

ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΔΥΤΙΚΗΣ ΕΛΛΑΔΑΣ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΜΕΘΟΔΟΙ ΠΡΟΒΛΕΨΕΩΝ



ΓΑΣΤΕΡΑΤΟΣ ΙΩΑΝΝΗΣ

ΣΠΑΤΟΥΛΑΣ ΠΕΤΡΟΣ

ΣΚΛΑΠΑΝΗΣ ΓΕΩΡΓΙΟΣ

ΕΠΟΠΤΕΥΩΝ ΚΑΘΗΓΗΤΗΣ: ΦΩΤΕΙΝΟΠΟΥΛΟΣ ΜΙΧΑΗΛ-ΑΝΑΣΤΑΣΙΟΣ

ΠΑΤΡΑ 2017

ΠΕΡΙΛΗΨΗ

Βασικός στόχος σχεδόν κάθε ανθρώπινης δραστηριότητας είναι η διενέργεια προβλέψεων, δηλαδή η χρησιμοποίηση ενός εκτιμώμενου μοντέλου για την πρόβλεψη των μελλοντικών τιμών των μεγεθών. Η πρόβλεψη είναι απαραίτητη γιατί αποτελεί το βασικό εργαλείο για κάθε μελλοντική εξέλιξη και απόφαση.

Παράλληλα οι προβλέψεις που γίνονται πρέπει να είναι έγκυρες και όσο πιο κοντά στο μέλλον που θα προκύψει. Η ποιότητα της πρόβλεψης εξαρτάται κυρίως από τον τρόπο που συλλέχτηκαν και αναλύθηκαν οι πληροφορίες. Έτσι, η ανάγκη για έγκυρες, άρα και πρακτικά χρήσιμες προβλέψεις οδήγησε στην ανάπτυξη πολλών τεχνικών και μεθόδων συλλογής και επεξεργασίας διαφόρων πρωτογενών πληροφοριών.

Τέτοιοι μέθοδοι, λοιπόν, προβλέψεων μπορεί να είναι ποιοτικές ή ποσοτικές. Με τις ποιοτικές μεθόδους ασχολούνται επιστήμες όπως η φιλοσοφία, θεολογία, οικονομία, πολιτικές επιστήμες κ.α., οι οποίες για την ανάλυση των δεδομένων χρησιμοποιούν την ‘ανθρώπινη’ λογική και τον τρόπο με τον οποίο οι διάφορες σχολές ανάλυσης αντιλαμβάνονται τον φυσικό κόσμο. Με τον όρο ‘ανθρώπινη’ λογική εννοείται κάθε απόφαση που λαμβάνει κανείς μη χρησιμοποιώντας μαθηματικές εξισώσεις. Για παράδειγμα, πρόβλεψη, η καλύτερα εκτίμηση, για την εξέλιξη μιας γλώσσας, της πολιτικής κατάστασης μιας χώρας, την πορεία της Ευρωπαϊκής Ένωσης και άλλα, είναι αποτέλεσμα λογικών διεργασιών σε διάφορα φόρουμ σκέψης. Σε όλες τις παραπάνω προβλέψεις δεν χρειάζεται καμία μαθηματική επεξεργασία. Αντίθετα, στις ποσοτικές μεθόδους η ανάλυση των δεδομένων γίνεται με την βοήθεια μαθηματικών εκφράσεων υποστηριζόμενες από υπολογιστές.

SUMMARY

The main goal almost every human activity is forecasting, that the use of an estimation model to predict future sizes price. The provision is necessary because it is the basic tool for future development and decision.

While predictions made must be accurate and as close to future arising. The quality of the forecast depends mainly on how the information collected and analyzed. Thus, the need for sound, and therefore practically useful predictions led to the development of many techniques for the collection and processing methods of various primary information.

Such methods are therefore estimates may be qualitative or quantitative. With qualitative methods involved disciplines such as philosophy, theology, economics, political science, etc., which for the analysis of data using the 'human' logic and the way in which various analytical faculties perceive the physical world. The term 'human' logic means any decision taken by someone not using mathematical equations. For example, forecasting, the best estimate for the development of a language, the political situation of a country, the European Union and other course is the result of logical thought processes in various forums. In all the above estimates no mathematical treatment needs. Instead, quantitative methods of data analysis is done with the aid of mathematical expressions supported computers.

ΕΙΚΟΝΕΣ

Εικόνα 1 - Κλάδοι οικονομετρίας.....	10
Εικόνα 2 Διαδικασία οικονομικής ανάλυσης	11
Εικόνα 3 - Υπόδειγμα συστήματος εξισώσεων	19
Εικόνα 4 -Πληθυσμιακή Καμπύλη Παλινδρόμησης	26
Εικόνα 5 Δειγματική Συνάρτηση της Παλινδρόμησης.....	31
Εικόνα 6 Δειγματική Συνάρτηση της Παλινδρόμησης.....	32
Εικόνα 7 Διαγράμματα Συναρτήσεων Απωλειών	34
Εικόνα 8 Διαγράμματα Συναρτήσεων Απωλειών	34
Εικόνα 9 Οι υποθέσεις του κλασσικού γραμμικού υποδείγματος παλινδρόμησης.....	39
Εικόνα 10 Οι υποθέσεις του κλασσικού γραμμικού υποδείγματος παλινδρόμησης.....	40
Εικόνα 11 Ομοσκεδαστικότητα.....	40
Εικόνα 12 Ετεροσκεδαστικότητα	41
Εικόνα 13 Τα Διαγράμματα Διασποράς (scatter plots)	92

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ	2
SUMMARY	3
ΕΙΚΟΝΕΣ.....	4
ΚΕΦΑΛΑΙΟ 1.....	8
ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΟΙΚΟΝΟΜΕΤΡΙΑ.....	8
1.1 Ορισμός της Οικονομετρίας.....	8
1.2 Αναδρομή σε Ιστορικά Στοιχεία	9
1.3 Οι Κλάδοι της Οικονομετρίας.....	10
1.4 Οικονομετρική Ανάλυση: η Διαδικασία της.....	11
1.5 Οικονομετρική ανάλυση: Αναφορά στα στάδια της.....	12
1.6 Υποδείγματα Οικονομετρίας.....	15
1.6.1 Υποδείγματα μιας εξισώσεως (single equation models)	15
1.6.2 Υποδείγματα πολλών εξισώσεων (simultaneous equation models).....	16
1.6.3 Παραδείγματα	17
1.6.4 Διάκριση ανάμεσα στις μεταβλητές ενός οικονομετρικού συστήματος εξισώσεων	18
1.6.5 Παράδειγμα εφαρμογής ενός οικονομετρικού υποδείγματος.....	19
ΚΕΦΑΛΑΙΟ 2.....	21
ΠΑΛΙΝΔΡΟΜΗΣΗ- ΟΙΚΟΝΟΜΕΤΡΙΚΑ ΥΠΟΔΕΙΓΜΑΤΑ	21
2.1 Τι είναι η παλινδρόμηση ;	21
2.2 «Νόμος» του Galton : Η Ερμηνεία του.....	22
2.3 Παλινδρόμηση- αιτιότητα	22
2.4 Παλινδρόμηση- συσχέτιση.....	23
2.5 Βασικές Έννοιες στην Απλή Παλινδρόμηση	24
2.6 Πληθυσμιακή Καμπύλη Παλινδρόμησης	26
2.7 Η Πληθυσμιακή Συνάρτηση της Παλινδρόμησης (population regression function PRF)	26
2.8 Στοχαστικές διαταραχές (Stochastic Disturbances).....	27

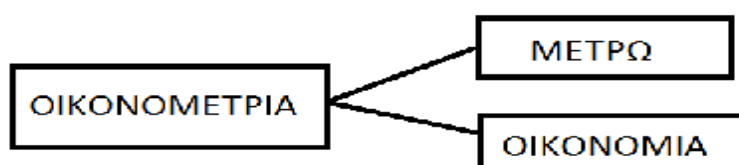
2.9	Η Δειγματική Συνάρτηση της Παλινδρόμησης.....	29
2.10	Απλή Παλινδρόμηση- Εκτίμηση.....	32
2.11	Διαγράμματα Συναρτήσεων Απωλειών	34
2.12	Οι υποθέσεις του κλασσικού γραμμικού υποδείγματος παλινδρόμησης.....	37
2.13	Εφαρμογές.....	42
ΚΕΦΑΛΑΙΟ 3.....		46
ΓΕΝΙΚΑ ΣΤΟΙΧΕΙΑ ΣΧΕΤΙΚΑ ΜΕ ΤΙΣ ΠΡΟΒΛΕΨΕΙΣ.....		46
3.1	Εισαγωγή.....	46
3.2	Βασικά Βήματα Προβλέψεων.....	47
3.3	Τα Πεδία Εφαρμογής των Προβλέψεων	49
ΚΕΦΑΛΑΙΟ 4.....		52
ΤΕΧΝΙΚΕΣ ΠΡΟΒΛΕΨΕΩΝ.....		52
4.1	Τεχνικές Προβλέψεων.....	52
4.2	Κατηγορίες των Μεθόδων Πρόβλεψης.....	52
4.2.1	Η Ποσοτική Πρόβλεψη (quantitative forecasting)	52
4.2.2	Η Ποιοτική Πρόβλεψη (qualitative ή judgmental forecasting) 56	
4.3	Οι Μέθοδοι Πρόβλεψης Συνεχούς Ζήτησης.....	58
4.3.1	Η Απλοϊκή μέθοδος – Naive.....	59
4.3.2	Οι Μέθοδοι Μέσου Όρου	59
4.3.3	Οι Μέθοδοι Εκθετικής Εξομάλυνσης.....	62
4.3.4	Η Απλή Γραμμική Παλινδρόμηση.....	75
4.4	Οι Δείκτες Αξιολόγησης Προβλέψεων	77
4.4.1	Η Βασική Στατιστική Ανάλυση.....	78
4.4.2	Η Στατιστική Ανάλυση Ακρίβειας Προβλέψεων	81
4.4.3	Ο Ρυθμός Ανάπτυξης.....	85
ΚΕΦΑΛΑΙΟ 5.....		86
ΟΙ ΧΡΟΝΟΣΕΙΡΕΣ.....		86
5.1	Η Έννοια της Χρονοσειράς.....	86

5.2 Τα Ποιοτικά Χαρακτηριστικά της Χρονοσειράς.....	87
5.3 Η Αναπαράσταση των Χρονοσειρών.....	89
ΣΥΜΠΕΡΑΣΜΑΤΑ	93
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	95

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΟΙΚΟΝΟΜΕΤΡΙΑ

1.1 Ορισμός της Οικονομετρίας



Οικονομετρία: Κλάδος της οικονομικής επιστήμης που εφαρμόζει τις μεθόδους της οικονομικής ανάλυσης σε στατιστικά δεδομένα για την διάγνωση και ερμηνεία ενός οικονομικού φαινομένου.

Η Οικονομετρία εφαρμόζεται σε όλο το φάσμα των κοινωνικών επιστημών όπως Οικονομικά, Διοίκηση Επιχειρήσεων, Λογιστική, Χρηματοοικονομικά, Μάρκετινγκ, Ψυχολογία, Κοινωνιολογία, Πολιτική Επιστήμη, κλπ.

Ο παραπάνω ορισμός της Οικονομετρίας (Οικονομία + Μέτρηση) δε συνεπάγεται ότι κάθε μέτρηση οικονομικού μεγέθους αποτελεί αντικείμενο της Οικονομετρίας. Π.χ. η μέτρηση του Εθνικού εισοδήματος, ή η κατασκευή και μέτρηση του δείκτη τιμών καταναλωτή δεν υπάγονται στα οικονομετρικά προβλήματα.

Ο έλεγχος των διαφόρων οικονομικών θεωριών και η σύγκριση μεταξύ τους χρησιμοποιώντας πραγματικά στοιχεία δεν είναι ο μοναδικός σκοπός της Οικονομετρίας. Με την οικονομετρική έρευνα μπορούμε να δημιουργήσουμε νέες θεωρίες, οι οποίες, μετά από συνεχείς εφαρμογές και επαληθεύσεις, να γίνουν γενικά παραδεκτές.

1.2 Αναδρομή σε Ιστορικά Στοιχεία

1926: Χρησιμοποιήθηκε για πρώτη φορά ο όρος «Οικονομετρία» από τον Ragnar Frisch.

1930: Δημιουργήθηκε η «Econometric Society»

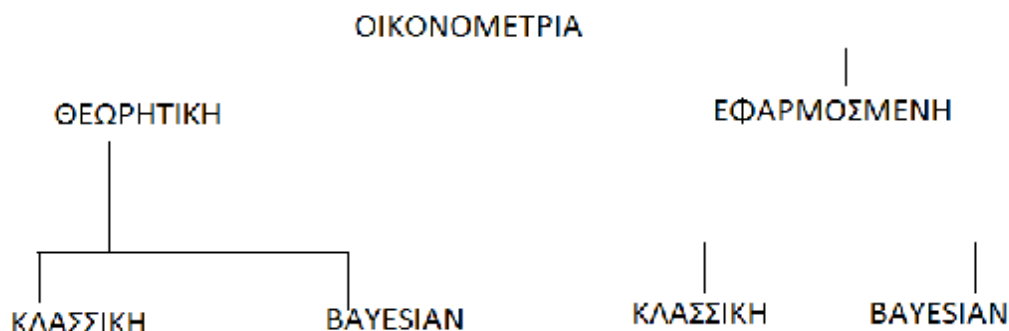
1933: Εκδίδεται το επιστημονικό περιοδικό «Econometrica» από την «Econometric Society»

1933: Irving Fisher : Υποστήριξε ότι η Οικονομική Επιστήμη θα καταστεί μια πλήρης επιστήμη όταν θα μπορεί να προβλέπει την εμφάνιση των οικονομικών κρίσεων.

1936: Η Κοινωνία των Εθνών ανέθεσε στον J. Tinbergen να εκπονήσει μελέτη σχετικά με τις οικονομικές κρίσεις και τις δυνατότητες ακριβούς πρόβλεψης τους.

1940: Haavelmo: Υποστήριξε ότι η συναρτησιακή σχέση μεταξύ δύο ή περισσότερων μεταβλητών δεν είναι ακριβής αλλά στοχαστική.

1.3 Οι Κλάδοι της Οικονομετρίας

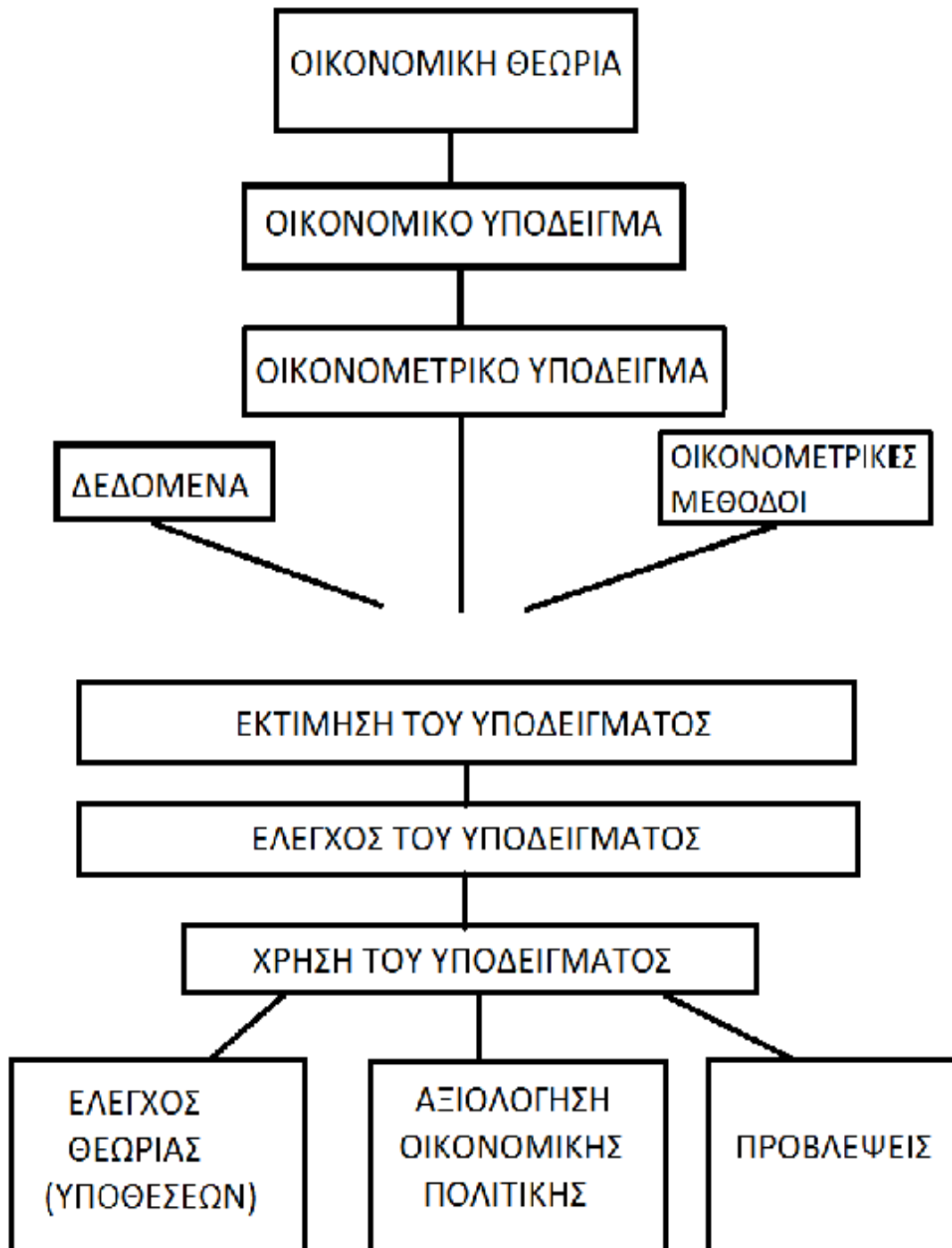


Εικόνα 1 - Κλάδοι οικονομετρίας

Η θεωρητική οικονομετρία έχει ως αντικείμενο την ανάπτυξη της μεθοδολογίας για την εκτίμηση των παραμέτρων των οικονομετρικών υποδειγμάτων και στηρίζεται στη μαθηματική στατιστική.

Η εφαρμοσμένη οικονομετρία χρησιμοποιεί τις μεθόδους της θεωρητικής οικονομετρίας για τη μελέτη ειδικών θεμάτων των οικονομικών επιστημών. Παραδείγματα που μπορούν να αναφερθούν είναι η συνάρτηση παραγωγής, η συνάρτηση κατανάλωσης, η συνάρτηση ζητήσεως, η συνάρτηση προσφοράς κλπ.

1.4 Οικονομετρική Ανάλυση: η Διαδικασία της



Εικόνα 2 Διαδικασία οικονομικής ανάλυσης

1.5 Οικονομετρική ανάλυση: Αναφορά στα στάδια της

Στάδιο 1^ο : Διατυπώνεται μια υπόθεση ή πρόταση με βάση την οικονομική θεωρία. Η διατύπωση αυτή είναι ποιοτικού και όχι ποσοτικού χαρακτήρα.

Παραδείγματα:

1. Κατά τον Keynes υφίσταται ένας θεμελιώδης ψυχολογικός νόμος σύμφωνα με τον οποίο οι άνθρωποι, κατά κανόνα και κατά μέσο όρο, έχουν την τάση να αυξάνουν την καταναλωτική τους δαπάνη καθώς το εισόδημά τους αυξάνεται, αλλά κατά ποσό μικρότερο από το ποσό που αυξάνει το εισόδημά τους.
2. Η ζητούμενη ποσότητα ενός αγαθού εξαρτάται κυρίως από την τιμή του, τις τιμές υποκατάστατων αγαθών, τις τιμές συμπληρωματικών αγαθών, το εισόδημα, τις προτιμήσεις των καταναλωτών, τη διανομή του εισοδήματος, το μέγεθος των περιουσιακών στοιχείων, τις προσδοκίες των καταναλωτών αναφορικά με τις τιμές και το εισόδημα κλπ.

Στάδιο 2^ο : Με τα οικονομικά μαθηματικά η οικονομική θεωρία διατυπώνεται με τη μορφή μαθηματικών εξισώσεων. Αρχικά γίνεται η επιλογή των οικονομικών μεταβλητών που αφορούν τη συγκεκριμένη οικονομική θεωρία που μελετάμε και στη συνέχεια καθορίζεται η μορφή των συναρτησιακών σχέσεων που συνδέουν τις μεταβλητές αυτές. Έτσι δημιουργείται ένα μαθηματικό υπόδειγμα που είναι καθαρά αιτιοκρατικό. Στο στάδιο αυτό δεν μας απασχολεί ακόμη η δυνατότητα, ή μη, μετρήσεως των μεταβλητών που υπεισέρχονται στο υπόδειγμα, καθώς και η διαθεσιμότητα των μεταβλητών αυτών. Επιπλέον δεν είναι δυνατό να ελεγχθεί και να επιβεβαιωθεί εμπειρικά η οικονομική θεωρία. Η

οικονομική ανάλυση εξακολουθεί να παραμένει περισσότερο ποιοτικής φύσεως.

Παραδείγματα:

1. $Y = \beta_1 + \beta_2 X$, $0 < \beta_2 < 1$, όπου

Y = καταναλωτικές δαπάνες

X = εισόδημα μετά από φόρους (διαθέσιμο εισόδημα)

B_2 = οριακή ροπή για κατανάλωση, δηλαδή η μέση μεταβολή των καταναλωτικών δαπανών που αντιστοιχούν στη μοναδιαία αύξηση του (διαθέσιμου) εισοδήματος^(*)

(*) Marginal propensity to consume (MPC)

2. Η συνάρτηση ζήτησεως ενός αγαθού που ζητήσαμε προηγουμένως μπορεί να εκφρασθεί με ένα γραμμικό-προσθετικό υπόδειγμα όπως παρακάτω:

$$Q = f(p, p_s, p_c, y) = \beta_0 + \beta_1 p + \beta_2 p_s + \beta_3 p_c + \beta_4 y$$

(-) (+) (-) (+)/(-)

Για τον καθορισμό των πρόσημων των συντελεστών $\beta_1, \beta_2, \beta_3, \beta_4$ μπορούμε να χρησιμοποιήσουμε την οικονομική θεωρία, όμως (ακόμη) δεν μπορούμε να έχουμε τις αριθμητικές τιμές τους. (Στην παραπάνω σχέση p =τιμή, p_s = τιμή υποκατάστατων αγαθών, p_c = τιμή συμπληρωματικών αγαθών, y =εισόδημα).

Στάδιο 3^ο : Η προηγούμενη μαθηματική σχέση που αφορά την κατανάλωση υπονοεί ότι (καθ' ότι αιτιοκρατική) για κάθε επίπεδο διαθέσιμου εισοδήματος αντιστοιχεί ένα και μόνο επίπεδο καταναλώσεως. Μια τέτοια συμπεριφορά δεν ανταποκρίνεται ποτέ σε πραγματικά δεδομένα. Οι παρατηρούμενες αποκλίσεις από τις προβλέψεις της αιτιοκρατικής σχέσης των μαθηματικών οικονομικών

μπορούν να ληφθούν υπόψη με την εισαγωγή στο υπόδειγμα και μιας τυχαίας μεταβλητής. Έτσι μια αιτιοκρατική σχέση μετατρέπεται σε στοχαστική. Με τη στοχαστική θεώρηση των οικονομικών σχέσεων τα υποδείγματα μετατρέπονται σε οικονομετρικά. Πέραν όμως της εισαγωγής στοχαστικών όρων στα υποδείγματα, σε ένα οικονομετρικό υπόδειγμα λαμβάνεται υπόψη η διαθεσιμότητα και η αξιοπιστία των δεδομένων που πρόκειται να χρησιμοποιηθούν. Η συλλογή, επεξεργασία και παρουσίαση των οικονομικών δεδομένων που χρησιμοποιεί η οικονομετρία, αποτελούν αντικείμενο της οικονομικής στατιστικής. Τέτοια δεδομένα είναι π.χ. μεταξύ άλλων και τα GNP, απασχόληση, τιμές, επενδύσεις, εισαγωγές, εξαγωγές, FDI (ξένες άμεσες επενδύσεις) κλπ. Για τους τύπους και τα χαρακτηριστικά των οικονομετρικών υποδειγμάτων γίνεται εκτενής αναφορά σε επόμενη ενότητα. Εδώ αναφέρουμε απλώς ότι οι εξισώσεις που απαρτίζουν ένα οικονομετρικό υπόδειγμα και μπορεί να είναι μια ή περισσότερες, ονομάζονται διαρθρωτικές εξισώσεις (structural equations) γιατί εξηγούν τη βασική διάρθρωση ενός οικονομικού συστήματος.

Στάδιο 4^ο : Αφού έχει καθοριστεί η διαρθρωτική δομή ενός οικονομετρικού υποδείγματος και είναι διαθέσιμα τα απαιτούμενα στοιχεία, γίνεται η εκτίμηση των παραμέτρων του υποδείγματος με τη χρήση κατάλληλων οικονομετρικών μεθόδων. Όπου απαιτείται γίνεται προσαρμογή των κλασσικών μεθόδων της μαθηματικής στατιστικής στα ιδιαίτερα χαρακτηριστικά των οικονομικών σχέσεων, λαμβάνοντας πάντα υπόψη ότι οι οικονομικές μετρήσεις δυνατόν να εμπεριέχουν σφάλματα.

Μέθοδοι οικονομετρικής εκτίμησης:

OLS 3SLS

GLS FIML

ML

GMM

ILS

2SLS

LIML

Στάδιο 5^ο : Για τον έλεγχο του υποδείγματος εκτός από τα καθαρά στατιστικά κριτήρια, χρησιμοποιούνται και οικονομικά κριτήρια. Πράγματι, σε πολλές περιπτώσεις η οικονομική θεωρία καθορίζει όχι μόνο τα πρόσημα αλλά και τα όρια των συντελεστών (π.χ. σε περιπτώσεις ελαστικοτήτων, οριακών ροπών, πολλαπλασιαστών). Τα οικονομικά κριτήρια είναι λοιπόν a priori κριτήρια.

Στάδιο 6^ο : Χρησιμοποίηση του υποδείγματος για : έλεγχο της θεωρίας, αξιολόγηση οικονομικής πολιτικής, προβλέψεις.

1.6 Υποδείγματα Οικονομετρίας

1.6.1 Υποδείγματα μιας εξισώσεως (single equation models)

Στα υποδείγματα αυτά υποθέτουμε ότι όποιες αιτιατές σχέσεις (casual relationships) είναι μόνο προς μια κατεύθυνση δηλαδή από τις επεξηγηματικές μεταβλητές προς την εξαρτημένη μεταβλητή. Με άλλα

λόγια οι επεξηγηματικές μεταβλητές αποτελούν πάντα τα αίτια και η εξαρτημένη μεταβλητή το αποτέλεσμα.

Παράδειγμα: Ένα υπόδειγμα ζήτησεως

$$Q = \beta_0 + \beta_1 p + \beta_2 Y + \beta_3 p_s + \beta_4 p_c + u$$

Όπου Q = ζητούμενη ποσότητα, p = τιμή, p_s = τιμή υποκατάστατων αγαθών, p_c = τιμή συμπληρωματικών αγαθών, Y = εισόδημα

1.6.2 Υποδείγματα πολλών εξισώσεων (simultaneous equation models)

Σε πολλές περιπτώσεις μια οικονομική μεταβλητή όχι μόνο επηρεάζει (προκαλεί) αλλά ταυτόχρονα επηρεάζεται (προκαλείται) από άλλη(ες) οικονομική(ές) μεταβλητή(ές). Γι' αυτό και τα υποδείγματα μιας εξισώσεως συχνά αποτελούν μέρος ενός συστήματος εξισώσεων. Οι εξισώσεις αυτές ονομάζονται διαρθρωτικές εξισώσεις (structural equations) καθώς περιγράφουν τη διάρθρωση (δομή) της οικονομίας. Ως “διάρθρωση” εννοούμε όλα τα χαρακτηριστικά του εξεταζόμενου προβλήματος που παραμένουν σταθερά κατά την εξεταζόμενη περίοδο. Τα χαρακτηριστικά αυτά από τεχνική άποψη είναι η μορφή των εξισώσεων, καθώς και οι αριθμητικές τιμές των συντελεστών που ονομάζονται διαρθρωτικοί συντελετές ή διαρθρωτικές παράμετροι.

Οι διαρθρωτικές εξισώσεις διακρίνονται σε:

- α) εξισώσεις συμπεριφοράς (behavioural equations) που περιγράφουν τη συμπεριφορά των οικονομικών μονάδων και
- β) ταυτότητες (identities) ή εξισώσεις ορισμού (definitional) που είναι σχέσεις που επαληθεύονται λογιστικά, δηλαδή δεν περιγράφουν διαδικασίες αποφάσεων ούτε αιτιατές σχέσεις μεταξύ οικονομικών μεταβλητών. Είναι πάντα αιτιοκρατικές εξισώσεις.

1.6.3 Παραδείγματα

1. Κεϋνσιανό υπόδειγμα προσδιορισμού του εισοδήματος:

$$C_t = \beta_0 + \beta_1 Y_t + U_t \quad (1) \quad \beta_0 > 0, 0 < \beta_1 < 1$$

$$Y_t = C_t + I_t \quad (2)$$

C_t = κατανάλωση, I_t = επένδυση, Y_t = εισόδημα

Η (1) είναι εξίσωση συμπεριφοράς ενώ η (2) είναι ταυτότητα.

2. Υπόδειγμα προσφοράς-ζήτησης

{ συμπεριφοράς } $Q_t^d = \alpha_0 + \alpha_1 p_t + u_{1t}$, $\alpha_1 < 0$ { συνάρτηση ζήτησης }

{ συμπεριφοράς } $Q_t^s = \beta_0 + \beta_1 p_t + u_{2t}$, $\beta_1 > 0$ { συνάρτηση προσφοράς }

{ ταυτότητα } $Q_t^d = Q_t^s$ { συνθήκη ισορροπίας }

Q_t^d = ζητούμενη ποσότητα

Q_t^s = προσφερόμενη ποσότητα

3. $Q_t^d = \alpha_0 + \alpha_1 p_t + \alpha_2 p_{t-1} + u_{1t}$ (1)

$$Q_t^s = \beta_0 + \beta_1 p_t + \beta_2 Y + u_{2t} \quad (2)$$

$$Q_t^d = Q_t^s$$

Τα παραδείγματα (1) και (2) είναι στατιστικά συστήματα, γιατί όλες οι μεταβλητές αναφέρονται στην ίδια χρονική περίοδο, ενώ το παράδειγμα (3) είναι ένα δυναμικό σύστημα γιατί στην εξίσωση συμπεριφοράς (1) η προσφερόμενη ποσότητα τη χρονική στιγμή t είναι συνάρτηση της τιμής τόσο κατά τη χρονική στιγμή t όσο και κατά τη χρονική στιγμή $t-1$.

1.6.4 Διάκριση ανάμεσα στις μεταβλητές ενός οικονομετρικού συστήματος εξισώσεων

Οι μεταβλητές αρχικά διακρίνονται σε συστηματικές και τυχαίες. Η αναγκαιότητα εισαγωγής τυχαίων μεταβλητών συζητήθηκε ήδη.

Οι συστηματικές μεταβλητές διακρίνονται σε ενδογενείς (endogenous) και εξωγενείς (exogenous). Οι ενδογενείς προσδιορίζονται από το σύστημα στο οποίο αναφέρονται, ενώ οι εξωγενείς θεωρούνται δεδομένες, δηλαδή προσδιορίζονται εκτός του συστήματος. Οι ενδογενείς μεταβλητές, μέσω των εξισώσεων του συστήματος, προσδιορίζουν άλλες ενδογενείς μεταβλητές, αλλά και προσδιορίζονται ταυτόχρονα από αυτές. Οι εξωγενείς μεταβλητές προσδιορίζουν τις ενδογενείς, αλλά δεν προσδιορίζονται από αυτές. Τέλος, οι εξωγενείς καθώς και οι ενδογενείς με χρονική υστέρηση μεταβλητές ονομάζονται προκαθορισμένες μεταβλητές (predetermined variables).

Παραδείγματα

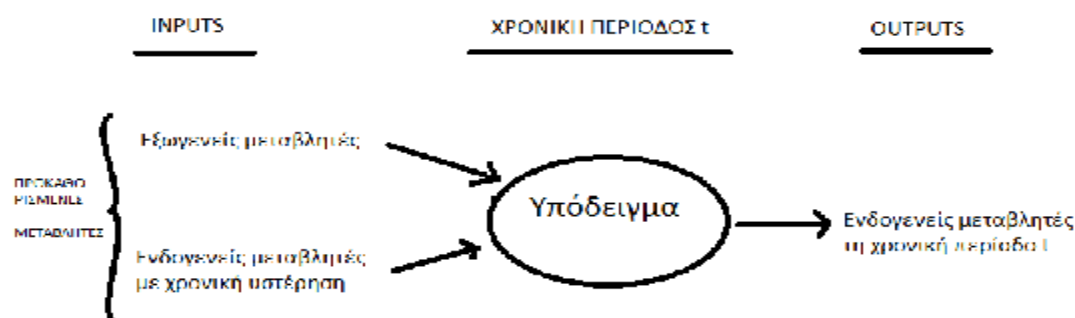
Στο υπόδειγμα προσδιορισμού του εισοδήματος οι C_t , Y_t είναι ενδογενείς μεταβλητές και η I_t εξωγενής.

Στο υπόδειγμα προσφοράς-ζήτησης (παράδειγμα 2) όλες οι μεταβλητές είναι ενδογενείς.

Τέλος, στο δυναμικό σύστημα (παράδειγμα 3) οι Q_t^d , Q_t^s , P_t είναι ενδογενείς μεταβλητές, η Y_t εξωγενής και η P_{t-1} προκαθορισμένη.

*Σημειώνεται ότι όσο πιο απλοποιημένο είναι ένα υπόδειγμα διαρθρωτικών εξισώσεων, τόσο ο αριθμός των εξωγενών μεταβλητών που υπεισέρχονται σε αυτό αυξάνει. Στο τέλει διαρθρωτικό υπόδειγμα όλες οι μεταβλητές θα ήταν ενδογενείς.

*Υπόδειγμα συστήματος ταυτόχρονων εξισώσεων χρησιμοποιήθηκε για πρώτη φορά στην Οικονομετρία το 1936 από τον Tinbergen. Αναφέρονταν στην οικονομία της Ολλανδίας και περιείχε 24 εξισώσεις.



Εικόνα 3 - Υπόδειγμα συστήματος εξισώσεων

1.6.5 Παράδειγμα εφαρμογής ενός οικονομετρικού υποδείγματος

Έστω το απλοποιημένο υπόδειγμα του Keynes:

$$(1) \quad C_t = \beta_0 + \beta_1 Y_t + U_t, \quad 0 < \beta_1 < 1$$

$$(2) \quad Y_t = C_t + I_t$$

Έστω ακόμα ότι με κάποια μέθοδο εκτίμησης βρίσκουμε $\beta_1 = 0,72$

Αν υποθέσουμε ότι $C_t =$ συνολική κατανάλωση (οπότε $Y_t =$ GNP) και επιβάλλονται νέοι φόροι ποία η επίπτωση στο εισόδημα από την αναμενόμενη μείωση των επενδύσεων;

$$Y_t = \beta_0 + \beta_1 Y_t + U_t + I_t \quad (\text{αντικατάσταση } C_t \text{ στη (2)})$$

$Y_t = \frac{\beta_0}{1-\beta_1} + \frac{1}{1-\beta_1} I_t + \frac{1}{1-\beta_1} U_t \quad \text{ó} \quad E(Y_t) = \frac{\beta_0}{1-\beta_1} + \frac{1}{1-\beta_1} I_t$ με $\beta_1 = \text{MPC}$, $1/(1-\beta_1)$ είναι ο πολλαπλασιαστής εισοδήματος (income multiplier) M .

Άρα για $\text{MPC} = 0,72$, $M = 1/(1-0,72) \approx 3,57$. Άρα στη μοναδιαία μείωση των επενδύσεων αντιστοιχεί περίπου 3,5 φορές μεγαλύτερη μείωση του εισοδήματος.

ΚΕΦΑΛΑΙΟ 2

ΠΑΛΙΝΔΡΟΜΗΣΗ- ΟΙΚΟΝΟΜΕΤΡΙΚΑ ΥΠΟΔΕΙΓΜΑΤΑ

2.1 Τι είναι η παλινδρόμηση ;

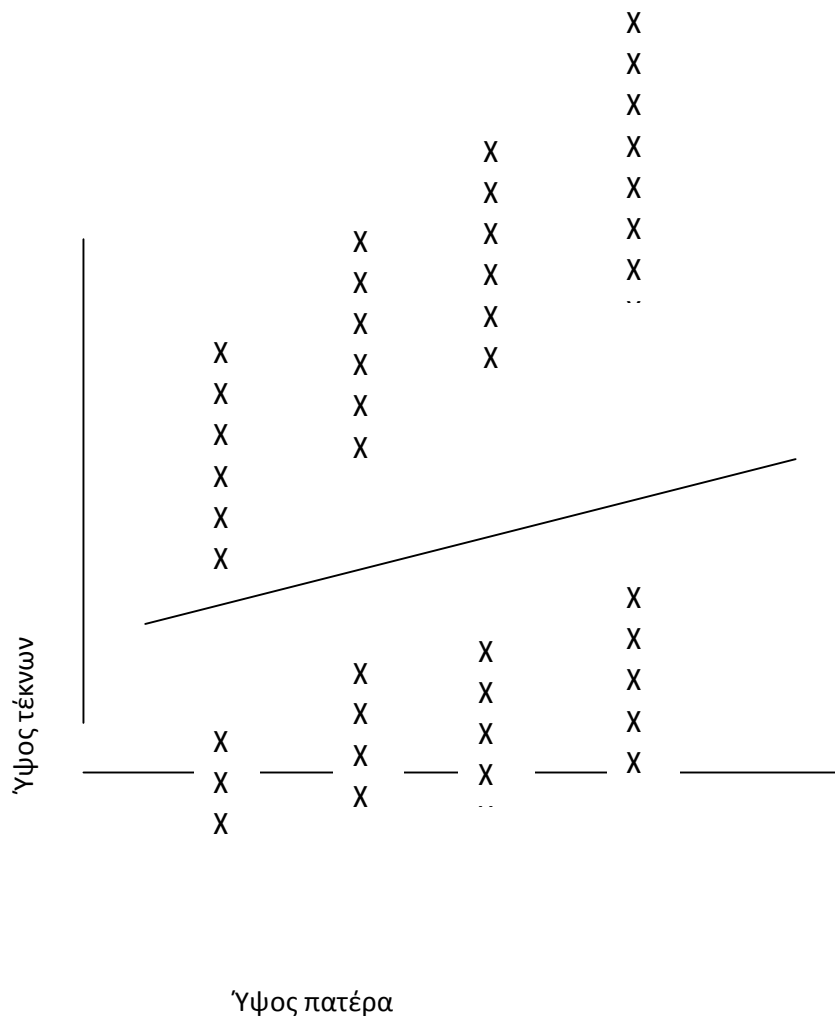
Ο όρος «παλινδρόμηση» (regression) χρησιμοποιήθηκε στη στατιστική αρχικά από τον Francis Galton το 1886. Ο Galton παρατήρησε ότι τα τέκνα πολύ υψηλών γονέων ήταν μεν και εκείνα υψηλά, αλλά όχι τόσο όσο οι γονείς τους. Ομοίως τα τέκνα πολύ κοντών γονέων ήταν μεν και εκείνα κοντά, αλλά όχι τόσο κοντά όσο οι γονείς τους. Με άλλα λόγια το ύψος των τέκνων ασυνήθιστα υψηλών ή ασυνήθιστα κοντών γονέων έτεινε προς το μέσο ύψος του πληθυσμού. Ο Galton ονόμασε τις διαπιστώσεις του «νόμο της παλινδρόμησης προς τη μετριότητα» (regression to mediocrity).

Η σημασία που ο Galton έδωσε στον όρο «παλινδρόμηση» δεν απέχει πολύ από τη σημασία που αποδίδουμε στη καθημερινή ζωή (Παλινδρόμηση: πορεία, κίνηση προς τα εμπρός και προς τα πίσω διαδοχικά).

Παλινδρομώ: πηγαίνω εμπρός-πίσω, αλλάζω γνώμη ή στάση ,δεν ξέρω τι θέλω (Ελληνικό λεξικό Γεωργόπουλου Φυτράκη 1993)).

Η σύγχρονη ερμηνεία του όρου παλινδρόμηση στη στατιστική είναι αρκετά διαφορετική: regression analysis is concerned with the study of the dependence of one variable , the dependent variable, on one or more other variables, the explanatory variables, with a view to estimating and/or predicting the (population) mean or average value of the former in terms of the known or fixed (in repeated sampling) values of the latter(Gujarati).

2.2 «Νόμος» του Galton : Η Ερμηνεία του



Για κάθε ύψος των πατέρων αντιστοιχεί μία κατανομή υψών των τεκνών. Όμως το μέσο ύψος των τέκνων αυξάνεται όσο αυξάνεται το ύψος των πατέρων. Η κόκκινη γραμμή παριστάνει ακριβώς αυτή την σχέση μέσου ύψους τέκνων – ύψος πατέρων. Το γεγονός ότι η κλίση της γραμμής αυτής είναι θετική αλλά <1 εξηγεί τον "νόμο" του Galton.

2.3 Παλινδρόμηση- αιτιότητα

Μια στατιστική σχέση όσο ισχυρή και αν είναι από μόνη της δεν μπορεί να συνεπάγεται αιτιότητα. Η ύπαρξη αιτιότητας θα πρέπει να βασίζεται

σε κάποια θεωρία εκτός της στατιστικής, ή έστω στη κοινή λογική. Π.χ. το ότι η κατανάλωση εξαρτάται από το εισόδημα και όχι το αντίθετο (δηλ. το εισόδημα να εξαρτάται από την κατανάλωση) δεν είναι κάτι που μου το λέει η στατιστική αλλά η οικονομική θεωρία. Σε ένα υπόδειγμα που προβλέπει τη γεωργική παραγωγή με βάση τη βροχόπτωση την ηλιοφάνεια το λίπασμα κλπ., το ότι η βροχόπτωση καθορίζει τη παραγωγή και όχι το αντίθετο δηλ. η παραγωγή να καθορίζει τη βροχόπτωση πάλι δεν το καταλαβαίνουμε από τη στατιστική αλλά από τη κοινή λογική.

2.4 Παλινδρόμηση- συσχέτιση

Η ανάλυση συσχέτισης μέσω των συντελεστών γραμμικής συσχέτισης μετρά την "ένταση" ή το βαθμό της συσχέτισης μεταξύ δύο μεταβλητών. Στην ανάλυση συσχέτισης μπορούμε να υποθέσουμε από και οι δύο μεταβλητές είναι στοχαστικές και δεν έχει σημασία ποια θεωρούμε πρώτη και ποια δεύτερη. Στην ανάλυση παλινδρόμησης υπάρχει ασυμμετρία στον τρόπο που χειριζόμαστε τις μεταβλητές καθώς η μία θεωρείται ως αποτέλεσμα και η άλλη (ή οι άλλες) ως το αίτιο. Επιπλέον η εξαρτημένη μεταβλητή θεωρείται στοχαστική ενώ η ανεξάρτητη αρχικά θεωρείται ως μη στοχαστική. Για Κάθε τιμή της επεξηγηματικής μεταβλητής θεωρούμε ότι υπάρχει μια κατανομή τιμών της εξαρτημένης μεταβλητής.

Ορολογία

Αποτέλεσμα

Αίτιο

Dependent variable	Independent variable
Explained Variable	Explanatory Variable
Predictand	predictor
Regressand	Regressor
Response	Stimulus (control variable)
Endogenous	Exogenous

2.5 Βασικές Έννοιες στην Απλή Παλινδρόμηση

Εβδομαδιαίο οικογενειακό διαθέσιμο εισόδημα X (σε €)

X \ Y	80	100	120	140	160	180	200	220	240	260
Εβδομαδιαία Δαπάνη Κατανάλωσης Y (σε €)	55	65	79	80	102	110	135	135	137	150
	60	70	84	93	107	115	136	137	145	152
	65	74	90	95	110	120	140	140	155	175
	70	80	94	103	116	130	144	152	165	178
	75	85	98	108	118	135	145	157	175	180
		88		113	125	140		160	189	185
				115				162		191
Σύνολο	325	462	445	707	678	750	685	1043	966	1211

Κάθε στήλη του παραπάνω πίνακα δίνει την κατανομή της εβδομαδιαίας καταναλωτικής δαπάνης που αντιστοιχεί σε δεδομένο επίπεδο διαθέσιμου εισοδήματος. Δηλαδή μας παρέχει την υπό συνθήκη κατανομή της Y για δεδομένες τιμές του X .

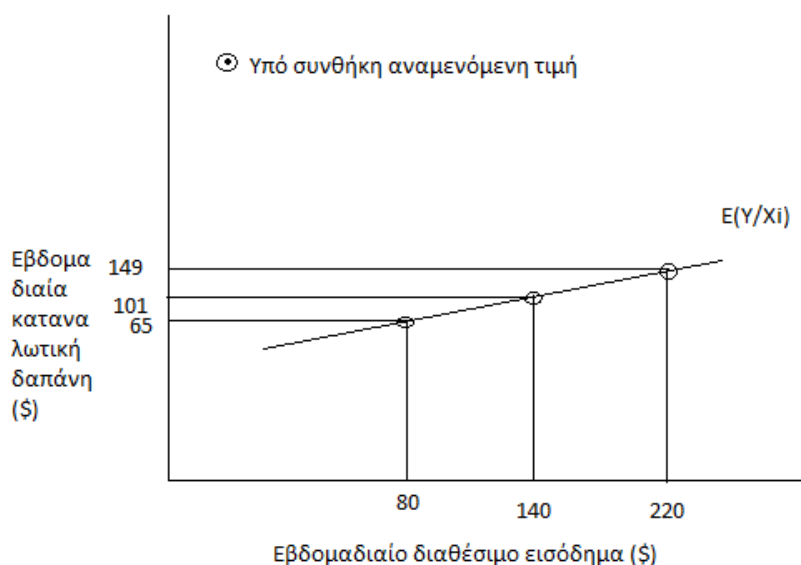
Εύκολα μπορούν να υπολογισθούν οι υπό συνθήκη πιθανότητες η (διακριτή) μεταβλητή Y να έχει τις τιμές Y_1, Y_2, \dots, Y_j , με δεδομένο ότι η (διακριτή) μεταβλητή X έχει αντίστοιχα τις τιμές X_1, X_2, \dots, X_i . Συμβολισμός: $P(Y = Y_j | X = X_i)$ ή για απλούστευση: $P(Y_j | X_i)$. Η υπό συνθήκη αναμενόμενη τιμή της Y για δεδομένο X θα είναι: $E(Y | X = X_j) = \sum Y_i * P(Y = Y_i | X = X_j)$.

π.χ. $E(Y | X = X_j) = 55 * \frac{1}{5} + 60 * \frac{1}{5} + 65 * \frac{1}{5} + 70 * \frac{1}{5} + 75 * \frac{1}{5} = 65$

Όμοια υπολογίζουμε τις για τις υπόλοιπες τιμές της X .

Εισόδημα	80	100	120	140	160	180	200	220	240	260
Υπό συνθήκη μέση τιμή της Y	65	77	89	101	113	125	137	149	161	173

2.6 Πληθυσμιακή Καμπύλη Παλινδρόμησης



Εικόνα 4 -Πληθυσμιακή Καμπύλη Παλινδρόμησης

Γεωμετρικά η πληθυσμιακή καμπύλη παλινδρόμησης είναι ο γεωμετρικός τόπος των υπό συνθήκη μέσων τιμών της εξαρτημένης μεταβλητής για δεδομένες τιμές της ανεξάρτητης μεταβλητής.

2.7 Η Πληθυσμιακή Συνάρτηση της Παλινδρόμησης (population regression function PRF)

Κάθε υπό συνθήκη μέση τιμή της Y είναι συνάρτηση των X_i :

$$E(Y|X_i) = f(X_i)$$

Η f μας φανερώνει πως το μέσο αποτέλεσμα μεταβάλλεται (ανταποκρίνεται) σε μεταβολή του αιτίου που το προκάλεσε.

Η συναρτησιακή μορφή της PRF πρέπει να είναι σύμφωνη με την υπάρχουσα (οικονομική) θεωρία. Αν δεν υπάρχει συγκεκριμένη θεωρία θα ξεκινάμε πάντα από ένα γραμμικό και προσθετικό παράδειγμα
Δηλ.:

$$E(Y_i|X_{ik}) = \beta_0 + \beta_1 * X_{i1} + \beta_2 * X_{i2} + \dots + \beta_k * X_{ik}$$

Για το συγκεκριμένο παράδειγμα :

$$E(Y|X_i) = \beta_1 + \beta_2 * X_i$$

Τα β_1, β_2 ονομάζονται συντελεστές παλινδρόμησης. Ο συντελεστής β_2 στο συγκεκριμένο παράδειγμα ονομάζεται οριακή ροπή για κατανάλωση (marginal propensity to consume), καθώς δίνει τη μέση μεταβολή στην καταναλωτική δαπάνη που αντιστοιχεί στη μοναδιαία μεταβολή του διαθέσιμου εισοδήματος.

2.8 Στοχαστικές διαταραχές (Stochastic Disturbances)

Το υπόδειγμα που περιγράφει την εξάρτηση της υπό συνθήκη αναμενόμενης τιμής της Y από την X δηλ. το: είναι ένα καθαρά αιτιοκρατικό υπόδειγμα.

Όμως από τα δεδομένα τον πίνακα είναι φανερό ότι σε επίπεδο οικογένειας το συγκεκριμένο υπόδειγμα δεν ακολουθείται αφού, όπως εύκολα διαπιστώνει κανείς, υπάρχουν και περιπτώσεις που η καταναλωτική δαπάνη μειώνεται όταν αυξάνεται το εισόδημα (θυμίζουμε ότι σύμφωνα με τη θεωρία του Keynes $0 < \beta_2 < 1$).

Η μετάβαση από τις υπό συνθήκη αναμενόμενες τιμές της καταναλωτικής δαπάνης για δεδομένο διαθέσιμο εισόδημα, σε συγκεκριμένες τιμές της καταναλωτικής δαπάνης καθιστά απαραίτητη την εισαγωγή ενός στοχαστικού όρου στο υπόδειγμα.

Η στοχαστική διαταραχή u_i ορίζεται σαν μια τυχαία μεταβλητή που εκφράζει την απόκλιση μιας συγκεκριμένης τιμής της Y από την αντίστοιχη υπό συνθήκη αναμενόμενη τιμή δηλ.:

$$u_i = Y_i - E(Y|X_i) \Rightarrow Y_i = E(Y|X_i) + u_i$$

Η εξίσωση αυτή εκφράζει μία συγκεκριμένη τιμή Y_i σαν άθροισμα δύο όρων: της υπό συνθήκη αναμενόμενης τιμής για δεδομένο X_i , που μας παρέχει το συστηματικό (αιτιοκρατικό) τρόπο εξέλιξης της Y με την X και της στοχαστικής διαταραχής, που εκφράζει τη μη συστηματική συνιστώσα της μεταβολής της Y .

Η συνιστώσα U_i είναι ένα υποκατάστατο όλων των μεταβλητών που μπορεί να επηρεάζουν την Y και για κάποιο λόγο δεν έχουν συμπεριληφθεί στην εξίσωση.

Υποθέτουμε και πάλι ότι η πληθυσμιακή συνάρτηση παλινδρόμησης είναι γραμμική. Τότε:

Η εξίσωση αυτή εκφράζει μία συγκεκριμένη τιμή Y_i σαν άθροισμα δύο όρων: της υπό συνθήκη αναμενόμενης τιμής για δεδομένο X_i , που μας παρέχει το συστηματικό (αιτιοκρατικό) τρόπο εξέλιξης της Y με την X και της στοχαστικής διαταραχής, που εκφράζει τη μη συστηματική συνιστώσα της μεταβολής της Y .

Η συνιστώσα U_i είναι ένα υποκατάστατο όλων των μεταβλητών που μπορεί να επηρεάζουν την Y και για κάποιο λόγο δεν έχουν συμπεριληφθεί στην εξίσωση.

Υποθέτουμε και πάλι ότι η πληθυσμιακή συνάρτηση παλινδρόμησης είναι γραμμική. Τότε:

Η εξίσωση αυτή εκφράζει μία συγκεκριμένη τιμή Y_i σαν άθροισμα δύο όρων: της υπό συνθήκη αναμενόμενης τιμής για δεδομένο X_i , που μας παρέχει το συστηματικό (αιτιοκρατικό) τρόπο εξέλιξης της Y με την X

και της στοχαστικής διαταραχής, που εκφράζει τη μη συστηματική συνιστώσα της μεταβολής της Y .

Η συνιστώσα U_i είναι ένα υποκατάστατο όλων των μεταβλητών που μπορεί να επηρεάζουν την Y και για κάποιο λόγο δεν έχουν συμπεριληφθεί στην εξίσωση.

Υποθέτουμε και πάλι ότι η πληθυσμιακή συνάρτηση παλινδρόμησης είναι γραμμική. Τότε:

$$Y_i = E(Y|X_i) + U_i = \beta_1 + \beta_2 * X_i + U_i$$

Και παίρνοντας αναμενόμενες τιμές:

$$E(Y_i|X_i) = E\{E(Y|X_i)\} + E(U_i|X_i) = E(Y|X_i) + E(U_i|X_i)$$

Οπότε δηλ. $E(U_i|X_i) = 0$, δηλ. η υπόθεση ότι η καμπύλη παλινδρόμησης διέρχεται από τις υπό συνθήκη αναμενόμενες τιμές τις συνεπάγεται ότι οι υπό συνθήκη αναμενόμενες της U_i για δεδομένο X_i είναι μηδέν.

2.9 Η Δειγματική Συνάρτηση της Παλινδρόμησης

Συνήθως δε μας είναι γνωστές όλες οι τιμές Y_i όπως υποθέσαμε μέχρι τώρα. Για κάθε X_i έχουμε μια τιμή Y_i , δηλ. κάθε φορά έχουμε ένα δείγμα από το πληθυσμό.

Σε αντιστοιχία με την PRF ορίζουμε τη δειγματική συνάρτηση παλινδρόμησης ως εξής:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 * X_i$$

Όπου, $\hat{Y}_i =$ εκτιμητής της $E(Y|X_i)$
 $\hat{\beta}_1 =$ εκτιμητής του β_1
 $\hat{\beta}_2 =$ εκτιμητής του β_2

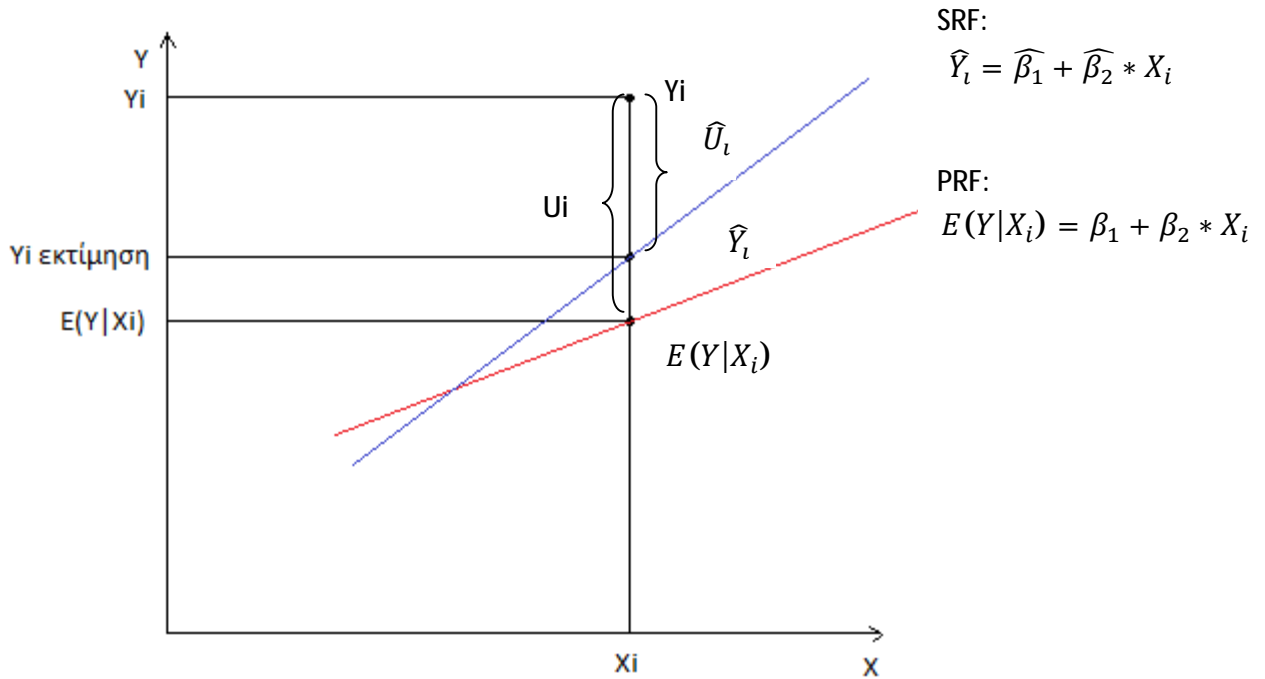
(Σημ. εκτιμητής ή (δειγματικό στατιστικό) είναι ένας κανόνας ή τύπος ή μέθοδος που μας υπολογίζει την αντίστοιχη παράμετρο του πληθυσμού από πληροφορίες που μας δίνονται από το δείγμα. Μια συγκεκριμένη τιμή που βρίσκεται σε μια εφαρμογή χρησιμοποιώντας έναν εκτιμητή ονομάζεται εκτίμηση).

Η πραγματική τιμή Y_i μπορεί επομένως να εκφρασθεί και ως:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 * X_i + \hat{u}_i$$

όπου \hat{u}_i η εκτίμηση για το U_i . Το \hat{u}_i ονομάζεται κατάλοιπο.

Το παρακάτω σχήμα βοηθά στο να ξεκαθαρισθούν οι έννοιες διαταραχές, κατάλοιπα, PRF, SRF.

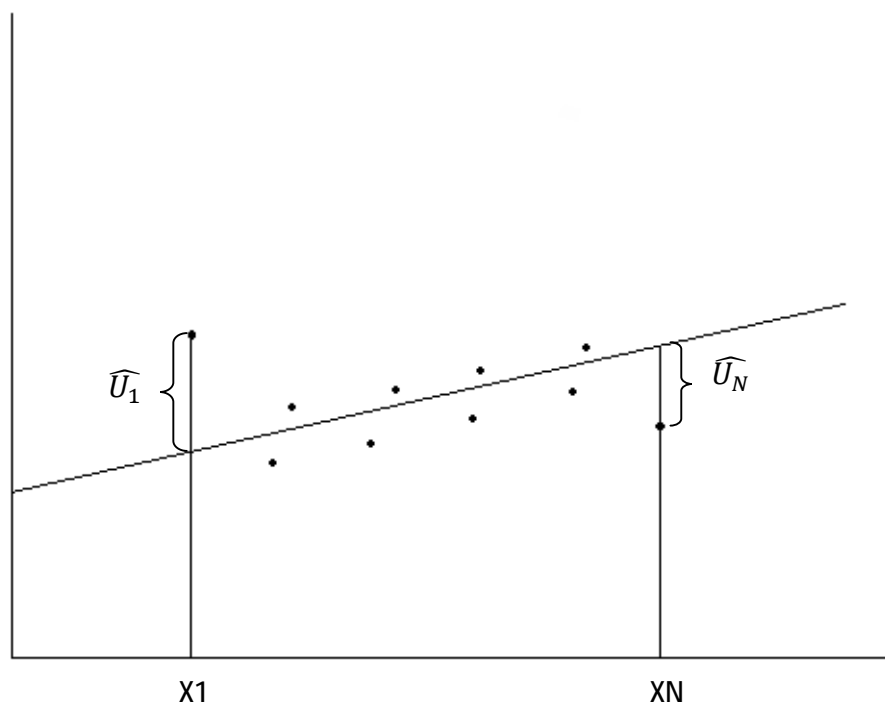


Εικόνα 5 Δειγματική Συνάρτηση της Παλινδρόμησης

Μερικές αιτίες σχετικά με την ύπαρξη του στοχαστικού όρου

- Η σχετική θεωρία είναι ασαφής
- Μη διαθεσιμότητα κάποιων υποψήφιων επεξηγηματικών μεταβλητών, π.χ. τα περιουσιακά στοιχεία για το παράδειγμα μας.
- Πολυπλοκότητα (intrinsic randomness) της ανθρώπινης συμπεριφοράς.
- Λάθος συναρτησιακή σχέση, ή/και σφάλματα στα δεδομένα
- Occam's razor (αρχή της απλότητας).

2.10 Απλή Παλινδρόμηση- Εκτίμηση



Εικόνα 6 Δειγματική Συνάρτηση της Παλινδρόμησης

$$\widehat{U}_i = Y_i - \widehat{Y}_i = Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 * X_i$$

Ζητούμενο: ένα κριτήριο με βάση το οποίο να προσδιορίσουμε τα β_1, β_2 .

Έστω ότι αρχικά επιλέγουμε: $\sum \widehat{U}_i = 0$

Τότε: $\sum \widehat{U}_i = \sum (Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 * X_i) = 0$ και διαιρώντας με N (όπου N το μέγεθος του δείγματος):

$$\frac{\sum Y_i}{N} - \frac{\sum \widehat{\beta}_1}{N} - \widehat{\beta}_2 * \frac{\sum X_i}{N} = 0 \Rightarrow \bar{Y} = \widehat{\beta}_1 + \widehat{\beta}_2 * \bar{X}$$

Με αυτό τον τρόπο τα $\widehat{\beta}_1, \widehat{\beta}_2$ εκλέγονται έτσι ώστε η SRF να περνά από το σημείο (\bar{Y}, \bar{X}) . Όμως έτσι τα $\widehat{\beta}_1, \widehat{\beta}_2$ δεν ορίζονται μονοσήμαντα αφού μια ευθεία δε μπορεί να ορισθεί από ένα μόνο σημείο.

Μια καλύτερη σκέψη θα ήταν να ελαχιστοποιήσουμε το άθροισμα των απόλυτων τιμών των αποκλίσεων των

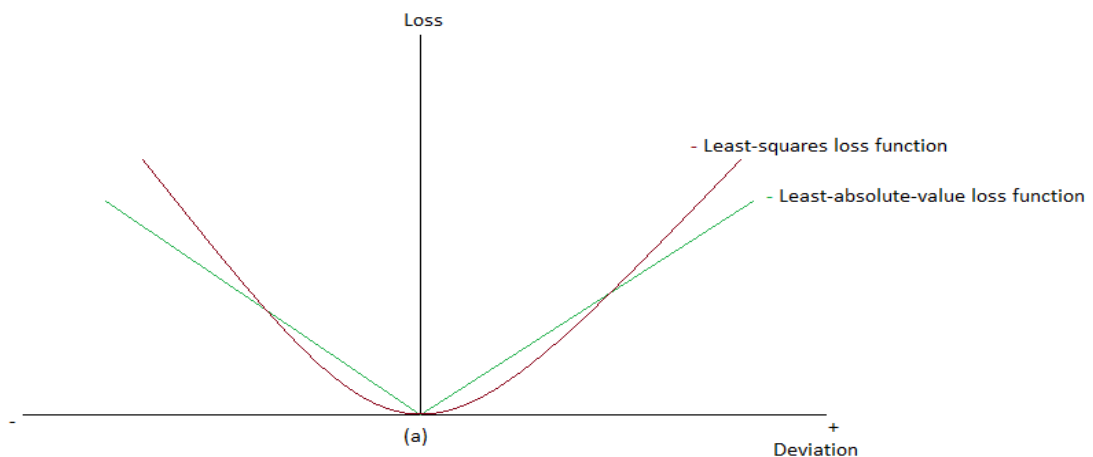
δειγματικών σημείων από τη γραμμή παλινδρόμησης. Με αυτόν τον τρόπο η βαρύτητα που αποδίδεται στις αποκλίσεις είναι ανάλογη με τις απόλυτες τιμές τους. Μια τέτοια διαδικασία ελαχιστοποίησης του αθροίσματος των απόλυτων τιμών των αποκλίσεων αφενός είναι υπολογιστικά σχετικά δύσκολη, αφετέρου θα επιθυμήσαμε η βαρύτητα που θα δίνονταν στις συγκριτικά μεγαλύτερες αποκλίσεις να ήταν μεγαλύτερη από αυτή που θα αντιστοιχούσε αν ίσχυε η αναλογικότητα. Μια τέτοια διαδικασία που αποδίδει στις μεγαλύτερες αποκλίσεις μεγαλύτερη βαρύτητα απ' ό,τι αναλογικά τους αντιστοιχεί είναι αυτή που ελαχιστοποιεί το άθροισμα των τετραγώνων των αποκλίσεων. Τα παραπάνω φαίνονται καλύτερα στο σχήμα της επόμενης σελίδας που παριστάνει τη λεγόμενη συνάρτηση απωλειών (loss function) σε συνάρτηση των αποκλίσεων.

Με τη διαδικασία ελαχιστοποίησης του αθροίσματος των απόλυτων τιμών των αποκλίσεων η συνάρτηση απωλειών για μια συγκεκριμένη απόκλιση (θετική ή αρνητική) ισούται με την απόλυτη τιμή της απόκλισης. Με τη διαδικασία ελαχιστοποίησης του αθροίσματος των τετραγώνων των αποκλίσεων η τιμή της συναρτήσεως απωλειών για μια συγκεκριμένη απόκλιση ισούται με το τετράγωνο της απόκλισης. Ένα πρόβλημα με τη διαδικασία ελαχιστοποίησης του αθροίσματος των ελαχίστων τετραγώνων αφορά την περίπτωση που έχουμε ακραίες τιμές στα δειγματικά δεδομένα. Σε μια τέτοια περίπτωση η ύπαρξη έστω και μίας ακραίας τιμής με μεγάλη απόκλιση είναι δυνατό να αλλάξει δραματικά τις τιμές των συντελεστών παλινδρόμησης. Μία τέτοια κατάσταση αντιμετωπίζεται είτε αφαιρώντας τελείως την ακραία τιμή

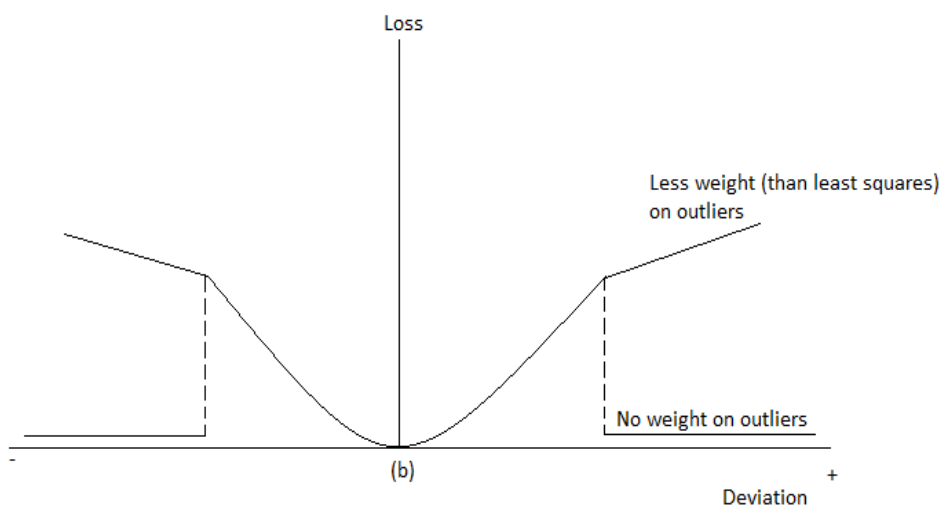
από τα δεδομένα, είτε τροποποιώντας τη μέθοδο της ελαχιστοποίησης του αθροίσματος των ελαχίστων τετραγώνων μειώνοντας τη σημασία των ακραίων τιμών με τη χρήση κατάλληλων συντελεστών σταθμίσεως. Δύο παραδείγματα τέτοιων εναλλακτικών μεθόδων φαίνονται στο σχήμα.

2.11 Διαγράμματα Συναρτήσεων Απωλειών

(a) Loss functions and (b) alternative loss functions



Εικόνα 7 Διαγράμματα Συναρτήσεων Απωλειών



Εικόνα 8 Διαγράμματα Συναρτήσεων Απωλειών

Με το κριτήριο των ελαχίστων τετραγώνων θέλουμε $\sum \widehat{U}_i^2 = \sum (Y_i - \widehat{Y}_i)^2 = \sum (Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 * X_i)^2 \rightarrow \min$ καθώς $\sum \widehat{U}_i^2 = f(\widehat{\beta}_1, \widehat{\beta}_2)$ πρέπει:

$$\frac{\partial f(\widehat{\beta}_1, \widehat{\beta}_2)}{\partial \widehat{\beta}_1} = \frac{\partial f(\widehat{\beta}_1, \widehat{\beta}_2)}{\partial \widehat{\beta}_2} = 0$$

$$\text{Έτσι, } -2 \sum (Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 * X_i) = 0 \Rightarrow \sum \widehat{U}_i = 0$$

$$-2 \sum (Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 * X_i) * X_i = 0 \Rightarrow \sum \widehat{U}_i * X_i = 0$$

$$\left. \begin{aligned} \text{Η } \sum Y_i &= N * \widehat{\beta}_1 + \widehat{\beta}_2 * \sum X_i \\ \sum Y_i * X_i &= \widehat{\beta}_1 * \sum X_i + \widehat{\beta}_2 * \sum X_i^2 \end{aligned} \right\} \text{Κανονικές εξισώσεις}$$

$$\text{Οπότε, } \widehat{\beta}_1 = \bar{Y} - \widehat{\beta}_2 * \bar{X}$$

$$\widehat{\beta}_2 = \frac{\sum (X_i - \bar{X}) * (Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{N * \sum X_i * Y_i - \sum X_i \sum Y_i}{N * \sum X_i^2 - (\sum X_i)^2}$$

$$\text{Η αν } x_i = X_i - \bar{X} \text{ και } y_i = Y_i - \bar{Y} \text{ τότε } \widehat{\beta}_2 = \frac{\sum x_i * y_i}{\sum x_i^2} = \frac{\sum x_i * Y_i}{\sum X_i^2 - N * \bar{X}^2}$$

$$\begin{aligned} \text{Σημ.: } \sum x_i^2 &= \sum (X_i - \bar{X})^2 = \sum X_i^2 - 2 * \sum X_i * \bar{X} + \sum \bar{X}^2 = \sum X_i^2 - 2 * \\ &\bar{X} * \sum X_i + \sum \bar{X}^2 = \sum X_i^2 - 2 * N * \bar{X}^2 + N * \bar{X}^2 \Rightarrow \sum x_i^2 = \sum X_i^2 - \\ &N * \bar{X}^2 \end{aligned}$$

Κατά τον ίδιο τρόπο

$$\sum x_i * y_i = \sum (X_i * Y_i) - N * \bar{X} * \bar{Y}.$$

Η SRL έχει τις παρακάτω ιδιότητες:

1) Διέρχεται από τα δειγματικά μέσα των Y_i, X_i

Πράγματι από τις κανονικές εξισώσεις

$$\bar{Y} = \widehat{\beta}_1 + \widehat{\beta}_2 * \bar{X}$$

2) Το μέσο των εκτιμήσεων \widehat{Y}_i ισούται με το μέσο των Y_i

Πράγματι

$$\widehat{Y}_i = \widehat{\beta}_1 + \widehat{\beta}_2 * X_i = (\bar{Y}_i - \widehat{\beta}_2 * \bar{X}_i) + \widehat{\beta}_2 * X_i = \bar{Y} + \widehat{\beta}_2 * (X_i - \bar{X}) \Rightarrow$$

$$\frac{\sum \widehat{Y}_i}{N} = \frac{\sum \bar{Y}}{N} + \widehat{\beta}_2 * \frac{\sum (X_i - \bar{X})}{N} \Rightarrow$$

$$\bar{\widehat{Y}}_i = \bar{Y}$$

3) $\sum \widehat{u}_i = 0$ (από κανονικές εξισώσεις)

4) Τα υπόλοιπα \widehat{u}_i δεν σχετίζονται με τις εκτιμήσεις \widehat{Y}_i .

Πράγματι: $\sum (Y_i - \bar{Y})(\widehat{u}_i - \bar{\widehat{u}}_i) = \sum y_i \widehat{u}_i = \widehat{\beta}_2 \sum x_i \widehat{u}_i$

(επειδή $\widehat{y}_i = \widehat{\beta}_2 x_i$ όπως θα δείξουμε παρακάτω)

$$= \widehat{\beta}_2 \sum x_i (y_i - \widehat{\beta}_2 x_i) = \widehat{\beta}_2 \sum x_i y_i - \widehat{\beta}_2^2 \sum x_i^2 =$$

$$= \widehat{\beta}_2^2 \sum x_i^2 - \widehat{\beta}_2^2 \sum x_i^2 = 0 \text{ (γιατί } \widehat{\beta}_2 = \sum x_i y_i / \sum x_i^2 \text{)}$$

5) Έκφραση της SRF με τη μορφή αποκλίσεων από τις μέσες τιμές.

$$\left. \begin{array}{l} Y_i = \widehat{\beta}_1 + \widehat{\beta}_2 X_i + \widehat{u}_i \\ \\ \bar{Y} = \widehat{\beta}_1 + \widehat{\beta}_2 \bar{X} \end{array} \right\} \begin{array}{l} Y_i - \bar{Y} = \widehat{\beta}_2 (X_i - \bar{X}) + u_i \\ \\ \Rightarrow \boxed{y_i = \widehat{\beta}_2 X_i + \widehat{u}_i} \end{array}$$

Οπότε και $\hat{y}_i = \widehat{\beta}_2 x_i$

2.12 Οι υποθέσεις του κλασσικού γραμμικού υποδείγματος παλινδρόμησης

Υπόθεση 1: Η επεξηγηματική μεταβλητή X είναι μη στοχαστική

Υπόθεση 2: $E(u_i/X_i) = 0$

Υπόθεση

3:

$Var(u_i/X_i) = E\{u_i - E(u_i)|X_i\}^2 = E(u_i^2/X_i) = \sigma^2$

(όχι ετεροσκεδαστικότητα).

Υπόθεση 4: $Cov(u_i, u_j/X_i, X_j) =$

$$= E(u_i - E(u_i)/X_i)(u_j - E(u_j)/X_j) = E(u_i/X_i)(u_j/X_j) = 0$$

(όχι αυτοσυσχέτιση μεταξύ των διαταραχών u)

Υπόθεση 5: $Cov(u_i, X_i) = E(\{u_i - E(u_i)\}\{X_i - E(X_i)\}) = E\{u_i(X_i - E(X_i))\} = E(u_i X_i) - E(u_i)E(X_i) = E(u_i X_i) = 0$

($E(X_i) = \text{σταθερά}$)

(όχι συνδιακύμανση μεταξύ u_i, X_i , για την περίπτωση που δεν ισχύει η Υπόθεση 1).

Υπόθεση 6: Το υπόδειγμα παλινδρόμησης πρέπει να είναι σωστά ορισμένο.

Λοιπές υποθέσεις (π.χ. μεταβλητότητα της $X \neq 0$, γραμμικότητα στις παραμέτρους κ.λπ.).

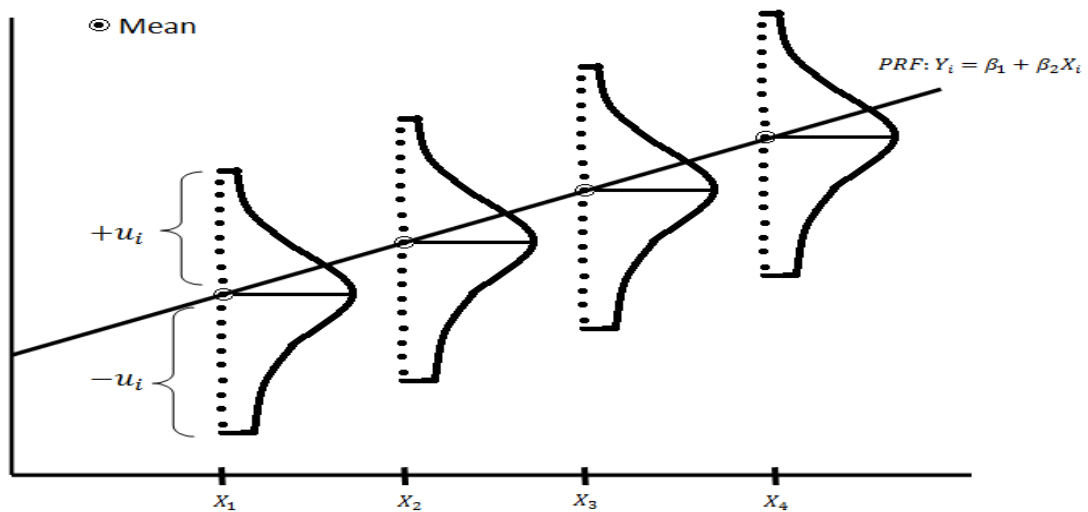
Επεξηγηματικά σχόλια.

✓ Η υπόθεση 2 ουσιαστικά σημαίνει ότι το μέσο αποτέλεσμα όλων των παραγόντων, που δεν υπεισέρχονται απευθείας στο υπόδειγμα αλλά ενσωματώνονται στις διαταραχές u_i , πάνω στην Y είναι μηδέν (βλέπε σχήμα).

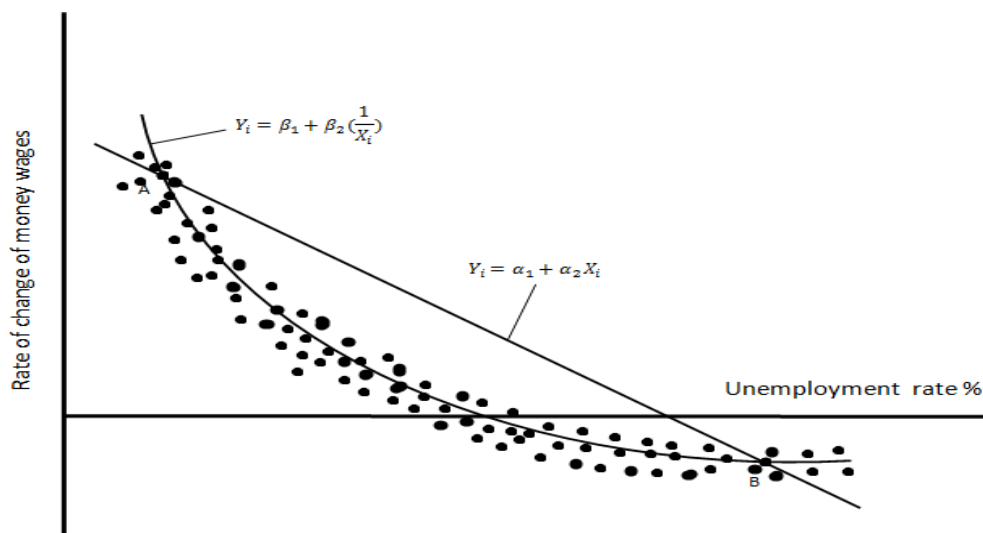
✓ Η υπόθεση 3 απαιτεί οι πληθυσμοί της Y που αντιστοιχούν σε κάθε X_i να έχουν την ίδια διακύμανση. Αν κάτι τέτοιο δεν ίσχυε, οι διαφορετικές τιμές Y_i δεν θα είχαν την ίδια αξιοπιστία, και θα έπρεπε να δώσουμε μεγαλύτερη σπουδαιότητα σε τιμές που προέρχονται από πληθυσμούς με συγκριτικά μικρότερη διακύμανση (βλέπε σχήμα επόμενης σελίδας).

✓ Η υπόθεση 4 απαιτεί οι διαταραχές $u_i, u_j, i \neq j$ να είναι ασυσχέτιστες. Αν κάτι τέτοιο δεν ισχύει τότε η Y_t δεν θα εξαρτάται μόνο από την X_t αλλά και από την u_{t-1} , καθώς η u_{t-1} θα επηρεάζει ως ένα βαθμό την u_t . (Σημ. ο δείκτης t στην προκειμένη περίπτωση αναφέρεται στην τάξη της παρατήρησης και δεν αφορά αναγκαστικά χρόνο).

✓ Η υπόθεση 5 (που ικανοποιείται αυτομάτως όταν ισχύει η υπόθεση (1)) εκφράζει το γεγονός ότι η επίδραση της X στην Y πρέπει να προστίθεται στην επίδραση της u στην Y (δηλαδή η κάθε μεταβλητή να δρα στην Y χωριστά από την άλλη). Κάτι τέτοιο προφανώς δεν θα ισχύει αν οι X, u συσχετίζονται).

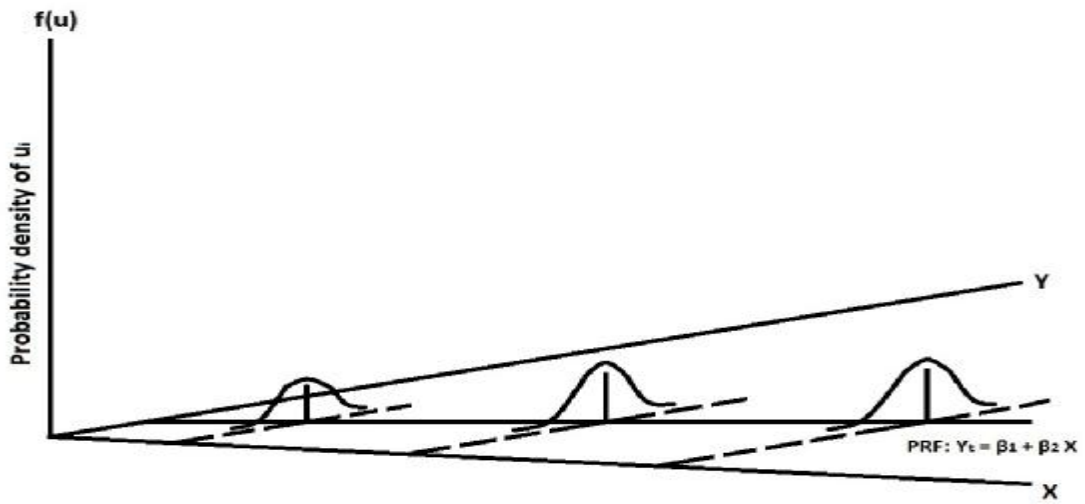


Υπό συνθήκη κατανομή των διαταραχών U_i



Εσφαλμένη εξειδίκευση του υποδείγματος παλινδρόμησης

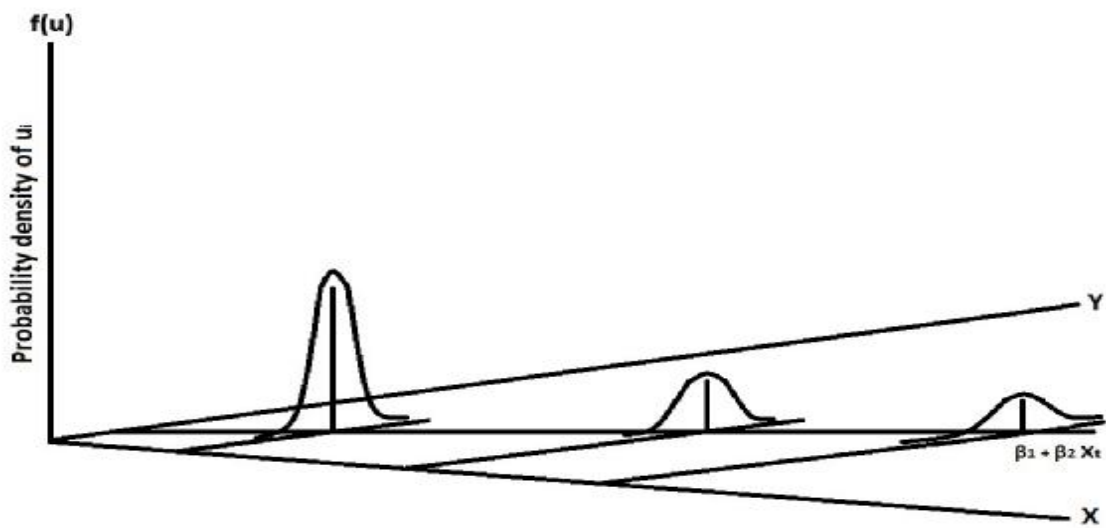
Εικόνα 9 Οι υποθέσεις του κλασικού γραμμικού υποδείγματος παλινδρόμησης



Εικόνα 10 Οι υποθέσεις του κλασικού γραμμικού υποδείγματος παλινδρόμησης

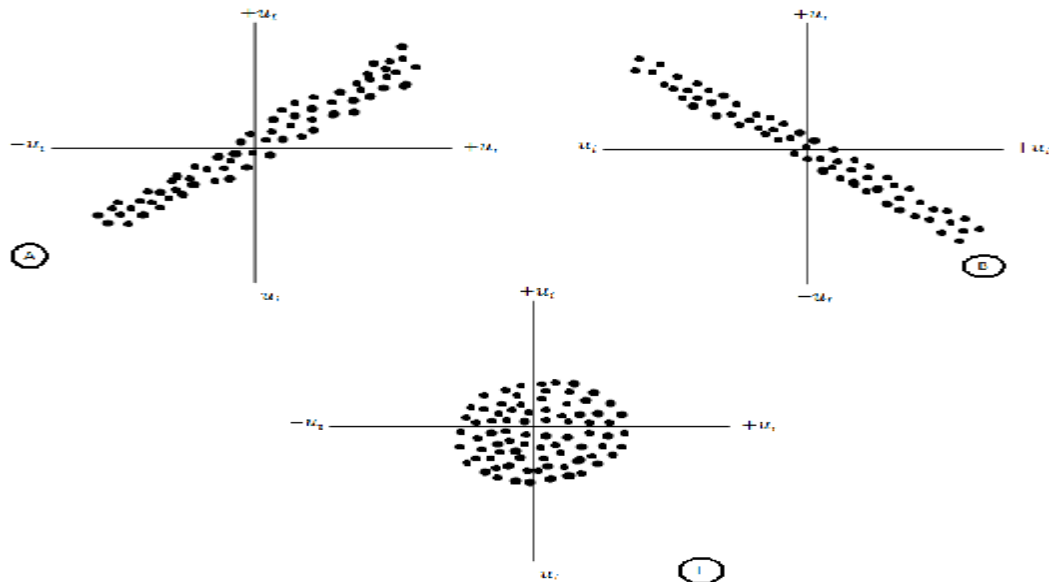
Ομοσκεδαστικότητα

$$\text{var}(u_i/X_i) = \sigma_i^2$$



Εικόνα 11 Ομοσκεδαστικότητα

Ετεροσκεδαστικότητα



Εικόνα 12 Ετεροσκεδαστικότητα

Αυτοσυσχέτιση στον διαταρακτικό όρο

- a) Θετική αυτοσυσχέτιση
- b) Αρνητική αυτοσυσχέτιση
- c) Μηδενική αυτοσυσχέτιση

Διευκρινήσεις θα πρέπει να δοθούν και σχετικά με τη σημασία του όρου “γραμμικότητα” που αναφέρεται στις λοιπές υποθέσεις. Σε ένα υπόδειγμα παλινδρόμησης διακρίνουμε γραμμικότητα ως προς τις μεταβλητές και γραμμικότητα ως προς τις παραμέτρους. Γραμμικότητα ως προς τις μεταβλητές σημαίνει ότι η υπό συνθήκη αναμενόμενη τιμή της εξαρτημένης μεταβλητής είναι γραμμική συνάρτηση της ανεξάρτητης

μεταβλητής. Γραμμικότητα ως προς τις παραμέτρους σημαίνει ότι η υπό συνθήκη αναμενόμενη τιμή της εξαρτημένης μεταβλητής είναι γραμμική συνάρτηση των παραμέτρων β_1, β_2 . Σημειώνεται με έμφαση ότι στην περίπτωση της γραμμικότητας ως προς τις παραμέτρους, το υπόδειγμα δεν είναι κατ' ανάγκη γραμμικό και ως προς την ανεξάρτητη μεταβλητή. Στο εξής ο όρος “γραμμικό” υπόδειγμα παλινδρόμησης θα δηλώνει ένα υπόδειγμα παλινδρόμησης γραμμικό ως προς τις παραμέτρους.

2.13 Εφαρμογές

Σαν μια πρώτη εφαρμογή στην εκτιμητική και συμπερασματολογία της απλής παλινδρόμησης θα χρησιμοποιηθούν τα δεδομένα του παρακάτω πίνακα (για παιδαγωγικούς λόγους χρησιμοποιούνται μόνο 5 ζεύγη τιμών Y_i, X_i έτσι ώστε οι φοιτητές να μπορούν να επεξεργασθούν την άσκηση βήμα- βήμα χωρίς τη βοήθεια H/Y):

X	Y	XY	X²	\hat{Y}	$\hat{u}=Y-\hat{Y}$	\hat{u}^2
2	4	8	4	4,50	-0,50	0,25
3	7	21	9	6,25	0,75	0,5625
1	3	3	1	2,75	0,25	0,0625
5	9	45	25	9,75	-0,75	0,5625
9	17	153	81	16,75	0,25	0,0625
<u>Άθροισμα:</u> 20	40	230	120	40,00	0	1,5

Οι κανονικές εξισώσεις θα είναι:

$$\Sigma Y_i = N\hat{\beta}_1 + \hat{\beta}_2 \Sigma X_i \qquad 40 = 5\hat{\beta}_1 + 20\hat{\beta}_2 \qquad \hat{\beta}_1 = 1$$

è

è

$$\sum X_i Y_i = \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2 \qquad 230 = 20\hat{\beta}_1 + 120\hat{\beta}_2 \qquad \hat{\beta}_2 = 1,75$$

Πίνακας αποκλίσεων από μέσες τιμές:

	$x = X - \bar{X}$	$y = Y - \bar{Y}$	xy	x^2	y^2
	-2	-4	8	4	16
	-1	-1	1	1	1
	-3	-5	15	9	25
	1	1	1	1	1
	5	9	45	25	81
<u>ΑΘΡΟΙΣΜΑ</u>	0	0	70	40	124

Από τα αποτελέσματα του πίνακα αποκλίσεων έχουμε:

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 8 - 1,75 * 4 = 1$$

$$\hat{\beta}_2 = \frac{\sum xy}{\sum x_i^2} = \frac{70}{40} = 1,75$$

$$ESS = \hat{\beta}_2 \sum xy = 1,75 * 70 = 122,5$$

$$TSS = \sum y^2 = 124$$

$$RSS = TSS - ESS = 124 - 122,5 = 1,5$$

$$R^2 = ESS / TSS = 0,988$$

ΥΠΟΛΟΓΙΣΜΟΣ $\hat{\sigma}^2$, $VAR(\hat{\beta}_1)$, $VAR(\hat{\beta}_2)$

$$\hat{\sigma}^2 = \sum \hat{u}^2 / (N-2) = 1,5 / (5-2) = 0,5$$

$$VAR(\hat{\beta}_1) = 0,5 / 40 = 0,0125$$

$$VAR(\hat{\beta}_2) = \hat{\sigma}^2 \left(\frac{1}{N} + \frac{\bar{X}^2}{\sum X_i^2} \right) = 0,5 \left(\frac{1}{5} + \frac{16}{40} \right) = \frac{12}{40} = 0,3$$

Άρα $s.e(\hat{\beta}_1)=\sqrt{0,3}=0,5477$, $s.e(\hat{\beta}_2)=\sqrt{0,0125} \approx 0,1118$

Για $N=5$ και 95% επίπεδο εμπιστοσύνης $t_{0,025}=3,182$

Άρα το διάστημα εμπιστοσύνης 95% για τον $\hat{\beta}_1$ θα είναι: $1 \pm 3,182 * 0,5477$, δηλαδή απο -0,74 μέχρι 2,74 και για τον $\hat{\beta}_2$ θα είναι: $1,75 \pm 3,182 * (0,1118)$ δηλαδή από 1,39 μέχρι 2,11. Παρατηρήστε 'τοι μέσα στο διάστημα εμπιστοσύνης για τον $\hat{\beta}_1$ συμπεριλαμβάνεται και το μηδέν, άρα ο $\hat{\beta}_1$ δεν είναι στατιστικά σημαντικός.

Αντίθετα το διάστημα εμπιστοσύνης για τον $\hat{\beta}_2$ δεν περιέχει το μηδέν , άρα ο $\hat{\beta}_2$ είναι στατιστικά σημαντικός, πάντα για 95% επίπεδο εμπιστοσύνης.

Ισοδύναμα στο ίδιο συμπέρασμα μπορούμε να καταλήξουμε και με το κλασσικό t-test για τους $\hat{\beta}_1, \hat{\beta}_2$: Πράγματι, για τον $\hat{\beta}_1$:

$t = \frac{\hat{\beta}_1}{s.e(\hat{\beta}_1)} = \frac{1}{0,5477} = 1,826 < 3,182$ άρα $\hat{\beta}_1$ όχι στατιστικά σημαντικός για το 95% επίπεδο εμπιστοσύνης.

Ομοίως για τον $\hat{\beta}_2$: $t = \frac{\hat{\beta}_2}{s.e(\hat{\beta}_2)} = \frac{1,75}{0,1118} \approx 15,7 > 3,182$ άρα $\hat{\beta}_2$ στατιστικά σημαντικός για το 95% επίπεδο εμπιστοσύνης.

ΣΗΜΕΙΩΣΗ: Παρατηρήστε ότι λόγω του πολύ μικρού μεγέθους του δείγματος η κρίσιμη τιμή για το t-test είναι σημαντικά μεγαλύτερη του 2.

Διάστημα εμπιστοσύνης για το σ^2

Θα έχουμε για το 95% διάστημα εμπιστοσύνης κατώτερη τιμή:

$$\frac{(N-2)\hat{\sigma}^2}{X_{0,975}^2} = \frac{3*0,5}{9,35} = \mathbf{0,16}$$

και κατώτερη τιμή : $\frac{(N-2)\hat{\sigma}^2}{X_{0,025}^2} = \frac{3*0,5}{0,216} = \mathbf{6,94}$

Έλεγχος με το στατιστικό F

$$F = \frac{ESS/1}{RSS/(N-2)} = \frac{122,5}{1,5/3} = \mathbf{245,0} . \text{ Με } F_{0,95}(1,3) = \mathbf{10,1} \text{ απορρίπτεται η } H_0$$

$$(\hat{\beta}_2 = 0) \text{ καθώς } F > F_{0,95}(1,3)$$

Έλεγχος για το συντελεστή συσχέτισης

$$t = \frac{R\sqrt{N-2}}{\sqrt{(1-R^2)}} = \frac{\sqrt{0,988*3}}{\sqrt{(1-0,988)}} \approx \mathbf{15,7} > \mathbf{3,182}$$

Άρα ο R στατιστικά σημαντικός.

Ισοδυναμία στατιστικών ελέγχων

$$F = \mathbf{245,0} \Leftrightarrow \sqrt{F} = \sqrt{\mathbf{245,0}} \approx \mathbf{15,7}$$

$$t = \frac{\hat{\beta}_2}{s.e.(\hat{\beta}_2)} \approx \mathbf{15,7}$$

$$t = \frac{R\sqrt{N-2}}{\sqrt{(1-R^2)}} \approx \mathbf{15,7}$$

ΚΕΦΑΛΑΙΟ 3

ΓΕΝΙΚΑ ΣΤΟΙΧΕΙΑ ΣΧΕΤΙΚΑ ΜΕ ΤΙΣ ΠΡΟΒΛΕΨΕΙΣ

3.1 Εισαγωγή

Οι προβλέψεις αποτελούν ένα αναπόσπαστο κομμάτι της καθημερινότητάς μας. Αυτό αφορά είτε τις επιχειρήσεις, είτε τα κράτη, ή ακόμα και το κάθε άτομο ξεχωριστά. Για αυτό το λόγο το πεδίο των προβλέψεων έχει γίνει αντικείμενο εκτεταμένης έρευνας και ανάπτυξης. Σε αυτό βοηθάει πολύ η ανάπτυξη της τεχνολογίας που παρέχει πλέον στους ερευνητές και στους στατιστικούς πολύτιμα στατιστικά εργαλεία που κάνει την υπολογιστική διαδικασία, σαφώς πιο εύκολη. Αυτό γίνεται πιο εμφανές αν αναλογιστεί κανείς πως τα τελευταία είκοσι χρόνια, παρατηρείται έντονη η τάση του σχεδιασμού και της δημιουργίας διάφορων πληροφοριακών συστημάτων για την εφαρμογή μεθόδων πρόβλεψης και την εξαγωγή προβλέψεων.

Φυσικά με την διάδοση του διαδικτύου υπάρχει μία προσπάθεια δημιουργίας διαδικτυακών εφαρμογών οι οποίες σχετίζονται με τον τομέα των προβλέψεων. Εξαιτίας των πλεονεκτημάτων των διαδικτυακών εφαρμογών δημιουργήθηκε η ιδέα της υλοποίησης μίας τέτοιας εφαρμογής χρησιμοποιώντας ένα στατιστικό περιβάλλον λογισμικού, το οποίο ενσωματώνει καινοτόμα χαρακτηριστικά, όπως είναι η αυτόματη σύγκριση των μεθόδων πρόβλεψης και η συνεχής αυτόματη ανανέωση των διαθέσιμων χρονοσειρών. Στη συνέχεια της παρούσας διπλωματικής αναλύεται το θεωρητικό υπόβαθρο που χρησιμοποιήθηκε στην υλοποίηση της εφαρμογής, αλλά και ενός πειράματος σχετικά με την σύγκριση των μεθόδων πρόβλεψης σε συγκεκριμένες χρονοσειρές.

3.2 Βασικά Βήματα Προβλέψεων

Για την επίτευξη της μεγαλύτερης δυνατής ακρίβειας και αξιοπιστίας κατά την διάρκεια της πρόβλεψης, είναι απαραίτητο ο σχεδιασμός και η εκτέλεση να γίνεται με ένα πολύ προσεκτικό και καλά σχεδιασμένο τρόπο.

Βήμα 1ο: Ορισμός του προβλήματος

Ο ορισμός του προβλήματος είναι αρκετές φορές το πιο δύσκολο μέρος στην διαδικασία πρόβλεψης. Κατά τη διάρκεια αυτού του βήματος θα πρέπει να γίνει απολύτως σαφές τι θέλουμε να προβλέψουμε, αλλά και να καθορίσουμε όλες τις παραμέτρους του προβλήματος που μελετάται. Επίσης θα πρέπει να γίνει αποτίμηση, όλων των συνιστωσών που πιθανώς επηρεάζουν το τελικό αποτέλεσμα.

Βήμα 2ο: Συλλογή των πληροφοριών

Γενικά υπάρχουν δύο τύποι δεδομένων, που μπορεί να χρησιμοποιηθούν σε προβλήματα προβλέψεων. Από τη μία είναι τα στατιστικά ιστορικά δεδομένα και από την άλλη η συσσωρευμένη εμπειρία και κριτική ικανότητα διάφορων ανθρώπων. Και

οι δύο τύποι πληροφοριών θα πρέπει να λαμβάνονται υπόψιν στις προβλέψεις. Η συλλογή των ιστορικών δεδομένων είναι πολύ σημαντική, διότι αποτελούν την βάση με τα οποία θα εφαρμόσουμε τις μεθόδους πρόβλεψης και θα εξάγουμε τις προβλέψεις μας για αυτά.

Βήμα 3ο: Προετοιμασία χρονοσειρών

Με αυτό το βήμα μπορούμε να αποκτήσουμε μία ολοκληρωμένη αίσθηση των διαθέσιμων δεδομένων, και έτσι να αναγνωρίσουμε πιθανά

λανθάνοντα πρότυπα, όπως σημαντική τάση ή εποχιακότητα και ασυνέχειες. Η ανάλυση αυτή βοηθά και στην αναγνώριση πιθανής ανάγκης για χρήση κριτικής αξιολόγησης από μέρους ειδικών. Η προσαρμογή των δεδομένων, η απομόνωση των συνιστωσών της χρονοσειράς και η αναγνώριση και απομόνωση ειδικών γεγονότων και ενεργειών μας οδηγεί σε μία εξομλυμένη σειρά δεδομένων, έτοιμη για την εφαρμογή μοντέλων πρόβλεψης.

Βήμα 4ο: Επιλογή μεθόδων πρόβλεψης

Η κατασκευή ενός μοντέλου πρόβλεψης περιλαμβάνει την προσαρμογή των δεδομένων που έχουν συλλεχθεί σε ένα υπάρχον μοντέλο, με στόχο την ελαχιστοποίηση των σφαλμάτων. Συνεπώς απαιτείται η κατάλληλη επιλογή μοντέλου που πρέπει ιδανικά να έχει τη κατάλληλη ισορροπία μεταξύ ενός πολύπλοκου μοντέλου και ενός πιο απλοϊκού. Και αυτό διότι το πιο πολύπλοκο μοντέλο μπορεί να έχει μεγαλύτερη ακρίβεια, αλλά το πιο απλό μοντέλο είναι πιο εύκολο να γίνει κατανοητό. Άρα επιτυγχάνεται η σωστή επιλογή των κατάλληλων μεθόδων πρόβλεψης, αλλά και επιλογή των παραμέτρων τους, σύμφωνα με τα ιστορικά δεδομένα τους.

Βήμα 5ο: Χρήση και αξιολόγηση των μοντέλων πρόβλεψης

Αφού επιλέξουμε το κατάλληλο μοντέλο πρόβλεψης και κάνουμε σωστή επιλογή των παραμέτρων του, στο τελευταίο στάδιο το επιλεγμένο μοντέλο χρησιμοποιείται ώστε να παραχθούν οι ζητούμενες προβλέψεις. Η αξιολόγηση του μοντέλου μπορεί να επιτευχθεί με το πέρασμα του χρόνου, καθώς νέα δεδομένα γίνονται διαθέσιμα. Και αυτό διότι έτσι μπορεί να γίνει σύγκριση ανάμεσα στις παραχθείσες προβλέψεις και στα καινούργια διαθέσιμα δεδομένα. Αυτή η αξιολόγηση είναι δυνατή

χρησιμοποιώντας εξειδικευμένους στατιστικούς δείκτες. Τέλος, η παρακολούθηση του προτύπου χρονοσειράς και των σφαλμάτων της πρόβλεψης συμβάλλει στην έγκαιρη αντιμετώπιση διαρθρωτικών αλλαγών και στην εξάλειψη της προκατάληψης στις προβλέψεις.

3.3 Τα Πεδία Εφαρμογής των Προβλέψεων

Όπως έχει ήδη αναφερθεί οι προβλέψεις είναι πάρα πολύ σημαντικές και βρίσκουν εφαρμογή σε πληθώρα επιστημονικών και ερευνητικών πεδίων, αλλά και σε καθημερινές πτυχές της ζωής μας. Παρακάτω παρατίθενται μερικά πεδία εφαρμογής των τεχνικών προβλέψεων.

Οικονομία

Οι κυβερνήσεις κυρίως, αλλά και άλλοι οργανισμοί έχουν ανάγκη να προβλέψουν βασικούς οικονομικούς δείκτες, όπως το Ακαθάριστο Εγχώριο Προϊόν, την ανεργία, τον πληθωρισμό κλπ. Επομένως οι προβλέψεις είναι απαραίτητες για τον καθορισμό των δημοσιονομικών και νομισματικών πολιτικών τους.

Χρηματοοικονομικά

Στον κόσμο των επιχειρήσεων, αλλά κυρίως των επενδύσεων είναι απαραίτητες οι προβλέψεις για τις τιμές των μετοχών, των επιτοκίων δανεισμού, της ισοτιμίας και των τιμών των αγαθών. Συνεπώς οι τεχνικές προβλέψεων έχουν εφαρμοσθεί αι μελετηθεί ευρέως στην πρόβλεψη γενικών τιμών χρηματιστηρίων αλλά και στην πρόβλεψη μεταβλητότητας.

Παραγωγή

Οι επιχειρήσεις έχουν ανάγκη να προβλέπουν, εκτός από τα επίπεδα τιμών που θα διαμορφωθούν μελλοντικά, αλλά και τις ποσότητες των πρώτων υλών που θα χρειαστούν για την παραγωγική διαδικασία. Αυτό για να γίνει θα πρέπει να μελετηθούν πρώτα από όλα οι προβλέψεις που αφορούν την ζήτηση.

Marketing

Και στον τομέα του marketing οι προβλέψεις παίζουν σπουδαίο ρόλο. Αποφάσεις για το κόστος της διαφήμισης εξαρτάται άμεσα από τις προβλέψεις για την ζήτηση στην αγορά.

Περιβάλλον και Κλίμα

Για θέματα όπως η ενεργειακή ζήτηση, η διαχείριση υδάτινων πόρων, η μετεωρολογία και η ρύπανση αντιμετωπίζονται συχνά με χρήση τεχνικών προβλέψεων.

Κοινωνία

Δημογραφικές προβλέψεις διεξάγονται συχνά με αντικείμενο τον πληθυσμό και την πληθυσμιακή σύνθεση διαφόρων περιοχών του κόσμου. Επίσης προβλέψεις γίνονται και τομείς όπως η εγκληματικότητα και η διάδοση επιδημιών.

Τουρισμός

Ο τουρισμός αποτελεί ως γνωστόν την βασικότερη υπηρεσία που εξάγει η χώρα μας. Για αυτό το λόγο χρησιμοποιούνται μοντέλα προβλέψεων για την εκτίμηση των συνολικών αφίξεων τουριστών, τις διανυκτερεύσεις ανά περιοχή και το ταξιδιωτικό συνάλλαγμα.

Μεταφορές και μετακινήσεις

Η πρόβλεψη του κυκλοφοριακού φόρτου και η πρόβλεψη καταληψιμότητας ανά μονάδα χρόνου σε διαφορετικά γεωγραφικά σημεία και προς συγκεκριμένη κατεύθυνση είναι μερικά προβλήματα που μπορούν να αντιμετωπιστούν με εξειδικευμένα μοντέλα και τεχνικές προβλέψεων.

ΚΕΦΑΛΑΙΟ 4

ΤΕΧΝΙΚΕΣ ΠΡΟΒΛΕΨΕΩΝ

4.1 Τεχνικές Προβλέψεων

Το πρώτο και σημαντικότερο βήμα, για την εξαγωγή προβλέψεων αλλά και γενικότερα για την στατιστική ανάλυση, είναι η συλλογή των απαραίτητων ιστορικών στοιχείων της μεταβλητής που θέλουμε να μελετήσουμε. Τα δεδομένα θα πρέπει σαφώς να είναι επικαιροποιημένα έτσι ώστε η πρόβλεψη να συμβαδίζει με βάση το τωρινό περιβάλλον της μεταβλητής για να είναι όσο το δυνατόν πιο ακριβής η πρόβλεψη. Το δεύτερο βήμα αφορά στη χρήση της κατάλληλης μεθόδου πρόβλεψης, με σκοπό την εξαγωγή προβλέψεων μεγαλύτερης ακρίβειας. Τελευταίο βήμα, είναι η αξιολόγηση των προβλέψεων που παρήχθησαν χρησιμοποιώντας κατάλληλους στατιστικούς δείκτες.

4.2 Κατηγορίες των Μεθόδων Πρόβλεψης

Υπάρχουν δύο μεγάλες κατηγορίες μεθόδων πρόβλεψης στις οποίες εντάσσονται οι τεχνικές προβλέψεων που έχουν αναπτυχθεί μέχρι σήμερα. Αυτές είναι οι εξής:

- *Ποσοτικές προβλέψεις* (quantitative forecasting)
- *Ποιοτικές προβλέψεις* (qualitative ή judgmental forecasting)

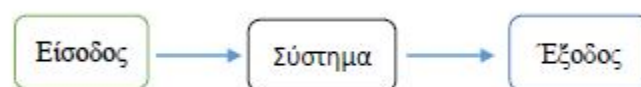
4.2.1 Η Ποσοτική Πρόβλεψη (quantitative forecasting)

Οι μέθοδοι αυτές βασίζονται στη μαθηματική μοντελοποίηση και άρα είναι αντικειμενικές και επαναλήψιμες (δηλαδή παράγουν το ίδιο αποτέλεσμα κάθε φορά που εισάγουμε τα ίδια δεδομένα). Οι ποσοτικές

μέθοδοι απαιτούν μία σειρά από αριθμητικά δεδομένα που όμως δεν είναι πάντα διαθέσιμα ή αξιόπιστα. Οι ποσοτικές μέθοδοι μπορούν να διακριθούν σε μοντέλα χρονοσειρών (*time series models*) και σε αυτές που βασίζονται σε αιτιοκρατικά ή επεξηγηματικά μοντέλα (*causal relationship* ή *explanatory models*).

4.2.1.1 Το Μοντέλο των Χρονοσειρών

Πρόκειται για το πιο διαδεδομένο είδος ποσοτικού μοντέλου πρόβλεψης. Το μοντέλο χρονοσειρών προϋποθέτει ότι η απαραίτητη πληροφορία για την πρόβλεψη περιέχεται στη χρονοσειρά των στοιχείων. Χρονοσειρά είναι μία σειρά παρατηρήσεων που λαμβάνονται σε κανονικά διαστήματα μέσα σε ένα καθορισμένο χρονικό διάστημα. Για παράδειγμα, αν καταγράφουμε τις μηνιαίες τιμές του χρυσού στο χρηματιστήριο μετάλλων του Λονδίνου για χρονικό διάστημα 5 ετών, τότε έχουμε στη διάθεσή μας μια χρονοσειρά των μηνιαίων τιμών του χρυσού. Η ανάλυση χρονοσειράς κάνει την υπόθεση ότι μπορεί να γίνει πρόβλεψη με βάση τα μοτίβα (*patterns*) των διαθέσιμων δεδομένων. Έτσι, η ανάλυση αυτή αναζητάει τάσεις, κυκλικότητα, περιοδικότητα κτλ. στα δεδομένα προκειμένου να δημιουργήσει ένα μοντέλο πρόβλεψης. . Στα μοντέλα χρονοσειρών περιλαμβάνονται οι μέθοδοι αποσύνθεσης, οι μέθοδοι εξομάλυνσης καθώς και οι αυτό-παλινδρομικές μέθοδοι κινητού μέσου όρου.



Μέθοδοι αποσύνθεσης

Για την μέθοδο της αποσύνθεσης εντοπίζουμε τις συνιστώσες των διαθέσιμων δεδομένων (δηλ., της χρονοσειράς) προκειμένου να καταλάβουμε τον τρόπο με τον οποίο συμπεριφέρονται. Ακριβώς σε αυτή τη συμπεριφορά βασίζεται η πρόβλεψη. Οι συνιστώσες είναι η τάση (T), ο κύκλος (C), η εποχικότητα (S) και η τυχαιότητα (R). Αν εξετάσουμε μια οποιαδήποτε χρονοσειρά, διαπιστώνουμε ότι αποτελεί συνδυασμό ενός ή περισσότερων από τα παραπάνω στοιχεία. Σκοπός των μεθόδων αποσύνθεσης είναι η απομόνωση των συνιστωσών αυτών με τη μέγιστη δυνατή ακρίβεια.

Μέθοδοι εξομάλυνσης

Οι μέθοδοι εξομάλυνσης μπορούν να εφαρμοστούν πολύ εύκολα και είναι κατάλληλες για προβλέψεις με βραχυπρόθεσμο ορίζοντα. Σε αυτές τις μεθόδους γίνεται χρήση των ιστορικών δεδομένων για τον προσδιορισμό μίας εξομαλυμένης τιμής για τις χρονοσειρές. Στη συνέχεια, αυτή η εξομαλυμένη τιμή προεκτείνεται ώστε να αποτελέσει πρόβλεψη για τη μελλοντική τιμή της χρονοσειράς. Η βασική ιδέα των μεθόδων εξομάλυνσης είναι ότι υπάρχει ένα λανθάνον πρότυπο συμπεριφοράς το οποίο ακολουθούν οι τιμές των μεταβλητών πρόβλεψης. Έτσι, ελαχιστοποιείται η τυχαιότητα που υπάρχει στην χρονοσειρά με την πρόβλεψη πλέον να βασίζεται σε ένα εξομαλυμένο πρότυπο συμπεριφοράς. Υπάρχουν δύο κατηγορίες μεθόδων εξομάλυνσης: οι κινητού μέσου όρου και οι εκθετικής εξομάλυνσης.

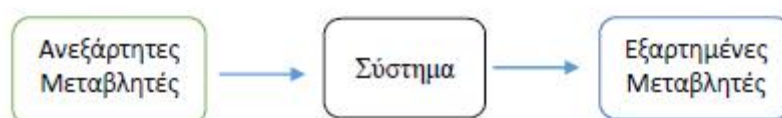
Αυτοπαλινδρομικές μέθοδοι κινητού μέσου όρου (ARMA)

Οι συγκεκριμένες μέθοδοι είναι στοχαστικά μαθηματικά μοντέλα με τα οποία περιγράφεται η χρονική εξέλιξη ενός φυσικού μεγέθους. Τα

μοντέλα περιέχουν το τυχαίο παράγοντα, τα ιστορικά δεδομένα και άλλους στοχαστικούς παράγοντες. Το μοντέλο που προκύπτει τελικά είναι ένας γραμμικός συνδυασμός των παραπάνω ποσοτήτων. Τα αυτοπαλινδρομικά μοντέλα βασίζονται στην παραδοχή της αλληλεξάρτησης μεταξύ των τιμών που λαμβάνει η χρονοσειρά τις διάφορες χρονικές στιγμές.

4.2.1.2 Το Αιτιοκρατικό ή Επεξηγηματικό Μοντέλο

Στα αιτιοκρατικά μοντέλα χρησιμοποιείται μία αρκετά διαφορετική προσέγγιση για την δημιουργία πρόβλεψης. Θεωρούν ότι η μεταβλητή για την οποία θέλουμε να κάνουμε πρόβλεψη είναι εξαρτημένη με κάποιο τρόπο από μία ή περισσότερες παραμέτρους. Η δυσκολία έγκειται στην εύρεση της μαθηματικής σχέσης με την οποία επηρεάζεται η ζητούμενη μεταβλητή από τις παραμέτρους αυτές. Για παράδειγμα, αν θεωρήσουμε ότι η ζήτηση για σίδηρο οπλισμού εξαρτάται: α) από τη χρηματική αξία των συμβάσεων για δημόσια έργα που υπογράφονται μεταξύ του ΥΠΕΧΩΔΕ και των κατασκευαστικών εταιρειών και β) από τον αριθμό των οικοδομικών αδειών που εκδίδονται από τις πολεοδομίες της χώρας, τότε αν βρούμε τη μαθηματική τους σχέση μπορούμε να κατασκευάσουμε ένα μαθηματικό μοντέλο πρόβλεψης για την ζήτηση σε σίδηρο οπλισμού. Είναι προφανές ότι τα αιτιοκρατικά μοντέλα μπορεί να είναι πολύ περίπλοκα, ειδικά στην περίπτωση που λαμβάνονται υπόψιν πολλές παράμετροι. Στα αιτιοκρατικά μοντέλα ανήκουν οι μέθοδοι παλινδρόμησης και οι οικονομετρικές μέθοδοι.



Μέθοδοι παλινδρόμησης

Η παλινδρόμηση περιλαμβάνει διαδικασίες και τεχνικές μοντελοποίησης και ανάλυσης διαφόρων μεταβλητών, όπου το ζητούμενο είναι εύρεση συσχετίσεων μεταξύ μίας εξαρτημένης και μίας ή και περισσότερων ανεξάρτητων μεταβλητών. Ειδικότερα, η ανάλυση της παλινδρόμησης μας βοηθά να καταλάβουμε με ποιο τρόπο αλλάζει η τιμή της εξαρτημένης μεταβλητής, όταν αλλάζουμε την τιμή μίας ανεξάρτητης μεταβλητής, ενώ οι υπόλοιπες διατηρούνται σταθερές. Στην περίπτωση ύπαρξης μίας ανεξάρτητης μεταβλητής η μέθοδος ονομάζεται “Απλή Παλινδρόμηση” ενώ στην περίπτωση περισσότερων, “Πολλαπλή Παλινδρόμηση”.

Οικονομετρικές μέθοδοι

Εάν οι ανεξάρτητες μεταβλητές συσχετίζονται μεταξύ τους τότε προκύπτει ένα σύστημα ταυτόχρονων εξισώσεων. Αυτό το σύστημα αποτελεί ένα μοντέλο, το οποίο ονομάζεται οικονομετρικό και συναντάται συχνά σε περιπτώσεις οικονομικών ή επιχειρησιακών σχέσεων.

4.2.2 Η Ποιοτική Πρόβλεψη (qualitative ή judgmental forecasting)

Στην κατηγορία αυτή περιλαμβάνονται μέθοδοι στις οποίες η πρόβλεψη γίνεται από έναν ή περισσότερους ειδικούς με βάση την γνώση, την εμπειρία και το ένστικτό τους. Αυτού του είδους η πρόβλεψη είναι υποκειμενική και συχνά περιλαμβάνει συγκεκριμένους τύπους μεροληψιών και προκαταλήψεων που έχουν αρνητική απόδοση στην ακρίβεια των προβλέψεων.

4.2.2.1 Πάνελ

Η μέθοδος αυτή βασίζεται στο γεγονός ότι μια πολυπληθής ομάδα ανθρώπων από διαφορετικές θέσεις μπορεί να κάνει μία πιο αξιόπιστη πρόβλεψη απ' ό,τι ένας μεμονωμένος ή λίγοι άνθρωποι. Έτσι, διοργανώνονται ανοιχτές συναντήσεις με ελεύθερη ανταλλαγή απόψεων μεταξύ ανθρώπων από όλο το φάσμα των θέσεων ενός οργανισμού. Ένα μειονέκτημα της μεθόδου αυτής είναι ότι η άποψη των υφιστάμενων «σκεδάζεται» ή υποβαθμίζεται από την άποψη των ανώτερων στην ιεραρχία. Αυτό το μειονέκτημα προσπαθεί να διορθώσει η μέθοδος Delphi.

4.2.2.2 Μέθοδος Delphi

Η μέθοδος αυτή είναι μια τεχνική πρόβλεψης, στόχος της οποίας είναι η προσέγγιση μιας συμφωνίας μεταξύ μιας ομάδας ειδικών, διατηρώντας την ανωνυμία τους. Η ιδέα πίσω από αυτήν είναι ότι ενώ οι ειδικοί δεν θα συμφωνήσουν σε όλα τα ζητήματα, εντούτοις σε ό,τι συμφωνήσουν αυτά κατά πάσα πιθανότητα που θα συμβούν. Σύμφωνα με τη μέθοδο αυτή:

1. γίνεται επιλογή των συμμετεχόντων, συνήθως ειδικών από διαφορετικές θέσεις ή επιστημονικά υπόβαθρα,
2. μέσω ενός ερωτηματολογίου συλλέγονται οι απόψεις όλων (χωρίς ο ένας να δει ή να γνωρίζει τους υπόλοιπους συμμετέχοντες),

3. οι απαντήσεις όλων ταξινομούνται και επανατροφοδοτούνται (feedback) στους συμμετέχοντες μαζί με ένα καινούργιο ερωτηματολόγιο,

4. το βήμα 3 επαναλαμβάνεται όσες φορές κρίνεται απαραίτητο προκειμένου να επιτευχθεί μια συμφωνία μεταξύ των συμμετεχόντων. Συνήθως 3 ή 4 «γύροι» είναι αρκετοί.

Λόγω της ανωνυμίας του καθενός και της ίδιας βαρύτητας όλων των απόψεων, με τη μέθοδο Delphi αποφεύγεται το μειονέκτημα της πρώτης μεθόδου (πάνελ). Από την άλλη, η μέθοδος είναι σχετικά χρονοβόρα.

4.3 Οι Μέθοδοι Πρόβλεψης Συνεχούς Ζήτησης

Στην παρούσα διπλωματική, χρησιμοποιήσαμε 8 μεθόδους πρόβλεψης χρονοσειρών, στις οποίες έγινε αργότερα σύγκριση για το ποια μέθοδος δίνει τα καλύτερα αποτελέσματα με βάση τα δεδομένα και τους στατιστικούς δείκτες. Επίσης εξετάστηκε και η σειρά κατάταξης της κάθε μεθόδου βάσει των δεικτών που εξετάστηκαν. Επομένως στην εφαρμογή ενσωματώθηκαν οι παρακάτω μέθοδοι πρόβλεψης:

1. Απλοϊκή μέθοδος - Naive
2. Κινητός μέσος όρος – KMO
3. Απλή γραμμική παλινδρόμηση – LRL
4. Απλή εκθετική εξομάλυνση – SES
5. Εκθετική εξομάλυνση γραμμικής τάσης – Holt
6. Εκθετική εξομάλυνση μη γραμμικής τάσης – Damped

7. Error-Trend-Seasonal ή Exponential Smoothing method – ETS

8. Κλασική μέθοδος Theta

Στις παρακάτω παραγράφους αναφέρονται αναλυτικά τα παραπάνω μοντέλα, αλλά και μερικά ακόμα που δεν έχουν ενσωματωθεί στην εφαρμογή.

4.3.1 Η Απλοϊκή μέθοδος – Naive

Στην ουσία η Naive είναι η πιο απλή μέθοδος πρόβλεψης. Η συγκεκριμένη στατιστική μέθοδος δίνει ως πρόβλεψη για την επόμενη χρονική περίοδο, την ίδια τιμή με την πραγματική τιμή της τωρινής περιόδου. Η αποδοτικότητά της είναι καλή για αποεποχικοποιημένες χρονοσειρές και προβλέψεις μίας περιόδου, διότι συνήθως η επόμενη τιμή της χρονοσειράς δεν διαφέρει σημαντικά από την προηγούμενη παρατήρηση που είναι διαθέσιμη. Συνήθως όμως δεν δίνει ακριβείς προβλέψεις και για αυτό το λόγο χρησιμοποιείται περισσότερο ως σημείο αναφοράς (benchmark) για άλλες πιο πολύπλοκες μεθόδους. Ο τύπος που χρησιμοποιείται για την περιγραφή της μεθόδου είναι:

$$F_t = Y_{t-1}$$

4.3.2 Οι Μέθοδοι Μέσου Όρου

Με τις μεθόδους των μέσων όρων γίνεται χρήση των ιστορικών δεδομένων, ώστε να προσδιοριστεί μία εξομαλυμένη τιμή για τις χρονοσειρές κάτι το οποίο ομαλοποιεί την χρονοσειρά και έτσι μπορούν να χρησιμοποιηθούν και ως μέθοδοι πρόβλεψης. Αυτό γίνεται πολύ

εύκολα, προεκτείνοντας αυτή τη τιμή για την χρονική περίοδο που θέλουμε να κάνουμε την πρόβλεψη, και ως συνέπεια αυτή η τιμή αποτελεί την μελλοντική τιμή της χρονοσειράς. Μερικοί από τους μέσους όρους που χρησιμοποιούμε για την παραγωγή προβλέψεων είναι ο απλός μέσος όρος και ο κινητός μέσος όρος.

4.3.2.1 Ο Απλός μέσος όρος

Η μέθοδος του απλού μέσου όρου βασίζεται στη λογική της εύρεσης του μέσου όρου όλων των διαθέσιμων ιστορικών δεδομένων και η χρησιμοποίηση αυτής της τιμής ως πρόβλεψη. Γενικά η χρήση αυτής της μεθόδου προκρίνεται για περιπτώσεις που τα δεδομένα δεν παρουσιάζουν τάση ή έντονη εποχιακότητα, αλλά εμφανίζουν μία σταθερότητα με την πάροδο του χρόνου. Επομένως ο τύπος που μας δίνει αυτή την πρόβλεψη είναι εξής:

$$F_{t+1} = \frac{1}{t} \cdot \sum_{i=1}^t Y_i$$

4.3.2.2 Ο Κινητός μέσος όρος

Για να είναι εφικτός ο υπολογισμός του κινητού μέσου όρου των δεδομένων μίας χρονοσειράς θα πρέπει να καθοριστεί αρχικά ο αριθμός των δεδομένων που θα συμπεριληφθεί στον υπολογισμό αυτό. Ο όρος κινητός μέσος όρος περιγράφει την διαδικασία κατά την οποία καθώς μία καινούργια παρατήρηση γίνεται διαθέσιμη, μπορεί να υπολογιστεί ένας καινούργιος μέσος όρος, στον οποίο παραλείπεται η πιο παλιά

παρατήρηση προκειμένου να συμπεριληφθεί η πιο πρόσφατη. Αυτός ο νέος

κινητός μέσος όρος θα είναι η πρόβλεψη για την επόμενη περίοδο. Θα πρέπει να τονιστεί πως ο αριθμός των παρατηρήσεων που χρησιμοποιούνται για την εξαγωγή του μέσου όρου πρέπει οπωσδήποτε να παραμείνει σταθερός καθ' όλη την διαδικασία πρόβλεψης και εμπεριέχει πάντα τις πιο πρόσφατες παρατηρήσεις. Αν η χρονοσειρά αποτελείται από n δεδομένα και το μήκος των παρατηρήσεων που θα χρησιμοποιηθούν για τον υπολογισμό κάθε μέσου όρου είναι k , τότε ονομάζεται κινητός μέσος όρος μήκους k ή ΚΜΟ(k) ή ΜΑ(k). Η σχέση που χρησιμοποιείται για την εφαρμογή της μεθόδου του κινητού μέσου όρου είναι:

$$F_{t+1} = \frac{1}{k} \cdot \sum_{i=t-k+1}^t Y_i$$

Είναι προφανές πως αν το k τεθεί ίσο με τη μονάδα, τότε ουσιαστικά η τιμή του ΚΜΟ(1) ταυτίζεται με την απλοϊκή μέθοδο (Naive). Είναι επίσης ξεκάθαρο πως ο κινητός μέσος όρος δίνει ιδιαίτερη σημασία στις τελευταίες ίσου μήκους παρατηρήσεις, αλλά δυστυχώς κάτι τέτοιο απαιτεί περισσότερο χώρο αποθήκευσης, διότι πρέπει να αποθηκεύονται όλες οι παρατηρήσεις από τις οποίες γίνεται η εξαγωγή του κινητού μέσου όρου κάθε φορά και όχι μόνο η τιμή του μέσου όρου.

Ένα από τα μειονεκτήματα των μεθόδων μέσων όρων είναι ότι δεν γίνεται να διαχειριστεί επιτυχώς ορισμένα κύρια χαρακτηριστικά των χρονοσειρών, όπως είναι η τάση και η εποχιακότητα. Για αυτό το λόγο το μοντέλο μέσων όρων χρησιμοποιείται κυρίως για την εξομάλυνση των συνιστωσών της εποχιακότητας και της τάσης των χρονοσειρών, ώστε να

προκύψει μία εκτίμηση της γραμμής τάσης-κύκλου. Άρα τα μοντέλα αυτά θα μπορούσαμε να πούμε ότι είναι κατάλληλα για εργαλεία αποσύνθεσης παρά για μεθόδους πρόβλεψης.

4.3.3 Οι Μέθοδοι Εκθετικής Εξομάλυνσης

Οι μέθοδοι πρόβλεψης της εκθετικής εξομάλυνσης είναι στην ουσία μία προέκταση του προτύπου των ιστορικών δεδομένων, όπως τάση και εποχιακότητα. Οι προβλέψεις παράγονται αφού γίνει η εξομάλυνση των παρατηρήσεων, προκειμένου να απομονωθούν τα πραγματικά πρότυπα από τις τυχαίες διακυμάνσεις. Ουσιαστικά, οι μέθοδοι εκθετικής εξομάλυνσης είναι μία προέκταση των κινητών μέσων όρων και ως εκ τούτου, έχουν ως βασική αρχή πως τα πιο πρόσφατα ιστορικά δεδομένα εμπεριέχουν και την πιο σημαντική πληροφορία. Άρα τα πιο πρόσφατα δεδομένα αποτελούν τον οδηγό για την εξαγωγή των προβλέψεων, ενώ στα πιο παλιά μειώνεται εκθετικά ο συντελεστής βαρύτητας. Για αυτό εξάλλου καλούνται και μέθοδοι εκθετικής εξομάλυνσης.

Οι μέθοδοι εκθετικής εξομάλυνσης παρουσιάστηκαν για πρώτη φορά στις αρχές του 1940 και γνώρισαν ευρεία εφαρμογή λόγω της επανάστασης της πληροφορικής το 1960. Οι μέθοδοι αυτοί είναι αρκετά δημοφιλείς στο πεδίο των προβλέψεων λόγω της απλότητας των μοντέλων που υιοθετούν, των περιορισμένων απαιτήσεών τους για αποθήκευση δεδομένων και το μειωμένο υπολογιστικό φόρτο που απαιτούν. Από ότι φαίνεται παρά την απλότητά τους, σύμφωνα με τα αποτελέσματα εμπειρικών μελετών, παρουσιάζουν πολύ καλά αποτελέσματα ακρίβειας σε σχέση με πιο πολύπλοκες μεθόδους. Αυτό συμβαίνει διότι οι μέθοδοι εκθετικής εξομάλυνσης δεν επηρεάζονται από τις ιδιομορφίες των προτύπων των δεδομένων, ούτε από εμφανιζόμενες

ακραίες τιμές, οι οποίες είναι συχνό φαινόμενο στα επιχειρησιακά δεδομένα.

Τα μοντέλα εκθετικής εξομάλυνσης κατηγοριοποιούνται σύμφωνα με τη γενική μορφή της γραφικής αναπαράστασης των δεδομένων με οριζόντιο άξονα τον χρόνο. Σύμφωνα με αυτό τον τρόπο της κατηγοριοποίησης προκύπτουν τέσσερα μοντέλα τάσης, τα μοντέλα σταθερού επιπέδου, γραμμικής τάσης, εκθετικής τάσης και φθίνουσας τάσης. Πριν προβούμε σε περαιτέρω ανάλυση του κάθε μοντέλου, κρίνεται σκόπιμο να αναφερθούν ορισμένα χαρακτηριστικά και για τα τέσσερα μοντέλα, ώστε να υπάρχει μία γενική εικόνα για την χρήση τους.

Αρχικά, το μοντέλο σταθερού επιπέδου υποθέτει την απουσία τάσης από τα δεδομένα. Συνήθως αυτά τα μοντέλα χρησιμοποιούνται ευρέως για προβλέψεις ενός βήματος ή όταν οι χρονοσειρές χαρακτηρίζονται από αυξημένο θόρυβο ή τυχαιότητα. Εν συνεχεία, το μοντέλο γραμμικής τάσης είναι πιο διαδεδομένο, διότι η πρόβλεψη γίνεται προκύπτει απλά από την προέκταση μίας ευθείας γραμμής για οποιαδήποτε χρονική στιγμή στο μέλλον. Ένα χαρακτηριστικό παράδειγμα που ενδείκνυται η χρήση του μοντέλου της εκθετικής τάσης είναι όταν βρισκόμαστε στην αρχή του κύκλου ζωής ενός προϊόντος, καθώς το ποσοστό αύξησης των πωλήσεων δύναται να μεγαλώνει συνεχώς.

Πολλές μελέτες έχουν δείξει πως τα μοντέλα γραμμικής και εκθετικής τάσης δίνουν γενικά υπεραισιόδοξα αποτελέσματα και οδηγούν σε ιδιαίτερα υψηλές τιμές προβλέψεων και συνεπώς σε αυξημένες τιμές του δείκτη Mean Error. Έτσι προέκυψε η ανάγκη εφαρμογής ενός άλλου μοντέλου, όπως το μοντέλο φθίνουσας τάσης, το οποίο στις περιπτώσεις μεσοπρόθεσμων και μακροπρόθεσμων προβλέψεων αποτελεί την καλύτερη επιλογή, διότι μειώνεται σταδιακά το μέγεθος κατά το οποίο

αυξάνονται οι τιμές της χρονοσειράς κάθε χρονική περίοδο. Παρακάτω αναλύονται και παρουσιάζονται τα μοντέλα για την καλύτερη κατανόησή τους.

4.3.3.1 Το Μοντέλο Σταθερού Επιπέδου (Simple Exponential Smoothing)

Το μοντέλο σταθερού επιπέδου ή αλλιώς απλή εκθετική εξομάλυνση, περιγράφεται από τις παρακάτω εξισώσεις:

$$et = Y_t - F_t$$

$$S_t = S_{t-1} + a \cdot et$$

$$F_{t+1} = S_t$$

Όπου et είναι το σφάλμα που προκύπτει από τη διαφορά της πραγματικής με την τιμή της πρόβλεψης για την ίδια χρονική περίοδο t . Το S_t δηλώνει το επίπεδο, η F_t είναι η πρόβλεψη που πραγματοποιείται και t είναι χρονική περίοδος. Η παράμετρος a αποτελεί τον συντελεστή εξομάλυνσης της μεθόδου και λαμβάνει τιμές στο διάστημα $[0, 1]$.

Από τις παραπάνω εξισώσεις βλέπουμε πως σε κάθε πρόβλεψη υπολογίζουμε το σφάλμα, διότι με αυτό κρατάμε την τιμή της πρόβλεψης αρκετά κοντά στο επίπεδο της πραγματικής χρονοσειράς. Αυτό βέβαια προϋποθέτει τον υπολογισμό της τιμής του επιπέδου από την πραγματική χρονοσειρά για την προηγούμενη χρονική στιγμή.

Αρχικό Επίπεδο

Βέβαια από τα παραπάνω προκύπτει το εξής ερώτημα σχετικά με την λειτουργία του μοντέλου. Αυτό σχετίζεται με το τι θα γίνει με την

πρόβλεψη για την πρώτη χρονική περίοδο για την οποία δεν υπάρχουν διαθέσιμα ιστορικά δεδομένα. Η απάντηση που δίνεται είναι ότι η πρώτη πρόβλεψη θα πάρει τη τιμή του αρχικού επιπέδου. Επομένως θα πρέπει να δοθεί ιδιαίτερη προσοχή στην σωστή επιλογή του αρχικού επιπέδου του μοντέλου πρόβλεψης για την εξαγωγή των προβλέψεων με όσο το δυνατόν μεγαλύτερη ακρίβεια.

Μερικές συνηθισμένες μεθοδολογίες για την πρώτη τιμή του επιπέδου (S_0) της απλής εκθετικής εξομάλυνσης είναι:

- Ο μέσος όρος όλων των διαθέσιμων ιστορικών δεδομένων
- Ο μέσος όρος όλων των n πρώτων παρατηρήσεων
- Η πρώτη παρατήρηση
- Το σταθερό επίπεδο από το μοντέλο της απλής γραμμικής παλινδρόμησης

Η επιλογή του αρχικού επιπέδου για την εφαρμογή του μοντέλου, αφήνεται στον ερευνητή και εξαρτάται από τα ποιοτικά χαρακτηριστικά της εκάστοτε χρονοσειράς. Είναι λογικό πως αν το αρχικό επίπεδο δεν ανταποκρίνεται στα δεδομένα, τότε θα έχει αρνητική επίδραση στις τιμές των δεδομένων. Επίσης, το αρχικό επίπεδο επηρεάζει ακολούθως την επιλογή του συντελεστή εξομάλυνσης.

Συντελεστής Εξομάλυνσης

Ο βέλτιστος συντελεστής εξομάλυνσης καθορίζεται από δύο παράγοντες. Ο πρώτος παράγοντας είναι το ποσοστό θορύβου στην χρονοσειρά. Αν υπάρχει, δηλαδή, μεγάλο ποσοστό θορύβου, τότε θα πρέπει να επιλεγεί μικρή τιμή για τον συντελεστή εξομάλυνσης για να αποφύγουμε την

υπερβολική αντίδραση στο θόρυβο. Ο δεύτερος παράγοντας είναι η σταθερότητα του μέσου όρου της χρονοσειράς, Όσο πιο πολύ μεταβάλλεται ο μέσος όρος, τόσο μεγαλύτερη πρέπει να είναι η τιμή του συντελεστή εξομάλυνσης, ώστε οι προβλέψεις να παρακολουθούν τις μεταβολές που παρουσιάζουν τα δεδομένα.

Οι τιμές που λαμβάνει ο συντελεστής εξομάλυνσης βρίσκονται στο διάστημα $[0, 1]$ και γενικά το κριτήριο προσδιορισμού είναι η ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος (MSE). Βέβαια η τιμή της παραμέτρου μπορεί να αλλάξει δεδομένου πως μπορεί να επιλεγθούν άλλα κριτήρια για την επιλογή του συντελεστή, όπως είναι η ελαχιστοποίηση κάποιου άλλου σφάλματος. Στην εφαρμογή χρησιμοποιήθηκε για την εύρεση τόσο του αρχικού επιπέδου, όσο και για την εύρεση του συντελεστή εξομάλυνσης, το κριτήριο της ελαχιστοποίησης του αθροίσματος του μέσου τετραγωνικού σφάλματος (SSE), του οποίου ο τύπος είναι:

$$SSE = \sum_{i=1}^n (Y_i - F_i)^2 = \sum_{i=1}^n e_i^2$$

Η εύρεση της σταθεράς εξομάλυνσης βρίσκεται με αλγοριθμικό τρόπο μέσω της γραμμικής αναζήτησης η οποία ελαχιστοποιεί το σφάλμα, είτε με κάποια άλλη βελτιστοποίηση. Το μειονέκτημα της γραμμικής αναζήτησης είναι ο χρόνος υπολογισμού ειδικά αν υπάρχουν διαθέσιμα πολλά ιστορικά δεδομένα.

Είναι σημαντικό να εξετάσουμε το τι ακριβώς συμβαίνει με τις ακραίες τιμές του συντελεστή εξομάλυνσης. Όταν η τιμή του συντελεστή είναι κοντά στη μονάδα, τότε το σφάλμα της τελευταίας περιόδου διαδραματίζει σημαντικό ρόλο στον υπολογισμό της επόμενης

πρόβλεψης, ενώ στην αντίθετη περίπτωση η νέα πρόβλεψη θα μεταβληθεί σε πολύ μικρό βαθμό. Στην ακραία περίπτωση όπου $a=1$, το μοντέλο ταυτίζεται με την απλοϊκή μέθοδο (Naive), ενώ όταν $a=0$ τότε κάθε πρόβλεψης είναι ίση με το αρχικό επίπεδο.

4.3.3.2 Το Μοντέλο Γραμμικής Τάσης (Holt Exponential Smoothing)

Το μοντέλο εξομάλυνσης για γραμμική τάση είναι στην ουσία μία επέκταση του μοντέλου της απλής εκθετικής εξομάλυνσης η οποία μπορεί επιπρόσθετα να διαχειριστεί την συνιστώσα της τάσης που συχνά συναντάμε στα επιχειρησιακά δεδομένα. Το μοντέλο αυτό προτάθηκε το 1957 από τον Holt και για το λόγο αυτό η μέθοδος διατηρεί μέχρι σήμερα το όνομά του. Η μέθοδος περιγράφεται μαθηματικά από τις εξής εξισώσεις:

$$et = Yt - Ft$$

$$St = St-1 + Tt-1 + a \cdot et$$

$$Tt = Tt-1 + \alpha \cdot \beta \cdot et$$

$$Ft+m = St+m \cdot Tt$$

Όπου et είναι το σφάλμα της πρόβλεψης το οποίο είναι η απόκλιση της πραγματικής τιμής από την πρόβλεψη για την ίδια χρονική περίοδο t . Το St , είναι το επίπεδο της χρονοσειράς και Tt η τάση. Η $Ft+m$ είναι η πρόβλεψη που πραγματοποιείται στο τέλος της περιόδου t και αναφέρεται σε m περιόδους μπροστά, t η χρονική περίοδος και m ο χρονικός ορίζοντας της πρόβλεψης. Η παράμετρος a αποτελεί τον συντελεστή εξομάλυνσης για το επίπεδο, ενώ η παράμετρος β είναι ο

συντελεστής εξομάλυνσης της τάσης, όπου και οι δύο συντελεστές παίρνουν τιμές στο διάστημα $[0, 1]$.

Αρχικό Επίπεδο και Αρχική Τάση

Όπως είναι εμφανές από την αναδρομικότητα του μοντέλου, είναι αναγκαίο να οριστεί ένα αρχικό επίπεδο (S_0) και μία αρχική τάση (T_0), ώστε να μπορέσει να εφαρμοστεί το μοντέλο πρόβλεψης αρχικά και κατόπιν, των ζητούμενων σημειακών προβλέψεων για δεδομένο χρονικό ορίζοντα. Το αρχικό επίπεδο (S_0) υπολογίζεται όπως και στην απλή εκθετική εξομάλυνση. Σαν αρχική τάση συνήθως χρησιμοποιείται:

- Η διαφορά της δεύτερης και πρώτης παρατήρησης ($Y_2 - Y_1$)
- Η διαφορά της νιοστής και πρώτης παρατήρησης διαιρεμένης με $n-1$ (π.χ. $(Y_4 - Y_1)/3$)
- Η σταθερά της κλίσης από το μοντέλο της απλής γραμμικής παλινδρόμησης

Και εδώ θα πρέπει να τονιστεί πως απαιτείται προσεκτική επιλογή τόσο του αρχικού επιπέδου όσο και της αρχικής τάσης, αφού όπως είναι λογικό οι τιμές αυτές επηρεάζουν το μοντέλο της πρόβλεψης και κατ' επέκταση τις σημειακές προβλέψεις. Η τελική επιλογή γίνεται πάντα σύμφωνα με τα ποιοτικά χαρακτηριστικά των δεδομένων της χρονοσειράς που θέλουμε να μελετήσουμε.

Συντελεστές Εξομάλυνσης

Όσον αφορά την εύρεση των τιμών των συντελεστών εξομάλυνσης, όπως στο μοντέλο της απλής εκθετικής εξομάλυνσης, οι τιμές που παίρνουν είναι στο διάστημα $[0, 1]$ και ποικίλουν ανάλογα με το κριτήριο επιλογής

που θα χρησιμοποιηθεί. Η πιο δημοφιλής διαδικασία που χρησιμοποιείται είναι η γραμμική αναζήτηση των συντελεστών που ελαχιστοποιούν το μέσο τετραγωνικό σφάλμα (*MSE*). Συνήθως η βέλτιστη τιμή του συντελεστή α για το επίπεδο είναι μεγαλύτερη από τη τιμή του συντελεστή β για την τάση. Ο λόγος είναι ότι η τιμή της τάσης είναι συνήθως πολύ μικρότερη από την τιμή του επιπέδου για κάθε περίοδο., καθώς αν προστεθεί το ίδιο ποσοστό σφάλματος και στη τάση και στο επίπεδο τότε οι προβλέψεις που προκύπτουν μπορεί να είναι ασταθείς. Για την ανάπτυξη της εφαρμογής σαν κριτήριο υπολογισμού των συντελεστών έχει χρησιμοποιηθεί το κριτήριο ελαχιστοποίησης του αθροίσματος των τετραγωνικών σφαλμάτων (*SSE*)

4.3.3.3 Το Μοντέλο Μη Γραμμικής Τάσης (Damped Trend Method)

Το μοντέλο γραμμικής τάσης που περιεγράφηκε παραπάνω, έχει μία τάση για θετική προκατάληψη ιδιαίτερα στις περιπτώσεις μεσοπρόθεσμων και μακροπρόθεσμων προβλέψεων. Επίσης το μοντέλο μπορεί να μεταβληθεί κατάλληλα ώστε να προσαρμόζεται και σε μη γραμμικές τάσεις. Αυτό μπορεί να επιτευχθεί με την προσθήκη μίας παραμέτρου η οποία ελέγχει το ρυθμό αύξησης των τιμών των προβλέψεων. Αυτή η παράμετρος ονομάζεται παράμετρος διόρθωσης τάσης (*trend-modification parameter*), συμβολίζεται με φ και περιεγράφηκε από τους Gardner και McKenzie το 1985. Οι εξισώσεις που περιγράφουν μαθηματικά το μοντέλο μη γραμμικής τάσης είναι:

$$\begin{aligned}
e_t &= Y_t - F_t \\
S_t &= S_{t-1} + \varphi \cdot T_{t-1} + a \cdot e_t \\
T_t &= \varphi \cdot T_{t-1} + \alpha \cdot \beta \cdot e_t \\
F_{t+m} &= S_t + \sum_{i=1}^m \varphi^i \cdot T_t
\end{aligned}$$

Όπως βλέπουμε οι παραπάνω εξισώσεις είναι παρόμοιες με αυτές του γραμμικού μοντέλου εκτός από την τελευταία, όπου αντί να υπολογίζεται μία γραμμική αύξηση της τάσης μέσω του συντελεστή m , γίνεται ένας μη γραμμικός υπολογισμός αυτής, γεγονός που οφείλεται στην παράμετρο φ . Η συγκεκριμένη παράμετρος δεν έχει κάποιο άνω όριο και μπορεί να λάβει οποιαδήποτε τιμή άνω του μηδενός. Βέβαια είναι πολύ σημαντική η επιβολή άνω και κάτω ορίων ανάλογα με την εκάστοτε περίπτωση. Το μοντέλο της μη γραμμικής τάσης, ανάλογα με την τιμή που παίρνει η παράμετρος εξομάλυνσης, μπορεί να πάρει τις μορφές:

- Αν $\varphi=0$, τότε προκύπτει το μοντέλο της απλής εκθετικής εξομάλυνσης, αφού η τάση δεν έχει πλέον καμία επίδραση στον καθορισμό των στατιστικών σημειακών προβλέψεων.
- Αν $0<\varphi<1$, τότε προκύπτει το μοντέλο της φθίνουσας τάσης, το οποίο αναλύουμε σε αυτή την παράγραφο και είναι κατάλληλο για την παραγωγή μεσοπρόθεσμων και μακροπρόθεσμων προβλέψεων.
- Αν $\varphi=1$, τότε προκύπτει το μοντέλο της γραμμικής τάσης, αφού στην εξίσωση υπολογισμού της πρόβλεψης, τη θέση του αθροίσματος παίρνει το γινόμενο της μεταβλητής χρονικού ορίζοντα m και της προηγούμενης τάσης T_t .
- Αν $\varphi>1$, τότε προκύπτει το μοντέλο της εκθετικής τάσης, το οποίο είναι κατάλληλο στις περιπτώσεις όπου το ζητούμενο είναι η πρόβλεψη

ζήτησης στην αρχή του κύκλου ζωής ενός προϊόντος ή μιας υπηρεσίας. Αυτό συμβαίνει, διότι το συγκεκριμένο μοντέλο χαρακτηρίζεται από μεγάλη θετική προκατάληψη.

Γενικά όσον αφορά την εύρεση των συντελεστών εξομάλυνσης και της παραμέτρου διόρθωσης τάσης, το συνηθέστερο κριτήριο που χρησιμοποιείται είναι η ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος. Είναι επόμενο πως όσο αυξάνεται το πλήθος των παραμέτρων, τόσο αυξάνεται και η υπολογιστική πολυπλοκότητα του προβλήματος και συνεπώς και του χρόνου υπολογισμού των παραμέτρων. Αναφορικά με την επιλογή του αρχικού επιπέδου και της αρχικής τάσης ισχύουν ακριβώς τα ίδια που περιεγράφηκαν και στην περίπτωση του μοντέλου γραμμικής τάσης.

Όμως σε περίπτωση λανθασμένης βελτιστοποίησης του συντελεστή φ , υπάρχει το σενάριο της παραγωγής προκατειλημμένων προβλέψεων. Για αυτό το λόγο, το εύρος τιμών του συντελεστή φ το περιορίζουμε στο διάστημα $[0, 1]$. Μέσα από εμπειρικές έρευνες έχει αποδειχθεί πως η χρήση του μοντέλου μη γραμμικής τάσης δίνει μεγαλύτερη ακρίβεια προβλέψεων, από το μοντέλο γραμμικής τάσης. Σε γενικές γραμμές, το μοντέλο μη γραμμικής τάσης δίνει καλύτερα αποτελέσματα σε περιπτώσεις όπου είναι αδύνατη η επιλογή κάποιου μοντέλου για περιγραφή της χρονοσειράς ή σε περιπτώσεις όπου απαιτείται ένα πλήρως αυτοματοποιημένο σύστημα για την παραγωγή προβλέψεων μεγάλου πλήθους χρονοσειρών. Επίσης, όσο μεγαλύτερος είναι ο ορίζοντας πρόβλεψης τόσο πλεονεκτεί το μοντέλο μη γραμμικής τάσης.

4.3.3.4 Error-Trend-Seasonal/ Exponential Smoothing method – ETS

Είναι ένα state space model, το οποίο δημιουργήθηκε από τον Hyndman και τους συνεργάτες του (2008). Το State Space Model (SSM) αναφέρεται σε μια κατηγορία πιθανολογικών γραφικών μοντέλων, που περιγράφει την πιθανολογική εξάρτηση μεταξύ της μεταβλητής λανθάνουσας κατάστασης και την παρατηρούμενη μέτρηση. Η μέτρηση μπορεί να είναι είτε συνεχής ή διακριτή. Ένα στατιστικό μοντέλο είναι μια στοχαστική (ή τυχαία) διαδικασία δημιουργίας δεδομένων, όπου μπορεί να παραχθεί μία ολόκληρη κατανομή προβλέψεων. Κάθε μοντέλο αποτελείται από μια εξίσωση που περιγράφει τα δεδομένα και μερικές εξισώσεις μετάβασης που δείχνουν πως τα βασικά στοιχεία της χρονοσειράς (επίπεδο, τάση, εποχιακότητα) αλλάζουν με την πάροδο του χρόνου. Ως εκ τούτου, αυτά αναφέρονται ως «state space models».

Για κάθε μέθοδο υπάρχουν δύο μοντέλα, ένα με προσθετικά σφάλματα (Additive errors) και ένα με πολλαπλασιαστικά σφάλματα (Multiplicative errors). Οι σημειακές προβλέψεις που παράγονται από τα μοντέλα είναι ίδια αν είναι ίδιες οι τιμές των παραμέτρων εξομάλυνσης. Θα μπορούν, ωστόσο, να δημιουργήσουν διαφορετικά χρονικά διαστήματα πρόβλεψης. Ουσιαστικά η μέθοδος ETS διαθέτει τρία βασικά στοιχεία για την εξαγωγή προβλέψεων, το είδος των σφαλμάτων, το είδος της τάσης και το είδος της εποχικότητας και επιλέγονται τα είδη αυτά με αυτόν τον τρόπο *ETS (Error,Trend,Seasonal)*. Το στοιχείο των σφαλμάτων και το στοιχείο της εποχικότητας μπορεί να είναι είτε προσθετικά (Additive-A), είτε πολλαπλασιαστικά (Multiplicative-M). Επίσης το στοιχείο της εποχικότητας μπορεί να μην λαμβάνεται υπόψη

(None-N). Αν ληφθεί υπόψη και το στοιχείο της τάσης, τότε μπορεί να είναι ένα από τα παρακάτω είδη:

- Δεν λαμβάνεται υπόψη το στοιχείο της τάσης-None (N)
- Προσθετική-Additive (A)
- Προσθετική φθίνουσα τάση-Additive damped (A_d)
- Πολλαπλασιαστική τάση-Multiplicative (M)
- Πολλαπλασιαστική φθίνουσα τάση-Multiplicative Damped (M_d)

Επομένως οι περιπτώσεις ανάλογα με αυτές τις συνιστώσες χωρίζονται αρχικά με βάση το είδος των σφαλμάτων, δηλαδή σε προσθετικά ή πολλαπλασιαστικά. Κατόπιν ο διαχωρισμός γίνεται με βάση τις περιπτώσεις των άλλων δύο συνιστωσών, της τάσης και της εποχικότητας, όπως φαίνεται στον παρακάτω πίνακα:

Trend	Seasonal		
	N (None)	A (Additive)	M (Multiplicative)
N (None)	(N, N)	(N, A)	(N, M)
A (Additive)	(A, N)	(A, A)	(A, M)
A_d (Additive Damped)	(A_d , N)	(A_d , A)	(A_d , M)
M (Multiplicative)	(M, N)	(M, A)	(M, M)
M_d (Multiplicative Damped)	(M_d , N)	(M_d , A)	(M_d , M)

Άρα υπάρχουν διαθέσιμα 30 state models, 15 με προσθετικά σφάλματα και 15 με πολλαπλασιαστικά.

Απλή Εκθετική Εξομάλυνση με Προσθετικά Σφάλματα-ETS(A, N, N)

Όπως έχουμε αναφέρει και παραπάνω για την απλή γραμμική παλινδρόμηση ισχύουν οι τύποι:

$$e_t = Y_t - F_t$$

$$S_t = S_{t-1} + a \cdot e_t$$

$$F_{t+1} = S_t$$

Όπου επιλύοντας τις παραπάνω εξισώσεις η εξίσωση της πραγματικής τιμής γράφεται ως εξής:

$$Y_t = S_{t-1} + e_t$$

Για να γίνει αυτό σε ένα state model, το μόνο που χρειάζεται είναι να καθοριστεί η κατανομή πιθανοτήτων για e_t . Στο μοντέλο με προσθετικά σφάλματα, υποθέτουμε ότι τα σφάλματα πρόβλεψης ενός βήματος (one-step forecast errors) e_t διανέμονται κανονικά με μέση τιμή 0 και διακύμανση σ^2 . Μία συντομογραφία για αυτό είναι το $e_t = e_t \sim NID(0, \sigma^2)$, όπου NID σημαίνει πως είναι κατανεμημένα με κανονικό και

ανεξάρτητο τρόπο (Normally and Independently Distributed). Επομένως τώρα οι εξισώσεις γράφονται:

$$Y_t = S_{t-1} + e_t$$

$$S_t = S_{t-1} + a \cdot e_t$$

Η πρώτη εξίσωση αναφέρεται ως εξίσωση μέτρησης και η δεύτερη ως εξίσωση κατάστασης ή μετάβασης. Αυτές οι δύο εξισώσεις, μαζί με την στατιστική κατανομή των σφαλμάτων, σχηματίζουν ένα πλήρως καθορισμένο στατιστικό μοντέλο. Συγκεκριμένα, αποτελούν ένα state model, οι οποίες είναι υποκείμενες στην απλή εκθετική εξομάλυνση.

Η εξίσωση μέτρησης δείχνει τη σχέση μεταξύ των παρατηρήσεων και των μη παρατηρούμενων καταστάσεων. Σε αυτή την περίπτωση η

παρατήρηση Yt είναι μια γραμμική συνάρτηση του επιπέδου St και το τυχαίο σφάλμα et , το απρόβλεπτο μέρος του Yt . Για τα άλλα μοντέλα, αυτή η σχέση μπορεί να είναι μη γραμμική.

Η εξίσωση της μετάβασης δείχνει την εξέλιξη της κατάστασης μέσα στο χρόνο. Η επίδραση της παραμέτρου εξομάλυνσης α είναι η ίδια όπως και για τις μεθόδους που συζητήθηκαν προηγουμένως. Για παράδειγμα η παράμετρος α ρυθμίζει τον βαθμό της αλλαγής στα διαδοχικά επίπεδα. Όσο υψηλότερη είναι η τιμή του α , τόσο πιο γρήγορες θα είναι οι μεταβολές του επιπέδου.

4.3.4 Η Απλή Γραμμική Παλινδρόμηση

Η παλινδρόμηση εμπεριέχει διαδικασίες και τεχνικές μοντελοποίησης και ανάλυσης διαφόρων μεταβλητών, όπου το ζητούμενο είναι η εύρεση συσχετίσεων μεταξύ μίας εξαρτημένης και μίας ή και περισσότερων ανεξάρτητων μεταβλητών. Ουσιαστικά με τη μέθοδο αυτή μελετάμε τον τρόπο συμπεριφοράς της εξαρτημένης μεταβλητής, όταν μεταβληθεί η τιμή μίας ανεξάρτητης μεταβλητής όταν οι υπόλοιπες ανεξάρτητες μεταβλητές παραμένουν σταθερές.

Επιπρόσθετα, με την χρήση της ανάλυσης παλινδρόμησης μπορούμε να κάνουμε μία εκτίμηση της τιμής της εξαρτημένης μεταβλητής, δεδομένου των τιμών των ανεξάρτητων μεταβλητών. Πρόκειται ουσιαστικά για την μέση τιμή της εξαρτημένης μεταβλητής, διατηρώντας σταθερές τις ανεξάρτητες μεταβλητές. Η πρόβλεψη της τιμής της εξαρτημένης μεταβλητής είναι μία μαθηματική συνάρτηση των ανεξάρτητων μεταβλητών η οποία ονομάζεται εξίσωση της παλινδρόμησης.

Παρόλο που η παλινδρόμηση χρησιμοποιείται σε θέματα πρόβλεψης, ο κυριότερος λόγος χρησιμοποίησης είναι η ανάλυση και η κατανόηση των σχέσεων που μπορεί να έχουν οι ανεξάρτητες και οι εξαρτημένες μεταβλητές. Πιο συγκεκριμένα μέσα από τη παλινδρόμηση μπορούμε να συμπεράνουμε ποιες από τις εξεταζόμενες μεταβλητές συσχετίζονται, αλλά και να βρούμε τη μορφή αυτής της συσχέτισης. Προκειμένου να επιτευχθεί η ανάλυση της παλινδρόμησης, έχουν αναπτυχθεί ποικίλες τεχνικές. Μία από τις πιο διαδεδομένες είναι και η απλή γραμμική παλινδρόμηση.

Η μέθοδος της απλής γραμμικής παλινδρόμησης βασίζεται στην υπόθεση ύπαρξης σχέσης ανάμεσα στη μεταβλητή πρόβλεψης (εξαρτημένη μεταβλητή) και σε μία άλλη μεταβλητή (ανεξάρτητη μεταβλητή). Εκτός από την παραπάνω υπόθεση της ύπαρξης της σχέσης, υποθέτουμε επιπλέον ότι η σχέση αυτή είναι και γραμμική. Σκοπός της απλής γραμμικής παλινδρόμησης είναι η έκφραση της σχέσης της μεταβλητής Y και μίας ανεξάρτητης μεταβλητής X με την εξίσωση της ευθείας γραμμής:

$$Y_i = a + b \cdot X_i$$

Η οποία χρησιμοποιείται στην πράξη. Όπου το a είναι η τεταγμένη του σημείου τομής της ευθείας με τον άξονα των εξαρτημένων μεταβλητών (για $b=0$) και b είναι η κλίση της ευθείας, δηλαδή ο ρυθμός μεταβολής του Y ανά μοναδιαία αύξηση του X .

Στόχος της απλής γραμμικής παλινδρόμησης είναι η εκτίμηση των παραμέτρων a και b έτσι ώστε η παραπάνω ευθεία να αποτελεί τη βέλτιστη, δηλαδή να εφαρμόζεται όσο γίνεται καλύτερα στα δεδομένα.

Το σφάλμα θεωρείται η κατακόρυφη απόκλιση της παρατήρησης από την ευθεία και δίδεται ως εξής:

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

Όπου η τιμή \hat{Y}_i αντιπροσωπεύει την εκτιμώμενη τιμή από την ευθεία της παλινδρόμησης και η τιμή Y_i αντιστοιχεί στην πραγματική παρατήρηση. Σαν βέλτιστη ευθεία παλινδρόμησης, επιλέγεται αυτή για την οποία το άθροισμα των τετραγώνων των σφαλμάτων γίνεται ελάχιστο. Κατά συνέπεια με τη χρήση της μεθόδου αυτής που είναι γνωστή και σαν μέθοδος των ελαχίστων τετραγώνων οι τιμές των συντελεστών a και b υπολογίζονται από τις ακόλουθες σχέσεις:

$$b = \frac{\frac{\sum_{i=1}^n X_i \cdot Y_i}{n} - \bar{X} \cdot \bar{Y}}{\frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2}$$

Η ισοδύναμα:

$$b = \frac{\sum_{i=1}^n [(X_i - \bar{X}) \cdot (Y_i - \bar{Y})]}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

$$a = \bar{Y} - b \cdot \bar{X}$$

Όπου \bar{X} και \bar{Y} οι μέσες τιμές των διανυσμάτων X και Y και n ο αριθμός των δεδομένων με βάση των οποίων υπολογίζεται η ευθεία παλινδρόμησης.

4.4 Οι Δείκτες Αξιολόγησης Προβλέψεων

Πέρα από την γραφική απεικόνιση των δεδομένων, είναι χρήσιμο να παρέχονται και κάποια αριθμητικά δεδομένα που με βάση αυτά μπορούμε να εξάγουμε συμπεράσματα και να αξιολογήσουμε με βάση τα διαθέσιμα ιστορικά δεδομένα. Επίσης με τον υπολογισμό των βασικών στατιστικών δεικτών μπορούμε να επιλέξουμε τις σωστές μεθοδολογίες

και μεθόδους πρόβλεψης. Παρακάτω αναλύονται οι δείκτες που χρησιμοποιήθηκαν για την πραγματοποίηση του πειράματος, αλλά και μερικοί ακόμα για λόγους βιβλιογραφικής πληρότητας.

4.4.1 Η Βασική Στατιστική Ανάλυση

Η κατηγορία αυτή περιλαμβάνει τους βασικούς στατιστικούς δείκτες οι οποίοι μπορούν να υπολογισθούν άμεσα σε δεδομένη χρονοσειρά Y μεγέθους n παρατηρήσεων.

Μέση τιμή (Average)

Αναφέρεται στον υπολογισμό της μέσης τιμής των παρατηρήσεων. Αυτό βοηθάει στον εντοπισμό του επιπέδου γύρω από το οποίο κινούνται οι πραγματικές τιμές της χρονοσειράς. Η μέση τιμή υπολογίζεται από τον ακόλουθο τύπο:

$$\bar{Y} = \frac{1}{n} \cdot \sum_{i=1}^n Y_i$$

Η Μέγιστη και ελάχιστη τιμή (Maximum και Minimum)

Αναφέρονται στις ακραίες τιμές μίας χρονοσειράς και αποτελούν μια εκτίμηση της διακύμανσης των δεδομένων της, αλλά και της τυχαιότητας που μπορεί να περιέχουν.

Η Τυπική απόκλιση (Standard Deviation)

Η Τυπική Απόκλιση προσπαθεί να δείξει ποια είναι η μέση απόσταση που απέχει ο κάθε αριθμός από το μέσο όρο του αριθμητικού συνόλου που ανήκει. Η τυπική απόκλιση δίνεται από τον ακόλουθο τύπο:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}}$$

$$\sigma_{\text{πληθυσμού}} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n}}$$

Η Διακύμανση (Variance)

Ορίζεται ως το τετράγωνο της τυπικής απόκλισης και είναι ένας ακόμα δείκτης ο οποίος χρησιμεύει για τον εντοπισμό της διασποράς των παρατηρήσεων γύρω από τη μέση τιμή.

Η Συνδιακύμανση

Αν υποθέσουμε πως έχουμε δύο διακριτές τυχαίες μεταβλητές X και Y, ορίζουμε ένα μέτρο σχέσης μεταξύ των δύο περιοχών δεδομένων ώστε να καθορισθεί κατά πόσο οι δύο μεταβλητές μεταβάλλονται ανάλογα (θετική συνδιακύμανση), αντιστρόφως ανάλογα (αρνητική συνδιακύμανση) ή κατά πόσο οι τιμές είναι ασυσχέτιστες (μηδενική συνδιακύμανση). Η συνδιακύμανση υπολογίζεται από τον ακόλουθο τύπο:

$$\text{Cov}(X, Y) = \frac{1}{n} \cdot \sum_{i=1}^n [(X_i - \bar{X}) \cdot (Y_i - \bar{Y})]$$

Ο Συντελεστής γραμμικής συσχέτισης (Linear Correlation Coefficient)

Ονομάζεται και συντελεστής Pearson και εκφράζει την συγκέντρωση των σημείων ενός διαγράμματος διασποράς γύρω από την γραμμή παλινδρόμησης, δίνοντας ένα μέτρο της γραμμικής συσχέτισης μεταξύ

των δύο μεταβλητών. Ο τύπος του συντελεστή γραμμικής συσχέτισης είναι ο ακόλουθος:

$$r_{XY} = \frac{\sum_{i=1}^n [(X_i - \bar{X}) \cdot (Y_i - \bar{Y})]}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \cdot \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

Ο συντελεστής Pearson παίρνει τιμές στο διάστημα [-1, 1] και αναλόγως με την τιμή έχουμε:

- Αν $r = \pm 1$, υπάρχει τέλεια γραμμική συσχέτιση.
- Αν $-0,3 < r < 0,3$, δεν υπάρχει γραμμική συσχέτιση.
- Αν $-0,5 < r \leq -0,3$ ή $0,3 \leq r < 0,5$, υπάρχει ασθενής γραμμική συσχέτιση.
- Αν $-0,7 < r \leq -0,5$ ή $0,5 \leq r < 0,7$, υπάρχει μέση γραμμική συσχέτιση.
- Αν $-0,8 < r \leq -0,7$ ή $0,7 \leq r < 0,8$, υπάρχει ισχυρή γραμμική συσχέτιση.
- Αν $-1 < r \leq -0,8$ ή $0,8 \leq r < 1$, Υπάρχει πολύ ισχυρή γραμμική συσχέτιση.

Ο Συντελεστής αυτοσυσχέτισης (Autocorrelation Coefficient)

Ο συντελεστής αυτός εκφράζει τη συσχέτιση μεταξύ παρατηρήσεων της ίδιας μεταβλητής με χρονική υστέρηση k περιόδους και παίρνει τιμές στο διάστημα [0, 1]. Αν η τιμή του συντελεστή τείνει στο μηδέν, τότε υπάρχει μηδενική συσχέτιση των παρατηρήσεων χρονικής υστέρησης k , ενώ αν είναι κοντά στην μονάδα, έχει μεγάλη συσχέτιση. Ο τύπος υπολογισμού του συντελεστή είναι ο εξής:

$$ACF_k = \frac{\sum_{i=1+k}^n [(Y_i - \bar{Y}) \cdot (Y_{i-k} - \bar{Y})]}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

Ο Συντελεστής μεταβλητότητας (Coefficient of Variation)

Πρόκειται για ένα κανονικοποιημένο μέτρο της διασποράς των παρατηρήσεων ενός δείγματος ή ενός πληθυσμού. Το πλεονέκτημά του, σε σχέση με την τυπική απόκλιση είναι ότι είναι απαλλαγμένο από την επίδραση του επιπέδου των παρατηρήσεων. Ο υπολογισμός μπορεί να γίνει χρησιμοποιώντας τον ακόλουθο τύπο:

$$c_v = \frac{\sigma}{\bar{Y}} \cdot 100 (\%)$$

Η Μέση τιμή διαστήματος μεταξύ ζητήσεων (Intermittent Demand Interval)

Ο δείκτης αυτός εφαρμόζεται σε χρονοσειρές διακοπτόμενης ζήτησης και εκφράζει τη μέση τιμή των αποστάσεων διαδοχικών περιόδων με μη μηδενική τιμή. Αν η τιμή του δείκτη προκύψει ίση με τη μονάδα, τότε αναφερόμαστε σε χρονοσειρά συνεχούς ζήτησης, ενώ όταν πρόκειται για χρονοσειρά διακοπτόμενης ζήτησης παίρνει τιμές μεγαλύτερης της μονάδας. Μεγάλη τιμή του δείκτη συνεπάγεται και μεγάλα, κατά μέσο όρο, μεσοδιαστήματα μεταξύ των μη μηδενικών παρατηρήσεων.

4.4.2 Η Στατιστική Ανάλυση Ακρίβειας Προβλέψεων

Σε αυτή την κατηγορία, σε αντίθεση με τη βασική στατιστική ανάλυση όπου απαιτείται μονάχα η πραγματική χρονοσειρά των παρατηρήσεων, ανήκουν οι στατιστικοί δείκτες που, για να υπολογιστούν, απαιτείται και μία σειρά πρόβλεψης, που προκύπτει από εφαρμογή κάποιας μεθόδου πρόβλεψης επί της πραγματικής χρονοσειράς. Η συγκεκριμένη κατηγορία έχει μεγάλη σημασία στον κλάδο των προβλέψεων καθώς αποτελεί ένα τρόπο αξιολόγησης των μεθόδων.

Οι στατιστικοί δείκτες της κατηγορίας αυτής μπορούν να υπολογιστούν, αφού πρώτα ορίσουμε το σφάλμα, δηλαδή τη διαφορά μεταξύ πραγματικής τιμής και πρόβλεψης για μία περίοδο:

$$e_i = Y_i - F_i$$

Είναι προφανές πως η τιμή του σφάλματος μπορεί να υπολογισθεί μόνο για τις τιμές όπου έχουμε υπολογίσει μοντέλο πρόβλεψης (in-sample error). Αφού όμως παρέλθει το απαραίτητο χρονικό διάστημα και αποκτήσουμε τα δεδομένα που χρειαζόμαστε, μπορούμε να υπολογίσουμε τόσο το σφάλμα του μοντέλου πρόβλεψης, όσο και το πραγματικό σφάλμα (out-of-sample error). Οι παρακάτω στατιστικοί δείκτες είναι ορισμένοι έτσι ώστε να αναφέρονται στο σφάλμα του μοντέλου πρόβλεψης για συγκεκριμένες περιόδους, αλλά με κατάλληλες αλλαγές στις περιόδους, μπορεί να εκφράσουν την ακρίβεια πρόβλεψης και πέρα των συγκεκριμένων περιόδων.

Το Μέσο σφάλμα (Mean Error)

Στην ουσία αποτελεί ένα απλό προσημασμένο μέσο όρο των σφαλμάτων και χρησιμοποιείται ως ένα μέτρο εκτίμησης της συστηματικότητας του σφάλματος. Πιο πρακτικά αν ο συντελεστής είναι κοντά στο μηδέν, τόσο τα σφάλματα είναι τυχαία και όχι συστηματικά. Αν παίρνει αρνητικές τιμές σημαίνει αισιοδοξία στις προβλέψεις, ενώ αν παίρνει θετικές τιμές δηλώνει απαισιοδοξία. Ο δείκτης μπορεί να υπολογιστεί με τον παρακάτω τύπο:

$$ME = \frac{1}{n} \sum_{i=1}^n (Y_i - F_i)$$

Το Μέσο απόλυτο σφάλμα (Mean Absolute Error)

Αποτελεί ένα μέτρο της ακρίβειας της πρόβλεψης έναντι των πραγματικών τιμών. Χρησιμοποιείται ως ένα μέτρο της αστοχίας της πρόβλεψης, χωρίς να δίνεται έμφαση στην κατεύθυνση της πρόβλεψης. Όσο πιο μεγάλη η τιμή του δείκτη, τόσο πιο μικρή η ακρίβεια της μεθόδου πρόβλεψης που εφαρμόστηκε. Υπολογίζεται ως εξής:

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - F_i|$$

Το Μέσο τετραγωνικό σφάλμα (Mean Squared Error)

Είναι ένα μέτρο ακρίβειας της πρόβλεψης, το οποίο δίνει πολύ μεγαλύτερο βάρος στα μεγάλα σφάλματα και μικρότερο βάρος στα μικρά. Χρησιμοποιείται συνήθως για τον υπολογισμό των βέλτιστων παραμέτρων εξομάλυνσης. Δίνεται από τον ακόλουθο τύπο:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - F_i)^2$$

Η Ρίζα μέσου τετραγωνικού σφάλματος (Root Mean Squared Error)

Προκύπτει από τον υπολογισμό της τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος. Ο τύπος είναι ο ακόλουθος:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - F_i)^2}$$

Το Μέσο απόλυτο ποσοστιαίο σφάλμα (Mean Absolute Percentage Error)

Αρκετές φορές είναι πιο χρήσιμος ο υπολογισμός των σφαλμάτων πρόβλεψης σε καθαρά ποσοστιαία μορφή. Αυτό γίνεται διότι διευκολύνει τη σύγκριση μίας μεθόδου πρόβλεψης, όταν εφαρμόζεται σε διαφορετικές χρονοσειρές. Το μέσο απόλυτο σφάλμα είναι εκφρασμένο επί τοις εκατό και δίνεται από τον ακόλουθο τύπο:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - F_i}{Y_i} \right| \cdot 100 (\%)$$

Το Συμμετρικό μέσο απόλυτο ποσοστιαίο σφάλμα (Symmetric Mean Absolute Percentage Error)

Πρόκειται για μία παραλλαγή του MAPE και υπολογίζεται με τον ακόλουθο τύπο:

$$sMAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - F_i}{\frac{Y_i + F_i}{2}} \right| \cdot 100 = \frac{1}{n} \sum_{i=1}^n \left| \frac{2 \cdot (Y_i - F_i)}{Y_i + F_i} \right| \cdot 100 (\%)$$

Το Μέσο απόλυτο κανονικοποιημένο σφάλμα (Mean Absolute Scaled Error)

Ο συντελεστής αυτός δημιουργήθηκε με σκοπό να αντιμετωπιστούν όσο το δυνατόν οι περιπτώσεις απροσδιοριστίας των δεικτών MAPE και sMAPE, αλλά και να δοθεί η ίδια βαρύτητα και για τα μικρά και για τα μεγάλα σφάλματα. Η παρακάτω εξίσωση μας δίνει την τιμή του σφάλματος:

$$MASE = \frac{\frac{1}{n} \sum_{i=1}^n |Y_i - F_i|}{\frac{1}{n-1} \sum_{i=2}^n |Y_i - Y_{i-1}|}$$

4.4.3 Ο Ρυθμός Ανάπτυξης

Ο δείκτης αυτός είναι ένα μέτρο το οποίο εκφράζει την αυξητική ή την φθίνουσα πορεία μίας σειράς δεδομένων μέσα σε ένα συγκεκριμένο χρονικό διάστημα. Η μαθηματική έκφραση του ρυθμού ανάπτυξης έχει ως εξής:

$$\text{Growth Rate} = \frac{\frac{1}{ppy} \cdot \sum_{i=n-ppy+1}^n Y_i - \frac{1}{n-ppy} \cdot \sum_{i=1}^{n-ppy} Y_i}{\frac{1}{n-ppy} \cdot \sum_{i=1}^{n-ppy} Y_i} \cdot 100 (\%)$$

Όπου Y το διάνυσμα των n παρατηρήσεων και ppy το πλήθος των περιόδων στο μήκος ενός έτους.

ΚΕΦΑΛΑΙΟ 5

ΟΙ ΧΡΟΝΟΣΕΙΡΕΣ

5.1 Η Έννοια της Χρονοσειράς

Οι χρονοσειρές αποτελούν ένα σύνολο διαδοχικών παρατηρήσεων της τιμής κάποιου φυσικού ή άλλου μεγέθους ανηγμένες στο χρόνο. Οι χρονοσειρές μπορούν να κατηγοριοποιηθούν με βάση τον τρόπο προσδιορισμού των μελλοντικών τιμών τους, σε ντετερμινιστικές και στοχαστικές χρονοσειρές αντίστοιχα. Όταν οι διαδοχικές παρατηρήσεις δεν είναι ανεξάρτητες μεταξύ τους και οι μελλοντικές τιμές μπορούν να προσδιοριστούν ακριβώς από τις προηγούμενες, τότε αυτές οι χρονοσειρές είναι ντετερμινιστικές. Ωστόσο, κάτι τέτοιο δεν συμβαίνει με τις πραγματικές χρονοσειρές, καθώς το μέλλον καθορίζεται μερικώς από το παρελθόν. Αυτό συμβαίνει διότι στον πραγματικό κόσμο οι χρονοσειρές, δεν περιγράφουν την διαχρονική εξέλιξη κάποιου φυσικού μεγέθους, για το οποίο υπάρχει πλήρης γνώση όλων των παραγόντων που το επηρεάζουν, αλλά αντιθέτως επηρεάζονται από τον λεγόμενο «τυχαίο παράγοντα», ο οποίος αντιπροσωπεύει μία στατιστική μεταβλητή.

Ωστόσο, ο διαχωρισμός των χρονοσειρών σύμφωνα με την παραπάνω κατηγοριοποίηση δεν είναι τόσο εύκολος και προφανής. Παρόλα αυτά, θα πρέπει να γίνει η κατηγοριοποίηση ώστε να γίνεται πιο εύκολα η αναγνώριση και η κατανόηση των παραμέτρων που παίζουν ρόλο στη χρονική εξέλιξη της χρονοσειράς ανεξαρτήτως των δεδομένων της.

5.2 Τα Ποιοτικά Χαρακτηριστικά της Χρονοσειράς

Οι χρονοσειρές δομούνται από κάποια βασικά χαρακτηριστικά, όπου με τις κατάλληλες τεχνικές μπορούν να αναλυθούν σε αυτά. Οι παραδοσιακές μέθοδοι ανάλυσης χρονοσειρών ασχολούνται με την αποσύνθεση της χρονοσειράς σε τέσσερα βασικά της στοιχεία: την τάση, την κυκλικότητα, την εποχιακότητα και τις μη κανονικές διακυμάνσεις. Αυτή η προσέγγιση είναι χρήσιμη όχι μόνο για την εφαρμογή της κλασικής μεθόδου αποσύνθεσης, αλλά και για την περαιτέρω ανάλυση των ποιοτικών χαρακτηριστικών της χρονοσειράς, ανεξαρτήτως της μεθόδου πρόβλεψης που θα χρησιμοποιηθεί.

Αρχικά, η **τάση** ορίζεται σαν μία «μακροπρόθεσμη» μεταβολή του μέσου επιπέδου τιμών της χρονοσειράς. Βέβαια, ενώ αυτός ο όρος είναι κοινά αποδεκτός, υπάρχει ένα πρόβλημα σχετικά με την αποσαφήνιση του ποια μεταβολή θεωρείται μακροπρόθεσμη. Για παράδειγμα βλέπουμε ότι οι περιβαλλοντολογικές και κλιματικές μεταβλητές εμφανίζουν μία κυκλική διακύμανση 50 ετών. Επομένως, αν κάποιος διαθέτει στοιχεία για τα πρώτα 20 έτη θα θεωρούσε την παρατηρούμενη αύξηση του μέσου επιπέδου σαν τάση, ενώ αν εξέταζε δεδομένα μερικών εκατοντάδων ετών θα συμπεραίνε πως πρόκειται, για κυκλική διακύμανση. Άρα, για να μπορέσει κανείς να διαπιστώσει την τάση θα πρέπει να έχει στην διάθεσή του έναν ικανό αριθμό παρατηρήσεων και ταυτόχρονα να εκτιμήσει το κατάλληλο μήκος της περιόδου μέσα στην οποία θα αναζητήσει την ύπαρξη της.

Η **κυκλικότητα** είναι το δεύτερο συστατικό μίας χρονοσειράς και αντιπροσωπεύει μια “κυματοειδή” μεταβολή που οφείλεται σε ειδικές εξωγενείς συνθήκες και εμφανίζεται κατά περιόδους, οι οποίες δεν είναι σταθερές και είναι συνήθως είναι μεγαλύτερες του ενός έτους.

Κυκλικότητα εμφανίζουν οι χρονοσειρές, κατά κύριο λόγο των περισσότερων οικονομικών μεγεθών όπως για παράδειγμα του Ακαθάριστου Εγχώριου Προϊόντος, των τιμών των μετοχών και των τιμών του πετρελαίου. Αυτές οι κυκλικές μεταβολές οφείλονται στις γενικότερες οικονομικές συνθήκες που επικρατούν και που χαρακτηρίζονται από περιόδους με διαδοχικές ανόδους και υφέσεις και είναι γνωστές με την ονομασία επιχειρηματικός κύκλος.

Η **εποχιακότητα** είναι μία περιοδική διακύμανση που έχει σταθερό μήκος και μικρότερο του έτους μήκος. Η εποχιακότητα μπορεί εύκολα να αναγνωριστεί παρατηρώντας την γραφική της αναπαράσταση. Επειδή οι αλλαγές που προκαλεί στα δεδομένα της χρονοσειράς είναι σταθερές με την πάροδο του χρόνου, είναι εύκολο να αντιμετωπίσουμε την επίδρασή της. Αυτό συμβαίνει με την εύρεση των δεικτών εποχιακότητας για τα αντίστοιχα χρονικά διαστήματα, όπου με την διαίρεση αυτών των δεικτών με τα πραγματικά δεδομένα παίρνουμε την λεγόμενη αποεποχικοποιημένη χρονοσειρά, η οποία είναι απαλλαγμένη από την επίδραση της εποχιακότητας. Ένα πολύ χαρακτηριστικό παράδειγμα εποχιακής χρονοσειράς είναι οι πωλήσεις παγωτών που εμφανίζουν μία έντονη άνοδο τους καλοκαιρινούς μήνες και οι πωλήσεις καυσίμων τους χειμερινούς.

Ως **ασυνέχειες** ορίζονται οι απομονωμένες παρατηρήσεις που εμφανίζονται στη γραφική αναπαράσταση κάποιας χρονοσειράς ως απότομες αλλαγές στο πρότυπο συμπεριφοράς της. Τέτοιες αλλαγές έχουν είτε παροδικό είτε μόνιμο χαρακτήρα. Αυτές που έχουν παροδικό χαρακτήρα και που η επίδρασή τους στη χρονοσειρά έχει μικρή χρονική διάρκεια ονομάζονται *outliers* ή *special events*. Πρέπει να τονιστεί εδώ πως η αναγνώρισή τους είναι δύσκολη και απαιτεί και θεωρητική γνώση,

αλλά και κριτική ικανότητα. Ένα *outlier* αποτελεί μία ασυνήθιστη παρατήρηση που οφείλεται σε κάποιο εξαιρετικό ή απρόβλεπτο γεγονός. Για παράδειγμα, μία απεργία μπορεί να προκαλέσει δραματική πτώση στη παραγωγική διαδικασία, το οποίο έχει σαν αποτέλεσμα μία μικρή πτώση στο γράφημα παραγόμενων προϊόντων της εταιρείας.

Στον αντίποδα, οι αλλαγές με μόνιμο χαρακτήρα ονομάζονται *level-shifts* αφού εμφανίζονται ως απότομες αλλαγές στο μέσο επίπεδο των τιμών της χρονοσειράς. Ένα παράδειγμα τέτοιας μόνιμης αλλαγής είναι η πτώση των πωλήσεων μίας εταιρείας εξαιτίας της εισόδου στην αγορά μίας ανταγωνίστριας εταιρείας. Με αυτές τις συνθήκες θα υπάρξει μία απότομη μείωση του μέσου επιπέδου τιμών, το οποίο θα σταθεροποιηθεί σε ένα πιο χαμηλό σημείο.

Μη κανονικές διακυμάνσεις θεωρούνται εκείνες που απομένουν όταν έχουν απομονωθεί όλα τα υπόλοιπα χαρακτηριστικά της χρονοσειράς. Οι διακυμάνσεις αυτές αντιπροσωπεύουν μία τυχαία μεταβλητή που εκφράζει τον τυχαίο παράγοντα μιας στοχαστικής διαδικασίας. Επίσης μπορεί να αντιπροσωπεύουν κάποια ασυνέχεια που συνδέεται με κάποιο εξαιρετικό γεγονός.

Τέλος, θα πρέπει να τονιστεί πως η οπτικοποίηση των δεδομένων είναι πολύ σημαντική για την ευκολότερη εύρεση των ποιοτικών χαρακτηριστικών των χρονοσειρών, όπως φαίνεται στα παρακάτω παραδείγματα.

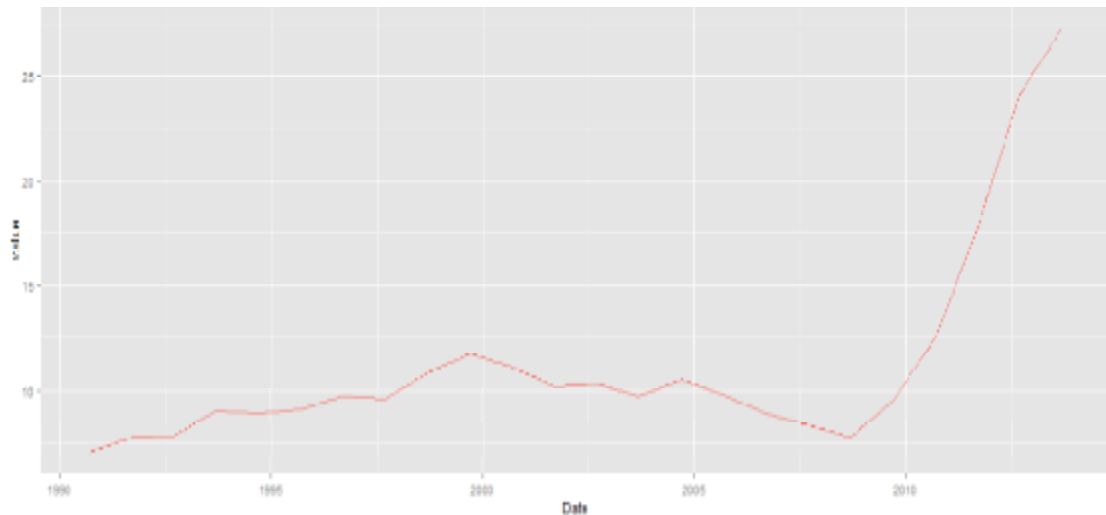
5.3 Η Αναπαράσταση των Χρονοσειρών

Η χρονοσειρά δεν είναι τίποτα παραπάνω από μία σειρά παρελθουσών τιμών για την περιγραφή ενός μεγέθους. Η γραφική αναπαράσταση των δεδομένων αποτελεί ένα

σημαντικό εργαλείο που βοηθάει στην ανάλυση της χρονοσειράς, αλλά και τη διαδικασία της πρόβλεψης. Από την αναπαράσταση των δεδομένων γίνονται πιο εμφανή τα βασικά στοιχεία της χρονοσειράς, κάτι το οποίο βοηθάει τον αναλυτή στην επιλογή της κατάλληλης μεθοδολογίας και εργαλείων με σκοπό να έχει τα βέλτιστα αποτελέσματα με το μικρότερο σφάλμα. Επίσης, με το γράφημα μπορεί να γίνει εύκολα ο εντοπισμός ακραίων εσφαλμένων τιμών, και έτσι ο αναλυτής μπορεί κάνοντας τις κατάλληλες κινήσεις να διορθώσει τις εσφαλμένες τιμές.

Η αναπαράσταση ουσιαστικά έγκειται σε δισδιάστατη γραφική απεικόνιση των πραγματικών τιμών των διαθέσιμων ιστορικών δεδομένων ως προς το χρόνο. Οι κύριοι τύποι γραφημάτων που χρησιμοποιούνται είναι:

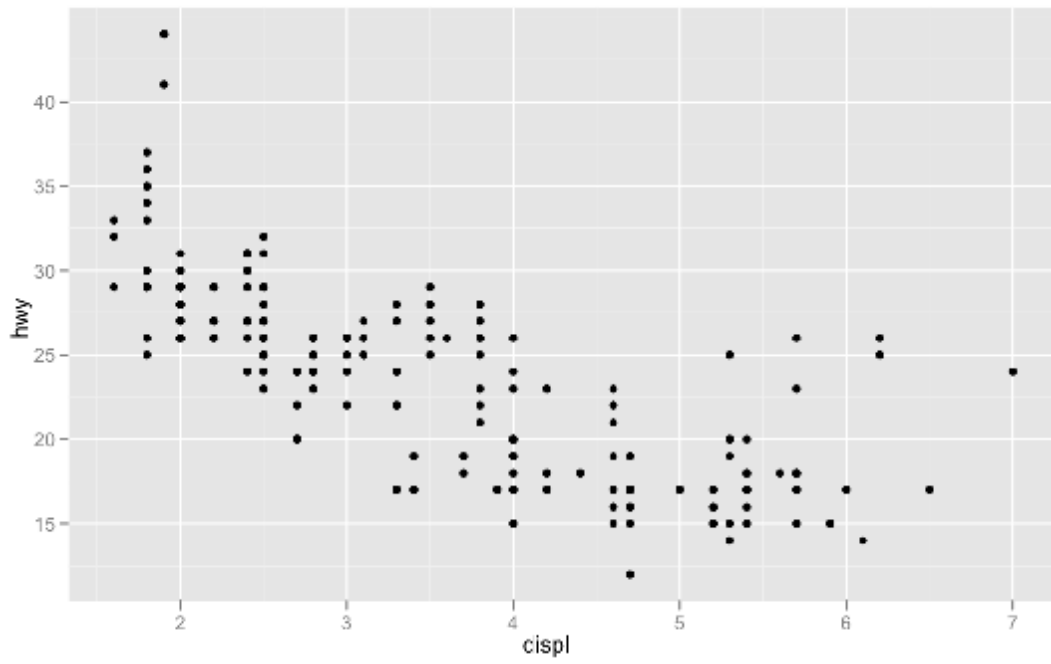
1. Τα Διαγράμματα Χρόνου (time plots): Είναι το πιο χρησιμοποιούμενο διάγραμμα και αναπαριστά τα διαθέσιμα ιστορικά δεδομένα στην πάροδο του χρόνου. Με την χρησιμοποίηση αυτών των διαγραμμάτων είναι πολύ πιο εύκολο να αντιληφθούμε τα βασικά χαρακτηριστικά των χρονοσειρών. Λόγω της ευκολίας της κατανόησης του συγκεκριμένου είδους γραφήματος από την πλειοψηφία του κόσμου, οδήγησε στην χρησιμοποίηση αυτού του τύπου γραφήματος για την απεικόνιση των δεδομένων στην εφαρμογή που κατασκευάστηκε.



Ποσοστό Ανεργίας στην Ελλάδα

2. Τα Εποχιακά Διαγράμματα (seasonal plots): Αυτά τα διαγράμματα χρησιμοποιούνται για χρονοσειρές που παρουσιάζουν έντονη εποχιακότητα.

3. Τα Διαγράμματα Διασποράς (scatter plots): Το διάγραμμα διασποράς είναι το κατάλληλο γράφημα που δημιουργούμε ως πρώτο βήμα προκειμένου να μελετηθεί η σχέση που υπάρχει μεταξύ δύο συνεχών αριθμητικών μεταβλητών, ιδιαίτερα αν αποσκοπούμε στη δημιουργία μοντέλου πρόγνωσης των τιμών της μίας μεταβλητής από την άλλη. Με το διάγραμμα διασποράς και μια έμπειρη στατιστική ματιά ανιχνεύεται η σχέση που ενδεχομένως να υπάρχει μεταξύ των δύο μεταβλητών.



Εικόνα 13 Τα Διαγράμματα Διασποράς (scatter plots)

ΣΥΜΠΕΡΑΣΜΑΤΑ

Οι προβλέψεις αποτελούν ένα αναπόσπαστο κομμάτι της καθημερινότητάς μας. Αυτό αφορά είτε τις επιχειρήσεις, είτε τα κράτη, ή ακόμα και το κάθε άτομο ξεχωριστά. Για αυτό το λόγο το πεδίο των προβλέψεων έχει γίνει αντικείμενο εκτεταμένης έρευνας και ανάπτυξης. Σε αυτό βοηθάει πολύ η ανάπτυξη της τεχνολογίας που παρέχει πλέον στους ερευνητές και στους στατιστικούς πολύτιμα στατιστικά εργαλεία που κάνει την υπολογιστική διαδικασία, σαφώς πιο εύκολη. Αυτό γίνεται πιο εμφανές αν αναλογιστεί κανείς πως τα τελευταία είκοσι χρόνια, παρατηρείται έντονη η τάση του σχεδιασμού και της δημιουργίας διάφορων πληροφοριακών συστημάτων για την εφαρμογή μεθόδων πρόβλεψης και την εξαγωγή προβλέψεων.

Στα οικονομικά, ένα αγαθό θεωρείται ως ένα εμπορεύσιμο προϊόν που παράγεται για να ικανοποιήσει τις επιθυμίες και τις ανάγκες. Σε αυτά μπορεί να περιλαμβάνονται αγαθά και υπηρεσίες. Πρόκειται για μία φυσική ουσία, όπως τα τρόφιμα, τα δημητριακά και τα μέταλλα, τα οποία είναι ανταλλάξιμα με άλλο προϊόν του ίδιου τύπου, και τα οποία οι επενδυτές αγοράζουν ή πωλούν, συνήθως μέσω συμβολαίων μελλοντικής εκπλήρωσης (future contracts). Η τιμή του εμπορεύματος εξαρτάται από την προσφορά και τη ζήτηση.

Με την πτυχιακή μας εργασία θέλαμε να αναδείξουμε την μεγάλη σημασία που έχει η πρόβλεψη πωλήσεων και οι μέθοδοι – τεχνικές στην σημερινή κοινωνία και ειδικότερα στον χώρο των επιχειρήσεων.

Παρουσιάσαμε βασικά στοιχεία για την επιστήμη της οικονομετρίας και ταυτόχρονα δείξαμε την χρησιμότητα της μέσα από πρακτικά

παραδείγματα. Επίπλεον αναφέραμε τα βασικά οικονομετρικά μοντέλα και ποια είναι η χρήση τους.

Σας γενικό συμπέρασμα μπορούμε να αναφέρουμε ότι οι τεχνικές προβλέψεων αποτελούν σημαντικό εργαλείο για κάθε επιχείρηση προσδίδοντας σημαντικό πλεονέκτημα σε σχέση με τις επιχειρήσεις που δεν κάνουν χρήση. Επίσης αποτελούν βάση για την λήψη αποφάσεων στρατηγικού περιεχόμενου, που έχουν να κάνουν με την ανάπτυξη και προοπτική των επιχειρήσεων.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Α.ΜΗΛΙΩΝΗΣ ΣΗΜΕΙΩΣΕΙΣ ΟΙΚΟΝΟΜΕΤΡΙΑΣ Μέρος Α΄

Δημελή, Σ., (2004) *Σύγχρονες μέθοδοι ανάλυσης χρονολογικών σειρών*.

Makridakis, S., Wheelwright, SC., and McGee, V.E., (1983). *Forecasting: Methods and Applications*, 2nd Ed. John Wiley and Sons, New York.

ΚΟΥΝΕΤΑΣ Κ. σημειώσεις μαθήματος τεχνικές προβλέψεων & ελέγχου
θεωρία & εργαστήριο Τεχνολογικό Εκπαιδευτικό Ίδρυμα Πάτρας τμήμα:
επιχειρηματικού σχεδιασμού & πληροφοριακών συστημάτων Πάτρα
Δεκέμβριος 2012