ως σκοπό να ελέγξουν αν η εξίσωση της παλινδρόμησης, στο σύνολό της, εξηγεί ένα σημαντικό μέρος των μεταβολών της εξαρτημένης μεταβλητής **Y** και εφόσον έχουμε περισσότερες από μία ανεξάρτητες μεταβλητές, να ελέγξουμε τη σημαντικότητα των συντελεστών παλινδρόμησης. Έτσι διαπιστώνουμε ποιες από τις ανεξάρτητες μεταβλητές μας ασκούν σημαντική επίδραση στην **Y** και ποιες δεν ασκούν.

1. Ο έλεγχος στατιστικής σημαντικότητας της εξίσωσης παλινδρόμησης ταυτίζεται με τον έλεγχο της στατιστικής σημαντικότητας του συντελεστή πολλαπλού προσδιορισμού $\mathbf{R^2}$. Με άλλα λόγια, θα ελέγξουμε αυτό που μετρά ο συντελεστής πολλαπλού προσδιορισμού, δηλαδή το ποσοστό των μεταβολών της **Y** που οφείλεται στις επιδράσεις που ασκούν οι ανεξάρτητες μεταβλητές $\mathbf{X_1}$, $\mathbf{X_2}$, ….., $\mathbf{X_k}$, και το οποίο είναι διάφορο του μηδενός. Έτσι, λοιπόν η μηδενική υπόθεση ($\mathbf{H_0}$) και η εναλλακτική υπόθεση ($\mathbf{H_1}$) διατυπώνονται ως εξής σύμφωνα με τον Ιωάννη Γ. Χαλικιά (2001), « Στατιστική – Μέθοδοι για Επιχειρηματικές Αποφάσεις », Αθήνα : Εκδόσεις Rosili.:

Υπόθεση μηδέν (H_0) : Υποθέτουμε ότι η εξίσωση παλινδρόμησης δεν εξηγεί καθόλου τις μεταβολές της **Y** (το ποσοστό της εξηγημένης διασποράς της **Y** είναι μηδέν) και επομένως $\mathbf{b_1 = b_2 = \ldots = b_k = 0}$.

Εναλλακτική υπόθεση (H_1) : Υποθέτουμε ότι η εξίσωση παλινδρόμησης εξηγεί ένα μέρος των μεταβολών της **Y** (το ποσοστό της εξηγημένης διασποράς της **Y** είναι μεγαλύτερο του μηδενός) και επομένως ένας από τους συντελεστές $\mathbf{b_i \neq 0}$.

Επομένως, θα συγκρίνουμε τις δύο συνιστώσες της **TSS**, την εξηγημένη (**RSS**) και την ανεξήγητη (**ESS**). Εάν η εξηγημένη (**RSS**) είναι σημαντικά μεγαλύτερη από την ανεξήγητη (**ESS**), αυτό σημαίνει ότι η επίδραση της εξίσωσης παλινδρόμησης είναι σημαντική. Ενώ αν συμβαίνει το αντίθετο, δηλαδή, η ανεξήγητη (**ESS**) είναι σημαντικά μεγαλύτερη από την εξηγημένη (**RSS**), σημαίνει ότι το ποσοστό της **TSS** που περιγράφεται από την εξίσωση είναι αμελητέο.

Τα **RSS** και **ESS** όπως ήδη έχει αναφερθεί σε προηγούμενο κεφάλαιο είναι αθροίσματα τετραγώνων αποκλίσεων, που όμως βασίζονται σε διαφορετικό αριθμό βαθμών ελευθερίας. Για αυτό το λόγο, η σύγκριση μεταξύ τους θα γίνει αφού διαιρεθούν με τους αντίστοιχους βαθμούς ελευθερίας (d.f. – degrees of freedom). Οι λόγοι που θα προκύψουν ονομάζονται μέσα τετράγωνα (MS – mean squares) και ο έλεγχος μεταξύ τους βασίζεται στην κατανομή F (υπάρχουν ειδικοί στατιστικοί πίνακες με αποτελέσματα που έχουν ως σκοπό την διευκόλυνση του αναλυτή).

Εάν η τιμή $\mathbf{F_{k,n-k-1}}$ είναι μεγαλύτερη της κριτικής τιμής $\mathbf{F_{(k,n-k-1),\alpha}}$ (όπου $\boldsymbol{\alpha}$ = επίπεδο σημαντικότητας), απορρίπτεται η μηδενική υπόθεση και αντίστροφα. Ο λόγος **ESS/(n-k-1)** ονομάζεται μέσο τετραγωνικό σφάλμα (mean square error) και συμβολίζεται με MSE. Έτσι λοιπόν το μέσο τετραγωνικό σφάλμα είναι το τετράγωνο του τυπικού σφάλματος εκτίμησης και ισούται με $\mathbf{s_e^2}$.

Τώρα, αν ο πρώτος έλεγχος της στατιστικής σημαντικότητας του συντελεστή πολλαπλού προσδιορισμού $\mathbf{R^2}$ δείξει ότι η εξίσωση παλινδρόμησης εξηγεί ένα σημαντικό μέρος των μεταβολών της εξαρτημένης μεταβλητής **Y**, το επόμενο βήμα είναι να ελέγξουμε τη στατιστική σημαντικότητα των συντελεστών μερικής παλινδρόμησης $\mathbf{b_i}$, (όπου i = 1, 2, ….**k**). Επειδή οι συντελεστές $\mathbf{b_i}$ υπόκεινται στα σφάλματα δειγματοληψίας θα πρέπει να γνωρίζουμε όχι μόνο αν οι συντελεστές αυτοί είναι διάφοροι του μηδενός αλλά και σε ποιο διάστημα εμπιστοσύνης βρίσκονται οι τιμές των συντελεστών μερικής παλινδρόμησης του πληθυσμού. Εξάλλου ως γνωστόν οι συντελεστές παλινδρόμησης $\mathbf{b_i}$ είναι εκείνοι που μας περιγράφουν τη σχέση εξάρτησης της **Y** με τις μεταβλητές **X**.

Το τυπικό σφάλμα της κατανομής δειγματοληψίας του συντελεστή $\mathbf{b_i}$ συμβολίζεται με $\mathbf{s_{bi}}$ άρα η μηδενική υπόθεση ($\mathbf{H_0}$) και η εναλλακτική υπόθεση ($\mathbf{H_1}$) διατυπώνονται ως εξής:

Υπόθεση μηδέν (H_0) : Υποθέτουμε ότι ο συντελεστής παλινδρόμησης ισούται με τον συντελεστή μερικής παλινδρόμησης, δηλαδή, $\mathbf{b_i = b_i^*}$.

Εναλλακτική υπόθεση (H_1) : Υποθέτουμε ότι ο συντελεστής παλινδρόμησης είναι διάφορος από τον συντελεστή μερικής παλινδρόμησης, δηλαδή, $\mathbf{b_i \neq b_i^*}$.

Ο έλεγχος γίνεται με το γνωστό κριτήριο **t** και **n-k-1** βαθμούς ελευθερίας και υπολογίζεται μέσω του τύπου:

$$\mathbf{t_{n-k-1} = \frac{| b_i - b_i^* |}{s_{bi}}} \qquad (3.15)$$

Εάν η τιμή | $\mathbf{t_{n\text{-}k\text{-}1}}$ | είναι μεγαλύτερη της κριτικής τιμής | $\mathbf{t_{n\text{-}k\text{-}1,\ \alpha/2}}$ |, απορρίπτεται η μηδενική υπόθεση και αντίστροφα. Επομένως, για επίπεδο σημαντικότητας **α**, η κριτική τιμή θα αντιστοιχεί στο κάτω ή το άνω **α/2** της αντίστοιχης κατανομής **t** με **n-k-1** βαθμούς ελευθερίας. Θα πρέπει να τονίσουμε ότι ο παραπάνω έλεγχος υποθέσεων που περιγράψαμε ισχύει για κάθε συντελεστή $\mathbf{b_i}$ με την προϋπόθεση ότι υπάρχουν και άλλες ανεξάρτητες μεταβλητές στο υπόδειγμα. Έτσι λοιπόν αυτός ο έλεγχος υποθέσεων μπορεί να εκφραστεί και ως εξής:

Υπόθεση μηδέν (H_0) : $\mathbf{b_i = 0}$, υπό την προϋπόθεση ότι συμπεριλαμβάνονται όλες οι ανεξάρτητες μεταβλητές στο υπόδειγμα.

Εναλλακτική υπόθεση (H_1) : $\mathbf{b_i \neq 0}$, υπό την προϋπόθεση ότι συμπεριλαμβάνονται όλες οι ανεξάρτητες μεταβλητές στο υπόδειγμα.

Για επίπεδο σημαντικότητας **α**, το διάστημα εμπιστοσύνης του **i** συντελεστή παλινδρόμησης ισούται με :

$$\mathbf{b_i - s_{bi} \cdot t_{n\text{-}k\text{-}1,\,\alpha/2} < b_i < b_i\ s_{bi} \cdot t_{n\text{-}k\text{-}1,\,\alpha/2}}$$

Επίσης και σε αυτό το κεφάλαιο οι τύποι υπολογισμού των τυπικών σφαλμάτων s_{bi} είναι πολύπλοκοι και δύσκολο να γίνουν οι απαραίτητοι υπολογισμοί με το χέρι, άρα, δεν κρίνεται σκόπιμο να αναφερθούν, εξάλλου υπάρχει πληθώρα προγραμμάτων για τον υπολογιστή, που διευκολύνουν τους υπολογισμούς μας και τα οποία είναι εύχρηστα για τον καθένα.

3.8 ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΙ ΣΤΗΝ ΠΕΡΙΠΤΩΣΗ ΑΠΟΡΡΙΨΗΣ ΤΗΣ ΥΠΟΘΕΣΗΣ ΤΗΣ ΓΡΑΜΜΙΚΟΤΗΤΑΣ

Στην περίπτωση που ο έλεγχος απορρίπτει την υπόθεση της γραμμικότητας και δεν γνωρίζουμε ποια είναι η αληθής μη – γραμμική σχέση στον πληθυσμό (ώστε είτε να χρησιμοποιήσουμε μη – γραμμική μέθοδο εκτίμησης, είτε να τη μετασχηματίσουμε σε γραμμική), τότε πρέπει να χρησιμοποιήσουμε εναλλακτικές οικονομετρικές εξειδικεύσεις ώστε το υπόδειγμα που θα εκτιμήσουμε να είναι γραμμικό και σωστά εξειδικευμένο. Αν γνωρίζουμε ποια ανεξάρτητη μεταβλητή συνδέεται με μη – γραμμική σχέση με την εξαρτημένη μεταβλητή, τότε θα μπορούσαμε να χρησιμοποιήσουμε τους μετασχηματισμούς **Box – Cox**, αντικαθιστώντας τη συγκεκριμένη ανεξάρτητη μεταβλητή X_{ik} με τον μετασχηματισμό της σύμφωνα με τον Ιωάννη Αλκιβ. Κασκαρέλη (1999), Ένδεκα Μαθήματα Οικονομετρίας », 2η Αναθεωρημένη Έκδοση, Αθήνα : Εκδόσεις Gutenberg :

$$\mathbf{X^{*}_{ik} = \frac{X^{d}_{ik} - 1}{d}} \qquad (3.16)$$

όπου **d** μια σταθερά, με $\mathbf{-1 \leq d \leq 1}$. Επομένως, θέτοντας

$\mathbf{d = -1 \rightarrow 1 - X^{*}_{ik} = 1/\,X^{*}_{ik}}$ (υπόδειγμα του αντίστροφου μετασχηματισμού)

$\mathbf{d = 0 \rightarrow X^{*}_{ik} = lnX_{ik}}$ (μέσω του κανόνα l'Hospital, λογαριθμικός μετασχηματισμός)

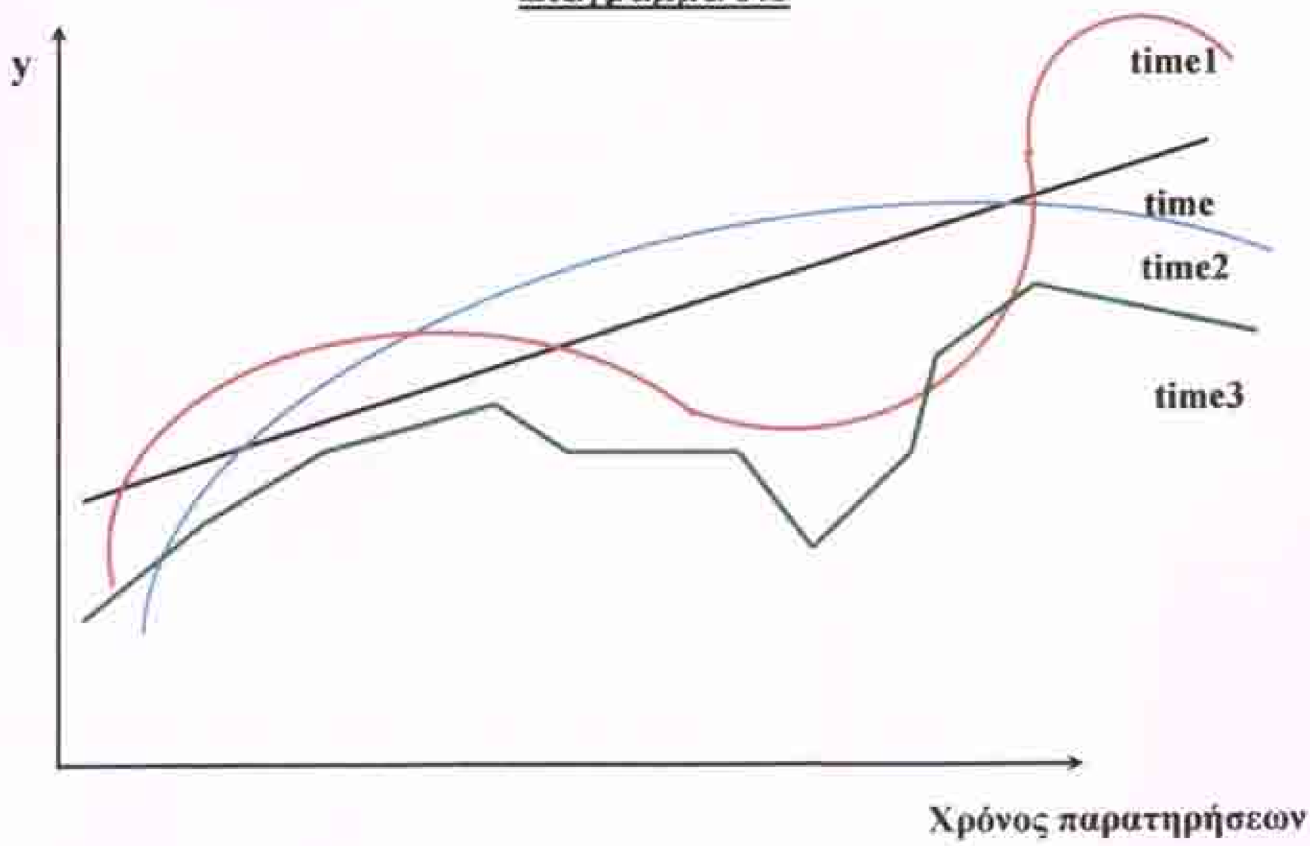
$\mathbf{d = 0.5 \rightarrow 1 + ½ * X^{*}_{ik} = \sqrt{X_{ik}}}$ **ή για** $\mathbf{d = -0.5 \rightarrow 1 - ½ * X^{*}_{ik} = 1\sqrt{X_{ik}}}$ κ.ο.κ

Άρα, θέτοντας τιμές στη σταθερά **d**, παίρνουμε και διαφορετικές μη – γραμμικές εξειδικεύσεις της μεταβλητής X_{ik}, αν και πολλές φορές αυτές οι εξειδικεύσεις είναι δύσκολο να περιγραφούν, και να ερμηνευτεί το οικονομικό νόημα που έχουν.

Γι' αυτό το λόγο οι οικονομολόγοι προτιμούν, για να προσεγγίσουν την αληθή μη – γραμμική σχέση, να εισάγουν μη – γραμμικές μορφές κάποιων ανεξάρτητων μεταβλητών, όπως επίσης και να εισάγουν νέες μεταβλητές που είναι γινόμενο ψευδομεταβλητών με κάποιες από τις ανεξάρτητες μεταβλητές, καθώς και οποιεσδήποτε άλλες μεταβλητές σε μη – γραμμική μορφή.

Με αυτόν τον τρόπο οι ερευνητές προσπαθούν να ερμηνεύσουν, και να αποτυπώσουν στη συνάρτηση παλινδρόμησης, πιθανές διαχρονικές μεταβολές των παρατηρήσεων της εξαρτημένης μεταβλητής **y**, που δεν ερμηνεύονται από τις άλλες ανεξάρτητες μεταβλητές που έχουν εισαχθεί γραμμικά στο υπόδειγμα (όπως φαίνεται στο παρακάτω διάγραμμα).

Διάγραμμα 3.1



3.9 ΕΠΙΛΟΓΗ ΜΕΤΑΒΛΗΤΩΝ

Το θέμα της επιλογής των μεταβλητών και το πρόβλημα της αλγεβρικής εξειδίκευσης του υποδείγματος είναι σημαντικά και βασικά. Από άποψη επιστημονικής πλευράς και σύμφωνα με τον Ανδρέα Δ. Κιντή (1999), « Στατιστικές και Οικονομετρικές Μέθοδοι », Αθήνα : Εκδόσεις Gutenberg τα δύο αυτά θέματα πρέπει να αντιμετωπίζονται ταυτόχρονα. Ωστόσο για λόγους απλούστευσης της δικής μα ανάλυσης, πρώτα θα πραγματευθούμε το πρόβλημα της επιλογής των μεταβλητών και στη συνέχεια θα αναφερθούμε στο θέμα της αλγεβρικής μορφής του υποδείγματος.

Για την εξήγηση της συμπεριφοράς της μεταβλητής **Y** υποθέτουμε ότι υποψήφιες ερμηνευτικές μεταβλητές είναι οι $\mathbf{X}_1$, $\mathbf{X}_2$, ……..$\mathbf{X}_m$ όπου $\mathbf{X}_{1i} = 1$ για όλες τις τιμές των άλλων μεταβλητών. Λόγοι λειτουργικότητας του υποδείγματος (αδυναμία εκτίμησης όλων των συντελεστών με ακρίβεια), καθώς και λόγοι περιορισμού του υπολογιστικού κόστους επιβάλλουν η σχετική ανάλυση να γίνει στα πλαίσια υποδείγματος, το οποίο θα περιλαμβάνει ένα υποσύνολο του αρχικού συνόλου των ερμηνευτικών μεταβλητών. Το υποσύνολο αυτό δεν είναι μοναδικό. Εξαρτάται από το σκοπό για το οποίο πρόκειται να χρησιμοποιηθεί το υπόδειγμα και από τη διαθεσιμότητα των αναγκαίων στοιχείων.

Οι σημαντικότερες χρήσεις ενός εμπειρικού υποδείγματος είναι :

i. Η περιγραφή της συμπεριφοράς ενός πολύπλοκου φαινομένου.
ii. Η πρόβλεψη (prediction) ή η πρόγνωση (forecasting) των τιμών της εξαρτημένης μεταβλητής **Y** σε περιοχές εκτός της περιοχής αναφοράς του δείγματος.
iii. Ο έλεγχος των τιμών της **Y**.

Στην περίπτωση που η ανάλυση αποβλέπει στην κατανόηση της συμπεριφοράς ενός φαινομένου, επιδίωξή μας πρέπει να είναι η ερμηνεία του μεγαλύτερου δυνατού μέρους της μεταβλητότητας των τιμών της εξαρτημένης μεταβλητής με το μικρότερο δυνατό αριθμό ανεξάρτητων μεταβλητών. Εξάλλου, όταν ο σκοπός της ανάλυσης είναι η πρόβλεψη ή η πρόγνωση, κριτήριο επιλογής των μεταβλητών πρέπει να είναι η ελαχιστοποίηση του μέσου σφάλματος τετραγώνου (**MSE**) της πρόβλεψης. Τέλος, στις περιπτώσεις που το υπόδειγμα πρόκειται να χρησιμοποιηθεί για τον προσδιορισμό της μεταβολής που πρέπει να επέλθει στις τιμές των ερμηνευτικών μεταβλητών, προκειμένου να επιτευχθούν ορισμένοι στόχοι (targets) αναφορικά με την εξαρτημένη μεταβλητή, απαιτείται μεγάλη ακρίβεια στους εκτιμημένους συντελεστές του υποδείγματος.

Αυτό, σε ικανοποιητικό βαθμό, διασφαλίζεται όταν τα τυπικά σφάλματα των συντελεστών είναι μικρά σε σχέση με τις τιμές των ίδιων των συντελεστών.

3.9.1 ΚΡΙΤΗΡΙΑ ΑΞΙΟΛΟΓΗΣΗΣ ΤΗΣ ΚΑΤΑΛΛΗΛΟΤΗΤΑΣ ΕΝΟΣ ΥΠΟΣΥΝΟΛΟΥ ΜΕΤΑΒΛΗΤΩΝ

Για την αξιολόγηση της επάρκειας εναλλακτικών υποδειγμάτων με διαφορετικά υποσύνολα ερμηνευτικών μεταβλητών, έχουν προταθεί διάφορα κριτήρια. Σε αυτό το σημείο όμως εμείς θα αναφερθούμε μόνο σε δύο από αυτά. Πρόκειται για τα κριτήρια που χρησιμοποιούνται συχνότερα στην εφαρμοσμένη έρευνα με κάποια επιτυχία.

3.9.1.1 ΤΟ ΚΡΙΤΗΡΙΟ ΤΟΥ ΜΕΣΟΥ ΤΩΝ ΤΕΤΡΑΓΩΝΩΝ ΤΩΝ ΚΑΤΑΛΟΙΠΩΝ [RESIDUAL MEAN SQUARE (RMS)]

Στην περίπτωση που το υπόδειγμα περιλαμβάνει λ συντελεστές, αυτό ορίζεται ως εξής :

$$\mathbf{(RMS)_\lambda = \frac{(ESS)_\lambda}{n - \lambda}} \qquad (3.17)$$

όπου το **(ESS)**$_\lambda$ = (error sum of squares) παριστάνει το άθροισμα των τετραγώνων των καταλοίπων σε ένα υπόδειγμα με λ παραμέτρους. Μεταξύ δύο υποδειγμάτων με διαφορετικό υποσύνολο μεταβλητών, προτιμάται εκείνο που έχει το μικρότερο (RMS).

Το **(RMS)**$_\lambda$ συνδέεται με τον συντελεστή προσδιορισμού $\mathbf{R^2}$ με την ακόλουθη σχέση :

$$\mathbf{R^2_\lambda = 1 - \frac{(ESS)_\lambda}{TSS} = 1 - (n - \lambda)\frac{(RMS)_\lambda}{TSS}}$$

όπου **(ESS)**$_\lambda$ **= (n – λ) (RMS)**$_\lambda$ **και** TSS = συνολικό άθροισμα των τετραγώνων των αποκλίσεων των τιμών της **Y** από το μέσο **Ỹ**. Στην περίπτωση του προσαρμοσμένου με τους αντίστοιχους βαθμούς ελευθερίας συντελεστή προσδιορισμού, η παραπάνω σχέση γίνεται :

$$\mathbf{(\text{προσαρμοσμένο})\ R^2_\lambda = 1 - (n - 1)\frac{(RMS)_\lambda}{TSS}} \qquad (3.18)$$

3.9.1.2 ΤΟ ΚΡΙΤΗΡΙΟ C

Από την εμπειρία μας ξέρουμε ότι, εφόσον από το προς εκτίμηση υπόδειγμα παραλείπονται ερμηνευτικές μεταβλητές, κατά κανόνα, οι εκτιμήσεις που παίρνουμε για τις παραμέτρους του υποδείγματος είναι μεροληπτικές.

Αυτό συνεπάγεται μεροληπτικές προβλέψεις για τις τιμές της **Y** που υπολογίζονται με βάση το εκτιμημένο υπόδειγμα. Στις περιπτώσεις αυτές, ως κριτήριο αξιολόγησης της ερμηνευτικής ικανότητας της εξίσωσης παλινδρόμησης χρησιμοποιείται το μέσο σφάλμα τετραγώνου της τιμής πρόβλεψης (predicted value) και όχι η διακύμανση. Έτσι, αν εκφράσουμε τις μεταβλητές στην τυποποιημένη τους μορφή, τότε παίρνουμε τον παρακάτω τύπο :

$$\mathbf{Z_\lambda = \frac{1}{s^2} \sum_{i=0}^{h} MSE(\hat{y}_i)} \qquad (3.19)$$

όπου το $\mathbf{Z_\lambda}$ παριστάνει το συνολικό μέσο σφάλμα τετραγώνου της πρόβλεψη το $\mathbf{MSE(\hat{y}_i)}$ μετράει το μέσο σφάλμα τετραγώνου που διαπράττεται κατά την πρόβλεψη της i της τιμής της **Y** από μια εξίσωση με λ παραμέτρους, το δε $\mathbf{s^2}$ εκφράζει τη διακύμανση του στοχαστικού όρου. Όπως είναι γνωστό, το $\mathbf{MSE(\hat{y}_i)}$ αποτελείται από δύο συνιστώσες, δηλαδή τη διακύμανση της πρόβλεψης και το τετράγωνο του μεροληπτικού σφάλματος, που οφείλεται στην παράλειψη ερμηνευτικών μεταβλητών. Αποδεικνύεται με τον Ανδρέα Δ. Κιντή (1999), « Στατιστικές και Οικονομετρικές Μέθοδοι », Αθήνα : Εκδόσεις Gutenberg.ότι ο προηγούμενος τύπος γράφεται ως εξής :

$$\mathbf{Z_\lambda = \frac{E(ESS)_\lambda}{s^2} + (2\lambda - n)} \qquad (3.20)$$

όπου το **$E(ESS)_\lambda$** αποτελεί αμερόληπτη εκτίμηση του **ESS**. Ο υπολογισμός του **Z_λ** προϋποθέτει τη γνώση του **s^2**.

Επειδή αυτό είναι άγνωστο, αντί του **Z_λ** υπολογίζεται το **C_λ** από τον ακόλουθο τύπο :

$$\mathbf{C_\lambda = \frac{(ESS)_\lambda}{\hat{s}^2} + (2\lambda - n)} \qquad (3.21)$$

όπου το **$\hat{s}^2$** αποτελεί εκτίμηση του **s^2** και υπολογίζεται από το υπόδειγμα που περιλαμβάνει όλες τις ερμηνευτικές μεταβλητές.

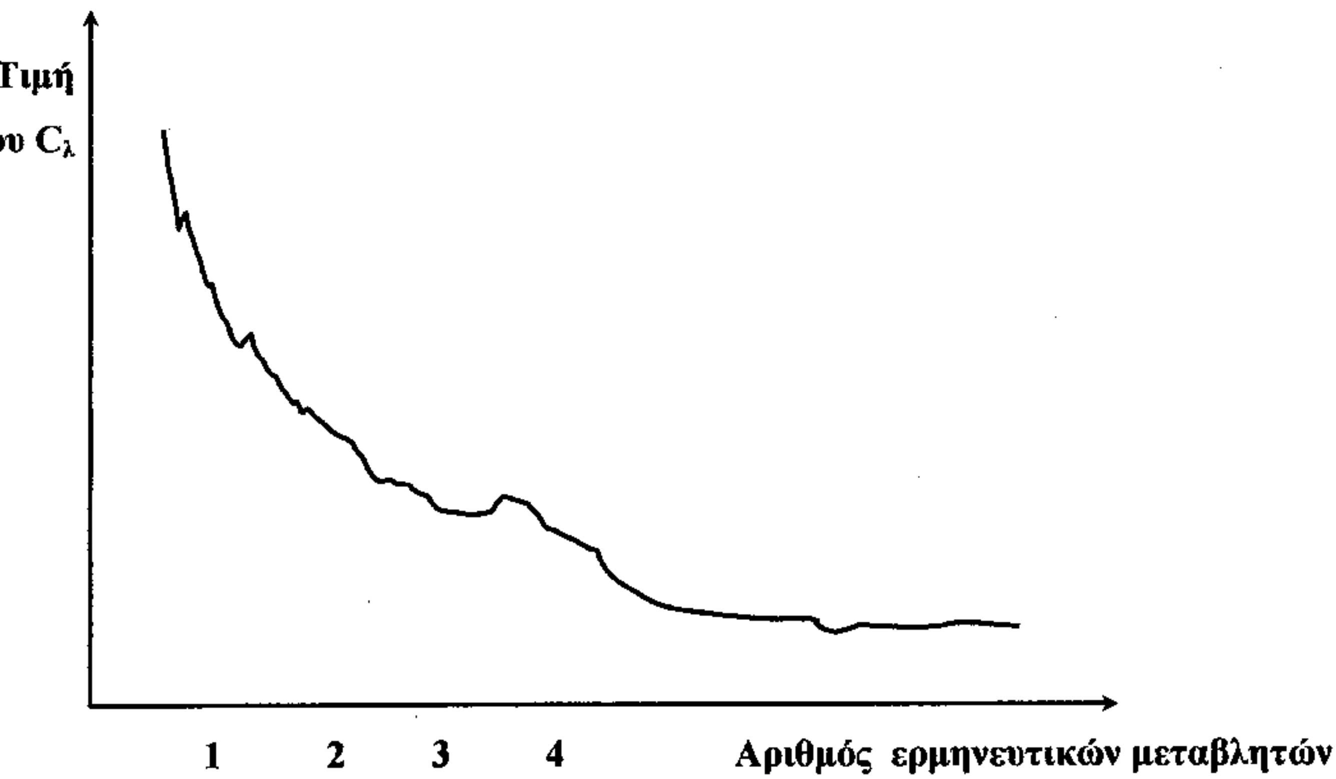
Αν αντικαταστήσουμε το **$\hat{s}^2$** με την αμερόληπτη εκτίμησή του, που ισούται με **ESS/ $n - \lambda$**, τότε ο προηγούμενος τύπος θα γίνει :

$$\mathbf{E(C_\lambda) = \frac{(ESS)_\lambda}{(ESS)_\lambda/n - \lambda} + 2\lambda - n = \lambda} \qquad (3.22)$$

Κατά συνέπεια, η διαφορά του **$C_\lambda - \lambda$** αποτελεί μέτρο του μεροληπτικού σφάλματος. Με άλλα λόγια, το κριτήριο **C_λ** μετράει την ερμηνευτική ικανότητα του υποδείγματος σε όρους του τυποποιημένου μέσου σφάλματος της πρόβλεψης. Συμπερασματικά λέμε, ότι το υποσύνολο των ερμηνευτικών μεταβλητών που δίνει τιμές για το **C_λ**, οι οποίες βρίσκονται πολύ κοντά στο **λ**, είναι το «καλό» υποσύνολο.

Το υποσύνολο αυτό μπορεί να προσδιοριστεί με τη βοήθεια ενός διαγράμματος.

Διάγραμμα 3.2



3.10 ΤΕΧΝΙΚΕΣ ΕΠΙΛΟΓΗΣ ΤΩΝ ΜΕΤΑΒΛΗΤΩΝ

Στη βιβλιογραφία έχουν προταθεί διάφορες τεχνικές για την επιλογή του κατάλληλου υποσυνόλου μεταβλητών, όταν κρίνεται ότι ο αρχικός αριθμός του πρέπει να περιοριστεί. Επίσης, κρίνεται απαραίτητο να διακρίνουμε τις περιπτώσεις που ανάμεσα στις ερμηνευτικές μεταβλητές υπάρχει ισχυρή σχέση εξάρτησης (πολυσυγγραμμικότητα) από εκείνες όπου οι ανεξάρτητες μεταβλητές είναι περίπου ορθογώνιες. Σχεδόν όλες οι προσεγγίσεις αποτυγχάνουν να δώσουν ικανοποιητική λύση, όταν ανάμεσα στις ερμηνευτικές μεταβλητές υπάρχει υψηλή εξάρτηση. Γι' αυτό είναι αναγκαίο πριν προχωρήσουμε στην διαδικασία επιλογής να εξεταστεί αν μεταξύ των ερμηνευτικών μεταβλητών υπάρχει ή όχι συγγραμμικότητα. Ένας σχετικά απλός τρόπος ελέγχου είναι ο υπολογισμός των χαρακτηριστικών ριζών της μήτρας συσχετίσεων των ανεξάρτητων μεταβλητών. Η ύπαρξη πολύ μικρών χαρακτηριστικών ριζών υποδηλώνει ισχυρό βαθμό εξάρτησης.

Οι πλέον γνωστές τεχνικές, που έχουν χρησιμοποιηθεί στην εφαρμοσμένη έρευνα, είναι :

i. Η εκτίμηση όλων των δυνατών παλινδρομήσεων.
ii. Η προσέγγιση της προς τα πίσω εγκατάλειψης (backward elimination) μεταβλητών.
iii. Η διαδικασία της προς τα εμπρός επιλογής (forward selection).
iv. Η μέθοδος της κατά «προσεκτικά βήματα» εκτίμησης (stepwise method).

3.10.1 ΕΚΤΙΜΗΣΗ ΟΛΩΝ ΤΩΝ ΔΥΝΑΤΩΝ ΠΑΛΙΝΔΡΟΜΗΣΕΩΝ

Η προσέγγιση αυτή συνίσταται στην εκτίμηση υποδειγμάτων με όλα τα δυνατά διακεκριμένα υποσύνολα μεταβλητών και έχει εφαρμογή τόσο στην περίπτωση που τα δεδομένα είναι ορθογώνια, όσο και σε καταστάσεις που ανάμεσα στις ερμηνευτικές μεταβλητές υπάρχει υψηλός βαθμός εξάρτησης.

Μια τέτοια προσέγγιση παρέχει στον ερευνητή το μέγιστο της πληροφόρησης που περιέχουν τα δεδομένα. Πρόβλημα δημιουργεί ο μεγάλος αριθμός υποδειγμάτων που πρέπει να εκτιμηθούν. Είναι αυτονόητο ότι κάτι τέτοιο δεν μπορεί να γίνει χωρίς τη βοήθεια ηλεκτρονικού υπολογιστή μεγάλης ταχύτητας.

Πιο συγκεκριμένα, αν ο συνολικός αριθμός των ερμηνευτικών μεταβλητών είναι m, ο αριθμός των παλινδρομήσεων που πρέπει να εκτιμηθούν είναι $\mathbf{2^m - 1}$. Για παράδειγμα, αν $\mathbf{m = 5}$ τότε θα εκτιμηθούν συνολικά $\mathbf{32 - 1 = 31(2^5 - 1)}$ εξισώσεις.

Το σύνολο των εξισώσεων αυτών υποδιαιρείται με υποσύνολα, που περιλαμβάνουν κάθε φορά **λ** ερμηνευτικές μεταβλητές (**λ** = 1, 2, ...,**κ**) και τα οποία ιεραρχούνται σύμφωνα με ένα από τα κριτήρια που αναφέρθηκαν πιο πάνω.

Έτσι λοιπόν, έχουμε $(^n{}_m)$ όπου ο συμβολισμός αυτός υποδηλώνει τους συνδυασμούς των n πραγματικών με m ερμηνευτικές μεταβλητές και υπολογίζεται από τον τύπο :

$$\binom{n}{m} = \frac{n!}{m!(n-m)!} \qquad (3.23)$$

Θα πρέπει όμως να τονιστεί ότι μια τέτοια μηχανιστική διαδικασία δεν οδηγεί πάντα στην καλύτερη επιλογή. Είναι απαραίτητο να γίνει αντιπαράθεση του υποδείγματος που επιλέγουμε με την καθιερωμένη θεωρία. Επίσης, θα πρέπει να αναφερθεί το γεγονός ότι η εμπειρία και γενικά οι ικανότητες του ερευνητή αποτελούν σημαντικούς παράγοντες στη διαδικασία επιλογής των μεταβλητών που τελικά θα χρησιμοποιηθούν στην ανάλυση του εκάστοτε φαινομένου. Ακόμη, πρέπει να σημειωθεί ότι αν ο αριθμός των υποψήφιων ερμηνευτικών μεταβλητών είναι μεγάλος, η προσέγγιση αυτή είναι ιδιαίτερα δαπανηρή σε χρόνο.

3.10.2 ΠΡΟΣΕΓΓΙΣΗ ΤΗΣ ΠΡΟΣ ΤΑ ΠΙΣΩ ΕΓΚΑΤΑΛΕΙΨΗΣ

Σύμφωνα με την προσέγγιση αυτή, αρχικά εκτιμούμε το υπόδειγμα με όλες τις μεταβλητές και στη συνέχεια εγκαταλείπουμε προοδευτικά τις μεταβλητές που εμφανίζουν τιμή για το κριτήριο t ή το **F** μικρότερη από κάποιο προκαθορισμένο επίπεδο [π.χ. **t<1 ή F<1**, σημειώνουμε ότι $\mathbf{F} = \mathbf{t}^2$]. Μετά την εγκατάλειψη της μεταβλητής με τιμή **t<1**, γίνεται επανεκτίμηση του υποδείγματος και η διαδικασία αυτή συνεχίζεται μέχρι που στο υπόδειγμα να μην υπάρχουν μεταβλητές με πολύ χαμηλές τιμές για το **t**. Στην περίπτωση που στο αρχικό υπόδειγμα ή σε κάποιο από τα επόμενα υπάρχουν περισσότερες μεταβλητές με **t<1** εγκαταλείπεται μόνο αυτή που έχει τη μικρότερη τιμή του t και το υπόδειγμα εκτιμάται με τις υπόλοιπες μεταβλητές. Η μέθοδος αυτή δεν ενδείκνυται να χρησιμοποιείται όταν υπάρχει υψηλή συγγραμμικότητα ανάμεσα στις ερμηνευτικές μεταβλητές.

3.10.3 Η ΔΙΑΔΙΚΑΣΙΑ ΤΗΣ ΠΡΟΣ ΤΑ ΕΜΠΡΟΣ ΕΠΙΛΟΓΗΣ

Πρώτα γίνεται παλινδρόμηση της εξαρτημένης μεταβλητής (**Y**) με κάθε μια από τις ερμηνευτικές μεταβλητές. Αρχή γίνεται με τη μεταβλητή που εμφανίζει τον υψηλότερο βαθμό συσχέτισης με την **Y** και έχει συντελεστή στατιστικά διαφορετικό από το μηδέν, ο οποίος έχει το αναμενόμενο πρόσημο με βάση τη θεωρία, σύμφωνα με τον Ανδρέα Δ. Κιντή (1999), « Στατιστικές και Οικονομετρικές Μέθοδοι », Αθήνα : Εκδόσεις Gutenberg. Στη συνέχεια και με βάση τους συντελεστές μερικής συσχέτισης της **Y** με κάθε μία από τις υπόλοιπες μεταβλητές, εισάγεται στο υπόδειγμα ως δεύτερη μεταβλητή εκείνη που έχει το μεγαλύτερο συντελεστή μερικής συσχέτισης και ελέγχεται η σημαντικότητά της.

Εάν είναι σημαντική, προχωρούμε στην εκτίμηση των συντελεστών μερικής συσχέτισης της **Y** με τις υπόλοιπες **m – 2** μεταβλητές, αφού προηγουμένως αφαιρεθεί τόσο από την **Y** όσο και από τις ανεξάρτητες μεταβλητές **m – 2**, η γραμμική επίδραση των δύο μεταβλητών, που ήδη υπάρχουν στο υπόδειγμα.

Η σχετική διαδικασία συνεχίζεται εφόσον η τιμή του προσαρμοσμένου $\mathbf{R^2}$ εξακολουθεί να αυξάνεται και η τιμή του **t**, που αντιστοιχεί στη νεοεισερχόμενη μεταβλητή, είναι μεγαλύτερη της προκαθορισμένης του τιμής (π.χ. **t<1**).

3.10.4 Η ΜΕΘΟΔΟΣ ΤΗΣ ΚΑΤΑ «ΠΡΟΣΕΚΤΙΚΑ ΒΗΜΑΤΑ» ΕΚΤΙΜΗΣΗΣ

Η προσέγγιση της προς τα εμπρός επιλογής εμφανίζει το εξής μειονέκτημα : μία μεταβλητή που εμφανίζεται ως «καλή» μεταβλητή σε ένα στάδιο, ενδέχεται σε μεταγενέστερο στάδιο να καταστεί περιττή με την έννοια ότι δεν είναι στατιστικά σημαντική. Το μειονέκτημα αυτό έχει σκοπό να διορθώσει η stepwise method. Έστω ότι η πρώτη μεταβλητή που εισέρχεται στο υπόδειγμα, σύμφωνα με την προηγούμενη μέθοδο, είναι η X_2 και ότι στη συνέχει εισάγεται η X_4. Η μέθοδος της κατά «προσεκτικά βήματα» επιλογής διερευνά αν πρέπει η X_2 να παραμείνει στο υπόδειγμα με κριτήριο τη σημαντικότητά της. Αν ναι, προχωρεί στην επιλογή της αμέσως επόμενης σημαντικής μεταβλητής. Αν όχι, εξαιρεί την X_2 και αναζητά την πιο σημαντική μεταβλητή από τις υπόλοιπες. Η μέθοδος αυτή συνδυάζει τις δύο προηγούμενες προκειμένου να καταλήξει στο ενδεδειγμένο υποσύνολο μεταβλητών. Βέβαια, αν υπάρχουν εξαρτήσεις ανάμεσα στις ερμηνευτικές μεταβλητές, τίποτε δεν διασφαλίζει ότι το υποσύνολο που τελικά επιλέγουμε είναι το καταλληλότερο.

3.11 ΠΟΛΥΣΥΓΓΡΑΜΜΙΚΟΤΗΤΑ

Μια από τις βασικές υποθέσεις που έχουν γίνει αναφέρει ότι οι ερμηνευτικές μεταβλητές είναι γραμμικά ανεξάρτητες. Όταν όμως κάποια από τις μεταβλητές εκφράζεται ως γραμμική συνάρτηση μιας ή περισσοτέρων από τις υπόλοιπες, η υπόθεση αυτή παραβιάζεται και το πρόβλημα που δημιουργείται είναι γνωστό στην οικονομετρία ως πρόβλημα της πολυσυγγραμμικότητας (multicollinearity problem). Η ύπαρξη πολυσυγγραμμικότητας δημιουργεί σοβαρά προβλήματα τόσο κατά τη στατιστική εκτίμηση του υποδείγματος, όσο και κατά τον έλεγχο και την ερμηνεία των αποτελεσμάτων. Παρακάτω θα αναφερθούμε στα εξής θέματα :

- Στη φύση της πολυσυγγραμμικότητας.
- Στις συνέπειες του προβλήματος.
- Λόγοι που προκαλούν την πολυσυγγραμμικότητα.
- Στη διαπίστωση και τη μέτρηση της πολυσυγγραμμικότητας.
- Στις πιθανές λύσεις του προβλήματος.

3.11.1 Η ΦΥΣΗ ΤΗΣ ΠΟΛΥΣΥΓΓΡΑΜΜΙΚΟΤΗΤΑΣ

Έστω το πολυμεταβλητό υπόδειγμα :

$$\mathbf{Y_i = b_1X_{1i} + b_2X_{2i} + b_3X_{3i} + \dots\dots\dots + b_\lambda X_{\lambda i} + u_i \qquad [i = 1,2,\dots,n]}$$

όπου $\mathbf{X_{1i}=1}$ για όλα τα $\mathbf{i}$ και $\mathbf{b_1}$ = σταθερός όρος.

Προκειμένου να διαπιστώσουμε αν μεταξύ των $\mathbf{X_j}$ ($\mathbf{j}$ = 1,2,....,λ) υπάρχει ακριβής εξάρτηση, εξετάζουμε την παρακάτω σχέση :

$$\mathbf{a_1X_{1i} + a_2X_{2i} + \dots\dots + a_\lambda X_{\lambda i} + e_i = 0} \qquad \text{για όλα τα } \mathbf{i} \qquad (3.24)$$

όπου τα $\mathbf{a_j}$ ($\mathbf{j}$ = 1,2,.....,λ) είναι σταθερές, το e_i παριστάνει το στοχαστικό όρο. Όταν η σχέση αυτή ισχύει μόνο για $\mathbf{a_j = 0}$ ($\mathbf{j}$ = 1,2,.....,λ) και $\mathbf{e_i = 0}$ ($\mathbf{j}$ = 1,2,.....,$\mathbf{n}$), τότε δεν υπάρχει σχέση γραμμικής εξάρτησης ανάμεσα στις $\mathbf{X_j}$. Στην περίπτωση αυτή οι ερμηνευτικές μεταβλητές ονομάζονται ορθογώνιες (δηλαδή η συνδιακυμανσή τους ισούται με το μηδέν). Τέτοιες καταστάσεις συναντώνται σπάνια στην εφαρμοσμένη οικονομετρική έρευνα.

Αντίθετα, όταν η παραπάνω σχέση ισχύει για ένα (ή περισσότερα) **a** διαφορετικό (διαφορετικά) από το μηδέν και $\mathbf{e}_i = \mathbf{0}$ για όλα τα **i**, τότε ανάμεσα στις ερμηνευτικές μεταβλητές υπάρχει ακριβής γραμμική εξάρτηση. Η κατάσταση αυτή είναι γνωστή σαν πλήρης πολυσυγγραμμικότητα . Πράγματι, αν $\mathbf{a}_3 \neq \mathbf{0}$ και $\mathbf{e}_i = \mathbf{0}$, τότε από την παραπάνω σχέση παίρνουμε :

$$\mathbf{X}_{3i} = -\frac{\mathbf{a}_1}{\mathbf{a}_3}\mathbf{X}_{1i} - \frac{\mathbf{a}_2}{\mathbf{a}_3}\mathbf{X}_{2i} - \ldots\ldots - \frac{\mathbf{a}_\lambda}{\mathbf{a}_3}\mathbf{X}_{\lambda i} \qquad (3.25)$$

δηλαδή η $\mathbf{X}_{3i}$ εκφράζεται σαν ακριβής γραμμική συνάρτηση των υπόλοιπων ερμηνευτικών μεταβλητών. Σ' αυτή την περίπτωση, είναι αδύνατη η εκτίμηση του υποδείγματος με τις γνωστές στατιστικές μεθόδους.

Και οι δύο αυτές περιπτώσεις που αναφέραμε πιο πάνω αποτελούν ακραίες καταστάσεις, οι οποίες συναντώνται σπάνια στην οικονομετρική έρευνα. Εκείνο που συμβαίνει συνήθως είναι ότι μεταξύ των ερμηνευτικών μεταβλητών υπάρχει μια μη πλήρης γραμμική εξάρτηση. Σε μια τέτοια περίπτωση η πρώτη σχέση ισχύει για ένα τουλάχιστον $\mathbf{a} \neq \mathbf{0}$ και $\mathbf{e}_i \neq \mathbf{0}$ για ένα τουλάχιστον **i**.

Αν για παράδειγμα $\mathbf{a}_3 \neq \mathbf{0}$, τότε θα έχουμε :

$$\mathbf{X}_{3i} = -\frac{\mathbf{a}_1}{\mathbf{a}_3}\mathbf{X}_{1i} - \frac{\mathbf{a}_2}{\mathbf{a}_3}\mathbf{X}_{2i} - \ldots\ldots - \frac{\mathbf{a}_\lambda}{\mathbf{a}_3}\mathbf{X}_{\lambda i} - \frac{\mathbf{1}}{\mathbf{a}_3}\mathbf{e}_i \qquad (3.26)$$

Η σχέση αυτή εκφράζει μια μη πλήρη γραμμική εξάρτηση μεταξύ της $\mathbf{X}_3$ και των υπολοίπων ερμηνευτικών μεταβλητών, δεδομένου ότι η $\mathbf{X}_3$ προσδιορίζεται και από τον στοχαστικό όρο $\mathbf{e}_i$. Εδώ έχουμε το φαινόμενο της μη πλήρους πολυσυγγραμμικότητας.

Πρόκειται για την πιο συνηθισμένη και πιο δύσκολη κατάσταση. Αξίζει να σημειωθεί ότι η πολυσυγγραμμικότητα αναφέρεται στην ύπαρξη γραμμικής εξάρτησης μεταξύ των ερμηνευτικών μεταβλητών. Δεν καλύπτει καταστάσεις όπου οι σχέσεις που συνδέουν τις μεταβλητές είναι μη γραμμικές.

3.11.2 ΣΥΝΕΠΕΙΕΣ ΤΗΣ ΠΟΛΥΣΥΓΓΡΑΜΜΙΚΟΤΗΤΑΣ

Μετά από την ανάλυση που προηγήθηκε σχετικά με τη φύση του προβλήματος της πολυσυγγραμμικότητας, ήρθε η ώρα να αναφερθούμε στις συνέπειες της παραβίασης της γραμμικής ανεξαρτησίας μεταξύ των ερμηνευτικών μεταβλητών ενός υποδείγματος :

1. Οι παράμετροι έχουν μεγάλες διακυμάνσεις και πολλά μεγαλύτερα τυπικά σφάλματα κάτι που σημαίνει πως η ακρίβεια και η σταθερότητα των δεδομένων είναι μειωμένη, με αποτέλεσμα πολλές φορές η προσθήκη ή η αφαίρεση περιορισμένου αριθμού παρατηρήσεων να προκαλεί σημαντικές μεταβολές στις τιμές των εκτιμηθέντων συντελεστών.

2. Επειδή έχουμε ευρύτερα διαστήματα εμπιστοσύνης, περιορίζεται η αξία των γνωστών κριτηρίων ελέγχου της στατιστικής σημαντικότητας των επιμέρους παραμέτρων και εμποδίζει τον προσδιορισμό των πραγματικών διαστημάτων εμπιστοσύνης.
3. Έχουμε λίγες μεταβλητές στατιστικά σημαντικές, παρόλα αυτά έχουν υψηλό $\mathbf{R}^2$ και στατιστικά σημαντικό.

4. Αρκετές από τις παραμέτρους μας μπορεί να έχουν λανθασμένα πρόσημα, με αποτέλεσμα να είναι δύσκολη η ερμηνεία των αποτελεσμάτων μας.

5. Υπάρχει δυσκολία στην εκτίμηση της συμβολής κάθε ερμηνευτικής μεταβλητής στο $\mathbf{R}^2$, κάτι το οποίο είναι δυνατό να οδηγήσει σε εσφαλμένη εξειδίκευση του υποδείγματος με την αυθαίρετη αφαίρεση ή πρόσθεση ανεξάρτητων μεταβλητών.

Για τους παραπάνω λόγους, η πολυσυγγραμμικότητα θεωρείται ένα από τα πιο σοβαρά προβλήματα, τα οποία αντιμετωπίζονται κατά την οικονομετρική ανάλυση των οικονομικών φαινομένων. Αξίζει να σημειωθεί ότι εφόσον οι αλληλεξαρτήσεις που υπάρχουν μεταξύ των ερμηνευτικών μεταβλητών παραμένουν σταθερές, η ύπαρξη πολυσυγγραμμικότητας δεν επηρεάζει την προγνωστική ικανότητα του υποδείγματος. Εξάλλου, όταν ο σκοπός του υποδείγματος είναι να υπολογίσουμε τη μελλοντική τιμή της εξαρτημένης μεταβλητής τότε η πολυσυγγραμμικότητα δεν αποτελεί πρόβλημα. Αν όμως ο σκοπός της έρευνας δεν είναι μόνο η πρόβλεψη αλλά και η αξιόπιστη εκτίμηση των παραμέτρων τότε το πρόβλημα πρέπει να λυθεί. Επίσης, εφόσον οι λοιπές υποθέσεις του κλασικού υποδείγματος δεν παραβιάζονται, η πολυσυγγραμμικότητα δεν επηρεάζει τις ιδιότητες των εκτιμητών ελαχίστων τετραγώνων. Με άλλα λόγια, παρά την ισχυρή γραμμική εξάρτηση που ενδέχεται να υπάρχει μεταξύ των ερμηνευτικών μεταβλητών , οι εκτιμητές ελαχίστων τετραγώνων εξακολουθούν να είναι **BLUE**.

3.11.3 ΛΟΓΟΙ ΠΟΥ ΠΡΟΚΑΛΟΥΝ ΤΗ ΠΟΛΥΣΥΓΓΡΑΜΜΙΚΟΤΗΤΑ

Στα πολυμεταβλητά οικονομετρικά υποδείγματα υπάρχει σχεδόν πάντοτε κάποια εξάρτηση μεταξύ δύο ή περισσότερων ερμηνευτικών μεταβλητών και εφόσον η πολυσυγγρμμικότητα αναφέρεται στην κατάσταση των ερμηνευτικών μεταβλητών πρόκειται για χαρακτηριστικό του δείγματος και όχι του πληθυσμού.

Η ύπαρξη πλήρους γραμμικής εξάρτησης μεταξύ δύο ή περισσότερων ερμηνευτικών μεταβλητών , υποδηλώνει ότι έχουν γίνει λάθη κατά τη διατύπωση της θεωρίας, δηλαδή, οι μεταβλητές που ερμηνεύουν το πραγματικό φαινόμενο είναι λιγότερες από αυτές που αναφέρει η θεωρία.

Η πολυσυγγραμμικότητα μπορεί να οφείλεται και στο γεγονός ότι οι ερμηνευτικές μεταβλητές έχουν δεχθεί επίδραση από εξωτερικούς παράγοντες. Επίσης, γραμμική εξάρτηση μπορούμε να έχουμε και στην περίπτωση που οι ανεξάρτητες μεταβλητές είναι στοχαστικές κάτι το οποίο δεν λαμβάνεται υπόψη κατά την εκτίμηση του υποδείγματος. Βέβαια, η πολυσυγγραμμικότητα μπορεί να προκληθεί και από τυχαίους λόγους. Θα πρέπει όμως να σημειωθεί ότι το πρόβλημα της πολυσυγγραμμικότητας εμφανίζεται κυρίως όταν στο προς εκτίμηση υπόδειγμα χρησιμοποιούνται δεδομένα χρονολογικών σειρών, παρά όταν έχουμε διακλαδικά δεδομένα.

3.11.4 ΔΙΑΠΙΣΤΩΣΗ ΚΑΙ ΜΕΤΡΗΣΗ ΤΗΣ ΠΟΛΥΣΥΓΓΡΑΜΜΙΚΟΤΗΤΑΣ

Όπως ήδη έχουμε αναφέρει η ύπαρξη πολυσυγγραμμικότητας είναι ένα από τα σοβαρότερα προβλήματα που υπάρχουν στην οικονομετρική έρευνα. Επίσης, από την εμπειρία μας ξέρουμε ότι στα πολυμεταβλητά υποδείγματα υπάρχει πάντα σε κάποιο βαθμό πολυσυγγραμμικότητα λόγω της αλληλεξάρτησης μεταξύ των ερμηνευτικών μεταβλητών. Έτσι λοιπόν η ανίχνευση και η μέτρηση της πολυσυγγραμμικότητας μπορεί να γίνει μέσω των εξής στοιχείων :

1) **Τον συντελεστή συσχέτισης μεταξύ των X :** Εφόσον η πολυσυγγραμμικότητα αναφέρεται στο βαθμό γραμμικής εξάρτησης μεταξύ των ερμηνευτικών μεταβλητών θα πρέπει να εξετάσουμε τις τιμές του συντελεστή συσχέτισης μεταξύ των μεταβλητών αυτών. Όμως, τα συμπεράσματα που πρόκειται να

βγάλουμε από τις τιμές του συντελεστή συσχέτισης διαφοροποιούνται ανάλογα με το πλήθος των ερμηνευτικών μεταβλητών που υπάρχουν στο υπόδειγμά μας. Αυτό σημαίνει ότι στα υποδείγματα με δύο ανεξάρτητες μεταβλητές η τιμή του συντελεστή απλής συσχέτισης μεταξύ των δύο ερμηνευτικών μεταβλητών αποτελεί ικανοποιητικό μέτρο του βαθμού γραμμικής τους εξάρτησης και όσο η τιμή αυτή πλησιάζει στη μονάδα, τόσο πιο σοβαρό είναι το πρόβλημα της πολυσυγγραμμικότητας, ενώ όταν η τιμή αυτή είναι χαμηλή δεν υπάρχει πρόβλημα πολυσυγγραμμικότητας. Όταν όμως στο υπόδειγμα υπάρχουν περισσότερες από δύο ερμηνευτικές μεταβλητές τότε ο έλεγχος γίνεται πιο πολύπλοκος και μια χαμηλή τιμή του συντελεστή συσχέτισης δεν σημαίνει απουσία πολυσυγραμμικότητας .

2) **Συντελεστής μερικής παλινδρόμησης :** Όπως ξέρουμε η πολυσυγγραμμικότητα δημιουργεί αστάθεια στις εκτιμήσεις των συντελεστών, μεγάλα τυπικά σφάλματα, κάποια από τα αλγεβρικά πρόσημα μπορεί αν μην συμφωνούν με την οικονομική θεωρία και μας δίνει συντελεστές στατιστικά μη σημαντικούς ενώ στην πραγματικότητα μπορεί να είναι αξιόπιστοι προσδιοριστικοί παράγοντες του υποδείγματός μας. Έτσι λοιπόν αν η προσθήκη μιας μεταβλητής έχει ως αποτέλεσμα την αύξηση της τιμής του $\mathbf{R}^2$, χωρίς να επηρεάζει την αξιοπιστία των υπόλοιπων συντελεστών των μεταβλητών που υπάρχουν στο υπόδειγμα, τότε δεν πρόκειται να δημιουργηθεί πολυσυγγραμμικότητα στο υπόδειγμα. Αν όμως η προσθήκη μιας μεταβλητής αυξήσει την τιμή του $\mathbf{R}^2$, τότε θα υπάρξουν σοβαρές μεταβολές στους συντελεστές των άλλων μεταβλητών με αποτέλεσμα να φαίνονται ως στατιστικά μη σημαντικοί και με ασυμβίβαστα πρόσημα σχετικά με την οικονομική θεωρία. Αυτό μας δείχνει ότι πρόκειται να υπάρξει πρόβλημα πολυσυγγραμμικότητας. Τώρα αν η προσθήκη μιας νέας μεταβλητής δεν επηρεάσει καθόλου την τιμή του $\mathbf{R}^2$, τότε η μεταβλητή αυτή θεωρείται περιττή και δεν υπάρχει λόγος να συμπεριληφθεί στο υπόδειγμά μας.

3)Σχέση μεταξύ των τιμών F και t : Ενδείξεις για την ύπαρξη πολυσυγγραμμικότητας μπορούμε να έχουμε και από την εξέταση των τιμών των κριτηρίων **F** και **t**. Όταν μεταξύ των ερμηνευτικών μεταβλητών υπάρχει ισχυρή εξάρτηση τότε είναι δυνατόν η τιμή του **F** να είναι στατιστικά σημαντική και συγχρόνως οι τιμές πολλών ή όλων των **t** να μην είναι στατιστικά διαφορετικές από το μηδέν δηλαδή στατιστικά μη σημαντικές.

Σημείωση **:** Για την ανίχνευση και μέτρηση της πολυσυγγραμμικότητας μπορούμε να χρησιμοποιήσουμε επίσης και δύο άλλους δείκτες τον **VIF** και **Tolerance**. Οι δείκτες αυτοί θεωρητικά υπολογίζονται από τους εξής τύπους :

$$\mathbf{VIF} = \frac{\mathbf{1}}{(\mathbf{1} - \mathbf{R}^2)} \quad (3.27) \qquad \text{και} \qquad \mathbf{Tolerance} = \mathbf{1} - \mathbf{R}^2 \quad (3.28)$$

Όταν η τιμή του δείκτη **VIF** είναι πάνω από τέσσερα (**VIF > 4**) τότε υπάρχει πιθανή πολυσυγγραμμικότητα και όταν η τιμή του δείκτη **Tolerance** είναι κάτω από **0.20**

(**Tolerance < 0.20**) τότε πάλι υπάρχει πιθανή πολυσυγγρμμικότητα. Αυτοί οι δείκτες μπορούν να υπολογιστούν και με τη βοήθεια προγράμματος ηλεκτρονικού υπολογιστή (π.χ **SPSS**).

3.11.5 ΠΙΘΑΝΕΣ ΛΥΣΕΙΣ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ ΤΗΣ ΠΟΛΥΣΥΓΓΡΑΜΜΙΚΟΤΗΤΑΣ

Όπως ήδη έχουμε αναφέρει, όταν μεταξύ των ερμηνευτικών μεταβλητών υπάρχει ισχυρή συσχέτιση, οι εκτιμήσεις των επιμέρους παραμέτρων του υποδείγματος είναι ανακριβείς, ασταθείς και αναξιόπιστοι. Έτσι λοιπόν είναι αδύνατον να προσδιοριστεί η πραγματική συμβολή των επιμέρους ερμηνευτικών μεταβλητών στη διαμόρφωση των τιμών της εξαρτημένης μεταβλητής Y. Βέβαια, η ερμηνευτική ικανότητα του υποδείγματος στο σύνολό του δεν επηρεάζεται από την ύπαρξη της πολυσυγγραμμικότητας αλλά από τον σκοπό της ανάλυσης.

Μερικές από της λύσεις του προβλήματος της πολυσυγγραμμικότητας είναι οι ακόλουθες σύμφωνα με τους Ανδρέα Δ. Κιντή (1999), « Στατιστικές και Οικονομετρικές Μέθοδοι », Αθήνα : Εκδόσεις Gutenberg και Damodar Gujarati (1999), « Essentials of Econometrics », 2nd Edition McGraw Hill :

1) **Αφαίρεση της μεταβλητής που δημιουργεί το πρόβλημα** :Αυτό μπορεί να είναι ικανοποιητική λύση από μαθηματικής πλευράς όχι όμως από πλευράς της οικονομετρίας, και αυτό γιατί η κάθε μεταβλητή αποτελεί πηγή πληροφοριών χρήσιμων για την εκάστοτε οικονομετρική ανάλυση. Έτσι λοιπόν η παράλειψη μιας μεταβλητής μπορεί να μειώνει τη σοβαρότητα της πολυσυγγραμικότητας, δημιουργεί όμως νέο πρόβλημα. Πρόκειται για την εισαγωγή μεροληπτικού σφάλματος στις εκτιμήσεις των διαφόρων παραμέτρων λόγω της παράλειψης ουσιώδους ερμηνευτικής μεταβλητής (specification error). Άρα λοιπόν δεν ενδείκνυται ως λύση η αφαίρεση κάποιας μεταβλητής.

2) **Αύξησης του μεγέθους του υποδείγματος** (προσθέτοντας κάποια μεταβλητή) **και επίτευξη ακρίβειας κατά τους υπολογισμούς**: Βασιζόμενοι στο γεγονός ότι η πολυσυγγραμμικότητα αναφέρεται στο δείγμα και όχι στον πληθυσμό, η αύξηση του μεγέθους του δείγματος κατά κανόνα συμβάλλει στη μείωση του βαθμού εξάρτησης μεταξύ των ερμηνευτικών μεταβλητών. Από την αύξηση του μεγέθους του δείγματος έχουμε και αύξηση των πληροφοριών που εμπεριέχονται στο δείγμα. Όμως αν οι νέες παρατηρήσεις αποτελούν απλά προέκταση των αρχικών, τότε η πρόσθεσή τους στο δείγμα δεν έχει νόημα. Εξάλλου, η αύξηση του αριθμού των παρατηρήσεων δεν είναι εύκολη υπόθεση.

3) **Η αλλαγή στην εξειδίκευση του υποδείγματος**: Στην συγκεκριμένη περίπτωση αντί να παραλείπουμε μια μεταβλητή, αντικαθιστούμε τη μεταβλητή που προκαλεί το πρόβλημα της πολυσυγγραμμικότητας με μια άλλη μεταβλητή που έχει παραπλήσιο περιεχόμενο με την πρώτη και να μην συσχετίζεται με τις υπόλοιπες μεταβλητές. Βέβαια το να βρει κανείς μια τέτοια μεταβλητή δεν είναι εύκολη υπόθεση. Άλλος τρόπος αλλαγής της εξειδίκευσης του υποδείγματος για να αντιμετωπιστεί η πολυσυγγραμμικότητα είναι η προσθήκη νέων εξισώσεων με τις οποίες να εκφράζονται οι πλέον ουσιώδεις σχέσεις μεταξύ των ερμηνευτικών μεταβλητών που συσχετίζονται. Οι νέες εξισώσεις να περιγράφουν πιθανόν και μεταβλητές που δεν υπάρχουν στο αρχικό υπόδειγμα. Βέβαια, η αλλαγή αυτή έχει ως αποτέλεσμα τη μετατροπή του αρχικού υποδείγματος από υπόδειγμα μιας εξίσωσης (single equation model) σε υπόδειγμα ταυτόχρονα προσδιοριζόμενων εξισώσεων (simultaneous – equation model).

4) **Η αλλαγή στη μέτρηση των μεταβλητών** : Βέβαια, μια τέτοια αλλαγή θα έχει νόημα μόνο εφόσον οι παράμετροι που θα εκτιμηθούν από το νέο υπόδειγμα εξακολουθούν να περιγράφουν τους ίδιους οικονομικούς μηχανισμούς με τις παραμέτρους του αρχικού υποδείγματος. Η αλλαγή αυτή μπορεί να γίνει με μετασχηματισμό ή ομαδοποίηση των μεταβλητών. Οι μετασχηματισμοί που έχουν χρησιμοποιηθεί αναφέρονται στη χρησιμοποίηση :

a) των πρώτων διαφορών, αντί των αρχικών τιμών των μεταβλητών.
b) των πρώτων διαφορών των λογαρίθμων των μεταβλητών.
c) του λόγου των μεταβλητών.

Θα πρέπει να τονιστεί όμως πως για κάθε μετασχηματισμό του υποδείγματος θα πρέπει να ισχύει ότι για να έχουν οι εκτιμητές ελαχίστων τετραγώνων τις γνωστές ιδιότητες θα πρέπει οι υποθέσεις που αφορούν τη στοχαστική μεταβλητή να ισχύουν στο υπόδειγμα που τελικά εκτιμάται και όχι στο αρχικό.

Άλλες προτεινόμενες λύσεις για το πρόβλημα της πολυσυγγραμμικότητας είναι η χρησιμοποίηση προγενέστερης πληροφόρησης, η παραγοντική ανάλυση (factor analysis), η μέθοδος της ραχοειδούς παλινδρόμησης (ridge regression) κ.α. Η ανάλυση αυτών των μεθόδων είναι πολύπλοκη, δηλαδή απαιτεί εξειδικευμένη και πολύ καλή γνώση στατιστικών εννοιών. Επίσης, είναι αρκετά χρονοβόρα μια τέτοια ανάλυση και δεν αποτελεί σκοπό της συγκεκριμένης εργασίας.

3.12 ΕΤΕΡΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑ

3.12.1 Η ΦΥΣΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ

Μια από τις βασικές υποθέσεις που έχουν γίνει αναφέρει ότι η διασπορά των επιμέρους τιμών της $\mathbf{u}_i$ γύρω από τον μέσο της είναι σταθερή και ανεξάρτητη των τιμών των ερμηνευτικών μεταβλητών. Με άλλα λόγια, οι τιμές της στοχαστικής μεταβλητής προέρχονται από μια κατανομή με μέσο μηδέν και σταθερή διακύμανση. Στην περίπτωση αυτή λέμε ότι οι τιμές του στοχαστικού όρου εμφανίζουν **ομοσκεδαστικότητα** (homoskedasticity). Όμως εξαιτίας πολλών προβλημάτων της πραγματικότητας η υπόθεση της ομοσκεδαστικότητας δεν είναι ρεαλιστική. Έτσι, πολύ συχνά στην πράξη αντιμετωπίζουμε καταστάσεις όπου η διακύμανση των $\mathbf{u}_i$ δεν είναι σταθερή. Στις περιπτώσεις αυτές λέμε ότι έχουμε **ετεροσκεδαστικότητα** (heteroskedasticity), αυτό σημαίνει πως οι τιμές προέρχονται από κατανομές με μηδενικό μέσο αλλά και με διαφορετική διακύμανση.

Το φαινόμενο της ετεροσκεδαστικότητας παρουσιάζεται στις περιπτώσεις που οι τιμές μιας ή περισσότερων ερμηνευτικών μεταβλητών εμφανίζουν μεγάλη μεταβλητότητα, δηλαδή, στις χαμηλές τιμές της ερμηνευτικής μεταβλητής **X**, η διακύμανση στις τιμές των u είναι μικρότερη από ότι είναι στις μεγάλες τιμές της **X**. Έτσι, στην περίπτωση του διμεταβλητού υποδείγματος έχουμε:

$$\mathbf{Y_i = a + bX_i + u_i} \quad (3.29)$$

Το φαινόμενο της ετεροσκεδαστικότητας εκφράζεται ως εξής:

$$\mathbf{Var(u_i) = E(u_i^2) = s_i^2 = f(X_i)} \quad (3.30)$$

Η παραπάνω σχέση δείχνει ότι η διακύμανση των $\mathbf{u_i}$ διαφέρει για κάθε **i**. Αυτό δεν είναι απαραίτητο να συμβαίνει πάντα, για να έχουμε ετεροσκεδαστικότητα αρκεί η διακύμανση να διαφέρει για ορισμένα **i**.

Στην περίπτωση του πολυμεταβλητού υποδείγματος έχουμε:

$$\mathbf{Y_i = b_0 + b_1X_{1i} + \ldots\ldots + b_\lambda X_{\lambda i} + u_i} \quad (3.31)$$

Εδώ δεν γνωρίζουμε ποια από τις $\mathbf{X_j}$ $\mathbf{(j = 1,2,\ldots..,\lambda)}$ προκαλεί την ετεροσκεδαστικότητα, και το φαινόμενο μπορεί να εκφραστεί ως εξής:

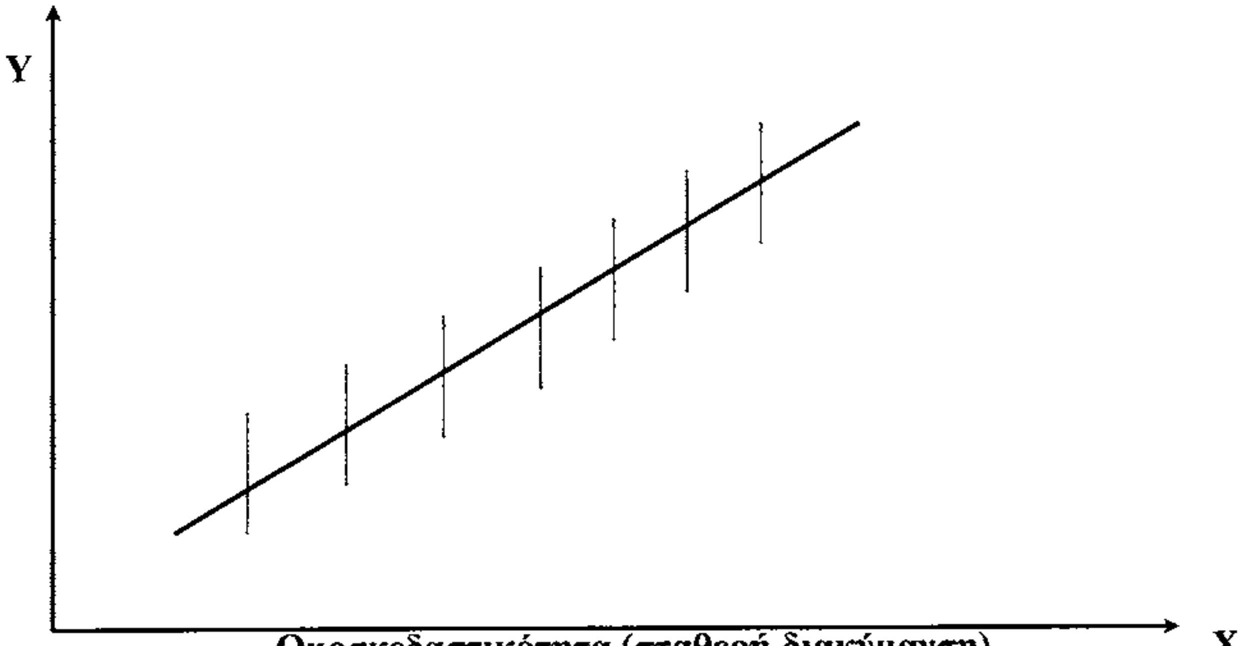
$$\mathbf{Var(u_i) = E(u_i^2) = s_i^2 = f(\hat{Y}_i)} \quad (3.32)$$

Όπου το $\mathbf{\hat{Y}_i}$ παριστάνει τις θεωρητικές τιμές της **Y** και υπολογίζεται από την εξής σχέση:

$$\mathbf{\hat{Y}_i = b_0 + b_1X_{1i} + \ldots\ldots + b_\lambda X_{\lambda i}} \quad (3.33)$$

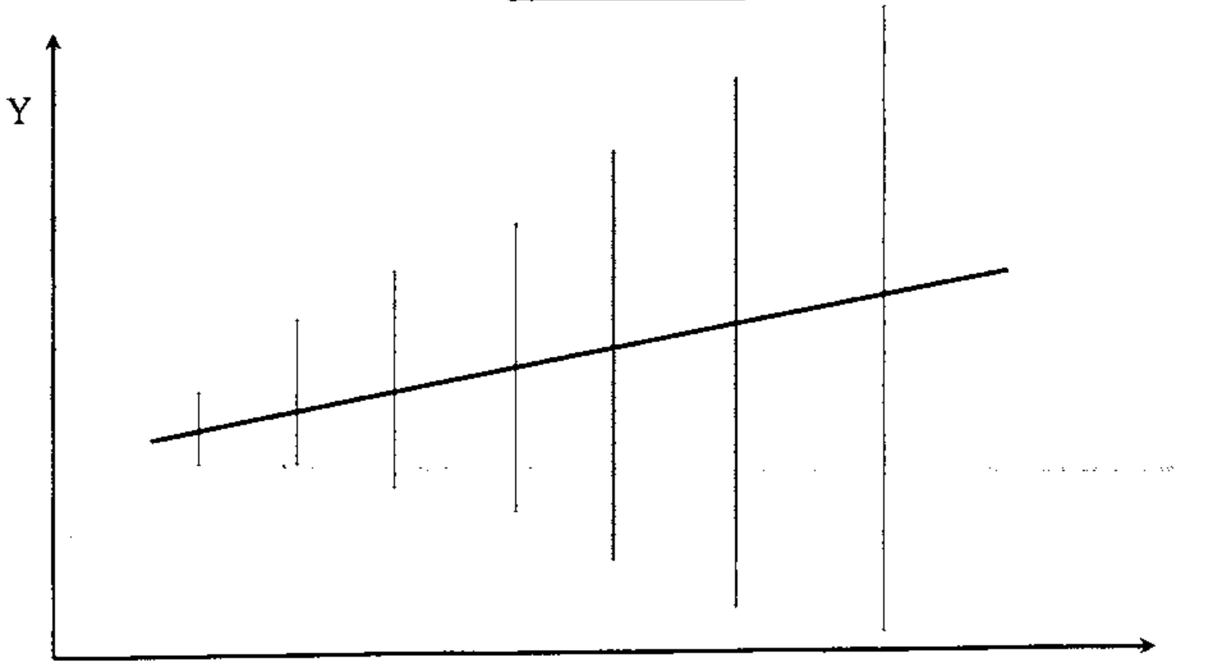
Εναλλακτικά, το φαινόμενο της ετεροσκεδαστικότητας μπορεί να παρασταθεί με διαγράμματα τα οποία μας δείχνουν πως μεταβάλλεται η διακύμανση.

Διάγραμμα 3.3



Ομοσκεδαστικότητα (σταθερή διακύμανση)

Διάγραμμα 3.4



Ετεροσκεδαστικότητα (μεταβαλλόμενη διακύμανση)

Αξίζει να τονιστεί ότι οι σχέσεις του διμεταβλητού και του πολυμεταβλητού υποδείγματος που εκφράζουν το πρόβλημα της ετεροσκεδαστικότητας, υποδηλώνουν ότι οι τιμές της στοχαστικής μεταβλητής δεν είναι ανεξάρτητες από τις τιμές των ερμηνευτικών μεταβλητών, αφού η διακύμανση των $\mathbf{u}_i$ εξαρτάται από τις τιμές της $\mathbf{X}_i$ ή της $\hat{\mathbf{Y}}_i$, με άλλα λόγια αυτό το πρόβλημα μας δείχνει ότι παραβιάζεται η ανεξαρτησία μεταξύ των ερμηνευτικών μεταβλητών και του στοχαστικού όρου. Στη συνέχεια έχουμε ως συνέπεια την παραβίαση τόσο των ιδιοτήτων όσο και της κατανομής των εκτιμητών των παραμέτρων του υποδείγματος. Για να ξεπεράσουμε το πρόβλημα της ετεροσκεδαστικότητας υποθέτουμε ότι ανεξάρτητα από τις τιμές των ερμηνευτικών μεταβλητών , η προσδοκώμενη τιμή της u ισούται με το μηδέν για κάθε παρατήρηση, δηλαδή, για κάθε $\mathbf{X}_s$ ή $\hat{\mathbf{Y}}_s$ έχουμε $\mathbf{E(u_i) = 0}$ για όλα τα i και όλα τα s. Αυτό διασφαλίζει ότι, καθώς η τιμή της **X** (ή της $\hat{\mathbf{Y}}$) μεγαλώνει, η προσδοκώμενη τιμή της **u** εξακολουθεί να ισούται με μηδέν και κατά συνέπεια:

$$\mathbf{Cov(u_i, X_i) = cov(u_i, \hat{Y}_i) = 0} \quad (3.34)$$

3.12.2 ΛΟΓΟΙ ΠΟΥ ΠΡΟΚΑΛΟΥΝ ΤΟ ΦΑΙΝΟΜΕΝΟ ΤΗΣ ΕΤΕΡΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑΣ

Οι κυριότεροι λόγοι που προκαλούν το φαινόμενο της ετεροσκεδαστικότητας στην εφαρμοσμένη οικονομετρική έρευνα σύμφωνα με τον Ανδρέα Δ. Κιντή (1999), « Στατιστικές και Οικονομετρικές Μέθοδοι », Αθήνα : Εκδόσεις Gutenberg, είναι :

1. Η μεταβολή στις εξωτερικές συνθήκες εφαρμογής του υποδείγματος.
2. Η χρησιμοποίηση διακλαδικών ή διαστρωματικών δεδομένων (cross – section data).

3. Η χρησιμοποίηση ομαδοποιημένων δεδομένων (grouped data).
4. Όταν οι συντελεστές του υποδείγματος δεν είναι σταθεροί σε ολόκληρο το πεδίο αναφοράς του δείγματος (varying coefficient models).

Α) Μεταβολή στις εξωτερικές συνθήκες

Κατά την οικονομετρική ανάλυση ενός υποδείγματος ένα σύνολο παραγόντων υποτίθεται ότι παραμένει αμετάβλητο για ολόκληρο το υπό μελέτη δείγμα. Πολλές φορές όμως αυτό δεν συμβαίνει. Τότε λέμε ότι έχουμε μεταβολή στις εξωτερικές συνθήκες εφαρμογής του υποδείγματος. Εδώ πρόκειται για αλλαγές στη νομοθεσία, στην οικονομική πολιτική, στην νομισματική πολιτική, στις προτιμήσεις των καταναλωτών, κ.λπ. Η επίδραση των μεταβολών αυτών εκφράζεται στις τιμές του διαταρακτικού όρου με αποτέλεσμα η διακύμανσή του να μην είναι σταθερή.

Β) Χρησιμοποίηση διακλαδικών ή διαστρωματικών δεδομένων

Έρευνες έχουν δείξει ότι το πρόβλημα της ετεροσκεδαστικότητας εμφανίζεται όταν έχουμε διακλαδικά ή διαστρωματικά δεδομένα. Για παράδειγμα, η εκτίμηση της παραγωγής σε μια βιομηχανία με διακλαδικά στοιχεία πάνω στο κεφάλαιο, την εργασία κ.α, υποθέτοντας ότι οι υπόλοιποι παράγοντες που επηρεάζουν την παραγωγή είναι ίδιοι για όλες τις βιομηχανίες. Η συμβολή όμως αυτών των παραγόντων που εκφράζονται μέσα από τις τιμές του στοχαστικού όρου, είναι περιορισμένη και δεν παρουσιάζει αξιόλογες διακυμάνσεις για τις μικρές επιχειρήσεις. Όμως η σημασία της για τις μεγάλες μονάδες είναι καθοριστική και εμφανίζει έντονες διαφοροποιήσεις από επιχείρηση σε επιχείρηση. Έτσι οι τιμές του διαταρακτικού όρου που αναφέρονται στις μεγάλες επιχειρήσεις θα έχουν διαφορετική διακύμανση από αυτές των μικρών επιχειρήσεων. Με άλλα λόγια, οι τιμές της στοχαστικής μεταβλητής θα εμφανίζουν ετεροσκεδαστικότητα.

Τα ίδια προβλήματα έχουμε και όταν χρησιμοποιούμε διαστρωματικά δεδομένα (οικογενειακός προϋπολογισμός).

Αξίζει να σημειωθεί όμως ότι όταν η εκτίμηση του υποδείγματος βασίζεται σε δεδομένα χρονολογικών σειρών, συνήθως δεν δημιουργείται πρόβλημα ετεροσκεδαστικότητας. Βέβαια, αν τα στοιχεία των χρονολογικών σειρών αναφέρονται σε μεγάλα χρονικά διαστήματα και υπάρχουν διαρθρωτικές μεταβολές στα επιμέρους δεδομένα τότε υπάρχει πιθανότητα να εμφανιστεί ετεροσκεδαστικότητα.

Γ) Χρησιμοποίηση ομαδοποιημένων δεδομένων

Πολλές φορές για την εκτίμηση ενός υποδείγματος χρησιμοποιούνται μέσοι όροι επιμέρους μεγεθών. Στις περιπτώσεις αυτές δημιουργείται πρόβλημα ετεροσκεδαστικότητας όταν οι ομάδες στις οποίες αναφέρονται οι μέσοι όροι είναι διαφορετικού μεγέθους.

Δ) Μη σταθεροί συντελεστές

Υπάρχουν υποδείγματα στα οποία ορισμένοι συντελεστές δεν είναι σταθεροί σε ολόκληρο το πεδίο αναφοράς του δείγματος. Έστω το υπόδειγμα :

$$\mathbf{Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + b_{3i}X_{3i} + u_i} \qquad (3.35)$$

όπου το $\mathbf{b_{3i}}$ μεταβάλλεται όταν μεταβάλλεται το $\mathbf{i}$.

Αν υποθέσουμε ότι οι μεταβολές στις τιμές του $\mathbf{b_{3i}}$ είναι τυχαίες, τότε μπορούμε να γράψουμε :

$$\mathbf{b_{3i} = b_3 + v_i} \quad (3.36)$$

όπου $\mathbf{v_i}$ = τυχαία μεταβλητή. Αντικαθιστώντας στην εξίσωση του υποδείγματος έχουμε :

$$\begin{aligned}\mathbf{Y_i} &= \mathbf{b_0 + b_1X_{1i} + b_2X_{2i} + (b_3 + v_i)X_{3i} + u_i} \\ &= \mathbf{b_0 + b_1X_{1i} + b_2X_{2i} + b_3X_{3i} + \omega_i \; [\omega_i = u_i + X_{3i}v_i]}\end{aligned}$$

Αν τα $\mathbf{u_i}$ και $\mathbf{v_i}$ είναι ανεξάρτητα τότε έχουμε :

$$\mathbf{Var(\omega_i) = var(u_i + X_{3i}v_i) = s^2_u + s^2_v X^2_{3i}} \quad (3.37)$$

Έτσι βλέπουμε ότι η διακύμανση του στοχαστικού όρου στο παραπάνω υπόδειγμα είναι συνάρτηση των τιμών της ερμηνευτικής μεταβλητής X_3 και κατά συνέπεια δεν είναι σταθερή.

3.12.3 ΣΥΝΕΠΕΙΕΣ ΤΗΣ ΕΤΕΡΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑΣ

Μετά την παραπάνω ανάλυση σχετικά με τους λόγους που προκαλούν το πρόβλημα της ετεροσκεδαστικότητας μπορούμε να φανταστούμε και να καταλάβουμε την σημαντικότητα των συνεπειών. Έτσι λοιπόν έχουμε τις εξής συνέπειες :

1. Οι εκτιμητές ελαχίστων τετραγώνων διατηρούν τις ιδιότητες της γραμμικότητας και της αμεροληψίας, δεν έχουν όμως πια την ελάχιστη διακύμανση και δεν είναι πλέον αποδοτικές.

2. Η αγνόηση του προβλήματος της ετεροσκεδαστικότητας και η εκτίμηση της διακύμανσης του **b*** από τον τύπο **var(b)** = s^2_u / Σx^2_i συνεπάγεται υποεκτίμηση της πραγματικής διακύμανσης και στη συνέχεια του τυπικού σφάλματος. Το γεγονός αυτό επηρεάζει τόσο τον έλεγχο της αξιοπιστίας των συντελεστών όσο και την κατασκευή των σωστών διαστημάτων εμπιστοσύνης. Ειδικότερα οι τιμές του **t** που δίνει το πρόγραμμα δείχνουν μεγαλύτερες από ότι στην πραγματικότητα είναι. Αυτό με τη σειρά του μπορεί να οδηγήσει στην απόρριψη της μηδενικής υπόθεσης ακόμα και όταν αυτή είναι σωστή. Με άλλα λόγια μπορούμε να πούμε ότι έχουμε μεγαλύτερα διαστήματα εμπιστοσύνης και ότι τα **F** και **t test** για έλεγχο στατιστικής σημαντικότητας δεν είναι αξιόπιστα.

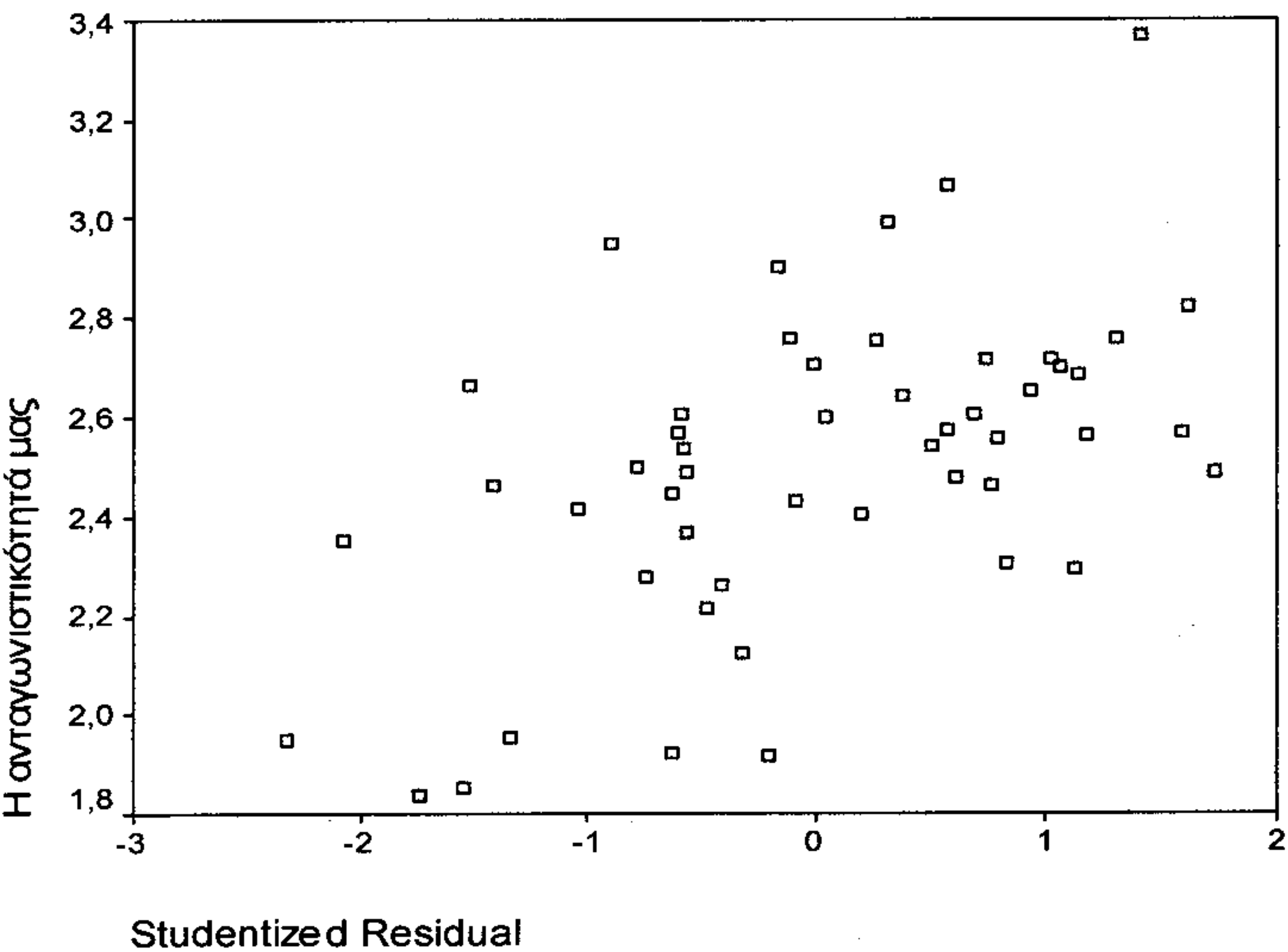
3.12.4 ΔΙΑΠΙΣΤΩΣΗ ΤΗΣ ΕΤΕΡΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑΣ

Αρχικά θα πρέπει να διευκρινίσουμε ότι γενικά δεν υπάρχουν ισχυρά στατιστικά κριτήρια για τη διαπίστωση της ετεροσκεδαστικότητας. Στην εφαρμοσμένη οικονομετρική έρευνα οι αποφάσεις για την ύπαρξη ή μη ετεροσκεδαστικότητας στηρίζονται στην ανάλυση της συμπεριφοράς των καταλοίπων ελαχίστων τετραγώνων. Η ανάλυση αυτή μπορεί να γίνει είτε με τη γραφική μέθοδο, είτε με ορισμένα αναλυτικά κριτήρια. Η επιλογή του κατάλληλου κριτηρίου ελέγχου εξαρτάται από τα διαθέσιμα στατιστικά στοιχεία.

α) Γραφική μέθοδος

Ενδείξεις για την ύπαρξη ετεροσκεδαστικότητας μπορούμε να έχουμε από τη γραφική παράσταση των τετραγώνων των καταλοίπων και των τιμών μιας από τις ερμηνευτικές μεταβλητές ή των θεωρητικών τιμών της εξαρτημένης μεταβλητής. Το παρακάτω διάγραμμα το οποίο έχει προκύψει με τη βοήθεια του προγράμματος **SPSS** μας δείχνει τη μορφή του διαγράμματος που φανερώνει μη ύπαρξη ετεροσκεδαστικότητας και αυτό το βλέπουμε διότι τα σημεία δεν ακολουθούν ένα συγκεκριμένο σχηματισμό.

Διάγραμμα 3.5



Εδώ θα πρέπει να τονίσουμε ότι από την απεικόνιση των διαγραμμάτων μόνο ενδείξεις μπορούμε να πάρουμε για την ύπαρξη ή μη ετεροσκεδαστικότητας. Δεν είναι δυνατό να βγουν οριστικά συμπεράσματα σχετικά με τη σταθερότητα ή μη της διακύμανσης του στοχαστικού όρου στο υπόδειγμα. Ο λόγος είναι ότι η διακύμανση των καταλοίπων ε_i προσδιορίζονται όχι μόνο από τη διακύμανση της στοχαστικής μεταβλητής, αλλά και από τη διακύμανση των ερμηνευτικών μεταβλητών.

β) Κριτήριο του Park

Το κριτήριο αυτό διαθέτει μια αλγεβρική μορφή που συνδέει τις τιμές των καταλοίπων με τις ανεξάρτητες μεταβλητές. Ειδικότερα, ο Park υιοθέτησε την ακόλουθη μορφή :

$$s^2_j = s^2 X^b_i e^{ui} \quad (3.38)$$

Στη θέση των s^2_i που είναι άγνωστες, ο Park χρησιμοποίησε τα τετράγωνα των καταλοίπων ελαχίστων τετραγώνων και εκτίμησε την παλινδρόμηση :

$$\log \varepsilon^2_i = a + b\log X_i + u_i \qquad [a = \log s^2] \quad (3.39)$$

Αν το **b** είναι στατιστικά διαφορετικό από το μηδέν δεχόμαστε ότι υπάρχει ετεροσκεδαστικότητα.

γ) Συντελεστής συσχέτισης του Spearman

Ο συντελεστής συσχέτισης κατά τάξεις του Spearman υπολογίζεται από τον τύπο :

$$r_{\varepsilon . x} = 1 - \frac{6\Sigma d^2_i}{n(n^2-1)} \quad (3.40)$$

όπου το **d** παριστάνει τις διαφορές των τάξεων ανάμεσα σε δύο χαρακτηριστικά ενός φαινόμενου και το n το μέγεθος του δείγματος. Στην περίπτωσή μας τα δύο χαρακτηριστικά είναι τα κατάλοιπα ελαχίστων τετραγώνων και οι αντίστοιχες τιμές της ερμηνευτικής μεταβλητής X_j με την οποία από ότι φαίνεται συμμεταβάλλεται η διακύμανση του στοχαστικού όρου.

Η εφαρμογή του κριτηρίου αυτού περιλαμβάνει τα εξής στάδια :

1. Υπολογίζονται τα κατάλοιπα ελαχίστων τετραγώνων από το αρχικό υπόδειγμα χρησιμοποιώντας το σύνολο των παρατηρήσεων.
2. Διατάσσονται οι τιμές της X_j κατά αύξουσα (ή φθίνουσα) τάξη μεγέθους και σε κάθε τιμή της X_j αντιστοιχείται η ανάλογη απόλυτη τιμή του ε_i.
3. Υπολογίζεται ο $r_{\varepsilon.x}$ και στη συνέχεια ελέγχεται η σημαντικότητά του με βάση την τιμή του t που λαμβάνεται από τον τύπο :

$$\mathbf{t} = \mathbf{r}_{\varepsilon.x}\sqrt{\mathbf{n}-\mathbf{2}}\ /\ \sqrt{\mathbf{1}-\mathbf{r}^2_{\varepsilon.x}} \qquad (3.41)$$

με **n – 2** βαθμούς ελευθερίας και το **n > 8**. Όταν η υπολογιζόμενη τιμή του t υπερβαίνει την κριτική τιμή αυτού, τότε συμπεραίνουμε ότι υπάρχει ετεροσκεδαστικότητα.

δ) Δείκτης Cook's Distance

Τέλος έχουμε τον δείκτη Cook's Distance όπου τον εντοπίζουμε μέσα από πρόγραμμα στατιστικής στον ηλεκτρονικό υπολογιστή. Δεν θα αναφερθούμε στον θεωρητικό υπολογισμό αυτού του δείκτη διότι είναι αρκετά πολύπλοκος. Όταν συναντούμε την τιμή του δείκτη την συγκρίνουμε με τον λόγο **4/(n – k – 1)**, όπου n ο αριθμός των παρατηρήσεων του δείγματος και k ο αριθμός των ανεξάρτητων μεταβλητών. Αν η τιμή του δείκτη Cook's Distance είναι μικρότερη από την τιμή του λόγου που αναφέραμε λίγο πριν(κρίσιμη τιμή) τότε δεν έχουμε πρόβλημα ετεροσκεδαστικότητας στο υπόδειγμά μας.

3.13 ΕΚΤΙΜΗΣΗ ΤΟΥ ΥΠΟΔΕΙΓΜΑΤΟΣ

Όταν με κάποιο από τα παραπάνω κριτήρια, αποδειχθεί ότι ο διαταρακτικός όρος είναι ετεροσκεδαστικός, δημιουργείται πρόβλημα εκτίμησης του υποδείγματος. Έτσι λοιπόν για να αντιμετωπίσουμε το πρόβλημα αυτό θα πρέπει να μετασχηματίσουμε το αρχικό υπόδειγμα ώστε η στοχαστική μεταβλητή να έχει σταθερή διακύμανση. Η λύση του προβλήματος είναι εύκολη όταν είναι γνωστές ή μπορούν να εκτιμηθούν οι διακυμάνσεις s^2_i. Όταν όμως οι s^2_i είναι άγνωστες, υιοθετούμε υποθέσεις σχετικά με τη μορφή της ετεροσκεδαστικότητας ή προσπαθούμε να κάνουμε εκτιμήσεις από τα δεδομένα του δείγματος, εφόσον αυτά το επιτρέπουν.

α) Σταθμισμένοι εκτιμητές ελαχίστων τετραγώνων

i) Η περίπτωση που οι s^2_i είναι γνωστές

Όταν οι διακυμάνσεις του στοχαστικού όρου είναι γνωστές, η λύση στο πρόβλημα της ετεροσκεδαστικότητας βρίσκεται στην εφαρμογή της **σταθμισμένης μεθόδου ελαχίστων τετραγώνων** (weighted least squares method). Έστω ότι έχουμε προς εκτίμηση το υπόδειγμα :

$$\mathbf{Y_i = a + bX_i + u_i} \quad (3.42)$$

στο οποίο ισχύουν όλες οι γνωστές υποθέσεις, εκτός από την υπόθεση της ομοσκεδαστικότητας, δηλαδή $\mathbf{E(u^2_i) = s^2_i}$.

Όπως είναι γνωστό, η μέθοδος ελαχίστων τετραγώνων στην αρχική της μορφή κατά την ελαχιστοποίηση του αθροίσματος των τετραγώνων των καταλοίπων δίνει το ίδιο βάρος σε όλες τις αποκλίσεις ε^2_i. Αυτό είναι σωστό στην περίπτωση που οι τιμές του στοχαστικού όρου εμφανίζουν ομοσκεδαστικότητα. Όταν όμως οι τιμές είναι ετεροσκεδαστικές, η άμεση εφαρμογή της μεθόδου ελαχίστων τετραγώνων οδηγεί σε εκτιμητές που δεν είναι αποτελεσματικοί. Η λύση στο πρόβλημα που δημιουργείται είναι να εφαρμοστεί η σταθμισμένη μέθοδος ελαχίστων τετραγώνων, η οποία δίνει διαφορετικά βάρη στις διάφορες τιμές του ε^2_i. Τα βάρη αυτά είναι αντιστρόφως ανάλογα προς το μέγεθος των διακυμάνσεων s^2_i.

Έτσι, αντί της ελαχιστοποίησης του $S = \Sigma\varepsilon^2_i$, ελαχιστοποιούμε το σταθμισμένο άθροισμα των ε^2_i :

$$\mathbf{S^* = \sum_i \omega_i \varepsilon^2_i = \sum_i \omega_i (Y_i - a^* - b^* X_i)^2} \quad (3.43)$$

όπου τα **a*** και **b*** είναι οι σταθμισμένοι εκτιμητές ελαχίστων τετραγώνων (WLSE) και $\omega_i = 1/s^2_i$. Αποδεικνύεται ότι οι **εκτιμητές αυτοί είναι blue**. Από την ελαχιστοποίηση της παραπάνω εξίσωσης παίρνουμε :

$$\mathbf{b^* = \frac{\sum_i \omega_i x_i y_i}{\sum_i \omega_i x^2_i}} \quad \text{και} \quad \mathbf{a^* = \tilde{Y}^* - b^* \bar{X}^*} \quad (3.44)$$

όπου $$\mathbf{\tilde{Y}^* = \frac{\sum_i \omega_i Y_i}{\sum_i \omega_i}} = \text{σταθμικός μέσος των } \mathbf{Y_i} \quad (3.45)$$

$$\mathbf{\bar{X}^* = \frac{\sum_i \omega_i X_i}{\sum_i \omega_i}} = \text{σταθμικός μέσος των } \mathbf{X_i} \quad (3.46)$$

$$\mathbf{y_i = Y_i - \tilde{Y}^*} \quad (3.47) \qquad \text{και} \qquad \mathbf{x_i = X_i - \bar{X}^*} \quad (3.48)$$

Επίσης, οι διακυμάνσεις των **a*** και **b*** δίνονται από τους τύπους :

$$\mathbf{Var(b^*)=} \frac{\sum_i \omega_i}{\sum_i \omega_i)(\sum_i \omega_i X^2_i) - (\sum_i \omega_i X_i)^2} \quad (3.49)$$

$$\mathbf{Var(a^*)=} \frac{\sum_i \omega_i X^2_i}{(\sum_i \omega_i)(\sum_i \omega_i X^2_i) - (\sum_i \omega_i X_i)^2} \quad (3.50)$$

Εδώ πρέπει να τονίσουμε, ότι οι τιμές του s^2_i σπάνια είναι γνωστές. Σε ορισμένες περιπτώσεις οι s^2_i μπορούν να εκτιμηθούν από τα δεδομένα του προβλήματος. Αυτό είναι δυνατόν όταν το μέγεθος του δείγματος είναι αρκετά μεγάλο και οι παρατηρήσεις μπορούν να ταξινομηθούν σε ομάδες, στις οποίες οι διακυμάνσεις να είναι περίπου σταθερές.

ii) **Η περίπτωση που οι s^2_i εκτιμώνται από τα δεδομένα του υποδείγματος**

Έστω ότι για κάθε τιμή της ερμηνευτικής μεταβλητής X_i **(i = 1, 2,….,m)** έχουμε N_i παρατηρήσεις για την εξαρτημένη μεταβλητή **Y**. Το προς εκτίμηση υπόδειγμα γράφεται ως εξής :

$$\mathbf{Y_{ij} = a + bX_i + u_{ij}} \qquad \mathbf{(j = 1, 2,\ldots..,N_i)} \quad (3.51)$$

και $\quad N = \sum_{i=1}^{m} N_i$ = σύνολο παρατηρήσεων της **Y**.

Τα δεδομένα για την Υ μας επιτρέπουν να εκτιμήσουμε τις παρακάτω διακυμάνσεις :

$$S^2_i = \frac{1}{N_i} \sum_{j=1}^{Ni} (Y_{ij} - \tilde{Y}_i)^2 \quad (3.52)$$

όπου

$$\tilde{Y}_i = \frac{1}{N_i} \sum_{j=1}^{Ni} Y_{ij} \quad (3.53)$$

Στη συνέχεια αντικαθιστούμε τις s^2_i με τις S^2 στη σχέση $\omega_i = 1/ s^2_i$ και εφαρμόζουμε τους τύπους του **b*** και **var(a*)** που αναφέρθηκαν στην προηγούμενη παράγραφο. Εδώ θα πρέπει να σημειώσουμε όμως ότι οι εκτιμητές **b*** και **a*** που παίρνουμε στην περίπτωση αυτή είναι μόνο ασύμπτωτα αποτελεσματικοί.

β) Λογαριθμικός μετασχηματισμός του υποδείγματος

Σε ορισμένες περιπτώσεις οι συνέπειες της ετεροσκεδαστικότητας μπορούν να περιοριστούν με το να κάνουμε λογαριθμικό μετασχηματισμό του υποδείγματος. Έτσι, αντί του αρχικού υποδείγματος, εκτιμούμε το παρακάτω :

$$\log Y_i = b_0 + b_1 \log X_i + u_i \quad (3.54)$$

Το υπόδειγμα στις αρχικές τιμές των μεταβλητών είναι της μορφής $Y_i = AX^b_i e^{ui}$, όπου e = βάση των φυσικών λογαρίθμων. Ο λόγος που ο μετασχηματισμός αυτός μειώνει τη σοβαρότητα του προβλήματος της ετεροσκεδαστικότητας είναι ότι έτσι μειώνεται αισθητά η κλίμακα στην οποία μετρούνται οι μεταβλητές. Είναι φανερό πως μια διαφορά 10 προς 1 στο αρχικό υπόδειγμα περιορίζεται με τον παραπάνω μετασχηματισμό σε διαφορά 2 προς 1 περίπου.

3.14 ΑΥΤΟΣΥΣΧΕΤΙΣΗ

3.14.1 Η ΦΥΣΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ ΚΑΙ ΟΙ ΛΟΓΟΙ ΠΟΥ ΤΟ ΠΡΟΚΑΛΟΥΝ

Στο κομμάτι αυτό θα ασχοληθούμε με τα προβλήματα που δημιουργούνται από την παραβίαση της ανεξαρτησίας μεταξύ των τιμών του στοχαστικού όρου. Με άλλα λόγια θα μελετήσουμε την περίπτωση που οι συνδιακυμάνσεις των τιμών της στοχαστικής μεταβλητής δεν ισούνται όλες με το μηδέν. Συμβολικά αυτό παριστάνεται ως εξής :

$$\mathbf{E(u_i\, u_j) \neq 0 \quad , \quad i \neq j}$$

Τέτοια φαινόμενα συναντιούνται κατά την ανάλυση πολλών οικονομικών δεδομένων. Η παραβίαση της ανεξαρτησίας των $\mathbf{u_i}$ οδηγεί στην εμφάνιση του προβλήματος της **αυτοσυσχέτισης** (autocorrelation) σύμφωνα με τον Δημήτρη Α. Ιωαννίδη (1999), « Στατιστικές Μέθοδοι », Τόμος πρώτος, Θεσσαλονίκη : Εκδόσεις Ζήτη. Όταν λέμε ότι έχουμε αυτοσυσχέτιση εννοούμε ότι υπάρχει σχέση εξάρτησης μεταξύ των διαδοχικών τιμών μιας μεταβλητής. Ο βαθμός της εξάρτησης αυτής μετριέται με τους **συντελεστές αυτοσυσχέτισης** (autocorrelation coefficients).

Επίσης, επειδή η αυτοσυσχέτιση είναι φαινόμενο που παρατηρείται συνήθως κατά την ανάλυση χρονολογικών σειρών, και για διευκόλυνση στους συμβολισμούς θα αντικαταστήσουμε το δείκτη i με το δείκτη **t (t = 1, 2, ...T)**.

Βέβαια, αυτοσυσχέτιση μπορεί να εμφανιστεί και κατά την ανάλυση διακλαδικών ή διαστρωματικών δεδομένων όταν οι παρατηρήσεις του δείγματος έχουν διαταχθεί σύμφωνα με ορισμένη τάξη.

Οι βασικότεροι λόγοι που προκαλούν την αυτοσυσχέτιση είναι :

i. **Η εσφαλμένη αλγεβρική εξειδίκευση του υποδείγματος**. Αν υποθέσουμε ότι η αρχική μαθηματική σχέση που συνδέει τις μεταβλητές **Υ** και **Χ** είναι δευτέρου βαθμού, ενώ κατά την εξειδίκευση του υποδείγματος χρησιμοποιήσαμε σχέση γραμμικής μορφής. Αν οι τιμές του **Χ** εμφανίζουν αυτοσυσχέτιση, τότε η παράλειψη του όρου $\mathbf{X}^2$ από το προς εκτίμηση υπόδειγμα μπορεί να προκαλέσει αυτοσυσχέτιση στις τιμές του διαταρακτικού όρου για το λόγο ότι αυτές περιλαμβάνουν και την επίδραση του όρου $\mathbf{X}^2$.

ii. **Η παράλειψη μιας ή περισσότερων ερμηνευτικών μεταβλητών**. Πολύ συχνά οι οικονομικές μεταβλητές εμφανίζουν κάποιο βαθμό αυτοσυσχέτιση με αποτέλεσμα πολλές φορές η παράλειψη μιας ή περισσότερων μεταβλητών από αυτές να προκαλεί αυτοσυσχέτιση στις τιμές του διαταρακτικού όρου. Στην περίπτωση αυτή όπως και στην προηγούμενη, η αυτοσυσχέτιση στις τιμές της **u** οφείλεται στην εσφαλμένη εξειδίκευση του υποδείγματος. Θα πρέπει να τονίσουμε σε αυτό το σημείο ότι η ερμηνευτική μεταβλητή που έχει παραλειφθεί πρέπει να είναι ανεξάρτητη από το στοχαστικό όρο. Στην αντίθετη περίπτωση είναι πολύ πιθανό να παραβιάζονται και άλλες υποθέσεις του υποδείγματος.

iii. **Η ύπαρξη σφαλμάτων μέτρησης στις μεταβλητές**. Η αυτοσυσχέτιση είναι ενδεχόμενο να οφείλεται στην ύπαρξη σφαλμάτων μέτρησης στις μεταβλητές, όταν αυτά εμφανίζουν συστηματικότητα.

iv. **Η εκτίμηση μέρους παρατηρήσεων με παρεμβολή**. Πολλές φορές γίνεται συμπλήρωση ελλειπουσών παρατηρήσεων από μια χρονολογική σειρά με τη μέθοδο της παρεμβολής. Το γεγονός αυτό είναι δυνατό να προκαλέσει αυτοσυσχέτιση στις τιμές της μεταβλητής **u**.

v. **Η κατανομή της επίδρασης ορισμένων τυχαίων γεγονότων σε περισσότερες από μια χρονική περίοδο**. Στην οικονομική ζωή μιας χώρας συμβαίνουν κάποια τυχαία γεγονότα, τα οποία ασκούν επίδραση πάνω στη διαμόρφωση των τιμών των διάφορων οικονομικών μεταβλητών. Η επίδραση αυτή, περιλαμβάνεται και στις τιμές της στοχαστικής μεταβλητής. Επίσης, συχνά συμβαίνει η επίδραση των τυχαίων παραγόντων να μην περιορίζεται μέσα σε μια χρονική περίοδο αλλά να επεκτείνεται και σε μελλοντικές περιόδους με αποτέλεσμα οι τιμές ορισμένων μεταβλητών να εμφανίζονται συστηματική υποεκτιμημένες ή υπερεκτιμημένες.

Μετά από την ανάλυση σχετικά με τους λόγους που προκαλούν την αυτοσυσχέτιση οδηγούμαστε στο συμπέρασμα ότι, κατά την ανάλυση πολλών οικονομικών χρονολογικών δεδομένων οι περισσότεροι από τους λόγους αυτούς πολύ πιθανό να συνυπάρχουν. Έτσι, το πρόβλημα της αυτοσυσχέτισης δημιουργείται κυρίως στις περιπτώσεις που η εκτίμηση του υποδείγματος γίνεται με στοιχεία χρονολογικών σειρών.

3.14.2 ΜΟΡΦΕΣ ΤΗΣ ΑΥΤΟΣΥΣΧΕΤΙΣΗΣ

Για την αντιμετώπιση του προβλήματος της αυτοσυσχέτισης είναι απαραίτητο να γνωρίζουμε τη μορφή της. Η πιο συνηθισμένη υπόθεση που γίνεται στην οικονομετρική έρευνα σε σχέση με τη συμπεριφορά των τιμών της $\mathbf{u}_t$ είναι ότι αυτές ακολουθούν το εξής πρότυπο :

$$\mathbf{u_t = \rho u_{t-1} + v_t} \qquad |\rho| < 1 \qquad (3.55)$$

όπου το $\mathbf{v_t}$ είναι τυχαία μεταβλητή με τις εξής ιδιότητες :

$\mathbf{E(v_t) = 0}$, για όλα τα t

$\mathbf{E(v^2_t) = s^2_v}$, δηλαδή σταθερή διακύμανση

$\mathbf{E(v_t\, v_s) = 0}$, για όλα τα $t \neq s$

Η παράμετρος **ρ** ονομάζεται **συντελεστής αυτοσυσχέτισης** (autocorrelation coefficient).Το πρότυπο $\mathbf{u_t = \rho u_{t-1} + v_t}$ είναι γνωστό ως **αυτοπαλίνδρομο σχήμα πρώτου βαθμού** (first – order autogressive sheme) ή αλλιώς ως **στοχαστική διαδικασία πρώτου βαθμού του Markov** όπως αναφέρεται στο βιβλίο του Γ.Κ. Χρήστου (2000), « Εισαγωγή στην Οικονομετρία », Τόμος πρώτος, Αθήνα : Εκδόσεις Gutenberg.

Κάτω από τις προηγούμενες σχέσης αποδεικνύεται ότι :

$$\mathbf{Var(u_t) = E(u^2_t) = \frac{s^2_u}{1-\rho^2} = s^2_u} \qquad (3.56)$$

και

$$\mathbf{Cov(u_t\, u_{t-1}) = E(u_t\, u_{t-1}) = \rho \frac{s^2_v}{1-\rho^2} = \rho s^2_u} \qquad (3.57)$$

όπου $\mathbf{s^2_v}$ = διακύμανση του $\mathbf{v_t}$ και **ρ** = συντελεστής αυτοσυσχέτισης.

Από τις σχέσεις αυτές είναι φανερό ότι κάτω από συνθήκες αυτοσυσχέτισης ο στοχαστικός όρος $\mathbf{u_t}$ στο υπόδειγμα :

$$\mathbf{Y_t = a + bX_t = u_t} \quad (3.58)$$

δεν διατηρεί τις κλασσικές υποθέσεις.

Αξίζει να τονιστεί πως για πολλά προβλήματα της πραγματικής ζωής το σχήμα $\mathbf{u_t = \rho u_{t-1} + v_t}$ δεν είναι επαρκές. Οι τιμές του u είναι δυνατό να ακολουθούν κάποιο αυτοπαλίνδρομο σχήμα ανώτερου βαθμού ή να ανήκουν σε μια διαφορετική στοχαστική διαδικασία, όπως, για παράδειγμα, είναι η διαδικασία του κινητού μέσου όρου (moving average process). Ακόμη, τα u είναι ενδεχόμενο να ακολουθούν κάποια διαδικασία μικτής μορφής, δηλαδή, ένα νέο σχήμα που συνδυάζει το αυτοπαλίνδρομο και το σχήμα του κινητού μέσου. Επειδή όμως τα αυτοπαλίνδρομα σχήματα χρησιμοποιούνται συχνότερα στην οικονομετρική έρευνα θα περιοριστούμε σε αυτά.

Τα πιο γνωστά αυτοπαλίνδρομα σχήματα ανώτερου βαθμού είναι τα εξής :

$$\mathbf{u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t} \quad (3.59)$$

$$\mathbf{u_t = \rho_4 u_{t-4} + \omega_t} \quad (3.60)$$

όπου τα $\boldsymbol{\varepsilon_t}$ και $\boldsymbol{\omega_t}$ είναι τυχαίες μεταβλητές και έχουν τις ίδιες ιδιότητες με την τυχαία μεταβλητή v_t στο σχήμα που έχουμε προαναφέρει. Το πρώτο σχήμα ονομάζεται **πλήρες αυτοπαλίνδρομο σχήμα δευτέρου βαθμού** και το δεύτερο σχήμα ονομάζεται **μη πλήρες σχήμα τετάρτου βαθμού**.

3.14.3 ΣΥΝΕΠΕΙΕΣ ΤΗΣ ΑΥΤΟΣΥΣΧΕΤΙΣΗΣ

Από την ανάλυση που προηγήθηκε συνεπάγεται ότι η ύπαρξη αυτοσυσχέτισης στις τιμές του στοχαστικού όρου έχει τις ακόλουθες συνέπειες :

1) Οι εκτιμητές ελαχίστων τετραγώνων εξακολουθούν να είναι γραμμικοί, αμερόληπτοι και συνεπείς, δεν είναι όμως αποτελεσματικοί.

2) Επειδή με την ύπαρξη αυτοσυσχέτισης διευρύνεται το διάστημα των τιμών που μπορεί να πάρουν οι εκτιμητές ελαχίστων τετραγώνων, δηλαδή, έχουν μεγαλύτερη διακύμανση, αυξάνεται η πιθανότητα να πάρουμε τιμές για τις παραμέτρους, που να απέχουν από την αληθινή τους τιμή.

3) Τα τυπικά σφάλματα των εκτιμητών των παραμέτρων του υποδείγματος που υπολογίζουμε, με βάση τα στοιχεία του δείγματος, από τους γνωστούς τύπους κατά κανόνα είναι υποεκτιμημένα. Αυτό επηρεάζει τις τιμές των στατιστικών ελέγχων και αυξάνει την πιθανότητα απόρριψης της μηδενικής υπόθεσης ακόμη και όταν αυτή είναι σωστή.

4) Η εκτίμηση της διακύμανσης του στοχαστικού όρου που προκύπτει από τον συνήθη τύπο με βάση τα κατάλοιπα ελαχίστων τετραγώνων είναι κατά κανόνα υποεκτιμημένη. Αυτό επηρεάζει προς τα πάνω την τιμή του συντελεστή προσδιορισμού.

5) Οι κριτικές τιμές των κριτηρίων ελέγχου κάτω από συνθήκες αυτοσυσχέτισης είναι φανερά μεγαλύτερες από αυτές που εμφανίζουν οι πίνακες, γεγονός που μπορεί να οδηγήσει σε λάθη.

6) Ακόμα αποδεικνύεται ότι οι προβλέψεις που γίνονται με βάση τους αρχικούς τύπους ελαχίστων τετραγώνων δεν είναι αποτελεσματικές.

Η συνδυασμένη λειτουργία των περιπτώσεων (3), (4) και (5) μπορεί να οδηγήσει σε σημαντικά σφάλματα σχετικά με την τιμή και την αξιοπιστία των επιμέρους συντελεστών και την κατασκευή των διαστημάτων εμπιστοσύνης. Για όλους αυτού τους λόγους είναι απαραίτητο να αναπτύξουμε κριτήρια ελέγχου της αυτοσυσχέτισης και τεχνικές εκτίμησης του υποδείγματος κάτω από τις νέες συνθήκες έτσι ώστε τα εμπειρικά αποτελέσματα να ανταποκρίνονται όσο είναι αυτό δυνατόν στην πραγματικότητα.

3.14.4 ΔΙΑΠΙΣΤΩΣΗ ΤΗΣ ΑΥΤΟΣΥΣΧΕΤΙΣΗΣ

Το κριτήριο που συνήθως χρησιμοποιείται στην εφαρμοσμένη έρευνα για τον έλεγχο της αυτοσυσχέτισης πρώτου βαθμού είναι αυτό των **Durbin – Watson**.

3.14.4.1 ΤΟ ΚΡΙΤΗΡΙΟ ΤΩΝ DURBIN – WATSON

Το κριτήριο αυτό δίνεται από τον τύπο :

$$\mathbf{d} = \frac{\sum_{t=2}^{n} (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1} \varepsilon^2_t} \qquad (3.61)$$

όπου τα ε_t παριστάνουν τα κατάλοιπα ελαχίστων τετραγώνων.

Οι **Durbin – Watson** κατόρθωσαν να δείξουν ότι η κριτική τιμή του **d** θα βρίσκεται πάντοτε μεταξύ των μεγεθών (στατιστικών) $\mathbf{d_L}$ και $\mathbf{d_U}$, τα οποία δεν εξαρτώνται από την **X**. Το $\mathbf{d_L}$ παριστάνει το κατώτερο όριο (lower) και το d_U το ανώτερο όριο (upper) των τιμών του **d**. Στη συνέχεια κατασκεύασαν κατανομές των $\mathbf{d_L}$ και $\mathbf{d_U}$ για διάφορα επίπεδα σημαντικότητας, δοσμένα μεγέθη δείγματος και διαφορετικό αριθμό μεταβλητών. Προτείνουν την απόρριψη της υπόθεσης $\mathbf{H_0}$ (ότι δεν υπάρχει αυτοσυσχέτιση) σε όλες τις περιπτώσεις που $\mathbf{d} < \mathbf{d_L}^*$ και αποδοχή της $\mathbf{H_0}$ όταν $\mathbf{d} > \mathbf{d_U}^*$, όπου ο αστερίσκος υποδηλώνει την κριτική τιμή.

Στην περίπτωση που ισχύει η σχέση $d_L < d < d_U$ το κριτήριο δεν οδηγεί σε συμπέρασμα.

Πριν προχωρήσουμε στον τρόπο χρησιμοποίησης του κριτηρίου, είναι χρήσιμο να αναφερθούμε θεωρητικά, χωρίς επεκτάσεις σε τύπους, στη σχέση που υπάρχει ανάμεσα στο **ρ** και στο **d**.

Ο συντελεστής αυτοσυσχέτισης **ρ** μπορεί να εκτιμηθεί από τον τύπο :

$$\rho = \frac{\sum_{t=2}^{n} \varepsilon_t \varepsilon_{t-1}}{\sqrt{\sum_{t=2} \varepsilon^2_t}\ \sqrt{\sum_{t=2} \varepsilon^2_{t-1}}} \qquad (3.62)$$

(*Σημείωση : Στην περίπτωση που τα δείγματα είναι μεγάλου μεγέθους το κριτήριο δίνεται από τον τύπο $\mathbf{d = 2\left(1 - \frac{\Sigma\varepsilon_t\,\varepsilon_{t-1}}{\Sigma\varepsilon^2_t}\right) = 2(1-\rho)}$ (3.63)

και ο συντελεστής αυτοσυσχέτισης υπολογίζεται από τον τύπο : $\mathbf{\rho = 1 - d / 2}$)

Έτσι έχουμε, ότι για $|\rho| < 1$, το d παίρνει τιμές μεταξύ **0** και **4**. Αναλυτικότερα :

i. Αν $\rho = 1$, τότε $d = 0$ (πλήρης θετική αυτοσυσχέτιση)
ii. Αν $\rho = 0$, τότε $d = 2$ (δεν υπάρχει αυτοσυσχέτιση)
iii. Αν $\rho = -1$, τότε $d = 4$ (πλήρης αρνητική αυτοσυσχέτιση)

Από τα παραπάνω προκύπτει ότι :

i. Αν $0 < \rho < 1$, τότε $0 < d < 2$ (περιοχή θετικής αυτοσυσχέτισης)
ii. Αν $-1 < \rho < 0$, τότε $2 < d < 4$ (περιοχή αρνητικής αυτοσυσχέτισης)

Είναι φανερό ότι η κατανομή του **d** είναι συμμετρική με μέσο το **2**. Τώρα μπορούμε να υποδιαιρέσουμε το σύνολο των δυνατών τιμών του d στα παρακάτω πέντε υποσύνολα:

i. Αν $\mathbf{0 < d < d_L}$: υπάρχει θετική αυτοσυσχέτιση
ii. Αν $\mathbf{d_L < d < d_U}$: το αποτέλεσμα είναι αβέβαιο
iii. Αν $\mathbf{d_L < d < 4 - d_U}$: δεν υπάρχει αυτοσυσχέτιση
iv. Αν $\mathbf{4 - d_U < d < 4 - d_L}$: το αποτέλεσμα είναι αβέβαιο
v. Αν $\mathbf{4 - d_L < d < 4}$: υπάρχει αρνητική αυτοσυσχέτιση

ΠΗΓΗ : Ανδρέας Δ. Κιντής (1999), « Στατιστικές και Οικονομετρικές Μέθοδοι », Αθήνα : Εκδόσεις Gutenberg.

Στη συνέχεια θα διατυπώσουμε τους ελέγχους υποθέσεων :

$\mathbf{H_0 : \rho = 0}$ (δεν υπάρχει αυτοσυσχέτιση)

έναντι των εναλλακτικών υποθέσεων :

$\mathbf{H_1 : \rho > 0}$ (υπάρχει θετική αυτοσυσχέτιση)
$\mathbf{H_1 : \rho < 0}$ (υπάρχει αρνητική αυτοσυσχέτιση)
$\mathbf{H_1 : \rho \neq 0}$ (υπάρχει απλώς αυτοσυσχέτιση)

Ο έλεγχος των υποθέσεων αυτών γίνεται με τη βοήθεια της τιμής του **d** και των τιμών $\mathbf{d_L}$ και $\mathbf{d_U}$ που παίρνουμε από τους σχετικούς πίνακες. Για την καλύτερη κατανόηση παραθέτουμε τον παρακάτω πίνακα :

Πίνακας (3.2)

Μέγεθος δείγματος	$K^* = 2$		$K^* = 4$	
	d_L	d_U	d_L	d_U
15	0.95	1.54	0.69	1.97
20	1.10	1.54	0.90	1.83
30	1.28	1.57	1.14	1.79
40	1.39	1.60	1.29	1.72

ΠΗΓΗ : Ανδρέας Δ. Κιντής (1999), « Στατιστικές και Οικονομετρικές Μέθοδοι », Αθήνα : Εκδόσεις Gutenberg.

(*) Το Κ παριστάνει τον αριθμό των ανεξάρτητων μεταβλητών.

(**)Τιμές των d_L και d_U σε επίπεδο σημαντικότητας 5%

Έστω, ότι κατά την εκτίμηση ενός υποδείγματος με δύο ανεξάρτητες μεταβλητές και n = 30, βρέθηκε ότι η τιμή του d ισούται με 1,10. Για να εξετάσουμε την ύπαρξη θετικής αυτοσυσχέτισης, τότε θα διατυπώσουμε τον έλεγχο υποθέσεων ως εξής :

$\mathbf{H_0 : \rho = 0}$ (δεν υπάρχει αυτοσυσχέτιση)

$\mathbf{H_1 : \rho > 0}$ (υπάρχει θετική αυτοσυσχέτιση)

Στη συγκεκριμένη περίπτωση έχουμε $\mathbf{d = 1,10 < d_L^* = 1,28}$. Η υπόθεση $\mathbf{H_0}$ απορρίπτεται και αποδεχόμαστε την H_1. Στην περίπτωση αρνητικής αυτοσυσχέτισης θα έχουμε :

$\mathbf{H_0 : \rho = 0}$ (δεν υπάρχει αυτοσυσχέτιση)

$\mathbf{H_1 : \rho < 0}$ (υπάρχει αρνητική αυτοσυσχέτιση)

Εάν η τιμή του **d** είναι **d = 2.90**, τότε με βάση τα δεδομένα του πίνακα θα έχουμε $\mathbf{4 - d_L^* = 2,72}$.

Κατά συνέπεια :

$$4 - d_L^* < d < 4$$

ή

$$2,72 < 2,90 < 4$$

Έτσι, συμπεραίνουμε ότι υπάρχει αρνητική αυτοσυσχέτιση. Τέλος για τον έλεγχο της υπόθεσης :

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

θα πρέπει να γνωρίζουμε τις τιμές d_L και d_U σε επίπεδο σημαντικότητας **2,5%.**

3.15 ΜΗ – ΓΡΑΜΜΙΚΑ ΩΣ ΠΡΟΣ ΤΙΣ ΜΕΤΑΒΛΗΤΕΣ ΥΠΟΔΕΙΓΜΑΤΑ ΠΟΛΛΑΠΛΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Πολλές φορές η σχέση που συνδέει την εξαρτημένη μεταβλητή με τις ανεξάρτητες μεταβλητές δεν είναι γραμμική. Από τα διαγράμματα των παρατηρήσεων του δείγματος και των ζευγών των παρατηρήσεων μεταξύ ερμηνευόμενης μεταβλητής και ερμηνευτικής (ή ερμηνευτικών) μεταβλητής μπορούμε συχνά να αντιληφθούμε αν το υπόδειγμα είναι ή όχι γραμμικό. Επίσης, η οικονομική θεωρία υποδεικνύει πολλές φορές ότι η σχέση μεταξύ της εξαρτημένης μεταβλητής και των ανεξάρτητων μεταβλητών είναι μη – γραμμική. Υπάρχουν περιπτώσεις όπου και τα μη – γραμμικά υποδείγματα μπορούν να μετασχηματιστούν με τέτοιο τρόπο, ώστε να πάρουμε μια γραμμική σχέση που να είναι μονότονος μετασχηματισμό τους, και στην οποία η αντιστοιχία μεταξύ των παραμέτρων του αρχικού και γραμμικού – μετασχηματισμένου υποδείγματος να είναι ακριβής. Αλλά όμως υποδείγματα δεν μπορούν να μετασχηματιστούν έτσι ώστε να πάρουμε μια ακριβή γραμμική σχέση τους, και τέτοια υποδείγματα και τρόπους εκτίμησής τους θα εξετάσουμε παρακάτω.

2. Μη – γραμμικά ως προς τις μεταβλητές υποδείγματα, που μέσω μετασχηματισμού τους είναι δυνατόν να εκτιμηθούν ως γραμμικά και να πάρουμε εκτιμήτριες για τις αρχικές παραμέτρους είδαμε στην απλή παλινδρόμηση. Εδώ θα αναφερθούμε σε δύο από αυτά προσαρμοσμένα στην πολλαπλή παλινδρόμηση σύμφωνα με τη μεθοδολογία που ακολουθείται από τον Ιωάννη Αλκιβ. Κασκαρέλη (1999), « Ένδεκα Μαθήματα Οικονομετρίας », 2η Αναθεωρημένη Έκδοση, Αθήνα : Εκδόσεις Gutenberg.

- ## ΤΟ ΠΟΛΥΩΝΥΜΙΚΟ ΥΠΟΔΕΙΓΜΑ

Το πρώτο είναι το ονομαζόμενο πολυωνυμικό υπόδειγμα, που μπορεί να παρουσιαστεί ως εξής :

$$\mathbf{Y_i = b_0 + b_1 Z_i + b_2 Z_i^2 + \ldots\ldots\ldots + b_k Z_i^k + \varepsilon_i} \qquad (3.64)$$

Εδώ βλέπουμε ότι η εξαρτημένη μεταβλητή $\mathbf{Y_i}$ παρουσιάζεται ως συνάρτηση ενός πολυωνύμου **κ – τάξης** της ερμηνευτικής μεταβλητής $\mathbf{Z_i}$. Επίσης οι δυνάμεις του $\mathbf{Z_i}$ εισέρχονται προσθετικά στο υπόδειγμα, το οποίο είναι γραμμικό ως προς τις παραμέτρους. Αν θέσουμε $\mathbf{Z_i = X_{i1}}$, $\mathbf{Z_i^2 = X_{i2}}$, $\mathbf{Z_i^k = X_{ik}}$, τότε το πολυωνυμικό υπόδειγμα μπορεί να γραφεί ως εξής :

$$\mathbf{Y_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + \ldots\ldots\ldots + b_k X_{i\kappa} + \varepsilon_i}$$

Αν τα σφάλματα πληρούν τις γνωστές υποθέσεις, τότε το υπόδειγμα αυτό μπορεί να εκτιμηθεί με την μέθοδο των ελαχίστων τετραγώνων. Πρέπει να επισημάνουμε ότι οι σχέσεις μεταξύ του $\mathbf{Z}_i$ και των δυνάμεών του είναι μη – γραμμικές, ότι ο πίνακας των ανεξάρτητων μεταβλητών είναι πλήρους βαθμού, και συνεπώς μπορούμε να βρούμε τις εκτιμήτριες των ελαχίστων τετραγώνων. Σε ποια δύναμη του $\mathbf{Z}_i$ θα σταματήσουμε, θα εξαρτηθεί από τους ελέγχους υποθέσεων σχετικά με τη στατιστική σημαντικότητα των συντελεστών των δυνάμεων του $\mathbf{Z}_i$. Πρέπει όμως να επισημάνουμε ότι, αν αποφασίσουμε για την τάξη του πολυωνύμου του $\mathbf{Z}_i$ μέσω των στατιστικών ελέγχων, θα πρέπει να διατηρήσουμε στην εξίσωση και όλες τις δυνάμεις του $\mathbf{Z}_i$ μικρότερης τάξης από το βαθμό του πολυωνύμου, είτε είναι στατιστικά σημαντικές είτε όχι.

Ένα παράδειγμα πολυωνυμικού υποδείγματος αποτελεί η συνάρτηση κόστους $\mathbf{C}_i$, η οποία αποτελεί ένα πολυώνυμο τρίτου βαθμού της παραγωγής $\mathbf{Q}_i$:

$$\mathbf{C_i = \alpha + \beta Q_i + \gamma Q_i^2 + \delta Q_i^3 + \varepsilon_i} \qquad (3.65)$$

- **<u>ΤΟ ΓΡΑΜΜΙΚΟ – ΛΟΓΑΡΙΘΜΙΚΟ ΥΠΟΔΕΙΓΜΑ</u>**

Μια άλλη πολύ γνωστή μη – γραμμική σχέση, η οποία προέρχεται από την οικονομική θεωρία, είναι η εκθετική συνάρτηση παραγωγής τύπου Cobb – Douglas, σύμφωνα με την οποία η παραγωγή $\mathbf{Q}_i$ εξαρτάται από την εργασία $\mathbf{N}_i$ και το κεφάλαιο $\mathbf{K}_i$ ως εξής :

$$\mathbf{Q_i = AN_i^{b1} * K_i^{b2} * e^{\varepsilon i}} \qquad (3.66)$$

Αυτό το υπόδειγμα είναι μη – γραμμικό ως προς τις μεταβλητές αλλά και τις παραμέτρους, είναι όμως δυνατόν να έχουμε μια ακριβή γραμμική σχέση του να πάρουμε τους λογαρίθμους :

$$\mathbf{lnQ_i = lnA + b_1lnN_i + b_2lnK_i + \varepsilon_i} \qquad (3.67)$$

Το γραμμικό – λογαριθμικό αυτό υπόδειγμα είναι γραμμικό πια ως προς τις παραμέτρους κλίσης, οι οποίες εκφράζουν οριακές αποδόσεις και συγχρόνως ελαστικότητα παραγωγής :

$$\mathbf{b_1 = \theta lnQ_i / \theta lnN_i = \frac{\theta Q_i / Q_i}{\theta N_i / N_i}} \qquad (3.68)$$

και

$$\mathbf{b_2 = \theta lnQ_i / \theta lnK_i = \frac{\theta Q_i / Q_i}{\theta K_i / K_i}} \qquad (3.69)$$

Οι μεταβλητές όμως είναι οι λογάριθμοι των αρχικών μεταβλητών.
Θέτοντας
$\mathbf{Y_i = lnQ_i}$, $\mathbf{X_{i1} = lnN_i}$ **και** $\mathbf{X_{i2} = lnKi}$, έχουμε το υπόδειγμα :

$$\mathbf{Y_i = lnA + b_1X_{i1} + b_2X_{i2} + \varepsilon_i} \qquad (3.70)$$

το οποίο, εφόσον τα σφάλματα πληρούν τις κλασικές υποθέσεις, μπορούμε να εκτιμήσουμε με τη μέθοδο των ελαχίστων τετραγώνων. Οι εκτιμήτριες που θα πάρουμε αποτελούν εκτιμήτριες των αρχικών παραμέτρων ως προς τις $\mathbf{b_1}$ και $\mathbf{b_2}$, ενώ η εκτιμήτρια του $\mathbf{e^{lnA}}$ αποτελεί την εκτιμήτρια του σταθερού όρου του αρχικού υποδείγματος.

Πρόβλημα στην εκτίμηση με τη μέθοδο των ελαχίστων τετραγώνων προκύπτει όταν ο όρος του σφάλματος εισάγεται πολλαπλασιαστικά στο αρχικό υπόδειγμα ως εξής :

$$\mathbf{Q_i = AN_i^{b1} * K_i^{b2} * \varepsilon_i}$$

Στην προκειμένη περίπτωση ο μέσος των σφαλμάτων στο μετασχηματισμένο υπόδειγμα, $\mathbf{E(ln\varepsilon_i)}$, δεν είναι μηδέν, αν έχουμε υποθέσει ότι ο μέσος των αρχικών σφαλμάτων είναι $E(\varepsilon_i) = 0$. Σε αυτή την περίπτωση το υπόδειγμα πρέπει να μετασχηματιστεί σε :

$$\mathbf{Y_i = [lnA + E(ln\varepsilon_i)] + b_1X_{i1} + b_2X_{i2} + [\ \varepsilon_i + E(ln\varepsilon_i)]} \quad (3.71)$$

όπου τώρα ο μέσος των σφαλμάτων του $\mathbf{u_i = ln\varepsilon_i - E(ln\varepsilon_i)}$ είναι $\mathbf{E(u_i) = 0}$. Βέβαια πριν καταφύγουμε σε αυτόν τον μετασχηματισμό καλό είναι, μέσω του γραμμικού – λογαριθμικού υποδείγματος, να ελέγξουμε κατά πόσο τα κατάλοιπα από την εκτίμησή του έχουν μέσο μηδέν. Αν πράγματι τα κατάλοιπα του εκτιμημένου γραμμικού – λογαριθμικού υποδείγματος έχουν μέσο μηδέν, τότε σημαίνει ότι η σωστή εξειδίκευση του όρου του σφάλματος, στην εκθετική συνάρτηση Cobb – Douglas, είναι να εισαχθεί σαν δύναμη του νεπέριου αριθμού.

Γνωρίζουμε ότι αυτό που λέμε στην οικονομική επιστήμη συνάρτηση παραγωγής Cobb – Douglas δεν είναι παρά μια μαθηματική εκθετική συνάρτηση με συγκεκριμένες ιδιότητες, οι οποίες μας διευκολύνουν αρκετά στην οικονομική ανάλυση. Έτσι, οι συναρτήσεις αυτές τύπου Cobb – Douglas δεν χρησιμοποιούνται στην οικονομική ανάλυση για να εκφράσουν μόνο συναρτήσεις παραγωγής, αλλά και πολλές άλλες οικονομικές σχέσεις. Οι σχέσεις τύπου Cobb – Douglas χαρακτηρίζονται σαν σχέσεις σταθερών ελαστικοτήτων και μοναδιαίας ελαστικότητας υποκατάστασης μεταξύ των συντελεστών παραγωγής (ή άλλων ανεξάρτητων μεταβλητών).Έτσι, για παράδειγμα,

Μπορεί να χρησιμοποιηθεί για να εκφράσει τη ζήτηση χρήματος $\mathbf{M_i}$ σαν συνάρτηση του εισοδήματος $\mathbf{I_i}$ και του επιτοκίου $\mathbf{R_i}$:

$$\mathbf{M_i = BR_i^{b1} * I_i^{b2} * e^{ui}} \quad (3.72)$$

ή τη ζήτηση προϊόντος Q_i, σαν συνάρτηση του εισοδήματος I_i και της τιμής του προϊόντος P_i :

$$Q_i = BP_i^{b2} * I_i^{b2} * e^{ui} \quad (3.73)$$

3.16 ΣΥΜΠΕΡΑΣΜΑ

Όπως λοιπόν εξετάσαμε παραπάνω οι μέθοδοι παλινδρόμησης που χρησιμοποιούν περισσότερες από μία ανεξάρτητες μεταβλητές ονομάζονται μέθοδοι πολλαπλής παλινδρόμησης. Στην πολλαπλή παλινδρόμηση έχουμε σφάλματα που οφείλονται στην ατελή θεωρία, σε ατελή εξειδίκευση και σε υπολογιστικά λάθη, πράγμα που είδαμε να συμβαίνει και στην απλή παλινδρόμηση.

Οι συντελεστές τις πολλαπλής παλινδρόμησης που εκτιμούνται με τη μέθοδο των ελαχίστων τετραγώνων μετρούν την καθαρή μεταβολή στην εξαρτημένη μεταβλητή που σχετίζεται με μία ανά μονάδα μεταβολή στην ανεξάρτητη μεταβλητή.

Ένα μέτρο που μας βοηθάει να δούμε τον βαθμό προσαρμοστικής ικανότητας του υποδείγματός μας είναι ο συντελεστής προσδιορισμού. Επίσης είναι πολύ σημαντικός και ο συντελεστής πολλαπλής συσχέτισης, ο οποίος είναι πάντα θετικός και αποτελεί μία γενίκευση της απλής συσχέτισης στην περίπτωση των δύο ή περισσοτέρων ανεξάρτητων μεταβλητών.

Για να ελέγξουμε τη συνολική σημαντικότητα του μοντέλου πολλαπλής παλινδρόμησης πρέπει να αναλύσουμε τη διακύμανση και να ελέγξουμε τη σημαντικότητα των επιμέρους μεταβλητών. Επίσης θα πρέπει να εξετάσουμε αν εμφανίζεται στο υπόδειγμά μας ετεροσκεδαστικότητα, πράγμα το οποίο σημαίνει ότι η διακύμανση των u_i δεν είναι σταθερή και αυτό είναι κάτι το οποίο δε θέλουμε για το υπόδειγμά μας, έπειτα θα ελέγξουμε αν εμφανίζεται αυτοσυσχέτιση, αν υπάρχει σημαίνει ότι τα **u_i** δεν είναι ανεξάρτητα μεταξύ τους και τυχαία, πράγμα το οποίο έρχεται σε αντίθεση με τις βασικές μας υποθέσεις και άρα το υπόδειγμα μας δεν είναι έγκυρο. Τέλος θα πρέπει να εξετάσουμε το φαινόμενο της πολυσυγγραμμικότητας οπού εμφανίζεται όταν δύο οι περισσότερες μεταβλητές σχετίζονται μεταξύ τους γραμμικά. Αυτό που μας ενδιαφέρει στην περίπτωση της πολυσυγγραμμικότητας είναι να δούμε όχι αν υπάρχει αλλά πόσο έντονη είναι. Η έντονη ύπαρξη πολυσυγγραμμικότητας δείχνει ότι το μοντέλο παλινδρόμησης δεν έχει την ικανότητα να διαχωρίζει την ακριβή συμβολή κάθε ανεξάρτητης μεταβλητής στην ερμηνεία της συμπεριφοράς της εξαρτημένης μεταβλητής.

Τελειώνοντας, η επιλογή του καλύτερου μοντέλου παλινδρόμησης απαιτεί ένα συμβιβασμό μεταξύ της επίτευξής του υψηλότερου συντελεστή προσδιορισμού **R_2** και της επιλογής ενός ελάχιστου αριθμού μεταβλητών πρόβλεψης. Το μοντέλο οριοθετείται από το αρχικό σύνολο των ανεξάρτητων μεταβλητών και τα επιμέρους επίπεδα σημαντικότητας που με τη βοήθεια τους απορρίπτουμε ή εισάγουμε μία ανεξάρτητη μεταβλητή έτσι ώστε στο τέλος να καταλήξουμε στην τελική εξίσωση πρόβλεψης.

ΚΕΦΑΛΑΙΟ 4ο

ΕΙΔΗ ΣΥΣΧΕΤΙΣΗΣ

4.1 ΕΙΣΑΓΩΓΗ

Τα υποδείγματα με τα οποία ασχοληθήκαμε στα προηγούμενα κεφάλαια αναφέρονται σε συναρτησιακές σχέσεις ανάμεσα σε δύο ή και περισσότερες μεταβλητές οι οποίες υποθέτουμε ότι είναι γραμμικές και ως προς τις μεταβλητές και ως προς τις παραμέτρους. Η μορφή όμως της συναρτησιακής σχέσης ανάμεσα σε δύο ή περισσότερες μεταβλητές, κυρίως οικονομικές, μπορεί να μην είναι γραμμική. Αυτό μπορεί να προκύπτει ως αποτέλεσμα των εμπειρικών δεδομένων ή από την οικονομική θεωρία. Αυτό που μας απασχολεί είναι το αν η ανάλυση και οι μέθοδοι εκτίμησης του γραμμικού υποδείγματος μπορούν να εφαρμοστούν και σε σχέσεις που δεν είναι γραμμικές. Όπως θα δούμε πολλές σχέσεις που δεν είναι γραμμικές μπορούν εύκολα να μετασχηματιστούν σε γραμμικές, οπότε οι μέθοδοι εκτίμησης του γραμμικού υποδείγματος μπορούν να εφαρμοστούν στις μετασχηματισμένες αλλά γραμμικές πια σχέσεις.

Στο κεφάλαιο αυτό θα ασχοληθούμε με τα διάφορα άλλα είδη συσχέτισης που μπορεί να παρουσιαστούν ανάμεσα σε μεταβλητές οι οποίες παρουσιάζουν κάποια αλληλεξάρτηση. Πιο συγκεκριμένα θα δούμε μη γραμμικές εξισώσεις παλινδρόμησης και θα δώσουμε αριθμητικά παραδείγματα καθώς και διαγράμματα έτσι ώστε να καταλάβουμε καλύτερα τη χρησιμότητα τους στην οικονομική θεωρία κυρίως. Επίσης θα προχωρήσουμε και στην ανάλυση της συσχέτισης δεδομένων που είναι χωρισμένα κατά τάξεις.

4.2 ΚΑΜΠΥΛΟΓΡΑΜΜΗ ΣΥΣΧΕΤΙΣΗ

Σύμφωνα με έρευνες που έχουν γίνει διαπιστώθηκε ότι το γραμμικό υπόδειγμα χρησιμοποιείται περισσότερο από τους οικονομικούς αναλυτές για δύο λόγους, ο πρώτος λόγος είναι ότι η πραγματική σχέση μεταξύ των δύο μεταβλητών είναι γραμμική και ο δεύτερος αφορά το εύρος των τιμών που καλύπτει το δείγμα. Ακόμα και σε περιπτώσεις που η πραγματική σχέση των μεταβλητών είναι καμπυλόγραμμη, εάν το δείγμα καλύπτει σχετικά περιορισμένο εύρος τιμών, οι τιμές του δείγματος δεν επαρκούν για να αποκαλύψουν τη μη-γραμμική σχέση και έτσι ο αναλυτής περιορίζεται στην εκτίμηση του απλού γραμμικού υποδείγματος.

Υπάρχουν βέβαια περιπτώσεις που τα δεδομένα μας αποκαλύπτουν μη-γραμμική σχέση μεταξύ των μεταβλητών. Υπάρχουν πολλές περιπτώσεις που η σχέση μεταξύ των μεγεθών είναι καμπυλόγραμμη. Για παράδειγμα η ζήτηση ηλεκτρικής ενέργειας αυξάνεται εκθετικά σε σχέση με τον πληθυσμό. Η αύξηση των πωλήσεων σε διάφορους κλάδους παρουσιάζει διαδοχικά μειούμενο ρυθμό μεταβολής σε σχέση με την αύξηση των δαπανών για διαφήμιση.

Η μαθηματική μορφή έκφρασης των μη γραμμικών σχέσεων μεταξύ δύο μεταβλητών παρουσιάζεται με το εξής υπόδειγμα :

$$\mathbf{Y} = \boldsymbol{\beta}_0 * \mathbf{X}^{\beta 1} * \mathbf{e}_u \qquad (4.1)$$

όπου :

Y = η τιμή της εξαρτημένης μεταβλητής

X = η τιμή της ανεξάρτητης μεταβλητής

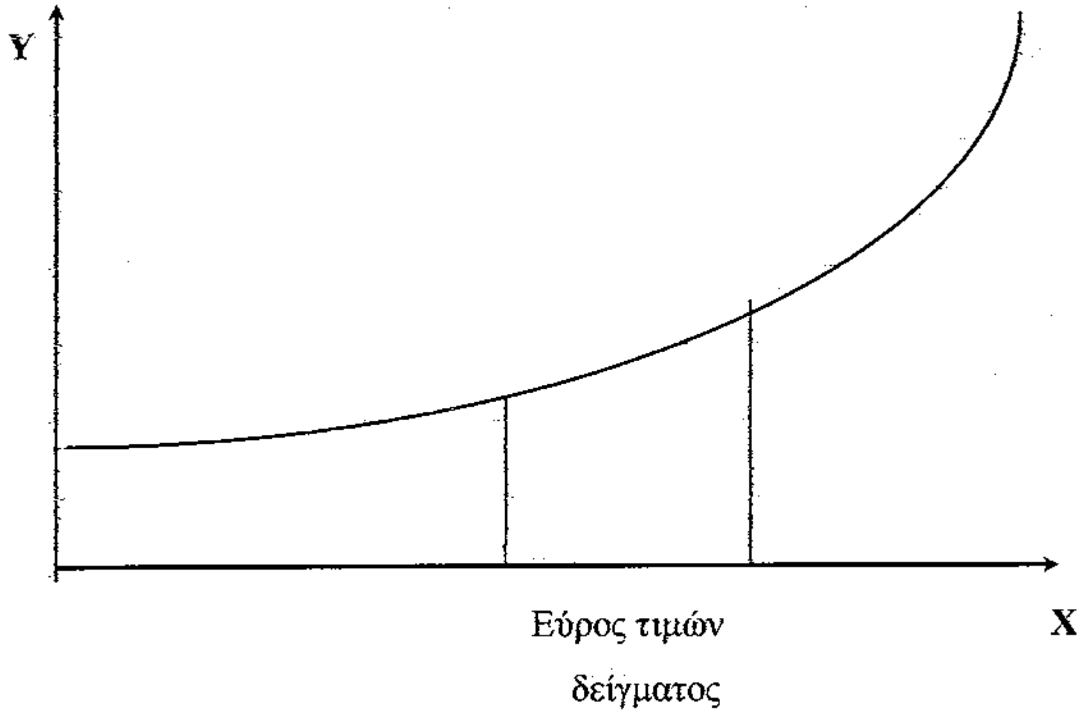
$\boldsymbol{\beta}_0$ = μία σταθερά

$\boldsymbol{\beta}_1$ = η τιμή του συντελεστή ελαστικότητας

e = η βάση των νεπέρειων λογαρίθμων

u = το σφάλμα ή κατάλοιπο (η διαφορά μεταξύ της πραγματικής τιμής της **ln(Y)** και της πρόβλεψης $\mathbf{ln(\hat{Y})}$ που προκύπτει από το υπόδειγμα) που κατανέμεται κανονικά με μέσο το $\mu = 0$ και διακύμανση σ^2_u

Διάγραμμα 4.1



Το κύριο χαρακτηριστικό του υποδείγματος (4.1) είναι ότι ο συντελεστής ελαστικότητας είναι σταθερός για όλες τις τιμές των μεταβλητών **X** και **Y**. Αυτό αποδεικνύεται ως εξής:

$$\eta_{Y/X} = \frac{dY}{dX} * \frac{X}{Y} = \beta_1 * \beta_0 * X^{\beta_1 - 1} * \frac{X}{Y} =$$

$$= \beta_1 * \beta_0 * X^{\beta_1} * \frac{1}{\beta_0 * X^{\beta_1}} = \beta_1 \qquad (4.2)$$

Από τη σχέση (4.2) προκύπτει ότι το υπόδειγμα (4.1) ενσωματώνει την υπόθεση ότι η ποσοστιαία μεταβολή της **Y** , που αντιστοιχεί σε μία ποσοστιαία μεταβολή της **X**, είναι ίδια ανεξάρτητα από την τιμή της **X**. Η υπόθεση της σταθερής ελαστικότητας συναντιέται συχνά στις σχέσεις των οικονομικών μεγεθών. Για παράδειγμα η ελαστικότητα ζήτησης ως προς το εισόδημα, η ελαστικότητα ζήτησης ή προσφοράς ως προς τις τιμές και άλλα πολλά.

Η μορφή της καμπύλης που αντιστοιχεί στην εξίσωση (4.1) εξαρτάται από την τιμή του συντελεστή β_1. Αν ο συντελεστής είναι μεγαλύτερος της μονάδας τότε η **Y** αυξάνεται εκθετικά σε σχέση με τη **X**. Τα παρακάτω διαγράμματα απεικονίζουν όλες τις μορφές της σχέσης (4.1)

Το υπόδειγμα :

$$\mathbf{Y} = \beta_0 * \mathbf{X}^{\beta 1} * \mathbf{e}^{u}$$

δε μπορεί να εκτιμηθεί με την μέθοδο των ελαχίστων τετραγώνων γιατί το υπόδειγμα είναι μη γραμμικό ως προς τις παραμέτρους β_0 και β_1 που σημαίνει ότι η παραγώγιση του αθροίσματος των τετραγώνων των αποκλίσεων ως προς τα β_0 και β_1 οδηγεί δε μη γραμμικό σύστημα εξισώσεων. Η εκτίμηση του υποδείγματος (4.1) γίνεται με έμμεσο τρόπο λογαριθμίζοντας και τα δύο μέλη της εξίσωσης.

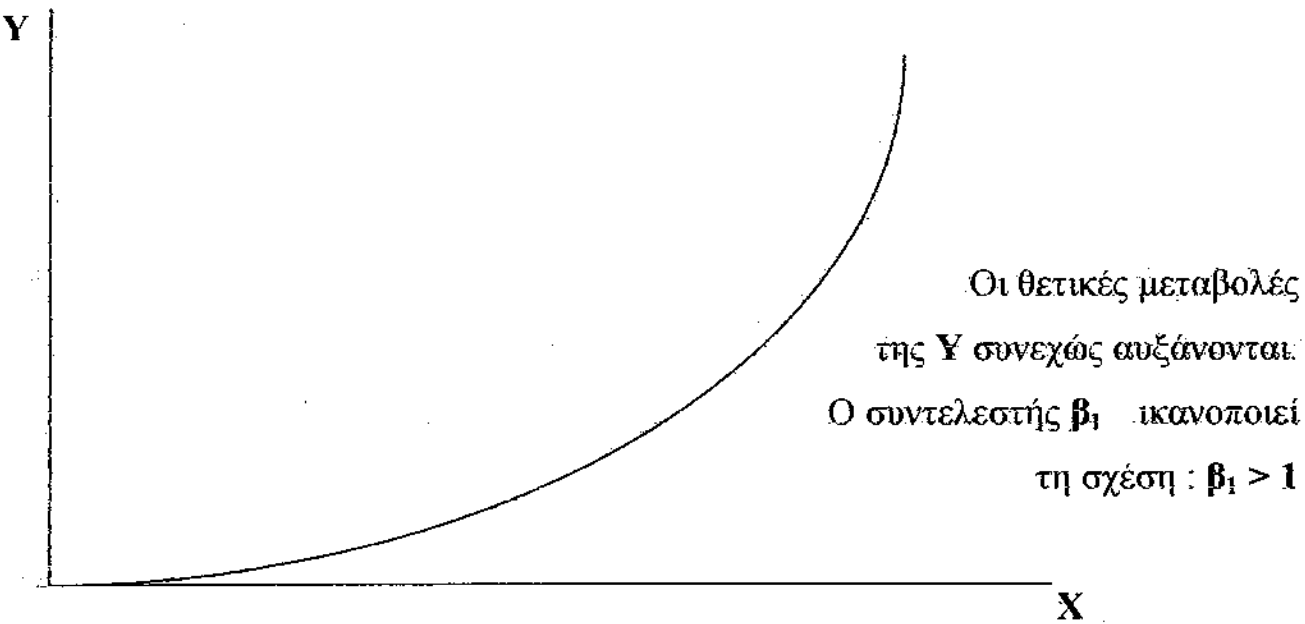
$$\ln(\mathbf{Y}) = \ln(\beta_0 * \mathbf{X}^{\beta 1} * \mathbf{e}^{u})$$

ή

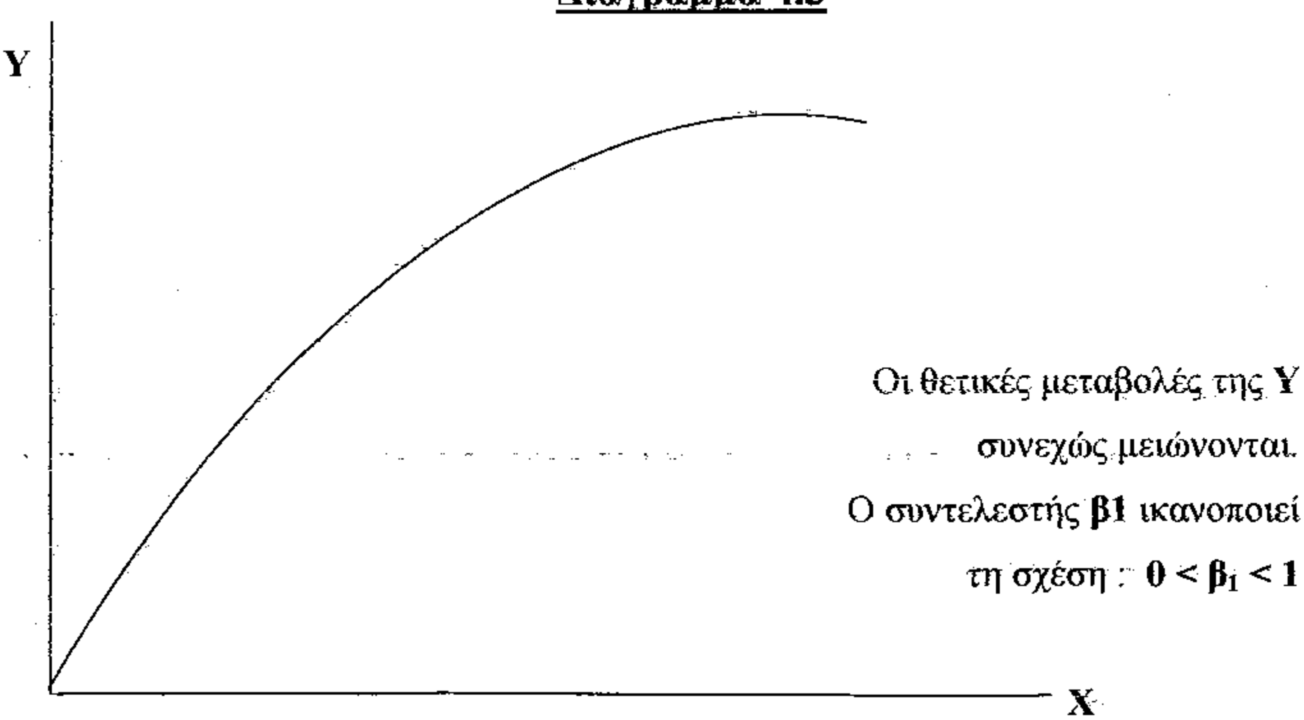
$$\ln(\mathbf{Y}) = \ln(\beta_0) + \beta_1 * \ln(\mathbf{X}) + \mathbf{u} \qquad (4.3)$$

4.2.1 ΜΟΡΦΗ ΤΗΣ ΣΧΕΣΗΣ $Y = \beta_0 * X^{\beta 1}$ ΓΙΑ ΔΙΑΦΟΡΕΣ ΤΙΜΕΣ ΤΟΥ ΣΥΝΤΕΛΕΣΤΗ β_1

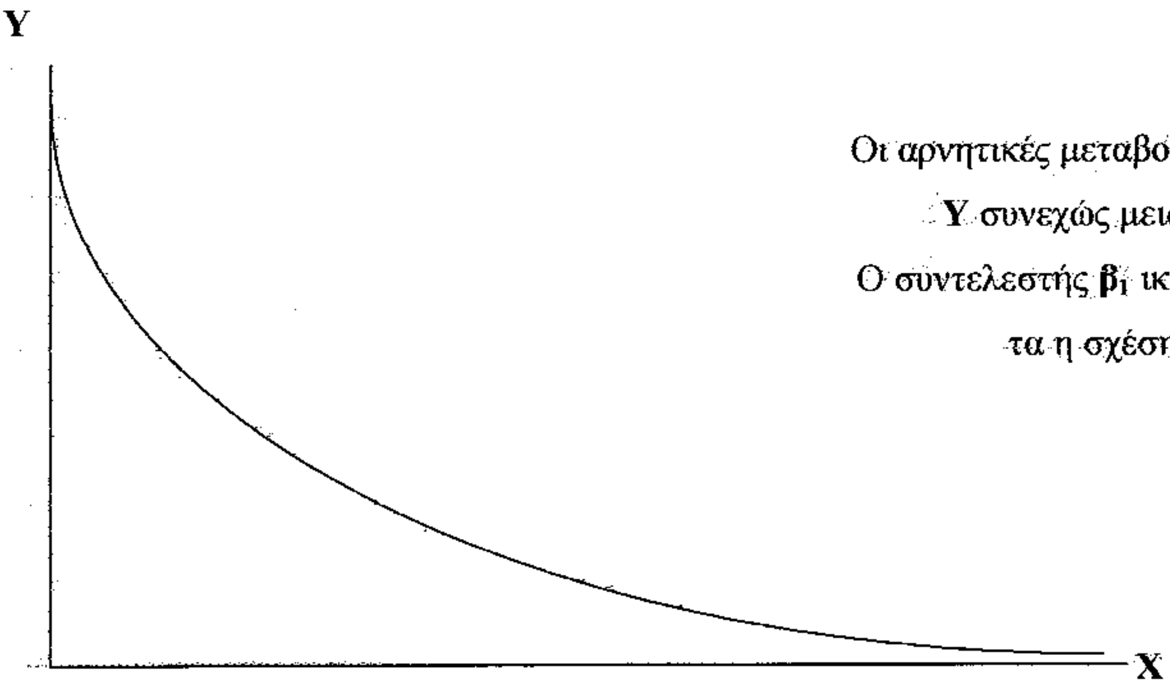
Διάγραμμα 4.2



Διάγραμμα 4.3



Διάγραμμα 4.4



Όλες οι υποθέσεις για την εφαρμογή της μεθόδου ελαχίστων τετραγώνων ισχύουν και σε αυτό το υπόδειγμα αφού το σφάλμα u κατανέμεται κανονικά με μέσο το μηδέν και σταθερή διακύμανση ίση με σ^2_u.

Άρα η προς εκτίμηση εξίσωση είναι :

$$\hat{Y} = b_0 * X^{b_1} \qquad (4.4)$$

ή

$$\ln(\hat{Y}) = \ln(b_0) + b_1 * \ln(X) \qquad (4.5)$$

Στη συνέχεια αν ακολουθήσουμε την εξής διαδικασία δηλαδή υπολογίσουμε το άθροισμα των τετραγώνων των αποκλίσεων μεταξύ των **ln(Y)** και **ln ($\hat{Y}$)** και προχωρήσουμε σε παραγώγιση ως προς **ln(b_0)** και **b_1** προκύπτει το παρακάτω σύστημα εξισώσεων :

$$\Sigma \ln(Y) = n \ln(b_0) + b_1 \Sigma \ln(X)$$

$$\Sigma \ln(Y)\ln(X) = \ln(b_0)\Sigma \ln(X) + b_1 \Sigma \ln(X)^2 \quad (4.6)$$

Στη συνεχεία αν επιλύσουμε το σύστημα μπορούμε να βρούμε το συντελεστή **ln(b_0)** αλλά και ότι άλλο θέλουμε.

4.2.2 ΕΦΑΡΜΟΓΗ

Ας προχωρήσουμε λοιπόν σε ένα παραδείγμα για την κατανόηση όσων είπαμε. Μία εταιρεία που πραγματοποιεί έρευνες αγορών ασχολείται με τη ζήτηση ενός είδους χυμού και από ένα τυχαίο δείγμα 14 οικογενειών που αποτελούνται από 3 άτομα συλλέχθηκαν τα εξής στοιχεία τα οποία φαίνονται στον παρακάτω πίνακα.

Πίνακας (4.1) Ποσότητα και Τιμή Αγοράς

Οικογένεια	Μέση Τιμή Αγοράς (σε ευρώ)	Ποσότητα (Μονάδες)
1	1,41	23
2	1,45	26
3	1,67	17
4	1,33	29
5	1,62	18
6	1,58	22
7	1,78	16
8	1,51	24
9	1,73	19
10	1,85	18
11	1,48	21
12	1,53	20
13	1,35	25
14	1,38	28

ΠΗΓΗ: Ιωάννης Γ. Χαλικιάς (2001), « Στατιστική – Μέθοδοι για Επιχειρηματικές Αποφάσεις », Αθήνα : Εκδόσεις Rosili.

Ο πίνακας 4.2 δίνει τον τρόπο υπολογισμού του συντελεστή συσχέτισης για τα παραπάνω δεδομένα.

Πίνακας (4.2) Εκτίμηση Συντελεστή Συσχέτισης Ποσότητας και Τιμής

Τιμή (X)	Ποσότητα (Y)	ln (X)	ln(Y)	$[\ln(X)]^2$	$[\ln(Y)]^2$	ln(X)*ln(Y)
1,41	23	0,344	3,135	0,118	9,831	1,077
1,45	26	0,372	3,258	0,138	10,615	1,211
1,67	17	0,513	2,833	0,263	8,027	1,453
1,33	29	0,285	3,367	0,081	11,339	0,960
1,62	18	0,482	2,890	0,233	8,354	1,394
1,58	22	0,457	3,091	0,209	9,555	1,414
1,78	16	0,577	2,773	0,332	7,687	1,599
1,51	24	0,412	3,178	0,170	10,100	1,310
1,73	19	0,548	2,944	0,300	8,670	1,614
1,85	18	0,615	2,890	0,378	8,354	1,778
1,48	21	0,392	3,045	0,154	9,269	1,194
1,53	20	0,425	2,996	0,181	8,974	1,274
1,35	25	0,300	3,219	0,090	10,361	0,966
1,38	28	0,322	3,332	0,104	11,104	1,073
	Σ =	6,045	42,952	2,752	132,241	18,317

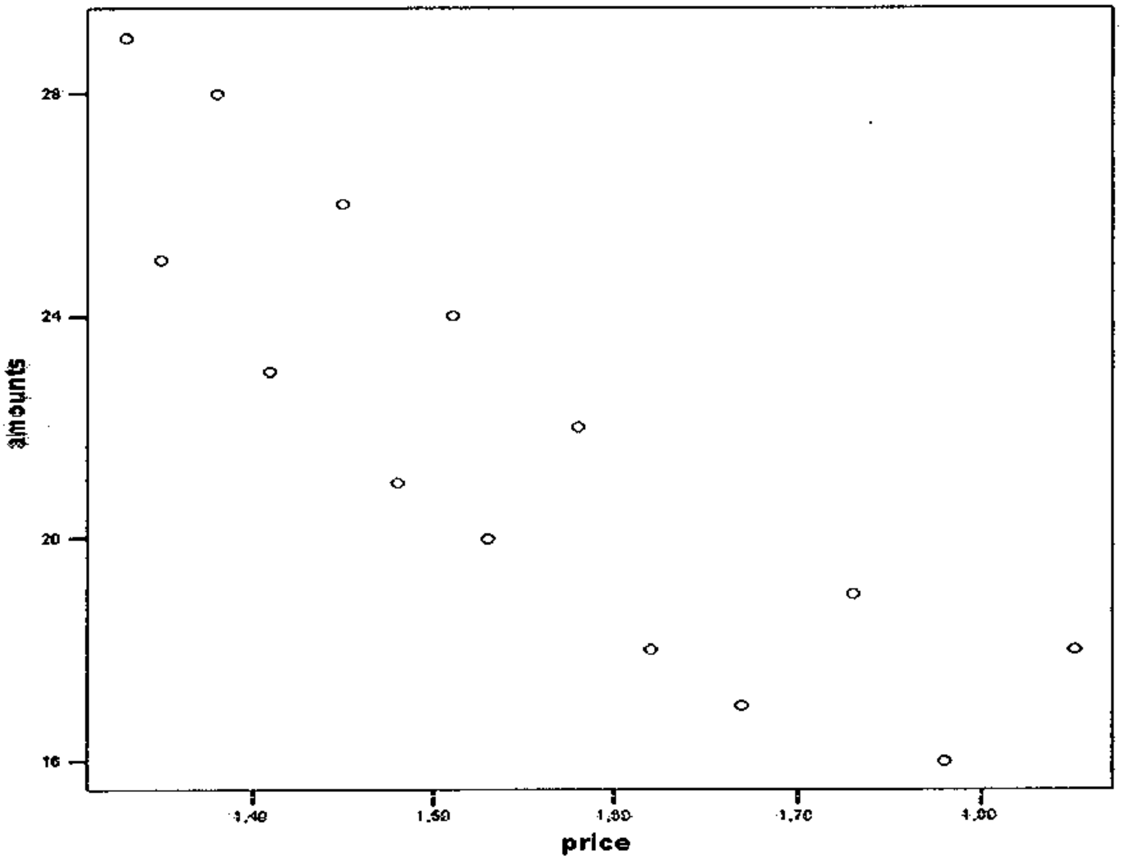
$$\mathbf{r} = \frac{n\Sigma\ln(X)\ln(Y) - \Sigma\ln(X)\Sigma\ln(Y)}{\sqrt{\{n[\Sigma\ln(X)^2\,] - \}\ \{n[\Sigma\ln(Y)^2\,] - [\Sigma\ln(Y)]^2\ \}}} =$$

$$= \frac{14(18{,}317) - (6{,}045)(42{,}953)}{\sqrt{[14\,(2{,}752) - (6{,}045)^2]\ [14\,(132{,}241) - (42{,}952)^2]}} =>$$

r = - 0,89

Όπως βλέπουμε ο συντελεστής συσχέτισης ανάμεσα στην ποσότητα και την τιμή είναι αρνητικός -0,89 και επειδή πλησιάζει τη μονάδα αυτό σημαίνει ότι υπάρχει μία **έντονη αρνητική γραμμική συσχέτιση μεταξύ των λογαρίθμων** των μεταβλητών ποσότητα και τιμή.

Διάγραμμα 4.5 Διάγραμμα Διασποράς Ποσότητας και Τιμής



Στο διάγραμμα 4.3 απεικονίζεται η διασπορά των ζευγών των παρατηρήσεων των μεταβλητών **X** και **Y**. Διαπιστώνουμε ότι η σχέση μεταξύ των μεταβλητών είναι μη γραμμική, άρα το πιο κατάλληλο υπόδειγμα για να αναλύσουμε τη σχέση μεταξύ της ποσότητας **(Y)** και της τιμής **(X)** είναι της μορφής του υποδείγματος που αναλύσαμε παραπάνω δηλαδή :

$$\mathbf{Y = \beta_0 * X^{\beta 1} * e_u}$$

Στη συνέχεια θα ασχοληθούμε με την εκτίμηση του γραμμικού λογαριθμικού υποδείγματος (4.5) χρησιμοποιώντας τα αποτελέσματα του πίνακα 4.2. Αν προχωρήσουμε στην επίλυση του συστήματος των εξισώσεων (4.6) προκύπτει το εξής :

$$\mathbf{b_1 = \frac{n\Sigma ln(X)\ ln(Y) - \Sigma ln(X)\ \Sigma ln(Y)}{n\ [\Sigma ln(X)^2\ ] - [\Sigma ln(X)]^2}} =$$

$$= \frac{14(18{,}317) - (6{,}045)\ (42{,}952)}{14(2{,}752) - (6{,}045)^2} =>$$

$$\mathbf{b_1 = -\ 1{,}604}$$

και

$$\mathbf{ln\ (b_0)\ = \Sigma ln(Y)/n - b1\Sigma ln(X)/n =}$$

$$= 42{,}952/14 - (-1{,}604) * 6{,}045/14 =>$$

$$\mathbf{ln\ (b_0) = 3{,}760}$$

Η εξίσωση παλινδρόμησης ως προς τους λογαρίθμους των μεταβλητών **X** και **Y** είναι :

$$\mathbf{ln\ (\hat{Y}) = 3{,}760 - 1{,}604\ ln(X)}$$

και το μη γραμμικό υπόδειγμα ως προς τις αρχικές μεταβλητές X και Y είναι :

$$\mathbf{\hat{Y} = e^{\ 3{,}76} * X^{\ -1{,}604}}$$

Αν τώρα προχωρήσουμε στην θεωρητική ερμηνεία της παραπάνω εξίσωσης προκύπτουν τα εξής : ότι ανεξάρτητα από το επίπεδο της τιμής του προϊόντος, αν η τιμή μεταβληθεί κατά 1% τότε η μέση ανά οικογένεια ποσότητα που καταναλώνεται θα μεταβληθεί κατά 1,604 %. Άρα για να το πούμε και με άλλα λόγια η σταθερή ελαστικότητα της ποσότητας ως προς την τιμή είναι – 1.6%.

4.3 ΑΛΛΕΣ ΣΥΝΑΡΤΗΣΙΑΚΕΣ ΜΟΡΦΕΣ ΥΠΟΔΕΙΓΜΑΤΩΝ

4.3.1 ΜΟΡΦΗ LOG – LINEAR (ΓΡΑΜΜΙΚΗ ΛΟΓΑΡΙΘΜΙΚΗ)

Με αυτό το είδος συνάρτησης μετράμε την ελαστικότητα κυρίως. Η μορφή της είναι η παρακάτω :

$$\mathbf{\ln Y_i = B_1 + B_2 \ln X_i + u_i} \quad (4.7)$$

* Παρατηρούμε ότι αυτή η μορφή έχει λογαρίθμους και στα δύο σκέλη της.

Προχωρώντας σε έναν μετασχηματισμό της παραπάνω μορφής προκύπτει η ακόλουθη συνάρτηση :

$$\mathbf{\ln Y_i = Y_i\,^*} \qquad\qquad \mathbf{\ln X_i = X_i\,^*}$$

άρα τότε έχουμε : $\mathbf{Y_i\,^* = B_1 + B_2\, X_i\,^* + u_i}$ **(1.8)**

Ας προχωρήσουμε όμως σε ένα παράδειγμα για να κατανοήσουμε τη λειτουργία της παραπάνω μορφής.

Παράδειγμα

Πίνακας 4.3 Ποσότητάς και τιμής

Ποσότητα Y	Τιμή X	ln Y	ln X
49	1	3,89	0
45	2	3,80	0,70
44	3	3,78	1,10
39	4	3,66	1,38
38	5	3,63	1,61
37	6	3,61	1,79
34	7	3,52	1,94
33	8	3,50	1,08
30	9	3,40	2,20
29	10	3,37	2,30

ΠΗΓΗ: Damodar Gujarati (1999), « Essentials of Econometrics », 2nd Edition McGraw Hill.

Η εξίσωση που προκύπτει από τα παραπάνω δεδομένα είναι η ακόλουθη :

$$\ln \hat{Y}_i = 3{,}96 - 0{,}227 \ln X_i$$

όπου **Y** = η ζητούμενη ποσότητα σε μονάδες

X = η τιμή σε euro

Αν προχωρήσουμε στην οικονομική ανάλυση της παραπάνω συνάρτησης παρατηρούμε ότι αν μεταβληθεί η τιμή κατά μία ποσοστιαία μονάδα θα μειωθεί η ζητούμενη ποσότητα κατά 2,27 %.

Για να αποφασίσουμε ποια συναρτησιακή μορφή θα χρησιμοποιήσουμε θα κάνουμε πρώτα ένα διάγραμμα διασποράς με τη βοήθεια του **SPSS** και αφού το μελετήσουμε θα διαπιστώσουμε αν οι μεταβλητές μας έχουν γραμμική σχέση. Αν δεν έχουν τότε θα λογαριθμίσουμε και το πιο πιθανό είναι αν ξανακάνουμε στη συνέχεια διάγραμμα διασποράς να μας δείξει ότι υπάρχει γραμμική σχέση.

4.2.2 ΜΟΡΦΗ COBB – DOUGLAS (ΣΥΝΑΡΤΗΣΗ ΠΑΡΑΓΩΓΗΣ)

Η συναρτησιακή μορφή της Cobb – Douglas είναι :

$$\ln Y_i = B_1 + B_2 \ln X_{2i} + B_3 \ln X_{3i} + u_i \quad (4.9)$$

Σε αυτή τη μορφή η μερική κλίση των συντελεστών B_2 και B_3 καλείται επίσης μερική ελαστικότητα των συντελεστών. Συνεπώς ο συντελεστής B_2 μετράει την ελαστικότητα του **Y** ως προς X_2 παραμένοντας η επίδραση της μεταβλητής X_3 σταθερή, έτσι μετράμε το ποσοστό μεταβολής της **Y** αν μεταβληθεί κατά **1%** η μεταβλητή X_2 παραμένοντας η επίδραση της μεταβλητής X_3 σταθερή. Επειδή η επίδραση της X_3 μένει σταθερή για αυτό ονομάζεται μερική ελαστικότητα. Ομοίως ο συντελεστής B_3 μετράει τη μερική ελαστικότητα του **Y** σε σχέση με τη μεταβλητή X_3 παραμένοντας η επίδραση της X_2 σταθερή. Εν συντομία, σε ένα πολλαπλό **Log – Linear** μοντέλο κάθε μερική κλίση κάθε συντελεστή μετράει τη μερική ελαστικότητα της εξαρτημένης μεταβλητής σε σχέση με την επεξηγηματική (ανεξάρτητη) μεταβλητή περί της οποίας γίνεται λόγος κάθε φορά, παραμένοντας η επίδραση των άλλων μεταβλητών σταθερή.

Ας προχωρήσουμε όμως σε ένα παράδειγμα για να κατανοήσουμε τη λειτουργία της παραπάνω μορφής καλύτερα. Θα ασχοληθούμε με τη συνολική παραγωγή σε σχέση με τις εισροές σε εργασία και το κεφάλαιο για το Μεξικό για τα έτη 1981 έως 2000.

Παράδειγμα

Πίνακας 4.4 Συνολική Παραγωγή σε σχέση με τις Εισροές σε Εργασία και το Κεφάλαιο

ΕΤΟΣ	ΣΥΝΟΛΙΚΗ ΠΑΡΑΓΩΓΗ (σε εκατ. πέσος)	ΕΡΓΑΣΙΑ (χιλιάδες άτομα)	ΚΕΦΑΛΑΙΟ (σε εκατ. πέσος)
1981	114043	8310	182113
1982	120410	8529	193749
1983	129187	8738	205192
1984	134705	8952	215130
1985	139960	9171	225021
1986	150511	9569	237026
1987	157897	9527	248897
1988	165286	9662	260661
1989	178491	10334	275466
1990	199457	10981	295378
1991	212323	11746	315715
1992	226977	11521	337642
1993	241194	11540	363599
1994	260881	12066	391847
1995	277498	12297	422382
1996	296530	12955	455049
1997	306712	13338	454677
1998	329030	13738	520553
1999	354057	15924	561531
2000	374977	14154	609825

ΠΗΓΗ : Damodar Gujarati (1999), « Essentials of Econometrics », 2nd Edition McGraw Hill.

Η εξίσωση που προκύπτει από τα παραπάνω δεδομένα είναι η ακόλουθη :

$$\ln\hat{Y}_t = -\,1{,}6524 + 0{,}3397\ \ln X_{2t} + 0{,}8460\ \ln X_{3t}$$

όπου **$\ln Y_t$** = η συνολική παραγωγή, **$\ln X_{2t}$** = οι εισροές σε εργασία,

$\ln X_{3t}$ = το κεφάλαιο

Αν προχωρήσουμε στην οικονομική ανάλυση της παραπάνω συνάρτησης παρατηρούμε ότι διατηρώντας σταθερή η επίδραση του κεφαλαίου, αν οι εισροές εργασίας αυξηθούν κατά **1%**, κατά μέσο όρο η συνολική παραγωγή θα αυξηθεί κατά **0,34%** . Παρομοίως διατηρώντας σταθερή την επίδραση των εισροών εργασίας, αν το κεφάλαιο αυξηθεί κατά **1%** , κατά μέσο όρο η συνολική παραγωγή θα αυξηθεί κατά **0,84%**. Αν προσθέσουμε των συντελεστών θα έχουμε μία οικονομικώς σημαντική παράμετρο η οποία ονομάζεται οικονομία κλίμακος. Και μας δίνει την ανταπόκριση της συνολικής παραγωγής ανάλογα με τις εισροές σε εργασία και κεφάλαιο. Αν το άθροισμα των δύο ελαστικοτήτων των συντελεστών είναι μονάδα έχουμε σταθερή οικονομία κλίμακος. Αν είναι μικρότερο της μονάδας έχουμε φθίνουσα οικονομία κλίμακος και αν είναι μεγαλύτερο τότε έχουμε αύξουσα οικονομία κλίμακος.

Για το Μεξικό που κάναμε την έρευνα αν κάνουμε την πρόσθεση προκύπτει ο αριθμός **1,1857 >1** άρα αυτό σημαίνει ότι η οικονομία του Μεξικού για την χρονική περίοδο που ερευνήσαμε παρουσίαζε αύξουσα οικονομία κλίμακος.

4.2.3 ΜΟΡΦΗ SEMI LOG (ΗΜΙΛΟΓΑΡΙΘΜΙΚΗ ΜΟΡΦΗ)

Τα **Semi log** μοντέλα χωρίζονται σε **log – lin** όπου ο λογάριθμος βρίσκεται στη μεταβλητή **Y** και στα **lin – log** όπου ο λογάριθμος βρίσκεται στη μεταβλητή **X**.

1. Ας αναφερθούμε λοιπόν πρώτα στη συναρτησιακή μορφή της **Semilog log – lin** η οποία είναι:

$$Y_t = Y_0 (1 + r)^t \quad (4.10)$$

Στα οικονομικά μαθηματικά ο παραπάνω τύπος είναι ο τύπος του ανατοκισμού.

όπου Y_0 = η αρχική τιμή της **Y**

Y_t = η τιμή της **Y** στο χρόνο **t**

r = ο ρυθμός αύξησης της **Y**

Προχωρώντας στον ανασχηματισμό του παραπάνω τύπου και πηγαίνοντας και στις δύο πλευρές το λογάριθμο έχουμε :

$$\ln Y_t = \ln Y_0 + t \ln (1+r)$$

Έστω ότι : $B_1 = \ln Y_0$, $B_2 = \ln (1+r)$

Έτσι προκύπτει το μοντέλο $\ln Y_t = B_1 + B_2 t + u_t$ **(4.11)**

Αυτό το μοντέλο είναι σαν κάθε άλλο γραμμικό μοντέλο παλινδρόμησης στο οποίο οι παράμετροι B_1 και B_2 είναι γραμμικοί. Η μόνη διαφορά είναι ότι η εξαρτημένη μεταβλητή είναι ο λογάριθμος της Υ και η ανεξάρτητη ή επεξηγηματική είναι ο χρόνος.

Μοντέλα σαν το παραπάνω ονομάζονται **Semilog** (ημιλογαριθμικά) επειδή μόνο μία μεταβλητή, στην συγκεκριμένη περίπτωση εμφανίζεται με λογαριθμική μορφή.

Ας προχωρήσουμε όμως σε ένα παράδειγμα για να κατανοήσουμε την παραπάνω μορφής. Θα ασχοληθούμε με το ρυθμό αύξησης των καταναλωτικών δανείων στις ΗΠΑ από το 1990 έως το 2004.

Παράδειγμα

Πίνακας 4.5 Ύψος καταναλωτικών δανείων στις ΗΠΑ από το 1990 έως το 2004

ΈΤΟΣ	Υ (σε εκατ. $)
1990	190,601
1991	199,365
1992	204,963
1993	228,162
1994	263,808
1995	308,272
1996	347,507
1997	349,386
1998	366,597
1999	381,115
2000	430,382
2001	511,768
2002	592,409
2003	646,055
2004	685,545

ΠΗΓΗ : Damodar Gujarati (1999), « Essentials of Econometrics », 2nd Edition McGraw Hill.

Η εξίσωση που προκύπτει από τα παραπάνω δεδομένα είναι η ακόλουθη :

$$\ln Y_t = 12{,}007 + 0{,}0946\, t$$

Αν προχωρήσουμε στην οικονομική ανάλυση της παραπάνω συνάρτησης παρατηρούμε ότι κάθε χρόνο το ύψος των καταναλωτικών δανείων αυξάνεται κατά **9,46%**.

Η κλίση στο μοντέλο **log – lin** μετράει τη σχετική μεταβολή της εξαρτημένης μεταβλητής **Y** για μία δεδομένη απόλυτη μεταβολή της ανεξάρτητης μεταβλητής **X**, που στη συγκεκριμένη περίπτωση είναι ο χρόνος.

2. Ας αναφερθούμε λοιπόν τώρα στη συναρτησιακή μορφή της **Semilog Lin – Lon** η οποία είναι:

Η συναρτησιακή μορφή της **Semilog Lin – Lon** είναι :

$$Y_t = B_1 + B_2 \ln X_{2t} + u_t \quad (4.12)$$

Βλέπουμε λοιπόν ότι σε αυτή τη μορφή η ανεξάρτητη μεταβλητή είναι σε λογαριθμική μορφή. Επίσης το B_2 εκφράζει την σχετική μεταβολή του **Y** ως προς την απόλυτη μεταβολή του **X**.

$$B_2 = \frac{\Delta_Y}{\Delta_X / X}$$

Ας προχωρήσουμε όμως σε ένα παράδειγμα για να κατανοήσουμε τη λειτουργία της παραπάνω μορφής. Θα εξετάσουμε λοιπόν τη σχέση ανάμεσα στο ΑΕΠ και την προσφορά χρήματος στις ΗΠΑ για τα έτη 1990 έως 2005 και θα προσπαθήσουμε να δούμε πόσο θα μεταβληθεί το ΑΕΠ αν μεταβληθεί κατά **1%** η προσφορά χρήματος.

Παράδειγμα

Πίνακας 4.6 ΑΕΠ και προσφορά χρήματος για τα έτη 1990 έως 2005

ΕΤΟΣ	ΑΕΠ (εκατ. $)	Προσφορά χρήματος
1990	1359,3	861,0
1991	1472,8	908,5
1992	1598,4	1023,2
1993	1782,8	1163,7
1994	1990,5	1286,7
1995	2249,7	1389,0
1996	2508,2	1500,2
1997	2723,0	1633,1
1998	3052,6	1795,5
1999	3166,0	1954,0
2000	3405,7	2385,2
2001	3772,2	2363,6
2002	4014,9	2562,6
2003	4240,3	2807,7
2004	4526,7	2901,0

ΠΗΓΗ : Damodar Gujarati (1999), « Essentials of Econometrics », 2nd Edition McGraw Hill.

Η εξίσωση που προκύπτει από τα παραπάνω δεδομένα είναι η ακόλουθη :

$$\mathbf{Y_t = -16329 + 2584 \ln X_t}$$

όπου $\mathbf{Y_t}$ = ΑΕΠ

$\mathbf{X_t}$ = προσφορά του χρήματος

Αν προχωρήσουμε στην οικονομική ανάλυση της παραπάνω συνάρτησης παρατηρούμε ότι αν μεταβληθεί η προσφορά του χρήματος κατά μία ποσοστιαία μονάδα θα αυξηθεί το ΑΕΠ κατά **25,84 %** εκατ **$**.

4.2.4 ΕΚΘΕΤΙΚΗ ΜΟΡΦΗ

Η συναρτησιακή μορφή της εκθετικής είναι η ακόλουθη :

$$\mathbf{Y = e^{\beta 0 + \beta 1X}} \qquad \mathbf{(4.13)}$$

Όμως μετατρέπεται εύκολα σε γραμμική αν προχωρήσουμε στον μετασχηματισμό της. Για να το επιτύχουμε το αυτό χρησιμοποιούμε το λογαριθμικό μετασχηματισμό και οδηγούμαστε στη εξής μορφή :

$$\mathbf{\ln Y = \ln \beta_0 + X \ln \beta_1}$$

Αν χρησιμοποιήσουμε την παραπάνω συνάρτηση βάζοντας στη θέση του **X** τον χρόνο **t** τότε μπορεί να μας βοηθήσει να ερμηνεύσουμε πολλά φαινόμενα που έχουν πρόοδο κατά τη διάρκεια του χρόνου. Σε αυτή την περίπτωση το $\mathbf{\beta_0}$ εκφράζει την εξαρτημένη μεταβλητή **Y** στην αρχή του χρόνου. Για την εκθετική μορφή δεν έχουμε να πούμε κάτι παραπάνω διότι συνήθως τη μετασχηματίζουμε σε λογαριθμική για την οποία μιλήσαμε ήδη.

4.2.5 ΛΟΓΙΣΤΙΚΗ ΜΟΡΦΗ

Η συναρτησιακή έκφραση της λογιστικής μορφής είναι η ακόλουθη :

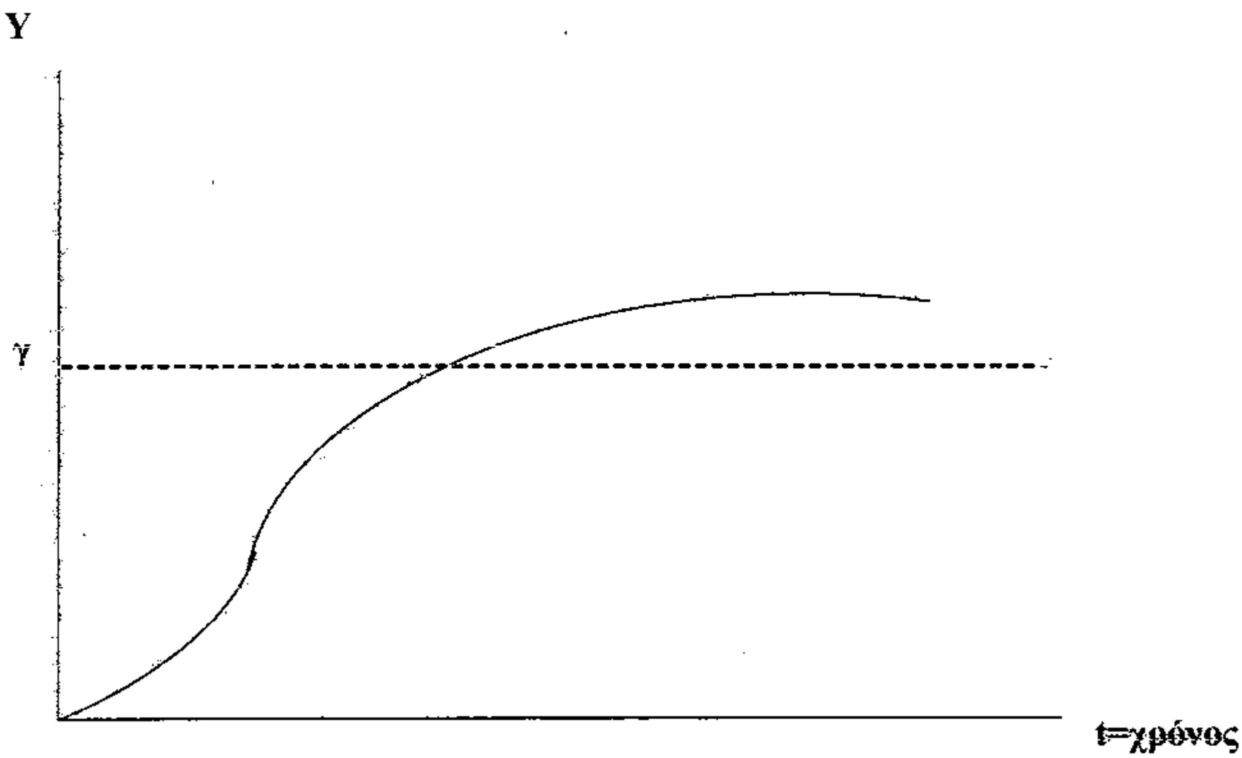
$$Y_t = \frac{\gamma}{1 + e^{\beta 0 + \beta 1 t}} \qquad (4.14)$$

Στην παραπάνω μορφή η παράμετρος γ είναι πάντα μεγαλύτερο του μηδενός και η παράμετρος β_1 πάντα μικρότερη του μηδενός. Η μορφή αυτή ονομάζεται και λογιστική καμπύλη και συνήθως χρησιμοποιείται για να εκφράσει την ανάπτυξη κάποιον φαινομένων ειδικά στην οικονομία.

Η συνάρτηση αυτή δεν είναι γραμμική ούτε ως προς τις παραμέτρους ούτε ως προς τις μεταβλητές οπότε αυτό έχει ως αποτέλεσμα να μη μπορούμε να χρησιμοποιήσουμε τις κλασικές μεθόδους που έχουμε δει για την εκτίμησή της. Η παράμετρος γ εκφράζει το ανώτατο όριο στο οποίο μπορεί να φτάσει η εξαρτημένη μεταβλητή **Y** και η παράμετρος β_1 μας δείχνει τον ρυθμό με τον οποίο η **Y** φτάνει σε αυτό το ανώτατο όριο. Στη λογιστική μορφή μπορούμε να χρησιμοποιήσουμε τη μέθοδο των ελαχίστων τετραγώνων που την έχουμε αναφέρει και σε άλλα κεφαλαία για να εκτιμήσουμε μόνο όμως τους συντελεστές β_0 και β_1.

Η γραφική απεικόνιση της λογιστικής μορφής είναι η παρακάτω:

Διάγραμμα 4.6



4.2.6 ΑΝΤΙΣΤΡΟΦΑ ΥΠΟΔΕΙΓΜΑΤΑ

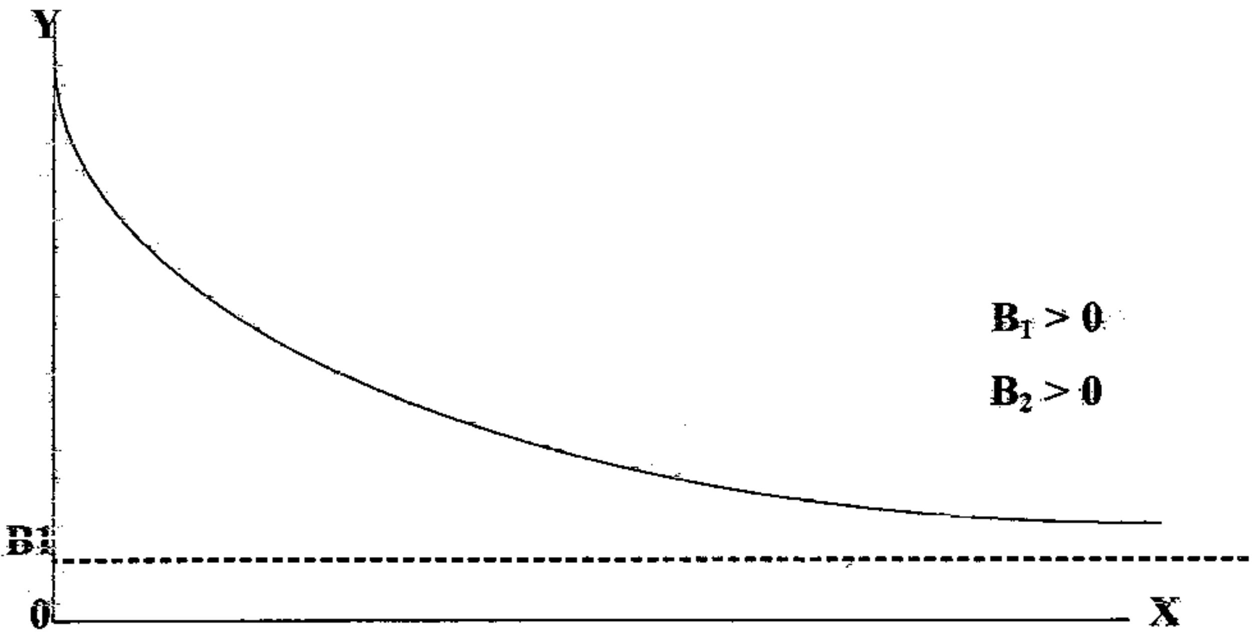
Μοντέλα με την ακόλουθη μορφή ονομάζονται αντίστροφα μοντέλα :

$$\mathbf{Y_i = B_1 + B_2\,(1/X_i) + u_i} \quad (4.15)$$

Αυτό το μοντέλο είναι μη γραμμικό ως προς **X** αλλά αποτελεί ένα γραμμικό μοντέλο παλινδρόμησης διότι οι παράμετροι $\mathbf{B_1}$ και $\mathbf{B_2}$ είναι γραμμικοί. Το αξιοπρόσεκτο χαρακτηριστικό αυτού του μοντέλου είναι ότι καθώς η ανεξάρτητη μεταβλητή **X** αυξάνει αόριστα, ο όρος $\mathbf{(1/X_i)}$ πλησιάζει το μηδέν και η εξαρτημένη μεταβλητή **Y** πλησιάζει την οριακή τιμή της $\mathbf{B_1}$.

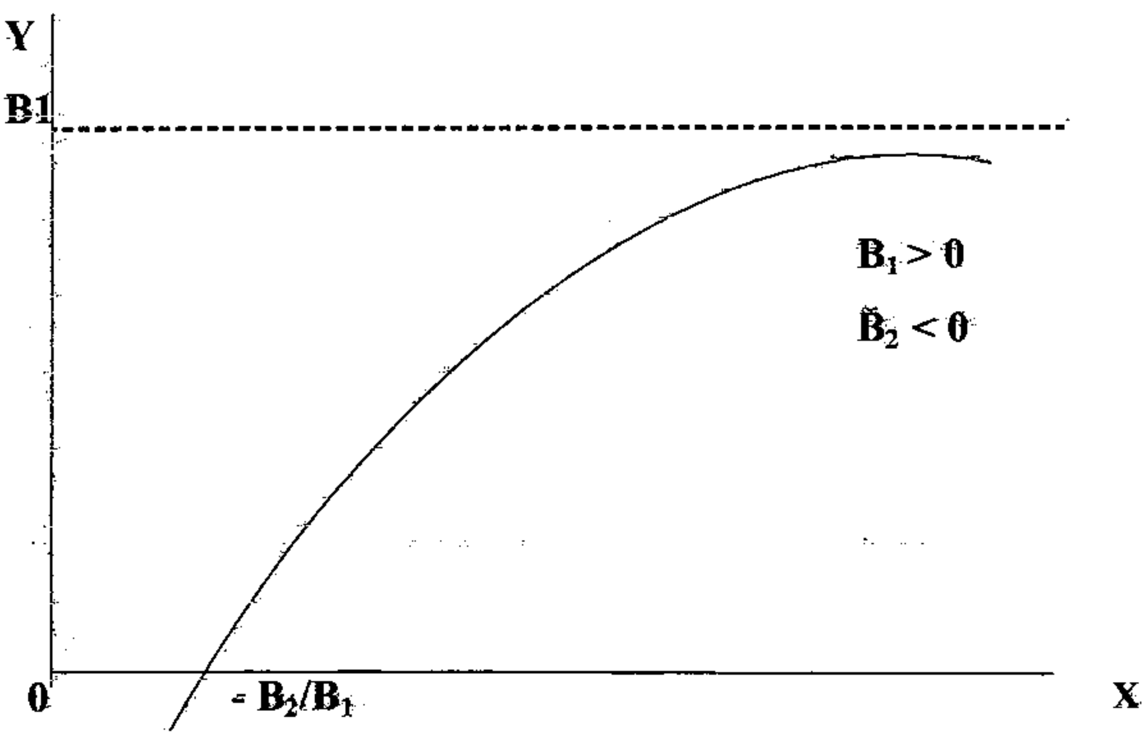
Υπάρχουν διάφορες μορφές διαγραμμάτων που αντιστοιχούν στο παραπάνω υπόδειγμα όπως :

Διάγραμμα 4.7



Γενική μορφή

Διάγραμμα 4.8

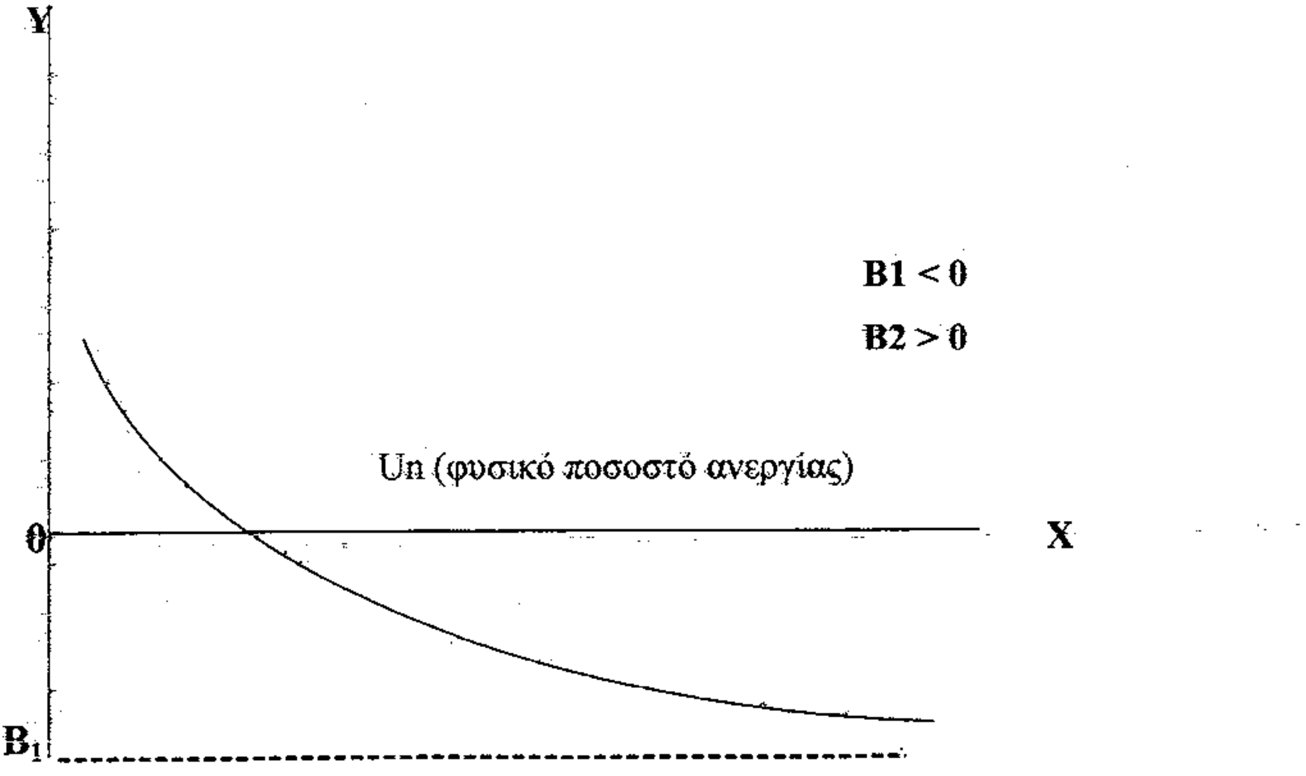


Καμπύλη του Engel

Η καμπύλη του Engel σχετίζεται με τις δαπάνες για την αγορά ενός αγαθού και με το διαθέσιμο εισόδημα. Υπάρχει ένα κρίσιμο σημείο στο διαθέσιμο εισόδημα κάτω από το οποίο το αγαθό δε μπορεί να αγοραστεί. Το σημείο αυτό είναι το - $\mathbf{B_2/B_1}$. Υπάρχει ένα σημείο κορεσμού της κατανάλωσης του αγαθού πέρα από το οποίο ο καταναλωτής δε θα το υπερβεί όσο και αν αυξηθεί το εισόδημά του. Αυτό το σημείο δεν είναι άλλο παρά η ασύμπτωτη της συνάρτησης $\mathbf{Y_i = B_1 + B_2\ (1/X_i) + u_i}$ όπως φαίνεται στο παραπάνω διάγραμμα.

Το συγκεκριμένο αντίστροφο μοντέλο είναι πολύ σημαντικό για την οικονομική θεωρία γιατί φανερώνει ένα ιδιαίτερο φαινόμενο που παρουσιάζεται στην αγορά από μέρους των καταναλωτών. Αν αναφερθούμε με πιο οικονομικούς όρους θα λέγαμε ότι η καμπύλη του **Engel** μας δείχνει την φθίνουσα οριακή χρησιμότητα που παρουσιάζει κάθε αγαθό πέρα από κάποιο κρίσιμο σημείο.

Διάγραμμα 4.9



Καμπύλη του Phillips

Η καμπύλη του **Phillips** αναφέρεται στην ποσοστιαία μεταβολή των ωρομίσθιων (αμοιβών) σε σχέση με την ποσοστιαία μεταβολή της ανεργίας. Αν παρατηρήσουμε το παραπάνω διάγραμμα βλέπουμε ότι όσο αυξάνεται η ανεργία ο μισθός μειώνεται. Αλλά για κάθε ποσοστιαία μονάδα που θα αυξηθεί η ανεργία, κάτω από το φυσικό όριο ανεργίας που είναι η ασύμπτωτη του συντελεστή $\mathbf{B_1}$, το ωρομίσθιο θα μειώνεται με μικρότερο ποσοστό.

Επειδή η παραπάνω μορφή είναι η πιο ενδιαφέρουσα από τις τρεις θα αναφέρουμε ένα παράδειγμα πάνω σε αυτή που θα εξετάζουμε τα παραπάνω μεγέθη για την περίοδο από το 1990 έως 2001 για τις ΗΠΑ.

Παράδειγμα

Πίνακας 4.6 Ποσοστό μεταβολής των ωρομισθίων και της ανεργίας για τα έτη 1990 έως 2001

Έτος	Υ(Ωρομίσθια)	Χ(Ανεργία)
1990	4,2	6,8
1991	3,5	5,5
1992	3,4	5,5
1993	3,0	6,7
1994	3,4	5,5
1995	2,8	5,7
1996	2,8	5,2
1997	3,6	4,5
1998	4,3	3,8
1999	5,0	3,8
2000	6,1	3,6
2001	6,7	3,5

ΠΗΓΗ : Damodar Gujarati (1999), « Essentials of Econometrics », 2nd Edition McGraw Hill.

Η εξίσωση που προκύπτει από τα παραπάνω δεδομένα είναι η ακόλουθη :

$$\mathbf{Y_t = -\,0{,}2594 + 20{,}5880\ (1/X_t)}$$

όπου $\mathbf{Y_t}$ = η ποσοστιαία μεταβολή των ωρομισθίων

$\mathbf{X_t}$ = η ποσοστιαία μεταβολή της ανεργίας

Αν προχωρήσουμε στην οικονομική ανάλυση της παραπάνω συνάρτησης παρατηρούμε ότι αν μεταβληθεί η ανεργία κατά μία ποσοστιαία μονάδα θα μειωθεί το ωρομίσθιο κατά **20,58 %** .

Το παραπάνω αντίστροφο μοντέλο μπορούμε να το μετατρέψουμε σε γραμμικό. Τότε θα προέκυπτε το εξής :

$$\mathbf{\hat{Y}_t = 8{,}0147 - 0{,}7883\ X_t}$$

Σε αυτή την περίπτωση ο συντελεστής $\mathbf{B_1}$ είναι αρνητικός και με αυτό τον τρόπο εκφράζεται η αντίστροφη ή αλλιώς αρνητική σχέση που υπάρχει ανάμεσα στα δύο μεγέθη.

4.2.7 ΠΟΛΥΩΝΥΜΙΚΑ ΜΟΝΤΕΛΑ

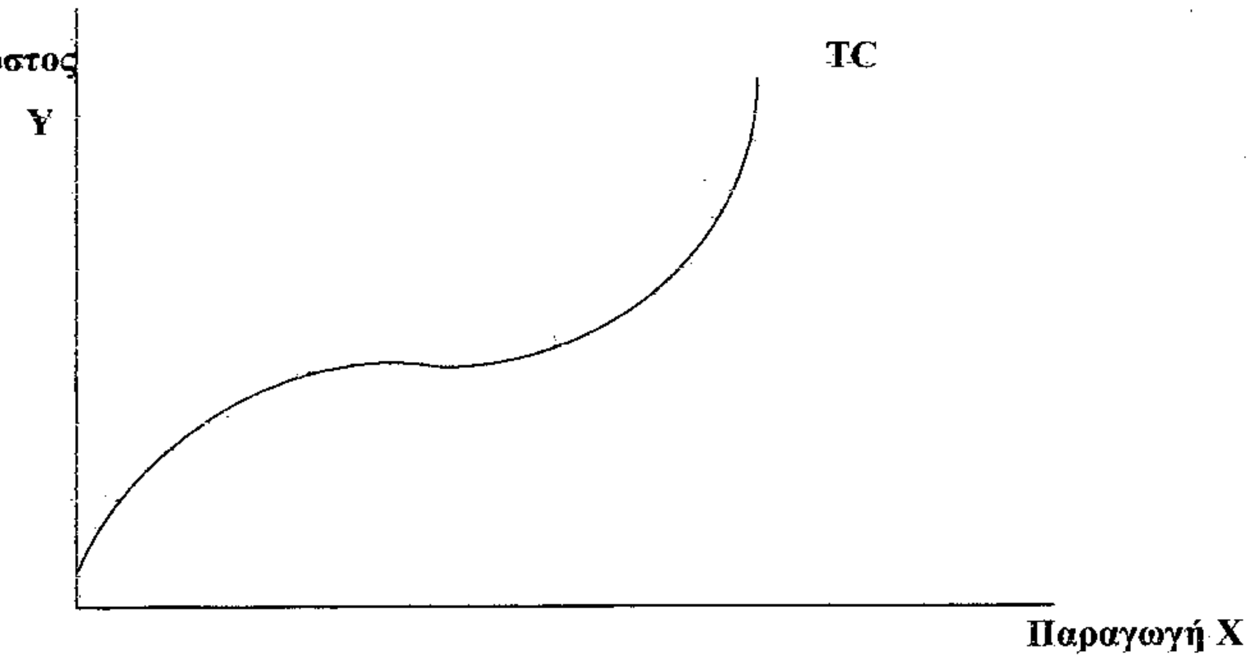
Τα πολυωνυμικά μοντέλα έχουν εκτεταμένη χρήση στην οικονομετρία και πιο συγκεκριμένα όταν αυτή σχετίζεται με την παραγωγή και τη συνάρτηση κόστους. Παρακάτω όπως θα δούμε στα διαγράμματα (4.8) και (4.9) απεικονίζεται το ολικό κόστος παραγωγής (TC) καθώς και το συσχετιζόμενο περιθώριο κόστους (MC) και το μέσο κόστος (AC).

Θεωρώντας ότι η **Y** αντιπροσωπεύει το TC και η **X** την παραγωγή, τότε η μαθηματική έκφραση της συνάρτησης του κόστους είναι η εξής :

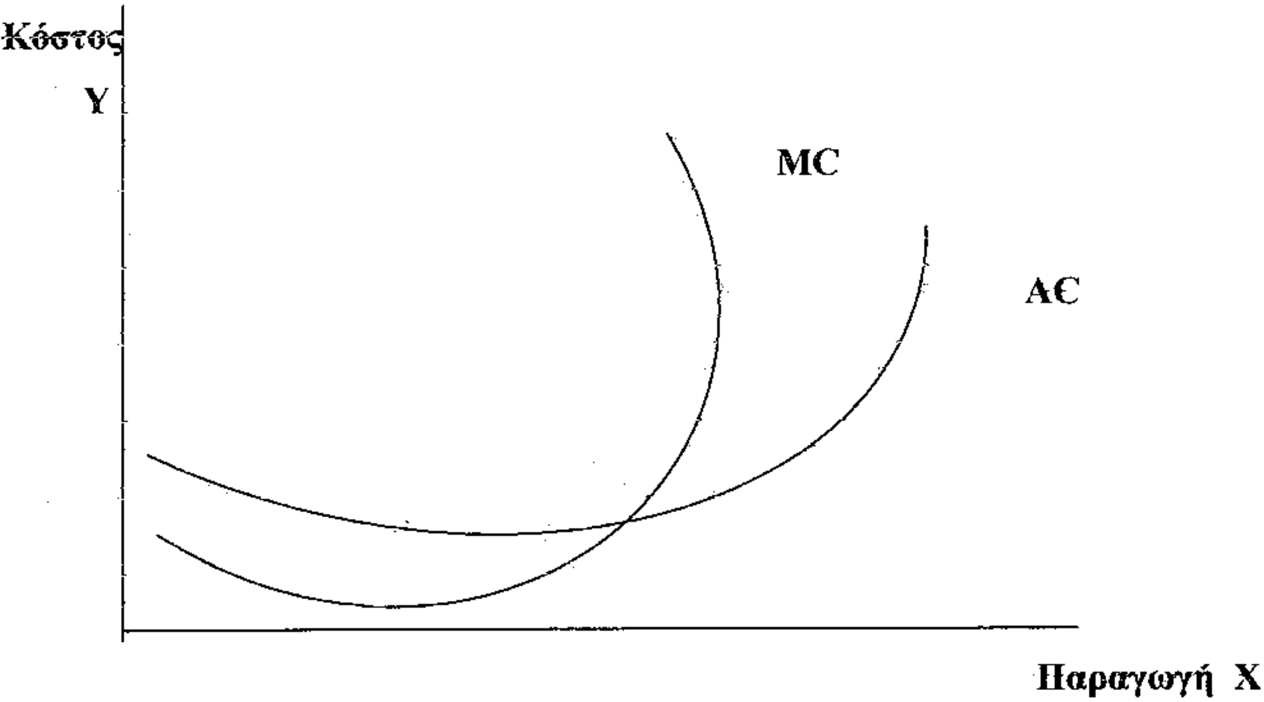
$$\mathbf{Y_i = B_1 + B_2X_i + B_3X_i^2 + B_4X_i^3 + u_i} \quad (4.16)$$

Η συνάρτηση αυτή ονομάζεται τρίτου βαθμού πολυωνυμική ως προς **X**. Η μεγαλύτερη δύναμη της **X** αντιπροσωπεύει το βαθμό της πολυωνυμικής συνάρτησης κάθε φορά. Σε αυτά τα μοντέλα υπάρχει μόνο μία εξαρτημένη (επεξηγηματική) μεταβλητή στο δεξί μέρος της συνάρτησης αλλά παρουσιάζεται με διάφορες δυνάμεις, κάνοντας τα μοντέλα αυτά πολυμεταβλητά.

Διάγραμμα 4.10



Διάγραμμα 4.11



Όταν τα μοντέλα αυτά έχουν τις μορφες που είδαμε στα παραπάνω διαγράμματα ισχύουν τα εξής για τις παραμέτρους :

1. **B_1 , B_2 και $B_4 > 0$**
2. **$B_3 < 0$**
3. **$B_3^2 < 3B_2B_4$**

Παρόλο που το μοντέλο (4.14) δεν είναι γραμμικό ως προς τη μεταβλητή **X**, είναι γραμμικό ως προς τις παραμέτρους **B** και είναι οπότε ένα γραμμικό παλινδρομικό μοντέλο. Τα μοντέλα αυτά μπορεί να παρουσιάσουν πολυσυγγραμικότητα αλλά αυτή η ανησυχία είναι περισσότερο φαινομενική. Γενικά όμως αυτού του είδους τα μοντέλα δεν παρουσιάζουν κάποια ιδιαίτερα προβλήματα για αυτό δε θα προχωρήσουμε σε κάποιο παράδειγμα.

4.3 ΑΝΑΛΥΣΗ ΣΥΣΧΕΤΙΣΗΣ ΚΑΤΑ ΤΑΞΕΙΣ

Οι τιμές των διαφόρων μεταβλητών μπορούν να μετρηθούν σε διάφορες κλίμακες. Έτσι οι μεταβλητές που είναι ποσοτικές εκφράζονται σε κλίμακα λόγου και αυτό διότι οι τιμές των μεταβλητών μπορούν να διαιρεθούν μεταξύ τους. Όπως η ποσότητα, το βάρος, το ύψος και άλλα. Οι μεταβλητές των οποίων οι παρατηρήσεις τοποθετούνται από τη μικρότερη στη μεγαλύτερη εκφράζονται σε κλίμακα ιεράρχησης επειδή οι τιμές των μεταβλητών μπορούν μόνο να ιεραρχηθούν και δε μπορούν να διαιρεθούν ή να αφαιρεθούν μεταξύ τους, παραδείγματος χάρη η βαθμολογία. Τέλος υπάρχει και η κλίμακα διαστήματος. Σε αυτή την κλίμακα οι τιμές των μεταβλητών μπορούν να συγκριθούν μόνο ως προς τη διαφορά τους. Συνήθως η θερμοκρασία εκφράζεται μέσω αυτής της κλίμακας.

Παρακάτω θα ασχοληθούμε με τη συσχέτιση μεταξύ δύο μεταβλητών που εκφράζονται σε κλίμακα ιεράρχησης ή μεταξύ ιεραρχημένης μεταβλητής και μίας ποσοτικής. Στις ιεραρχημένες μεταβλητές ο συντελεστής παλινδρόμησης δεν εκφράζει κάποιο χαρακτηριστικό.

Το Υπουργείο Εργασίας ζήτησε από μία ομάδα ειδικών να ερευνήσει αν υπάρχει σχέση μεταξύ του της αμοιβής και της απόδοσης του εργαζομένου στην εργασία του. Η έρευνα έγινε σε 12 εργαζομένους.

Παράδειγμα

Πίνακας 4.7 Αμοιβή και Απόδοση του Εργαζομένου

α/α	Ετήσιες Αμοιβές (χιλ. ευρώ)	Μέση Απόδοση (σε βαθμούς)
1	7,5	14,2
2	17,4	15,8
3	31,5	16,1
4	25,2	17,3
5	36,9	19,0
6	29,1	17,2
7	40,2	18,5
8	20,4	17,8
9	13,5	14,3
10	16,5	16,9
11	9,6	16,8
12	14,1	15,5

ΠΗΓΗ : Ιωάννης Γ. Χαλικιάς (2001), « Στατιστική – Μέθοδοι για Επιχειρηματικές Αποφάσεις », Αθήνα : Εκδόσεις Rosili.

Όπως είναι αναμενόμενο τα παραπάνω δεδομένα δε μπορούν να αναλυθούν με βάση κλασική παλινδρόμηση. Αν παραδείγματος χάρη προέκυπτε μία εξίσωση όπως **Y = 15 + 0,2X με Y = η *μέση απόδοση*, X= οι ετήσιες αμοιβές**. Ο συντελεστής $\mathbf{b_1=0,2}$ υπό άλλες συνθήκες θα μας έδειχνε ότι μία αύξηση των ετήσιων αμοιβών του εργαζομένου κατά χίλια ευρώ θα προκαλούσε μία αύξηση της μέσης απόδοσης κατά **0,2** βαθμούς.

Όμως στη δική μας περίπτωση αυτό που θέλουμε να δούμε είναι αν η τάξη μεγέθους των ετήσιων αμοιβών σχετίζεται με την τάξη μεγέθους της απόδοσης.

Άρα αυτό μου μας ενδιαφέρει είναι να ερευνήσουμε αν οι υψηλότερες αποδόσεις επιτυγχάνονται από τους εργαζομένους που έχουν υψηλότερες αμοιβές. Για αυτό το λόγο θα δημιουργήσουμε δύο άλλες μεταβλητές που θα βασίζονται όμως στις αρχικές και θα μας βοηθήσουν να κάνουμε την ιεράρχηση των τιμών των αρχικών μεταβλητών. Έτσι λοιπόν έχουμε τις νέες μεταβλητές R_X και R_Y και οι τιμές τους κυμαίνονται από το 1, η οποία αντιστοιχεί στην πιο μικρή τιμή της κάθε μεταβλητής και μετά αυξάνεται έως n. Παραδείγματος χάρη η μεταβλητή R_X έχει την τιμή 1 για την μικρότερη ετήσια αμοιβή (7,5 χιλ. ευρώ), 2 για την αμέσως μεγαλύτερη τιμή και φτάνει έως το 12 για την υψηλότερη τιμή των ετήσιων αμοιβών (40,2 χιλ. ευρώ). Αντίστοιχα το ίδιο συμβαίνει και για τη μεταβλητή $\mathbf{R_Y}$.

Ο συντελεστής συσχέτισης σε αυτή την περίπτωση έχει την εξής μορφή :

$$\mathbf{r_s} = \frac{\mathbf{1 - 6\sum_{i=1}^{n} d_i^2}}{\mathbf{n\,(n^2 - 1)}} \qquad (4.17)$$

όπου $\mathbf{r_s}$ = ο συντελεστής συσχέτισης κατά τάξεις του Spearman

$\mathbf{d_i = R_X - R_Y}$ (η διαφορά ανάμεσα στις τάξεις μεγέθους για κάθε ζεύγος παρατηρήσεων)

$\mathbf{n}$ = μέγεθος δείγματος

Πίνακας 4.8 Εκτίμηση Συντελεστή Συσχέτισης κατά Τάξεις μεταξύ Ετήσιων Αμοιβών και Απόδοση του Εργαζομένου

Ετήσιες Αμοιβές (χιλ. ευρώ)	Μέση Απόδοση (σε βαθμούς)	R_X	R_Y	d	d^2
7,5	14,2	1	1	0	0
17,4	15,8	6	4	2	4
31,5	16,1	10	5	5	25
25,2	17,3	8	9	-1	1
36,9	19,0	11	12	-1	1
29,1	17,2	9	8	1	1
40,2	18,5	12	11	1	1
20,4	17,8	7	10	-3	9
13,5	14,3	3	2	1	1
16,5	16,9	5	7	-2	4
9,6	16,8	2	6	-4	16
14,1	15,5	4	3	1	1
				Σ =	64

Με απλή αντικατάσταση του τύπου 4.15 προκύπτει ότι ο συντελεστής συσχέτισης $\mathbf{r_S = 0,776}$.

Η τιμή του σύντελεστής συσχέτισης μας δείχνει ότι υπάρχει έντονη θετική συσχέτιση μεταξύ της αμοιβής και της απόδοσης του εργαζομένου στην εργασία του. Αυτό σημαίνει με άλλα λόγια ότι τα άτομα που πληρώνονται περισσότερο αποδίδουν και καλύτερα στην εργασία τους.

Ο έλεγχος στατιστικής σημαντικότητας του r_s γίνεται με την κατανομή **t** ο τύπος είναι ο παρακάτω :

$$t_{n-2} = r_s\,(n-2)^{1/2} \,/\, (1-r_s^2)^{1/2} \quad (4.18)$$

Άρα κάνοντας απλή αντικατάσταση στον παραπάνω τύπο προκύπτει ότι :

$$t_{12-2} = 3{,}89$$

Το κρίσιμο σημείο της κατανομής **t** με επίπεδο σημαντικότητας **α=5%** είναι το εξής :

$$|t_{n-2,\alpha/2}| = |t_{10,\,0{,}025}| = 2{,}228$$

Παρατηρούμε $|t_{n-2}| > |t_{n-2,\alpha/2}|$ λοιπόν ότι οπότε οδηγούμαστε στο εξής συμπέρασμα, ότι η μηδενική υπόθεση απορρίπτεται διότι προκύπτει η παραπάνω ανίσωση που υποδηλώνει ότι αποδεχόμαστε την εναλλακτική υπόθεση δηλαδή ότι $r_s \neq 0$. Πράγμα που σημαίνει ότι υπάρχει σχέση ανάμεσα στην αμοιβή και τη απόδοση του εργαζομένου στην εργασία του.

Θα προχωρήσουμε όμως ακόμα σε ένα παράδειγμα για την καλύτερη κατανόηση της ιεράρχησης κατά τάξεις. Θα ασχοληθούμε με μία εταιρεία που κατασκευάζει μονωτικό υλικό και θα ερευνήσουμε τη σχέση που μπορεί να υπάρχει ανάμεσα στη διάρκεια ζωής του μονωτικού και η απόδοση του, δηλαδή κατά πόσο το μονωτικό υλικό συνεχίζει να έχει την ίδια μονωτική ιδιότητα με το πέρασμα του χρόνου καθώς αυτός το αλλοιώνει.

Παράδειγμα

Πίνακας 4.9 Διάρκεια Ζωής και Απόδοση του Μονωτικού Υλικου

α/α	Σύνθεση Μονωτικού Υλικού	Διάρκεια Ζωής (μέρες)	Απόδοση
1	A11	370	5
2	Δ23	550	2
3	Φ75	400	3
4	K86	250	7
5	K097	200	9
6	Π67	440	5
7	N57	650	1
8	O36	330	8
9	P32	400	6
10	O21	130	10

ΠΗΓΗ : Ιωάννης Γ. Χαλικιάς (2001), « Στατιστική – Μέθοδοι για Επιχειρηματικές Αποφάσεις », Αθήνα : Εκδόσεις Rosili.

Στη συνέχεια για να υπολογίσουμε τον συντελεστή συσχέτισης κατά τάξεις του **Spearman** θα κάνουμε τον παρακάτω πίνακα έτσι ώστε να διευκολυνθούμε. Ένα πολύ σημαντικό πλεονέκτημα αυτού του συντελεστή είναι ότι δεν είναι απαραίτητο οι μεταβλητές να ακολουθούν την κανονική κατανομή αλλά βασίζεται μόνο στις τάξεις μεγέθους. Δυστυχώς όμως τα συμπεράσματα που βγάζουμε από τη μελέτη αυτού του συντελεστή συσχέτισης σχετίζονται μόνο με τις τάξεις μεγέθους και δεν μπορούμε να εξάγουμε άλλα συμπεράσματα που να έχουν να κάνουν με τις αρχικές τιμές των μεταβλητών.

Πίνακας 4.8 Εκτίμηση Συντελεστή Συσχέτισης κατά Τάξεις μεταξύ Ετήσιων Διάρκειας Ζωής και Απόδοση του Μονωτικού Υλικού

α/α	Σύνθεση Μονωτικού Υλικού	Διάρκεια Ζωής (μέρες)	Απόδοση	R_X	R_Y	d	d^2
1	A11	370	5	5,0	4,5	0,5	0,25
2	Δ23	550	2	9,0	2,0	7,0	49,00
3	Φ75	400	3	6,5	3,0	3,5	12,25
4	K86	250	7	3,0	7,0	-4,0	16,00
5	K097	200	9	2,0	9,0	-7,0	49,00
6	Π67	440	5	8,0	4,5	3,5	12,25
7	N57	650	1	10,0	1,0	9,0	81,00
8	O36	330	8	4,0	8,0	-4,0	16,00
9	P32	400	6	6,5	6,0	0,5	0,25
10	Θ21	130	10	1,0	10,0	-9,0	81,00
						Σ =	317,00

Προχωρώντας σε απλή αντικατάσταση του τύπου 4.15 προκύπτει ότι ο συντελεστής συσχέτισης $\mathbf{r_s = -0{,}921}$.

Η τιμή του συντελεστής συσχέτισης μας δείχνει ότι υπάρχει έντονη αρνητική συσχέτιση μεταξύ της διάρκειας ζωής του μονωτικού υλικού και της απόδοσης του. Αυτό σημαίνει με άλλα λόγια ότι όσο περνάει ο καιρός που χρησιμοποιείται το μονωτικό υλικό τόσο μειώνεται και η απόδοση του.

Όσον αφορά και σε αυτό το παράδειγμα τη στατιστική σημαντικότητα έχουμε, κάνοντας απλή αντικατάσταση στον παραπάνω τύπο (4.16) προκύπτει ότι :

$$\mathbf{t_{10-2} = -6{,}687}$$

Το κρίσιμο με επίπεδο σημαντικότητας **α=5%** είναι το εξής :

$$|\mathbf{t_{n-2,\alpha/2}}| = |\mathbf{T_{8,\,0{,}025}}| = |-2{,}306|$$

Παρατηρούμε $|\mathbf{t_{n-2}}| > |\mathbf{t_{n-2,\alpha/2}}|$ λοιπόν ότι οπότε οδηγούμαστε στο εξής συμπέρασμα, ότι η μηδενική υπόθεση απορρίπτεται διότι προκύπτει η παραπάνω ανίσωση που υποδηλώνει ότι αποδεχόμαστε την εναλλακτική υπόθεση δηλαδή ότι $\mathbf{r_s \neq 0}$. Πράγμα που σημαίνει ότι υπάρχει σχέση ανάμεσα στην διάρκεια ζωής του μονωτικού υλικού και της απόδοσης του.

4.4 ΣΥΜΠΕΡΑΣΜΑ

Σε αυτό το κεφάλαιο ασχοληθήκαμε με διάφορα μοντέλα μερικά από τα οποία παρόλο που δεν ήταν γραμμικά ως προς τις μεταβλητές ήταν ως προς τις παραμέτρους **Β**. Βέβαια κάθε ερευνητής μπορεί να συνδυάσει τα χαρακτηριστικά ενός ή και περισσότερων μοντέλων.

Αν μας δοθεί ένα συγκεκριμένο δείγμα παρατηρήσεων για τις μεταβλητές **Y**, $\mathbf{X}_i$ μπορούμε πάντα να προσαρμόσουμε στις παρατηρήσεις του υποδείγματος με τη μέθοδο των ελαχίστων τετραγώνων $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ και να υπολογίσουμε τον συντελεστή πολλαπλού προσδιορισμού ακόμα και αν η πραγματική σχέση που συνδέει τις μεταβλητές είναι μη γραμμική ή δεν είναι αυτή ακριβώς η γραμμική σχέση που έχουμε αποφασίσει να εκτιμήσουμε. Με τον όρο γραμμικό υπόδειγμα αναφερόμαστε σε διάφορους τύπους συναρτήσεων που μπορεί να μη προέρχονται απαραίτητα από το γραμμικό μαθηματικό υπόδειγμα τις οικονομικής θεωρίας. Επίσης γνωρίζουμε ότι η οικονομική θεωρία προσδιορίζει ακριβείς σχέσεις μεταξύ των οικονομικών μεταβλητών άρα όσο προσεκτικά και αν συγκεντρώσουμε τα στοιχεία πάντα θα έχουμε κάποιες αποκλίσεις της προσαρμοσμένης ευθείας από την ακριβή σχέση. Τα κυριότερα σφάλματα εξειδίκευσης που μπορεί να γίνουν κατά τον προσδιορισμό του υποδείγματος είναι τα ακόλουθα :

1. Από ένα γραμμικό υπόδειγμα να έχουμε παραλείψει σημαντικές ερμηνευτικές μεταβλητές ή να έχουμε κάνει το εντελώς αντίθετο δηλαδή να έχουμε εισαγάγει μερικές άσχετες ερμηνευτικές μεταβλητές.
2. Το υπόδειγμα να είναι γραμμικό αλλά η σχέση να μεταβάλλεται από παρατήρηση σε παρατήρηση.
3. Το υπόδειγμα να μην είναι γραμμικό και εμείς να το εκτιμούμε ως γραμμικό. Σε αυτή τη περίπτωση πρέπει να ξεχωρίσουμε τα υποδείγματα που μέσω μετασχηματισμών μετατρέπονται σε γραμμικά από τα μη γραμμικά που πρέπει να εκτιμηθούν με άλλες μεθόδους αριστοποίησης.

Ο συντελεστής προσδιορισμού $\mathbf{R}^2$ για τον οποίο έχουμε μιλήσει σε προηγούμενα κεφάλαια μπορεί να μας βοηθήσει στην επιλογή υποδειγμάτων διαφορετικών αλγεβρικών μορφών. Το $\mathbf{R}^2$ μπορεί να χρησιμοποιηθεί ως κριτήριο επιλογής μόνο μεταξύ υποδειγμάτων που έχουν τον ίδιο αριθμό διαφορετικών ανεξάρτητων μεταβλητών. Σε δύο ή και περισσότερα υποδείγματα με διαφορετική αλγεβρική μορφή και τον ίδιο αριθμό μεταβλητών οι τιμές του $\mathbf{R}^2$ είναι συγκρίσιμες μόνο εφόσον η εξαρτημένη μεταβλητή εκφράζεται στις ίδιες μονάδες σε όλα τα υποδείγματα.

Για παράδειγμα στα παρακάτω υποδείγματα :

$$Y_i = B_1 + B_2 X_{2i} + B_3 X_{3i} + u_i \quad (R_i^2)$$

$$\ln Y_i = B_1 + B_2 \ln X_{2i} + B_3 \ln X_{3i} + u_i \quad (R_2^2)$$

οι τιμές των R_i^2 και R_2^2 δεν είναι συγκρίσιμες γιατί το R_i^2 μετράει το ποσοστό της διακύμανσης στο Y που ερμηνεύεται από τις μεταβλητές X_2, X_3 ενώ το R_2^2 μετράει το ποσοστό της διακύμανσης στο lnY που ερμηνεύεται από τις ίδιες μεταβλητές.

Συνήθως το R^2 όταν έχει υψηλή τιμή σημαίνει ότι το υπόδειγμα που επιλέχθηκε και εκτιμήθηκε, ερμηνεύει ικανοποιητικά το φαινόμενο που μελετάται. Όμως πολλές φορές συμβαίνει υποδείγματα με υψηλές τιμές για το R^2 να έχουν μη σημαντικούς συντελεστές σε βασικές ερμηνευτικές μεταβλητές. Επίσης είναι πιθανό να εμφανιστούν μεταβλητές με πρόσημα αντίθετα από αυτά που αναμένονται με βάση την οικονομική θεωρία. Τέλος κατά την ερμηνεία του R^2 δεν πρέπει να μας διαφεύγει το γεγονός ότι υψηλή συσχέτιση δε σημαίνει οπωσδήποτε και αιτιώδη εξάρτηση. Αυτό θα εξαρτηθεί από το αν και κατά πόσο σύμφωνα με τη θεωρία υπάρχει αιτιώδης σχέση μεταξύ τις εξαρτημένης και των ανεξάρτητων μεταβλητών.

Γενικά η μοντελοποίηση απαιτεί την κατάλληλη ισορροπία όσον αφορά την θεωρία, διαθεσιμότητα κατάλληλων δεδομένων, μία καλή ερμηνεία των στατιστικών προδιαγραφών των διαφόρων μοντέλων και η πρακτική κρίση. Δυστυχώς σύμφωνα με τη θεωρία δεν υπάρχει ποτέ τέλειο μοντέλο. Αυτό που προσπαθούμε πάντα να πετύχουμε είναι να βρούμε ένα μοντέλο το οποίο να είναι σε γενικές γραμμές καλό και να ικανοποιεί όλα απαραίτητα κριτήρια.

ΚΕΦΑΛΑΙΟ 5ο

ΓΕΝΙΚΗ ΕΦΑΡΜΟΓΗ

5.1 ΕΙΣΑΓΩΓΗ

Λίγα λόγια για την Οικονομετρία: Η οικονομετρία είναι η επιστήμη της οικονομικής μέτρησης που επιδιώκει να θεμελιώσει ποσοτικές σχέσεις ανάμεσα σε οικονομικές μεταβλητές με τη βοήθεια των οικονομικών, των μαθηματικών και της στατιστικής θεωρίας, για να διεξάγει – δημιουργήσει πλούτο. Τα οικονομετρικά υποδείγματα είναι συστήματα ταυτόχρονα προσδιοριζόμενων εξισώσεων που συνδέουν τις οικονομικές μεταβλητές, με σκοπό την πρόβλεψη και την εξήγηση της οικονομικής συμπεριφοράς. Τα δεδομένα που χρησιμοποιούνται στα διάφορα οικονομετρικά υποδείγματα είναι πάντα εργαστηριακά και όχι πειραματικά. Αυτή η επιστήμη αποτελεί το βασικό εργαλείο για την μελέτη και την πρόβλεψη ενός φαινομένου.

Αυτό συμβαίνει διότι ένα οικονομετρικό υπόδειγμα παρέχει σε αυτόν που κάνει προβλέψεις, ρητά εκφρασμένο σύστημα, στο οποίο κάποιος μπορεί να συγκεντρώσει και να σταθμίσει οικονομικές πληροφορίες με ένα συστηματικό και αξιόπιστο τρόπο. Οι οικονομετρικές προβλέψεις εφοδιάζουν τον προβλέποντα με μια λεπτομερή εικόνα των επιπτώσεων των προβλέψεών του. Επίσης, τα μοντέλα αυτά επιτρέπουν την αξιολόγηση και τη βελτίωση να ελεγχθούν υποθέσεις οι οποίες αφορούν τη θεωρία της επιχειρηματικής συμπεριφοράς.

Τα υποδείγματα όμως κάποιες φορές υπόκεινται σε περιορισμούς. Στους περιορισμού των οικονομετρικών υποδειγμάτων περιλαμβάνονται η ετροσκεδαστικότητα (Heteroscedasticity) δηλαδή η ύπαρξη κυμαινόμενης διακύμανσης στα τυπικά σφάλματα U_i, η πολυσυγγραμικότητα (Multicollinearity)των μεταβλητών δηλαδή η τέλεια γραμμική συσχέτιση ανάμεσα σε δύο ή περισσότερες μεταβλητές, και τέλος η αυτοσυσχέτιση (Autocorrelation) δηλαδή η ύπαρξη συσχέτισης μεταξύ των παρατηρήσεων (το σφάλμα που σχετίζεται με την ταυτόχρονη εκτίμηση των εξισώσεων του μοντέλου). Επίσης, οικονομετρικά υποδείγματα μεγάλης κλίμακας, που είναι έντονα αθροιστικά, δεν ενσωματώνουν διαθέσιμες λεπτομερείς πληροφορίες που αφορούν εξελίξεις σε ξεχωριστούς τομείς της οικονομίας. Τέλος, τα μοντέλα αυτά χρειάζονται συνεχή παρακολούθηση σε σχέση με την καταλληλότητά τους να διαγνώσουν τις αναγκαίες περιοδικές μεταβολές.

Συχνά, οι επιχειρήσεις υιοθετούν οικονομετρικά υποδείγματα, συνδέοντας μια πρόβλεψη για κάποιο οικονομικό μέγεθος που τους ενδιαφέρει με έναν ή περισσότερους οικονομικούς δείκτες που έχουν δημιουργηθεί από κατάλληλο οικονομετρικό μοντέλο. Η σύνδεση αυτή, επιτρέπει στις εταιρίες να εφαρμόζουν υποδείγματα μεγάλης κλίμακας χωρίς να χρειάζεται κάποιο ειδικά για τον εαυτό τους.

Στη συγκεκριμένη περίπτωση θα ασχοληθούμε με τη μελέτη των ανεξάρτητων (ερμηνευτικών) μεταβλητών που επηρεάζουν και κατ'επέκταση ερμηνεύουν καλύτερα την εξαρτημένη μας μεταβλητή που είναι η ανταγωνιστικότητα της επιχείρησής μας. Για να προβούμε στην εύρεση του κατάλληλου οικονομετρικού υποδείγματος θα χρησιμοποιήσουμε όπως προαναφέραμε τα εργαλεία της στατιστικής. Πιο συγκεκριμένα οι ανεξάρτητες μεταβλητές που θα χρησιμοποιήσουμε στην παρακάτω εφαρμογή μας με βάση τη σειρά που εμφανίζονται στο παράρτημα είναι οι ακόλουθες : χαμηλότερες τιμές ανταγωνιστικών προϊόντων, περισσότερο ελεύθερες αγορές, καλύτερη ποιότητα ανταγωνιστικών προϊόντων, χαμηλότερη ποιότητα ανθρώπινου δυναμικού μας, εντονότερη χρήση τεχνολογιών πληροφορικής, καλύτερη προστασία περιβάλλοντος, μεγαλύτερη διαπραγματευτική δύναμη προμηθευτών, μεγαλύτερη διαπραγματευτική δύναμη πελατών, υψηλότερο κόστος εργασίας, υψηλότερο κόστος πρώτων υλών, διαθεσιμότητα κατάλληλου εργασιακού δυναμικού, μεγαλύτερη καινοτομικότητα, καλύτερη φήμη μας, ύπαρξη κοντινών υποκατάστατων προϊόντων, χαμηλότερη τελική τιμή μας, υψηλότερο κόστος μεταφοράς προϊόντων, μεγαλύτερα γραφειοκρατικά εμπόδια, καλύτερα κανάλια διανομής, μεγαλύτερες δυσκολίες πρόσβασης σε χρηματοδότηση, υψηλότερη φορολόγηση, εντονότερη διαφήμισή μας και τέλος η ανεξάρτητη μεταβλητή μεγαλύτερο μέγεθος τοπικής αγοράς.

Επειδή η ανταγωνιστικότητα είναι ένα οικονομικό μέγεθος που εκτός από τους παράγοντες που παραθέτονται μπορεί να επηρεάζεται και από άλλους παράγοντες, εμείς αυτό που θα προσπαθήσουμε να καταφέρουμε είναι να βρούμε ένα ικανοποιητικό υπόδειγμα, αν όχι το βέλτιστο. Τα δεδομένα τα οποία θα χρησιμοποιήσουμε είναι διαστρωματικά και προέρχονται από 51 επιχειρήσεις. Στο παράρτημα υπάρχει ο πίνακας με τα δεδομένα μας αναλυτικά. Τα δεδομένα της εφαρμογής μας προέρχονται από το internet και πιο συγκεκριμένα από την τράπεζα δεδομένων που διαθέτει το ίδιο το πρόγραμμα στατιστικής SPSS.

5.2 ΑΝΤΑΓΩΝΙΣΤΙΚΟΤΗΤΑ

Ένα από τα ζητήματα (θέματα) πάνω στα οποία οι επιχειρήσεις προσπαθούν να αναπτύξουν – εφαρμόσουν ένα οικονομετρικό υπόδειγμα είναι η « Ανταγωνιστικότητα ».

Ο Ανταγωνισμός είναι ένας από τους πιο σημαντικούς παράγοντες που προσδιορίζουν τις οικονομικές συνθήκες των επιχειρήσεων όπως αναφέρει και στο βιβλίο του « Επιχειρησιακή Στρατηγική » ο Π. Ευθύμογλου. Με άλλα λόγια, ο ανταγωνισμός που υπάρχει σε μια αγορά αποτελεί τη βάση για την αξιολόγηση της καταλληλότητας συγκεκριμένων επιχειρηματικών δραστηριοτήτων από την άποψη της συμβολής τους στην επίτευξη αποδοτικής λειτουργίας και αποτελεσματικής ανάπτυξης των επιχειρήσεων στην αγορά αυτή.

Ο σκοπός της στρατηγικής ανταγωνισμού που εφαρμόζει μια επιχείρηση είναι η επίτευξη μιας πλεονεκτικής θέσης στην αγορά.(Π.Ευθύμογλου,1990, 21)

Η υλοποίηση των στρατηγικών αποφάσεων της επιχείρησης επισημαίνει πως μπορεί να επηρεάσει τόσο το βαθμό ελκυστικότητας της αγοράς όσο και την ανταγωνιστική της θέση. Για αυτό το λόγο η επιλογή της στρατηγικής ανταγωνισμού αποτελεί κρίσιμη απόφαση της επιχείρησης.

Η στρατηγική ανταγωνισμού που ακολουθεί η επιχείρηση, πρέπει να προκύπτει από μια πλήρη γνώση των συνθηκών ανταγωνισμού. Ο τελικός σκοπός της ανταγωνιστικής στρατηγικής είναι από τη μια πλευρά η προσαρμογή των δραστηριοτήτων της επιχείρησης προς τις συνθήκες ανταγωνισμού της αγοράς και από την άλλη πλευρά, η μεταβολή των συνθηκών αυτών προς όφελος της επιχείρησης.

5.3 ΜΕΘΟΔΟΛΟΓΙΑ ΤΗΣ ΟΙΚΟΝΟΜΕΤΡΙΚΗΣ ΜΕΛΕΤΗΣ

5.3.1 ΔΗΜΙΟΥΡΓΙΑ ΤΗΣ ΘΕΩΡΙΑΣ Η ΥΠΟΘΕΣΗΣ

a) Σύμφωνα με την οικονομική θεωρία ένας ουσιαστικός παράγοντας που επηρεάζει την ανταγωνιστικότητα είναι η καινοτομικότητα, όσο μεγαλύτερη είναι η ικανότητά μας να είμαστε καινοτόμοι τόσο μεγαλύτερη θα είναι και η ανταγωνιστικότητά μας διότι σε σχέση με τους ανταγωνιστές μας, εμείς θα είμαστε πρωτοπόροι.

b) Επιπλέον είναι αποδεδειγμένο ότι η ύπαρξη γραφειοκρατικών εμποδίων έχει επιπτώσεις στην ανταγωνιστικότητα των επιχειρήσεων ενός κλάδου.

Έτσι λοιπόν όσο πιο μεγάλα είναι τα γραφειοκρατικά εμπόδια για τις επιχειρήσεις τόσο μειώνεται η ανταγωνιστικότητα τους διότι τις παρεμποδίζουν να προχωρήσουν γρήγορα και εύκολα στις ενέργειες που επιθυμούν και έτσι όποιος προγραμματισμός γίνεται από μέρους τους καθυστερεί και δημιουργούνται προβλήματα.

c) Ένας ανασταλτικός παράγοντας για όλες τις επιχειρήσεις αναμφίβολα είναι οι δυσκολίες πρόσβασης σε χρηματοδότηση, όσο πιο δύσκολο είναι για κάποια επιχείρηση να βρει πηγές χρηματοδότησης τόσο μειώνεται η ανταγωνιστικότητά της διότι χωρίς την απαιτούμενη ρευστότητα η επιχείρηση αντιμετωπίζει πάρα πολλά προβλήματα πράγμα που προφανώς την καθιστά λιγότερο ανταγωνιστική.

d) Τέλος η φορολόγηση μπορεί να δημιουργήσει πολλά προβλήματα στις επιχειρήσεις γιατί όσο υψηλότερη είναι η φορολόγηση για τις επιχειρήσεις που ανήκουν προφανώς στον ίδιο κλάδο τόσο η ανταγωνιστικότητα της κάθε επιχείρησης θα μειώνεται, άρα θεωρούμε ότι η ερμηνευτική μεταβλητή «υψηλότερη φορολόγηση» επηρεάζει σημαντικά την ανταγωνιστικότητά μας

5.3.2 ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ

Τα δεδομένα που θα χρησιμοποιήσουμε στην παρακάτω εφαρμογή είναι διαστρωματικά, δηλαδή συγκεντρώθηκαν σε μία δεδομένη χρονική στιγμή και αφορούν πολλές μεταβλητές. Η κλίμακα μέτρησης των δεδομένων που χρησιμοποιήθηκε είναι τύπου **Livert** (είναι μονός ο αριθμός των παρατηρήσεων για να μπορούμε να βρούμε εύκολα το 50%) και ο αριθμός του δείγματος είναι 51 επιχειρήσεις. (βλέπε παράρτημα πίνακας με δεδομένα)

5.3.3 ΠΡΟΣΔΙΟΡΙΣΜΟΣ ΤΗΣ ΜΑΘΗΜΑΤΙΚΗΣ ΜΟΡΦΗΣ ΤΗΣ ΘΕΩΡΙΑΣ

Η μαθηματική μορφή η οποία προκύπτει από τις οικονομικές θεωρίες που διατυπώσαμε είναι η εξής :

$$\mathbf{Y=B_1+B_2X_1+B_3X_2+B_4X_3+B_5X_4} \quad (5.1)$$

Με **Y**= η ανταγωνιστικότητά μας

X_1= η μεγαλύτερη καινοτομικότητα

X_2= τα μεγαλύτερα γραφειοκρατικά εμπόδια

X_3= οι μεγαλύτερες δυσκολίες πρόσβασης σε χρηματοδότηση

X_4= η υψηλότερη φορολόγηση

5.3.4 ΠΡΟΣΔΙΟΡΙΣΜΟΣ ΤΗΣ ΟΙΚΟΝΟΜΕΤΡΙΚΗΣ ΜΟΡΦΗΣ ΤΗΣ ΘΕΩΡΙΑΣ

Η οικονομετρική μορφή αποτελείται από δύο μέρη, από το συστηματικό ή ντετερμινιστικό μέρος και από το στοχαστικό μέρος το οποίο απεικονίζεται με το $\mathbf{u_i}$ που ονομάζεται διαταρακτικός όρος και αντιπροσωπεύει την επίδραση κάποιων μεταβλητών που δεν χρησιμοποιήθηκαν στο υπόδειγμα .

$$\mathbf{Y=B_1+B_2X_1+B_3X_2+B_4X_3+B_5X_4+u_i} \quad (5.2)$$

Με **Y**= η ανταγωνιστικότητά μας

X_1= η μεγαλύτερη καινοτομικότητα

X_2= τα μεγαλύτερα γραφειοκρατικά εμπόδια

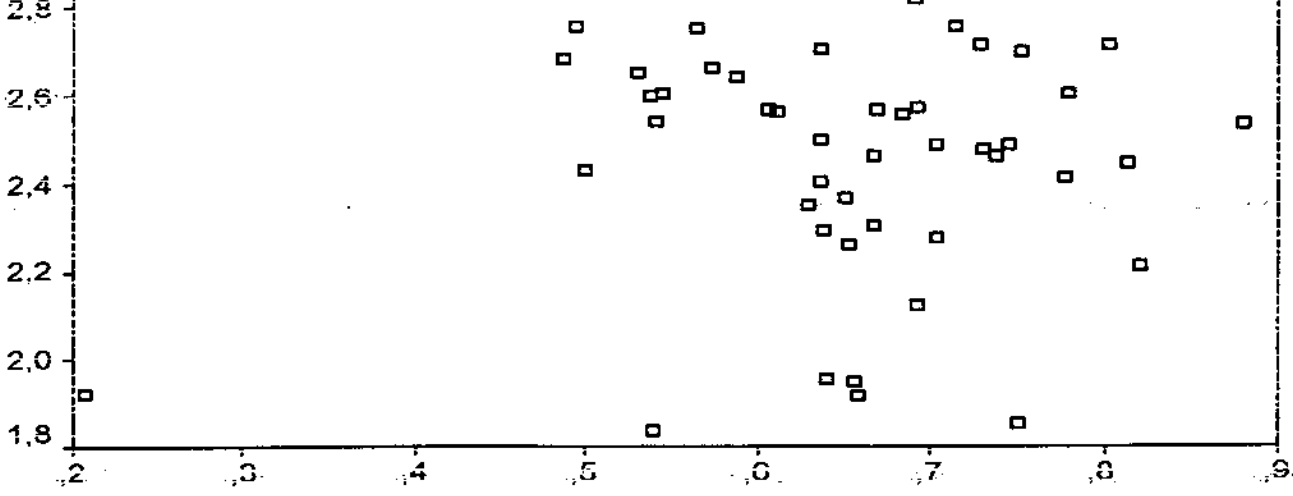
X_3= οι μεγαλύτερες δυσκολίες πρόσβασης σε χρηματοδότηση

X_4= η υψηλότερη φορολόγηση

5.3.5 ΕΚΤΙΜΗΣΗ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΚΑΙ ΕΛΕΓΧΟΣ ΤΟΥ ΥΠΟΔΕΙΓΜΑΤΟΣ

Στη συνέχεια θα προχωρήσουμε στην εκτίμηση των παραμέτρων του υποδείγματος με τις μεθόδους που έχουμε αναλύσει στη θεωρία καθώς και θα κάνουμε και έλεγχο στο υπόδειγμά μας για την εμφάνιση πολυσυγγραμμικότητας, ετεροσκεδαστικότητας και αυτοσυσχέτισης.

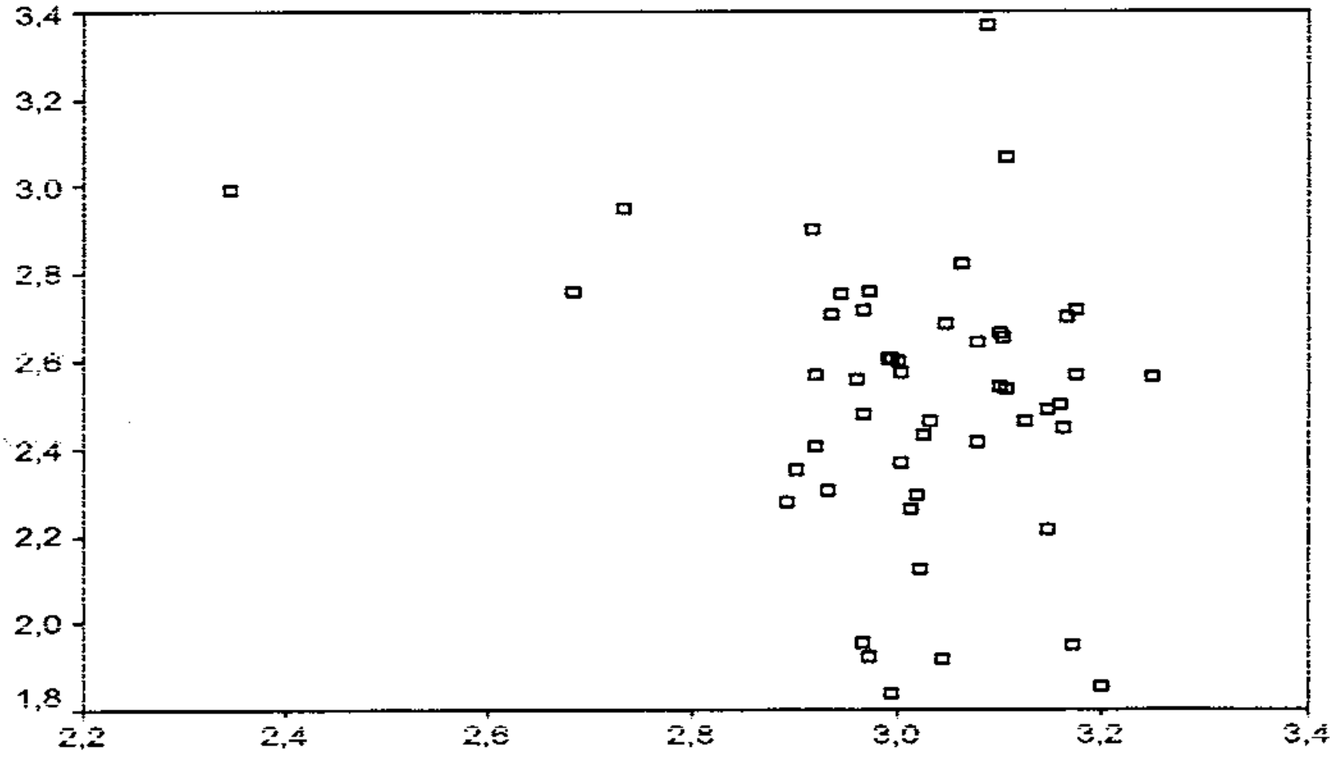
5.3.5.1 ΔΙΑΓΡΑΜΜΑ ΔΙΑΣΠΟΡΑΣ ΑΝΤΑΓΩΝΙΣΤΙΚΟΤΗΤΑΣ ΚΑΙ ΜΕΓΑΛΥΤΕΡΗΣ ΚΑΙΝΟΤΟΜΙΚΟΤΗΤΑΣ



Μεγαλύτερη καινοτομικότητα

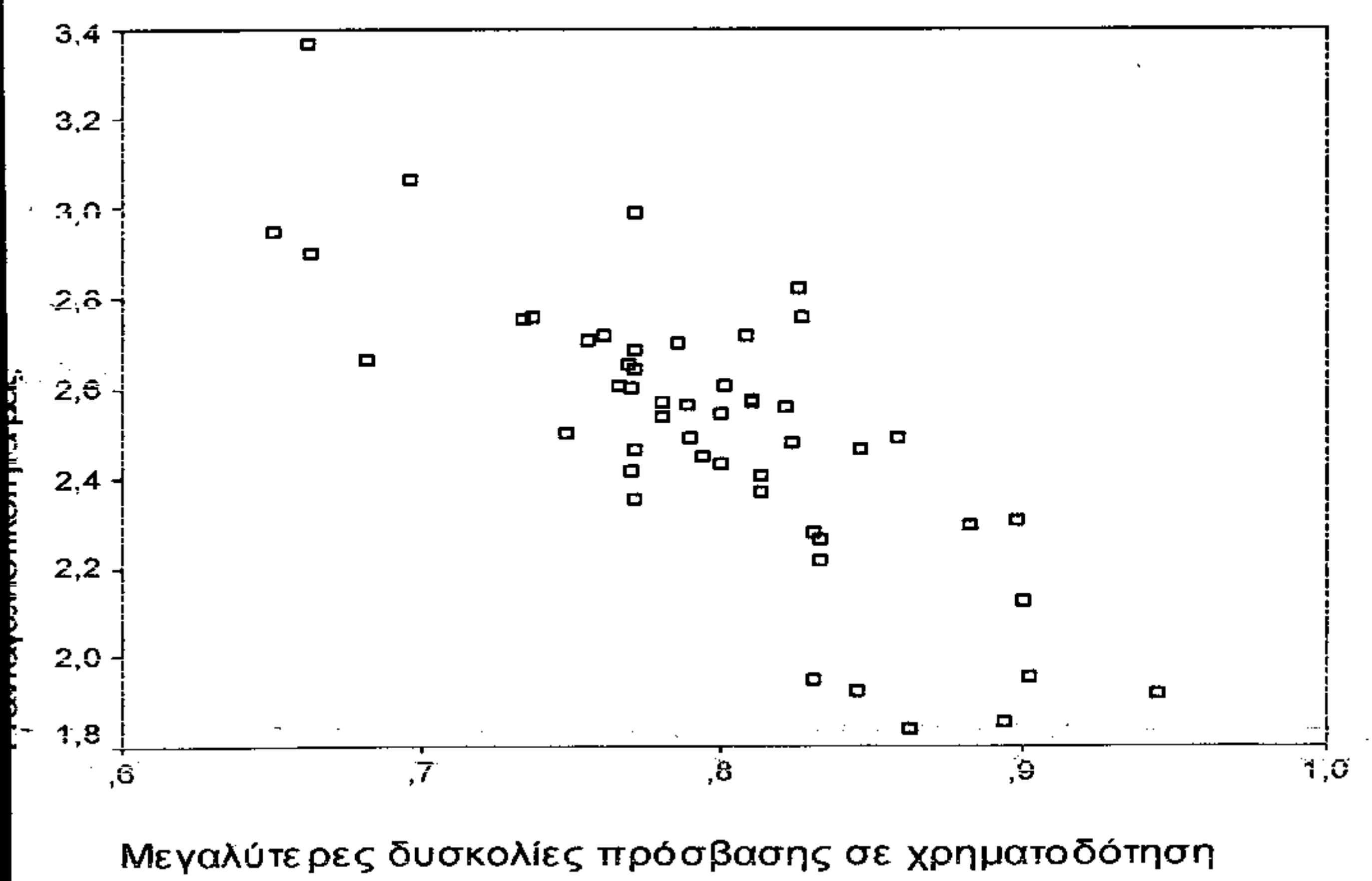
Παρατηρούμε στο παραπάνω διάγραμμα διασποράς ότι η σχέση ανάμεσα στα δύο μεγέθη (ανταγωνιστικότητα – καινοτομικότητα) δεν μπορεί να προσδιοριστεί εύκολα αλλά θα μπορούσαμε να διακρίνουμε με επιφύλαξη ότι υπάρχει ανάμεσά τους μία **ασθενής θετική** γραμμική συσχέτιση. Δηλαδή ότι η αύξηση της καινοτομικότητας μπορεί να επιφέρει και αύξηση της ανταγωνιστικότητάς μας.

5.3.5.2 ΔΙΑΓΡΑΜΜΑ ΔΙΑΣΠΟΡΑΣ ΑΝΤΑΓΩΝΙΣΤΙΚΟΤΗΤΑΣ ΚΑΙ ΜΕΓΑΛΥΤΕΡΩΝ ΓΡΑΦΕΙΟΚΡΑΤΙΚΩΝ ΕΜΠΟΔΙΩΝ



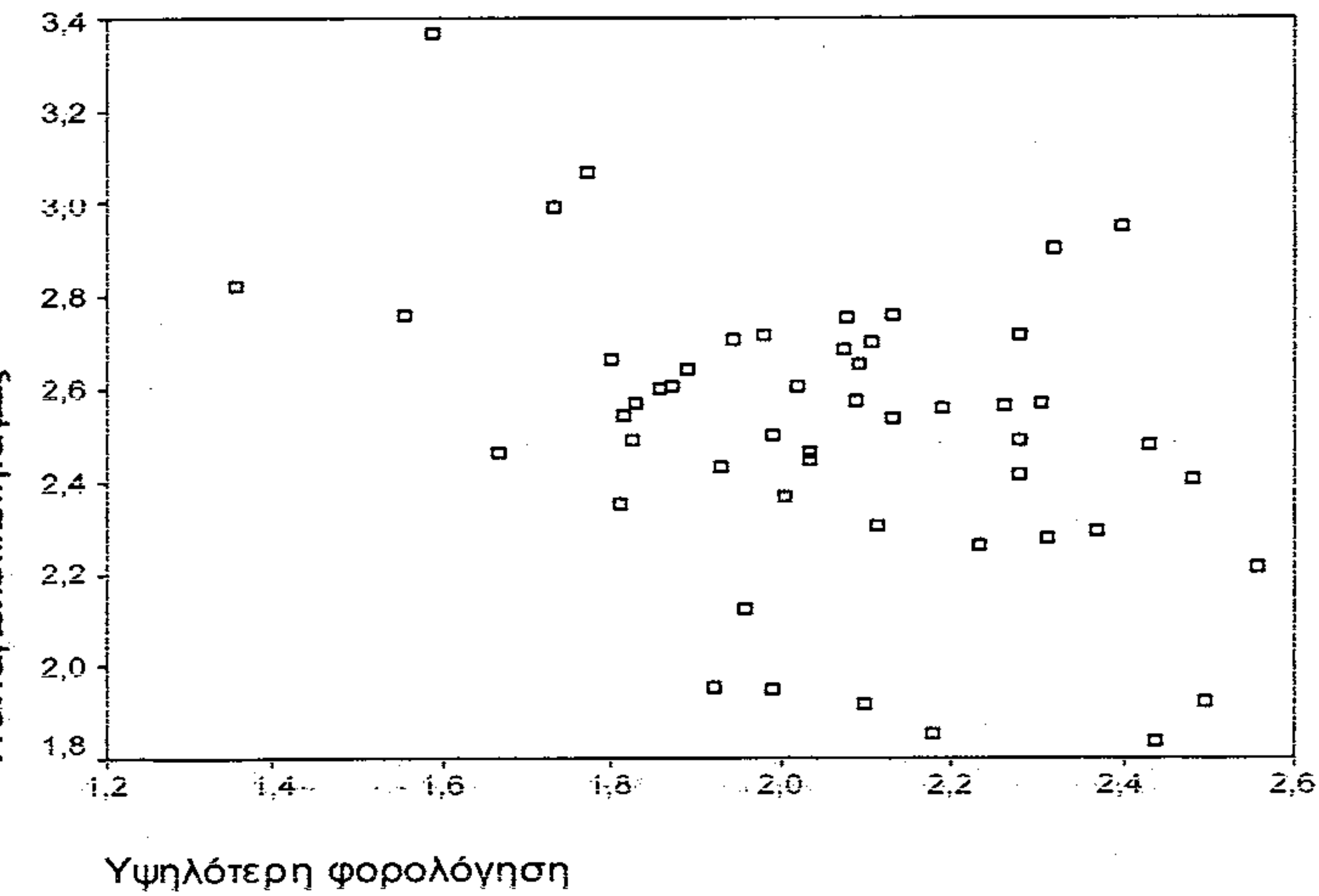
Βλέπουμε ότι στο παραπάνω διάγραμμα διασποράς (ανταγωνιστικότητα - γραφειοκρατικά εμπόδια) δεν φαίνεται καθαρά η σχέση που υπάρχει μεταξύ των δύο συγκεκριμένων μεγεθών αλλά αυτό μπορεί να αποδοθεί στο ότι οι παρατηρήσεις προέρχονται από ένα μικρό δείγμα (51 παρατηρήσεις) και η ανομοιομορφία μπορεί να οφείλεται στο ότι οι επιχειρήσεις μάλλον διαφέρουν μεταξύ τους, υποθέτουμε ως προς το μέγεθος. Όμως η μεταβλητή 'γραφειοκρατικά εμπόδια', σύμφωνα με την οικονομική μας θεωρία επηρεάζει σημαντικά στην ανταγωνιστικότητα άρα κρίνουμε ότι είναι απαραίτητο να χρησιμοποιηθεί στο υπόδειγμα που καταλήξαμε και δεν κρίνεται απαραίτητο να προχωρήσουμε σε αλλαγές.

5.3.5.3 ΔΙΑΓΡΑΜΜΑ ΔΙΑΣΠΟΡΑΣ ΑΝΤΑΓΩΝΙΣΤΙΚΟΤΗΤΑΣ ΚΑΙ ΜΕΓΑΛΥΤΕΡΩΝ ΔΥΣΚΟΛΙΩΝ ΠΡΟΣΒΑΣΗΣ ΣΕ ΧΡΗΜΑΤΟΔΟΤΗΣΗ



Είναι απολύτως εμφανές στο παραπάνω διάγραμμα διασποράς ότι ανάμεσα στην ανταγωνιστικότητα και στις μεγαλύτερες δυσκολίες πρόσβασης σε χρηματοδότηση υπάρχει μία **<u>έντονη αρνητική</u>** γραμμική συσχέτιση. Αυτό προφανώς σημαίνει ότι όσο αυξάνονται οι δυσκολίες πρόσβασης σε χρηματοδότηση τόσο μειώνεται η ανταγωνιστικότητα της επιχείρησής μας.

5.3.5.4 ΔΙΑΓΡΑΜΜΑ ΔΙΑΣΠΟΡΑΣ ΑΝΤΑΓΩΝΙΣΤΙΚΟΤΗΤΑΣ ΚΑΙ ΥΨΗΛΟΤΕΡΗΣ ΦΟΡΟΛΟΓΗΣΗΣ



Επίσης και στο παραπάνω διάγραμμα διασποράς που όμως σε αυτή την περίπτωση αφορά την ανταγωνιστικότητα και την υψηλή φορολόγηση βλέπουμε ότι πάλι υπάρχει **αρνητική** γραμμική συσχέτιση μεταξύ των μεγεθών που όμως δε θα μπορούσαμε να πούμε ότι είναι ιδιαίτερα έντονη όπως στο αμέσως προηγούμενο διάγραμμα διασποράς. Επομένως, όσο υψηλότερη είναι η φορολόγηση τόσο θα μειώνεται η ανταγωνιστικότητά μας.

5.3.6. ΕΛΕΓΧΟΣ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΤΟΥ ΥΠΟΔΕΙΓΜΑΤΟΣ

Κάνοντας χρήση της πολλαπλής παλινδρόμησης μέσα από το SPSS προκύπτει η παρακάτω εξίσωση παλινδρόμησης, οι συντελεστές φαίνονται και στον πίνακα **5.2 coefficients** και συγκεκριμένα στη στήλη **unstandardized coefficients (B)** και βρίσκονται σε κόκκινο χρώμα:

$$\mathbf{Y = 6{,}794 + 0{,}530X_1 - 0{,}413X_2 - 3{,}622X_3 - 0{,}245X_4}$$

Με **Y**= η ανταγωνιστικότητά μας

X_1= η μεγαλύτερη καινοτομικότητα

X_2= τα μεγαλύτερα γραφειοκρατικά εμπόδια

X_3= οι μεγαλύτερες δυσκολίες πρόσβασης σε χρηματοδότηση

X_4= η υψηλότερη φορολόγηση

Για να ξεκινήσουμε τον έλεγχο του υποδείγματος πρώτα από όλα θα ελέγξουμε αν τα πρόσημα των παραπάνω μεταβλητών συμβαδίζουν με την αρχική μας οικονομική θεωρία. Παρατηρούμε λοιπόν στον πίνακα **5.2 coefficients** και συγκεκριμένα στη στήλη **unstandardized coefficients (B)** ότι τα πρόσημα που προέκυψαν είναι όντως σωστά. Δηλαδή αν αυξηθεί η καινοτομικότητα θα αυξηθεί η ανταγωνιστικότητά μας ενώ αν αυξηθούν τα γραφειοκρατικά εμπόδια, οι δυσκολίες πρόσβασης σε χρηματοδότηση και η φορολόγηση τότε θα μειωθεί η ανταγωνιστικότητά μας.

Στη συνέχεια εξετάζουμε την στατιστική σημαντικότητα κάθε μεταβλητής ξεχωριστά με τη βοήθεια του **T- Test**. Ο έλεγχος που κάνουμε είναι ο εξής $\mathbf{H_0 : B_1 = 0}$ **vs.** $\mathbf{H_1 : B_1 \neq 0}$ σε επίπεδο σημαντικότητας $\mathbf{\alpha = 5\%}$. Αν κάνουμε αυτό τον έλεγχο και για τα υπόλοιπα **B** (παραμέτρους) θα δούμε ότι όλες οι μεταβλητές είναι στατιστικά σημαντικές διότι το **p_value** ή αλλιώς **Sig** της κάθε μίας είναι μικρότερο από το **0,05**. Τις τιμές για να κάνουμε τον έλεγχο τις παίρνουμε από τον πίνακα **5.2 coefficients** και συγκεκριμένα από τη στήλη **Sig**.

Το επόμενο βήμα είναι να δούμε με τη χρήση του **F – Test** αν έχουμε στατιστική σημαντικότητα σε όλες μαζί τις μεταβλητές ταυτόχρονα. Βλέπουμε λοιπόν ότι το **Sig** στον πίνακα **5.4 ANOVA** είναι και αυτό μικρότερο από το **0,05**, άρα έχοντας κάνει πριν τον έλεγχο : $\mathbf{H_0 : B_0,B_1,B_2,B_3,B_4 = 0}$ **vs.** $\mathbf{H_1 : B_0,B_1,B_2,B_3,B_4 \neq 0}$ καταλήγουμε στο συμπέρασμα ότι και όλες οι μεταβλητές μαζί είναι στατιστικά σημαντικές.

Πίνακας 5.1

Variables Entered/Removed

Model	Variables Entered	Variables Removed	Method
1	Υψηλότερη φορολόγηση, Μεγαλύτερη καινοτομικότητα, Μεγαλύτερα γραφειοκρατικά εμπόδια, Μεγαλύτερες δυσκολίες πρόσβασης σε χρηματοδότηση	,	Enter

a All requested variables entered.

b Dependent Variable: Η ανταγωνιστικότητά μας

Πίνακας 5.2

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
	B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
Constant)	6,794	,582		11,682	,000					
εγαλύτερη ινοτομικότητα	,530	,219	,189	2,417	,020	,176	,336	,186	,972	1,029
εγαλύτερα αφειοκρατ ά εμπόδια	-,413	,168	-,192	-2,452	,018	-,232	-,340	-,189	,965	1,036
γαλύτερες υσκολίες οόσβασης σε ηματοδότηση	-3,622	,406	-,717	-8,930	,000	-,791	-,796	-,688	,923	1,084
ψηλότερη ρολόγηση	-,245	,097	-,202	-2,530	,015	-,405	-,350	-,195	,929	1,077

Dependent Variable: Η ανταγωνιστικότητά μας

Πίνακας 5.3

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
				R Square Change	F Change	df1	df2	Sig. F Change	
,852	,727	,703	,1714	,727	30,568	4	46	,000	2,084

ictors: (Constant), Μεγαλύτερη καινοτομικότητα, Μεγαλύτερες δυσκολίες πρόσβασης σε

τοδότηση, Μεγαλύτερα γραφειοκρατικά εμπόδια, Υψηλότερη φορολόγηση

endent Variable: Η ανταγωνιστικότητά μας

Πίνακας 5.4

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3,593	4	,898	30,568	,000
	Residual	1,352	46	2,939E-02		
	Total	4,945	50			

a Predictors: (Constant), Μεγαλύτερη καινοτομικότητα, Μεγαλύτερες δυσκολίες πρόσβασης σε χρηματοδότηση, Μεγαλύτερα γραφειοκρατικά εμπόδια, Υψηλότερη φορολόγηση

b Dependent Variable: Η ανταγωνιστικότητά μας

5.3.7 ΕΛΕΓΧΟΣ ΤΟΥ ΥΠΟΔΕΙΓΜΑΤΟΣ ΓΙΑ ΠΟΛΥΣΥΓΓΡΑΜΜΙΚΟΤΗΤΑ

Η πολυσυγγραμμικότητα εμφανίζεται όταν δύο ή περισσότερες μεταβλητές σχετίζονται μεταξύ τους γραμμικά και κατά συνέπεια μπορεί η μία να αντικαταστήσει την άλλη. Μας ενδιαφέρει να εξετάσουμε όχι αν υπάρχει αλλά κατά πόσο έντονη είναι και είναι χαρακτηριστικό του δείγματος και όχι του πληθυσμού.

Για να ανιχνεύσουμε την πολυσυγγραμμικότητα υπάρχουν διάφοροι τρόποι.

- Εμφανίζεται όταν έχουμε υψηλό R_2 ενώ ταυτόχρονα λίγες από τις μεταβλητές μας είναι στατιστικά σημαντικές. Στο συγκεκριμένο υπόδειγμα παρατηρούμε ότι δεν παρουσιάζεται αυτό το φαινόμενο αφού έχουμε και υψηλό **R_2** και όλες οι μεταβλητές είναι στατιστικά σημαντικές. Αυτό φαίνεται στους πίνακες **5.2 Coefficients** και **5.3 Model Summary** που είδαμε πιο πάνω.

- Οι υψηλοί συντελεστές συσχέτισης μεταξύ ζευγών ανεξάρτητων μεταβλητών φανερώνουν ότι υπάρχει πρόβλημα πολυσυγγραμμικότητας. Όμως παρατηρώντας τον παρακάτω πίνακα **5.5 coefficient correlations** βλέπουμε ότι η συσχέτιση μεταξύ των μεταβλητών είναι πολύ μικρή, άρα θα μπορούσαμε να πούμε ότι δεν παρουσιάζεται ιδιαίτερο πρόβλημα συγγραμμικότητας.

Πίνακας 5.5

Coefficient Correlations

	Υψηλότερη φορολόγηση	**Μεγαλύτερη καινοτομικότητα**	**Μεγαλύτερα γραφειοκρατικά εμπόδια**	**Μεγαλύτερες δυσκολίες πρόσβασης σε χρηματοδότηση**
Υψηλότερη φορολόγηση	1,000	,026	-,016	-,263
Μεγαλύτερη αινοτομικότητα	,026	1,000	-,164	,025
Μεγαλύτερα ραφειοκρατικά εμπόδια	-,016	-,164	1,000	-,085
Μεγαλύτερες δυσκολίες πρόσβασης σε	-,263	,025	-,085	1,000

ependent Variable: Η ανταγωνιστικότητά μας

- Ελέγχουμε την τιμή του **Tolerance** η οποία αν είναι κάτω από **0,20** τότε πιθανώς έχουμε πρόβλημα πολυσυγγραμμικότητας. Μετά ελέγχουμε και την τιμή του **VIF** η οποία αν ξεπερνά το **4** τότε επίσης έχουμε πρόβλημα πολυσυγγραμμικότητας. Στο υπόδειγμά μας βλέπομε ότι οι τιμές του **Tolerance** και **VIF** είναι τέτοιες που δεν φανερώνουν ότι υπάρχει πρόβλημα πολυσυγγραμμικότητας. Οι τιμές αυτές βρίσκονται στον πίνακα **5.6 coefficients** και συγκεκριμένα στη στήλη **collinearity statistics.**

Πίνακας 5.6

Coefficients

		Collinearity Statistics	
Model		**Tolerance**	**VIF**
1	(Constant)		
	Μεγαλύτερη καινοτομικότητα	**,972**	**1,029**
	Μεγαλύτερα γραφειοκρατικά εμπόδια	**,965**	**1,036**
	Μεγαλύτερες δυσκολίες πρόσβασης σε χρηματοδότηση	**,923**	**1,084**
	Υψηλότερη φορολόγηση	**,929**	**1,077**

a Dependent Variable: Η ανταγωνιστικότητά μας

- Στη συνέχεια ελέγχουμε την τιμή του **Eigenvalue** για κάθε μία μεταβλητή, όσο περισσότερα **Eigenvalue** βρίσκονται κοντά στο 0 τόσο πιθανή είναι η ύπαρξη πολυσυγγραμμικότητας. Όπως βλέπουμε στον παρακάτω πίνακα **5.7 collinearity diagnostics** όλα τα **Eigenvalue** απέχουν από το μηδέν άρα δεν έχουμε πρόβλημα πολυσυγγραμμικότητας.

Πίνακας 5.7

Collinearity Diagnostics

		Eigenvalue	Condition Index	Variance Proportions				
Model	Dimension			(Constant)	Μεγαλύτερη καινοτομικότητα	Μεγαλύτερα γραφειοκρατικά εμπόδια	Μεγαλύτερες δυσκολίες πρόσβασης σε χρηματοδότηση	Υψηλότερη η φορολόγηση
1	1	4,960	1,000	,00	,00	,00	,00	,00
	2	2,472E-02	14,165	,00	,82	,00	,01	,11
	3	1,008E-02	22,178	,02	,13	,03	,06	,87
	4	3,985E-03	35,282	,04	,05	,15	,87	,01
	5	1,070E-03	68,077	,94	,00	,82	,06	,01

a Dependent Variable: Η ανταγωνιστικότητά μας

- Τέλος χρησιμοποιούμε τις βοηθητικές εξισώσεις ή αλλιώς τις βοηθητικές παλινδρομήσεις ανάμεσα στις ανεξάρτητες μεταβλητές και ελέγχουμε και σε αυτές τους δείκτες **Tolerance** και **VIF**. Αυτό το κάνουμε για κάθε ανεξάρτητη μεταβλητή ξεχωριστά κάθε φορά.

Πίνακας 5.8

Coefficients

		Unstandardized Coefficients	Sig.	Collinearity Statistics	
Model		B		Tolerance	VIF
1	(Constant)	4,289	,000		
	Μεγαλύτερες δυσκολίες πρόσβασης σε χρηματοδότηση	-,834	,129	,355	2,816
	Υψηλότερη φορολόγηση	-6,055E-02	,478	,824	1,213
	Μεγαλύτερη καινοτομικότητα	,338	,075	,925	1,081
	Η ανταγωνιστικότητά μας	-,280	,018	,309	3,235

a Dependent Variable: Μεγαλύτερα γραφειοκρατικά εμπόδια

Έχοντας λοιπόν ως εξαρτημένη μεταβλητή τα μεγαλύτερα γραφειοκρατικά εμπόδια βλέπουμε στον πίνακα **5.8 coefficients** στη στήλη **collinearity statistics** ότι οι τιμές των δεικτών είναι ικανοποιητικές, το ίδιο παρατηρείται και στους παρακάτω πίνακες 5.9, 5,10, 5.11 που η εξαρτημένη μεταβλητή αλλάζει σε κάθε περίπτωση. Άρα πάλι συμπεραίνουμε ότι δεν υπάρχει πρόβλημα πολυσυγγραμμικότητας.

Πίνακας 5.9

Coefficients

		Unstandardized Coefficients	Sig.	Collinearity Statistics	
Model		B		Tolerance	VIF
1	(Constant)	1,398	,000		
	Υψηλότερη φορολόγηση	-1,994E-02	,381	,829	1,206
	Μεγαλύτερη καινοτομικότητα	8,795E-02	,083	,922	1,085
	Η ανταγωνιστικότητά μας	-,175	,000	,747	1,338
	Μεγαλύτερα γραφειοκρατικά εμπόδια	-5,932E-02	,129	,898	1,113

a Dependent Variable: Μεγαλύτερες δυσκολίες πρόσβασης σε χρηματοδότηση

Πίνακας 5.10

Coefficients

		Unstandardized Coefficients	Sig.	Collinearity Statistics	
Model		**B**		**Tolerance**	**VIF**
1	**(Constant)**	**4,384**	**,006**		
	Μεγαλύτερη καινοτομικότητα	**,213**	**,524**	**,870**	**1,149**
	Η ανταγωνιστικότητά μας	**-,499**	**,015**	**,311**	**3,211**
	Μεγαλύτερα γραφειοκρατικά εμπόδια	**-,182**	**,478**	**,863**	**1,158**
	Μεγαλύτερες δυσκολίες πρόσβασης σε χρηματοδότηση	**-,840**	**,381**	**,343**	**2,913**

a Dependent Variable: Υψηλότερη φορολόγηση

Πίνακας 5.11

Coefficients

		Unstandardized Coefficients	Sig.	Collinearity Statistics	
Model		B		Tolerance	VIF
	(Constant)	-1,146	,115		
	Η ανταγωνιστικότητά μας	,212	,020	,308	3,246
	Μεγαλύτερα γραφειοκρατικά εμπόδια	,199	,075	,915	1,092
	Μεγαλύτερες δυσκολίες πρόσβασης σε χρηματοδότηση	,729	,083	,361	2,773
	Υψηλότερη φορολόγηση	4,185E-02	,524	,823	1,216

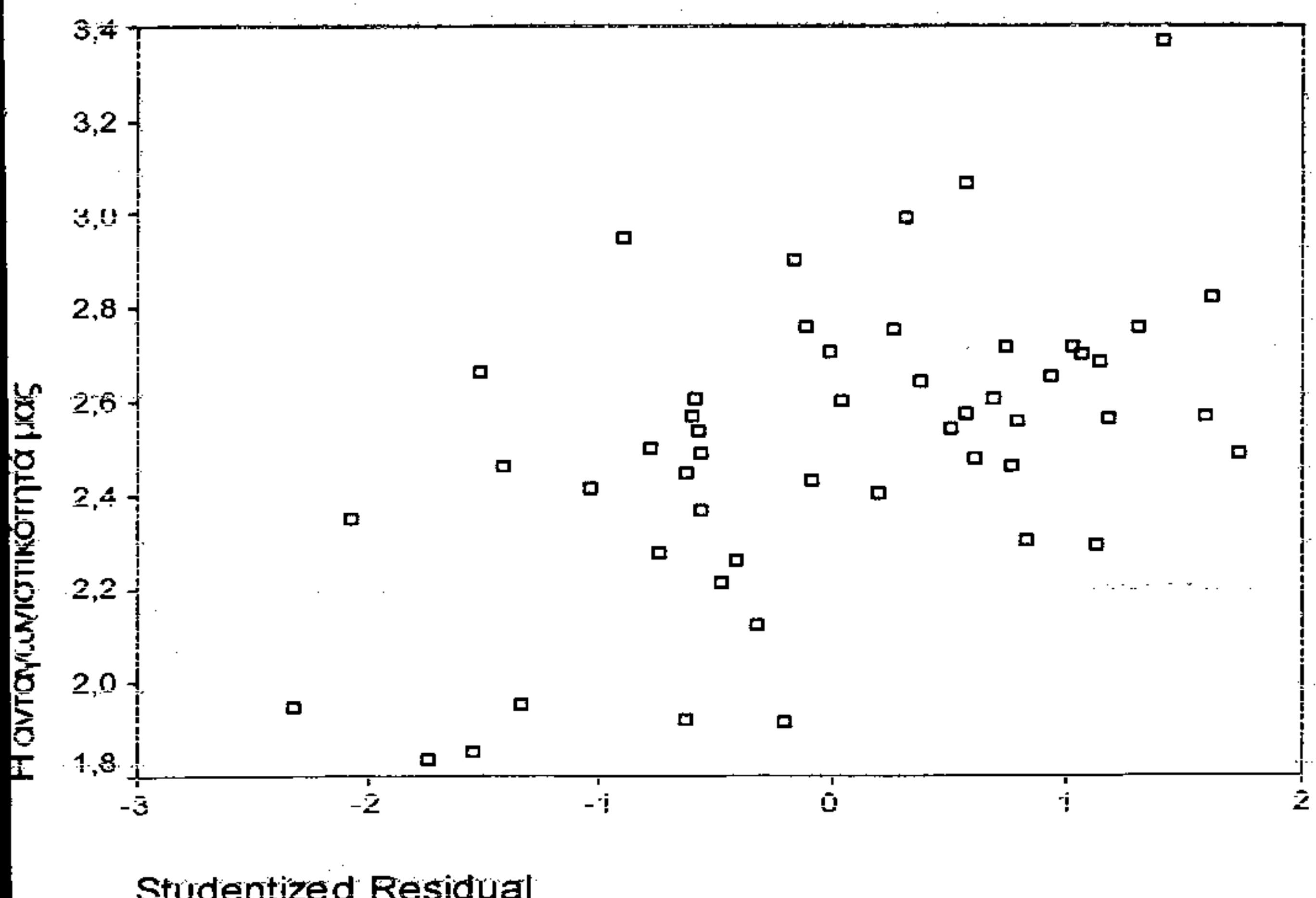
a Dependent Variable: Μεγαλύτερη καινοτομικότητα

5.3.8 ΕΛΕΓΧΟΣ ΤΟΥ ΥΠΟΔΕΙΓΜΑΤΟΣ ΓΙΑ ΕΤΕΡΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑ

Η ετεροσκεδαστικότητα παρουσιάζεται όταν η διακύμανση των U_i στο υπόδειγμα δεν είναι σταθερή.Το πρόβλημα της ετεροσκεδαστικότητας συναντάται συνήθως στα διαστρωματικά δεδομένα όπως είναι και τα δεδομένα του υποδείγματός μας. Οφείλεται στην ύπαρξη ανομοιογενών μονάδων ή ετερογενών μονάδων στο υπόδειγμα.

Την ετεροσκεδαστικότητα μπορούμε να την ανιχνεύσουμε με τους εξής τρόπους :

- Αν τα $\mathbf{e}_i$ (residuals) έχουν μία συγκεκριμένη μορφή και όχι τυχαία και φαίνεται ότι η διακύμανση δεν είναι σταθερή τότε έχουμε πρόβλημα ετεροσκεδαστικότητας.



Παρατηρούμε λοιπόν στο παραπάνω διάγραμμα διασποράς ανάμεσα στην ανταγωνιστικότητα και τα **Studentized Residual** ότι η διακύμανση των **residuals** είναι σταθερή και κυμαίνεται από **[-2, 2]** επίσης θα μπορούσαμε να πούμε ότι τα $\mathbf{e_i}$ δεν ακολουθούν κάποιο συγκεκριμένο σχήμα οπότε δεν έχουμε πρόβλημα ετεροσκεδαστικότητας. Το ότι υπάρχουν ένα ή δύο $\mathbf{e_i}$ εκτός του παραπάνω διαστήματος δεν μας επηρεάζει.

- Ελέγχουμε τον δείκτη **Cook's Distance** και τον συγκρίνουμε με τον εξής λόγο : **4/(n – k – 1)**, όπου **n** ο αριθμός των παρατηρήσεων και **k** ο αριθμός των ανεξάρτητων μεταβλητών. Όταν το αποτέλεσμα του δείκτη **Cook's Distance** είναι μικρότερο από το αποτέλεσμα της διαίρεσης (κρίσιμη τιμή) τότε δεν έχουμε πρόβλημα ετεροσκεδαστικότητας. Στο υπόδειγμά μας η κρίση τιμή που προκύπτει είναι **4/(51 – 4 – 1) = 0,087.** Αν κάνουμε τον έλεγχο με την τιμή του **Cook's Distance** από τον πίνακα **5.12 residual statistics** που είναι **0,21** βλέπουμε ότι δεν έχουμε στο υπόδειγμά μας πρόβλημα ετεροσκεδαστικότητας.

Πίνακας 5.12

Residuals Statistics

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	1,9521	3,1505	2,5051	,2681	51
Std. Predicted Value	-2,063	2,407	,000	1,000	51
Standard Error of Predicted Value	2,591E-02	,1181	5,048E-02	1,844E-02	51
Adjusted Predicted Value	1,9573	3,1323	2,5053	,2655	51
Residual	-,3878	,2864	6,357E-16	,1644	51
Std. Residual	-2,262	1,671	,000	,959	51
Stud. Residual	-2,320	1,724	,000	1,006	51
Deleted Residual	-,4080	,3098	-1,8398E-04	,1814	51
Stud. Deleted Residual	-2,442	1,763	-,004	1,022	51
Mahal. Distance	,162	22,754	3,922	4,246	51
Cook's Distance	,000	,129	,021	,027	51
Centered Leverage Value	,003	,455	,078	,085	51

a Dependent Variable: Η ανταγωνιστικότητά μας

- Τέλος χρησιμοποιούμε το **Spearman's Runk Correlation Test** στο οποίο όσο πιο πολλά αστεράκια υπάρχουν τόσο οι συσχετίσεις μεταξύ των ανεξάρτητων μεταβλητών είναι στατιστικά σημαντικές και άρα έχουμε πρόβλημα ετεροσκεδαστικότητας. Στον παρακάτω πίνακα **5.13 correlations** βλέπουμε ότι τα αστεράκια είναι ελάχιστα άρα δεν έχουμε πρόβλημα ετεροσκεδαστικότητας.

Πίνακας 5.13

Correlations

			Η ανταγωνιστικότητά μας	Μεγαλύτερη καινοτομικότητα	Μεγαλύτερα γραφειοκρατικά εμπόδια	Μεγαλύτερες δυσκολίες πρόσβασης σε χρηματοδότηση	Υψηλότερη φορολόγηση
Spearman's rho	Η ανταγωνιστικότητά μας	Correlation Coefficient	1,000	,007	-,123	-,755**	-,365**
		Sig. (2-tailed)	,	,961	,391	,000	,008
		N	51	51	51	51	51
	Μεγαλύτερη καινοτομικότητα	Correlation Coefficient	,007	1,000	,261	,068	,091
		Sig. (2-tailed)	,961	,	,064	,636	,526
		N	51	51	51	51	51
	Μεγαλύτερα γραφειοκρατικά εμπόδια	Correlation Coefficient	-,123	,261	1,000	-,003	-,037

Συνέχεια Πίνακα 5.13

Correlations

	Sig. (2-tailed)	,391	,064	,	,985	,799
	N	51	51	51	51	51
Μεγαλύτερες δυσκολίες πρόσβασης σε χρηματοδότηση	Correlation Coefficient	-,755**	,068	-,003	1,000	,350*
	Sig. (2-tailed)	,000	,636	,985	,	,012
	N	51	51	51	51	51
Υψηλότερη φορολόγηση	Correlation Coefficient	-,365**	,091	-,037	,350*	1,000
	Sig. (2-tailed)	,008	,526	,799	,012	,
	N	51	51	51	51	51

Correlation is significant at the .01 level (2-tailed).

orrelation is significant at the .05 level (2-tailed).

Όμως επειδή το παραπάνω **Test** μέσω του **SPSS** δεν είμαστε σίγουροι κατά πόσο είναι αξιόπιστο προτιμότερο θα ήταν να κάναμε και τον έλεγχο εμείς οι ίδιοι με το χέρι αλλά δυστυχώς μία τέτοια ενέργεια είναι χρονοβόρα γι'αυτό το λόγο τελικά δε θα προβούμε στην υλοποίησή της. Το πιο πιθανό είναι τα αποτελέσματα που θα έβγαιναν να επαλήθευαν τον παραπάνω ισχυρισμό μας, δηλαδή ότι δεν υπάρχει πρόβλημα ετεροσκεδαστικότητας.

5.3.9 ΕΛΕΓΧΟΣ ΤΟΥ ΥΠΟΔΕΙΓΜΑΤΟΣ ΓΙΑ ΑΥΤΟΣΥΣΧΕΤΙΣΗ

Η αυτοσυσχέτιση είναι η συσχέτιση παρατηρήσεων οι οποίες είναι διατεταγμένες στο χρόνο όπως οι χρονοσειρές ή στο χώρο όπως τα διαστρωματικά δεδομένα. Κυρίως όμως συναντάται σε χρονοσειρές. Στο δεδομένο υπόδειγμα τα δεδομένα μας είναι διαστρωματικά και το πιο πιθανό είναι να μη συναντήσουμε το πρόβλημα της αυτοσυσχέτισης, σε περίπτωση όμως που αντιμετωπίσουμε πρόβλημα τότε η αυτοσυσχέτιση στα διαστρωματικά δεδομένα ονομάζεται χωρική. Η αυτοσυσχέτιση συναντάται όταν η αναμενόμενη τιμή δύο διαφορετικών U δεν είναι μηδέν δηλαδή, το σφάλμα το οποίο συνδέεται με κάποια παρατήρηση σχετίζεται και επηρεάζεται από το σφάλμα που συνδέεται με οποιαδήποτε άλλη παρατήρηση. Με άλλα λόγια τα **U** δεν είναι ανεξάρτητα και τυχαία.

Οι τρόποι ανίχνευσης της αυτοσυσχέτισης είναι οι παρακάτω :

- Ελέγχουμε τη διασπορά των $\mathbf{e}_i$ και επειδή η αυτοσυσχέτιση αναφέρεται όπως είπαμε κυρίως σε χρονοσειρές στον κάθετο άξονα θα βάζαμε τον χρόνο **t** και στον οριζόντιο τα $\mathbf{e}_i$ αλλά επειδή τα δεδομένα μας είναι διαστρωματικά δε θα προχωρήσουμε στον παραπάνω έλεγχο.

- Προχωράμε στη διεξαγωγή του **Runs Test**, αν η τιμή του **Sig** είναι μεγαλύτερη από **α=0,05** τότε αυτό σημαίνει ότι δεν έχουμε αυτοσυσχέτιση. Αυτό προκύπτει από τον εξής έλεγχο :

  $\mathbf{H_0: E(U_i\ U_j)=0}$ **vs.** $\mathbf{H_1: E\ (U_i\ U_j) \neq 0}$. Όπως βλέπουμε στον ακριβώς από κάτω πίνακα **5.14 runs test** η τιμή που προκύπτει είναι **0,673** αυτό μας δείχνει ότι δεν έχουμε πρόβλημα αυτοσυσχέτισης.

Πίνακας 5.14

Runs Test

	RESIDUAL
Test Value	1,882E-03
Cases < Test Value	26
Cases >= Test Value	25
Total Cases	51
Number of Runs	25
Z	-,422
Asymp. Sig. (2-tailed)	,673

a. Mean

Πίνακας 5.15

Casewise Diagnostics

Case Number	Std. Residual	Η ανταγωνιστικότητά μας	Predicted Value	Residual
1	1,124	2,56	2,3691	,1927
2	,487	2,54	2,4586	8,345E-02
3	1,001	2,71	2,5431	,1716
4	-,717	2,28	2,3994	-,1229
5	-,544	2,54	2,6297	-9,3261E-02
6	,699	2,72	2,5971	,1198
7	-,581	2,60	2,7044	-9,9654E-02
8	-1,649	1,84	2,1218	-,2827
9	1,671	2,49	2,2031	,2864
10	-1,357	2,46	2,6957	-,2326
11	,745	2,46	2,3326	,1277
12	-1,002	2,42	2,5877	-,1717
13	-,762	2,95	3,0813	-,1307
14	1,098	2,68	2,4946	,1882
15	,254	2,75	2,7109	4,357E-02
16	1,538	2,57	2,3069	,2637
17	,532	3,07	2,9763	9,114E-02

Συνέχεια Πίνακα 5.15 Casewise Diagnostics				
18	-,155	2,90	2,9281	-2,6554E-02
19	-,611	2,44	2,5485	-,1047
20	,586	2,48	2,3773	,1004
21	1,036	2,70	2,5251	,1776
22	-1,272	1,95	2,1713	-,2180
23	-,487	1,92	2,0037	-8,3556E-02
24	-1,428	2,66	2,9093	-,2448
25	,368	2,64	2,5804	6,312E-02
26	-,085	2,43	2,4433	-1,4530E-02
27	-,437	2,21	2,2881	-7,4850E-02
28	-,764	2,50	2,6299	-,1310
29	-2,018	2,35	2,6946	-,3459
30	-,102	2,76	2,7773	-1,7501E-02
31	-,299	2,12	2,1747	,1927
32	1,450	2,82	2,5762	8,345E-02
33	,790	2,30	2,1677	,1716
34	-,012	2,71	2,7075	-,1229
35	-1,461	1,85	2,1014	-9,3261E-02
36	-,554	2,37	2,4605	,1198
37	1,201	2,76	2,5520	-9,9654E-02
38	,900	2,65	2,4983	-,2827
39	-2,262	1,95	2,3356	,2864
40	-,406	2,26	2,3326	-,2326
41	1,281	3,37	3,1505	,1277
42	,230	2,99	2,9551	-,1717
43	-,584	2,57	2,6669	-,1307

Συνέχεια Πίνακα 5.15

Casewise Diagnostics

44	,670	2,60	2,4880	,1882
45	-,195	1,92	1,9521	4,357E-02
46	,565	2,57	2,4741	,2637
47	,182	2,40	2,3716	9,114E-02
48	,777	2,56	2,4227	-2,6554E-02
49	,034	2,60	2,5958	-,1047
50	-,553	2,49	2,5823	,1004
51	1,079	2,29	2,1076	,1776

a Dependent Variable: Η ανταγωνιστικότητά μας

* **Υποσημείωση** : Επειδή ο αριθμός των παρατηρήσεων είναι παραπάνω από 40 και κατ'επέκταση και ο αριθμός των **Residual**, τότε στα μη παραμετρικά test δεν κάνουμε καμία υπόθεση για την κατανομή των δεδομένων.

- Τέλος ελέγχουμε την τιμή του δείκτη **Durbin – Watson** με τη βοήθεια του πίνακα **5.16 model summary** και του **Durbin – Watson d statistic** που βρίσκεται στο παράρτημα. . Αν ο δείκτης αυτός κυμαίνεται από **[0-4]** τότε μάλλον δεν έχουμε πρόβλημα αυτοσυσχέτισης. Όμως εμείς θα πρέπει να ελέγξουμε τις κρίσιμες τιμές που προκύπτουν για το υπόδειγμά μας με την τιμή του δείκτη **Durbin – Watson**. Το d_L είναι **1,378** και το d_U είναι **1,721**. Αυτές οι τιμές προκύπτουν από τον πίνακα **Durbin – Watson d statistic** για **n=51≈50** και **κ' = 4**. Ο δείκτης είναι **2,084** και βρίσκεται ανάμεσα στο διάστημα $d_U \leq$ **2,084** $\leq$ **4-** d_U => **1,721≤2,084≤2,279** άρα σύμφωνα με τη θεωρία αυτή η ανίσωση φανερώνει ότι σίγουρα δεν υπάρχει αυτοσυσχέτιση.

Πίνακας 5.16

Model Summary

	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
Model					
	,852	,727	,703	,1714	2,084

a Predictors: (Constant), Μεγαλύτερη καινοτομικότητα, Μεγαλύτερες δυσκολίες πρόσβασης σε χρηματοδότηση, Μεγαλύτερα γραφειοκρατικά εμπόδια, Υψηλότερη φορολόγηση

b Dependent Variable: Η ανταγωνιστικότητά μας

5.4 ΣΥΜΠΕΡΑΣΜΑ

Αφού λοιπόν κάναμε όλους τους αναγκαίους ελέγχους στο υπόδειγμά μας, καταλήξαμε τελικά ότι δεν παρουσιάζει κανένα πρόβλημα πολυσυγγραμμικότητας, ετεροσκεδαστικότητας και αυτοσυσχέτισης. Οπότε το παραπάνω υπόδειγμα που προέκυψε ύστερα από πολλές προσπάθειες και συγκρίσεις μπορεί να θεωρηθεί ότι είναι ικανό να ερμηνεύσει σε μεγάλο βαθμό την ανταγωνιστικότητά μας που είναι η εξαρτημένη μας μεταβλητή. Άρα το υπόδειγμα είναι δυνατόν να χρησιμοποιηθεί για να κάνουμε πρόβλεψη.

ΚΕΦΑΛΑΙΟ 6ο

ΓΕΝΙΚΑ ΣΥΜΠΕΡΑΣΜΑΤΑ

Όπως, ήδη έχουμε δει στα κεφάλαια 2 και 3 αναφέραμε κριτήρια για την αξιολόγηση της συνολικής ικανότητας ενός εκτιμώμενου υποδείγματος. Αρχικά, ξεκινήσαμε με τη μέθοδο της απλής παλινδρόμησης, όπου ο σκοπός της είναι να αναλύσει και να εκτιμήσει κυρίως τη σχέση που μπορεί να υπάρχει ανάμεσα σε μία εξαρτημένη μεταβλητή **Y** και μία ανεξάρτητη μεταβλητή **X**. Μια τέτοια ανάλυση μπορεί να μας δείξει το βαθμό εξάρτησης μεταξύ των δύο μεταβλητών.

Το μοντέλο της απλής παλινδρόμησης όπως έχουμε δει στηρίζεται σε κάποιες υποθέσεις που αν παραβιαστούν τα αποτελέσματα και τα τελικά συμπεράσματα θα γίνουν αναξιόπιστα επηρεάζοντας έτσι τις τελικές εκτιμήσεις μας σχετικά με την ερμηνεία του μοντέλου.

Η γραφική παράσταση της απλής γραμμικής παλινδρόμησης είναι ένα χρήσιμο εργαλείο και παίζει σπουδαίο ρόλο στη διεξαγωγή συμπερασμάτων. Για αυτό το λόγο θα πρέπει να είναι όσο το δυνατόν πιο αντιπροσωπευτική της πραγματικότητας και με τη βοήθεια του συντελεστή προσδιορισμού, του συντελεστή συσχέτισης και των διαστημάτων εμπιστοσύνης να μπορέσουμε να έχουμε προβλέψεις όπου να είναι έγκυρες και αξιόπιστες οδηγώντας μας σε ασφαλή συμπεράσματα σχετικά με το φαινόμενο που εξετάζουμε κάθε φορά.

Από την άλλη πλευρά για πολυμεταβλητά μοντέλα χρησιμοποιούμε τη μέθοδο της πολλαπλής παλινδρόμησης. Όμως όπως στην απλή παλινδρόμηση έτσι και στην πολλαπλή παλινδρόμηση υπάρχουν δυσκολίες στην εκτίμηση των εκάστοτε παραμέτρων διότι μπορεί να υπάρξουν σφάλματα που οφείλονται στην ατελή θεωρία, σε ατελή εξειδίκευση και σε υπολογιστικά λάθη.

Οι συντελεστές τις πολλαπλής παλινδρόμησης που εκτιμούνται με τη μέθοδο των ελαχίστων τετραγώνων όπως είδαμε δείχνουν την καθαρή μεταβολή που επέρχεται στην εξαρτημένη μεταβλητή όταν έχουμε μοναδιαία μεταβολή στην ανεξάρτητη μεταβλητή.

Ο βαθμός της προσαρμοστικής ικανότητας του υποδείγματός μας φαίνεται με τον συντελεστή πολλαπλού προσδιορισμού. Επίσης, σημαντικό ρόλο στις εκτιμήσεις μας παίζει και ο συντελεστής πολλαπλής συσχέτισης, ο οποίος είναι πάντα θετικός και αποτελεί μία γενίκευση της απλής συσχέτισης στην περίπτωση των δύο ή περισσοτέρων ανεξάρτητων μεταβλητών.

Για να ελεγχθεί η συνολική σημαντικότητα του μοντέλου πολλαπλής παλινδρόμησης όπως είδαμε θα πρέπει να αναλύσουμε τη διακύμανση και να ελέγξουμε τη σημαντικότητα των επιμέρους μεταβλητών. Επίσης θα πρέπει να παρατηρήσουμε αν το υπόδειγμά μας βρίσκεται κάτω από την επιρροή των τριών προβλημάτων που μπορούν να παρουσιαστούν κατά την εφαρμογή της μεθόδου ελαχίστων τετραγώνων στην πολλαπλή παλινδρόμηση. Αυτά τα προβλήματα έχουν κατονομαστεί και είναι η **ετεροσκεδαστικότητα**, που σημαίνει ότι η διακύμανση των $\mathbf{u}_i$ δεν είναι σταθερή δημιουργώντας μας τα προβλήματα που έχουμε προαναφέρει στο αντίστοιχο κεφάλαιο, η **αυτοσυσχέτιση**, δηλαδή τα $\mathbf{u}_i$ δεν είναι ανεξάρτητα μεταξύ τους και τυχαία, άρα το υπόδειγμα μας δεν μπορεί να θεωρηθεί ότι είναι έγκυρο. Τέλος έχουμε το πρόβλημα της **πολυσυγγραμμικότητας** το οποίο εμφανίζεται όταν δύο οι περισσότερες μεταβλητές σχετίζονται μεταξύ τους γραμμικά. Βέβαια, αυτό που μας ενδιαφέρει στην περίπτωση της πολυσυγγραμμικότητας είναι να δούμε όχι αν υπάρχει αλλά πόσο έντονη είναι, έτσι ώστε να δούμε πως μπορούμε να την αντιμετωπίσουμε.

Η επιλογή του καλύτερου μοντέλου παλινδρόμησης είδαμε ότι απαιτεί έναν συμβιβασμό μεταξύ της επίτευξής του υψηλότερου συντελεστή προσδιορισμού $\mathbf{R_2}$ και της επιλογής ενός ελάχιστου αριθμού μεταβλητών πρόβλεψης. Το $\mathbf{R^2}$ μετρά το ποσοστό της απόκλισης της εξαρτημένης μεταβλητής το οποίο μπορεί να εξηγηθεί από την παλινδρόμηση. Η τιμή του $\mathbf{R^2}$ κυμαίνεται πάντα από 0 έως 1. Το προσαρμοσμένο $\mathbf{R^2}$ είναι σταθμισμένο με τους βαθμούς ελευθερίας κάθε αντίστοιχου αθροίσματος τετραγώνων. Άρα το προσαρμοσμένο $\mathbf{R^2}$ αυξάνεται (βέβαια με την προσθήκη άλλης μιας μεταβλητής δεν αυξάνεται αναγκαστικά διότι κάτι τέτοιο θα αύξανε την τιμή του **k**) ή μειώνεται ανάλογα με το αν η συμβολή της νέας ερμηνευτικής μεταβλητής υπεραντισταθμίζει ή όχι την απώλεια των βαθμών ελευθερίας.

Τώρα περνώντας στο κεφάλαιο 4, είδαμε και ασχοληθήκαμε με διάφορα μοντέλα που δεν ήταν γραμμικά ως προς τις μεταβλητές αλλά ως προς τις παραμέτρους **B**. Βέβαια κάθε ερευνητής μπορεί να συνδυάσει τα χαρακτηριστικά ενός ή και περισσότερων μοντέλων.

Με τον όρο γραμμικό υπόδειγμα αναφερόμαστε σε διάφορους τύπους συναρτήσεων που μπορεί να μη προέρχονται απαραίτητα από το γραμμικό μαθηματικό υπόδειγμα τις οικονομικής θεωρίας. Επίσης, όπως είναι γνωστό, η οικονομική θεωρία προσδιορίζει ακριβείς σχέσεις μεταξύ των οικονομικών μεταβλητών άρα όσο προσεκτικά και αν συγκεντρώσουμε τα στοιχεία πάντα θα έχουμε κάποιες αποκλίσεις της προσαρμοσμένης ευθείας από την αρχική σχέση. Τα σφάλματα εξειδίκευσης που μπορεί να γίνουν κατά τον προσδιορισμό του υποδείγματος πολλές φορές είναι σοβαρά, η λεπτομερή ανάλυσή τους στο 4^{o} κεφάλαιο της εργασίας το αποδεικνύει αυτό με αρκετά κατανοητό τρόπο.

Ο συντελεστής προσδιορισμού $\mathbf{R^2}$ για τον οποίο έχουμε μιλήσει διεξοδικά στα προηγούμενα κεφάλαια μπορεί να μας βοηθήσει στην επιλογή υποδειγμάτων διαφορετικών αλγεβρικών μορφών. Όμως, το $\mathbf{R^2}$ μπορεί να χρησιμοποιηθεί ως κριτήριο επιλογής μόνο μεταξύ υποδειγμάτων που έχουν τον ίδιο αριθμό διαφορετικών ανεξάρτητων μεταβλητών. Σε δύο ή και περισσότερα υποδείγματα με διαφορετική αλγεβρική μορφή και τον ίδιο αριθμό μεταβλητών οι τιμές του $\mathbf{R}^2$ είναι συγκρίσιμες μόνο εφόσον η εξαρτημένη μεταβλητή εκφράζεται στις ίδιες μονάδες σε όλα τα υποδείγματα. Κάτι τέτοιο είναι απόλυτα λογικό και κατανοητό η περαιτέρω ανάλυσή του μπορεί να θεωρηθεί υπερβολή και επανάληψη εννοιών που έχουν αναλυθεί πλήρως στα αντίστοιχα κεφάλαια.

Τελειώνοντας λοιπόν τα γενικά συμπεράσματα της εργασίας θα πρέπει να τονίσουμε ότι η μοντελοποίηση διάφορων φαινόμενων απαιτεί την κατάλληλη ισορροπία όσον αφορά την θεωρία, τη διαθεσιμότητα κατάλληλων δεδομένων, μία καλή ερμηνεία των στατιστικών προδιαγραφών των διαφόρων μοντέλων και η πρακτική κρίση. Βέβαια είναι πολύ δύσκολο να πετύχουμε ένα τέλειο μοντέλο που θα ερμηνεύει και θα δίνει πλήρη στοιχεία για την πρόβλεψη κάποιων φαινόμενων . Αυτό που προσπαθούμε πάντα να πετύχουμε είναι να βρούμε ένα μοντέλο το οποίο να είναι σε γενικές γραμμές καλό και να ικανοποιεί όλα τα απαραίτητα κριτήρια και τις ιδιότητες ενός καλού υποδείγματος.

ΠΑΡΑΡΤΗΜΑ

	ανταγωνι	τιμέσ_αν	αγορέσ	ποιότητα	ανθρώπιν	πληροφορ	προστασί
1	2,56	4,77	6,53	10,05	6,02	7,91	5,80
2	2,54	4,16	5,95	9,16	4,47	7,20	5,57
3	2,71	4,65	5,93	8,91	4,59	7,30	5,46
4	2,28	3,71	5,73	9,00	5,04	7,03	5,10
5	2,54	9,31	9,91	12,08	7,57	11,59	10,05
6	2,72	6,11	7,35	10,12	5,61	8,66	6,90
7	2,60	4,75	6,26	9,36	5,08	7,44	5,83
8	1,84	2,71	5,09	8,22	3,47	6,22	4,34
9	2,49	3,99	6,05	9,06	4,51	7,29	5,43
10	2,46	5,06	6,30	9,74	5,48	8,05	6,57
11	2,46	4,36	6,38	9,50	5,74	7,59	5,76
12	2,42	4,82	6,73	9,80	5,38	8,07	6,19
13	2,95	2,56	4,55	7,64	3,22	5,76	3,83
14	2,68	2,89	4,56	8,22	3,40	6,12	3,69
15	2,75	4,30	6,02	9,74	5,29	7,48	5,66
16	2,57	4,26	6,43	9,55	5,32	7,59	5,90
17	3,07	5,72	7,30	10,07	5,75	8,53	6,70
18	2,90	2,40	5,00	8,54	3,91	6,34	4,01
19	2,44	7,79	8,74	11,10	6,44	10,19	8,83
20	2,48	5,59	6,50	9,64	4,83	8,17	6,12
21	2,70	4,74	6,70	9,45	4,73	7,79	6,10
22	1,95	3,30	6,07	9,44	5,34	7,37	5,16
23	1,92	3,00	5,16	8,73	3,50	6,56	4,57
24	2,66	4,44	5,91	9,36	4,74	7,48	5,66
25	2,64	3,61	4,80	7,72	3,50	6,35	4,23
26	2,43	3,18	5,79	8,83	4,63	6,53	4,98
27	2,21	4,09	6,77	9,58	4,68	7,78	5,91
28	2,50	4,80	6,43	9,42	4,87	7,61	6,25
29	2,35	4,49	5,76	9,18	4,76	7,43	5,58

	ανταγωνι	τιμέσ_αν	αγορέσ	ποιότητα	ανθρώπιν	πληροφορ	προστασί
30	2,76	3,99	5,45	8,89	4,70	7,00	5,02
31	2,12	5,29	7,34	10,18	5,73	8,40	6,66
32	2,82	3,78	5,80	8,57	4,26	7,04	5,07
33	2,30	4,37	5,81	9,08	5,02	7,28	5,38
34	2,71	2,64	4,41	7,50	2,94	5,80	3,97
35	1,85	5,28	7,06	9,73	5,27	8,21	6,42
36	2,37	4,51	6,23	9,38	5,25	7,71	5,76
37	2,76	4,25	5,74	9,10	6,40	7,25	5,25
38	2,65	3,74	6,20	9,55	5,24	7,33	5,50
39	1,95	4,11	6,24	9,39	5,08	7,42	5,53
40	2,26	3,30	5,52	8,82	4,22	6,73	5,14
41	3,37	4,16	5,94	8,56	3,99	7,09	5,05
42	2,99	3,81	5,75	9,25	6,69	7,20	5,21
43	2,57	3,58	4,90	8,10	3,58	6,60	4,32
44	2,60	4,47	6,59	9,82	5,73	7,87	6,14
45	1,92	3,56	6,14	9,62	5,14	7,61	5,51
46	2,57	4,84	6,64	9,53	5,09	7,81	6,17
47	2,40	2,94	5,17	8,53	4,13	6,57	4,52
48	2,56	4,09	5,38	8,05	3,40	6,43	4,80
49	2,60	4,17	5,93	9,17	4,37	7,07	5,25
50	2,49	5,12	6,68	9,16	4,56	7,75	5,90
51	2,29	3,83	4,96	8,21	3,78	7,02	4,78

	προμηθευ	πελάτεσ	κόστοσ_ε	κόστοσ_υ	διαθεσιμ	καινοτομ	φήμη
1	7,70	9,83	9,12	10,05	11,64	,61	10,55
2	6,97	8,95	8,48	9,18	10,79	,54	9,62
3	7,10	8,79	8,39	9,57	11,18	,73	9,35
4	6,73	8,76	8,16	9,12	10,87	,70	8,45
5	11,38	8,29	12,24	12,94	,00	,88	,00
6	8,49	9,26	9,76	10,42	11,43	,80	9,18
7	7,30	8,76	9,25	9,43	10,76	,78	10,78
8	5,89	7,83	7,35	8,47	10,02	,54	9,57
9	7,09	8,44	8,72	9,48	10,44	,70	10,08
10	7,84	8,01	9,37	9,72	10,68	,74	10,74
11	7,44	9,45	8,79	9,62	11,04	,67	10,26
12	7,93	8,83	9,68	10,18	11,40	,78	10,93
13	5,70	7,19	6,94	8,14	9,81	,67	8,73
14	5,97	7,93	7,35	8,13	10,01	,49	,00
15	7,51	9,47	8,79	9,83	11,53	,56	10,40
16	7,23	9,30	9,05	9,39	10,88	,61	9,57
17	8,22	9,78	9,57	10,30	11,54	,79	10,30
18	6,18	7,99	7,69	8,53	10,37	,58	9,41
19	9,78	9,57	11,23	11,64	11,27	,81	11,69
20	7,68	8,73	9,15	10,01	11,32	,73	8,99
21	7,56	8,80	9,13	9,84	10,90	,75	10,07
22	7,06	9,46	8,46	9,42	11,24	,64	9,92
23	6,00	7,59	6,62	6,51	10,24	,21	9,09
24	7,33	8,36	8,75	9,46	11,05	,57	8,36
25	5,77	7,84	7,53	8,41	9,98	,59	9,29
26	6,56	8,84	8,25	9,08	10,88	,50	9,66
27	7,38	8,54	9,48	9,69	11,23	,82	9,79
28	7,54	9,11	8,95	9,53	11,08	,64	10,81
29	7,01	8,28	8,87	9,59	11,01	,63	9,78

	προμηθευ	πελάτεσ	κόστοσ_ε	κόστοσ_υ	διαθεσιμ	καινοτομ	φήμη
30	6,70	9,14	8,02	9,55	11,07	,49	9,66
31	8,13	9,88	9,67	10,05	11,42	,69	10,91
32	6,82	8,74	8,19	9,19	10,59	,69	10,35
33	7,16	8,63	8,61	9,80	11,06	,67	9,72
34	5,40	7,32	6,82	8,03	9,57	,64	8,81
35	7,90	8,82	9,51	10,12	10,84	,75	10,35
36	7,52	9,43	8,94	10,00	11,41	,65	10,28
37	6,92	8,92	8,51	9,04	10,62	,71	9,44
38	7,06	9,62	8,73	9,40	11,08	,53	10,33
39	6,93	9,04	8,77	9,27	10,71	,66	10,11
40	6,47	8,41	7,97	8,66	10,46	,65	9,06
41	6,75	8,54	8,04	9,17	10,66	,78	7,74
42	6,84	9,50	8,32	9,14	10,90	,65	9,09
43	6,18	7,65	7,85	8,91	10,14	,67	9,72
44	7,52	9,77	9,12	10,00	11,54	,54	10,58
45	7,28	9,20	8,83	9,57	11,24	,66	9,54
46	7,71	9,28	9,10	9,89	11,28	,69	10,77
47	6,35	8,18	7,74	8,68	10,52	,64	8,13
48	6,28	7,35	7,83	8,84	10,32	,68	9,54
49	6,94	8,67	8,82	9,26	10,86	,54	10,64
50	7,43	8,84	8,87	9,76	10,98	,75	8,01
51	6,29	7,14	7,88	9,18	10,01	,64	,00

	υποκατάσ	τιμή_μασ	μεταφορι	νομοθεσί	κανάλια	χρηματοδ	φορολόγη
1	11,24	3,74	5,35	3,25	3,25	,79	2,26
2	10,44	3,81	5,60	3,10	3,10	,80	1,81
3	10,02	3,18	5,07	2,97	2,97	,81	1,98
4	9,97	3,87	5,03	2,89	2,89	,83	2,31
5	14,94	8,81	5,67	3,11	3,11	,78	2,13
6	12,20	4,52	5,50	3,18	3,18	,76	2,28
7	10,57	3,81	5,12	2,99	2,99	,77	2,02
8	,00	2,77	5,32	3,00	3,00	,86	2,44
9	10,56	3,33	5,40	3,15	3,15	,86	2,28
10	11,21	4,09	5,22	3,13	3,13	,77	1,66
11	10,88	3,53	5,37	3,03	3,03	,85	2,04
12	11,05	3,91	5,50	3,08	3,08	,77	2,28
13	,00	2,56	4,42	2,73	2,73	,65	2,40
14	9,23	4,38	5,04	3,05	3,05	,77	2,07
15	10,70	4,23	5,15	2,95	2,95	,73	2,08
16	11,19	4,41	5,58	3,17	3,17	,81	2,30
17	11,79	4,61	5,64	3,11	3,11	,70	1,77
18	,00	3,37	5,12	2,92	2,92	,66	2,32
19	13,53	5,55	5,73	3,16	3,16	,79	2,03
20	11,13	3,47	5,13	2,97	2,97	,82	2,43
21	10,97	4,16	5,47	3,17	3,17	,79	2,10
22	10,32	3,87	5,21	2,97	2,97	,90	1,92
23	9,66	3,43	4,95	2,97	2,97	,85	2,49
24	10,61	5,12	5,39	3,10	3,10	,68	1,80
25	,00	3,58	4,85	3,08	3,08	,77	1,89
26	9,48	3,47	5,21	3,02	3,02	,80	1,93
27	10,95	3,76	5,29	3,15	3,15	,83	2,56
28	10,22	4,13	5,42	3,16	3,16	,75	1,99
29	9,68	3,61	4,79	2,90	2,90	,77	1,81

	υποκατάσ	τιμή_μασ	μεταφορι	νομοθεσί	κανάλια	χρηματοδ	φορολόγη
30	9,65	3,26	5,24	2,97	2,97	,74	1,55
31	11,74	3,91	5,44	3,02	3,02	,90	1,96
32	,00	3,66	5,32	3,06	3,06	,83	1,35
33	10,12	3,89	5,05	2,93	2,93	,90	2,11
34	,00	4,08	4,84	2,94	2,94	,76	1,94
35	11,66	4,32	5,57	3,20	3,20	,89	2,18
36	10,77	4,03	5,23	3,00	3,00	,81	2,01
37	10,53	3,93	5,12	2,68	2,68	,83	2,13
38	10,67	4,01	5,50	3,10	3,10	,77	2,09
39	10,76	4,34	5,55	3,17	3,17	,83	1,99
40	9,62	4,04	5,49	3,01	3,01	,83	2,23
41	10,13	3,85	5,26	3,09	3,09	,66	1,59
42	10,59	3,71	5,41	2,35	2,35	,77	1,73
43	,00	3,99	5,06	2,92	2,92	,78	1,83
44	10,83	3,89	5,17	2,99	2,99	,80	1,87
45	10,80	3,71	5,47	3,04	3,04	,94	2,10
46	10,69	3,66	5,25	3,00	3,00	,81	2,09
47	9,44	3,33	4,98	2,92	2,92	,81	2,48
48	,00	3,04	4,75	2,96	2,96	,82	2,19
49	,00	3,47	5,10	3,00	3,00	,77	1,86
50	11,19	4,03	5,57	3,15	3,15	,79	1,82
51	10,31	4,06	5,17	3,02	3,02	,88	2,37

	διαφήμισ	μέγεθοσ
1	2,64	12,24
2	2,89	11,54
3	3,47	11,50
4	2,94	10,86
5	3,22	15,19
6	3,43	12,56
7	4,09	12,38
8	3,14	10,20
9	3,14	11,50
10	2,56	12,20
11	3,18	11,77
12	3,66	12,38
13	3,09	9,84
14	2,77	10,38
15	2,48	11,91
16	3,40	11,94
17	2,48	12,64
18	3,26	10,31
19	3,43	13,84
20	3,14	11,65
21	3,81	12,01
22	2,56	11,81
23	3,40	10,55
24	2,40	11,59
25	3,00	10,22
26	3,26	11,37
27	4,06	12,24
28	3,37	12,16
29	3,00	11,56

	διαφήμισ	μέγεθοσ
30	2,56	11,22
31	3,22	12,50
32	2,56	11,32
33	2,71	11,30
34	2,77	9,65
35	3,33	12,36
36	2,89	11,88
37	3,47	11,22
38	3,18	11,90
39	2,83	11,59
40	3,22	10,76
41	2,83	11,18
42	2,94	11,14
43	3,09	10,42
44	2,71	12,02
45	3,00	11,61
46	3,22	12,10
47	3,37	10,78
48	3,56	10,68
49	3,43	11,51
50	2,48	11,86
51	3,00	10,40

Πίνακας Durbin - Watson d stastistic : Significance points of d_L and d_U at 0,05 level of significance

	κ' = 1		κ' = 2		κ' = 3		κ' = 4		κ' = 5	
n	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU
6	0,610	1,400								
7	0,700	1,356	0,467	1,896						
8	0,763	1,332	0,559	1,777	0,368	2,287				
9	0,624	1,320	0,629	1,699	0,455	2,126	0,296	2,588		
10	0,879	1,320	0,697	1,641	0,525	2,016	0,376	2,414	0,243	2,822
11	0,927	1,324	0,658	1,604	0,595	1,928	0,444	2,283	0,316	2,645
12	0,971	1,331	0,812	1,579	0,658	1,864	0,512	2,177	0,379	2,506
13	1,010	1,340	0,861	1,562	0,715	1,816	0,574	2,094	0,445	2,390
14	1,045	1,350	0,905	1,551	0,767	1,779	0,632	2,030	0,505	2,296
15	1,077	1,361	0,946	1,543	0,814	1,750	0,685	1,977	0,562	2,220
16	1,106	1,371	0,982	1,539	0,857	1,728	0,734	1,935	0,615	2,157
17	1,133	1,381	1,015	1,536	0,894	1,710	0,779	1,900	0,664	2,104
18	1,158	1,391	1,046	1,535	0,933	1,696	0,820	1,872	0,710	2,060
19	1,160	1,180	1,074	1,536	0,967	1,665	0,859	1,848	0,752	2,023
20	1,201	1,411	1,100	1,537	0,998	1,676	0,894	1,828	0,792	1,991
21	1,221	1,420	1,125	1,538	1,026	1,669	0,927	1,812	0,829	1,964
22	1,239	1,429	1,147	1,541	1,053	1,664	0,958	1,797	0,863	1,940
23	1,257	1,437	1,268	1,543	1,078	1,660	0,986	1,785	0,895	1,920
24	1,273	1,446	1,188	1,546	1,101	1,656	1,013	1,775	0,925	1,902
25	1,288	1,454	1,206	1,550	1,123	1,654	1,038	1,767	0,953	1,886
26	1,302	1,461	1,224	1,553	1,143	1,652	1,062	1,759	0,979	1,873
27	1,316	1,469	1,240	1,556	1,162	1,651	1,084	1,753	1,004	1,861
28	1,328	1,476	1,255	1,560	1,181	1,650	1,104	1,747	1,028	1,850
29	1,341	1,483	1,270	1,563	1,198	1,650	1,124	1,743	1,050	1,841
30	1,352	1,489	1,284	1,567	1,214	1,650	1,143	1,739	1,071	1,833
31	1,363	1,496	1,297	1,570	1,229	1,650	1,160	1,735	1,090	1,825
32	1,373	1,502	1,309	1,574	1,244	1,650	1,177	1,732	1,109	1,519
33	1,383	1,508	1,321	1,577	1,258	1,651	1,193	1,730	1,127	1,813
34	1,393	1,514	1,333	1,580	1,271	1,652	1,208	1,728	1,144	1,808
35	1,402	1,519	1,343	1,584	1,283	1,653	1,222	1,726	1,160	1,803
36	1,411	1,525	1,354	1,587	1,295	1,654	1,236	1,724	1,175	1,799
37	1,419	1,530	1,364	1,590	1,307	1,655	1,249	1,723	1,190	1,795
38	1,427	1,535	1,373	1,594	1,318	1,656	1,261	1,722	1,204	1,792
39	1,435	1,540	1,382	1,597	1,328	1,658	1,273	1,722	1,218	1,789
40	1,442	1,544	1,391	1,600	1,338	1,659	1,285	1,721	1,230	1,786
45	1,475	1,566	1,430	1,615	1,383	1,666	1,336	1,720	1,287	1,776
50	1,503	1,585	1,462	1,628	1,421	1,674	1,378	1,721	1,335	1,771
55	1,528	1,601	1,490	1,641	1,452	1,681	1,414	1,724	1,374	1,766
60	1,549	1,616	1,514	1,652	1,480	1,689	1,444	1,727	1,408	1,767
65	1,567	1,629	1,536	1,662	1,503	1,696	1,741	1,731	1,438	1,767
70	1,583	1,641	1,554	1,672	1,525	1,703	1,494	1,735	1,464	1,768
75	1,598	1,652	1,571	1,680	1,543	1,709	1,515	1,739	1,487	1,770
80	1,611	1,662	1,586	1,688	1,560	1,715	1,534	1,743	1,507	1,772
85	1,624	1,671	1,600	1,696	1,575	1,721	1,550	1,747	1,525	1,774
90	1,635	1,679	1,612	1,703	1,589	1,726	1,566	1,751	1,542	1,776
95	1,645	1,687	1,623	1,709	1,602	1,732	1,579	1,755	1,557	1,778
100	1,654	1,694	1,634	1,715	1,613	1,736	1,592	1,758	1,571	1,780
150	1,720	1,746	1,706	1,760	1,693	1,774	1,679	1,788	1,665	1,802
200	1,758	1,788	1,748	1,789	1,738	1,799	1,728	1,810	1,718	1,820

ΠΗΓΗ : Damodar Gujarati (1999), « Essentials of Econometrics », 2nd Edition McGraw Hill.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Dawning & Clark (1998), « Στατιστική των Επιχειρήσεων », 3η Αμερικανική Έκδοση, Αθήνα : Εκδόσεις Barron's Κλειδάριθμος.
2. Κοσμάς Μιλτ. Μαρκάτος , « Στατιστική Επιχειρήσεων », Εκδόσεις ''Ελληνικός Λόγος''.
3. Τάκης Παπαϊωάννου – Σωτήρης Λουκάς(1990), « Εισαγωγή στη Στατιστική », Ιωάννινα.
4. Δημήτρης Α. Ιωαννίδης (1999), « Στατιστικές Μέθοδοι », Τόμος πρώτος, Θεσσαλονίκη : Εκδόσεις Ζήτη.
5. Ιωάννης Γ. Χαλικιάς (2001), « Στατιστική – Μέθοδοι για Επιχειρηματικές Αποφάσεις », Αθήνα : Εκδόσεις Rosili.
6. Ιωάννης Αλκιβ. Κασκαρέλης (1999), « Ένδεκα Μαθήματα Οικονομετρίας », 2η Αναθεωρημένη Έκδοση, Αθήνα : Εκδόσεις Gutenberg.
7. Γ.Κ. Χρήστου (2000), « Εισαγωγή στην Οικονομετρία », Τόμος πρώτος, Αθήνα : Εκδόσεις Gutenberg.
8. Γ. Δρόσος – Δ. Καραπιστόλας (1994), « Στατιστική Επιχειρήσεων », Αθήνα : Εκδόσεις Ελλήν.
9. Δ.Π. Ψωϊνός (1999), « Στατιστική », Θεσσαλονίκη : Εκδόσεις Ζήτη.
10. Ανδρέας Δ. Κιντής (1999), « Στατιστικές και Οικονομετρικές Μέθοδοι », Αθήνα : Εκδόσεις Gutenberg.
11. Πέτρος Α. Κιόχος (1993), « Στατιστική », Αθήνα : Εκδόσεις Interbooks.

12. Damodar Gujarati (1999), « Essentials of Econometrics », 2nd Edition McGraw Hill.

13. Jeffrey Jarrett (2002), « Μέθοδοι Προβλέψεων », Αθήνα : Εκδόσεις Gutenberg.

14. Πρόδρομος Γ. Ευθύμογλου (1990), « Επιχειρησιακή Στρατηγική », Τεύχος Α Θεσσαλονίκη : Φωτ/θεσία, μοντάζ, Σοφία Σπυρίδου, εκτύπωση Typooffset.