

ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΔΥΤΙΚΗΣ ΕΛΛΑΔΑΣ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ
ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**ΑΝΑΛΥΣΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ &
ΟΙΚΟΝΟΜΙΚΕΣ ΕΦΑΡΜΟΓΕΣ**

ΕΚΠΟΝΗΣΗ:

ΓΕΩΡΓΟΠΟΥΛΟΥ ΑΘΑΝΑΣΙΑ

ΚΑΛΥΒΙΤΗ ΠΑΥΛΙΝΑ

ΜΙΕΤΣΙΚΑ ΟΛΓΑ

ΕΠΟΠΤΕΥΩΝ ΚΑΘΗΓΗΤΗΣ

ΦΩΤΕΙΝΟΠΟΥΛΟΣ ΜΙΧΑΛΗΣ

ΠΑΤΡΑ - 2016

Περίληψη

Κύριος σκοπός αυτής της εργασίας είναι η θεωρητική και εμπειρική προσέγγιση της ανάλυσης με την χρήση παλινδρομήσεων. Τα διάφορα δεδομένα μπορούν να αναλυθούν με την χρήση οικονομετρικών υποδειγμάτων. Τα υποδείγματα μπορούν να διαφέρουν ως προς τον αριθμό των ανεξάρτητων μεταβλητών, τον τρόπο που εισάγονται στο υπόδειγμα κλπ.

Στην αρχή της εργασίας γίνεται αναλυτική θεωρητική παρουσίαση της απλής παλινδρόμησης. Στο πρώτο κομμάτι αυτής της ανάλυσης γίνεται πλήρης ανάλυση του απλού γραμμικού μοντέλου παλινδρόμησης καθώς και των υποθέσεων που το διέπουν. Στο τέλος του κεφαλαίου παρουσιάζεται η πρακτική εφαρμογή μιας απλής παλινδρόμησης με τη βοήθεια του οικονομετρικού πακέτου Eviews και αναλύονται τα αποτελέσματα της.

Στην συνέχεια της εργασίας ακολουθεί η ανάλυση της πολλαπλής παλινδρόμησης. Όπως και προηγουμένως στην περίπτωση της απλής παλινδρόμησης γίνεται εκτενής ανάλυση του θεωρητικού υπόβαθρου της πολλαπλής παλινδρόμησης, των ελέγχων που μπορούμε να εξάγουμε και του τρόπου που ερμηνεύονται τα αποτελέσματα των ελέγχων. Και εδώ στο τέλος του κεφαλαίου γίνεται μια ανάλυση ενός μοντέλου πολλαπλής παλινδρόμησης μέσω της χρήσης του οικονομετρικού πακέτου Eviews και ανάλυση των αποτελεσμάτων που μας δίνει καθώς και ερμηνεία των ελέγχων που εξάγονται.

Τέλος, ακολουθούν τα συμπεράσματα που εξάγονται και η βιβλιογραφία που χρησιμοποιήθηκε για την εκπόνηση της εργασίας.

Summary

The main purpose of this final paper is to analyze various economic data by using regressions.

Various data sets can be analyzed through econometric models. Models vary in the number of independent variables and the way they are included in the analysis.

At the beginning of this final paper a detailed theoretical presentation of simple regression is presented and the basic assumptions of it are highlighted. At the end of the chapter we present a practical application of a simple regression through the Eviews econometric program and we analyze the results that are carried out.

Then we continue with the analysis of the multiple regression. One more time as in the case of the simple regression we analyze extensively the theoretical background of multiple regression, the test hypotheses that can be drawn and how can we interpret the results of the tests.

At the end of the chapter an analysis of a multiple regression model is carried out through the use of Eviews econometric program and we analyze the results.

Finally, the conclusions and the literature used for the preparation of this work are presented.

Περιεχόμενα

Περίληψη	2
Summary	3
Εισαγωγή	8
ΚΕΦΑΛΑΙΟ 1 : Απλή Γραμμική Παλινδρόμηση	11
1.1 Η Ευθεία Παλινδρόμησης.....	12
1.2 Προσδιορισμός του υποδείγματος	14
1.3 Υποθέσεις του απλού γραμμικού υποδείγματος παλινδρόμησης	15
1.4 Βήματα Απλής Γραμμικής Παλινδρόμησης	17
ΚΕΦΑΛΑΙΟ 2 Πολλαπλή Γραμμική Παλινδρόμηση	25
2.1 Το Στατιστικό Μοντέλο Πολλαπλής Παλινδρόμησης.....	25
2.2. Υποθέσεις Πολλαπλής Γραμμικής Παλινδρόμησης.....	26
2.3 Βήματα της Πολλαπλής Γραμμικής Παλινδρόμησης.....	27
ΚΕΦΑΛΑΙΟ 3: Βασικές αρχές χρήσης Eviews	33
ΚΕΦΑΛΑΙΟ 4: Εφαρμογές	37
4.1 Κατανάλωση νοικοκυριού	37
4.2 Σωματικά χαρακτηριστικά φοιτητών.....	42

4.3 Αποδόσεις τραπεζικών μετοχών	46
4.4 Σωματικό λίπος	54
4.4.1 Σύγκριση αποτελεσμάτων απλής και πολλαπλής παλινδρόμησης	60
Συμπεράσματα	63
Βιβλιογραφία	64

Εισαγωγή

Η οικονομετρία έχει ως αντικείμενο την εξέταση και εκτίμηση διαφόρων οικονομικών σχέσεων μέσω της εξέτασης διαφόρων μεταβλητών. Απαραίτητη ωστόσο προϋπόθεση είναι η γνώση της στατιστικής θεωρίας. Με τη χρήση του κατάλληλου υποδείγματος αλλά και τα απαραίτητα δεδομένα είναι δυνατόν να ελεγχθεί η σχέση με την οποία συνδέονται διάφορες μεταβλητές.

Πέρα από την εκτίμηση μοντέλων για την εκτίμηση σχέσεων η οικονομετρία μπορεί να χρησιμοποιηθεί και για την διενέργεια προβλέψεων. Οι προβλέψεις αποτελούν πολύ σημαντικό κομμάτι της οικονομετρικής ανάλυσης καθώς μπορούν να μας δώσουν μια εικόνα για το πώς αναμένεται να κινηθούν διάφορα οικονομικά μεγέθη στο μέλλον. Για παράδειγμα, οι αποδόσεις του χρηματιστηριακού δείκτη τιμών, οι τιμές των πρώτων υλών κλπ.

Τα μοντέλα που είναι διαθέσιμα στον ερευνητή για να επιλέξει είναι πάρα πολλά και καλύπτουν το σύνολο των απαιτήσεων. Έτσι η πιο απλή μορφή που συναντάται η απλή γραμμική παλινδρόμηση στην οποία μέσω μιας ανεξάρτητης μεταβλητής μελετούμε την εξαρτημένη μεταβλητή ενώ επίσης υπάρχει και η δυνατότητα πολλαπλής γραμμικής παλινδρόμησης με τη χρήση πολλών ερμηνευτικών μεταβλητών.

Τα στάδια τα οποία ακολουθεί η οικονομετρική ανάλυση είναι τα εξής.

Αρχικά, γίνεται η επιλογή του υποδείγματος, δηλαδή γίνεται η επιλογή του μοντέλου καθώς και το ποιες μεταβλητές θα περιέχονται σε αυτό, ποια θα είναι η εξαρτημένη και ποιες οι ανεξάρτητες καθώς και η μαθηματική του διατύπωση. Στη συνέχεια, επιλέγεται με ποιο τρόπο θα γίνει η εκτίμηση του υποδείγματος, δηλαδή η εκτίμηση των συντελεστών των επιλεγμένων ερμηνευτικών μεταβλητών. Συνήθως επιλέγεται η μέθοδος των ελαχίστων τετραγώνων.

Τέλος μετά την εκτίμηση του υποδείγματος ακολουθούν οι διάφοροι έλεγχοι που μπορούν να πραγματοποιηθούν προκειμένου να ελεγχθεί η αξιοπιστία του υποδείγματος και των αποτελεσμάτων που παράγει προκειμένου ο ερευνητής να είναι σε θέση να γνωρίζει αν τα ληφθέντα αποτελέσματα παρουσιάζουν με σωστό τρόπο την σχέση μεταξύ των ανεξάρτητων μεταβλητών και της εξαρτημένης του εκτιμημένου υποδείγματος.

Περιεχόμενα Εικόνων

Εικόνα1: Τα στάδια της Ανάλυσης Παλινδρόμησης.....	12
Εικόνα 2 Διάγραμμα διασποράς με ευθεία παλινδρόμησης.....	15
Εικόνα 3 Διάγραμμα διασποράς.....	16
Εικόνα 4 Αποστάσεις από την ευθεία παλινδρόμησης.....	18
Εικόνα 5 Δημιουργία αρχείου εργασίας.....	33
Εικόνα 6 Δημιουργία αρχείου εργασίας, τυχαίες παρατηρήσεις.....	34
Εικόνα 7 Κεντρική οθόνη Eviews.....	35
Εικόνα 8 Εκτίμηση συνάρτησης.....	36
Εικόνα 9 Διάγραμμα διασποράς.....	38
Διάγραμμα 10 Διάγραμμα Διασποράς.....	42
Εικόνα 15 Εισαγωγή δεδομένων.....	47
Εικόνα 16 Εκτίμηση εξίσωσης παλινδρόμησης.....	48

Περιεχόμενα Πινάκων

Πίνακας 6 Πίνακας συσχετίσεων.....	32
Πίνακας 1 Πίνακας Δεδομένων	38
Πίνακας 2 Αποτελέσματα παλινδρόμησης	40
Πίνακας 3 Περιγραφικά στοιχεία μεταβλητών	43
Πίνακας 4 Πίνακας Συσχετίσεων.....	44
Πίνακας 5 Αποτελέσματα παλινδρόμησης	44
Πίνακας 7 Αποτελέσματα παλινδρόμησης	49
Πίνακας 8 Περιγραφικά μέτρα μεταβλητών.....	52
Πίνακας 9 Συσχετίσεις.....	53
Πίνακας 10 Περιγραφικά μέτρα μεταβλητών.....	54
Πίνακας 11 Πίνακας συσχετίσεων.....	55
Πίνακας 12 Αποτελέσματα παλινδρόμησης	556
Πίνακας 13 Αποτελέσματαπαλινδρόμησης	557
Πίνακας 14 Αποτελέσματα παλινδρόμησης	599

ΚΕΦΑΛΑΙΟ 1 : Απλή Γραμμική Παλινδρόμηση

Στην απλή γραμμική παλινδρόμηση έχουμε να κάνουμε με μια μέθοδο η οποία χρησιμοποιείται προκειμένου να μας φανερώσει μέσω μιας γραμμικής σχέσης τη σχέση που συνδέει μια εξαρτημένη μεταβλητή Y και μια ανεξάρτητη μεταβλητή X , οι οποίες θεωρούμε ότι συνδέονται μέσω μιας σχέσης. Αυτό μπορούμε να το υποθέσουμε είτε με βάση την οικονομική θεωρία είτε γιατί θέλουμε να δούμε αν υπάρχει αυτή η σχέση. Σε κάθε εκτιμώμενο μοντέλο γραμμικής παλινδρόμησης διακρίνουμε αυτά τα δύο είδη μεταβλητών: τις ανεξάρτητες (independent variables) και τις εξαρτημένες μεταβλητές (dependent variables). Ως ανεξάρτητη μεταβλητή X ονομάζεται αυτή η οποία ελέγχεται από τον ερευνητή όπως για παράδειγμα το ύψος της δαπάνης για διατροφή ενός ατόμου, οι τιμές ενός προϊόντος, ο αριθμός των εργαζομένων σε μια επιχείρηση, η ποσότητα φαρμάκου που χορηγείται σε έναν ασθενή, η θερμοκρασία επεξεργασίας ενός προϊόντος κλπ.

Εξαρτημένη μεταβλητή Y ονομάζεται εκείνη η μεταβλητή στην οποία απεικονίζονται οι μεταβολές στην επιλεγμένη ανεξάρτητη μεταβλητή (π.χ. η ζήτηση ενός προϊόντος, ο χρόνος αναμονής των πελατών ενός καταστήματος, οι μέρες νοσηλείας ενός ασθενούς, η αντοχή ενός υλικού).

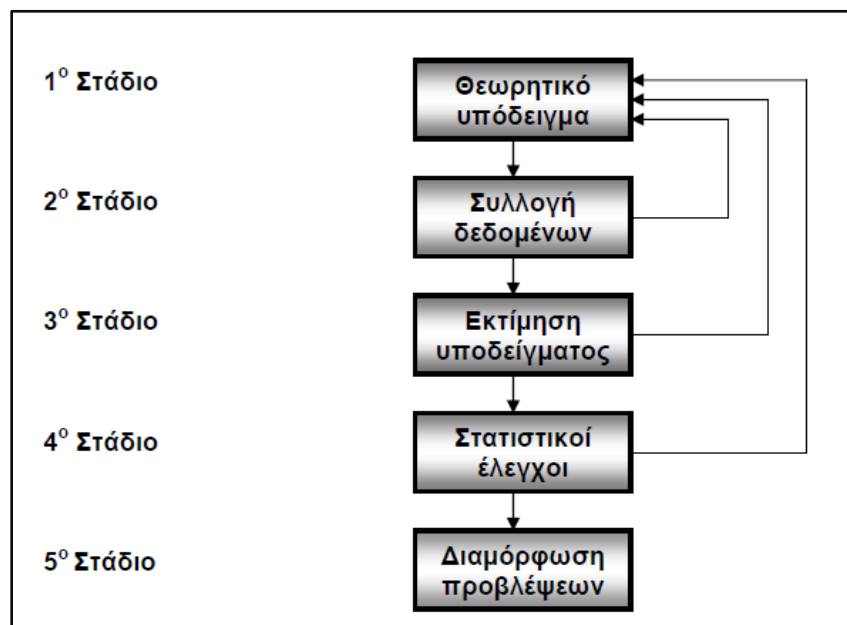
Στις επόμενες ενότητες γίνεται αναλυτική περιγραφή του τρόπου με τον οποίο βρίσκεται η ζητούμενη γραμμική σχέση μεταξύ των μεταξύ ανεξάρτητης και εξαρτημένης μεταβλητής και πώς ερμηνεύονται τα αποτελέσματα καθώς και πότε αυτά είναι αξιοποιήσιμα και έχουν στατιστική αξία.

1.1 Η Ευθεία Παλινδρόμησης

Η ανάλυση παλινδρόμησης εστιάζεται στην ταυτόχρονη μελέτη και ερμηνεία δύο ή περισσότερων μεταβλητών για να βρεθεί πως αυτές οι μεταβλητές σχετίζονται. Μέσω αυτής της διαδικασίας προσδιορίζεται ποσοτικά η σχέση αυτή. Η ανάλυση παλινδρόμησης βασίζεται σε στατιστικές μεθόδους, οι οποίες χρησιμοποιούνται για την εκτίμηση των σχέσεων μεταξύ των μεταβλητών που εξετάζονται. Αναλυτικότερα, με την ανάλυση παλινδρόμησης προσδιορίζεται η ποσοτική σχέση μεταξύ της εξαρτημένης και μιας ή περισσότερων ανεξάρτητων μεταβλητών με σκοπό την πρόβλεψη των τιμών της εξαρτημένης μεταβλητής με βάση συγκεκριμένες τιμές των ανεξάρτητων μεταβλητών¹.

Μπορούμε να παρουσιάσουμε πιο αναλυτικά τη παλινδρόμηση στα παρακάτω πέντε στάδια.

Τα στάδια αυτά αναφέρονται στο παρακάτω σχήμα με την σειρά με την οποία εφαρμόζονται και μπορούν να κάνουν πιο κατανοητό τον τρόπο με τον οποίο εφαρμόζεται η μεθοδολογία της ανάλυσης με την χρήση παλινδρόμησης².



Εικόνα1: Τα στάδια της Ανάλυσης Παλινδρόμησης

¹Γ. Παπαδόπουλος, Ανάλυση Παλινδρόμησης, Εργαστήριο Μαθηματικών και Στατιστικής. Πηγή: www.aua.gr/gpapadopoulos

²Αγιακλόγλου Χ.Ν. και Οικονόμου Γ.Σ. (2002), Μέθοδοι προβλέψεων και Ανάλυσης αποφάσεων, εκδ. Γ. Μπένου, Αθήνα.

Αρχικά κατασκευάζουμε το θεωρητικό υπόδειγμα στο οποίο περιλαμβάνονται ανεξάρτητες μεταβλητές μέσω των οποίων γίνεται η προσπάθεια για την ερμηνεία της εξαρτημένης.

Ακολούθως συλλέγονται τα κατάλληλα δεδομένα για την ανάλυση της παλινδρόμησης. Στο τρίτο στάδιο ο ερευνητής μπορεί να προβεί στην εκτίμηση του υποδείγματος, δηλαδή να προσδιοριστούν οι συντελεστές του σταθερού όρου και του ή των ανεξάρτητων μεταβλητών.

Στη συνέχεια σε αυτό που στο διάγραμμα παρουσιάζεται ως τέταρτο στάδιο εφαρμόζονται διάφοροι στατιστικοί έλεγχοι έτσι ώστε να ελεγχθεί το εκτιμημένο υπόδειγμα που διατυπώθηκε αρχικά. Σε περίπτωση που προκύψει κάποιο πρόβλημα από τους στατιστικούς ελέγχους όσον αφορά στην στατιστική σημαντικότητα κάποιου ή κάποιων από τους συντελεστές είναι δυνατόν να διορθωθεί η μορφή του υποδείγματος με κάποια άλλη που ο ερευνητής θεωρεί πιο κατάλληλη.

Τέλος, στο τελευταίο στάδιο μπορεί να χρησιμοποιηθεί το εκτιμημένο υπόδειγμα για την πρόβλεψη τιμών της εξαρτημένης μεταβλητής εάν κάτι τέτοιο απαιτείται.

1.2 Προσδιορισμός του υποδείγματος

Η μαθηματική μορφή που λαμβάνει το υπόδειγμα της απλής γραμμικής παλινδρόμησης για τις μεταβλητές X και Y είναι η ακόλουθη :

$$Y = \alpha + \beta X + \varepsilon$$

Όπου :

Y = η εξαρτημένη μεταβλητή

X = η ανεξάρτητη μεταβλητή

α = ο σταθερός όρος

β = η κλίση του γραμμικού υποδείγματος

ε = το τυχαίο σφάλμα

Τα α και β είναι ονομάζονται συντελεστές της παλινδρόμησης (regression coefficients), και τις τιμές αυτών ο ερευνητής προσπαθεί να προσδιορίσει. Το τμήμα $\alpha + \beta X$ καλείται προσδιοριστικό μέρος της εξίσωσης της απλής γραμμικής παλινδρόμησης και είναι τα κατάλοιπα ή τυχαίο σφάλμα. Το α , δηλαδή ο σταθερός όρος, είναι το σημείο στο οποίο τέμνει τον άξονα Y η ευθεία που δημιουργούμε μεταξύ δυο σημείων που ορίζονται από τις τιμές των παρατηρήσεων των δυο μεταβλητών. Το β όπως αναφέρθηκε και προηγουμένως είναι η κλίση της ευθείας ή αλλιώς το ποσό κατά το οποίο μεταβάλλεται η Y όταν η X μεταβάλλεται κατά μια μονάδα. Τέλος το ε αφορά τα κατάλοιπα δηλαδή τις αποκλίσεις από την ευθεία που δημιουργείται από τα συγκεκριμένα α και β . Όσο μικρότερη είναι η διασπορά των καταλοίπων τόσο καλύτερα μπορούμε να πούμε ότι προσαρμόζεται η εξίσωση γραμμικής παλινδρόμησης στα δεδομένα.

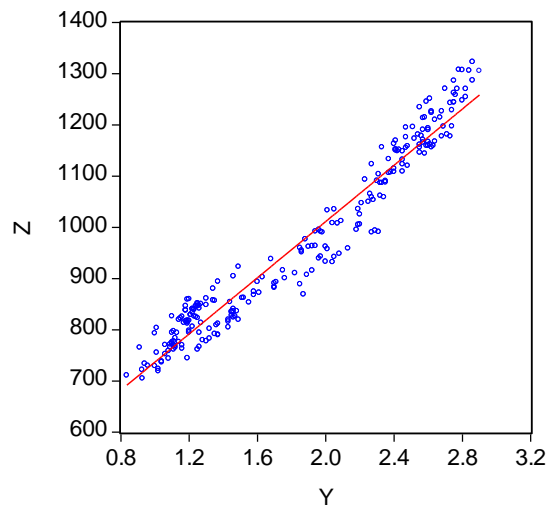
1.3 Υποθέσεις του γραμμικού υποδείγματος παλινδρόμησης

απλού

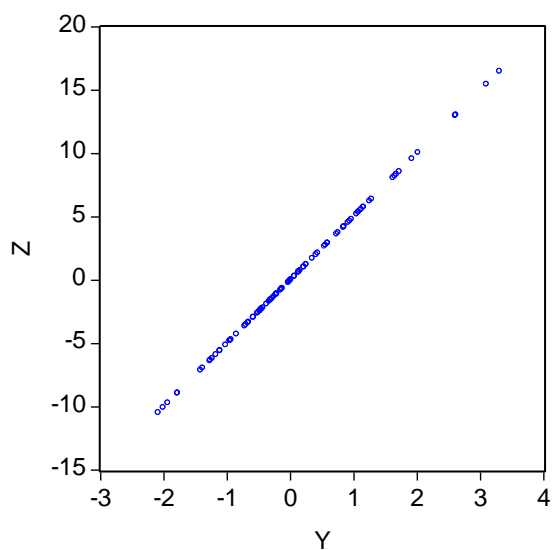
Το απλό γραμμικό υπόδειγμα παλινδρόμησης διέπεται από κάποιες υποθέσεις οι οποίες είναι οι εξής:

1. Η συναρτησιακή μορφή του υποδείγματος είναι γραμμική.

Αυτό σημαίνει ότι προσαρμόζουμε μια ευθεία στα δεδομένα χωρίς αυτό να σημαίνει ότι τα δεδομένα βρίσκονται επί της ευθείας. Στην εικόνα 2 που ακολουθεί απεικονίζεται ένα διάγραμμα από ένα τυχαίο δείγμα μεταβλητών οι οποίες δημιουργήθηκαν με βάση μια γραμμική σχέση. Τα κατάλοιπα αποτελούν τις αποστάσεις από τις μπλέ κουκκίδες στην ευθεία που φαίνεται στο διάγραμμα της εικόνας. Στο διάγραμμα της εικόνας 3 είναι ότι τα κατάλοιπα είναι μηδενικά.



Εικόνα 2 Διάγραμμα διασποράς με ευθεία παλινδρόμησης



Εικόνα 3 Διάγραμμα διασποράς

2. Ο μέσος του όρου σφάλματος είναι μηδέν: $E(\varepsilon) = 0$. Αυτό σημαίνει ότι η τυχαία μεταβλητή μπορεί να παίρνει αρνητικές αλλά και θετικές τιμές όμως η μέση της τιμή της είναι μηδέν.

3. Η διακύμανση του σφάλματος είναι σταθερή. $Var(\varepsilon) = \sigma^2$. Αυτή η υπόθεση μας φανερώνει ότι η διασπορά των τιμών του σφάλματος δεν μεταβάλλεται όταν αλλάζει η τιμή της ανεξάρτητης μεταβλητής X .

4. Η συνδιακύμανση μεταξύ των σφαλμάτων είναι μηδέν: $Cov(\varepsilon_i, \varepsilon_j) = 0$. Σε περίπτωση που παραβιάζεται αυτή η υπόθεση δημιουργείται πρόβλημα της αυτοσυσχέτισης.

5. Η συνδιακύμανση των καταλοίπων και της ανεξάρτητης μεταβλητής είναι πάντα μηδέν: $Cov(\varepsilon, X) = 0$. Η υπόθεση αυτή μας τονίζει η ανεξάρτητη μεταβλητή X δεν είναι τυχαία.

6. Οι όροι σφάλματος, ανεξάρτητοι μεταξύ τους, κατανομονται σύμφωνα με την κανονική κατανομή με μέσο 0 και διακύμανση σ^2 .

1.4 Βήματα Απλής Γραμμικής Παλινδρόμησης

Προκειμένου να εκτιμήσουμε μια αξίωση απλής γραμμικής παλινδρόμησης ακολουθούνται τα παρακάτω βήματα:

1) Ορισμός της ανεξάρτητης και την εξαρτημένης μεταβλητής

Όπως αναφέρθηκε και προηγουμένως ανεξάρτητη μεταβλητή X ονομάζεται αυτή η οποία ελέγχεται από τον ερευνητή όπως για παράδειγμα το ύψος της δαπάνης για διατροφή ενός ατόμου, οι τιμές ενός προϊόντος, κλπ. Αντίθετα εξαρτημένη μεταβλητή Y ονομάζεται εκείνη η μεταβλητή στην οποία απεικονίζονται οι μεταβολές στην επιλεγμένη ανεξάρτητη μεταβλητή όπως για παράδειγμα (π.χ. η ζήτηση ενός προϊόντος, ο χρόνος αναμονής των πελατών ενός καταστήματος κ.λ.π.)

2) Προσδιορισμός του συντελεστή συσχέτισης μεταξύ εξαρτημένης και ανεξάρτητης μεταβλητής

Όταν υποπτευόμαστε ότι ίσως μεταξύ δυο μεταβλητών υπάρχει κάποια σχέση είτε με βάση την οικονομική θεωρία είτε γιατί εμπειρικά πιστεύουμε ότι υπάρχει αυτή η σχέση υπολογίζουμε το συντελεστή συσχέτισης γίνεται σύμφωνα με τον τύπο του Pearson που αναφέρθηκε και προηγουμένως.

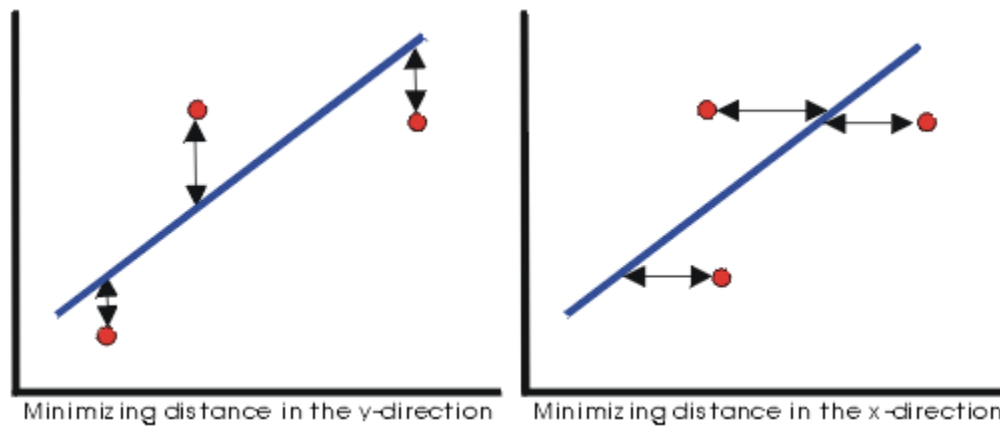
3) Έλεγχος στατιστικής σημαντικότητας συντελεστή συσχέτισης.

Σε αυτό το σημείο γίνεται έλεγχος του συντελεστή συσχέτισης προκειμένου να εξακριβωθεί αν έχει στατιστική αξία ή ισούται με το μηδέν σύμφωνα με τον έλεγχο που αναφέρθηκε σε προηγούμενη ενότητα.

4) Υπολογισμός ευθείας παλινδρόμησης με την μέθοδο των ελαχίστων τετραγώνων.

Σκοπός, όπως έχει ήδη αναφερθεί, είναι να εκτιμηθούν οι παράμετροι υποδείγματος δηλαδή ο συντελεστής $\hat{\alpha}$ του σταθερού όρου και ο συντελεστής $\hat{\beta}$ της ανεξάρτητης

μεταβλητής. Μέσω της μεθόδου των ελαχίστων τετραγώνων δημιουργείται μια ευθεία γραμμή οι οποία έχει την ιδιότητα να ελαχιστοποιεί το τετράγωνο της απόστασης ενός συνόλου σημείων από την γραμμή αυτή. Αυτή η ευθεία παλινδρόμησης πρέπει να πληροί την προϋπόθεση ότι θα πρέπει να περνά κοντά από τα σημεία που προκύπτουν από τα ζεύγη παρατηρήσεων X και Y με τέτοιο τρόπο ώστε τα σφάλματα να ελαχιστοποιούνται.³



Εικόνα 4 Αποστάσεις από την ευθεία παλινδρόμησης

Με τη μέθοδο ελαχίστων τετραγώνων ο συντελεστής του σταθερού όρου υπολογίζεται από τον τύπο:

$$\hat{a} = \bar{Y} - \hat{\beta}\bar{X}$$

Ενώ ο συντελεστή της κλίσης της ευθείας από τον εξής τύπο:

$$\hat{\beta} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

³Βενέτης Α. Ιωάννης (2009), Εισαγωγικές διαλέξεις στην Οικονομετρία.

5) Έλεγχος στατιστικής σημαντικότητας του συντελεστή β

Σε όλη την διαδικασία της απλής γραμμικής παλινδρόμησης αυτό που δίνει αξία στα αποτελέσματα που λαμβάνονται είναι το είναι στατιστικά σημαντικά. Το ίδιο ισχύει και για το συντελεστή της κλίσης της ευθείας. Αν ο συντελεστής είναι ίσος με το μηδέν (δηλαδή δεν έχει στατιστική αξία) τότε δεν έχει νόημα η ανάλυση. Ο έλεγχος του β γίνεται ως εξής: Αρχικά διατυπώνουμε την αρχική υπόθεση

$$H_0: \beta=0 \text{ έναντι της εναλλακτικής}$$

$$H_1: \beta \neq 0$$

Στην περίπτωση που απορριφθεί η αρχική υπόθεση (δηλαδή ο συντελεστής β δεν είναι μηδέν) τότε έχει νόημα να συνεχίσουμε την ανάλυση των δεδομένων μας ανεξάρτητα από το πρόσημο του.

6) Έλεγχος στατιστικής σημαντικότητας του μοντέλου συνολικά.

Όπως αναφέρθηκε και στο προηγούμενο έλεγχο η στατιστική σημαντικότητα του συντελεστή της κλίσης της ευθείας είναι δείγμα ότι η ανάλυση μας έχει νόημα. Ωστόσο το ίδιο συμβαίνει και με τον έλεγχο του μοντέλου συνολικά. Ο έλεγχος που μας δείχνει το πόσο έγκυρο είναι ένα μοντέλο είναι ο έλεγχος F. Προκειμένου να γίνει ο παραπάνω έλεγχος πρέπει να δημιουργηθεί ο πίνακας που προκύπτει από την ανάλυση της διακύμανσης (ANOVA). Όπως και στον έλεγχο για ένα μεμονωμένο συντελεστή έτσι και εδώ η μηδενική $H_0: \beta=0$ με εναλλακτική $H_1: \beta \neq 0$ ακολουθεί την $F(1, n-2)$ με $n-2$ βαθμούς ελευθερίας. Στη περίπτωση της απλής γραμμικής παλινδρόμησης ο έλεγχος F και ο έλεγχος για την σημαντικότητα του β είναι ισοδύναμοι.

Πίνακας ANOVA ⁴				
	B.E.	Άθροισμα τετραγώνων	Μέσο άθροισμα τετραγώνων	Στοιχείο Ελέγχου
Παλινδρόμηση	1	SSR	MSR	F=MSR/MSE
Κατάλοιπα	n-2	SSE	MSE	
Σύνολο	n-1	Μεταβλητότητα y		

7) Προσδιορισμός του συντελεστή προσδιορισμού R^2

Ο συντελεστής προσδιορισμού R^2 (coefficient of determination) μετρά το ποσοστό της διακύμανσης της μεταβλητής Y που εξηγείται από την μεταβλητή X και δίνεται από τον τύπο: $R^2 = \frac{SSR}{SST} 0 \leq R^2 \leq 1$ όπου SSE το άθροισμα των «εξηγημένων» τετραγώνων και SST το συνολικό άθροισμα τετραγώνων της εξαρτημένης μεταβλητής.

Ο συντελεστής προσδιορισμού είναι ένα μέτρο προσαρμογής του υποδείγματος στα δεδομένα. Δεν έχει μονάδα μέτρησης και το εύρος των τιμών του είναι μεταξύ 0 και 1. Το 0 σημαίνει η μεταβλητή X εξηγεί το 0% της μεταβλητότητας της εξαρτημένης μεταβλητής Y . Αυτό κατ' επέκταση σημαίνει ότι η εν λόγω μεταβλητή δεν είναι χρήσιμη για το υπόδειγμα καθώς δεν μας δίνει καθόλου πληροφόρηση για το πως μεταβάλλονται οι τιμές της Y . Αντίθετα αν η τιμή του συντελεστή προσδιορισμού R^2 είναι 1 αυτό σημαίνει ότι η μεταβλητή X ερμηνεύει το σύνολο της μεταβλητότητας (100%) της εξαρτημένης μεταβλητής Y και άρα μπορούμε να πούμε πως αρκεί η εξέταση μόνο της μεταβλητής X για να καταλάβουμε πως μεταβάλλεται η μεταβλητή Y . Γενικότερα καθώς η τιμή του R^2 απομακρύνεται από το 0 προς την τιμή 1 θεωρείται ότι η προσαρμογή είναι ολοένα και καλύτερη.

Ωστόσο, δεν είναι σωστό να είμαστε ιδιαίτερα αυστηροί στην κρίση μας ιδιαίτερα όταν πρόκειται για μικροοικονομικές μεταβλητές π.χ. ωρομίσθια και έτη εκπαίδευσης ή ωρομίσθια και ηλικία. Σε τέτοια δεδομένα αν λάβουμε για παράδειγμα $R^2=0,14$ σε μια

⁴ Μπαρδάκη Θ. – Λακουμέντα Ι. (2015), Πολλαπλή Παλινδρόμηση και Οικονομικές Εφαρμογές, Τ.Ε.Ι. Δυτικής Ελλάδος

παλινδρόμηση της μορφής $Y_i = \alpha + \beta X_i + u_i$, όπου X_i η ηλικία και Y_i το ωρομίσθιο τότε μια λανθασμένη ερμηνεία θα έλεγε ότι το 0,14 δείχνει ότι το υπόδειγμα δεν έχει καλή προσαρμογή. Μια άλλη ερμηνεία είναι ότι η ηλικία ερμηνεύει μόνο το 14% της μεταβλητότητας του ωρομισθίου. Αν αντίθετα το R^2 στο παραπάνω παράδειγμα ήταν 0,99 δηλαδή αρκούσε η ηλικία για να ερμηνευτεί το 99% της μεταβλητότητας των ωρομισθίων ή αλλιώς τα ωρομίσθια καθορίζονται κατά 99% από την ηλικία του εργαζόμενου, αυτό θα σήμαινε ότι άλλα χαρακτηριστικά όπως: η εκπαίδευση, η ικανότητα και η εμπειρία και διάφοροι άλλοι παράγοντες δεν έχουν καμία σχέση της μεταβλητότητας του ωρομισθίου. Αυτό φυσικά φαντάζει περίεργο ή αλλιώς υπονοεί πρόβλημα με τα δεδομένα. Κάτι τέτοιο σε πραγματικά δεδομένα είναι πολύ δύσκολο να συμβεί καθώς οι μεταβολές που αφορούν μια εξαρτημένη μεταβλητή π.χ. απόδοση μιας καλλιεργήσιμης έκτασης δεν εξαρτάται μόνο από μια μεταβλητή όπως για παράδειγμα το ύψος της βροχόπτωσης αλλά και από πολλές άλλες παραμέτρους όπως οι μέρες ηλιοφάνειας η περιεκτικότητα του εδάφους σε θρεπτικά συστατικά, η ποσότητα των φυτοφαρμάκων που χρησιμοποιούνται κλπ. οι οποίες περιέχουν ένα ποσοστό της πληροφόρησης με αποτέλεσμα το R^2 να είναι χαμηλότερο της μονάδας.

Γενικότερα όσο υψηλότερο είναι το R^2 τόσο καλύτερα μπορούμε να πούμε ότι προσαρμόζεται στα δεδομένα και τόσο καλύτερα ερμηνεύει την εξαρτημένη μεταβλητή το υπόδειγμα που εξετάζουμε βέβαια όμως δεν πρέπει να είμαστε απόλυτοι καθώς όπως προαναφέρθηκε μπορεί να οδηγηθούμε σε εσφαλμένα συμπεράσματα.

8) Προσδιορισμός των καταλοίπων

Τα κατάλοιπα αφορούν όλες τις παρατηρήσεις του δείγματος που εξετάζεται συνολικά και αφορά την διαφορά που υπάρχει από την παρατήρηση με την εκτιμώμενη τιμή της Y που λαμβάνουμε από την μέθοδο των ελαχίστων τετραγώνων.

$$\varepsilon = Y_i - \hat{Y}_i$$

Όπως έχει αναφερθεί και στην μέθοδο των ελαχίστων τετραγώνων απαραίτητη προϋπόθεση είναι να βρούμε τους συντελεστές των α και β με τέτοιο τρόπο ώστε να ελαχιστοποιούνται τα τετράγωνα των σφαλμάτων.

9) Έλεγχος για την κανονικότητα των καταλοίπων

Μια ακόμα σημαντική υπόθεση που γίνεται είναι αυτή που αφορά στα κατάλοιπα και κυρίως ότι αυτά κατανέμονται κανονικά, με μηδενική μέση τιμή και σταθερή διακύμανση. Δηλαδή, προσθέτουμε την κανονικότητα της κατανομής στις υποθέσεις που έχουν ήδη γίνει.

$$E[X \sim N[0, \sigma^2]]$$

Η παραπάνω σχέση σημαίνει τα κατάλοιπα είναι ανεξάρτητα και ασυσχέτιστα. Για τον έλεγχο της κανονικότητας των σφαλμάτων συχνά χρησιμοποιείται ο έλεγχος Jarque-Bera. Οι υποθέσεις στον εν λόγω έλεγχο διατυπώνονται ως εξής:

$$H_0: \text{Τα κατάλοιπα κατανέμονται κανονικά}$$
$$H_1: \text{Τα κατάλοιπα δεν κατανέμονται κανονικά}$$

10) Έλεγχος για την ύπαρξη αυτοσυσχέτισης στα κατάλοιπα.

Εδώ γίνεται η υπόθεση περί της πλήρους τυχαιότητας των καταλοίπων, δηλαδή ότι τα σφάλματα δεν συσχετίζονται μεταξύ τους.

Δηλαδή, $E[u_i | u_j, \dots, u_n] = 0$. Οι παραβιάσεις της ανεξαρτησίας $E[u_i, u_j] = 0$ είναι επίσης πολύ συχνές σε μοντέλα παλινδρόμησης με χρονοσειρές. Η ύπαρξη σειριακής συσχέτισης στα κατάλοιπα είναι συχνά σύμπτωμα ενός κακώς διαμορφωμένου (λάθος εξειδικευμένου) μοντέλου. Οι μορφές με τις οποίες εμφανίζεται η αυτοσυσχέτιση είναι πολυποικίλες με συνηθέστερη εκείνη της αυτοσυσχέτισης πρώτου βαθμού: της συσχέτισης, δηλαδή, των τιμών του τυχαίου σφάλματος οι οποίες απέχουν μεταξύ τους μία χρονική περίοδο.

Το πρόβλημα της αυτοσυσχέτισης εμφανίζεται στην ανάλυση της παλινδρόμησης κάθε φορά που παραβιάζεται η υπόθεση της ανεξαρτησίας των τιμών του τυχαίου σφάλματος ενός υποδείγματος (Καραμάνης2005)⁵.

Οι βασικότερες αιτίες που προκαλούν αυτοσυσχέτιση είναι οι ακόλουθες:

1. Οι περισσότερες οικονομικές χρονολογικές σειρές, όπως για παράδειγμα το ΑΕΠ, η κατανάλωση, η απασχόληση κ.α, παρουσιάζουν αδράνεια. Αυτό συμβαίνει επειδή τα εν λόγω μεγέθη ακολουθούν τους οικονομικούς κύκλους. Ξεκινώντας από τα χαμηλά μιας περιόδου ύφεσης οι τιμές αυτών των μεταβλητών μεταβάλλονται θετικά, δηλαδή η τιμή της επόμενης περιόδου είναι μεγαλύτερη από τη τιμή της προηγούμενης. Το φαινόμενο αυτό συνεχίζεται μέχρις ότου ξαναεμφανιστεί ύφεση και η μεταβολή επιβραδυνθεί ή και αναστραφεί.
2. Το φαινόμενο της αυτοσυσχέτισης μπορεί επίσης να οφείλεται στο ότι παρουσιάζουν κάποιες ανεξάρτητες μεταβλητές από το υπόδειγμα οι οποίες επηρεάζουν την εξαρτημένη μεταβλητή ή ότι το υπόδειγμα δεν έχει τη σωστή συναρτησιακή μορφή.
3. Η τρέχουσα τιμή της εξαρτημένης μεταβλητής μπορεί να εξαρτάται, πέραν όλων των άλλων, και από τις παρελθούσες τιμές αυτής. Αν λοιπόν οι παρελθούσες τιμές απουσιάζουν από τις ανεξάρτητες μεταβλητές του υπόδειγμα θα παρουσιάσει αυτοσυσχέτιση.

Οι τιμές των τυπικών σφαλμάτων των εκτιμητών των συντελεστών ενός υποδείγματος επηρεάζονται σημαντικά από το πρόβλημα της αυτοσυσχέτισης με συνέπεια όλοι οι στατιστικοί έλεγχοι που διενεργούνται επί αυτού του μοντέλου να μην δίνουν αξιόπιστα αποτελέσματα.

⁵Καραμάνης Ι. Δημήτριος (2005), Τμήμα Οργάνωσης και Διοίκησης Επιχειρήσεων Πανεπιστήμιο Πειραιά, Μελέτη υποδειγμάτων ARCH και εφαρμογή στον πληθωρισμό της Ελλάδος.

Οι μορφές με τις οποίες εμφανίζεται η αυτοσυσχέτιση είναι ποικίλες με συνηθέστερη εκείνη της αυτοσυσχέτισης πρώτου βαθμού: της συσχέτισης, δηλαδή, των τιμών του τυχαίου σφάλματος οι οποίες απέχουν μεταξύ τους μία χρονική περίοδο. Το πρόβλημα της αυτοσυσχέτισης – όπως επίσης και εκείνο της ετεροσκεδαστικότητας – θα πρέπει να λαμβάνεται σοβαρά υπ' όψιν στην ανάλυση της παλινδρόμησης ενός υποδείγματος, ώστε, τελικά, να μην αμφισβητείται η αξιοπιστία των αποτελεσμάτων της εκτίμησης των παραμέτρων του και να καθίσταται, παράλληλα, δυνατό να διενεργούνται με ακρίβεια οι απαραίτητοι στατιστικοί έλεγχοι.

ΚΕΦΑΛΑΙΟ 2 Πολλαπλή Γραμμική Παλινδρόμηση

2.1 Το Στατιστικό Μοντέλο Πολλαπλής Παλινδρόμησης

Το πολλαπλό υπόδειγμα γραμμικής παλινδρόμησης αποτελεί την γενίκευση του απλού γραμμικού υποδείγματος εισάγοντας περισσότερες από μία ερμηνευτικές μεταβλητές στα δεξιά του υποδείγματος. Για παράδειγμα ένα k -μεταβλητό υπόδειγμα μπορεί να γραφεί ως

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_k X_{ik} + u_i$$

Παρακάτω αναφέρεται ένας αριθμός παραδειγμάτων πολυμεταβλητών γραμμικών υποδειγμάτων από την εφαρμοσμένη οικονομική ανάλυση.

Παράδειγμα 1. Αναφορικά με την εξέταση μιας εξίσωσης μισθών όπου θέλουμε να δούμε τι επηρεάζει το ύψος του μισθού πέραν της εκπαίδευσης μπορούμε να ελέγξουμε την επίδραση των χρόνων εμπειρίας με βάση το υπόδειγμα

$$w = \alpha + \beta \cdot \text{εκπ} + \gamma \cdot \text{εμπ} + u$$

Παράδειγμα 2. Η απόδοση μιας καλλιέργειας επηρεάζεται από πολλούς παράγοντες. Παρακάτω παρουσιάζεται ένα παράδειγμα με το οποίο μπορούν να εξεταστούν κάποιοι από αυτούς όπως οι μέρες ηλιοφάνειας και το μέσο ύψος βροχόπτωσης.

$$R = \alpha + \beta_1 H + \beta_2 B + u_t$$

Όπου H οι μέρες ηλιοφάνειας ανά μήνα και B το μέσο ύψος βροχόπτωσης ανά μήνα.

2.2. Υποθέσεις Πολλαπλής Γραμμικής Παλινδρόμησης

Οι κλασικές υποθέσεις της απλής γραμμικής παλινδρόμησης με μια ανεξάρτητη μεταβλητή απλώς εμπλουτίζονται στο πολλαπλό υπόδειγμα παλινδρόμησης. Αυτές παρουσιάζονται αναλυτικά στη συνέχεια.

α. Η τυχαία μεταβλητή u έχει αναμενόμενη τιμή η μέση τιμή ίση με το μηδέν: $E(u) = 0$

β. Η σχέση μεταξύ των μεταβλητών Y και X είναι γραμμική.

γ. Η διακύμανση του διαταρακτικού όρου (σ^2) είναι σταθερή για όλες τις τιμές των ανεξάρτητων μεταβλητών: $Var(u_i) = E(u_i^2) = \sigma^2$

δ. Οι διαταρακτικοί όροι που αντιστοιχούν σε διαφορετικές παρατηρήσεις είναι ασυσχέτιστοι μεταξύ τους: $Cov(u_i, u_j) = 0$

ε. Μεταξύ του διαταρακτικού όρου και των ανεξάρτητων μεταβλητών δεν υπάρχει συσχέτιση

στ. Αφού ο διαταρακτικός όρος κατανέμεται κανονικά και η εξαρτημένη μεταβλητή Y_i κατανέμεται επίσης κανονικά

ζ. Οι ανεξάρτητες μεταβλητές δεν έχουν κανένα είδος γραμμικής συσχέτισης άρα το υπόδειγμα δεν εμφανίζει πολυσυγγραμμικότητα.

Οι παραπάνω υποθέσεις περιγράφουν τόσο τη μορφή του μοντέλου όσο και τι άλλο θα πρέπει να τηρείται μεταξύ των μερών του προκειμένου να είναι κατάλληλο το υπόδειγμα για εκτίμηση και διαμέσου αυτού να υπάρχει η δυνατότητα για την εκτίμηση αξιόπιστων αποτελεσμάτων.

2.3 Βήματα της Πολλαπλής Γραμμικής Παλινδρόμησης

Στη διαδικασία εκτίμησης ενός υποδείγματος πολλαπλής γραμμικής παλινδρόμησης ακολουθούνται τα παρακάτω βήματα:

1) Επιλογή ανεξάρτητων μεταβλητών και της εξαρτημένης

Η μοναδική διαφορά που υφίσταται μεταξύ της απλής γραμμικής παλινδρόμησης και της πολλαπλής γραμμικής παλινδρόμησης στο θέμα της επιλογής μεταβλητών είναι ότι στην πολλαπλή γραμμική παλινδρόμηση οι ανεξάρτητες μεταβλητές είναι πάνω από μια. Έτσι όπως και στην απλή γραμμική παλινδρόμηση και εδώ επιλέγονται μεταβλητές οι οποίες θεωρούμε ότι έχουν σχέση με την εξαρτημένη μεταβλητή, είτε γιατί αυτό προκύπτει από την οικονομική θεωρία είτε γιατί το υποπτευόμαστε. Πολύ σημαντικό είναι να μην εισάγουμε στο υπόδειγμα μεγάλο αριθμό ανεξάρτητων μεταβλητών ιδιαίτερα όταν έχουμε μικρό δείγμα να εξετάσουμε.

2) Υπολογισμός ανά ζεύγη των συντελεστών συσχέτισης μεταξύ εξαρτημένης μεταβλητής και ανεξάρτητων μεταβλητών

Ο υπολογισμός όπως αναφέρεται και στο τίτλο ο υπολογισμός των συντελεστών συσχέτισης γίνεται για όλα τα ζεύγη ανεξάρτητων και εξαρτημένης μεταβλητής. Για κανονικά κατανομημένες μεταβλητές χρησιμοποιείται ο τύπος του Pearson ενώ αν οι μεταβλητές μας δεν κατανέμονται κανονικά χρησιμοποιείται ο συντελεστής συσχέτισης του Spearman.

3) Υπολογισμός της ευθείας παλινδρόμησης με την μέθοδο των ελαχίστων τετραγώνων

Όπως ήδη αναφέρθηκε, η μέθοδος εκτίμησης μέσω των ελαχίστων τετραγώνων δημιουργεί μια ευθεία τέτοια ώστε η ευθεία αυτή να είναι η βέλτιστη όσον αφορά την

ελαχιστοποίηση των τετραγώνων των κάθετων αποστάσεων των σημείων που αναπαρίστανται γραφικά από τα σημεία της ευθείας αυτής. Με λίγα λόγια ζητάμε την ευθεία η οποία ελαχιστοποιεί το άθροισμα τετραγώνων των καταλοίπων.

4) Έλεγχος στατιστικής σημαντικότητας των συντελεστών β για κάθε μεταβλητή ξεχωριστά.

Ο έλεγχος για την στατιστική σημαντικότητα είναι σημαντικός για να δούμε την στατιστική αξία του μοντέλου μας και μπορεί να μας βοηθήσει στην ανάλυση των μεταβλητών που έχουμε επιλέξει σε σχέση με την εξαρτημένη μεταβλητή. Για κάθε συντελεστή ανεξάρτητης μεταβλητής γίνεται έλεγχος σημαντικότητας με μηδενική υπόθεση $H_0: \beta=0$. Η στατιστική ελέγχου είναι η $t = \frac{\hat{\beta}}{se(\hat{\beta})}$. Εάν $|t| > t_{\alpha/2, (n-2)}$, τότε απορρίπτεται η μηδενική υπόθεση.

5) Έλεγχος F για την σημαντικότητα του μοντέλου συνολικά

Για το συνολικό έλεγχο του υποδείγματος και για να εξακριβωθεί η στατιστική του εγκυρότητα πραγματοποιούμε έναν έλεγχο F. Η μηδενική υπόθεση αναφέρει ότι όλοι οι συντελεστές των ανεξάρτητων μεταβλητών είναι από κοινού 0.

$H_0: \beta_2=\beta_3=\dots=\beta_k=0$ έναντι της εναλλακτικής H_1 : «δεν είναι όλοι οι συντελεστές ταυτόχρονα ίσοι με το μηδέν», μπορούμε να χρησιμοποιήσουμε τη στατιστική ελέγχου:

$$F = \frac{SSE/(k-1)}{SSR/(n-k)} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$
 που ακολουθεί την κατανομή F με $k-1$ και $n-k$

βαθμούς ελευθερίας. Για να υπολογίσουμε την τιμή F απαιτείται να κατασκευάσουμε τον πίνακα ANOVA.

Πίνακας ANOVA				
	B.E.	Άθροισμα τετραγώνων	Μέσο άθροισμα τετραγώνων	Στοιχείο Ελέγχου
Παλινδρόμηση	1	SSR	MSR	F=MSR/MSE
Κατάλοιπα	n-2	SSE	MSE	
Σύνολο	n-1	Μεταβλητότητα y		

6) Υπολογισμός του συντελεστή προσδιορισμού

Ένα κριτήριο καλής προσαρμογής του υποδείγματος στα δεδομένα, είναι ο συντελεστής πολλαπλού προσδιορισμού (coefficient of multiple determination), R^2 , ο οποίος παίρνει τιμές στο διάστημα $0 < R^2 < 1$. Η προσαρμογή είναι καλή αν η τιμή του R^2 είναι κοντά στη μονάδα. Ωστόσο ο χαμηλές τιμές του R^2 δεν πρέπει να μας οδηγούν στο συμπέρασμα ότι η παλινδρόμηση είναι άχρηστη (Wooldridge, 2006). Αξίζει να σημειωθεί ότι ο R^2 είναι κατάλληλος για τη σύγκριση της προσαρμογής δύο υποδειγμάτων που έχουν την ίδια εξαρτημένη μεταβλητή και τον ίδιο αριθμό ερμηνευτικών μεταβλητών, οι οποίες μπορεί βεβαίως να διαφέρουν στα δύο υποδείγματα.

Ο διορθωμένος συντελεστής προσδιορισμού

Από το τρόπο που ορίστηκε ο συντελεστής προσδιορισμού, όπως είδαμε παραπάνω, γίνεται αντιληπτό ότι μια ενδεχόμενη προσθήκη περαιτέρω επεξηγηματικών μεταβλητών θα αύξανε την τιμή του R^2 . Η εξήγηση για αυτό είναι ότι με τη προσθήκη επεξηγηματικών μεταβλητών η τιμή για το RSS θα μειωθεί ή θα μείνει αμετάβλητη, ενώ η τιμή του TSS παραμένει αμετάβλητη. Αυτό μπορεί να δημιουργήσει πρόβλημα κατά την χρήση ως μέτρο καλής προσαρμογής του υποδείγματος. Ένα καλύτερο κριτήριο για τη σύγκριση της προσαρμογής δύο υποδειγμάτων, τα οποία έχουν μεν την ίδια εξαρτημένη μεταβλητή, αλλά διαφέρουν ως προς τον αριθμό των ερμηνευτικών

μεταβλητών, είναι ο διορθωμένος συντελεστής προσδιορισμού (adjusted coefficient of determination), \bar{R}^2

7) Υπολογισμός Αθροίσματος Καταλοίπων

Με δεδομένο ότι με τον συντελεστή R^2 έχουμε λάβει το ποσοστό της μεταβλητότητας που ερμηνεύεται από το μοντέλο έχει νόημα να δούμε τι συμβαίνει με το ανερμήνευτο κομμάτι ή αλλιώς a κατάλοιπα. Το άθροισμα των τετραγώνων των καταλοίπων προκύπτει αν από τα SST (η συνολική μεταβλητότητα της Y) αφαιρέσουμε το άθροισμα SSR (η μεταβλητότητα της Y που ερμηνεύεται από την παλινδρόμηση).

8) Έλεγχος για την κανονικότητα των καταλοίπων

Μια ακόμα σημαντική υπόθεση που γίνεται είναι αυτή που αφορά στα κατάλοιπα και κυρίως ότι αυτά κατανέμονται κανονικά, με μηδενική μέση τιμή και σταθερή διακύμανση. Όπως αναφέρεται στη βιβλιογραφία⁶, «παραβιάσεις της κανονικότητας οφείλονται μερικές φορές στην παρουσία μερικών μεγάλων παρατηρήσεων που δεν προέρχονται από το δείγμα. Αφού για την εκτίμηση των παραμέτρων χρησιμοποιείται η μέθοδος των ελαχίστων τετραγώνων, η ύπαρξη ακραίων παρατηρήσεων μπορεί να επηρεάσει κατά τρόπο δυσανάλογο τις εκτιμήσεις των παραμέτρων».

9) Έλεγχος αυτοσυσχέτισης των καταλοίπων

Σημαντικό είναι επίσης να μην αυτοσυσχετίζονται τα κατάλοιπα. Το πρόβλημα της αυτοσυσχέτισης μπορεί να οφείλεται (i) σε καθαρή αυτοσυσχέτιση στο υπόδειγμα, (ii) σε παράλειψη σημαντικής ανεξάρτητης μεταβλητής και (iii) σε κακή εξειδίκευση του υποδείγματος (συναρτησιακή μορφή).

⁶Κιουφεντζή Ο. (2006), Το Κλασικό Πολλαπλό Γραμμικό Μοντέλο Παλινδρόμησης – Στατιστικά Συμπεράσματα και Εκτιμήσεις.

10) Έλεγχος για την ανεξαρτησία των μεταβλητών

Μια ακόμα ιδιαίτερα σημαντική υπόθεση που γίνεται είναι αυτή που αφορά στην ανεξαρτησία των μεταβλητών. Δηλαδή υποθέτουμε ότι δεν υπάρχει κάποια ακριβής γραμμική σχέση μεταξύ των μεταβλητών, μια γραμμική σχέση που με οποιονδήποτε τρόπο να μπορεί να περιγράψει την σχέση τους. Η χρήση ενός υποδείγματος πολλαπλής παλινδρόμησης για την εξέταση ενός φαινομένου μας δίνει πολύ καλύτερα αποτελέσματα εκτίμησης όταν οι ανεξάρτητες μεταβλητές που χρησιμοποιούνται στο μοντέλο δεν συσχετίζονται μεταξύ τους. Στην αντίθετη περίπτωση το υπόδειγμα δεν μπορεί να μας δώσει την απαραίτητη πληροφόρηση. Όταν οι ανεξάρτητες μεταβλητές συσχετίζονται μεταξύ τους είναι πιθανό οι συντελεστές της παλινδρόμησης που θα ληφθούν μετά την εκτίμησή τους να παρουσιάζουν αστάθεια και οι τιμές ή το πρόσημό τους να μεταβάλλονται ύστερα από την πρόσθεση ή την αφαίρεση κάποιας μεταβλητής από το υπόδειγμα. Επίσης είναι δυνατόν ακόμα και μεταβολές στις τιμές των δεδομένων να επηρεάζουν σημαντικά τις τιμές των συντελεστών. Για παράδειγμα μπορεί μια μεταβλητή να φαίνεται να έχει θετική επιρροή πάνω στην εξαρτημένη μεταβλητή ενώ με την προσθήκη μιας νέας μεταβλητής μπορεί το πρόσημό της να αλλάξει και να γίνει αρνητικό κάτι που σημαίνει ότι δημιουργείται ασάφεια όσον αφορά το πώς πραγματικά συνδέονται οι δυο αυτές μεταβλητές. Η προβληματική κατάσταση κατά την οποία υπάρχουν ισχυρές συσχετίσεις μεταξύ των ανεξάρτητων μεταβλητών στην πολλαπλή παλινδρόμηση ονομάζεται πολυσυγγραμμικότητα (multicollinearity).

Όπου υφίσταται αυτό το πρόβλημα, απαιτείται ιδιαίτερη προσοχή στην ερμηνεία όλων των εκτιμητριών που προκύπτουν από το μοντέλο αυτό.

Ο ερευνητής μπορεί μέσα από μια σειρά ενδείξεων να υποπτευθεί την ύπαρξη πολυσυγγραμμικότητας. Ο πιο ασφαλής τρόπος για την ανίχνευση του προβλήματος της πολυσυγγραμμικότητας είναι μέσω του πίνακα των συντελεστών συσχέτισης των ανεξάρτητων μεταβλητών.

	X	Y	Z
X	1	0.99	-0.92
Y	0.99	1	0.12
Z	-0.92	0.12	1

Πίνακας 1 Πίνακας συσχετίσεων

Αν για παράδειγμα πάρουμε τα αποτελέσματα που παρουσιάζονται στο παραπάνω πίνακα, θα παρατηρήσει κανείς ότι εμφανίζονται μεγάλες θετικές και αρνητικές τιμές οι οποίες βρίσκονται κοντά στην μονάδα. Αυτές οι τιμές μας φανερώνουν ότι οι μεταβλητές έχουν μεταξύ τους ισχυρό βαθμό συσχέτισης. Ένα συμπέρασμα που προκύπτει από τα παραπάνω αποτελέσματα είναι ότι κάποιες από τις χρησιμοποιούμενες μεταβλητές έχουν πολύ μικρή συνεισφορά στην εκτίμηση της εξαρτημένης μεταβλητής οπότε θα πρέπει να αφαιρεθούν από το μοντέλο τελείως ή να αντικατασταθούν από άλλες που δεν παρουσιάζουν αυτό το πρόβλημα. Άλλες ενδείξεις για την ενδεχόμενη ύπαρξη πολυσυγγραμικότητας είναι:

α) Η ύπαρξη αντίθετων πρόσημων σε ορισμένους συντελεστές στην παλινδρόμηση από αυτά που είναι αναμενόμενα είτε λόγω της οικονομικής θεωρίας είτε λόγω της φύσης του προβλήματος

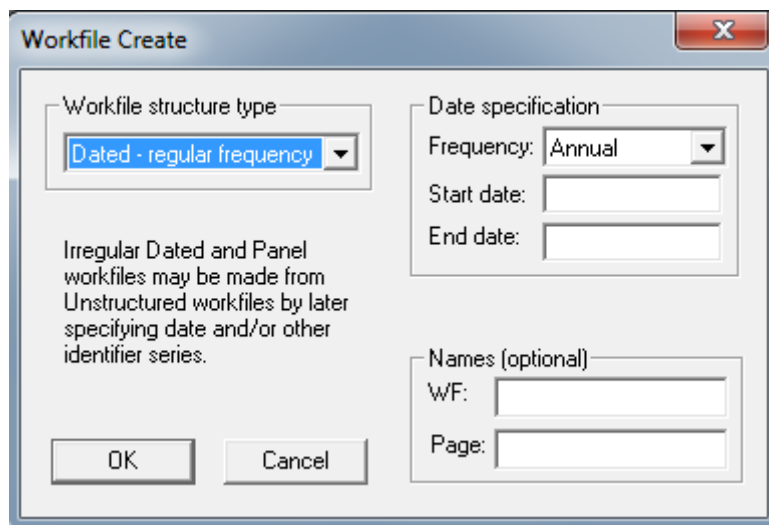
β) Εάν εμφανίζονται μεγάλες τιμές στις τυπικές αποκλίσεις των στατιστικά σημαντικών συντελεστών της παλινδρόμησης

Και οι δύο παραπάνω ενδείξεις είναι ανησυχητικές για την ενδεχόμενη ύπαρξη πολυσυγγραμικότητας και θα πρέπει ο εκάστοτε ερευνητής να προχωρήσει σε περαιτέρω έρευνα για την ανίχνευσή της.

ΚΕΦΑΛΑΙΟ 3: Βασικές αρχές χρήσης Eviews

Το πακέτο Eviews είναι μια εφαρμογή για περιβάλλον Windows της οποίας σκοπός είναι η ανάλυση οικονομικών δεδομένων, η παρουσίαση των αποτελεσμάτων και η κατασκευή οικονομετρικών υποδειγμάτων. Το πακέτο Eviews χρησιμοποιείται κυρίως όταν θέλουμε να αναλύσουμε οικονομικές χρονολογικές σειρές.

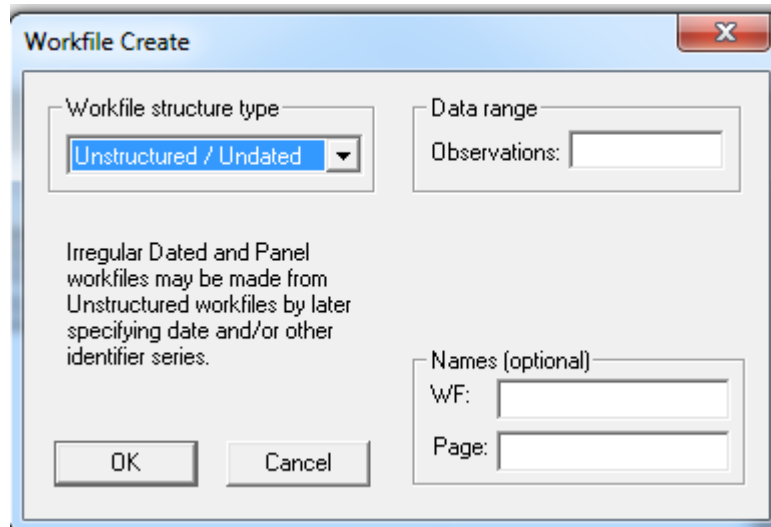
Για την εισαγωγή των στοιχείων στο Eviews θα πρέπει να επιλέξουμε την σειρά επιλογών, File>New>Workfile με την οποία δημιουργείται ένα νέο αρχείο εργασίας στο οποίο καλούμαστε πρώτα από όλα να επιλέξουμε το είδος των δεδομένων που θα χρησιμοποιήσουμε και στην συνέχεια να τα εισάγουμε με τις επιλογές που είναι διαθέσιμες. Αξίζει να αναφερθεί πως η διαδικασία για την εκτίμηση ενός απλού οικονομετρικού υποδείγματος με το εν λόγω στατιστικό πακέτο είναι ιδιαίτερα απλή και δεν απαιτεί ιδιαίτερες γνώσεις του προγράμματος παρά μόνο μια καλή ανάγνωση του εγχειριδίου χρήσης του το οποίο το συνοδεύει.



Εικόνα5 Δημιουργία αρχείου εργασίας

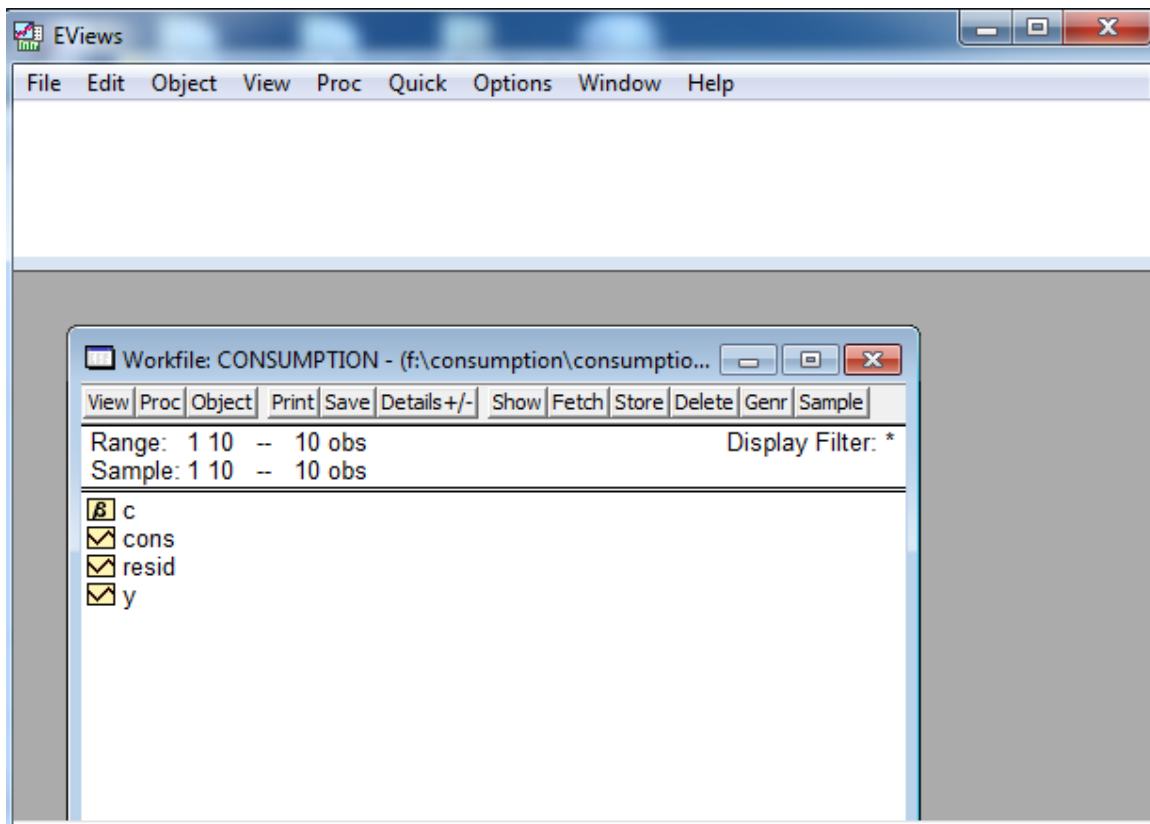
Συνεχίζοντας, με τις επιλογές αυτές δηλώνουμε ότι θα πρέπει να δημιουργηθεί νέο αρχείο εργασίας. Στην συνέχεια βλέπουμε τον πίνακα επιλογών Workfile Range του πακέτου στον οποίο θα πρέπει να ορίσουμε το είδος των σειρών που έχουμε καθώς

επίσης και το εύρος στο οποίο κινούνται (πχ ημερομηνίες αν πρόκειται για χρονολογικά στοιχεία) ή τυχαίες παρατηρήσεις όπως φαίνονται στη παρακάτω εικόνα.



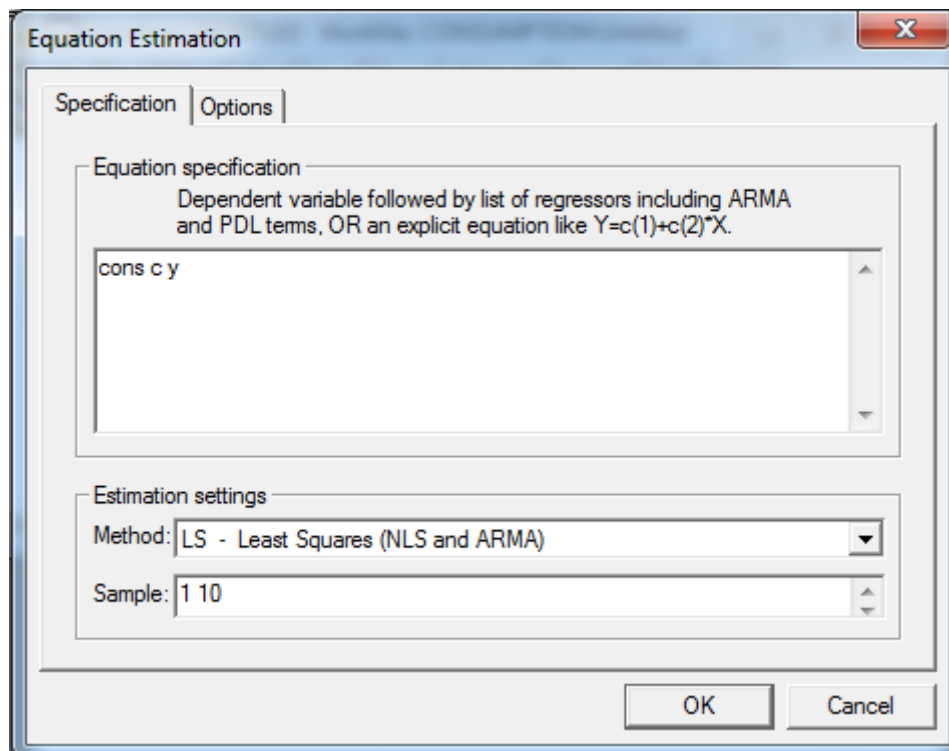
Εικόνα Δημιουργία αρχείου εργασίας, τυχαίες παρατηρήσεις

Στην συγκεκριμένη περίπτωση θα επιλέξουμε Unstructured/ Undated και θα εισάγουμε 10 παρατηρήσεις στο πεδίο Observations. Με τον τρόπο αυτό το πακέτο αντιλαμβάνεται ότι έχουμε 10 παρατηρήσεις που αναφέρονται σε μη χρονολογικά στοιχεία. Στη συνέχεια ονομάζουμε και τις μεταβλητές που θα χρησιμοποιηθούν και ακολουθεί η εισαγωγή των δεδομένων.



Εικόνα7 Κεντρική οθόνη Eviews

Η εισαγωγή των δεδομένων γίνεται με την επιλογή File>Import>ReadText-Lotus-Excel: Με την εντολή αυτή μπορούμε να εισάγουμε στοιχεία στο πακέτο σε μια ποικιλία μορφών, όπως πχ Excel. Στην εντολή αυτή μπορούμε επίσης να εισάγουμε ορισμένες σειρές από μια βάση δεδομένων την οποία έχουμε ήδη δημιουργήσει όπως για παράδειγμα στο Excel. Στο εν λόγω παράδειγμα έχουμε εισάγει 10 παρατηρήσεις (10 obs) και οι μεταβλητές είναι: η σταθερά c , η μεταβλητή της κατανάλωσης $cons$, η μεταβλητή του εισοδήματος y και τα κατάλοιπα $resid$ τα οποία δημιουργεί το πρόγραμμα αυτομάτως.



Εικόνα 8 Εκτίμηση συνάρτησης

Τέλος, προκειμένου να εκτιμήσουμε ένα απλό γραμμικό υπόδειγμα επιλέγουμε Quick>Estimate equation όπου όπως φαίνεται και στην ανωτέρω εικόνα γίνεται η εκτίμηση της επιθυμητής συνάρτησης. Στο πεδίο Equation specification δίνεται η μορφή της συνάρτησης που είναι επιθυμητή. Εδώ αναγράφονται με την σειρά η εξαρτημένη μεταβλητή (cons), η σταθερά (c) και η ανεξάρτητη μεταβλητή (y). Στο επόμενο πεδίο ακολουθεί η επιλογή της μεθόδου εκτίμησης του υποδείγματος. Προεπιλογή είναι η μέθοδος των ελαχίστων τετραγώνων η οποία είναι και η επιθυμητή ενώ στο τελευταίο πεδίο παρουσιάζεται το μέγεθος του δείγματος. Τα αποτελέσματα της παλινδρόμησης καθώς και κάποια διαγράμματα και ελέγχους που μας δίνει την δυνατότητα να κάνουμε το στατιστικό αυτό πακέτο παρουσιάζονται στην συνέχεια.

ΚΕΦΑΛΑΙΟ 4: Εφαρμογές

Στα παραδείγματα που ακολουθούν τόσο για την απλή γραμμική παλινδρόμηση όσο και για την πολλαπλή γραμμική παλινδρόμηση γίνεται μια προσπάθεια για πρακτική αποτύπωση όλων όσων προηγήθηκαν μέσω απλών εφαρμογών. Τα υποδείγματα αφορούν περιπτώσεις όπου υπάρχει θεωρητικό υπόβαθρο από την οικονομική θεωρία όπως η σχέση εισοδήματος και κατανάλωσης, αποδόσεις τραπεζικών μετοχών σε σχέση με την απόδοση του Γ.Δ. του χρηματιστηρίου αλλά και ένα παράδειγμα που αφορά στην σχέση του σωματικού βάρους με το ύψος και τη σχέση του ποσοστού του σωματικού λίπους μιας ομάδας ατόμων σε σχέση με διάφορα σωματικά χαρακτηριστικά τους.

4.1 Κατανάλωση νοικοκυριού

Αφού έγινε η ανάλυση του θεωρητικού υποβάθρου του απλού γραμμικού υποδείγματος τώρα θα ακολουθήσουν κάποιες εφαρμογές του για την καλύτερη κατανόησή του.

Με το υπόδειγμα που θα εκτιμήσουμε θα προσπαθήσουμε να δούμε την σχέση μεταξύ εισοδήματος και κατανάλωσης όσον αφορά ένα νοικοκυριό. Δηλαδή θα προσπαθήσουμε να δούμε τι επίδραση έχει το ύψος του εισοδήματος στην συνολική κατανάλωση. Σύμφωνα με τον Γερμανό στατιστικό Ernst Engel (1821-1896) όσο το εισόδημα ενός νοικοκυριού αυξάνει τόσο η κατανάλωση σε φαγητό αυξάνει κατ' αναλογία ή η εισοδηματική ελαστικότητα για την ζήτηση για φαγητό είναι μεγαλύτερη του μηδενός αλλά μικρότερη της μονάδας. Αυτό σημαίνει ότι καθώς αυξάνεται το εισόδημα, το ποσοστό του εισοδήματος που δαπανάται για τρόφιμα μειώνεται, ακόμη και αν οι πραγματικές δαπάνες τρόφιμα αυξάνονται.

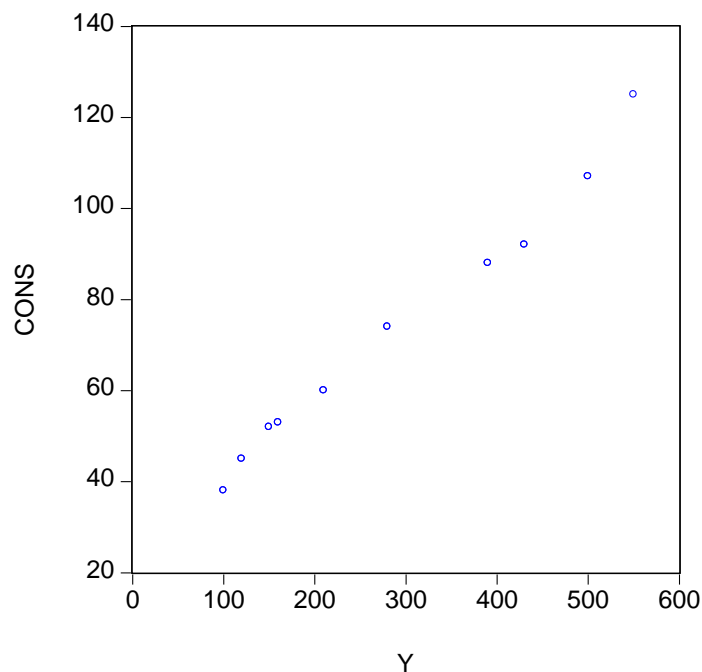
Το οικονομικό μοντέλο που θα εξετάσουμε για να περιγράψουμε αυτή τη σχέση είναι το παρακάτω.

Κατανάλωση σε Φαγητό $Y_i = \alpha + \beta * X_i + \varepsilon_i$, όπου $\alpha > 0$ και $0 < \beta < 1$.

Έστω ότι έχουμε τα παρακάτω δεδομένα:

Εισόδημα	100	120	150	160	210	280	390	430	500	550
Κατανάλωση	38	45	52	53	60	74	88	92	107	125

Πίνακας2 Πίνακας Δεδομένων



Εικόνα9 Διάγραμμα διασποράς

Παραπάνω δίνεται ένα διάγραμμα διασποράς των δεδομένων. Η επεξηγηματική μεταβλητή Y (εισόδημα) απεικονίζεται στον οριζόντιο άξονα ενώ η εξαρτημένη μεταβλητή CONS (κατανάλωση) στον κάθετο. Κάθε σημείο στο διάγραμμα απεικονίζει τις προτιμήσεις ενός νοικοκυριού. Μπορεί κανείς να πει ότι με μια πρώτη ματιά ότι υπάρχει θετική σχέση ανάμεσα στη κατανάλωση των νοικοκυριών και το εισόδημα τους. Το οικονομετρικό μοντέλο που θαεκτιμηθεί είναι το εξής:

$$\hat{Y}_i = \hat{a} + \hat{\beta}X_i + \hat{u}_i \text{ για } i=1, \dots, n \text{ όπου}$$

Y_i : η κατανάλωση του νοικοκυριού i .

α : σταθερός όρος.

β : κλίση.

n : το μέγεθος του δείγματος.

X_i : το εισόδημα του νοικοκυριού i .

ε_i : ο διαταρακτικός όρος του νοικοκυριού i .

Σκοπός της εφαρμογής είναι η εκτίμηση των συντελεστών α και β του υποδείγματος μας. Η μέθοδος η οποία και θα χρησιμοποιηθεί είναι η μέθοδος (OLS, ordinary least square) ελαχίστων τετραγώνων. Σύμφωνα με αυτήν οι εκτιμήσεις των παραμέτρων α και β δίνονται από τους παρακάτω τύπους:

$$\hat{a} = \bar{Y} - \hat{\beta}\bar{X}$$

$$\hat{\beta} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Με την χρήση του οικονομετρικού πακέτου Eviewσεκτιμούμε την παραπάνω σχέση και παίρνουμε τα εξής αποτελέσματα:

DependentVariable: CONS				
Method: LeastSquares				
Date: 08/24/15 Time: 20:50				
Sample: 1 10				
Includedobservations: 10				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	23.77765	2.498524	9.516679	0.0000
X	0.171704	0.007587	22.63137	0.0000
R-squared	0.984621	Meandependentvar		73.40000
Adjusted R-squared	0.982698	S.D. dependentvar		28.79892
S.E. of regression	3.788098	Akaikcinfocriterion		5.678461
Sum squaredresid	114.7975	Schwarzcriterion		5.738978
Loglikelihood	-26.39231	Hannan-Quinncrier.		5.612074
F-statistic	512.1787	Durbin-Watsonstat		1.330353
Prob(F-statistic)	0.000000			

Πίνακας3 Αποτελέσματα παλινδρόμησης

Από το παραπάνω πίνακα παρατηρούμε ότι το μοντέλο που αρχικώς εκτιμήθηκε $\hat{Y}_i = \hat{a} + \hat{\beta}X_i + \hat{u}_i$ παίρνει τις τιμές $\hat{Y}_i = 23,77 + 0,171X_i + \hat{u}_i$. Δηλαδή ο σταθερός συντελεστής είναι θετικός και παίρνει την τιμή 23,77 ενώ ο συντελεστής X (εισόδημα) παίρνει την τιμή 0,171.

Αρχικά, πριν αναφερθούμε στην ερμηνεία των συντελεστών αξίζει να επισημάνουμε ότι τόσο ο σταθερός όρος όσο και η ανεξάρτητη μεταβλητή του εισοδήματος είναι στατιστικά σημαντικές. Αυτό προκύπτει από την τιμή του p-valueπου και για τον σταθερό όρο και για την μεταβλητή X (εισόδημα) είναι 0.000.

Αυτό σημαίνει ότι απορρίπτεται η μηδενική υπόθεση $H_0 : C=0$ και $H_0 : X=0$ έναντι των εναλλακτικών $H1: C \neq 0$ και $H1: X \neq 0$ αντίστοιχα.

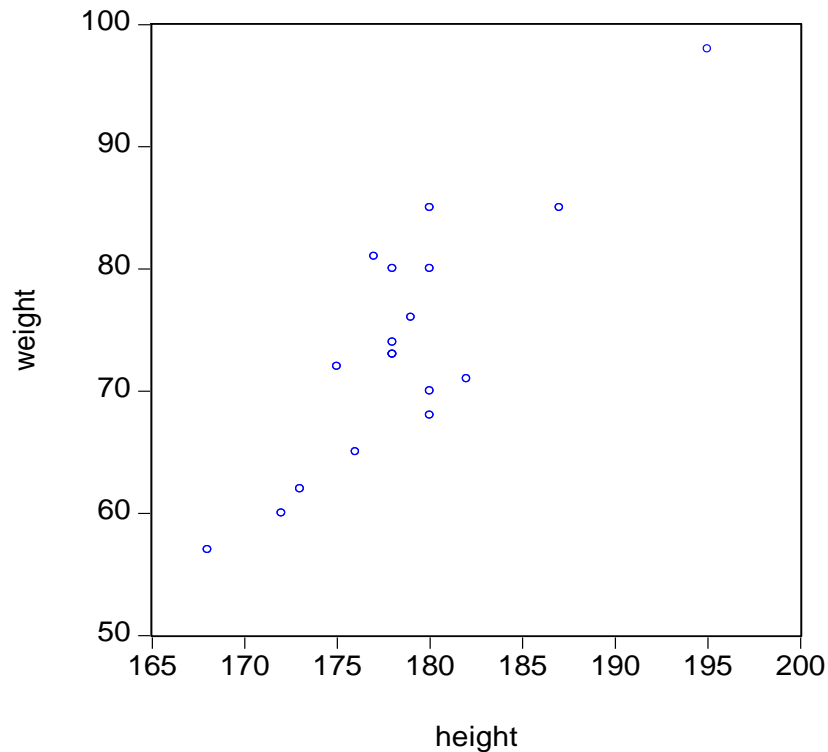
Η ερμηνεία του συντελεστή εισόδημα μπορεί να διατυπωθεί ως εξής: Για κάθε μονάδα αύξησης του εισοδήματος αυξάνεται η συνολική κατανάλωση των εξεταζόμενων νοικοκυριών κατά 0,17 μονάδες. Αν π.χ. η μονάδα μέτρησης του εισοδήματος είναι σε € μπορούμε να πούμε ότι για κάθε ένα ευρώ αύξησης του εισοδήματος αυξάνεται η

κατανάλωση κατά 0,17 ευρώ. Δηλαδή παρατηρείται μικρότερη αύξηση στο ρυθμό κατανάλωσης από το ρυθμό αύξησης του εισοδήματος.

Αξίζει να αναφερθεί επίσης τι το R^2 (συντελεστής συσχέτισης) του υποδείγματος όπως φαίνεται και στο παραπάνω πίνακα παίρνει την τιμή 0,98 πράγμα που σημαίνει ότι ερμηνεύεται το 98% της μεταβλητότητας της εξαρτημένης μεταβλητής, δηλαδή ότι το εκτιμημένο υπόδειγμα προσαρμόζεται πολύ καλά στα δεδομένα.

4.2 Σωματικά χαρακτηριστικά φοιτητών

Σε ένα ακόμα παράδειγμα εξετάζουμε ένα τυχαίο δείγμα 18 φοιτητών όσον αφορά το βάρος και το ύψος τους. Σε αυτό το παράδειγμα θεωρούμε ως ανεξάρτητη μεταβλητή το ύψος του κάθε φοιτητή και ως εξαρτημένη μεταβλητή το βάρος του. Το ερώτημα που τίθεται εδώ είναι να δούμε αν το ύψος επηρεάζει το βάρος των εξεταζόμενων φοιτητών.



Διάγραμμα 10 Διάγραμμα Διασποράς

Για παράδειγμα από το παραπάνω διάγραμμα μπορούμε να παρατηρήσουμε ότι πιο βαρείς είναι κυρίως οι πιο ψηλοί φοιτητές. Απ την άλλη παρατηρείται επίσης το φαινόμενο πολλοί φοιτητές με το ίδιο ύψος να έχουν πολύ διαφορετικό βάρος. Από το παραπάνω διάγραμμα διασποράς είναι προφανές ότι το ύψος και το βάρος των εξεταζόμενων φοιτητών σχετίζονται. Φαίνεται δηλαδή, να είναι δυνατόν να προσεγγιστεί γραμμικά η σχέση αυτή. Αν παραστήσουμε τα ζεύγη των παρατηρήσεων ύψους και

βάρους σε ένα διάγραμμα, θα προκύψει το παραπάνω διάγραμμα διασποράς. (Διάγραμμα 9). Η προσεκτική παρατήρηση του διαγράμματος διασποράς μπορεί να μας δώσει σημαντικές πληροφορίες για τη σχέση που ενδεχομένως υπάρχει μεταξύ των μεταβλητών τις οποίες εξετάζουμε. Στη συγκεκριμένη περίπτωση είναι λογικό κάποιος να υποπτευθεί την ύπαρξη μιας σχέσης μεταξύ ύψους και βάρους ενός ατόμου

	W	Y
Mean	73.88889	178.6667
Median	73.00000	178.0000
Maximum	98.00000	195.0000
Minimum	57.00000	168.0000
Std. Dev.	10.05800	5.790459
Skewness	0.440881	1.002836
Kurtosis	3.165411	5.114718
Jarque-Bera	0.603648	6.371067
Probability	0.739468	0.041356
Sum	1330.000	3216.000
Sum Sq. Dev.	1719.778	570.0000
Observations	18	18

Πίνακας 4 Περιγραφικά στοιχεία μεταβλητών

Στον πίνακα 3 παρουσιάζονται τα περιγραφικά στοιχεία των μεταβλητών. Από αυτόν τον πίνακα αντλούμε πολύ χρήσιμα στοιχεία για να έχουμε καλύτερη εικόνα για τις μεταβλητές που χρησιμοποιούνται. Εδώ βλέπουμε ότι το μέσο βάρος των φοιτητών που εξετάζονται είναι 73,8 κιλά ενώ το μέσο ύψος είναι 1,78. Στη συνέχεια ακολουθεί ένας πίνακας συσχετίσεων από τον οποίο βλέπουμε ότι επιβεβαιώνεται η αρχική μας υποψία για την ύπαρξη σχέσης μεταξύ ύψους και βάρους καθώς ο συντελεστής συσχέτισης έχει τιμή 0,85.

	W	Y
W	1	0.85
Y	0.85	1

Πίνακας 5 Πίνακας Συσχετίσεων

Το επόμενο βήμα είναι να εξετάσουμε ποιά είναι η ακριβής αυτή γραμμική σχέση θα εκτιμήσουμε την ευθεία παλινδρόμησης με την μέθοδο των ελαχίστων τετραγώνων. Η εξίσωση της ευθείας παλινδρόμησης δίνεται από τη σχέση:

$$w = \alpha + \beta y + u_i$$

όπου α είναι η σταθερά της σχέσης και β είναι ο συντελεστής για την ανεξάρτητη μεταβλητή μας. Με την χρήση του οικονομετρικού πακέτου Eview εκτιμούμε την παραπάνω σχέση και παίρνουμε τα εξής αποτελέσματα:

Dependent Variable: W				
Method: Least Squares				
Date: 04/05/16 Time: 15:04				
Sample: 1 18				
Included observations: 18				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-190.4542	40.66777	-4.683173	0.0002
Y	1.479532	0.227505	6.503287	0.0000
R-squared	0.725523	Mean dependent var		73.88889
Adjusted R-squared	0.708368	S.D. dependent var		10.05800
S.E. of regression	5.431615	Akaike info criterion		6.326789
Sum squared resid	472.0390	Schwarz criterion		6.425719
Log likelihood	-54.94110	Hannan-Quinn criter.		6.340430
F-statistic	42.29274	Durbin-Watson stat		2.973083
Prob(F-statistic)	0.000007			

Πίνακας 6 Αποτελέσματα παλινδρόμησης

Από το παραπάνω πίνακα βλέπουμε ότι η σταθερά παίρνει την τιμή -190,4 ενώ ο συντελεστής του ύψους εκτιμάται στο 1,47. Επίσης τόσο η σταθερά όσο και η μεταβλητή του ύψους είναι στατιστικά σημαντικές. Επίσης παρατηρούμε ότι ο συντελεστής

προσδιορισμού (R-squared) είναι 0,72 πράγμα που σημαίνει ότι το 72% του τρόπου με τον οποίο μεταβάλλεται η εξαρτημένη μεταβλητή (βάρος) εξαρτάται από την μεταβολή της ανεξάρτητης μεταβλητής (ύψος). Άλλα χρήσιμα όπως η F-statisticη οποία ελέγχει ότι όλοι οι συντελεστές του υποδείγματος (εκτός του σταθερού όρου) είναι 0. Η πιθανότητα για να επιβεβαιωθεί ο παραπάνω έλεγχος δίνεται από τη τιμή Prob(F-Static) όπου εδώ απορρίπτεται η αρχική υπόθεση ότι η τιμή του συντελεστή της μεταβλητής του ύψους είναι 0.

Στο πίνακα επίσης παρατηρούμε και άλλα χρήσιμα στοιχεία όπως τη μέση τιμή της εξαρτημένης μεταβλητής (Meandependentvar)την τυπική της απόκλιση (S.D. dependentvar), στατιστικές για την ύπαρξη γραμμικής συσχέτισης στα κατάλοιπα (Durbin-Watsonstat) όπου εδώ δεν φαίνεται να υπάρχει θετική γραμμική συσχέτιση (Durbin-Watsonstat > 2).

Συνοψίζοντας, μπορούμε να πούμε ότι ορθά επιλέξαμε την μεταβλητή του ύψους για να ερμηνεύσουμε το βάρος των εν λόγω φοιτητών καθώς κατά 72% περίπου όπως είδαμε και από τα αποτελέσματα της παλινδρόμησης το βάρος επηρεάζεται από το ύψος του φοιτητή. Όμως βλέπουμε ότι υπάρχουν και άλλοι παράγοντες οι οποίοι θα μπορούσαν να εξεταστούν προκειμένου να λάβουμε μια πιο σαφή εικόνα για το βάρος των φοιτητών. Αυτοί οι παράγοντες θα αποτελούσαν στην ουσία νέες μεταβλητές για το υπόδειγμα μας όπως για παράδειγμα η ημερήσια πρόσληψη θερμίδων, οι ώρες ημερήσιας άσκησης κλπ. Αυτό θα απαιτούσε μια άλλη μορφή υποδείγματος. Αυτή η μορφή υποδείγματος καθώς και εφαρμογές του παρουσιάζονται στο επόμενο κεφάλαιο.

4.3 Αποδόσεις τραπεζικών μετοχών

Μια πολύ σημαντική εφαρμογή των προαναφερθέντων αγορά στην ανάλυση οικονομικών δεδομένων και κυρίως δια μέσου χρονολογικών σειρών. Η ανάλυση αποδόσεων μετοχών είναι συνήθης στην προσπάθεια κατανόησης του τρόπου με τον οποίο κινείται μια συγκεκριμένη μετοχή ή ένα σύνολο μετοχών.

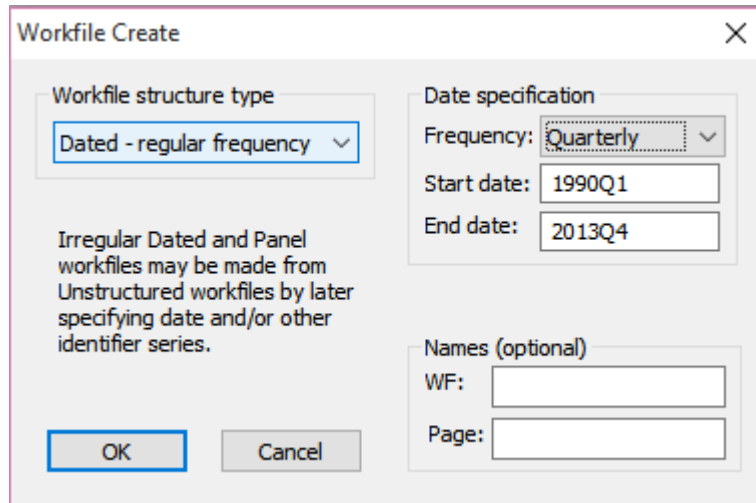
Στο παρακάτω υπόδειγμα που κατασκευάσαμε χρησιμοποιείται ως εξαρτημένη μεταβλητή η απόδοση του ΓΔ(GDR) του Χρηματιστηρίου Αθηνών και ως ανεξάρτητες μεταβλητές οι ημερήσιες αποδόσεις τραπεζικών μετοχών. Αυτές είναι η Τράπεζα Ελλάδος (ELLR), η Εθνική Τράπεζα (ETER) και η Τράπεζα Alpha (ALPHAR). Αυτό θα μπορούσε να είναι χρήσιμο για έναν επενδυτή ο οποίος θα ήθελε να δει πως αντιδρούν έως τώρα ορισμένες μετοχές που θα ήθελε να προσθέσει στο χαρτοφυλάκιο του.

Τα δεδομένα αφορούν μια περίοδο ενός περίπου έτους και τα χρησιμοποιείται ένα μοντέλο πολλαπλής παλινδρόμησης. Με τη βοήθεια του Eviews παίρνουμε τα εξής αποτελέσματα:

Το υπόδειγμα έχει την μορφή:

$$GDR = \alpha + \beta_1 * ELLR + \beta_2 * ETER + \beta_3 * ALPHAR + u_i$$

Αρχικά εισάγουμε στο Eviewστα δεδομένα όπως έχει ήδη αναφερθεί και διαλέγουμε τον τύπο των δεδομένων όπου εδώ είναι



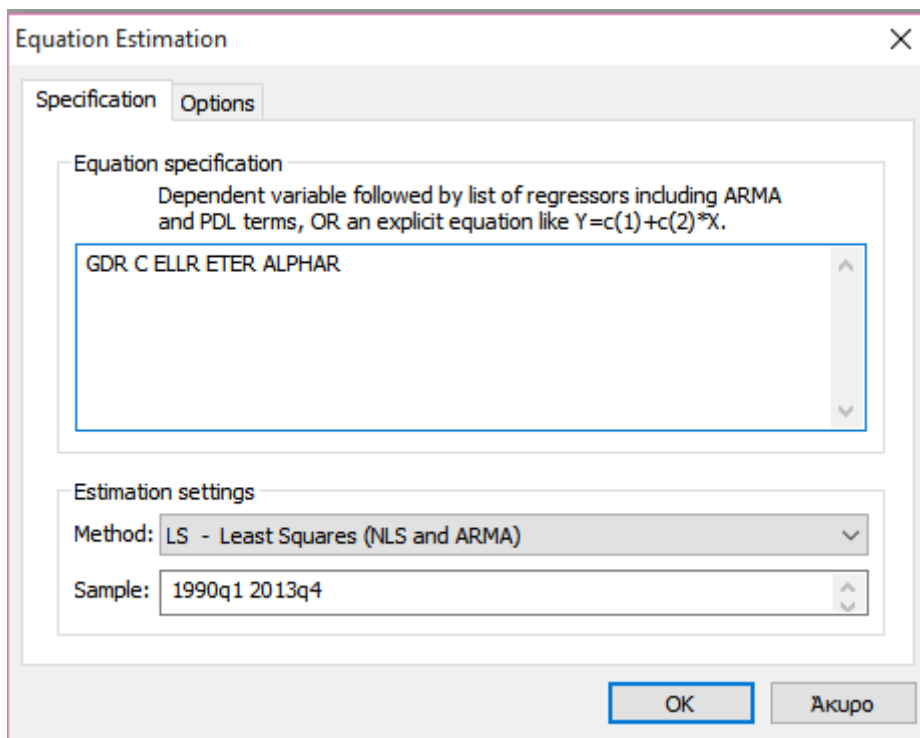
Εικόνα11 Εισαγωγή δεδομένων

Χρονοσειρές οπότε επιλέγουμε το πεδίο Dated-regularfrequency και στην συνέχεια όσον αφορά την συχνότητα των παρατηρήσεων επιλέγουμε Quarterly αφού πρόκειται για τριμηνιαία δεδομένα. Εφόσον έχουμε συγκεκριμένο εύρος δεδομένων γράφουμε το εύρος των ημερομηνιών που θα χρησιμοποιηθούν με την μορφή που φαίνεται στην ανωτέρω εικόνα και είναι της μορφής Χρόνος-Q-τρίμηνο, δηλαδή ο όρος 2013Q4 για παράδειγμα αναφέρεται στο τέταρτο τρίμηνο του 2013.

Στην συνέχεια εκτιμούμε το πολλαπλό μοντέλο παλινδρόμησης:

$$GDR = \alpha + \beta_1 * ELLR + \beta_2 * ETER + \beta_3 * ALPHAR + u_i$$

Επιλέγοντας Quick- Estimate Equation, όπου μας εμφανίζεται το παρακάτω παράθυρο:



Εικόνα.12 Εκτίμηση εξίσωσης παλινδρόμησης

Στο πεδίο Equation Estimation, όπως ακριβώς κάναμε και στην απλή παλινδρόμηση, αναγράφουμε με τη σειρά την ονομασία της εξαρτημένης μεταβλητής (GDR), το σταθερό όρο (C) και στην συνέχεια, αφού εδώ πρόκειται για μοντέλο πολλαπλής παλινδρόμησης τις ανεξάρτητες μεταβλητές μας (εδώ ELLR, ETER, ALPHAR) και πατάμε το κουμπί OK.

Το Eviews στην συνέχεια μας δίνει τον παρακάτω πίνακα με τα αποτελέσματα της παλινδρόμησης.

Dependent Variable: GDR				
Method: Least Squares				
Date: 06/09/15 Time: 11:20				
Sample : 1 246				
Included observations: 246 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.060728	0.073932	0.821396	0.4122
ELLR	0.259370	0.046495	5.578437	0.0000
ETER	0.175567	0.026870	6.533969	0.0000
ALPHAR	0.037611	0.029076	1.293504	0.1971
R-squared	0.833504	Mean dependent var		-0.198924
Adjusted R-squared	0.828587	S.D. dependent var		2.776196
S.E. of regression	1.149403	Akaike info criterion		3.148470
Sum squared resid	313.1071	Schwarz criterion		3.262797
Log likelihood	-377.6876	Hannan-Quinn criter.		3.194510
F-statistic	169.4944	Durbin-Watson stat		2.049752
Prob(F-statistic)	0.000000			

Πίνακας7 Αποτελέσματα παλινδρόμησης

Στο παραπάνω πίνακα παρατηρούμε διάφορα πεδία τα οποία εμπεριέχουν διάφορες αριθμητικές τιμές. Από αυτά, τα πεδία που μας αφορούν περισσότερο είναι:

- **DependentVariable:** GDR , στο πεδίο αυτό παρουσιάζεται το όνομα της μεταβλητής
- **Method:** LeastSquares , αναφέρεται η μέθοδος εκτίμησης, εδώ είναι η μέθοδος ελαχίστων τετραγώνων
- **Date:** 06/09/15 **Time:** 11:20. Η ημερομηνία και η ώρα εμφάνισης των αποτελεσμάτων
- **Sample 1 246:** Το μέγεθος του δείγματος. Εδώ το δείγμα έχει 246 παρατηρήσεις.
- **Included observations:** Το δείγμα που χρησιμοποιήθηκε στην εκτίμηση. Εδώ είναι όλες οι τιμές του δείγματος
- **Variable:** Σε αυτή τη στήλη παρουσιάζονται τα ονόματα των μεταβλητών.
- **Coefficient:** Στη στήλη αυτή παρουσιάζονται οι εκτιμήσεις των συντελεστών
- **St. Error:** Στη στήλη αυτή παρουσιάζονται τα τυπικά σφάλματα των συντελεστών
- **t-Statistic:** Σε αυτή τη στήλη παρουσιάζεται ο έλεγχος της υπόθεσης $H_0: \beta_i=0$ για κάθε συντελεστή του υποδείγματος $t = \frac{\widehat{\beta}_i - 0}{S_{\widehat{\beta}_i}}$
- **Prob:** Η πιθανότητα λάθους αν απορρίψουμε την H_0 .
- **R-Squared:** Παρουσιάζεται ο συντελεστής προσδιορισμού $R^2 = \frac{\sum \hat{u}^2}{\sum (Y - \hat{Y})^2}$
- **Adjusted R-Squared** Διορθωμένος συντελεστής προσδιορισμού

$$\bar{R}^2 = 1 - (1 - R^2) \frac{T - 1}{T - k}$$
- **S.E. of regression** Υπολογίζεται το τυπικό σφάλμα της παλινδρόμησης $S = \sqrt{\frac{\sum \hat{u}^2}{T - k}}$
- **Sumsquared resid:** Υπολογίζεται το άθροισμα τετραγώνων των καταλοίπων $\sum \hat{u}^2$
- **Durbin-Watson stat:** Η στατιστική Durbin Watson $DW = \frac{\sum (\hat{u}_t - \hat{u}_{t-1})^2}{\sum \hat{u}_t^2}$
- **Meandependentvar:** Αριθμητικός Μέσος της εξαρτημένης μεταβλητής $\bar{Y} = \frac{\sum Y}{T}$
- **S.D. dependentvar:** Η τυπική απόκλιση της εξαρτημένης μεταβλητής

$$S = \sqrt{\frac{\sum(Y - \bar{Y})^2}{T - k}}$$

- F-statistic Έλεγχος ότι από κοινού συντελεστές είναι μηδέν $\beta_1 = \beta_2 = \dots = \beta_k = 0$
- Prob(Fstatistic) Η πιθανότητα λάθους αν απορριφθεί η υπόθεση ότι από κοινού συντελεστές είναι μηδέν $\beta_1 = \beta_2 = \dots = \beta_k = 0$

Από τα αποτελέσματα της παραπάνω παλινδρόμησης προκύπτει ότι το υπόδειγμα προσαρμόζεται πολύ καλά στα δεδομένα (R-squared = 0,83) πράγμα που σημαίνει ότι ερμηνεύεται το 83% της μεταβλητότητας των αποδόσεων του Γενικού Δείκτη (GDR). Επίσης παρατηρείται ότι η μεταβλητή ALPHAR δεν είναι στατιστικά σημαντική για το υπόδειγμα. Αντίθετα οι μεταβλητές ELLR, ETER είναι **στατιστικά σημαντικές** και έχουν τις εξής **θετικές** επιδράσεις πάνω στην απόδοση του ΓΔ:

-Για 1% αύξηση της απόδοσης της μετοχής της **ELLR** αυξάνεται η απόδοση του ΓΔ του ΧΑ κατά 0.259%. Με τον ίδιο τρόπο ερμηνεύονται και οι υπόλοιπες μεταβλητές δηλαδή:

-Για 1% της απόδοσης της μετοχής της **ETE** αυξάνεται η απόδοση του ΓΔ του ΧΑ κατά 0.170,175%.

Στο παρακάτω πίνακα φαίνονται κάποια περιγραφικά μέτρα των 3 ανεξάρτητων μεταβλητών που χρησιμοποιήθηκαν στην έως τώρα ανάλυση καθώς και της εξαρτημένης μεταβλητής.

	GDR	ELLR	ETER	ALPHAR
Mean	-0.198924	-0.186040	-0.506402	-0.577321
Median	-0.098763	-0.302115	0.000000	-0.313480
Maximum	10.13003	7.373737	17.21311	17.75701
Minimum	-14.64711	-11.12245	-34.13174	-35.00000
Std. Dev.	2.776196	2.167778	5.987259	5.927192
Skewness	-0.362760	-0.301532	-0.913142	-0.633521
Kurtosis	6.941425	6.327115	8.181236	7.581123
Jarque-Bera	163.9582	116.7158	308.0928	230.6276
Probability	0.000000	0.000000	0.000000	0.000000
Sum	-48.73640	-45.57990	-124.0684	-141.4437
Sum Sq. Dev.	1880.572	1146.619	8746.733	8572.112
Observations	245	245	245	245

Πίνακας 8 Περιγραφικά μέτρα μεταβλητών

Από το παραπάνω πίνακα βλέπουμε ότι οι μέσες αποδόσεις των τραπεζικών μετοχών αλλά και του ΓΔ είναι αρνητικές για το διάστημα που εξετάζεται ενώ επίσης βλέπουμε έντονες διαφοροποιήσεις ανάμεσα στις μεταβλητές για τις μέγιστες και ελάχιστες τιμές που έλαβαν ενώ παρατηρείται όπως είναι φυσικό και αντίστοιχη διαφοροποίηση στις τυπικές τους αποκλίσεις (ETER, ALPHAR). Αυτό θα μπορούσε να

είναι χρήσιμο για έναν επενδυτή ο οποίος θα ήθελε να δει πως αντιδρούν έως τώρα ορισμένες μετοχές που θα ήθελε να προσθέσει στο χαρτοφυλάκιο του. Υπάρχει αρνητική ασυμμετρία που μπορεί να οφείλεται είτε μακριά ουρά αριστερά ή αρνητικές τιμές. Επίσης βλέπουμε ότι οι τιμές των μεταβλητών δεν κατανέμονται κανονικά.

Στην συνέχεια ελέγχουμε αν οι ανεξάρτητες μεταβλητές που θα χρησιμοποιούνται για την επεξήγηση της συμπεριφοράς της εξαρτημένης μεταβλητής συνδέονται γραμμικά με την εξαρτημένη. Την ύπαρξη ή όχι μιας τέτοιας σχέσης μπορούμε να την δούμε είτε γραφικά σε διαγράμματα ενώ επιβεβαιώνεται στατιστικά μέσω των πινάκων συσχετίσεων. Και στις 3 μεταβλητές μας παρατηρούμε ισχυρή γραμμική σχέση μεταξύ των μεταβλητών. Από το πίνακα συσχετίσεων παρατηρούμε ότι η μεγαλύτερη γραμμική σχέση είναι αυτή των αποδόσεων της εθνικής τράπεζας και της απόδοσης του ΓΔ. (ETER και GDR με συντελεστή συσχέτισης 0,86) Ακολουθεί η ALPHAR (0,80) ενώ και η ELLR έχει αρκετά καλή σχέση (0,70).

	GDR
GDR	1
ETER	0.86
ELLR	0.70
EUROBR	0.80

Πίνακας9 Συσχετίσεις

Όσο μεγαλύτερη είναι η συσχέτιση τόσο μεγαλύτερο μέρος πληροφοριών της εξαρτημένης μεταβλητής εξηγεί η ανεξάρτητη. Ωστόσο, όταν οι ανεξάρτητες μεταβλητές είναι περισσότερες από μία είναι σκόπιμο, οι ανεξάρτητες μεταβλητές, μεταξύ τους να μην είναι ισχυρά συσχετισμένες. Όταν οι ανεξάρτητες μεταβλητές είναι συσχετισμένες επεξηγούν το ίδιο μέρος της διακύμανσης της εξαρτημένης μεταβλητής. Η συμμετοχή δύο ανεξάρτητων συσχετισμένων μεταβλητών αυξάνει στην ουσία το λάθος του μοντέλου.

4.4 Σωματικό λίπος

Στη συνέχεια ακολουθούν κάποιες εφαρμογές της απλής γραμμικής παλινδρόμησης αλλά και μια εφαρμογή πολλαπλής παλινδρόμησης με σκοπό να δούμε τι αποτελέσματα λαμβάνουμε από τις εφαρμογές αυτές και να κάνουμε άμεση σύγκριση τους με σκοπό να δούμε αν υπάρχουν διαφοροποιήσεις στα αποτελέσματα. Τα δεδομένα που συγκεντρώθηκαν αφορούν στην εξέταση του τρόπου με τον οποίο επηρεάζεται το σωματικό λίπος μιας ομάδας ατόμων σε σχέση με τη περίμετρο του μπράτσου και του μηρού τους. Με την χρήση της γραμμικής παλινδρόμησης (απλής και πολλαπλής) προσπαθούμε να ερμηνεύσουμε την εξαρτημένη μεταβλητή, βάση των τιμών κάποιων επεξηγηματικών (ανεξάρτητων) μεταβλητών. Εδώ η εξαρτημένη μεταβλητή είναι το ποσοστό του σωματικού λίπους και οι ανεξάρτητες μεταβλητές η περιφέρεια του μηρού και του μπράτσου. Το δείγμα αποτελείται από 20 άτομα από τα οποία λήφθηκαν τα παρακάτω στοιχεία:

	Μπράτσο(ARM)	Λίπος (FAT)	Μηρός (LEG)
Mean	27.60500	20.28500	49.79400
Median	27.80000	21.20000	49.75000
Maximum	37.00000	27.80000	58.58000
Minimum	21.30000	11.80000	42.20000
Std. Dev.	3.640774	5.124582	5.144428
Skewness	0.464866	-0.383050	0.048369
Kurtosis	3.467132	1.973566	1.708530
Jarque-Bera	0.902178	1.367065	1.397711
Probability	0.636934	0.504831	0.497154
Sum	552.1000	405.7000	995.8800
Sum Sq. Dev.	251.8495	498.9655	502.8377
Observations	20	20	20

Πίνακας 10 Περιγραφικά μέτρα μεταβλητών

Από τον παραπάνω πίνακα παρατηρούμε ότι όσον αφορά την μεταβλητή της περιμέτρου του μπράτσου ο μ.ο. είναι 27,6 με μέγιστη παρατηρούμενη τιμή τα 37 εκ. και ελάχιστη τα 21,3 εκ. Αντίστοιχα για το λίπος η μέση τιμή είναι 20,2% ποσοστό σωματικού λίπους

με ελάχιστο το 11,8% και μέγιστο το 27,8%. Τέλος για την περίμετρο του μηρού η μέση τιμή είναι 49,7 εκ. με μέγιστη τα 58,5 εκ. και ελάχιστη τα 42,2 εκ. Στη συνέχεια ακολουθεί ο πίνακας που μας φανερώνει την σχέση μεταξύ των μεταβλητών που συμμετέχουν στην ανάλυση.

	Μπράτσο	Λίπος	Μηρός
Μπράτσο	1	0.15	-0.44
Λίπος	0.15	1	-0.25
Μηρός	-0.44	-0.25	1

Πίνακας 11 Πίνακας συσχετίσεων

Από τον πίνακα συσχετίσεων παρατηρούμε ότι όσον αφορά τις ανεξάρτητες μεταβλητές δεν παρουσιάζουν έντονα θετική συσχέτιση με την εξαρτημένη μεταβλητή του Λίπους. Αντίθετα η μεταβλητή της περιφέρειας του μπράτσου είναι μόλις κατά 0,15 συσχετισμένη με αυτήν του σωματικού λίπους ενώ η μεταβλητή του μηρού έχει αρνητική συσχέτιση -0,25.

Όπως έχει ήδη αναφερθεί στην περίπτωση της απλής παλινδρόμησης, όσο μεγαλύτερη είναι η συσχέτιση τόσο μεγαλύτερο μέρος του τρόπου που μεταβάλλεται η εξαρτημένη μεταβλητή εξηγεί η ανεξάρτητη.

Στη συνέχεια παρουσιάζονται τα αποτελέσματα του πρώτου υποδείγματος που εξετάστηκε. Σε αυτό, όπως και σε κάθε υπόδειγμα που θα εξεταστεί στη συνέχεια, η εξαρτημένη μεταβλητή είναι το ποσοστό του σωματικού λίπους, ενώ η ανεξάρτητη μεταβλητή είναι η περίμετρος του μπράτσου (ARM).

Dependent Variable: FAT				
Method: Least Squares				
Date: 04/05/16 Time: 15:01				
Sample: 1 20				
Included observations: 20				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	14.42075	9.127919	1.579851	0.1316
ARM	0.212434	0.327963	0.647739	0.0253
R-squared	0.622778	Mean dependent var		20.28500
Adjusted R-squared	0.631512	S.D. dependent var		5.124582
S.E. of regression	5.204698	Akaike info criterion		6.231640
Sum squared resid	487.5999	Schwarz criterion		6.331213
Log likelihood	-60.31640	Hannan-Quinn criter.		6.251078
F-statistic	0.419565	Durbin-Watson stat		1.869979
Prob(F-statistic)	0.525330			

Πίνακας 12 Αποτελέσματα Παλινδρόμησης

Όπως παρατηρείται στο παραπάνω πίνακα, η μεταβλητή της περιφέρειας του μπράτσου (ARM) είναι στατιστικά σημαντική καθώς η τιμή pvalue για την μεταβλητή αυτή είναι 0,025 συνεπώς σε επίπεδο σημαντικότητας 5% απορρίπτεται η υπόθεση ότι η μεταβλητή αυτή είναι 0. Αυτό σημαίνει ότι η μεταβλητή αυτή μπορεί να μας προσφέρει χρήσιμη πληροφόρηση για το πώς επηρεάζεται η εξαρτημένη μεταβλητή του ποσοστού του σωματικού λίπους. Αντίθετα ο σταθερός όρος στο υπόδειγμα δεν είναι στατιστικά σημαντικός. Η μεταβλητή της περιφέρειας του μπράτσου (ARM) έχει τιμή 0,21 πράγμα που σημαίνει ότι για κάθε 1 εκατοστό αύξησης της περιφέρειας του μπράτσου αυξάνεται το ποσοστό του σωματικού λίπους (FAT) κατά 0,21%.

Αυτό είναι λογικό να συμβαίνει καθώς αναμένουμε πως μια αύξηση του ποσοστού του σωματικού λίπους θα επηρεάσει και την περιφέρεια του μπράτσου του ατόμου που εξετάζεται, με δεδομένο ότι το μπράτσο είναι ένα σημείο του σώματος στο οποίο γίνεται αποθήκευση λίπους.

Αξίζει να σημειωθεί ότι ο δείκτης R-Squared που αποτελεί ένδειξη του ποσοστού της διακύμανσης της εξαρτημένης μεταβλητής που εξηγείται από το μοντέλο παίρνει την τιμή 0,62. Αυτό σημαίνει ότι το μοντέλο που επιλέχθηκε επεξηγεί το 62% της συνολικής διακύμανσης του δείγματος. Αυτή η τιμή του συντελεστή προσδιορισμού σημαίνει ότι το

μοντέλο είναι μέτρια ικανοποιητικό για την περιγραφή της μεταβλητότητας της εξαρτημένης μεταβλητής. Συνεπώς μια προσθήκη μιας επιπλέον μεταβλητής ή μια τροποποίηση της μορφής του μοντέλου που εξετάζεται μπορεί να βελτιώσει τα αποτελέσματά του. Αυτό θα εξεταστεί στη συνέχεια.

Ακολούθως παρουσιάζονται τα αποτελέσματα του δεύτερου μοντέλου απλής γραμμικής παλινδρόμησης που εξετάζεται. Σε αυτό χρησιμοποιείται σαν εξαρτημένη μεταβλητή το ποσοστό του σωματικού λίπους (FAT), ενώ η ανεξάρτητη μεταβλητή είναι η περίμετρος του μηρού (LEG).

Dependent Variable: FAT				
Method: Least Squares				
Date: 04/05/16 Time: 15:03				
Sample: 1 20				
Included observations: 20				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	33.12687	11.34977	2.918726	0.0092
LEG	-0.257900	0.226788	-1.137187	0.2704
R-squared	0.067029	Mean dependent var	20.28500	
Adjusted R-squared	0.015197	S.D. dependent var	5.124582	
S.E. of regression	5.085494	Akaike info criterion	6.185301	
Sum squared resid	465.5206	Schwarz criterion	6.284874	
Log likelihood	-59.85301	Hannan-Quinn criter.	6.204739	
F-statistic	1.293195	Durbin-Watson stat	1.732023	
Prob(F-statistic)	0.270372			

Πίνακας 13 Αποτελέσματα παλινδρόμησης

Από τα αποτελέσματα του παραπάνω πίνακα βλέπουμε ότι η μεταβλητή της περιφέρειας του μηρού (LEG) είναι στατιστικά μη σημαντική καθώς η τιμή pvalueγια την μεταβλητή αυτή είναι 0,279 συνεπώς για κάθε επίπεδο σημαντικότητας δεν απορρίπτεται η υπόθεση ότι η μεταβλητή αυτή είναι 0. Αυτό σημαίνει ότι η μεταβλητή αυτή δεν μπορεί να μας προσφέρει χρήσιμη πληροφόρηση για το πώς επηρεάζεται η εξαρτημένη μεταβλητή του ποσοστού του σωματικού λίπους και κατά συνέπεια το υπόδειγμα συνολικά δεν μας είναι χρήσιμο. Αντίθετα ο σταθερός όρος στο υπόδειγμα είναι στατιστικά σημαντικός. Η μεταβλητή της περιφέρειας του μηρού (LEG)έχει τιμή -0,25 αλλά λόγω του γεγονότος

ότι δεν είναι στατιστικά σημαντική δεν έχει νόημα να προβούμε στην ερμηνεία της για το πώς επηρεάζει το ποσοστό του σωματικού λίπους (FAT).

Αξίζει να σημειωθεί ότι ο δείκτης R-Squared που αποτελεί ένδειξη του ποσοστού της διακύμανσης της εξαρτημένης μεταβλητής που εξηγείται από το μοντέλο παίρνει την τιμή 0,06. Αυτό σημαίνει ότι το μοντέλο που επιλέχθηκε επεξηγεί το 6% της συνολικής διακύμανσης του δείγματος. Αυτή η τιμή του συντελεστή προσδιορισμού σημαίνει ότι το μοντέλο είναι πολύ κακό για την περιγραφή της μεταβλητότητας της εξαρτημένης μεταβλητής. Συνεπώς πρέπει να τροποποιηθεί για να μπορέσει να βελτιώσει τα αποτελέσματά του.

Στα δυο μοντέλα απλής γραμμικής παλινδρόμησης που εξετάστηκαν τα αποτελέσματα είναι αντίθετα. Στην πρώτη περίπτωση υπάρχει στατιστική σημαντικότητα της ανεξάρτητης μεταβλητής και κατά συνέπεια και του μοντέλου ενώ στο δεύτερο τα αποτελέσματα δείχνουν ακριβώς το αντίθετο. Προκειμένου η μελέτη των δεδομένων αυτών να είναι πιο πλήρης στη συνέχεια θα ακολουθήσει η μελέτη ενός υποδείγματος πολλαπλής γραμμικής παλινδρόμησης.

Στο μοντέλο αυτό παραμένει ως εξαρτημένη μεταβλητή το ποσοστό του σωματικού λίπους (FAT), ενώ χρησιμοποιούνται δυο ανεξάρτητες μεταβλητές, η περίμετρος του μπράτσου (ARM) και η περίμετρος του μηρού (LEG).

Στην πολλαπλή παλινδρόμηση είναι προτιμότερο, οι ανεξάρτητες μεταβλητές, να μην είναι ισχυρά συσχετισμένες. Και αυτό γιατί θέλουμε να αποφύγουμε οι ανεξάρτητες μεταβλητές να επεξηγούν το ίδιο μέρος της διακύμανσης της εξαρτημένης μεταβλητής. Για τον λόγο αυτό πρέπει πριν την εξέταση του μοντέλου της γραμμικής παλινδρόμησης να εξετάζουμε τις συσχετίσεις των μεταβλητών. Από τον πίνακα 11 που παρουσιάστηκε προηγουμένως είδαμε ότι όσον αφορά τις ανεξάρτητες μεταβλητές δεν παρουσιάζουν έντονα θετική συσχέτιση με την εξαρτημένη μεταβλητή του Λίπους (FAT). Αντίθετα η μεταβλητή της περιφέρειας του μπράτσου (ARM) είναι μόλις κατά 0,15 συσχετισμένη με αυτήν του σωματικού λίπους (FAT) ενώ η μεταβλητή του μηρού (LEG) έχει αρνητική συσχέτιση -0,25.

Dependent Variable: FAT				
Method: Least Squares				
Date: 04/05/16 Time: 15:12				
Sample: 1 20				
Included observations: 20				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.648437	9.943836	0.769164	0.4530
ARM	0.433150	0.188533	-2.297481	0.0354
LEG	-0.017087	0.128463	-0.133012	0.8958
R-squared	0.657147	Mean dependent var		20.28500
Adjusted R-squared	0.659112	S.D. dependent var		5.124582
S.E. of regression	2.515162	Akaike info criterion		4.859408
Sum squared resid	101.2166	Schwarz criterion		5.058554
Log likelihood	-44.59408	Hannan-Quinn criter.		4.898283
F-statistic	20.95829	Durbin-Watson stat		2.320870
Prob(F-statistic)	0.000009			

Πίνακας124 Αποτελέσματα παλινδρόμησης

Στο παραπάνω πίνακα βλέπουμε τα αποτελέσματα στο μοντέλο που εξετάστηκε. Όπως παρατηρεί εύκολα κανείς, η μεταβλητή της περιφέρειας του μηρού (LEG) είναι στατιστικά μη-σημαντική καθώς η τιμή pvalueγια την μεταβλητή αυτή είναι 0,89 συνεπώς σε κάθε επίπεδο σημαντικότητας δεν απορρίπτεται η υπόθεση ότι η μεταβλητή αυτή είναι 0, άρα δεν μας βοηθά να ερμηνεύσουμε την εξαρτημένη μεταβλητή.

Αντίθετα η μεταβλητή της περιφέρειας του μπράτσου (ARM)είναι στατιστικά σημαντική σε επίπεδο σημαντικότητας 5% και άρα συνεισφέρει στο υπόδειγμα. Μάλιστα έχει την τιμή 0,43 πράγμα που σημαίνει ότι για κάθε 1 εκατοστό αύξησης της περιφέρειας του μηρού αυξάνεται το ποσοστό του σωματικού λίπους (FAT) κατά 0,43%.

Μια πιθανή εξήγηση γι' αυτή την θετική σχέση είναι ότι ίσως στο συγκεκριμένο σημείο του σώματος αποθηκεύεται εύκολα λίπος και συνεπώς αποτελεί ένδειξη για αύξηση του σωματικού λίπους μια αύξηση της περιφέρειας του μηρού.

Αξίζει να σημειωθεί ότι ο δείκτης R-Squaredπου αποτελεί ένδειξη του ποσοστού της διακύμανσης της εξαρτημένηςμεταβλητής που εξηγείταιαπό το μοντέλο παίρνει την τιμή 0,79. Αυτό σημαίνει ότι το μοντέλο που επιλέχθηκε επεξηγεί το 79% της συνολικής διακύμανσης του δείγματος. Αυτή η υψηλή τιμή του συντελεστή προσδιορισμού

σημαίνει ότι είναι ικανοποιητικό το μοντέλο για την περιγραφή της κίνησης της εξαρτημένης μεταβλητής.

4.4.1 Σύγκριση αποτελεσμάτων απλής και πολλαπλής παλινδρόμησης

Στην προηγούμενη ενότητα αναφέρθηκαν τρία παραδείγματα παλινδρομήσεων εκ των οποίων τα δυο είχαν την μορφή της απλής παλινδρόμησης ενώ το τρίτο περιείχε δυο ερμηνευτικές μεταβλητές (πολλαπλή παλινδρόμηση). Τα αποτελέσματα που προέκυψαν με την μέθοδο των ελαχίστων τετραγώνων διαφέρουν σημαντικά τόσο όσον αφορά τη στατιστική σημαντικότητά τους όσο και για τις τιμές που παίρνει η κάθε μεταβλητή στην απλή γραμμική παλινδρόμηση και στο πολλαπλό υπόδειγμα.

Ξεκινώντας από το πρώτο υπόδειγμα στο οποίο η εξαρτημένη μεταβλητή είναι το ποσοστό του σωματικού λίπους (FAT), ενώ η ανεξάρτητη μεταβλητή είναι η περίμετρος του μπράτσου (ARM) είδαμε ότι η μεταβλητή της περιφέρειας του μπράτσου (ARM) βρέθηκε στατιστικά σημαντική σε επίπεδο σημαντικότητας 5%.

Αυτό όπως ερμηνεύθηκε και στην αντίστοιχη ενότητα σημαίνει ότι η μεταβλητή αυτή μπορεί να μας προσφέρει χρήσιμη πληροφορία για το πώς επηρεάζεται η εξαρτημένη μεταβλητή του ποσοστού του σωματικού λίπους. Αντίθετα ο σταθερός όρος στο υπόδειγμα δεν είναι στατιστικά σημαντικός.

Όσον αφορά το συντελεστή προσδιορισμού R-Squared που αποτελεί ένδειξη του ποσοστού της διακύμανσης της εξαρτημένης μεταβλητής που εξηγείται από το μοντέλο είδαμε ότι το μοντέλο που επιλέχθηκε επεξηγεί το 62% της συνολικής διακύμανσης του δείγματος.

Στο δεύτερο υπόδειγμα απλής γραμμικής παλινδρόμησης χρησιμοποιήθηκε σαν ανεξάρτητη μεταβλητή η περίμετρος του μηρού (LEG) ενώ εξαρτημένη μεταβλητή παραμένει το ποσοστό του σωματικού λίπους (FAT). Τα αποτελέσματα του υποδείγματος αυτού όπως είδαμε παραπάνω φανερώνουν πως η μεταβλητή της

περιφέρειας του μηρού (LEG) είναι στατιστικά μη σημαντική καθώς για κάθε επίπεδο σημαντικότητας. Αυτό σημαίνει πως συνολικά το μοντέλο αυτό δεν μπορεί να μας προσφέρει χρήσιμη πληροφόρηση για το πώς επηρεάζεται η εξαρτημένη μεταβλητή του ποσοστού του σωματικού λίπους. Επίσης ο συντελεστής R-Squared έχει τιμή 0,06 και είναι ιδιαίτερα χαμηλός.

Έχοντας κάνει αναφορά στα αποτελέσματα των δυο υποδειγμάτων αξίζει να αναφέρουμε τα αποτελέσματα του υποδείγματος πολλαπλής παλινδρόμησης προκειμένου να εντοπιστούν οι διαφορές που υπάρχουν. Όπως αναφέρθηκε στη προηγούμενη ενότητα παραμένει ως εξαρτημένη μεταβλητή το ποσοστό του σωματικού λίπους (FAT), ενώ χρησιμοποιούνται δυο ανεξάρτητες μεταβλητές, η περίμετρος του μπράτσου (ARM) και η περίμετρος του μηρού (LEG).

Τα αποτελέσματα αυτής φανερώνουν πως η μεταβλητή της περιφέρειας του μηρού (LEG) είναι στατιστικά μη-σημαντική για κάθε επίπεδο σημαντικότητας. Αυτό έρχεται σε συμφωνία με τα αποτελέσματα της εκτίμησης του απλού υποδείγματος όπου και εκεί τα αποτελέσματα έδειξαν ότι η μεταβλητή της περιφέρειας του μηρού είναι στατιστικά μη-σημαντική.

Αντίθετα η μεταβλητή της περιφέρειας του μπράτσου (ARM) είναι στατιστικά σημαντική σε επίπεδο σημαντικότητας 5% και στα δυο υποδείγματα.

Όσον αφορά την τιμή που λαμβάνει στα δυο υποδείγματα προκύπτει ότι και στις δυο περιπτώσεις έχει θετικό πρόσημο κάτι που σημαίνει ότι όσο αυξάνεται η περιφέρεια του μπράτσου αυξάνεται κατά ένα ποσοστό και το ποσοστό του σωματικού λίπους. Η τιμή της μεταβλητής στο απλό υπόδειγμα ήταν 0,21 ενώ στην πολλαπλή παλινδρόμηση 0,43. Αυτό σημαίνει ότι στο υπόδειγμα με τις δυο μεταβλητές φαίνεται να διπλασιάζεται η επίδραση της ανεξάρτητης μεταβλητής πάνω στην εξαρτημένη.

Όσον αφορά το συντελεστή προσδιορισμού R-Squared στο απλό υπόδειγμα έχει τιμή 0,62 (62%) ενώ στο υπόδειγμα με τις 2 μεταβλητές έχει τιμή 0,65 (65%). Η τιμή παρατηρείται ότι έχει αυξηθεί ελαφρώς στο υπόδειγμα με τις 2 μεταβλητές αλλά η διαφορά είναι μικρή.

Συνολικά μπορούμε να πούμε ότι τα αποτελέσματα παραμένουν σταθερά μεταξύ της απλής και της πολλαπλής παλινδρόμησης καθώς τόσο η στατιστική σημαντικότητα των ερμηνευτικών μεταβλητών και στις δυο περιπτώσεις παραμένει σταθερή ενώ και το πρόσημο της μεταβλητής περιφέρειας του μπράτσου (ARM) παραμένει σταθερή. Το μόνο που παρουσιάζει αξιοσημείωτη αλλαγή είναι η τιμή της μεταβλητής η οποία σχεδόν διπλασιάζεται στη περίπτωση του πολλαπλού υποδείγματος.

Συμπεράσματα

Στην εργασία αυτή γίνεται μια προσπάθεια για ανάλυση της θεωρίας που αφορά στις παλινδρομήσεις και στο τρόπο που αξιοποιούνται στις διάφορες οικονομικές εφαρμογές.

Τα διάφορα είδη παλινδρομήσεων όπως γίνεται εύκολα αντιληπτό αποτελούν σπουδαία εργαλεία για την εύρεση των σχέσεων που υπάρχουν μεταξύ των διάφορων οικονομικών δεδομένων και την αποτύπωση αυτών. Με αυτό τον τρόπο ο κάθε ερευνητής μπορεί να ερμηνεύσει τον τρόπο με τον οποίο κινούνται στο χρόνο οι μεταβλητές που τον ενδιαφέρουν.

Οι δυνατότητες εξειδίκευσης των υποδειγμάτων είναι πάρα πολλές και αυτό κάνει την επιλογή του κατάλληλου μοντέλου για την κάθε περίπτωση ιδιαίτερα κρίσιμη. Επίσης πέρα από την λήψη αποτελεσμάτων που ερμηνεύουν την έως τώρα σχέση των εκάστοτε ερμηνευτικών και εξαρτημένων μεταβλητών είναι δυνατόν ο ερευνητής να κάνει και πρόβλεψη για τον μελλοντικό τρόπο με τον οποίο αναμένεται να κινηθεί η υπό εξέταση μεταβλητή.

Σε αυτή την εργασία πέρα από την θεωρητική προσέγγιση τόσο της απλής όσο και της πολλαπλής παλινδρόμησης γίνεται και εκτίμηση ενός σε κάθε περίπτωση μοντέλου μέσω του οικονομετρικού πακέτου Eviewsto οποίο βοηθά στην καλύτερη αντίληψη της θεωρίας και της σημαντικότητας των ελέγχων που πρέπει ο ερευνητής να εξετάζει προκειμένου να λαμβάνονται κάθε φορά αξιόπιστα αποτελέσματα.

Στα αποτελέσματα των υποδειγμάτων που γίνεται εκτίμηση πραγματοποιείται και σύγκριση προκειμένου να βρεθούν τυχόν διαφορές μεταξύ απλής και πολλαπλής παλινδρόμησης. Κάτι τέτοιο δεν εντοπίζεται τουλάχιστον σε σημαντικό βαθμό και μπορούμε να πούμε πως τα αποτελέσματα παρουσιάζουν σταθερότητα.

Βιβλιογραφία

1. ΑΤΕΙ Πατρών, Τμήμα Πληροφορικής στην οικονομία και στη διοίκηση, Σημειώσεις Οικονομετρίας.
2. Βενέτης Α. Ιωάννης (2009), Εισαγωγικές διαλέξεις στην Οικονομετρία.
3. Boutsikas M.V. (2004), Σημειώσεις μαθήματος «Στατιστικά Προγράμματα» Τμήμα Στατ. & Ασφ. Επιστήμης, Πανεπιστήμιο Πειραιώς.
4. Bruce e. Hansen (2010), Econometrics, University of Wisconsin.
5. Δρ. Γούλας Ελευθέριος (2010), Μακροοικονομική, Σημειώσεις.
6. Γρηγοριάδου Π. Μαρία (2014), Παραβιάσεις των βασικών υποθέσεων του γραμμικού μοντέλου παλινδρόμησης.
7. Διαμαντόπουλος Επαμεινώνδας, ΙΕΚ Ξάνθης – Ειδικότητα Μηχανογραφημένου Λογιστηρίου Σημειώσεις για το μάθημα Στατιστική II.
8. Καραμάνης Ι. Δημήτριος (2005), Τμήμα Οργάνωσης και Διοίκησης Επιχειρήσεων Πανεπιστήμιο Πειραιά, Μελέτη υποδειγμάτων ARCH και εφαρμογή στον πληθωρισμό της Ελλάδος.
9. Καρβέλης Χαράλαμπος (2008), Ανάλυση χρονολογικών σειρών: προβλέποντας το μέλλον, κατανοώντας το παρελθόν.
10. Κιουφεντζή Ο. (2006), Το Κλασσικό Πολλαπλό Γραμμικό Μοντέλο Παλινδρόμησης – Στατιστικά Συμπεράσματα και Εκτιμήσεις.

11. Κυριαζόπουλος-Τσέλης (2012), Απασχόληση και ΑΕΠ στην Ελλάδα: Μια εμπειρική προσέγγιση.
12. Κύριος Θ. Ηλίας Θ. (2010), Μελέτη για εκτίμηση πολυμεταβλητών στην ανάλυση χρονοσειρών.
13. Παναγιώτου Σ. Γεώργιος Σ. (2005), Πωλήσεις Ι.Χ. Αυτοκινήτων σε δεκαπέντε χώρες-μέλη της Ευρωπαϊκής Ένωσης, Πανεπιστήμιο Πειραιώς.
14. JackJohnston- JohnDinardo (2004), Οικονομετρικές Μέθοδοι, επιμ. Βενέτης Ιωάννης, μτφ. Παπαϊωάννου Τρισεύγενη.

Ηλεκτρονική Βιβλιογραφία:

1. <https://research.stlouisfed.org/> Federal Reserve Economic Data
2. <http://ec.europa.eu/eurostat/data/database> Βάση δεδομένων Eurostat