



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ  
ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ  
ΜΟΝΑΔΑ ΣΥΣΤΗΜΑΤΩΝ ΠΡΟΒΛΕΨΗΣ ΚΑΙ ΠΡΟΟΠΤΙΚΗΣ

**Καταγραφή και Παρουσίαση των Παραδοσιακών  
και Σύγχρονων Μεθόδων Πρόβλεψης  
Επισκόπηση και Εκτίμηση Λογισμικών Προβλέψεων**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

των

**ΒΑΣΙΛΕΙΟΥ ΣΟΥΚΟΥ  
ΔΗΜΗΤΡΙΟΥ ΚΑΝΤΖΙΟΥ**

**Επιβλέπων :** Β. Ασημακόπουλος  
Καθηγητής Ε.Μ.Π

Αθήνα, Αύγουστος 2006





# ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ  
ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ  
ΜΟΝΑΔΑ ΣΥΣΤΗΜΑΤΩΝ ΠΡΟΒΛΕΨΗΣ ΚΑΙ ΠΡΟΟΠΤΙΚΗΣ

## **Καταγραφή και Παρουσίαση των Παραδοσιακών και Σύγχρονων Μεθόδων Πρόβλεψης Επισκόπηση και Εκτίμηση Λογισμικών Προβλέψεων**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ  
των  
**ΒΑΣΙΛΕΙΟΥ ΣΟΥΚΟΥ  
ΔΗΜΗΤΡΙΟΥ ΚΑΝΤΖΙΟΥ**

**Επιβλέπων:** Β. Ασημακόπουλος  
Καθηγητής Ε.Μ.Π

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 24<sup>η</sup> Αυγούστου 2006.

.....  
Βασίλειος Ασημακόπουλος  
Καθηγητής Ε.Μ.Π.

.....  
Ιωάννης Ψαρράς  
Καθηγητής Ε.Μ.Π.

.....  
Δημήτριος Ασκούνης  
Επικ. Καθηγητής Ε.Μ.Π.

Αθήνα, Αύγουστος 2006

.....  
**ΒΑΣΙΛΕΙΟΣ Κ. ΣΟΥΚΟΣ**

Διπλωματούχος Ηλεκτρολόγος Μηχανικός  
και Μηχανικός Υπολογιστών Ε.Μ.Π.

.....  
**ΔΗΜΗΤΡΙΟΣ Γ. ΚΑΝΤΖΙΟΣ**

Διπλωματούχος Ηλεκτρολόγος Μηχανικός  
και Μηχανικός Υπολογιστών Ε.Μ.Π.

Με επιφύλαξη παντός δικαιώματος. © 2006 – All rights reserved

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ' ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τους συγγραφείς.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τους συγγραφείς και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

# ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

<b>ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ</b> .....	11
<b>ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ-ΓΡΑΦΗΜΑΤΩΝ</b> .....	14
<b>ΠΡΟΛΟΓΟΣ</b> .....	17
<b>ΠΕΡΙΛΗΨΗ</b> .....	18

## *ΜΕΡΟΣ ΠΡΩΤΟ*

<b>ΚΑΤΑΓΡΑΦΗ ΜΕΘΟΔΩΝ ΠΡΟΒΛΕΨΗΣ</b> .....	32
--	----

### **ΚΕΦΑΛΑΙΟ 1ο**

<b><i>ΤΕΧΝΙΚΕΣ ΧΡΟΝΟΣΕΙΡΩΝ</i></b> .....	33
<b>1.1 Μέθοδοι Εξομάλυνσης</b> .....	33
1.1.1 Εισαγωγή.....	33
1.1.2 Απλοί Κινητοί Μέσοι.....	34
1.1.3 Εκθετική Εξομάλυνση μιας Παραμέτρου .....	37
1.1.4 Αξιολόγηση του Σφάλματος στην Πρόβλεψη.....	41
1.1.5 Γραμμικός Κινητός Μέσος .....	46
1.1.6 Εφαρμογή της Εκθετικής Εξομάλυνσης σε Εποχικά Δεδομένα .....	49
<b>1.2 Μέθοδοι Διαχωρισμού και Ανάλυσης Χρονοσειρών</b> .....	55
1.2.1 Εισαγωγή.....	55
1.2.2 Η Φύση των Οικονομικών Χρονοσειρών .....	56
1.2.3 Αποσύνθεση Χρονοσειρών.....	58
1.2.4 Ανάλυση και Διαχωρισμός της Εποχιακότητας.....	59
1.2.5 Προσαρμογή και Ανάλυση της Τάσης.....	65
1.2.6 Το Μοντέλο Γραμμικής Τάσης.....	68
1.2.7 Εκτίμηση της Τάσης με Μηνιαία ή Τριμηνιαία Δεδομένα .....	71
1.2.8 Απομόνωση της Κυκλικότητας .....	72
1.2.9 Αποτελέσματα Ανάλυσης Χρονοσειρών και Πρόβλεψη .....	76

## ΚΕΦΑΛΑΙΟ 2ο

<b>ΑΙΤΙΟΚΡΑΤΙΚΕΣ ΜΕΘΟΔΟΙ ΠΡΟΒΛΕΨΗΣ</b> .....	78
<b>2.1 Γραμμική Παλινδρόμηση και Συσχέτιση</b> .....	78
2.1.1 Εισαγωγή .....	78
2.1.2 Το Μοντέλο της Απλής Γραμμικής Παλινδρόμησης.....	78
2.1.3 Εκτίμηση του Μοντέλου της Απλής Γραμμικής Παλινδρόμησης.....	82
2.1.4 Τυπικό Σφάλμα Παλινδρόμησης .....	84
2.1.5 Στατιστική Σημαντικότητα της Παλινδρόμησης .....	86
2.1.6 Σημαντικότητα της Σταθεράς Παλινδρόμησης.....	89
2.1.7 Βαθμός Συσχέτισης.....	90
2.1.8 Διαδοχική Συσχέτιση (Αυτοσυσχέτιση) .....	93
2.1.9 Διόρθωση για Αυτοσυσχέτιση .....	95
2.1.10 Το Απλό Γραμμικό Παλινδρομικό Μοντέλο και Πρόβλεψη.....	96
<b>2.2 Μέθοδοι Πολλαπλής Παλινδρόμησης</b> .....	101
2.2.1 Εισαγωγή .....	101
2.2.2 Προϋποθέσεις του Μοντέλου Πολλαπλής Γραμμικής Παλινδρόμησης.....	101
2.2.3 Εκτίμηση Πολλαπλής Γραμμικής Παλινδρόμησης .....	103
2.2.4 Λύσεις Ελαχίστων Τετραγώνων και Έλεγχος Σημαντικότητας .....	105
2.2.5 Πολυσυγγραμμικότητα .....	109
2.2.6 Ψευδομεταβλητές (dummy variables) στα Μοντέλα Παλινδρόμησης .....	110
2.2.7 Η «'Αριστη» Εξίσωση Παλινδρόμησης .....	111
<b>2.3 Οικονομετρικά Μοντέλα</b> .....	113
2.3.1 Εισαγωγή .....	113
2.3.2 Εξισώσεις Οικονομετρικών Μοντέλων .....	114
2.3.3 Οφέλη και Περιορισμοί των Οικονομετρικών Μοντέλων στις Προβλέψεις.....	117
2.3.4 Πρόβλημα Οικονομετρικής Εκτίμησης .....	119
2.3.5 Διαδικασίες Οικονομετρικής Εκτίμησης.....	120
2.3.6 Το Πρόβλημα της Εξειδίκευσης .....	123
2.3.7 Το Πρόβλημα της Ταυτοποίησης .....	123
2.3.8 Μια Εφαρμογή Οικονομετρικών Μοντέλων .....	126
2.3.9 Προσαρμογή Οικονομετρικού Μοντέλου στην Πρόβλεψη .....	129

## ΚΕΦΑΛΑΙΟ 3ο

<b>ΠΡΟΧΩΡΗΜΕΝΑ ΘΕΜΑΤΑ ΑΝΑΛΥΣΗΣ ΧΡΟΝΟΣΕΙΡΩΝ</b> .....	130
<b>3.1 Αυτοσυσχέτιση, Αυτοπαλινδρομικά Μοντέλα και Ανάλυση Χρονοσειρών</b> .....	130
3.1.1 Εισαγωγή .....	130
3.1.2 Χαρακτηριστικά Χρονοσειρών.....	132
3.1.3 Σημείωση για τα Μοντέλα ARIMA.....	136
3.1.4 Η Συνάρτηση Αυτοσυσχέτισης.....	136

3.1.5	Κατανομή Δειγματοληψίας Αυτοσυσχετίσεων .....	139
3.1.6	Είναι τα Δεδομένα Τυχαία;.....	143
3.1.7	Είναι τα Δεδομένα Μη-Σταθερά;.....	146
3.1.8	Εξάλειψη της Τάσης στις Χρονοσειρές.....	147
3.1.9	Είναι τα Δεδομένα Εποχικά;.....	151
3.1.10	Συντελεστές Μερικής Αυτοσυσχέτισης.....	152
3.1.11	Ανάλυση Αυτοσυσχέτισης – Σύνοψη .....	154
<b>3.2</b>	<b>Πρόσθετα Μοντέλα Χρονοσειρών .....</b>	<b>155</b>
3.2.1	Εισαγωγή .....	155
3.2.2	Αυτοπαλινδρομικά Μοντέλα .....	155
3.2.3	Εφαρμογή της Μεθόδου Adaptive Filtering .....	160
3.2.4	Μοντέλα Κινητών Μέσων .....	166
3.2.5	Αυτοπαλινδρομικά Μοντέλα Κινητού Μέσου .....	169
3.2.6	Σύνοψη για τη Μοντελοποίηση Μη-Εποχικών Μοντέλων Χρονοσειρών.....	171
3.2.7	Περισσότερα για την Εκτίμηση με τη Μέθοδο Adaptive Filtering .....	174
3.2.8	Τελεστές Backshift για τα Μοντέλα MA και ARMA.....	177
<b>3.3</b>	<b>Οι Μέθοδοι Box- Jenkins.....</b>	<b>177</b>
3.3.1	Εισαγωγή .....	177
3.3.2	Ταυτοποίηση: Προκαταρκτικά Σχόλια .....	178
3.3.3	Εκτίμηση Μοντέλου και Διαγνωστικός Έλεγχος .....	184
3.3.4	Η Πρόβλεψη με Μοντέλα ARIMA .....	185
3.3.5	Πρόβλεψη Χρονοσειρών με Εποχιακότητα.....	197
<b>3.4</b>	<b>Προβλέψεις με Χρήση Ασαφούς Λογικής (Fuzzy Logic) .....</b>	<b>200</b>
3.4.1	Μοντέλο Πρόβλεψης βασισμένο σε Ασαφή Λογική .....	200
3.4.1.1	Εισαγωγή .....	200
3.4.1.2	Θεωρία Ασαφούς Συνόλου και Ασαφής λογική .....	201
3.4.1.3	Συμπέρασμα του βαθμού ιδιότητας μέλους.....	202
3.4.1.4	Γλωσσικό πλαίσιο εξισώσεων .....	202
3.4.1.5	Η δομή του συμβουλευτικού εργαλείου ασαφούς λογικής (Fuzzy Logic Advisory Tool-FLAT) .....	204
3.4.1.6	Η ενότητα δημιουργίας λειτουργιών ιδιότητας μέλους .....	206
3.4.1.7	Ενότητα "δημιουργίας ασάφειας" (fuzzification module).....	210
3.4.1.8	Ενότητα συμπεράσματος .....	210
3.4.1.9	Ενότητα Αναίρεσης Ασάφειας (defuzzification module) .....	214
3.4.1.10	Συμπεράσματα .....	215
3.4.1.11	Αξιολόγηση της επιλεγμένης μεθόδου .....	216
3.4.1.12	Προσαρμοστικότητα μέσω του συντονισμού του ασαφούς μοντέλου .....	217
3.4.1.13	Η δημιουργία των λειτουργιών ιδιότητας μέλους.....	217
3.4.1.14	Ικανότητα επεξεργασίας των δεδομένων οκτώ διαστάσεων .....	218

3.4.2	Ασαφής Πρόβλεψη Εποχιακότητας.....	219
3.4.2.1	Εισαγωγή .....	219
3.4.2.2	Καταδεικνύοντας την ασαφή τάση μέσω της ασαφούς ανάλυσης παλινδρόμησης.....	219
3.4.2.3	Υπολογισμός των συνόλων των ασαφών δεικτών εποχιακότητας .....	222
3.4.2.4	Κάνοντας προβλέψεις.....	224
3.4.2.5	Εποχιακή ασάφεια και τάσεις.....	226
3.4.2.6	Εφαρμογή σε πρόβλημα πρόβλεψης πωλήσεων.....	227
3.4.2.7	Επίλογος – Συμπέρασμα .....	232
3.4.3	Διαχείριση Προβλημάτων Πρόβλεψης με Χρήση Ασαφών Χρονοσειρών .....	232
3.4.3.1	Εισαγωγή .....	232
3.4.3.2	Έννοιες ασαφών χρονοσειρών.....	233
3.4.3.3	Μια νέα μέθοδος για το χειρισμό προβλημάτων πρόβλεψης βασισμένα σε ασαφείς χρονοσειρές.....	235
3.4.3.4	Ανάλυση απόδοσης.....	243
3.4.3.5	Συμπεράσματα .....	246
<b>3.5</b>	<b>Μέθοδοι Προβλέψεων με Χρήση Νευρωνικών Δικτύων (Neural Networks)</b> .....	<b>247</b>
3.5.1	Εισαγωγή .....	247
3.5.2	Υποκειμενική – Κριτική Επέμβαση στις Προβλέψεις και τα Ιστορικά Δεδομένα Χρονοσειρών (Judgmental Adjustment) με Χρήση Νευρωνικών Δικτύων - Βασικές Έννοιες για τα Ιστορικά Δεδομένα και τη Κριτική Επέμβαση .....	250
3.5.3	Τεχνητά Νευρωνικά Δίκτυα .....	252
3.5.3.1	Γενικά .....	252
3.5.3.2	Βιολογικοί Νευρώνες – Τεχνητοί Νευρώνες.....	253
3.5.4	Μοντέλο Πρόβλεψης.....	255
3.5.5	Πληροφοριακό Σύστημα Κριτικών και Στατιστικών Προβλέψεων με Χρήση Τ.Ν.Δ. σε Στατιστικές Προβλέψεις .....	263
3.5.5.1	Νευρωνικό Δίκτυο για την Κριτική και Στατιστική Πρόβλεψη .....	263
3.5.6	Χρήση των Νευρωνικών Δικτύων για Προβλέψεις σε Συστήματα Εμπορικών Συναλλαγών.....	269
3.5.6.1	Διαχείριση Κινδύνου-Κίνδυνος Χρεοκοπίας.....	272
3.5.6.2	Καταμερισμός Κεφαλαίου .....	276
3.5.6.3	Διαφοροποίηση .....	277
	<b>Βιβλιογραφία 1<sup>ο</sup> Μέρους</b> .....	<b>280</b>



## **ΜΕΡΟΣ ΔΕΥΤΕΡΟ**

<b>ΚΑΤΑΓΡΑΦΗ ΚΑΙ ΕΚΤΙΜΗΣΗ ΛΟΓΙΣΜΙΚΩΝ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΓΙΑ ΤΗΝ ΠΑΡΑΓΩΓΗ ΠΡΟΒΛΕΨΕΩΝ .....</b>	<b>282</b>
---	------------

### **ΚΕΦΑΛΑΙΟ 1ο**

<b>ΚΑΤΑΓΡΑΦΗ ΤΩΝ ΛΟΓΙΣΜΙΚΩΝ ΚΑΙ ΣΥΛΛΟΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ .....</b>	<b>283</b>
<b>1.1 Εισαγωγή .....</b>	<b>283</b>
<b>1.2 Επισκόπηση βιβλιογραφίας-δημοσιεύσεων .....</b>	<b>283</b>
<b>1.3 Καταγραφή λογισμικών προβλέψεων .....</b>	<b>283</b>
<b>1.4 Αρχική αξιολόγηση των λογισμικών .....</b>	<b>285</b>

### **ΚΕΦΑΛΑΙΟ 2ο**

<b>ΤΑΞΙΝΟΜΗΣΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ .....</b>	<b>286</b>
<b>2.1 Εισαγωγή .....</b>	<b>286</b>
<b>2.2 Αναζήτηση και συλλογή ειδικών χαρακτηριστικών .....</b>	<b>286</b>
<b>2.3 Κατάρτιση Συγκριτικών Πινάκων .....</b>	<b>286</b>
2.3.1 Στοιχεία εταιρειών λογισμικού (Company Information) .....	287
2.3.2 Απαιτήσεις σε Hardware/Software (Hardware/Software Requirements) .....	291
2.3.3 Μέθοδοι προβλέψεων των λογισμικών (Selected Forecasting Techniques).....	293
2.3.4 Δυνατότητες των Λογισμικών (Miscellaneous Capabilities).....	302

### **ΚΕΦΑΛΑΙΟ 3ο**

<b>ΑΞΙΟΛΟΓΗΣΗ ΛΟΓΙΣΜΙΚΩΝ ΚΑΙ ΣΥΓΚΡΙΣΗ ΕΠΙΚΡΑΤΕΣΤΕΡΩΝ ΜΕ ΤΟ ΛΟΓΙΣΜΙΚΟ ΡΥΘΜΙΑ .....</b>	<b>308</b>
<b>3.1 Εισαγωγή .....</b>	<b>308</b>
<b>3.2 Ποιοτική ανάλυση των συγκριτικών πινάκων και επιλογή των επικρατέστερων λογισμικών .....</b>	<b>308</b>
<b>3.3 Σύντομη παρουσίαση επικρατέστερων λογισμικών .....</b>	<b>309</b>
3.3.1 Παρουσίαση του Autobox.....	309
3.3.2 Παρουσίαση του Eviews.....	312
3.3.3 Παρουσίαση του ForecastPro .....	317
3.3.4 Παρουσίαση του Statgraphics Centurion.....	319
3.3.5 Παρουσίαση του Minitab .....	322
3.3.6 Παρουσίαση του SPSS.....	325
3.3.7 Παρουσίαση του SAS .....	328
3.3.8 Παρουσίαση του mySAP SCM.....	330
3.3.9 Παρουσίαση του MicrosoftForecaster .....	332

3.3.10	Παρουσίαση του JD Edwards EnterpriseOne Demand Forecasting and Advanced Forecast Modeling (Oracle).....	333
<b>3.4</b>	<b>Παρουσίαση του ΡΥΘΙΑ.....</b>	<b>336</b>
<b>3.5</b>	<b>Σύγκριση επικρατέστερων λογισμικών με το ΡΥΘΙΑ .....</b>	<b>342</b>
<b>3.6</b>	<b>Προοπτική Εξέλιξης-Επέκτασης της Έρευνας .....</b>	<b>349</b>
	<b>Βιβλιογραφία 2<sup>ο</sup> Μέρους .....</b>	<b>350</b>

#### ***ΠΑΡΑΡΤΗΜΑΤΑ***

	ΠΑΡΑΡΤΗΜΑ “Α”-Επιστολή προς τις Εταιρείες Λογισμικών .....	353
	ΠΑΡΑΡΤΗΜΑ “Β”-Στοιχεία που Ζητήθηκαν από τις Εταιρείες Λογισμικών .....	354

# ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

## 1<sup>ο</sup> ΜΕΡΟΣ

Πίνακας	Τίτλος	Σελίδα
1.1	Πρόβλεψη με κινητούς μέσους όρους	35
1.2	Πρόβλεψη με τη μέθοδο της εκθετικής εξομάλυνσης μιας παραμέτρου	39
1.3	Υπολογισμός του MSE για μοντέλο εκθετικής εξομάλυνσης, όπου $\alpha = 0,2$	43
1.4	Αποτελέσματα ανάλυσης ευαισθησίας	45
1.5	Διενέργεια προβλέψεων με τη μέθοδο των κινητών μέσων ( $n=3$ )	46
1.6	Πρόβλεψη με γραμμικούς κινητούς μέσους - χρονοσειράς με τάση χωρίς καμιά τυχαία συνιστώσα	48
1.7	Εφαρμογή της εκθετικής εξομάλυνσης τριπλής παραμέτρου (Μοντέλο Winters)	51
1.8	Εύρεση αρχικών εκτιμήσεων των εποχικών συντελεστών ως εισροές στο μοντέλο Winters	53
1.9	Συνολική απασχόληση στον μη-γεωργικό τομέα (δεδομένα για δύο μήνες του 1980)	60
1.10	Υπολογισμοί σε στάδια 1 έως 4 για την εύρεση εποχικών δεικτών	61
1.11	Υπολογισμοί που αφορούν τα στάδια 5 και 6 προς εύρεση εποχικών δεικτών	63
1.12	Μέσες εβδομαδιαίες αποδοχές εργατών	66
1.13	Σχετικές αυξήσεις εβδομαδιαίων αποδοχών	66
1.14	Υπολογισμοί ελαχίστων τετραγώνων	69
1.15	Απομόνωση της κυκλικής συνιστώσας	74
2.1	Υπολογισμοί για την παλινδρόμηση	83
2.2	Πραγματικές, εκτιμημένες και τιμές καταλοίπων από την παλινδρόμηση του καθαρού εισοδήματος πάνω στο χρόνο της American Hospital Supply, Inc. για την περίοδο 1965 ως 1980	95
2.3	Προκαταρκτικός πίνακας για πολλαπλή παλινδρόμηση	104
2.4	Η ανάλυση διακύμανσης ενυπόθηκων δανείων που έχουν παλινδρομηθεί πάνω σε δύο ανεξάρτητες μεταβλητές	106
2.5	Πραγματικά και προβλεφθέντα αποτελέσματα, σφάλματα, τυπικά σφάλματα και ποσοστιαίο σφάλμα για το οικονομικό μοντέλο	128
3.1	Χρονοσειρές κερδών ανά μετοχή ( $Y_t$ ) της Gates Learjet Corporation	138
3.2	Χρονοσειρά 40 τυχαίων παρατηρήσεων	140
3.3	Κατάλοιπα από την πρόβλεψη της χρονοσειράς δανείων ακίνητης περιουσίας (\$ δις)	144
3.4	Πρώτες διαφορές επί των δεδομένων για τα κέρδη ανά μέρισμα της Chase Manhattan Corporation	149
3.5	Μεταφορές αποθεμάτων χωρίς προσαρμογή για απόσβεση κεφαλαίου, τριμηνιαία αθροίσματα μη-εποχικά προσαρμοσμένα (\$ χιλιάδες)	151
3.6	Καθαρό κέρδος ιδιωτικών εγχώριων επενδύσεων	158

Πίνακας	Τίτλος	Σελίδα
3.7	Στάδια της adapting filtering για μια αυτοπαλίνδρομη διαδικασία	161
3.8	Τυποποιημένες τιμές μετά από διαίρεση των τιμών με τη μεγαλύτερη τιμή	163
3.9	Μέθοδος τυποποίησης με διαίρεση κάθε ομάδας τιμών δια της τετραγωνικής ρίζας του αθροίσματος των τετραγώνων των τιμών	164
3.10	Χρονοσειρά με σταθερό συντελεστή αύξησης για διάστημα έξι περιόδων	166
3.11	Ταυτοποίηση γενικών μοντέλων χρονοσειρών	172
3.12	Ειδικά μοντέλα χρονοσειρών	173
3.13	Δείκτης Βιομηχανικής παραγωγής	186
3.14	Σύγκριση δύο μοντέλων AR(1)	194
3.15	Μηνιαίες Πωλήσεις Γαλακτοκομικού Προϊόντος (σε κουτιά κατά την περίοδο 1992 - 1994)	228
3.16	Ασαφείς παράμετροι $A=(A0, A1, A2);(h=0)$	228
3.17	Συγκεκριμένα σύνολα δεικτών εποχιακότητας και συγκεκριμένοι δείκτες εποχιακότητας	229
3.18	Συγκεκριμένες προβλέψεις των μηνιαίων πωλήσεων για το 1995	230
3.19	Δείκτες Παραγωγού $I_k$ ,	231
3.20	Πραγματικές εγγραφές και διακυμάνσεις των ιστορικών στοιχείων	238
3.21	Ασαφοποιημένες ιστορικές εγγραφές	239
3.22	Προβλεπόμενες διακυμάνσεις με τη βάση παραθύρου $w = 5$	241
3.23	Προβλεπόμενα αποτελέσματα με χρήση της προτεινόμενης ασαφούς μεθόδου χρονοσειρών με βάση παραθύρου $w = 5$	242
3.24	Προβλεπόμενες εγγραφές με διαφορετικές βάσεις παραθύρου	244
3.25	Λάθη πρόβλεψης με διαφορετικές βάσεις παραθύρου	244
3.26	Τα αποτελέσματα πρόβλεψης διαφορετικών μεθόδων πρόβλεψης	245
3.27	Μια σύγκριση των μέσων λαθών πρόβλεψης των διαφορετικών μεθόδων πρόβλεψης	245
3.28	Μια σύγκριση των μέσων λαθών πρόβλεψης της προτεινόμενης μεθόδου και της μεθόδου Song Chissom	246
3.29	Η πρόβλεψη για 1 μήνα μετά έχει μικρότερο σφάλμα από αυτή για 3 μήνες μετά.	268
3.30	Επιλεγμένες περιπτώσεις για πρόβλημα χρεοκοπίας με $R_p=1$ και αρχικό κεφάλαιο $z=5000$ . Οι συναλλαγές παύουν με απώλεια $z$ (χρεοκοπία) ή κέρδος $\alpha-z$ (κεφάλαιο $\alpha$ )	275
3.31	Παραδείγματα για πρόβλημα χρεοκοπίας με λόγο εξόφλησης $R_p=2$ και αρχικό κεφάλαιο $z=5000$ .	276

# **ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ**

## **2<sup>ο</sup> ΜΕΡΟΣ**

<b>Πίνακας</b>	<b>Τίτλος</b>	<b>Σελίδα</b>
1.1	Κατάλογος λογισμικών που εκτελούν προβλέψεις	284
1.2	Κατάλογος επιλεγμένων λογισμικών	285
2.1	Στοιχεία εταιρειών λογισμικού (company information)	287
2.2	Απαιτήσεις σε υλικό/ λογισμικό (hardware/software requirements)	291
2.3	Κινητοί μέσοι όροι-εκθετική εξομάλυνση (moving averages-exponential smoothing)	293
2.4	Μέθοδοι αποσύνθεσης (decomposition methods)	295
2.5	Νευρωνικά δίκτυα-οικονομετρικά μοντέλα (neural networks-econometric models)	297
2.6	Μέθοδοι παλινδρόμησης (regression methods)	299
2.7	Δυνατότητες λογισμικών (miscellaneous capabilities)	302
2.8	Ειδικές λειτουργίες λογισμικών (special features)	304
3.1	Απαιτήσεις σε υλικό/ λογισμικό (hardware/software requirements) των επικρατέστερων λογισμικών	342
3.2	Κινητοί μέσοι όροι-εκθετική εξομάλυνση των επικρατέστερων λογισμικών (moving averages-exponential smoothing)	343
3.3	Μέθοδοι αποσύνθεσης των επικρατέστερων λογισμικών (decomposition methods)	344
3.4	Νευρωνικά δίκτυα-οικονομετρικά μοντέλα των επικρατέστερων λογισμικών (neural networks-econometric models)	345
3.5	Μέθοδοι παλινδρόμησης των επικρατέστερων λογισμικών (regression methods)	346
3.6	Ειδικές λειτουργίες (special features) των επικρατέστερων λογισμικών	347

# ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ

## 1<sup>ο</sup> ΜΕΡΟΣ

Σχήμα	Τίτλος	Σελίδα
1.1	Παρατηρηθείσες τιμές και προβλέψεις με τη μέθοδο της απλής εξομάλυνσης	36
1.2	Παρατηρηθείσες τιμές και προβλέψεις με τη μέθοδο της απλής εκθετικής εξομάλυνσης	40
1.3	Αρχικές Απαιτήσεις Επιδομάτων Ανεργίας. Πολιτειακά Προγράμματα (εκατοντάδες)	56
1.4	Οι εβδομαδιαίες αποδοχές σε γραμμική και λογαριθμική κλίμακα	67
1.5	Γραφική Παράσταση του κόστους θέρμανσης έναντι χρόνου	69
1.6	Παρατηρηθείσες και προβλεφθείσες τιμές	71
2.1	Ενυπόθηκα δάνεια και ύψος επιτοκίου (παρατηρηθέντα και προβλεφθέντα)	84
2.2	Παρατηρηθέντα, προβλεφθέντα και δύο όρια τυπικών σφαλμάτων	86
2.3	(α) Κίνηση κατά μήκος της καμπύλης προσφοράς. (β) Κίνηση κατά μήκος της καμπύλης ζήτησης	124
3.1	Συντελεστές Αυτοσυσχέτισης ARIMA (1,0,0) (αποτελέσματα σύμφωνα με το πρόγραμμα Minitab)	142
3.2	Οι συντελεστές αυτοσυσχέτισης των καταλοίπων από την πρόβλεψη δανείων ακίνητης περιουσίας με το μοντέλο απλής εκθετικής εξομάλυνσης	145
3.3	Συντελεστές αυτοσυσχέτισης των καταλοίπων από πρόβλεψη των δανείων ακίνητης περιουσίας με το μοντέλο Brown για την τάση	145
3.4	Συντελεστές αυτοσυσχέτισης για τη χρονοσειρά των δεύτερων διαφορών	150
3.5	Συντελεστές αυτοσυσχέτισης φορτίων για δεδομένα που αφορούν αποθέματα	152
3.6	Συντελεστές αυτοσυσχέτισης των καταλοίπων από μοντέλο αυτοπαλινδρόμησης	159
3.7	Συντελεστές αυτοσυσχέτισης των καθαρών κερδών ιδιωτικών εγχώριων επενδύσεων μη-οικιακού διαρκούς εξοπλισμού	163
3.8	(α) ACF για μοντέλο MA(1) με $\theta_1 > 0$ , (β) PACF για μοντέλο MA(1) με $\theta_1 < 0$	169
3.9	(α) ACF για το μοντέλο ARMA (1,1) (β) PACF για το μοντέλο ARMA (1,1)	171
3.10	Η διαδικασία Box-Jenkins	178
3.11	(α) Μοντέλο AR(1), (β) Μοντέλο AR(2)	181
3.12	(α) Μοντέλο MA(1), (β) Μοντέλο MA(2)	182
3.13	(α) Μοντέλο ARMA(1,1), (β) Μοντέλο ARMA(1,1)	183

Σχήμα	Τίτλος	Σελίδα
3.14	Συντελεστές αυτοσυσχέτισης για το δείκτη βιομηχανικής παραγωγής χρονοσειράς	187
3.15	Συντελεστές μερικής αυτοσυσχέτισης του δείκτη βιομηχανικής παραγωγής χρονοσειράς	188
3.16	Συντελεστές αυτοσυσχέτισης για χρονοσειρά πρώτων διαφορών	189
3.17	Συντελεστές μερικής αυτοσυσχέτισης για χρονοσειρά πρώτων διαφορών	190
3.18	Συντελεστές αυτοσυσχέτισης για τα κατάλοιπα του εκτιμημένου μοντέλου ARIMA (1,0,0)	191
3.19	Συντελεστές Συσχέτισης για τα κατάλοιπα εκτιμημένου μοντέλου ARIMA(1,0,0)	193
3.20	Η απεικόνιση των γλωσσικών σχέσεων από τις γλωσσικές εξισώσεις	203
3.21	Συμβουλευτικό εργαλείο ασαφούς λογικής (fuzzy logic advisory tool-FLAT)	204
3.22	Η προ-επεξεργασία των on-line δεδομένων	207
3.23	Ασαφής διανομή	208
3.24	Κατανομή της ασαφούς διανομής	209
3.25	Διαδικασία " Fuzzification " μιας μεταβλητής εισαγωγής με τιμή $e$	210
3.26	Η λειτουργία εξαγωγής συμπεράσματος στο FLAT	212
3.27	Σύγκριση της πραγματικής ζήτησης, της ζήτησης που προβλέπεται από τον πελάτη και τον εμπειρογνώμονα μαζί και της ζήτησης που προβλέπεται από το FLAT κατά τη διάρκεια της πρώτης περιόδου.	215
3.28	Σύγκριση της πραγματικής ζήτησης, της ζήτησης που προβλέπεται από τον πελάτη και τον εμπειρογνώμονα μαζί και της ζήτησης που προβλέπεται από το FLAT κατά τη διάρκεια της δεύτερης περιόδου.	216
3.29	Διαδικασία αναφοράς σε μεγάλη κλίμακα.	218
3.30	Χαρακτηριστικό παραδείγμα της ασαφούς τάσης χρονοσειρών που αναλύεται με την ασαφή ανάλυση παλινδρόμησης	221
3.31	Χαρακτηριστικό παραδείγμα της ασαφούς τάσης χρονοσειρών που αναλύεται με την ασαφή ανάλυση παλινδρόμησης	221
3.32	Χαρακτηριστικό παραδείγμα της ασαφούς τάσης χρονοσειρών που αναλύεται με την ασαφή ανάλυση παλινδρόμησης	222
3.33	Χαρακτηριστικό παραδείγμα της ασαφούς τάσης χρονοσειρών που αναλύεται με την ασαφή ανάλυση παλινδρόμησης	222
3.34	Χαρακτηριστικό παραδείγμα της ασαφούς τάσης χρονοσειρών που αναλύεται με την ασαφή ανάλυση παλινδρόμησης	222
3.35	«Σταθερή» εποχιακή διακύμανση	223
3.36	Σχεδίαση των κατ' εκτίμηση στοιχείων	227
3.37	Αποτελέσματα της ανάλυσης ασαφούς παλινδρόμησης των πωλήσεων γαλακτοκομικού προϊόντος κατά τα έτη 1992-1994	229

<b>Σχήμα</b>	<b>Τίτλος</b>	<b>Σελίδα</b>
3.38	Ασαφείς προβλέψεις για τις πωλήσεις του γαλακτοκομικού προϊόντος για το έτος 1995	231
3.39	Modal Trend Analysis	231
3.40	Καμπύλες των πραγματικών και των προβλεπόμενων εγγραφών	242
3.41	Παροδική (transient) και μεταφερόμενη (transferring) επίδραση	251
3.42	Επαναληπτικό Δίκτυο με ένα κρυμμένο επίπεδο	253
3.43	Αντιστοιχία βιολογικού-τεχνητού νευρώνα	254
3.44	Λογιστική συνάρτηση $f(u)=1/(1+\exp(-gu))$	254
3.45	Υπερβολική Εφασπτομένη $f(u)=\tanh(gu)$	254
3.46	Μοντέλο Πρόβλεψης κατά Jae Kyu Lee και Chang Seon Yum	257
3.47	Χρονοσειρά Φιλτραρισμένη από τα Υποκειμενικά Γεγονότα	259
3.48	Τμήμα Βάσης Δεδομένων Υποκειμενικών Σεναρίων	260
3.49	Νευρωνικό Δίκτυο επεξεργασίας Υποκειμενικών Σεναρίων	261
3.50	Είσοδοι και έξοδοι Μοντέλου Νευρωνικού Δικτύου για την πρόβλεψη της Βασικής Τάσης. Τμήμα της Βάσης Δεδομένων Σει Εκπαίδευσης	262
3.51	Νευρωνικό δίκτυο για την εφαρμογή πρόβλεψης της επίδρασης Κριτικών Γεγονότων	264
3.52	Ροή σημάτων σε ένα τμήμα Τ.Ν.Δ. που τρέχει BackPropagation	265
3.53	Γραφική αναπαράσταση της ροής των υπολογισμών	266

## **2<sup>ο</sup> ΜΕΡΟΣ**

<b>Σχήμα</b>	<b>Τίτλος</b>	<b>Σελίδα</b>
2.1	Διαδικασία “Bottom – Up ”	307
2.2	Διαδικασία “Top-Down”	307



### ΠΡΟΛΟΓΟΣ

Η διπλωματική αυτή εργασία εκπονήθηκε στα πλαίσια των ερευνητικών δραστηριοτήτων της Μονάδας Συστημάτων Προβλέψεων και Προοπτικής, που λειτουργεί στο τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Ηλεκτρονικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου, κατά την περίοδο Δεκεμβρίου 2005-Αυγούστου 2006.

Στα πλαίσια αυτής της εργασίας καταβλήθηκε κάθε δυνατή προσπάθεια για την αρτιότερη και ακριβέστερη παρουσίαση του αντικειμένου, υπό την επίβλεψη και καθοδήγηση του καθηγητή κ. Βασίλη Ασημακόπουλου, τον οποίο και θα θέλαμε να ευχαριστήσουμε θερμά για την ευκαιρία που μας έδωσε να ασχοληθούμε με το συγκεκριμένο αντικείμενο. Υπό την επίβλεψη του, η Μονάδα Συστημάτων Προβλέψεων και Προοπτικής αποτελεί ένα πρότυπο περιβάλλον επικοινωνιακής εργασίας, το οποίο συνέβαλε επικοινωνιακά στην ολοκλήρωση της έρευνάς μας.

Θα θέλαμε επίσης να ευχαριστήσουμε και τα υπόλοιπα μέλη της τριμελούς επιτροπής, τον καθηγητή κ. Ιωάννη Ψαρρά και τον επίκουρο καθηγητή κ. Δημήτριο Ασκούνη.

Επιπλέον, θα ήταν παράλειψη να μην αναφερθούμε στη βοήθεια και καθοδήγηση που μας προσέφερε η διδάκτωρ Έλλη Παγουρτζή, η οποία παρά τον φόρτο εργασίας της, μας βοήθησε πρόθυμα όποτε τη χρειαστήκαμε. Η συνεργασία μας διεξήχθη πάντα σε άριστο κλίμα. Τέλος, να ευχαριστήσουμε τα υπόλοιπα μέλη του εργαστηρίου, και ιδιαίτερα το Νίκο Μπουγιούκο, τον Φώτη Πετρόπουλο και τον Βαγγέλη Σκαρογιάννη, για τις πολύτιμες συμβουλές τους κατά την εκπόνηση της εργασίας μας..

Αθήνα, Αύγουστος 2006

### ΠΕΡΙΛΗΨΗ

Ο σκοπός της παρούσης διπλωματικής εργασίας είναι η καταγραφή και παρουσίαση όλων των μεθόδων πρόβλεψης, παραδοσιακών και σύγχρονων, καθώς και η παρουσίαση και εκτίμηση των λογισμικών που εκτελούν προβλέψεις.

Η εργασία, η οποία χωρίζεται σε δύο μέρη, εκπονήθηκε με στόχο αρχικά την σύντομη παρουσίαση των προηγμένων τεχνικών που εφαρμόζονται στη διαδικασία προβλέψεων, τεχνικές που βασίζονται στην ασαφή λογική και τα νευρωνικά δίκτυα. Επιπρόσθετα γίνεται και παρουσίαση των παραδοσιακά χρησιμοποιούμενων τεχνικών. Όλα τα παραπάνω γίνονται πάντα με τη βοήθεια παραδειγμάτων και με τη χρήση ευανάγνωστων πινάκων και σχημάτων. Στη συνέχεια επιχειρήθηκε η αναζήτηση και η καταγραφή όλων των λογισμικών και των εφαρμογών που χρησιμοποιούνται για παραγωγή προβλέψεων. Ο μεγάλος αριθμός λογισμικών που προέκυψε- καταγράφηκαν συνολικά 82 λογισμικά- αξιολογήθηκε ως προς συγκεκριμένα χαρακτηριστικά, με αποτέλεσμα την αρχική επιλογή 42, για τα οποία καταρτίστηκαν αναλυτικοί συγκριτικοί πίνακες βασισμένοι πάνω σε ειδικά επιλεγμένα χαρακτηριστικά. Στο τελευταίο στάδιο της εργασίας, κάνοντας χρήση αυτών των συγκριτικών πινάκων, καταλήξαμε στην τελική επιλογή των 10 επικρατέστερων - κατά την άποψή μας- λογισμικών. Αφού επιχειρείται μια σύντομη παρουσίασή τους, στη συνέχεια τα λογισμικά αυτά συγκρίνονται με το λογισμικό ΡΥΤΗΙΑ (λογισμικό που δημιουργήθηκε από τη Μονάδα Συστημάτων Πρόβλεψης και Προοπτικής του ΕΜΠ της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Ηλεκτρονικών Υπολογιστών).

Επιπλέον, σαν ευρύτερη προοπτική εξέλιξης της έρευνας που επιχειρείται στο 2<sup>ο</sup> μέρος, προτείνεται η περαιτέρω διερεύνηση των δυνατοτήτων των 10 επικρατέστερων λογισμικών, κάνοντας πια, χρήση των λογισμικών επί συγκεκριμένων χρονοσειρών και μελετώντας τα τεχνικά εγχειρίδιά τους (technical manuals και demos), καταλήγοντας έτσι σε αναλυτική κριτική παρουσίασή τους, σύμφωνα με ανάλογες αναφορές σχετικών παλαιότερων δημοσιεύσεων ([4]).

Αναλυτικότερα, για κάθε μέρος της εργασίας, έχουμε τα εξής:

**Στο πρώτο μέρος** της εργασίας γίνεται μια προσπάθεια καταγραφής και αναλυτικής παρουσίασης των παραδοσιακών και σύγχρονων μεθόδων πρόβλεψης. Ταυτόχρονα επιχειρείται η ταξινόμησή τους στις κάτωθι κατηγορίες:

- Μέθοδοι Χρονοσειρών
- Αιτιοκρατικές Μέθοδοι Πρόβλεψης
- Προηγμένα Μοντέλα Ανάλυσης Χρονοσειρών

Στο πρώτο κεφάλαιο της εργασίας αναλύονται τα απλούστερα μοντέλα πρόβλεψης, δηλ. οι μέθοδοι χρονοσειρών. Τα μοντέλα αυτά περιλαμβάνουν τον κινητό μέσο όρο, την εκθετική εξομάλυνση, καθώς και προεκτάσεις του κινητού μέσου όρου και της εκθετικής εξομάλυνσης. Κατόπιν, γίνεται μία εισαγωγή στις μεθόδους διαχωρισμού κλασικών χρονοσειρών. Ειδικότερα:

## ΠΕΡΙΛΗΨΗ

---

➤ Στη παράγραφο 1.1 εισάγονται δύο βασικοί τύποι μεθόδων πρόβλεψης: ο κινητός μέσος όρος και η εξομάλυνση. Ο κινητός μέσος όρος βασίζεται στον μέσο όρο σταθμισμένων παρατηρήσεων του παρελθόντος. Οι μέθοδοι εξομάλυνσης βασίζονται στη διαδικασία καθορισμού του μέσου όρου παρατηρήσεων μιας χρονοσειράς που διαμορφώθηκαν στο παρελθόν κατά φθίνοντα ή εκθετικό τρόπο.

Από τη στιγμή που βρίσκονται στη διάθεσή μας οι προβλεφθείσες τιμές, μπορούν να συγκριθούν με γνωστές τιμές και στη συνέχεια να υπολογιστεί το σφάλμα πρόβλεψης,  $e$ .

Για να γίνει μια πρόβλεψη πρέπει να ακολουθήσουμε τα εξής βήματα:

1. Επιλέγουμε μια μέθοδο πρόβλεψης βασισμένη σε ήδη αποκτημένη γνώση γύρω από το παρατηρηθέν μοντέλο της χρονοσειράς.
2. Χρησιμοποιούμε τη μέθοδο πρόβλεψης για την απόκτηση εκτιμημένης τιμής των δεδομένων.
3. Υπολογίζουμε το σφάλμα πρόβλεψης.
4. Παίρνουμε απόφαση για την καταλληλότητα του μοντέλου που βασίζεται στη μέτρηση του σφάλματος πρόβλεψης.

Αρχίζουμε με τις απλούστερες μεθόδους του απλού κινητού μέσου όρου και της εκθετικής εξομάλυνσης μιας παραμέτρου, καθώς και με την αξιολόγηση του σφάλματος στη πρόβλεψη.

Αν το πρότυπο των δεδομένων έχει τάση, η μέθοδος του γραμμικού κινητού μέσου όρου προσαρμόζει τον απλό κινητό μέσο στην τάση. Η μέθοδος Holt είναι μια άλλη μέθοδος εξομάλυνσης με την οποία επιτυγχάνεται προσαρμογή στην τάση. Βεβαίως αυτή αποτελεί μοντέλο με δύο παραμέτρους. Τέλος, η μέθοδος Winter είναι μέθοδος εξομάλυνσης με προσαρμογές στην τάση και την εποχικότητα. Περιλαμβάνει τρεις παραμέτρους και αποτελεί επέκταση του μοντέλου διπλής παραμέτρου γραμμικής εξομάλυνσης του Holt.

➤ Στη παράγραφο 1.2 γίνεται ανάλυση των μεθόδων διαχωρισμού και ανάλυσης χρονοσειρών.

Στις τεχνικές που αναφέρθηκαν προηγουμένως χρησιμοποιήσαμε απλές μεθόδους προεκβολής προκειμένου να προβλέψουμε μελλοντικές τιμές μιας χρονοσειράς από παρατηρηθείσες τιμές. Η χρονική περίοδος εκφραζόταν σε έτη, τρίμηνα ή μήνες. Οι χρονοσειρές μελετώνται κάτω από την υπόθεση ότι με βάση τη συμπεριφορά στο παρελθόν μπορεί να προβλεφθεί η μελλοντική πορεία της διακύμανσης της παρατηρούμενης μεταβλητής. Μολονότι η πρόβλεψη μελλοντικών καταστάσεων δεν γίνεται χωρίς σφάλμα, η ανάλυση και ο διαχωρισμός (decomposition) των χρονοσειρών έχει αξία στη διαδικασία της πρόβλεψης γιατί έτσι μπορεί να μειωθεί το σφάλμα πρόβλεψης.

Επισημαίνεται ότι είναι αδύνατο να προβλεφθεί το μέλλον με απόλυτη ακρίβεια και ότι πάντοτε θα υπάρχει, ως ένα βαθμό, σφάλμα στην πρόβλεψη. Οι μέθοδοι για την ανάλυση και το διαχωρισμό χρονοσειρών είναι στατιστικές τεχνικές βελτίωσης των προβλέψεων μέσω της διάσπασης του προτύπου της χρονοσειράς σε υπο-πρότυπα που ονομάζονται συνιστώσες.

Στη παράγραφο που αφορά τις μεθόδους διαχωρισμού και ανάλυσης χρονοσειρών μελετούμε τη

## ΠΕΡΙΛΗΨΗ

---

φύση και την έκταση αυτών των συνιστώσων. Οι τεχνικές διαχωρισμού που χρησιμοποιούνται για τη μελέτη αυτών των συνιστώσων δεν αποτελούν ενιαίο σώμα της στατιστικής θεωρίας. Επίσης, δεν πληρούνται συνήθως όλες οι προϋποθέσεις για επιστημονική στατιστική διερεύνηση, όπως για παράδειγμα, η ύπαρξη τυχαίας δειγματοληψίας. Επομένως, οι διαδικασίες διάσπασης των χρονοσειρών συχνά αξιολογούνται με κριτήρια που επηρεάζονται από την κρίση εκείνου που διενεργεί την πρόβλεψη.

Ο διαχωρισμός σε συνιστώσες υποθέτει ότι τα δεδομένα της χρονοσειράς διαμορφώνονται από κάποιο πρότυπο (ή πρότυπα) και το σφάλμα. Αυτά τα πρότυπα ονομάζονται συστηματικές συνιστώσες. Είναι η τάση, η εποχιακότητα και η κυκλικότητα.

Η εποχιακότητα ή εποχικές διακυμάνσεις, είναι μεταβολές που λαμβάνουν χώρα με κανονικότητα κατά τη διάρκεια ενός έτους και είναι το αποτέλεσμα των συνηθειών και των κλιματολογικών συνθηκών. Η τάση μιας χρονοσειράς είναι το αποτέλεσμα μακροχρόνιων οικονομικών δυνάμεων που επιδρούν στη διαμόρφωση των τιμών της υπό μελέτη χρονοσειράς. Οι κύκλοι είναι κυμάνσεις που καλύπτουν μεγαλύτερο του ενός έτους διάστημα και προέρχονται από μεταβολές στη συνολική επιχειρηματική και οικονομική δραστηριότητα.

Οι χρονοσειρές υπόκεινται επίσης σε τυχαίες ή μη-κανονικές διακυμάνσεις. Η μέθοδος *ratio-to-moving-average* ελαχιστοποιεί την επίδραση της τάσης των κυκλικών και των μη-συστηματικών στοιχείων από τα δεδομένα της αρχικής χρονοσειράς και διασφαλίζει ότι τελικά παραμένει μόνον η επίδραση του εποχικού παράγοντα.

Οι εποχικοί δείκτες χρησιμοποιούνται για να εξαλείψουν τον εποχικό συντελεστή από τα δεδομένα της χρονοσειράς. Αυτοί οι δείκτες μπορεί να είναι σταθεροί ή κινητοί, όπως επίσης μπορεί να εκφράζονται ως απόλυτοι ή σχετικοί αριθμοί.

Η τάση μιας χρονοσειράς μπορεί γραφικά να παρασταθεί σε αριθμητική ή σε ημι-λογαριθμική κλίμακα. Πολλές είναι οι μέθοδοι που μπορούν να χρησιμοποιηθούν για την εκτίμηση της τάσης μιας χρονοσειράς. Μια τέτοια είναι η μέθοδος των ελαχίστων τετραγώνων. Οι εκτιμήσεις της τάσης για ένα χρόνο μπορούν να τροποποιηθούν σε εκτιμήσεις τάσης για περιόδους μικρότερες του έτους.

Η κυκλική συνιστώσα μπορεί να προβλεφθεί καλύτερα αν γνωρίζουμε καλά την τρέχουσα κατάσταση της οικονομίας. Ωστόσο, μπορούμε να εξαλείψουμε την κυκλική συνιστώσα με τη διαδικασία των κινητών μέσων για δεδομένα που έχουν προηγουμένως απαλλαγεί από την εποχιακότητα.

Το υπόλοιπο της παραγράφου αφιερώνεται σε παρατηρήσεις σχετικά με την ανάλυση χρονοσειρών και τις εφαρμογές της στη διαδικασία της πρόβλεψης. Απαιτείται καλή αξιολογική κρίση προκειμένου να χρησιμοποιηθούν οι μέθοδοι αυτές σε προβλήματα πρόβλεψης. Εν τούτοις, επισημαίνοντας τους στατιστικούς περιορισμούς αυτών των τεχνικών, μπορούμε να μειώσουμε το διάστημα κατά το οποίο θα χρειαστεί να λειτουργήσει η υποκειμενική κρίση κι επομένως να βελτιώσουμε τη διαδικασία πρόβλεψης.

Στο δεύτερο κεφάλαιο εξετάζεται το πώς ο ασχολούμενος με την πρόβλεψη μπορεί να

## ΠΕΡΙΛΗΨΗ

---

μελετήσει τη σχέση ανάμεσα στις μεταβλητές. Κάτι τέτοιο τον καθιστά ικανό να θεμελιώσει παλινδρομικά μοντέλα για πρόβλεψη. Εισάγονται αρχικά τα απλά γραμμικά μοντέλα. Ακολουθεί η πολλαπλή παλινδρόμηση και τα καμπυλόγραμμα μοντέλα. Στη συνέχεια εισάγονται τα οικονομικά μοντέλα που χρησιμοποιούνται στην πρόβλεψη ώστε να ολοκληρωθούν οι διαδικασίες που βασίζονται στην παλινδρόμηση. Ειδικότερα:

➤ Στη παράγραφο 2.1 μελετούμε με την απλή παλινδρόμηση. Ο σκοπός της απλής παλινδρομικής ανάλυσης είναι η εκτίμηση της φύσης της σχέσης που υπάρχει ανάμεσα σε μια ανεξάρτητη και μια εξαρτημένη μεταβλητή. Ο σκοπός ανάλυσης της συσχέτισης είναι η απάντηση στο ερώτημα που αφορά το βαθμό εξάρτησης μεταξύ δύο μεταβλητών. Αν και μπορεί να υπάρχουν ευλογοφανείς αιτίες για μια σχέση, η ύπαρξη υψηλού βαθμού συσχέτισης δεν μας λέει τίποτα για αυτές τις αιτίες. Δείχνει μόνο το εάν δύο μεταβλητές συμμεταβάλλονται ή όχι.

Η πρώτη υπόθεση που γίνεται κατά την ανάλυση του απλού παλινδρομικού μοντέλου είναι ότι η σχέση είναι γραμμική. Τα σφάλματα πρόβλεψης μπορεί να προκύψουν από:

(α) ατελή θεωρία-ορισμένες μεταβλητές σχετιζόμενες με εξαρτημένες μεταβλητές μπορεί να έχουν εξαιρεθεί από την παλινδρόμηση.

(β) ατελή εξειδίκευση-εκτίμηση γραμμικής σχέσης όταν στην πραγματικότητα η σχέση είναι μη-γραμμική.

(γ) σφάλματα μέτρησης-τέτοια σφάλματα μπορεί να προκύψουν κατά τη μέτρηση των μεταβλητών.

Άλλες υποθέσεις του μοντέλου παλινδρόμησης είναι:

1. Η ίδια κατανομή των τιμών  $X$  μπορεί να εμφανιστεί σε οποιαδήποτε επανάληψη του πειράματος.
2. Ο όρος του τυχαίου σφάλματος έχει μέσο μηδέν.
3. Η διακύμανση του όρου του τυχαίου σφάλματος είναι σταθερή για κάθε τιμή του  $X$  (ομοσκεδαστικότητα).
4. Οι όροι του σφάλματος είναι ανεξάρτητοι μεταξύ τους, δηλαδή δεν υπάρχει αυτοσυσχέτιση.
5. Ο όρος του σφάλματος για κάθε τιμή του  $X$  κατανέμεται κανονικά.

Το τυπικό σφάλμα της παλινδρόμησης παρέχει μια εκτίμηση της ακριβείας της εξίσωσης που εκτιμάται. Είναι η τετραγωνική ρίζα του αθροίσματος των τετραγώνων των αποκλίσεων των καταλοίπων από τη γραμμή παλινδρόμησης διαιρεμένο διά  $n-k$ , όπου  $k$  είναι ο αριθμός των προς εκτίμηση συντελεστών στην εξίσωση παλινδρόμησης, δηλαδή η τετραγωνική ρίζα του μέσου "σφάλματος τετραγώνου.

Μια παλινδρόμηση έχει έννοια εάν η κλίση της γραμμής δείχνει να είναι διαφορετική από το μηδέν. Επίσης, μια υψηλή τιμή του  $F$  δείχνει ότι η διακύμανση που οφείλεται στην παλινδρόμηση είναι μεγαλύτερη από τη διακύμανση που δεν οφείλεται στην παλινδρόμηση. Η σταθερά του γραμμικού μοντέλου δείχνει την τιμή του  $Y$  όταν  $X$  ισούται με μηδέν.

Ο συντελεστής προσδιορισμού αποτελεί μέτρηση του ποσοστού της διακύμανσης της εξαρτημένης μεταβλητής που ερμηνεύεται από την παλινδρόμηση της ανεξάρτητης μεταβλητής. Η συσχέτιση είναι θετική εάν οι τιμές των  $X$  και  $Y$  μεταβάλλονται προς την ίδια κατεύθυνση. Εάν μετα-

## ΠΕΡΙΛΗΨΗ

---

βάλλονται προς αντίθετες κατευθύνσεις, η συσχέτιση είναι αρνητική.

Η γραφική παράσταση των πραγματικών και των προβλεφθεισών τιμών και των καταλοίπων παρέχει ένα πολύ χρήσιμο τρόπο διαπίστωσης τυχόν αποκλίσεων από τις παραδοχές (υποθέσεις) του μοντέλου παλινδρόμησης. Διάφορα πρότυπα καταλοίπων μπορούν να δείξουν αυτοσυσχέτιση, μη-γραμμικότητα ή ετεροσκεδαστικότητα.

Το τεστ Darbin-Watson δείχνει το εάν διαδοχικές τιμές σφάλματος αυτοσυσχετίζονται. Υπάρχουν διάφορες διαδικασίες για τη διόρθωση της αυτοσυσχέτισης. Μεταξύ αυτών περιλαμβάνεται η διόρθωση με υπολογισμούς με το χέρι.

Το ενδιαφέρον συχνά επικεντρώνεται στο παρατηρούμενο πρότυπο της διασποράς γύρω από τους εκτιμηθέντες συντελεστές παλινδρόμησης. Τα διαστήματα εμπιστοσύνης για τη μέση τιμή και την ατομική τιμή, επιτρέπουν στον προβλεπόμενα να εξηγήσει τη διασπορά στις εκτιμήσεις.

➤ Στη παράγραφο 2.2 μελετούμε με την πολλαπλή παλινδρόμηση. Γενικά, οι μέθοδοι παλινδρόμησης που χρησιμοποιούν περισσότερες από μία ανεξάρτητες μεταβλητές ονομάζονται μέθοδοι πολλαπλής παλινδρόμησης. Στην πολλαπλή παλινδρόμηση έχουμε σφάλματα που οφείλονται στην ατελή θεωρία, σε ατελή εξειδίκευση και σε υπολογιστικά λάθη.

Οι συντελεστές που εκτιμώνται στην πολλαπλή παλινδρόμηση μετρούν την καθαρή μεταβολή στην εξαρτημένη μεταβλητή που σχετίζεται με μια ανά μονάδα μεταβολή στην ανεξάρτητη μεταβλητή.

Η μέθοδος των ελαχίστων τετραγώνων για την εκτίμηση της εξίσωσης πολλαπλής παλινδρόμησης παρέχει την καλύτερη προσαρμογή ελαχιστοποιώντας τα αθροίσματα των κάθετων αποκλίσεων ανάμεσα στις παρατηρηθείσες τιμές της εξαρτημένης μεταβλητής και τις εκτιμηθείσες τιμές.

Ένα μέτρο του βαθμού προσαρμοστικής ικανότητας του μοντέλου είναι ο συντελεστής προσδιορισμού. Αυτό είναι το άθροισμα των τετραγώνων που οφείλεται στην παλινδρόμηση διά του συνολικού αθροίσματος των τετραγώνων ή τη συνολική διακύμανση στις τιμές του  $Y$ . Ο συντελεστής πολλαπλής συσχέτισης είναι πάντοτε μη αρνητικός και αποτελεί μια απευθείας γενίκευση της απλής συσχέτισης στην περίπτωση των δύο ή περισσότερων ανεξάρτητων μεταβλητών.

Για να ελέγξουμε τη συνολική σημαντικότητα του μοντέλου πολλαπλής παλινδρόμησης, γίνεται ανάλυση της διακύμανσης (ANOVA) προκειμένου να εξεταστεί αν το άθροισμα τετραγώνων του μοντέλου είναι ή δεν είναι σημαντικά μεγαλύτερο από το άθροισμα τετραγώνων του σφάλματος.

Εκτός από τον έλεγχο της συνολικής ερμηνευτικής ικανότητας του μοντέλου, υπάρχουν κριτήρια για τον έλεγχο της σημαντικότητας των επιμέρους μεταβλητών, της συμβολής μιας πρόσθετης μεταβλητής κλπ.

Τα γραφήματα των καταλοίπων στην ανάλυση της πολλαπλής παλινδρόμησης δείχνουν το ίδιο φαινόμενο, όπως και στην ανάλυση της απλής γραμμικής παλινδρόμησης.

Η πολυσυγγραμμικότητα επιδρά όταν ζεύγη ανεξάρτητων μεταβλητών σχετίζονται μεταξύ τους και οι ανεξάρτητες μεταβλητές συνεισφέρουν σε πλασματικές πληροφορίες. Η ύπαρξη πολυσυγγραμμικότητας δείχνει ότι το μοντέλο παλινδρόμησης δεν έχει την ικανότητα να διαχωρίζει

## ΠΕΡΙΛΗΨΗ

---

την ακριβή συμβολή κάθε ανεξάρτητης μεταβλητής στην ερμηνεία της συμπεριφοράς της εξαρτημένης μεταβλητής.

Η επιλογή του καλύτερου μοντέλου παλινδρόμησης απαιτεί ένα συμβιβασμό μεταξύ της επίτευξης του υψηλότερου συντελεστή προσδιορισμού και της επιλογής ενός ελάχιστου αριθμού μεταβλητών πρόβλεψης.

Η μέθοδος *stepwise* για την επιλογή του καλύτερου μοντέλου πρόβλεψης εξετάζει όλες τις μεταβλητές και απορρίπτει οποιαδήποτε μεταβλητή δεν δίδει ένα σημαντικό F-στατιστικό. Το τελικό μοντέλο πρόβλεψης είναι το καλύτερο υπό περιορισμένη έννοια. Το μοντέλο οριοθετείται από το αρχικό σύνολο των ανεξάρτητων μεταβλητών-τα επιμέρους επίπεδα σημαντικότητας, που επιλέγονται για την εισαγωγή ή την απόρριψη των ανεξάρτητων μεταβλητών, εν μέρει καθορίζουν την τελική εξίσωση πρόβλεψης μια εναλλακτική τεχνική επιλογής ίσως θα μπορούσε να αποφέρει ένα διαφορετικό μοντέλο πρόβλεψης.

➤ Στη παράγραφο 2.3 μελετούμε τα οικονομετρικά μοντέλα. Οικονομετρία είναι η επιστήμη της οικονομικής μέτρησης που επιδιώκει να θεμελιώσει ποσοτικές σχέσεις ανάμεσα σε οικονομικές μεταβλητές με τη βοήθεια της στατιστικής θεωρίας. Τα οικονομετρικά μοντέλα είναι συστήματα ταυτόχρονα προσδιοριζόμενων εξισώσεων που συνδέουν τις οικονομικές μεταβλητές, με σκοπό την πρόβλεψη και την εξήγηση της οικονομικής συμπεριφοράς:

Μια συνδυασμένα, εξαρτημένη μεταβλητή ονομάζεται *ενδογενής* και μια προκαθορισμένη μεταβλητή ονομάζεται *εξωγενής*. Το θεωρητικό σχήμα του οικονομετρικού μοντέλου καθορίζει αν μια μεταβλητή ταξινομείται σαν εξωγενής ή σαν ενδογενής.

Ένα πλήρες σύστημα ταυτόχρονα προσδιοριζόμενων εξισώσεων θα περιέχει τόσες εξισώσεις όσες και ενδογενείς μεταβλητές.

Οι διαρθρωτικές εξισώσεις δείχνουν τις σχέσεις μεταξύ μεταβλητών σε όρους μιας ήδη ισχύουσας ή μιας νέας οικονομικής θεωρίας. Οι συντελεστές αυτών των εξισώσεων ονομάζονται διαρθρωτικοί συντελεστές.

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

Στους περιορισμούς των οικονομετρικών μοντέλων περιλαμβάνονται η πολυσυγγραμμικότητα των μεταβλητών και το σφάλμα που σχετίζεται με την ταυτόχρονη εκτίμηση των εξισώσεων του μοντέλου. Επίσης, τα οικονομετρικά μοντέλα μεγάλης κλίμακας, που είναι έντονα αθροιστικά, δεν ενσωματώνουν διαθέσιμες λεπτομερείς πληροφορίες που αφορούν εξελίξεις σε ξεχωριστούς τομείς της οικονομίας. Τέλος, τα μοντέλα αυτά χρειάζονται συνεχή παρακολούθηση σε σχέση με την καταλληλότητά τους να διαγνώσουν τις αναγκαίες περιοδικές μεταβολές.

Για να αποφύγουμε το μεροληπτικό σφάλμα των ταυτόχρονα προσδιοριζόμενων εξισώσεων,

## ΠΕΡΙΛΗΨΗ

---

εφαρμόζουμε τη μία ή την άλλη από τις τεχνικές που ακολουθούν: FIML, LIML, 2SLS ή 3SLS ή ILS. Η 2SLS είναι η τεχνική που χρησιμοποιείται ευρύτερα.

Άλλα προβλήματα που σχετίζονται με την κατασκευή οικονομετρικών μοντέλων είναι εκείνα της εξειδίκευσης και της ταυτοποίησης. Συχνά, οι επιχειρήσεις υιοθετούν οικονομετρικά μοντέλα, συνδέοντας μια πρόβλεψη επί των πωλήσεων με έναν ή περισσότερους οικονομικούς δείκτες, που δημιουργήθηκαν από κατάλληλο οικονομετρικό μοντέλο κάποιας συμβουλευτικής εταιρίας. Η σύνδεση αυτή, επιτρέπει στις εταιρίες να εφαρμόζουν μοντέλα μεγάλης κλίμακας χωρίς να χρειάζεται να κατασκευάσουν κάποιο ειδικά για τον εαυτό τους.

Σήμερα, δεν υπάρχει ξεκάθαρος και απλός τρόπος για την αξιολόγηση της συνολικής προβλεπτικής ικανότητας, στα πλαίσια ενός τόσο μεγάλου αριθμού ειδικών ασχολουμένων με την οικονομετρική πρόβλεψη.

Στο τρίτο κεφάλαιο αναπτύσσονται οι μέθοδοι αυτοπαλινδρόμησης κινητού μέσου (ARMA) χρονοσειρών. Αυτές οι μέθοδοι πρόβλεψης περιλαμβάνουν γενικευμένες μεθόδους αναπροσαρμοζόμενου φιλτραρίσματος και τεχνικές Box-Jenkins. Επίσης, αναλύεται η χρήση της ασαφούς λογικής και των νευρωνικών δικτύων στη διαδικασία της πρόβλεψης. Ειδικότερα:

➤ Ο σκοπός της παραγράφου 3.1 είναι να εισαγάγει τον αναγνώστη σε χρήσιμες έννοιες της ανάλυσης χρονοσειρών και στην ταυτοποίηση μοντέλων που θα εξυπηρετούν τους σκοπούς της πρόβλεψης. Επίσης, επικεντρώνουμε την προσοχή μας στις μεθόδους και τις αρχές παλινδρόμησης, όπως εφαρμόζονται απευθείας στη μελέτη χρονοσειρών.

Έχουμε αυτοσυσχέτιση (AR) όταν οι μεταβλητές της δεξιάς πλευράς της εξίσωσης είναι απλά οι τιμές των εξαρτημένων μεταβλητών στις προηγούμενες περιόδους και που είναι χρονικά υστερημένες παρατηρήσεις.

Έχουμε εντελώς τυχαία χρονοσειρά ή white noise όταν δεν υπάρχει κανένα ορατό πρότυπο σ' αυτή. Με τη μέτρηση και εξέταση των αυτοσυσχετίσεων για χρονικές υστερήσεις περισσότερες της μιας περιόδων, παρέχεται ένδειξη για το πώς σχετίζονται οι τιμές μιας δεδομένης χρονοσειράς.

Εάν μια στοχαστική διαδικασία δεν μεταβάλλεται ως προς το χρόνο (σταθερή), μπορούμε εύκολα να μοντελοποιήσουμε τη χρονοσειρά μέσω μιας εξίσωσης με σταθερούς συντελεστές που έχουν υπολογιστεί από ιστορικές παρατηρήσεις. Μια μη-σταθερή χρονοσειρά είναι εκείνη όπου οι διαρθρωτικές σχέσεις ενός μοντέλου αλλάζουν μέσα στο χρόνο. Εάν σε μία μη-σταθερή χρονοσειρά πάρουμε τις διαφορές μία ή περισσότερες φορές και η προκύπτουσα χρονοσειρά είναι σταθερή, λέμε ότι η χρονοσειρά είναι ομογενής.

Η συνάρτηση αυτοσυσχέτισης είναι εξαιρετικά χρήσιμη για την περιγραφή της διαδικασίας θεμελιώδους ανάπτυξης ενός προβλεπτικού μοντέλου. Η αυτοσυσχέτιση πρώτης τάξης είναι η αυτοσυσχέτιση χρονοσειράς σε περίοδο  $t$  με τις παρατηρήσεις της χρονοσειράς στην περίοδο  $t-1$ .



## ΠΕΡΙΛΗΨΗ

---

Όμοια, η αυτοσυσχέτιση υψηλότερων τάξεων αναφέρεται σε συσχέτιση με περιόδους  $t-2$ ,  $t-3$ . κ.λπ.

Σημαντικά ερωτήματα που δημιουργούνται γύρω από μια χρονοσειρά είναι τα εξής:

1. Είναι τυχαία τα δεδομένα;
2. Είναι τα δεδομένα μη-σταθερά;
3. Είναι τα δεδομένα εποχικά;

Οι απαντήσεις στα παραπάνω ερωτήματα μας δίνουν τη δυνατότητα καλύτερων προβλέψεων.

Ο συντελεστής μερικής αυτοσυσχέτισης μετράει το βαθμό της σχέσης μεταξύ μιας παρατήρησης  $Y_t$ , και της  $Y_{t-k}$ , όταν η επίδραση των άλλων χρονικών υστερήσεων παραμένει σταθερά. Μελετάμε τις μερικές αυτοσυσχετίσεις όταν δεν γνωρίζουμε την κατάλληλη τάξη της διαδικασίας αυτοπαλινδρόμησης που ταιριάζει στη χρονοσειρά.

➤ Στη παράγραφο 3.2, μελετούμε προσεγγίσεις στην πρόβλεψη χρονοσειρών, όπου η γνώση των ιδιοτήτων των δεδομένων της χρονοσειράς εξυπηρετεί καλύτερες προβλέψεις. Η εν λόγω παράγραφος επικεντρώνεται στην εξέταση πρόσθετων μοντέλων χρονοσειρών και στη μελέτη των μεθόδων του αναπροσαρμοζόμενου φίλτραρίσματος (adaptive filtering). Η βάση της μεθόδου είναι ότι οι σταθμικοί όροι που χρησιμοποιούνται στη συγκεκριμένη τεχνική αλλάζουν όταν αποκτηθούν νέες παρατηρήσεις.

Ειδικότερα η μέθοδος του adaptive filtering είναι μια μέθοδος πρόβλεψης χρονοσειρών, όπου η γνώση των ιδιοτήτων της χρονοσειράς χρησιμοποιείται για την παραγωγή καλύτερων προβλέψεων. Η αιτιολογία για τη χρησιμοποίηση της μεθόδου αυτής είναι ότι μια πρόβλεψη μπορεί να αυτοπροσαρμοστεί σε ένα υπό μελέτη πρότυπο δεδομένων.

Ένα αυτοπαλίνδρομο μοντέλο (AR) διαφέρει από τα γενικά γραμμικά παλινδρομικά μοντέλα στο ότι οι ανεξάρτητες μεταβλητές είναι ιστορικές τιμές της εξαρτημένης μεταβλητής. Η τάξη  $P$ , ενός μοντέλου AR προσδιορίζει τον αριθμό των παραμέτρων του μοντέλου.

Η μέθοδος του adaptive filtering εφαρμοζόμενη στα μοντέλα AR, χρησιμοποιεί τις πληροφορίες που περιέχονται σε παρελθούσες παρατηρήσεις για την εύρεση του καλύτερου συνόλου συντελεστών στάθμισης. Οι σταθμίσεις αυτές επικαιροποιούνται (updated) συνεχώς, όταν μεταβάλλεται το πρότυπο της χρονοσειράς.

Το μοντέλο κινητού μέσου (MA) δίνει προβλέψεις που βασίζονται σε ένα γραμμικό συνδυασμό παρελθόντων σφαλμάτων πρόβλεψης και επομένως, είναι παρόμοιο με τα μοντέλα εκθετικής εξομάλυνσης. Η τάξη  $q$ , ενός μοντέλου MA αναφέρεται στον αριθμό των παραμέτρων του μοντέλου.

Εάν το πρότυπο αυτοσυσχέτισης στα κατάλοιπα της χρονοσειράς απαιτεί ένα πιο σύνθετο μοντέλο, μπορούμε να υπολογίσουμε ένα μικτό μοντέλο. Ένα αυτοπαλίνδρομο μοντέλο κινητού μέσου ARMA, τάξης  $(p,q)$  έχει αυτοσυσχετίσεις που μειώνονται, καθώς αυξάνεται η απόσταση μεταξύ των καταλοίπων. Επίσης, τα πρότυπα των χρονοσειρών που μπορούν να περιγραφούν από διαδικασίες ARMA είναι πιο γενικά από εκείνο των μοντέλων AR(p) ή των MA(q).

Η ταυτοποίηση των AR(p), MA(q) και των μικτών χρονοσειρών ARMA  $(p.,q)$  απαιτεί τον

## ΠΕΡΙΛΗΨΗ

---

υπολογισμό και τη γραφική αποτύπωση της συνάρτησης αυτοσυσχέτισης (συντελεστές) του δείγματος και της συνάρτησης μερικής αυτοσυσχέτισης (συντελεστές) των δεδομένων της χρονοσειράς. Εάν οι συντελεστές αυτοσυσχέτισης και οι συντελεστές μερικής αυτοσυσχέτισης του δείγματος για μια χρονοσειρά φαίνεται να διαμορφώνονται από κάποιο συγκεκριμένο μοντέλο χρονοσειράς, τότε είναι πιθανόν ότι αυτό το ιδιαίτερο μοντέλο χρονοσειράς διαμόρφωσε τα δεδομένα της χρονοσειράς.

➤ Στη παράγραφο 3.3 αναλύουμε τη προσέγγιση των Box-Jenkins, η οποία είναι μια στατιστικά θεμελιωμένη μέθοδος ανάλυσης και κατασκευής μοντέλων πρόβλεψης που αναπαριστούν ικανοποιητικά μια χρονοσειρά.

Οι ιδιότητες μιας χρονοσειράς καθορίζονται από τη μελέτη των αρχικών δεδομένων και των γραφημάτων των απλών και μερικών αυτοσυσχετίσεων. Η μελέτη αυτή επιτρέπει στον προβλέποντα να τακτοποιήσει την τάση, να αντιμετωπίσει την εποχιακότητα και να ελέγξει τη σταθερότητα της διακύμανσης στη χρονοσειρά. Ο σκοπός της ταυτοποίησης είναι η επιλογή του συγκεκριμένου μοντέλου ARMA που έχει διαμορφώσει την παρατηρηθείσα χρονοσειρά.

Τα μοντέλα ARMA εκτιμούνται με τη μη-γραμμική μέθοδο ελαχίστων τετραγώνων. Για την εκτίμηση περισσότερων μοντέλων ARMA χαμηλής τάξης, υπάρχουν έτοιμα προγράμματα (Minitab, SAS, SPSS).

Για τον έλεγχο της καταλληλότητας του μοντέλου, οι όροι του σφάλματος εξετάζονται διαγνωστικά, ώστε να καθαριστεί αν αυτοί είναι white noise (τυχαίοι). Μετά από την ταυτοποίηση και την αξιολόγηση της ισχύος ενός μοντέλου ARMA, μπορούν να γίνουν προβλέψεις για μία ή περισσότερες περιόδους στο μέλλον. Καθώς αποκτώνται νέες παρατηρήσεις, μπορεί να χρησιμοποιηθεί το ίδιο μοντέλο για την αναθεώρηση των προβλέψεων μέσω της επιλογής άλλης χρονικής στιγμής ως χρόνου έναρξης.

Μια μη-στάσιμη χρονοσειρά ως προς τον μέσο, συχνά μετασχηματίζεται σε σταθερή αν πάρουμε *d* διαφορίσεις. Το αποτέλεσμα θα είναι ένα μοντέλο ARMA για τη *διαφορισμένη χρονοσειρά*. Εάν η χρονοσειρά είναι μη-στάσιμη ως προς τη διακύμανση, η σειρά μπορεί να σταθεροποιηθεί με το μετασχηματισμό των αρχικών δεδομένων. Η αρχική χρονοσειρά αναφέρεται ως ένα μοντέλο ARIMA. Το σύμβολο I χρησιμοποιείται για την ολοκλήρωση (I= integrated) και υποδηλώνει μη-διαφόριση.

Για την εξυπηρέτηση των σκοπών της πρόβλεψης, αρχικά, ταυτοποιούμε το ενδεδειγμένο μοντέλο ARIMA και διαφορίζουμε τα δεδομένα ώστε να επιτύχουμε στασιμότητα. Έτσι, το μοντέλο μετασχηματίζεται σε μοντέλο της κατηγορίας ARMA.

Πολλές χρονοσειρές είναι εποχικές. Η εποχιακότητα ποικίλλει από έτος σε έτος δείχνοντας ότι τα μοντέλα που είναι βασισμένα σε μη-προσαρμοσμένα δεδομένα παρά σε εποχικά προσαρμοσμένα δεδομένα, είναι περισσότερο πιθανό να αποδειχθούν ευέλικτα και χρήσιμα για τους σκοπούς της ταυτοποίησης.

Σε ένα πολλαπλασιαστικό μοντέλο ARIMA, οι συνήθεις και οι εποχικές συνιστώσες

## ΠΕΡΙΛΗΨΗ

αυτοπαλινδρόμησης, οι συνιστώσες διαφόρισης και κινητού μέσου όρου πολλαπλασιάζονται ταυτόχρονα στο γενικό μοντέλο. Ένας πολύ χρήσιμος συμβολισμός για την περιγραφή των τάξεων των διαφόρων συνιστωσών στο πολλαπλασιαστικό μοντέλο δίνεται από

$$(p,d,q) \times (P,D,Q)^s$$

Για λόγους που αφορούν την ταυτοποίηση μιας εποχικής διαδικασίας ARMA, αγνοούμε τη μη-εποχική διαδικασία και καθορίζουμε το εάν η εποχιακότητα εκφράζεται με μια διαδικασία AR ή MA.

➤ Στη παράγραφο 3.4 αναπτύσσονται η έννοιες της ασαφούς λογικής και της θεωρίας ασαφών συνόλων. Δίδονται οι σχέσεις που τις διέπουν και αναλύεται ο τρόπος εφαρμογής τους στη παραγωγή προβλέψεων.

Γενικά, οι οργανισμοί, που είτε παράγουν αγαθά είτε παρέχουν υπηρεσίες, απαιτούν προβλέψεις των απαιτήσεων για τις αποφάσεις τους. Επομένως, ένας από τους κύριους σκοπούς (στόχους) των λειτουργιών της διοικήσεως περιλαμβάνει τεχνικές πρόβλεψης του τύπου των απαιτήσεων που πρέπει να είναι έγκαιρες, εύλογα ακριβείς και αξιόπιστες. Υπάρχουν δύο κατηγορίες ποσοτικών τεχνικών πρόβλεψης που μπορούν να χρησιμοποιηθούν: τα αιτιώδη πρότυπα και οι χρονοσειρές. Τα αιτιώδη πρότυπα στοχεύουν στη διευκρίνιση των παραγόντων που επηρεάζουν και στις σχέσεις τους με τα αποτελέσματα. Από την άλλη, τα πρότυπα των χρονοσειρών παράγουν προβλέψεις από μια σειρά στοιχείων που παρατηρούνται σε κανονικά (τακτά) χρονικά διαστήματα, βασισμένα σε μελλοντικά δρώμενα που προκύπτουν από το σχέδιο (πρότυπο) του παρελθόντος.

Από τη δημιουργία της από τον καθηγητή Zadeh, η θεωρία ασαφών συνόλων αναγνωρίζεται όλο και περισσότερο ως χρήσιμο εργαλείο στο χειρισμό της ασάφειας της ανθρώπινης γνώσης. Η ασαφής ανάλυση παλινδρόμησης έχει εισαχθεί και μπορεί να χρησιμοποιηθεί για την καθιέρωση των συγκεχυμένων (ασαφών) αιτιωδών προτύπων. Η ασαφής ανάλυση χρονοσειρών έχει εισαχθεί με την εισαγωγή των αποκαλούμενων "τεχνικών ασαφούς διαμόρφωσης". Σε δεδομένα χρονοσειρών, τυπικά πρότυπα όπως η τάση, η εποχιακότητα, ακόμη και ο κύκλος μπορούν να υπάρξουν.

Επίσης, στη συγκεκριμένη παράγραφο αναπτύσσεται μια ασαφής τεχνική για τις τάσεις και την πρόβλεψη της εποχιακότητας. Μέσω της χρήσης της ασαφούς παλινδρόμησης, αναλύεται η ασαφής τάση των χρονοσειρών. Κατόπιν, η *ασαφής εποχιακότητα* καθορίζεται με τη χρησιμοποίηση των βαθμών ιδιότητας μέλους των εποχών στο πρότυπο παλινδρόμησης. Χρησιμοποιούνται τα ακόλουθα βήματα:

- (1) ανάλυση της ασαφούς τάσης με χρήση της ασαφούς ανάλυσης παλινδρόμησης.
- (2) υπολογισμός των *συνόλων των δεικτών ασαφούς εποχιακότητας* για τις εποχές.
- (3) χρήση των αποτελεσμάτων των βημάτων (1) και (2) για την παραγωγή συγκεχυμένων (ασαφών) ή/και σύντομων και αποφασιστικών προβλέψεων.
- (4) ανάλυση εποχιακής ασάφειας και εποχιακών τάσεων.

## ΠΕΡΙΛΗΨΗ

---

➤ Στη παράγραφο 3.5 μελετούμε τη παραγωγή προβλέψεων με χρήση των νευρωνικών δικτύων. Το όνομα "νευρωνικό δίκτυο" προέρχεται από τη νευρωνική δομή στον εγκέφαλο, και ορίστηκε έτσι από τους βιολόγους και τους φυσιολόγους που προσπάθησαν να μιμηθούν και να τυποποιήσουν τους νευρώνες στον εγκέφαλο. Το θέμα είναι σήμερα διεπιστημονικό, κυμαινόμενο από τους βιολόγους και τους ψυχολόγους, στόχος των οποίων είναι να γίνει κατανοητή καλύτερα η πραγματική διαδικασία υπολογισμού στον εγκέφαλο, από τους φυσικούς που έχουν διατυπώσει επιτυχώς τα στατιστικά μηχανικά πρότυπα των δικτύων, από τους μηχανικούς που ενδιαφέρονται για τα παράλληλα σχέδια φίλτρων, από τους μαθηματικούς, τους επιστήμονες υπολογιστών και τους στατιστικούς, για τους οποίους η λογική της αναγνώρισης του συλλογισμού και της ομιλίας είναι ενεργοί τομείς της έρευνας. Κάθε ένας από αυτούς τους επιστημονικούς κλάδους αντιμετωπίζει το πρόβλημα της τεχνητής νοημοσύνης από τη μοναδική οπτική του γωνία και ταυτόχρονα εμπνέεται από τις νευρο-φυσιολογικές εξελίξεις

Όσον αφορά την εφαρμογή των νευρωνικών δικτύων στη πρόβλεψη, επισημαίνονται τα παρακάτω. Ενώ τα γραμμικά μοντέλα, όπως εκείνα που είναι βασισμένα στις τεχνικές παλινδρόμησης, ήταν η βάση των παραδοσιακών στατιστικών μεθόδων πρόβλεψης, τα μειονεκτήματά τους έχουν οδηγήσει στην αυξανόμενη δραστηριότητα στη μη γραμμική διαμόρφωση. Τα νευρωνικά δίκτυα είναι μη γραμμικά μοντέλα που μπορούν να εκπαιδευθούν για να χαρτογραφήσουν τις προηγούμενες και μελλοντικές τιμές μιας χρονοσειράς και με αυτόν τον τρόπο εξάγουν την κρυμμένη δομή και τις σχέσεις που διέπουν τα δεδομένα.

Η πρόσφατη αυξητική τάση στην ερευνητική δραστηριότητα των νευρωνικών δικτύων είναι μια συνέπεια της διαπίστωσης ότι τα νευρωνικά δίκτυα έχουν μεγάλες δυνατότητες αναγνώρισης σχημάτων, που ξεπερνούν εκείνες άλλων τεχνικών σε πολλές εφαρμογές - για την ακρίβεια ή/και για την ταχύτητα υπολογισμού.

Αν και οι περιορισμένες μη γραμμικές τεχνικές παλινδρόμησης αναπτύσσονται, οι απεριόριστες μη γραμμικές ιδιότητες του νευρωνικού δικτύου το κάνουν προτιμότερο υποψήφιο στο τομέα της πρόβλεψης χρονοσειρών. Για να τοποθετήσουμε τα νευρωνικά δίκτυα στα πλαίσια των παραδοσιακών στατιστικών μεθόδων, πρέπει να θεωρηθούν ως "τεχνική εξαγωγής συμπεράσματος, μη γραμμική, μη παραμετρική και πολλών μεταβλητών". Το "*πολλών μεταβλητών*" αναφέρεται στο ότι τα δεδομένα εισαγωγής του νευρωνικού δικτύου έχουν πολλές διαφορετικές μεταβλητές των οποίων οι αλληλεξαρτήσεις και αλληλεπιδράσεις χρησιμοποιούνται για τη πρόβλεψη της μελλοντικής συμπεριφορά μιας χρονικής ακολουθίας. Το "*μη παραμετρικό*", είναι συνέπεια της έλλειψης οποιωνδήποτε υποθέσεων σχετικά με τη σχέση μεταξύ των μεταβλητών εισόδου και των παρεκτάσεων στο μέλλον. Μάλλον το δίκτυο «εκπαιδεύεται» από την προσαρμογή των *ελεύθερων παραμέτρων* για να ανακαλύψει οποιεσδήποτε πιθανές σχέσεις, απαλλαγμένο από του περιορισμούς των μοντέλων, καθώς διαμορφώνεται και σχηματίζεται απλώς από τα δεδομένα εισόδου.

Οι *ελεύθεροι παράμετροι* είναι βάρη που συνδέονται με τις γραμμές επικοινωνίας σημάτων μεταξύ των νευρώνων, και οι οποίοι μειώνουν τα σήματα που διέρχονται, ή τα δεδομένα.

## ΠΕΡΙΛΗΨΗ

---

Μία σημαντική πτυχή των νευρωνικών δικτύων είναι ικανότητάς τους να κατασκευάζουν *μη γραμμικές* σχέσεις μεταξύ των δεδομένων εισόδου και στόχου στην έξοδο. Η γραμμικότητα, ( $p_t = \alpha_1 p_{t-1} + \alpha_2 \dots$ , για μια χρονοσειρά τιμών  $p_t$  με σταθερές  $\alpha_i$ ) περιγράφει την απλούστερη σχέση που μπορεί να εμφανιστεί σε μια χρονοσειρά.

Κατά συνέπεια τα νευρωνικά δίκτυα καλύπτουν πολλά επιθυμητά χαρακτηριστικά γνωρίσματα, ως εργαλείο ανάλυσης στοιχείων μέσα σε ένα σχετικά αποδοτικό σχέδιο εφαρμογής-από την άποψη των απαιτήσεων μνήμης και ταχύτητας υπολογισμών.

Το νευρωνικό δίκτυο μπορεί να θεωρηθεί ως τεχνική επεξεργασίας δεδομένων που καθοδηγεί, ή συσχετίζει, κάποιο τύπο ρεύματος εισαγωγής πληροφοριών με ένα ρεύμα εξαγωγής δεδομένων. Παραδείγματος χάριν, η είσοδος μπορεί να είναι υπό μορφή διδιάστατης εικόνας με ελλείποντα ή διαστρεβλωμένα τμήματα, και η επεξεργασμένη έξοδος μπορεί να παραγάγει μια διορθωμένη εικόνα. Στις εφαρμογές χρονοσειρών που ενδιαφέρουν εδώ, η εισαγωγή μπορεί να είναι χρόνος μιας διάστασης, και η έξοδος η παραγωγή της καλύτερης εκτίμησης του επόμενου στοιχείου στη σειρά. Γενικά οι στόχοι των νευρωνικών δικτύων μπορούν να διαιρεθούν σε τέσσερις τύπους ευδιάκριτων εφαρμογών:

1. *Κατάταξη*: Απόφαση για την κατηγορία στην οποία περιέρχεται ένα σχέδιο εισαγωγής.
2. *Συσχέτιση*: Ενεργεί ως προσπελάσιμη μνήμη, που ανακαλεί ένα δεδομένο εξόδου δίδοντάς του κάποιο μέρος από αυτό ως είσοδο.
3. *Κωδικοποίηση*: Η κωδικοποίηση συμπιέζει ένα δεδομένο εισόδου με την παραγωγή ενός δεδομένου εξόδου με μειωμένη διάσταση. Ο αντίθετος στόχος, που αποκωδικοποιεί, μπορεί επίσης να είναι ενδιαφέρον.
4. *Προσομοίωση*: Η δημιουργία ενός δεδομένου εξόδου για ένα δεδομένο εισόδου που ενεργεί ως ερέθισμα, όταν το δίκτυο έχει εκτεθεί σε ένα δείγμα των πιθανών ερεθισμάτων.

Οποιοσδήποτε ιδιαίτερος στόχος μπορεί να περιλάβει διάφορες ανωτέρω λειτουργίες. Τα νευρωνικά δίκτυα μπορούν να εφαρμοστούν σε καθένα υλικό λογισμικού (software) ή υλικού (hardware).

Μερικοί στόχοι επεξεργασίας που εκτελούνται από τα νευρωνικά δίκτυα μπορούν επίσης να εκπληρωθούν από τις παραδοσιακές τεχνικές, όπως η ανάλυση βάσεων δεδομένων ή η ανάλυση παλινδρόμησης. Εντούτοις, κανένας δεν μοιράζεται τις μοναδικές δυνατότητες των νευρωνικών δικτύων, τα πλεονεκτήματα των οποίων μπορούν να συνοψιστούν ως εξής:

1. *Γενίκευση*: Τα νευρωνικά δίκτυα καθιερώνονται αρχικά μέσω μιας φάσης "κατάρτισης", στην οποία παρουσιάζονται εισαγωγές παραδειγμάτων και το δίκτυο εκπαιδεύεται για να εξαγάγει τις σχετικές πληροφορίες από αυτά τα πρότυπα. Στη συνέχεια το δίκτυο έχει την ικανότητα να γενικοποιήσει έτσι ώστε, μια έως τώρα απαρατήρητη εισαγωγή προτύπου να μπορεί επίσης υποβληθεί σε επεξεργασία.
2. *Ευελιξία*: Η σειρά των στόχων στην οποία το νευρωνικό δίκτυο μπορεί να εφαρμοστεί, υπερβαίνει κατά πολύ οποιαδήποτε παραδοσιακή τεχνική.
3. *Μη γραμμική "διαμόρφωση"*: Η διαδικασία χαρτογράφησης δικτύων περιλαμβάνει τις μη γραμμικές λειτουργίες που μπορούν συνεπώς να καλύψουν μια μεγαλύτερη σειρά της

## ΠΕΡΙΛΗΨΗ

---

πολυπλοκότητας του προβλήματος. Αν και υπάρχουν και άλλες μη γραμμικές τεχνικές, το νευρωνικό δίκτυο είναι ανώτερο στη γενικότητα και την πρακτική ευκολία του στην εφαρμογή.

**Στο δεύτερο μέρος** της εργασίας επιχειρείται μια ενημέρωση και επικαιροποίηση (update) παλαιότερων δημοσιεύσεων, σχετικές με την εκτίμηση των λογισμικών για προβλέψεις (forecasting software evaluation) ([1], [2], [3], [4], [5]). Οι δημοσιεύσεις αυτές ήταν βασισμένες σε συγκεκριμένα χαρακτηριστικά. Ο τρόπος με τον οποίο αρχικά συλλέξαμε τα στοιχεία και στη συνέχεια ο τρόπος που τα αξιολογήσαμε περιγράφεται παρακάτω:

- Επισκόπηση βιβλιογραφίας-δημοσιεύσεων

Αναζητήθηκαν στις μηχανές αναζήτησης του διαδικτύου ([6],[7]) καθώς και σε σελίδες επιστημονικού περιεχομένου ([8]-[13]), δημοσιεύσεις σχετικές με λογισμικά προβλέψεων και ειδικότερα δημοσιεύσεις που αφορούσαν την εκτίμηση και αξιολόγηση αυτών των λογισμικών. Συγκεντρώθηκαν αρκετές, εκ των οποίων αξιολογήθηκαν και μελετήθηκαν εκείνες που επικεντρώνονταν στην συγκριτική παράθεση λογισμικών βάσει συγκεκριμένων στοιχείων .

- Καταγραφή λογισμικών προβλέψεων.

Αφού έγινε επισκόπηση των δημοσιεύσεων, αναζητήθηκε εκ νέου στις μηχανές αναζήτησης του διαδικτύου ([6],[7]), κάθε είδους λογισμικό (είτε αυτό αποτελεί στατιστικό ή μαθηματικό εργαλείο ή καθαρά εργαλείο προβλέψεων) που εκτελεί προβλέψεις. Απώτερος στόχος μας ήταν η ενημέρωσή μας για τυχόν νεότερες εκδόσεις παλαιότερων εφαρμογών ή κυκλοφορία νέων προϊόντων. Συγκεντρώθηκαν γενικά στοιχεία και πληροφορίες για 82 λογισμικά πακέτα και εφαρμογές, που σχετίζονται καθ' οιονδήποτε τρόπο με τις διαδικασίες προβλέψεων.

- Αρχική αξιολόγηση των λογισμικών

Από το σύνολο των εφαρμογών που συγκεντρώθηκαν επιλέχθηκαν 42 λογισμικά πακέτα, για τα οποία υπήρχε η δυνατότητα συλλογής μεγάλου αριθμού στοιχείων από το διαδίκτυο. Επισημαίνεται, ότι πολλές εταιρείες δεν παρέχουν πληροφορίες για τα χαρακτηριστικά των προϊόντων τους, με αποτέλεσμα να καθίσταται αδύνατη η αξιολόγηση τους.

- Επιλογή συγκεκριμένων χαρακτηριστικών για την τελική αξιολόγηση.

Σε συνεργασία με τον επιβλέποντα την εργασία και το προσωπικό του εργαστηρίου Μονάδας Συστημάτων Πρόβλεψης και Προοπτικής καταρτίστηκε λίστα με συγκεκριμένα χαρακτηριστικά, προκειμένου να αποτελέσουν τη κοινή βάση αξιολόγησης των λογισμικών που συγκεντρώθηκαν. Η λίστα αυτή συντάχθηκε κατά τέτοιο τρόπο ώστε να επιτρέψει, αφενός μεν την ανάδειξη των επικρατέστερων λογισμικών (σύμφωνα πάντα με τις απαιτήσεις που καθορίστηκαν από το εργαστήριο) αφετέρου δε την τελική αξιολόγηση και σύγκριση αυτών (των επικρατέστερων λογισμικών που θα προκύψουν από την αναζήτηση), με τα χαρακτηριστικά του λογισμικού ΡΥΤΗΙΑ (λογισμικό που δημιουργήθηκε από τη Μονάδα Συστημάτων Πρόβλεψης και Προοπτικής).

- Αναζήτηση και συλλογή των συγκεκριμένων χαρακτηριστικών

Η συγκέντρωση και καταγραφή των χαρακτηριστικών έγινε με τους παρακάτω τρόπους:

1. Κατόπιν ευρύτερης αναζήτησεως στο διαδίκτυο ([6], [7], [8]).
2. Κατόπιν αναζήτησεως στις ιστοσελίδες των εταιρειών ([6]-[36]).
3. Κατόπιν επισκέψεως στα γραφεία αντιπροσωπειών των εταιρειών στην Ελλάδα και αναζήτησεως των χαρακτηριστικών στα τεχνικά εγχειρίδια (Technical Manual) των λογισμικών.
4. Με αποστολή της επιστολής του Παραρτήματος “Α”, επισυνάπτοντας τη λίστα των επιλεγμένων χαρακτηριστικών του Παραρτήματος “Β”, μέσω ηλεκτρονικού ταχυδρομείου στα τμήματα τεχνικής υποστήριξης και πωλήσεων των 26 εταιρειών που παράγουν τα 42 λογισμικά πακέτα. Στην επιστολή που εστάλη δεν ανταποκρίθηκαν όλες οι εταιρείες.

- Κατάρτιση Συγκριτικών Πινάκων

Με βάση όσα από τα χαρακτηριστικά καταφέραμε να συγκεντρώσουμε καταρτίστηκαν πίνακες, οι οποίοι μπορούν να ταξινομηθούν σε τέσσερις κύριες κατηγορίες :

1. Πληροφορίες για την εταιρεία παραγωγής του λογισμικού.
2. Απαιτήσεις σε υλικό/λογισμικό (Hardawre/Software).
3. Μέθοδοι προβλέψεων που χρησιμοποιεί κάθε λογισμικό.
4. Διάφορες δυνατότητες των λογισμικών.

Η μορφή των πινάκων βασίστηκε στους αντίστοιχους πίνακες σχετικής παλαιότερης δημοσίευσης [1].

- Ποσοτική και ποιοτική ανάλυση των συγκριτικών πινάκων και επιλογή των επικρατέστερων λογισμικών.

Λαμβάνοντας υπόψη τα συγκριτικά στοιχεία που παρέχουν οι πίνακες επιλέχθηκαν τελικά 10 λογισμικά με τα παρακάτω κριτήρια:

1. Ανταποκρίνονται στα περισσότερα από τα χαρακτηριστικά της λίστας, που καταρτίστηκε σε συνεργασία με τον επιβλέποντα την εργασία (Παράρτημα “Β”) καθώς και με βάση αντίστοιχες παλιότερες δημοσιεύσεις.
2. Εφαρμόζουν πέραν των παραδοσιακών μεθόδων πρόβλεψης - τις οποίες χρησιμοποιεί η πλειονότητα των λογισμικών - και τις προηγμένες τεχνικές προβλέψεων.
3. Το περιβάλλον εργασίας τους (user interface) ήταν το πλέον φιλικό και προσιτό προς το μέσο χρήστη παρόμοιων λογισμικών.
4. Οι εταιρείες που τα παράγουν , στις ιστοσελίδες τους, παραθέτουν πλήθος στοιχείων. Αυτό είχε σαν αποτέλεσμα την πιο εύκολη για μας συλλογή -ανά πάσα στιγμή- των απαιτούμενων για τους σκοπούς της εργασίας χαρακτηριστικών. Επιπλέον οι εταιρείες αυτές ανταποκρίθηκαν άμεσα και πρόθυμα στην επιστολή με την οποία τους ζητήσαμε την αποστολή στοιχείων για τα προϊόντα τους.

## **ΜΕΡΟΣ ΠΡΩΤΟ**

### **ΚΑΤΑΓΡΑΦΗ ΜΕΘΟΔΩΝ ΠΡΟΒΛΕΨΗΣ**



### ΚΕΦΑΛΑΙΟ 1ο : ΤΕΧΝΙΚΕΣ ΧΡΟΝΟΣΕΙΡΩΝ

#### 1.1 ΜΕΘΟΔΟΙ ΕΞΟΜΑΛΥΝΣΗΣ [1], [2], [6]

##### 1.1.1 Εισαγωγή

Η στατιστική θεωρία διδάσκει ότι ο αριθμητικός μέσος είναι ένας εκτιμητής που ελαχιστοποιεί το λάθος εκτίμησης της τιμής του μέσου του πληθυσμού. Ειδικότερα, ο μέσος είναι ένας αμερόληπτος εκτιμητής και είναι ο καλύτερος που υπάρχει στις περιπτώσεις που αναλύονται τα αποτελέσματα ενός απλού τυχαίου δείγματος<sup>1</sup>. Εν τούτοις, υπάρχουν περιπτώσεις όπου ο μέσος δεν είναι επιθυμητός για τη διενέργεια πρόβλεψης. Για παράδειγμα, αν κάποιος πρέπει να εκτιμήσει την έκτη τιμή μιας απλής σειράς 1,2,3,4,5, ο μέσος είναι ίσος με 3 και είναι, φυσικά, αναξιόπιστος εκτιμητής. Αν εφαρμοστεί ο μέσος σ' αυτή την περίπτωση σαν εκτιμητής της σειράς, είναι φανερό πως το αποτέλεσμα θα οδηγήσει σε συστηματικά σφάλματα πρόβλεψης και ακόμη ότι τα σφάλματα αυτά δεν κατανέμονται τυχαία γύρω από το μηδέν.

Εάν ο αριθμητικός μέσος πρόκειται να χρησιμοποιηθεί στη διενέργεια πρόβλεψης, τότε είναι κρίσιμο να καταλάβουμε τις συνθήκες που καθορίζουν τη χρησιμότητά του για το σκοπό αυτό. Οι συνθήκες αυτές προϋποθέτουν ότι τα δεδομένα κατανέμονται τυχαία και ακολουθούν σταθερή ανέλιξη ως προς το χρόνο. Λέγοντας στάσιμα δεδομένα, αναφερόμαστε στο πεδίο των χρονοσειρών που εμφανίζεται όμοιο με μια οριζόντια ευθεία γραμμή με τη μεταβλητή του χρόνου να βρίσκεται στον άξονα των  $X$ . Εάν τα δεδομένα δεν είναι τυχαία, τότε υπάρχει συστηματική συμπεριφορά ή τάση σε αυτά. Η ύπαρξη προτύπου συμπεριφοράς υποδηλώνει ότι η μέθοδος πρόβλεψης που θα χρησιμοποιηθεί μπορεί να προβλέψει τη συμπεριφορά αυτή με μεγαλύτερη ακρίβεια από τον μέσο. Στη δεύτερη περίπτωση, ο αριθμητικός μέσος μπορεί να δώσει ακριβή πρόβλεψη μόνον όταν η σειρά είναι στάσιμη. Για μια μη-στάσιμη σειρά, όπως η 1,2,3,...,9, ο αριθμητικός μέσος θα υπερεκτιμήσει σχετικά το πρώτο μισό και θα υποεκτιμήσει το δεύτερο. Επομένως, ο μέσος είναι επαρκής εκτιμητής για στατικές τυχαίες σειρές, δηλαδή για σειρές που δεν παρουσιάζουν ένα πρότυπο ή κάποιο στοιχείο περιοδικότητας. Καθώς θα επαυξάνονται οι γνώσεις μας γύρω από τις τεχνικές πρόβλεψης, θα χρησιμοποιούμε τη μέθοδο πρόβλεψης με τον αριθμητικό μέσο ως σημείο σύγκρισης όταν θα πρέπει να αναλυθούν άλλες μέθοδοι.

Ένας περιορισμός της χρησιμοποίησης του αριθμητικού μέσου ως εργαλείου πρόβλεψης είναι ότι απαιτείται μεγάλος αριθμός παρατηρήσεων ενός δείγματος, προκειμένου ο μέσος να χρησιμοποιείται ως εκτιμητής. Για τα δεδομένα των χρονοσειρών, το μέγεθος του δείγματος αυξάνεται συνεχώς με το χρόνο καθώς καταγράφονται νέες παρατηρήσεις. Οι παρατηρήσεις του νέου δείγματος συμπεριλαμβάνονται στον υπολογισμό του αριθμητικού μέσου, δημιουργώντας προβλήματα καταχώρισης, καθώς και υπολογιστικά όταν χρειάζονται προβλέψεις που αφορούν μεγάλο αριθμό μεγεθών. Τέλος, ο αριθμητικός μέσος μιας χρονοσειράς δεν συλλαμβάνει τις μεταβολές της μέσης τιμής διαχρονικά. Για παράδειγμα, αν υπάρξει μια μεταβολή στα δεδομένα, ο αριθμητικός μέσος δεν θα τη

λάβει υπόψη και η μέθοδος πρόβλεψης θα είναι σχετικά ανακριβής για μια αρκετά μεγάλη χρονική περίοδο. Μια προσέγγιση για την ελαχιστοποίηση της ανακρίβειας είναι το να κρατήσουμε τον ίδιο αριθμό παρατηρήσεων στο δείγμα κατά τον υπολογισμό του μέσου. Το αποτέλεσμα θα είναι μια διαδικασία υπολογισμού του μέσου των πιο πρόσφατων σταθερών παρατηρήσεων του δείγματος. Η διαδικασία αυτή ονομάζεται μέθοδος του κινητού μέσου όρου.

Σ' αυτό το κεφάλαιο εισάγονται δύο βασικοί τύποι μεθόδων πρόβλεψης: ο κινητός μέσος όρος και η εξομάλυνση. Ο κινητός μέσος όρος βασίζεται στον μέσο όρο σταθμισμένων παρατηρήσεων του παρελθόντος. Οι μέθοδοι εξομάλυνσης βασίζονται στη διαδικασία καθορισμού του μέσου όρου παρατηρήσεων μιας χρονοσειράς που διαμορφώθηκαν στο παρελθόν κατά φθίνοντα ή εκθετικό τρόπο.

Για να γίνει μια πρόβλεψη πρέπει να ακολουθήσουμε τα εξής βήματα:

1. Επιλέγουμε μια μέθοδο πρόβλεψης βασισμένη σε ήδη αποκτημένη γνώση γύρω από το παρατηρηθέν μοντέλο της χρονοσειράς.
2. Χρησιμοποιούμε τη μέθοδο πρόβλεψης για την απόκτηση εκτιμημένης τιμής των δεδομένων.
3. Υπολογίζουμε το σφάλμα πρόβλεψης.
4. Παίρνουμε απόφαση για την καταλληλότητα του μοντέλου που βασίζεται στη μέτρηση του σφάλματος πρόβλεψης.

Αρχίζουμε με τις απλούστερες μεθόδους του απλού κινητού μέσου όρου και της εκθετικής εξομάλυνσης μιας παραμέτρου. Η ARSES αριστοποιεί την παράμετρο της απλής εκθετικής εξομάλυνσης και των μοντέλων της εκθετικής εξομάλυνσης δευτέρου βαθμού εάν το πρότυπο στη χρονοσειρά είναι δευτέρου βαθμού.

Αν το πρότυπο των δεδομένων έχει τάση, η μέθοδος του γραμμικού κινητού μέσου όρου προσαρμόζει τον απλό κινητό μέσο στην τάση. Παρόμοια, η μέθοδος του Brown προσαρμόζει την απλή εκθετική εξομάλυνση στην τάση. Η μέθοδος Holt είναι μια άλλη μέθοδος εξομάλυνσης με την οποία επιτυγχάνεται προσαρμογή στην τάση. Βεβαίως αυτή αποτελεί μοντέλο με δύο παραμέτρους. Τέλος, η μέθοδος Winter είναι μέθοδος εξομάλυνσης με προσαρμογές στην τάση και την εποχιακότητα.

Περιλαμβάνει τρεις παραμέτρους

### 1.1.2 Απλοί Κινητοί Μέσοι

Η πρόβλεψη χρονοσειρών με τη μέθοδο των κινητών μέσων εμπεριέχει τον υπολογισμό του μέσου όρου του δείγματος παρατηρήσεων, καθώς και τη χρησιμοποίηση αυτού του μέσου σαν πρόβλεψη για την επόμενη περίοδο. Ο αριθμός των παρατηρήσεων του δείγματος που συμπεριλαμβάνονται στον υπολογισμό του μέσου προσδιορίζεται στην αρχή της διαδικασίας

πρόβλεψης. Χρησιμοποιούμε τον όρο *κινητός μέσος* γιατί, καθώς κάθε νέα παρατήρηση γίνεται διαθέσιμη στο δείγμα, υπολογίζεται ένας νέος μέσος απορρίπτοντας την παλαιότερη παρατήρηση του δείγματος από τον μέσο και συμπεριλαμβάνοντας τη νεότερη. Καθώς υπολογίζεται κάθε νέος μέσος, γίνεται πρόβλεψη για την επόμενη περίοδο. Συνεπώς, κάθε πρόβλεψη χρησιμοποιεί τον ίδιο αριθμό

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 1ο

παρατηρήσεων δείγματος από τη χρονοσειρά και περιλαμβάνει μόνο την πλέον πρόσφατη παρατήρηση.

### Παράδειγμα 1.1

Έχουν συλλεγεί δεδομένα που αφορούν βιομηχανικά και εμπορικά αποθέματα σε δολάριο σταθερής αξίας [εποχικά προσαρμοσμένο, σύνολο (\$ δις)] για τις ΗΠΑ για το 1981. Η πρόβλεψη έχει αποφασιστεί να γίνει με τη μέθοδο των κινητών μέσων. Εφόσον δεν είναι γνωστή η κατάλληλη μέθοδος για να χρησιμοποιηθεί, θα υπολογιστούν και οι τριμηνιαίοι και οι κινητοί μέσοι πέντε μηνών και θα χρησιμοποιηθούν στην πρόβλεψη.

Ο σκοπός είναι να αναλυθεί και να καθοριστεί η χρησιμότητα και οι περιορισμοί αυτής της μεθόδου για το σύνολο των δεδομένων που παρατίθενται στον παρακάτω Πίνακα.

**ΠΙΝΑΚΑΣ 1.1.** Πρόβλεψη με κινητούς μέσους όρους

(1)	(2)	(3)	(4)	(5)
<i>Μήνας</i>	<i>Χρονική περίοδος</i>	<i>Τιμές που παρατηρήθηκαν</i>	<i>Κινητός μ.ό. τριών μηνών</i>	<i>Κινητός μ.ό. πέντε μηνών</i>
Ιαν.	1	262.8		
Φεβρ.	2	262.9		
Μάρτ.	3	262.6		
Απρ.	4	263.2	262.8	
Μάιος	5	163.9	262.9	
Ιουν.	6	265.4	263.2	263.1
Ιούλ.	7	266.5	264.2	263.6
Αύγ.	8	267.1	265.3	264.3
Σεπτ.	9	268.5	266.3	265.2
Οκτ.	10	269.7	267.4	266.3
Νοέμ.	11	270.4	268.4	267.4
Δεκέμ.	12	269.4	269.5	268.4

Παρατηρηθείσες τιμές βιομηχανικών και εμπορικών αποθεμάτων σε σταθερές τιμές δολαρίου, τέλος του έτους (εποχικά προσαρμοσμένες) σύνολο (\$ δις) για το 1981.

Οι στήλες 2,3 και 4 δίνουν τα αποτελέσματα που προκύπτουν από τη χρησιμοποίηση κινητού μέσου όρου τριών μηνών. ΠΗΓΗ: Survey of Current Business, Μάρτιος 1982, σσ. 5-3.

Στον Πίνακα 1.1 δείχνεται η εφαρμογή της τεχνικής του κινητού μέσου όρου στη χρονοσειρά των βιομηχανικών και εμπορικών αποθεμάτων στις ΗΠΑ για το 1981 (εποχικά προσαρμοσμένες). Υπολογίζεται ο κινητός μέσος τριών και πέντε μηνών.

Ο κινητός μέσος όρος 3 μηνών για τον Απρίλιο είναι απλά το άθροισμα των παρατηρήσεων για

τους μήνες Ιανουάριο, Φεβρουάριο και Μάρτιο, δηλαδή  $262.8 + 262.9 + 262.6 = 788.3$  διαρούμενο με το τρία και ισούται με 262.8. Έτσι, ο μέσος αυτός όρος αποτελεί την πρόβλεψη για τον Απρίλιο.

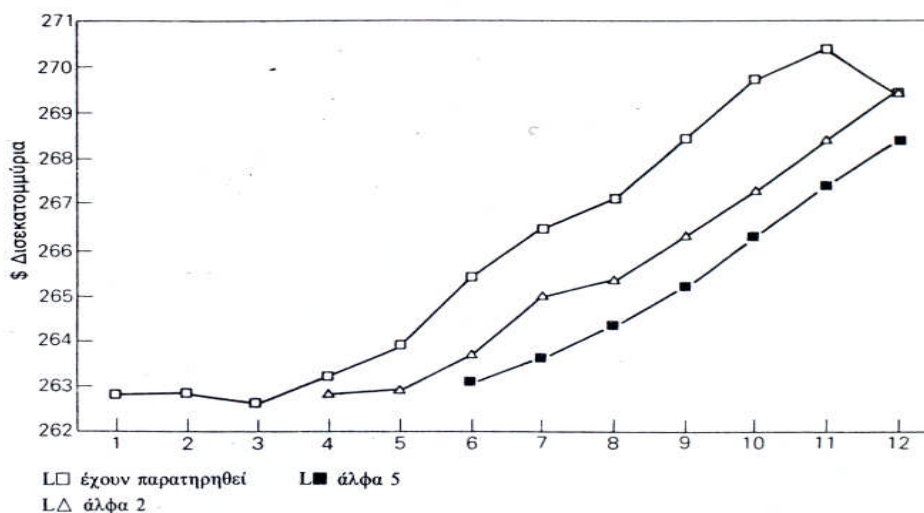
Για να κάνουμε την πρόβλεψη του Μαΐου, αφαιρούμε την τιμή για τον Ιανουάριο (262.8) και προσθέτουμε την πραγματική τιμή για τον Απρίλιο (263.2) στο σύνολο (788.3), το οποίο τώρα ισούται με 788.7. Διαιρούμε το νέο άθροισμα διά τρία και βρίσκουμε την πρόβλεψη για τον Μάιο, δηλαδή 262.9. Η μέθοδος αυτή συνεχίζεται και έτσι κάνουμε προβλέψεις για όλους τους μήνες του έτους. Ο τελευταίος αριθμός στη στήλη 4, δηλαδή 0269.5, είναι ο μέσος όρος για την περίοδο από Σεπτέμβριο έως Νοέμβριο. Τέλος, η πραγματική τιμή για τον Δεκέμβριο δεν θα έπρεπε να χρησιμοποιηθεί στους υπολογισμούς μέχρι να γίνει πρόβλεψη για τον Ιανουάριο του 1982.

Με όμοιο τρόπο, το αποτέλεσμα της εφαρμογής ενός κινητού μέσου όρου πέντε μηνών λαμβάνεται στη στήλη 5. Η πρόβλεψη για τον Ιούνιο είναι το άθροισμα των παρατηρηθεισών τιμών από τον Ιανουάριο μέχρι τον Μάιο διαιρεμένο διά 5. Η τελευταία καταχώριση της στήλης 5 είναι το άθροισμα των τιμών από τον Ιούλιο μέχρι τον Νοέμβριο. Τέλος, καθώς κάθε νέα παρατήρηση των βιομηχανικών και εμπορικών αποθεμάτων γίνεται διαθέσιμη, ο κινητός μέσος μπορεί εύκολα να επαναυπολογιστεί και να επικαιροποιηθεί.

Στο σχήμα 1.1 είναι αποτυπωμένες (α) οι τιμές για τα βιομηχανικά και εμπορικά αποθέματα προϊόντων, (β) ο μέσος όρος τριών μηνών (MA 3) και (γ) ο μέσος όρος πέντε μηνών (MA 5). Σημειώστε πως ο κινητός μέσος όρος 5 μηνών έχει μεγαλύτερο εξομαλυντικό αποτέλεσμα από τον τριμηνιαίο. Είναι Φανερό πως αν χρησιμοποιούνταν όλες οι παρατηρήσεις προκειμένου να γίνει η πρόβλεψη (ένας κινητός μέσος όρος 20 μηνών), ο αριθμητικός μέσος όλων αυτών των δεδομένων θα ήταν εκείνος που θα χρησιμοποιούνταν για να γίνει η πρόβλεψη. Εξάλλου, όταν χρησιμοποιούμε μόνον ένα μήνα για την πρόβλεψη, η πιο πρόσφατη τιμή είναι η προβλεφθείσα τιμή.

Γενικά, ο αριθμητικός μέσος είναι ο καλύτερος εκτιμητής όταν οι παρατηρήσεις του δείγματος είναι τυχαίες. Επομένως, ένας μεγάλος αριθμός περιόδων θα μπορούσε να χρησιμοποιηθεί όταν οι πραγματικές τιμές είναι τυχαίες. Το αποτέλεσμα της χρήσης μεγάλου αριθμού περιόδων είναι η χρησιμοποίηση μιας σχεδόν οριζόντιας γραμμής για να προβλέψουμε κάτι που ομαλοποιεί τις διακυμάνσεις που σχετίζονται με την τυχαιότητα των παρατηρήσεων.

**Σχήμα 1.1** Παρατηρηθείσες τιμές και προβλέψεις με τη μέθοδο της απλής εξομάλυνσης



Όμοια, ένας μικρός αριθμός περιόδων μπορεί να χρησιμοποιηθεί όταν στις παρατηρήσεις του δείγματος υπάρχει συγκεκριμένο πρότυπο συμπεριφοράς. Στις περιπτώσεις αυτές η χρησιμοποίηση λίγων μόνο περιόδων για την πρόβλεψη επιτρέπει στον κινητό μέσο όρο να προσεγγίσει και να ακολουθήσει το πρότυπο των παρατηρήσεων. Οι προβλέψεις αυτές θα ακολουθήσουν το πρότυπο στα πραγματικά δεδομένα επιβραδύνοντας τις κινήσεις κατά αρκετές περιόδους.

Η μέθοδος του κινητού μέσου όρου, σε αλγεβρική μορφή, διατυπώνεται ως εξής:

$$F_{t+1} = \frac{X_t + X_{t+1} + \dots + X_{t-n+1}}{n} = \frac{1}{n} \left( \sum_{i=t-n+1}^t X_i \right) \quad (1.1)$$

όπου  $t$  είναι η πιο πρόσφατη παρατήρηση και  $t+1$  είναι η επόμενη περίοδος. Η Φόρμουλα (ο τύπος) αυτή απαιτεί ο προβλέπων να διαθέτει τις τιμές των παρατηρήσεων του παρελθόντος.

Με την προσθήκη μιας νέας παρατήρησης και την εξάλειψη της παλαιότερης, μπορούμε να επαναδιατυπώσουμε τον τύπο ως εξής:

$$F_{t+1} = \frac{1}{n} \left( \sum_{i=t-n}^{t-1} X_i \right) + \frac{1}{n} (X_t - X_{t-n}) = F_t + \frac{X_t}{n} - \frac{X_{t-n}}{n} \quad (1.2)$$

Ο τύπος του κινητού μέσου όρου εκφράζει απλώς μια προσαρμογή στην πρόβλεψη ( $F$ ) που έγινε την προηγούμενη περίοδο. Είναι φανερό πως αν αυξηθεί το  $n$ , το εξομαλυντικό αποτέλεσμα είναι μεγαλύτερο γιατί γίνεται μια πολύ μικρότερη προσαρμογή για κάθε νέα χρονική περίοδο. Εάν τα δεδομένα είναι πολύ άστατα, το γεγονός αυτό αποτελεί ένα πολύ ελκυστικό χαρακτηριστικό.

Η μέθοδος του κινητού μέσου όρου είναι μια πολύ εύκολη διαδικασία πρόβλεψης, αφού οι απαιτήσεις σε στατιστικά στοιχεία είναι μικρές. Εν τούτοις, η τεχνική αυτή δεν χρησιμοποιείται ευρέως, δεδομένου ότι η προϋπόθεση να είναι τα δεδομένα στάσιμα είναι πολύ περιορισμένη. Πάντως, η μέθοδος του κινητού μέσου όρου είναι πιο αποτελεσματική από τη χρησιμοποίηση του αριθμητικού μέσου μιας ολόκληρης χρονοσειράς. Είναι πιο εύκαμπτη γιατί το  $n$  μπορεί να μεταβληθεί ώστε να ανταποκρίνεται στα παρατηρούμενα πρότυπα των δεδομένων.

### 1.1.3 Εκθετική Εξομάλυνση μιας Παραμέτρου

Υπάρχουν σοβαροί περιορισμοί στη χρήση του κινητού μέσου όρου. Πρώτον, για τους σχετικούς υπολογισμούς απαιτούνται οι παρελθούσες  $n$  παρατηρήσεις του δείγματος. Αν πρόκειται να προβλεφθεί ένας μεγάλος αριθμός μεγεθών (μεταβλητών), τα δεδομένα απαιτούν μεγάλο αποθηκευτικό χώρο. Δεύτερον, δίνεται ίση στάθμιση σε όλες τις παρατηρήσεις που αφορούν το παρελθόν, χωρίς

να δίνεται καμία στάθμιση στις παρατηρήσεις που είναι πριν από την περίοδο  $(t-n+1)$ . Οι πρόσφατες παρατηρήσεις, εν τούτοις, μπορεί να περιέχουν περισσότερες πληροφορίες από τις παλαιότερες, πράγμα που έχει σημασία για τις μελλοντικές προβλέψεις. Έτσι, κάποιος μπορεί να θέλει να αποδώσει μεγαλύτερη στάθμιση στην πρόβλεψη η οποία βασίζεται στις πιο πρόσφατες παρατηρήσεις παρά στις παλαιότερες.

Μια μέθοδος που απλοποιεί τους υπολογισμούς της πρόβλεψης και έχει μικρές απαιτήσεις σε δεδομένα, ονομάζεται *εκθετική εξομάλυνση*. Επίσης, η μέθοδος αυτή δίνει αυτοδιορθούμενες προβλέψεις αφού διαθέτει μηχανισμό ενσωματωμένων προσαρμογών, ο οποίος ρυθμίζει τις τιμές αλλάζοντας τις προς την αντίθετη κατεύθυνση από εκείνη που κινήθηκαν τα λάθη προηγούμενων περιόδων.

Για να αναπτύξουμε αυτή τη μέθοδο αρχίζουμε με την εξίσωση (1.2) και την ξαναγράφουμε ως εξής:

$$F_{t+1} = \frac{X_t}{n} - \frac{F_t}{n} + F_t \quad (1.3)$$

εάν κάνουμε την αντικατάσταση  $F_t = X_{t-n}$ . Παραπέρα, η εξίσωση αυτή μπορεί να γραφεί ως:

$$F_{t+1} = \frac{1}{n} X_t + \left(1 - \frac{1}{n}\right) F_t \quad (1.4)$$

Η εξίσωση (2.4) δίνει μια πρόβλεψη βασισμένη στη στάθμιση της πιο πρόσφατης παρατήρησης με συντελεστή  $1/n$  και της πιο πρόσφατης πρόβλεψης με συντελεστή στάθμισης  $[1-(1/n)]$ . Από τη στιγμή που ο αριθμός των περιόδων,  $n$ , είναι σταθερός, το κλάσμα  $1/n$  πρέπει να είναι μεγαλύτερο από το μηδέν και μικρότερο από τη μονάδα. Αν αντικαταστήσουμε το  $1/n$  με  $a$ , το βασικό μοντέλο γράφεται ως εξής:

$$F_{t+1} = aX_t + (1 - a)F_t \quad (1.5)$$

όπου  $t$  είναι η τρέχουσα περίοδος, τα  $F_{t+1}$  και  $F_t$  είναι τιμές πρόβλεψης για την επόμενη και την τρέχουσα περίοδο και  $X_t$  είναι η τιμή που παρατηρήθηκε την τρέχουσα περίοδο. Το  $a$  ονομάζεται *σταθερή εξομάλυνση* και παίρνει τιμές από 0 έως 1. Αφού η παραπάνω εξίσωση περιλαμβάνει μόνο μία σταθερά, το μοντέλο αυτό είναι μοντέλο εκθετικής εξομάλυνσης μιας παραμέτρου.

### Παράδειγμα 1.2

Δεδομένα βιομηχανικών αποθεμάτων σε σταθερές τιμές ανά μήνα [εποχικά προσαρμοσμένες, συνολικά (\$ δις)] για τις ΗΠΑ το 1981, έχουν συλλεγεί από το *Survey of Current Business*. Τα δεδομένα αυτά πρόκειται να χρησιμοποιηθούν για προβλέψεις με τη μέθοδο της εκθετικής εξομά-

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 1ο

λυνσης μιας παραμέτρου. Δεδομένου ότι δεν γνωρίζουμε την καλύτερη τιμή της ( $\alpha$ ) που θα χρησιμοποιηθεί στην πρόβλεψη, επιλέγονται τρεις τιμές (0.2, 0.5 και 0.8). Τα δεδομένα αυτά και οι προβλέψεις περιέχονται στον Πίνακα 1.2.

**ΠΙΝΑΚΑΣ 1.2 :** Πρόβλεψη με τη μέθοδο της εκθετικής εξομάλυνσης μιας παραμέτρου.

(1)		(2)	(3)	Εκθετικά εξομαλυμένα		
				(4)	(5)	(6)
Χρόνος	Μήνας	Χρονική περίοδος	Παρατηρηθείσες τιμές $X_t$	Άλφα 2 $\alpha=0.2$	Άλφα 5 $\alpha=0.5$	Άλφα 8 $\alpha=0.8$
1981	Ιαν.	1	145.5			
	Φεβρ.	2	145.8	145.5	145.5	145.5
	Μάρτ.	3	146.1	145.6	145.7	145.7
	Απρ.	4	146.4	145.7	145.9	146.0
	Μάιος	5	146.6	145.8	146.2	146.3
	Ιούν.	6	146.3	146.0	146.4	146.5
	Ιούλ.	7	146.8	146.1	146.2	146.3
	Αυγ.	8	146.9	146.2	146.5	146.7
	Σεπτ.	9	147.7	146.3	146.7	146.9
	Οκτ.	10	148.1	146.6	147.2	147.5
	Νοέμ.	11	148.1	146.9	147.7	148.0
	Δεκέμ.	12	147.5	147.1	147.9	148.1

Παρατηρηθείσες τιμές βιομηχανικών αποθεμάτων σε σταθερές τιμές δολαρίου ανά μήνα (εποχικά προσαρμοσμένες), σύνολο (\$ δις) για το 1981. ΠΗΓΗ: *Survey of Current Business*. Μάρτιος 1982. σσ. 5-3.

Για τον μήνα Μάρτιο 1981, υποθέτοντας ότι  $\alpha=0.2$ , η πρόβλεψη είναι

$$\begin{aligned} F &= \alpha(X_{\text{Φεβρ}}) + (1-\alpha)F_{\text{Φεβρ}} \\ &= 0.2(145.8) + (1-0.2)(145.5) \\ &= 145.6 \end{aligned}$$

Σημειωτέον επίσης ότι η αρχική πρόβλεψη (Φεβρουάριος) είναι η τιμή της παρατήρησης κατά την

προηγούμενη χρονική περίοδο (Ιανουάριος).

Η εκθετική εξομάλυνση μιας παραμέτρου είναι πολύ απλή μέθοδος, αφού μόνο μια τιμή, η πρόβλεψη της τελευταίας περιόδου, είναι αυτή που πρέπει να διασωθεί. Στην ουσία, ολόκληρη η χρονοσειρά εμπεριέχεται σ' αυτή την πρόβλεψη.

Εάν εκφράσουμε το  $F_t$  σε όρους της προηγούμενης παρατήρησης  $X_{t-1}$  και των τιμών της πρόβλεψης  $F_{t-1}$ , τότε το ισοδύναμο για την πρόβλεψη της επόμενης περιόδου γίνεται:

$$F_{t+1} = \alpha X_t + (1-\alpha)[\alpha X_{t-1} + (1-\alpha)F_{t-1}] \quad (1.6)$$

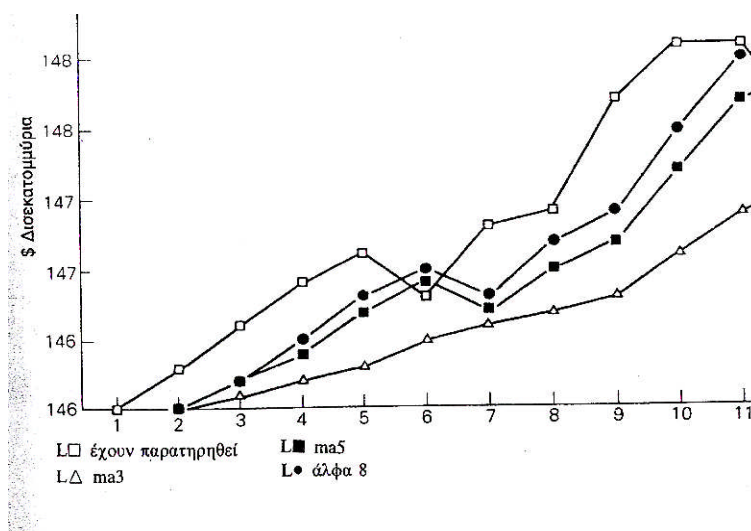
που απλοποιείται σε

$$F_{t+1} = \alpha X_t + \alpha(1-\alpha) X_{t-1} + (1-\alpha)^2 F_{t-1} \quad (1.7)$$

Η νέα αυτή εξίσωση είναι μοντέλο δευτεροβάθμιας εκθετικής εξομάλυνσης μιας παραμέτρου. Μπορούμε να συνεχίσουμε έτσι για έναν αριθμό προηγούμενων περιόδων, πράγμα που δείχνει ότι όλες οι προηγούμενες τιμές του  $X$  αντανακλώνται στην τρέχουσα πρόβλεψη. Έτσι, το όνομα αυτής της διαδικασίας προέρχεται από τις διαδοχικές σταθμίσεις  $\alpha, \alpha(1-\alpha), \alpha(1-\alpha)^2, \alpha(1-\alpha)^3, \dots$ , οι οποίες μειώνονται εκθετικά. Οι πιο πρόσφατες περίοδοι στη χρονοσειρά λαμβάνουν μεγαλύτερη στάθμιση στον υπολογισμό της πρόβλεψης. Προοδευτικά, αναγνωρίζεται μικρότερη προβλεπτική σημασία στις παλαιότερες τιμές του  $X$ . Πρακτικά, οι αρκετά παλιές τιμές της  $X$  εξαιρούνται. Η διαδικασία πρόβλεψης μπορεί να τροποποιηθεί οποιαδήποτε στιγμή με τη μεταβολή της τιμής της  $\alpha$ .

Την επίδραση της τιμής της  $\alpha$  στο μέγεθος της εξομάλυνσης που γίνεται μπορούμε να παρατηρήσουμε στο σχήμα 1.2.

**Σχήμα 1.2** Παρατηρηθείσες τιμές και προβλέψεις με τη μέθοδο της απλής εκθετικής εξομάλυνσης





## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 1ο

Μια μεγάλη τιμή της  $\alpha$  (0,8) αποδίδει μικρή εξομάλυνση στην πρόβλεψη. Ανάλογα, μια μικρή τιμή της  $\alpha$  (ίσως 0.20, 0.10 ή 0.05) αποφέρει σημαντική εξομάλυνση.

Μπορούμε να ξαναγράψουμε την εξίσωση (1.5) ως εξής:

$$F_{t+1} = \alpha F_t + \alpha (X_t - F_t) \quad (1.8)$$

και πιο απλά ως

$$F_{t+1} = F_t + \alpha e_t \quad (1.9)$$

όπου  $e_t$  το σφάλμα πρόβλεψης για την περίοδο  $t$ , είναι η πραγματική τιμή μείον την τιμή της πρόβλεψης. Επομένως, βλέπουμε ότι η πρόβλεψη που δίνεται από την εκθετική εξομάλυνση είναι η παλαιά πρόβλεψη συν μια προσαρμογή για το σφάλμα που έγινε στην τελευταία πρόβλεψη.

Όταν το  $\alpha$  βρίσκεται πλησίον του 1, η νέα πρόβλεψη περιέχει μια ουσιώδη προσαρμογή για το σφάλμα της προηγούμενης πρόβλεψης. Αντίθετα, εάν το  $\alpha$  βρίσκεται πολύ κοντά στο 0, η νέα πρόβλεψη θα περιέχει μικρή μόνο προσαρμογή για το σφάλμα. Επομένως, το αποτέλεσμα του μεγέθους του  $\alpha$  είναι όμοιο με τα αποτελέσματα των διαφόρων τιμών για τον αριθμό των παρατηρήσεων του δείγματος όταν υπολογίζουμε τον κινητό μέσο όρο. Τέλος, η εκθετική εξομάλυνση μιας παραμέτρου δίνει προβλέψεις που ακολουθούν το πρότυπο στα δεδομένα ενός δείγματος. Αυτό συμβαίνει γιατί η διαδικασία προσαρμόζει μόνο την επόμενη πρόβλεψη ως προς κάποιο ποσοστό τού πιο πρόσφατου σφάλματος πρόβλεψης, και δεν μπορεί να προβλέψει μεταβολές στην κατεύθυνση της χρονοσειράς.

### 1.1.4 Αξιολόγηση του Σφάλματος στην Πρόβλεψη

Το κύριο προτέρημα της μεθόδου εκθετικής εξομάλυνσης μιας παραμέτρου είναι οι περιορισμένες απαιτήσεις σε δεδομένα και η σχετική της απλότητα. Παρά την ελκτικότητα που εμφανίζει η μέθοδος της εκθετικής εξομάλυνσης παρατηρούνται σε αυτή ορισμένοι σημαντικοί περιορισμοί. Πρώτον, προβλέψεις που προέρχονται από εκθετική εξομάλυνση είναι ευαίσθητες στην εξειδίκευση της σταθεράς της εξομάλυνσης. Η επιλογή του  $\alpha$  βασίζεται σε μια διαδικασία δοκιμασίας και εκτίμησης (trial-and-error) του σφάλματος. Δεύτερον, οι τεχνικές της εκθετικής εξομάλυνσης οδηγούν σε προβλέψεις που καθυστερούν ως προς τα σημεία καμπής των πραγματικών δεδομένων της χρονοσειράς. Για να αξιολογήσουμε τη χρησιμότητα αυτής της τεχνικής πρέπει να εκτιμήσουμε το λάθος πρόβλεψης.

Στην ισότητα (1.9), το σφάλμα την περίοδο  $t$  ορίστηκε ως η πραγματική τιμή,  $X_t$  μείον την προβλεφθείσα τιμή,  $F_t$ :

$$e_t = X_t - F_t \quad (1.10)$$

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 1ο

Μια εξέταση του σφάλματος στην πρόβλεψη επιτρέπει στον προβλέποντα να αξιολογήσει αν η τεχνική ιχνηλατεί ικανοποιητικά τα πρότυπα που ακολουθούν οι παρατηρήσεις του δείγματος. Αξιολόγηση της αξιοπιστίας οποιασδήποτε τεχνικής απαιτεί τον προσδιορισμό των κριτηρίων. Όμως, δεν υπάρχει γενικά παραδεχτό άριστο μέτρο. Υπάρχει ένας αριθμός δεικτών που χρησιμοποιούνται για την αξιολόγηση της ακρίβειας της πρόβλεψης. Μια τεχνική που είναι γενικά παραδεκτή για την αξιολόγηση των μεθόδων της εκθετικής εξομάλυνσης (και πολλών άλλων) είναι το μέσο σφάλμα τετραγώνου (mean square error: MSE):

$$MSE = \frac{\sum (e_t)^2}{n} = \frac{\sum (X_t - F_t)^2}{n} \quad (1.11)$$

Το μέτρο αυτό ορίζει το σφάλμα ως το άθροισμα των τετραγώνων των σφαλμάτων της πρόβλεψης,  $e_t$ , διαιρεμένο με το μέγεθος του δείγματος, δηλαδή τον αριθμό των περιόδων της πρόβλεψης. Για να δείξουμε τον υπολογισμό του MSE θεωρούμε το επόμενο παράδειγμα.

### Παράδειγμα 1.3

Συνεχίζοντας με το παράδειγμα 1.2, ο προβλέπων (forecaster) καλείται να αξιολογήσει τις προβλέψεις που προετοιμάστηκαν με τη μέθοδο της εκθετικής εξομάλυνσης. Κάθε σύνολο προβλέψεων πρόκειται να αξιολογηθεί ώστε να ορίσει ποια τιμή της σταθεράς εξομάλυνσης,  $\alpha$ , είναι η καλύτερη, δηλαδή, ποια έχει το μικρότερο μετρήσιμο σφάλμα. Ο Πίνακας 1.3 επεξηγεί τον υπολογισμό του MSE για  $\alpha=0.2$

Χρησιμοποιώντας τον τύπο του MSE στα δεδομένα του Πίνακα 1.3, βρίσκουμε:

$$MSE = \frac{\sum (e_t)^2}{n} = \frac{8.35}{11} = 0.759$$

Επομένως, για την πρόβλεψη που βασίζεται σε τιμή σταθεράς εξομάλυνσης  $\alpha=0.2$ , το άθροισμα των τετραγώνων των σφαλμάτων πρόβλεψης διαιρούμενο διά του αριθμού των προβλέψεων (MSE) είναι 0,759.

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 1ο

**ΠΙΝΑΚΑΣ 1.3.** Υπολογισμός του MSE για μοντέλο εκθετικής εξομάλυνσης, όπου  $\alpha = 0,2$

Περίοδος	(1)	(2)	(3)	(4)
	Παρατηρηθείσα Τιμή $X_t$	Πρόβλεψη $F_t$	Σφάλμα $e_t$	Τετράγωνο σφάλματος $(e_t)^2$
1	145.6			
2	145.8	145.5	0.3	0.09
3	146.1	145.6	0.5	0.25
4	146.4	145.7	0.7	0.49
5	146.6	145.8	0.8	0.64
6	146.3	146.0	0.3	0.09
7	146.8	146.1	0.7	0.49
8	146.9	146.2	0.7	0.49
9	147.7	146.3	1.4	1.96
10	148.1	146.6	1.5	2.25
11	148.1	146.9	1.2	1.44
12	147.5	147.1	0.4	0.16
				$8.35 = S(e_t)^2$
				$\hat{\eta} = \sum (X_t - F_t)^2$

Τα δεδομένα έχουν παρθεί από τον Πίνακα 1.2

Αν υπολογίσουμε το MSE ως προς τις άλλες τιμές της σταθεράς, βρίσκουμε:

$$MSE=0.302 \text{ όταν } \alpha=0.5$$

και

$$MSE=0.200 \text{ όταν } \alpha=0.8$$

Το σχετικά ευρύ φάσμα των τιμών του MSE καταδεικνύει τον σημαντικό ρόλο της  $\alpha$  στη μέθοδο της εκθετικής εξομάλυνσης.

Μια δεύτερη τεχνική υπολογισμού του σφάλματος πρόβλεψης, είναι η *απόλυτη μέση απόκλιση* (Mean Absolute Deviation=MAD) και βασίζεται στην υπόθεση ότι η σοβαρότητα του σφάλματος σχετίζεται γραμμικά με το μέγεθός του. Αυτή ορίζεται από τη σχέση:

$$MAD = \frac{\sum |e|}{n} \quad (1.12)$$

Είναι το άθροισμα των απολύτων τιμών του σφάλματος (στήλη 3 του Πίνακα 1.3) διαιρούμενο διά του αριθμού των προβλέψεων. Για τα δεδομένα του Πίνακα 1.3, έχουμε:

$$MAD = \frac{\sum |e|}{n} = \frac{8.5}{11} = 0.77$$

Σαν μέτρο του σφάλματος πρόβλεψης είναι, θεωρητικά, λιγότερο αξιόπιστο από το MSE, εκτός από την περίπτωση της απλουστευτικής υπόθεσης που κάναμε παραπάνω. Πολλά προγράμματα υπολογιστών, όμως, δίνουν αυτόματα υπολογισμούς της MAD. Στις περισσότερες περιπτώσεις το MSE, είναι πιο χρήσιμο και στατιστικά είναι πιο θεμελιωμένος δείκτης για τη μέτρηση του σφάλματος πρόβλεψης.

Όπως σημειώνεται παραπάνω, μια πολύ χρήσιμη εφαρμογή για τη μέτρηση της ακριβείας της πρόβλεψης είναι ο προσδιορισμός της κατάλληλης τιμής της σταθεράς εξομάλυνσης  $\alpha$ , που χρησιμοποιείται στη μέθοδο εκθετικής εξομάλυνσης. Η σταθερά αυτή καθορίζει την έκταση που οι προβλέψεις του παρελθόντος επηρεάζουν την πρόβλεψη. Οι μικρές τιμές της  $\alpha$  αποσβήνουν αργά τις απομακρυσμένες παρατηρήσεις στη χρονοσειρά, με αποτέλεσμα την αργή ανταπόκριση στις μεταβολές των παραμέτρων που περιγράφουν το μέσο επίπεδο της χρονοσειράς. Μια πρόβλεψη που βασίζεται στην επιλογή μεγάλης τιμής της  $\alpha$ , από την άλλη πλευρά, αποσβήνει γρήγορα τις απομακρυσμένες παρατηρήσεις της χρονοσειράς με αποτέλεσμα μια πιο γρήγορη ανταπόκριση. Δυστυχώς, κάτι τέτοιο μπορεί να οδηγήσει τη διαδικασία της πρόβλεψης να ανταποκρίνεται σε ανώμαλες κινήσεις της χρονοσειράς. Αυτές οι ανώμαλες κινήσεις δεν αντανακλούν μεταβολές στις παραμέτρους που περιγράφουν τη χρονοσειρά. Επομένως, μια τέτοια κατάσταση δεν είναι καλύτερη από μια διαδικασία πρόβλεψης που αντιδρά πολύ αργά στις μεταβολές των παραμέτρων της χρονοσειράς.

Για να λύσουμε το πρόβλημα της επιλογής εφαρμόζουμε την ανάλυση ευαισθησίας στις ιστορικές χρονοσειρές, χρησιμοποιώντας διαφορετικές τιμές για τη σταθερά της εξομάλυνσης. Για κάθε τιμή πάνω από ένα εύρος τιμών, προετοιμάζεται ένα πρότυπο πρόβλεψης με τη μέθοδο της εκθετικής εξομάλυνσης και υπολογίζεται το κατάλληλο μέτρο της ακριβείας της πρόβλεψης. Στην πράξη, μελέτες που έγιναν, δείχνουν ότι οι τιμές από 0.05 έως 0.30 ταιριάζουν πολύ καλά στα πρότυπα της εκθετικής εξομάλυνσης. Τιμές της  $\alpha$  που είναι μεγαλύτερες από 0.30 συνήθως δείχνουν ότι κάποιο εναλλακτικό μοντέλο πρόβλεψης θα είναι περισσότερο κατάλληλο.

Για παράδειγμα, η ανάλυση ευαισθησίας εφαρμόστηκε σε ένα σύνολο ιστορικών δεδομένων προκειμένου να καθοριστεί η κατάλληλη σταθερά εξομάλυνσης. Τα αποτελέσματα δίνονται στον Πίνακα 1.4

**ΠΙΝΑΚΑΣ 1.4** Αποτελέσματα ανάλυσης ευαισθησίας

<b><math>\alpha</math></b>	<b>MSE</b>
0.02	39.3
0.04	39.2
0.06	38.6
0.08	38.4
0.10	38.2
0.12	38.1
0.14	37.9
0.16	37.4
0.18	37.4
0.20	37.2
0.22	37.6
0.24	37.8
0.26	38.9
0.28	38.1
0.30	38.3

Στον Πίνακα 1.4, το χαμηλότερο MSE προκύπτει για  $\alpha=0.20$ . Αυτό, θα μπορούσε να θεωρηθεί ως ένδειξη ότι η τιμή 0.20 για τη «σταθερά εξομάλυνσης» είναι η «καλύτερη» για τις υπό εξέταση ιστορικές χρονοσειρές. Εάν η «καλύτερη» τιμή της σταθεράς εξομάλυνσης ήταν μεγαλύτερη του 0,3, αυτό ίσως έδειχνε ότι είναι παρούσα η συμπεριφορά συγκεκριμένου τύπου χρονοσειράς, όπως της εποχικής. Επομένως, τα μοντέλα πρόβλεψης που αναγνωρίζουν αυτές τις άλλες μορφές συμπεριφοράς, θα μπορούσαν να δώσουν μια καλύτερη πρόβλεψη.

Συνοψίζοντας, η διαδικασία της πρόβλεψης ολοκληρώνεται πλήρως με τη μέτρηση της ακριβείας της πρόβλεψης. Όπως δείχνεται παραπάνω, η διαδικασία απαιτεί από τον προβλέποντα να δημιουργήσει σειρά προβλέψεων μέσα σε ένα διάστημα και να υπολογίσει το MSE για την κάθε πρόβλεψη. Η σταθερά εξομάλυνσης που απορρέει από το τελευταίο MSE είναι η τιμή που θα έπρεπε να είχε επιλεγεί για την πρόβλεψη. Καθώς παίρνονται όλο και νέα δεδομένα, η διαδικασία επαναλαμβάνεται για να δημιουργηθούν καινούριες «καλύτερες» σταθερές εξομάλυνσης. Εάν η «καλύτερη» σταθερά εξομάλυνσης είναι μεγαλύτερη από 0.3, τότε προτείνεται να ακολουθηθεί άλλο μοντέλο πρόβλεψης.

1.1.5 Γραμμικός Κινητός Μέσος

Εφαρμόζοντας τη μέθοδο των κινητών μέσων σε ένα σύνολο παρατηρήσεων δείγματος (χρονοσειρά) που ακολουθεί κάποιο πρότυπο, παίρνουμε προβλέψεις οι οποίες υποεκτιμούν τις εκ παρατηρήσεως τιμές. Η επανεξέταση του παραδείγματος 1.1 και του σχήματος 1.1 δείχνει πως μια χρονοσειρά που ακολουθεί ένα συνεπές ανοδικό πρότυπο οδηγεί σε προβλέψεις με τη μέθοδο των κινητών μέσων όρων που υποεκτιμούν την παρατηρηθείσα τιμή. Στο παράδειγμα 1.1 υπάρχει τυχαίος παράγων. Εν τούτοις, για να παρατηρήσουμε καλύτερα το πώς ένα πρότυπο γραμμικής τάσης επηρεάζει τα αποτελέσματα, θεωρούμε το επόμενο παράδειγμα όπου δεν υπάρχει ο τυχαίος παράγων.

Παράδειγμα 1.4

Μια υποθετική σειρά, η οποία ακολουθεί ανοδικό εξελικτικό πρότυπο, πρόκειται να χρησιμοποιηθεί για τη διενέργεια προβλέψεων με τη μέθοδο των κινητών μέσων ( $n=3$ ) (βλέπε Πίνακα 1.5).

Στο παραπάνω παράδειγμα, η πρόβλεψη υστερεί της πραγματοποίησης κατά δύο περιόδους. Το σφάλμα για κάθε περίοδο για την οποία γίνεται πρόβλεψη είναι το ίδιο, 5. Μια τέλεια γραμμική τάση στη σειρά έχει σαν αποτέλεσμα ένα συστηματικό σφάλμα που δεν μπορούμε ν' αποφύγουμε, εκτός αν εφαρμόσουμε κάποια άλλη εναλλακτική τεχνική. Με σκοπό να αποφύγουν αυτό το σφάλμα, οι ενεργούντες προβλέψεις πολλές φορές εφαρμόζουν τη μέθοδο του *γραμμικού κινητού μέσου*. Η μέθοδος αυτή απαιτεί τον υπολογισμένους δεύτερου κινητού μέσου,

**ΠΙΝΑΚΑΣ 1.5** Διενέργεια προβλέψεων με τη μέθοδο των κινητών μέσων ( $n=3$ )

Περίοδος	Τιμή	Πρόβλεψη	Σφάλμα
1	2.5		
2	5.0		
3	7.5		
4	10.0	5.0	5.0
5	12.5	7.5	5.0
6	15.0	10.0	5.0
7	17.5	12.5	5.0
8	20.0	15.0	5.0
9	22.5	17.5	5.0
10	25.0	20.0	5.0

δηλαδή έναν κινητό μέσο του κινητού μέσου των τιμών που παρατηρήθηκαν. Αυτοί οι διπλοί κινητοί μέσοι (double moving averages=DMA) υστερούν σε σύγκριση με τον απλό κινητό μέσο (SMA) κατά το ίδιο ποσό που ο απλός κινητός μέσος υστερεί σε σχέση με τις παρατηρηθείσες τιμές. Έτσι, οι διαφορές ανάμεσα στις παρατηρηθείσες τιμές και στον απλό κινητό μέσο θα είναι ίδιες με τις διαφορές ανάμεσα στον απλό και τον διπλό κινητό μέσο. Για να προβλέψουμε χωρίς την παρουσία

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 1ο

συστηματικού σφάλματος, εφαρμόζουμε τους ακόλουθους τύπους:

$$(SMA) \quad S_t = \sum_{i=t}^{t-n+1} \frac{X_i}{n} \quad (1.13)$$

$$DMA \quad S'_t = \sum_{i=t}^{t-n+1} \frac{S_i}{n} \quad (1.14)$$

$$(\text{Συνολική Πρόβλεψη}) \quad \alpha_t = S_t + (S_t - S'_t) = 2S_t - S'_t \quad (1.15)$$

και

$$(\text{με τάση στις παρατηρήσεις}) \quad b_t = \frac{2}{n-1}(S_t - S'_t) \quad (1.16)$$

Ο Πίνακας 1.6 περιλαμβάνει την εφαρμογή της διαδικασίας του γραμμικού κινητού μέσου στα υποθετικά δεδομένα του παραδείγματος 1.4. Σημειωτέον ότι στην τρίτη στήλη βρίσκεται το αποτέλεσμα της εφαρμογής του απλού κινητού μέσου ( $n=4$ ) στην παρατηρηθείσα χρονοσειρά. Η στήλη 4 αποτελεί το σφάλμα που σχετίζεται με αυτή την αρχική πρόβλεψη.

Η εξίσωση (1.14) είναι παρόμοια στη μορφή με τον υπολογισμό του απλού κινητού μέσου, αλλά αποτελεί την εξίσωση του διπλού κινητού μέσου και το αποτέλεσμα βρίσκεται στη στήλη 5. Η εξίσωση (1.15) επιτρέπει την πρόβλεψη μέσω της πρόσθεσης στον απλό κινητό μέσο της διαφοράς ανάμεσα στον απλό και τον διπλό κινητό μέσο. Η εξίσωση (1.16) δίνει έναν απευθείας υπολογισμό της τάσης στα δεδομένα. Οι κινητοί μέσοι αποτελούν τον μέσο των  $n$  παρατηρήσεων που βρίσκονται συγκεντρωμένες στο μέσον των  $n$  παρατηρήσεων  $(n-1)/2$ . Εάν ο απλός κινητός μέσος έχει χρησιμοποιηθεί για την πρόβλεψη, θα ακολουθούσε την παρατηρηθείσα τιμή με  $(n-1)/2b$ , όπου  $b$  είναι η μεταβολή της τάσης ή του προτύπου για κάθε περίοδο. Όμως,

$$\frac{n-1}{2}b_t = S_t - S'_t$$

δηλαδή η διαφορά ανάμεσα στους δύο κινητούς μέσους. Μπορούμε να ξαναγράψουμε τον τύπο για  $20b_t$  με τη μορφή της εξίσωσης (1.16).

**ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 1ο**

**ΠΙΝΑΚΑΣ 1.6.** Πρόβλεψη με γραμμικούς κινητούς μέσους - χρονοσειράς με τάση χωρίς καμιά τυχαία συνιστώσα

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Περίοδος	Απαντηθείσα Τιμή	SMA n=3	Σφάλμα	DMA n=3	Συνολική		
					Σφάλμα	Πρόβλεψη η	Σφάλμα α
1	2.5						
2	5.0						
3	7.5						
4	10.0	5.0	5.0				
5	12.5	7.5	5.0				
6	15.0	10.0	5.0				
7	17.5	12.5	5.0	7.5	5.0	17.5	0
8	20.0	15.0	5.0	10.0	5.0	20.0	0
9	22.5	17.5	5.0	12.5	5.0	22.5	0
10	25.0	20.0	5.0	15.0	5.0	25.0	0
Κλειδί τύπων	$X_t$	$S_t$	$e_t$	$S'_t$	$e'_t$	$a_t$	

Στήλη 4 = στήλη 2 - στήλη 3

Στήλη 6 = στήλη 3 - στήλη 5

Στήλη 7 = στήλη 3 + στήλη 6

Στήλη 8 = στήλη 2 - στήλη 7

Τέλος, ο τύπος για την πρόβλεψη της τιμής οποιασδήποτε περιόδου, είναι:

$$F_{t+m} = a_t + b_t \times m \quad (1.17)$$

όπου m είναι ο αριθμός των μελλοντικών περιόδων που πρέπει να προβλεφθούν.



1.1.6 Εφαρμογή της Εκθετικής Εξομάλυνσης σε Εποχικά Δεδομένα

Οι οικονομικοί στατιστικοί έχουν καταλήξει ότι τα δεδομένα χρονοσειρών εμφανίζουν κανονικά επαναλαμβανόμενα πρότυπα κατά τη διάρκεια ενός ημερολογιακού έτους. Αυτό το εποχικό πρότυπο επαναλαμβάνεται με συνέπεια από έτος σε έτος.

Ο Winters (1960) ανέπτυξε μια μέθοδο για την προσαρμογή της εποχικής ή περιοδικής κίνησης μέσα στο πλαίσιο της γραμμικής εκθετικής εξομάλυνσης. Επομένως, η διαδικασία του Winters μπορεί να εφαρμοστεί για προβλέψεις με βάση μια χρονοσειρά που εμφανίζει και τάση και εποχικό πρότυπο. Η επέκταση συνίσταται στην ύπαρξη επιπλέον εξίσωσης για τον υπολογισμό της εποχικής συνιστώσας της χρονοσειράς. Προκειμένου να κάνουμε πρόβλεψη με αυτό το μοντέλο χρειαζόμαστε τέσσερις εξισώσεις:

1. Κάνουμε επίκαιρη (we update) την εκθετικά εξομαλυνθείσα σειρά

$$S_t = a \frac{X_t}{I_{t-L}} + (1-a)(S_{t-1} + b_{t-1}) \quad (1.18)$$

όπου L είναι το εύρος της εποχιακότητας, δηλαδή ο αριθμός των τριμήνων ή μηνών του έτους.

2. Κάνουμε επίκαιρη την εκτίμηση της εποχιακότητας

$$I_t = \beta \frac{X_t}{S_t} + (1-\beta)I_{t-L} \quad (1.19)$$

όπου I είναι ο παράγοντας προσαρμογής της εποχιακότητας.

3. Κάνουμε επίκαιρο τον εκτιμητή της τάσης

$$b_t = \nu(S_t - S_{t-1}) + (1-\nu)b_{t-1} \quad (1.20)$$

4. Προβλέπουμε m περιόδους στο μέλλον

$$F_{t+m} = (S_t + b_t \times m)I_{t-L+m} \quad (1.21)$$

Με την εξίσωση (1.18) γίνονται επίκαιρες οι εξομαλυνθείσες τιμές της σειράς. Στην εξίσωση (1.18), το  $X_t$  διαιρείται δια του  $I_{t-L}$  που προσαρμόζει τις αρχικές παρατηρήσεις,  $X_t$  για εποχιακότητα και στη συνέχεια αναιρεί τις επιδράσεις της εποχιακότητας όσο καλύτερα αυτές μπορεί να μετρηθούν από τη χρονολογική σειρά.

Η εξίσωση (1.19) δίνει την εκτίμηση της εποχικής συνιστώσας,  $X_t / S_t$  πολλαπλασιασμένη επί τη σταθερά  $\beta$  συν την παλαιά εποχική εκτίμηση,  $I_{t-L}$  πολλαπλασιασμένη επί  $(1-\beta)$ . Επομένως, η επικαιροποίηση (updating) των εποχικών εκτιμήσεων είναι από μόνη της μια διαδικασία εκθετικής εξομάλυνσης. Επίσης, ο  $X_t$  διαιρείται διά  $S_t$  προκειμένου να εκφραστεί

η τιμή ως δείκτης παρά ως απόλυτο μέγεθος. Αυτό επιτρέπει την εύρεση του μέσου όρου των νέων εποχικών εκτιμήσεων με βάση τον εποχικό δείκτη της προηγούμενης περιόδου.

Η εξίσωση (1.20) εκφράζει τη σύγχρονη τιμή της συνιστώσας της τάσης που επιτυγχάνεται με τη

συνηθισμένη διαδικασία εκθετικής εξομάλυνσης.

Τέλος, μετά την εξίσωση (1.18), λαμβάνουμε την (1.19) για τις μελλοντικές περιόδους από την εξίσωση (1.21). Η διαφορά είναι ότι αυτή η εκτίμηση για τη μελλοντική περίοδο,  $t+m$ , πολλαπλασιάζεται επί  $I_{t-L+m}$ . Αυτός είναι ο τελικός διαθέσιμος εποχικός δείκτης και αποτελεί την προσαρμογή της πρόβλεψης για εποχιακότητα.

Η ευελιξία και η δύναμη της εποχικής εκθετικής εξομάλυνσης του Winter μπορεί να εξηγηθεί καλύτερα με μια εφαρμογή σε μια χρονοσειρά που έχει και εποχική συνιστώσα και συνιστώσα τάσης.

### Παράδειγμα 1.5

Τα δεδομένα που αφορούν δαπάνες για νέες βιομηχανικές εγκαταστάσεις και εξοπλισμό στο σύνολο του μη αγροτικού τομέα των ΗΠΑ (\$ δισεκατομμύρια χωρίς εξομάλυνση) εμφανίζουν τόσο εποχική συμπεριφορά όσο και κάποια τάση.

Τα δεδομένα αυτά παρατίθενται στον Πίνακα 1.7 για την περίοδο που καλύπτει το πρώτο τρίμηνο του 1979 μέχρι το δεύτερο τρίμηνο του 1982. Σημειώστε ότι το πρώτο τρίμηνο κάθε έτους έχει τη μικρότερη τιμή, ενώ η περίοδος 4 έχει τη μεγαλύτερη τιμή για το ίδιο έτος.

Για να φέρουμε σε πέρας αυτή την τεχνική εκθετικής εξομάλυνσης είναι απαραίτητο να προσδιοριστεί η φάση αρχικής εκκίνησης (initial phase). Στη φάση αυτή χρησιμοποιούνται δεδομένα του παρελθόντος για να αρχίσουμε τη διαδικασία της πρόβλεψης. Μετά τον προσδιορισμό της φάσης εκκίνησης, αρχίζει η φάση «επικαιροποίησης» όπου οι προβλέψεις αναπροσαρμόζονται, καθώς έχουμε στη διάθεσή μας νέες παρατηρήσεις.

Ο σκοπός του προσδιορισμού της φάσης εκκίνησης είναι να γίνει έναρξη του συστήματος πρόβλεψης με αξιόπιστες εκτιμήσεις των  $S_t$ ,  $I_t$  και  $b_t$ . Παραπέρα, η μέθοδος προσδιορισμού της φάσης εκκίνησης που παρουσιάζεται εδώ αποτελεί μια διαφοροποίηση της μεθόδου του Winters (1960).

1. Πρώτα, υπολογίζουμε την τάση εκτιμώντας τον μέσο της χρονοσειράς

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 1ο

**ΠΙΝΑΚΑΣ 1.7.** Εφαρμογή της εκθετικής εξομάλυνσης τριπλής παραμέτρου στα δεδομένα των δαπανών για νέες βιομηχανικές εγκαταστάσεις και εξοπλισμό (\$ δισεκατομμύρια) (Μοντέλο Winters)

		(1)	(2)	(3)	(4)	(5)	(6)
Έτος	Τρίμηνο	Περίοδος	Παρατηρηθείσες τιμές $X_t$	Απλή Εξομάλ. $S_t$	Εποχικοί δείκτες $I_t$	Ροπή $b_t$	Πρόβλεψη $F_{t+m}$
		-3			0.905		
		-2			1.003		
		-1			0.998		
		0		63.2	1.094	1.6	
1979	1	1	57.3	64.6	0.894	1.6	
	2	2	66.8	66.3	1.006	1.6	66.4
	3	3	68.4	66.8	1.025	1.4	67.8
	4	4	78.0	68.8	1.118	1.5	74.6
1980	1	5	65.2	70.8	0.910	1.6	62.8
	2	6	74.0	72.2	1.017	1.6	72.8
	3	7	74.1	73.5	1.015	1.5	75.6
	4	8	82.3	74.7	1.108	1.3	83.9
1981	1	9	69.8	77.2	0.906	1.5	69.2
	2	10	79.6	78.6	1.014	1.5	80.0
	3	11	81.8	80.2	1.018	1.5	81.3
	4	12	90.4	81.7	1.107	1.5	90.5
1982	1	13	73.8				75.4

Τα δεδομένα αφορούν έξοδα βιομηχανικών εγκαταστάσεων και εξοπλισμού, στο σύνολο του μη-αγροτικού τομέα των ΗΠΑ (\$ δισεκατομμύρια, χωρίς εξομάλυνση). Οι σταθερές εξομάλυνσης είναι  $\alpha=0.2$ ,  $\beta=0.6$  και  $\nu=0.2$  και  $MSE=2.29$ . ΠΗΓΗ: *Survey of Current Business*, διάφορα θέματα, σσ. S-II για το πρώτο και το τελευταίο έτος της παρατηρηθείσας τιμής,  $X_t$  πρώτο έτος:

$$\frac{1}{4}(X_1 + X_2 + X_3 + X_4) = \frac{1}{4}(57.3 + 66.8 + 68.4 + 78.0) = 67.6$$

τελευταίο έτος

$$\frac{1}{4}(X_9 + X_{10} + X_{11} + X_{12}) = \frac{1}{4}(69.8 + 79.6 + 81.8 + 90.4) = 80.4$$

Για να εκτιμήσουμε την αρχική κλίση της εξίσωσης τάσης,  $b_0$ , υπολογίζουμε

$$b_0 = \frac{80.4 - 67.6}{8} = 1.6$$

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 1ο

---

2. Ας σημειωθεί ότι η μεταβολή από τον μέσον του πρώτου έτους (1979) έως τον μέσον του τρίτου έτους (1981) συμβαίνει για 8 τρίμηνα. Ο μέσος του πρώτου έτους είναι το τρίμηνο 2.5, δηλαδή  $(1+2+3+4)/4=2.5$ .

Επομένως, η απάλειψη της εποχιακότητας σε χρόνο  $t=0$  υπολογίζεται από

$$S_0 = \frac{X_1 + X_2 + X_3 + X_4}{4} - 2.5b_0 = 67.6 - 2.5 \times 1.6 = 63.6$$

Επομένως, η αρχική γραμμική τάσης είναι

$$\hat{S}_t = S_0 + b_0 \times t = 63.6 + 1.6t \quad (1.22)$$

3. Για να υπολογίσουμε τους αρχικούς εποχικούς συντελεστές σχηματίζουμε τον Πίνακα 1.8. Στη στήλη 1 παρατίθενται τα αρχικά δεδομένα,  $X_t$ . Στη στήλη 2 υπάρχουν τα αποτελέσματα της χρήσης της εξίσωσης (1.22) για την εύρεση των τιμών τάσης για κάθε περίοδο. Οι τιμές τάσης συνάγονται κατά πανομοιότυπο τρόπο. Για παράδειγμα, θεωρούμε την περίοδο 3:

$$\hat{S}_3 = 63.6 + 1.6 \times 3 = 68.4$$

Οι εποχικοί συντελεστές καθορίζονται ως εξής: πρώτα διαιρούμε τις παρατηρηθείσες τιμές,  $X_t$ , με τις υπολογισθείσες τιμές τάσης,  $S_t$ . Οι σχετικοί λόγοι παρουσιάζονται στη στήλη 3. Το επόμενο βήμα είναι να βρούμε τον μέσο όρο αυτών των λόγων για κάθε τρίμηνο, προκειμένου να καθορίσουμε τους αρχικούς συντελεστές. Για αυτό το παράδειγμα, ο αρχικός εποχικός συντελεστής για το τρίμηνο I είναι:

$$\frac{0.912 + 0.911 + 0.895}{3} = 0.906$$

Οι συντελεστές αυτοί παρουσιάζονται στη στήλη 4.

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 1ο

**ΠΙΝΑΚΑΣ 1.8.** Εύρεση αρχικών εκτιμήσεων των εποχικών συντελεστών ως εισροές στο μοντέλο Winters

	(1)	(2)	(3)	(4)	(5)
Περίοδος	$X_t$	$S_t$	$X_t/S_t$	Αρχικοί Εποχικοί Συντελεστές	Αναθεωρημένοι Εποχικοί Συντελεστές
0		63.6			
1	57.3	65.2	0.912	0.906	0.905
2	66.8	66.8	1.000	1.004	1.003
3	68.4	68.4	1.000	0.999	0.998
4	78.0	70.0	1.114	1.095	1.094
5	65.2	71.6	0.911		
6	74.0	73.2	1.011		
7	74.1	74.8	0.991		
8	82.3	76.4	1.078		
9	69.8	78.0	0.895		
10	79.6	79.6	1.000		
11	81.8	81.2	1.007		
12	90.4	82.8	1.092		

4. Τέλος, το άθροισμα των εποχικών συντελεστών πρέπει να είναι 4 (4 φορές το 1,0). Εφόσον οι συντελεστές στη στήλη 4 έχουν άθροισμα 4,004, αναθεωρούμε τους αρχικούς συντελεστές με βάση το λόγο 4/4,004. Η αναλογία αυτή (0,999001) πολλαπλασιάζεται επί κάθε τριμηνιαίο εποχικό συντελεστή ώστε να έχουμε τους αναθεωρημένους αρχικούς εποχικούς συντελεστές της στήλης 5. Αυτοί οι τελευταίοι εποχικοί συντελεστές θα είναι μέρος της εισροής στο μοντέλο πρόβλεψης. Από τη στιγμή που έχουμε τις αρχικές εκτιμήσεις της τάσης και της εποχιακότητας, το μοντέλο Winters τους ενσωματώνει στη διαδικασία εκθετικής εξομάλυνσης των εξισώσεων (1.18), (1.19), (1.20) και (1.21). Επομένως, οι αρχικές εκτιμήσεις της τάσης, η παρατηρηθείσα τιμή σε χρόνο  $t=0$  και οι εποχικοί συντελεστές μπορούν συνεχώς να «επικαιροποιούνται». Επιστρέφοντας στον Πίνακα 1.7, μπορούμε να εξηγήσουμε τη διαδικασία «επικαιροποίησης».

Για την περίοδο 1, η επικαιροποίηση γίνεται ως εξής:

$$S_1 = a \frac{X_1}{I_{-3}} + (1-a)(S_0 + b_0) \quad (\text{στήλη 3})$$

Με  $a=0,2$  έχουμε

$$S_1 = 0.2 \frac{57.3}{0.905} + 0.8(63.2 + 1.6) = 64.6$$

$$I_1 = \beta \frac{X_1}{S_1} + (1-\beta)I_0 \quad (\text{στήλη 4})$$

$$\mu\epsilon \beta = 0.6$$

$$I_1 = 0.6 \frac{57.3}{64.6} + 0.4 \times 0.905 = 0.894$$

$$b_1 = v(S_1 - S_0) + (1-v) \times b_0$$

$$\mu\epsilon v = 0.2$$

$$b_1 = 0.2(64.6 - 63.2) + 0.8 \times 1.6 = 1.6$$

Τέλος για προβλέψεις μιας περιόδου μπροστά

$$F_2 = a[S_1 + b(1)] \times I_2 = (64.6 + 1.6)(1.003) = 66.4$$

Η διαδικασία επικαιροποίησης συνεχίζεται μέχρι να εξαντληθούν όλα τα δεδομένα. Καθώς συγκεντρώνονται νέες παρατηρήσεις, συνεχίζεται και η διαδικασία επικαιροποίησης. Το MSE ήταν 2.29, το οποίο είναι μικρότερο από τα αποτελέσματα των μοντέλων μη-εποχικής εξομάλυνσης.

Οι παρακάτω είναι δύο βασικοί περιορισμοί που σχετίζονται με αυτή την τρι-παραμετρική διαδικασία εκθετικής εξομάλυνσης. Πρώτα, όπως γίνεται και με οποιαδήποτε άλλη διαδικασία εκθετικής εξομάλυνσης, η επιλογή των σταθερών εξομάλυνσης ( $\alpha$ ,  $\beta$  και  $v$ ) γίνεται μέσω μιας διαδικασίας προσπάθειας-σφάλματος (trial-and-error). Από τη στιγμή που υπάρχουν τρεις παράμετροι, η διαδικασία αυτή μπορεί να αποδειχτεί πολύξοδη. Οι σταθερές που χρησιμοποιούνται στον Πίνακα 1.7 είναι αυτές που συνιστώνται από τον Winters, αν και δεν είναι αποδεκτές από όλους. Δεύτερον, υπάρχει μια κυκλική συνιστώσα σε πολλές παρατηρηθείσες χρονοσειρές. Είναι εξαιρετικά δύσκολο να εισάγουμε έναν κυκλικό συντελεστή στο μοντέλο Winters. Επομένως, αυτή η πηγή διακύμανσης δεν χρησιμοποιείται παρά μόνο για να δηλώσει το επίπεδο  $s$  και το ποσοστό μεταβολής  $b$ .

**1.2.1 Εισαγωγή**

Στην Παράγραφο 1.2 χρησιμοποιήσαμε απλές μεθόδους προεκβολής προκειμένου να προβλέψουμε μελλοντικές τιμές μιας χρονοσειράς από παρατηρηθείσες τιμές. Η χρονική περίοδος εκφραζόταν σε έτη, τρίμηνα ή μήνες. Οι χρονοσειρές μελετώνται κάτω από την υπόθεση ότι με βάση τη συμπεριφορά στο παρελθόν μπορεί να προβλεφθεί η μελλοντική πορεία της διακύμανσης της παρατηρούμενης μεταβλητής. Μολονότι η πρόβλεψη μελλοντικών καταστάσεων δεν γίνεται χωρίς σφάλμα, η ανάλυση και ο διαχωρισμός (decomposition) των χρονοσειρών έχει αξία στη διαδικασία της πρόβλεψης γιατί έτσι μπορεί να μειωθεί το σφάλμα πρόβλεψης.

Επισημαίνεται ότι είναι αδύνατο να προβλεφθεί το μέλλον με απόλυτη ακρίβεια και ότι πάντοτε θα υπάρχει, ως ένα βαθμό, σφάλμα στην πρόβλεψη. Οι μέθοδοι για την ανάλυση και το διαχωρισμό χρονοσειρών είναι στατιστικές τεχνικές βελτίωσης των προβλέψεων μέσω της διάσπασης του προτύπου της χρονοσειράς σε υπο-πρότυπα που ονομάζονται συνιστώσες.

Για το διαχωρισμό των χρονοσειρών δεχόμαστε ότι τα δεδομένα εμφανίζουν την εξής εικόνα:

$$\begin{aligned}\text{Δεδομένα} &= \text{πρότυπο} + \text{σφάλμα} \\ &= f(\text{ συστηματικές συνιστώσες}) + \text{σφάλμα}\end{aligned}$$

Το σφάλμα αυτό ή υπολειπόμενη (residual) απόκλιση, υποθέτουμε ότι μπορεί να εκφραστεί ως η διαφορά ανάμεσα στο συνδυασμένο αποτελεσμάτων υπο-μοντέλων της χρονοσειράς (συστηματικές συνιστώσες) και των παρατηρηθεισών τιμών της. Σ' αυτήν την παράγραφο θα μελετήσουμε τη φύση και την έκταση (scope) αυτών των συνιστωσών. Οι τεχνικές διαχωρισμού που χρησιμοποιούνται για τη μελέτη αυτών των συνιστωσών δεν αποτελούν ενιαίο σώμα της στατιστικής θεωρίας. Επίσης, δεν πληρούνται συνήθως όλες οι προϋποθέσεις για επιστημονική στατιστική διερεύνηση, όπως, για παράδειγμα, η ύπαρξη τυχαίας δειγματοληψίας. Επομένως, οι διαδικασίες διάσπασης των χρονοσειρών συχνά αξιολογούνται με κριτήρια που επηρεάζονται από την κρίση εκείνου που διενεργεί την πρόβλεψη.

Το υπόλοιπο της παραγράφου αφιερώνεται σε συζητήσεις γύρω από την ανάλυση χρονοσειρών και τις εφαρμογές της στη διαδικασία της πρόβλεψης. Σε αντίθεση με τις μεθόδους εξομάλυνσης, δεν υπάρχει σώμα στατιστικής θεωρίας που να έχει αναπτυχθεί γι' αυτό ακριβώς το ζήτημα. Επομένως, απαιτείται καλή αξιολογική κρίση προκειμένου να χρησιμοποιηθούν οι μέθοδοι αυτές σε προβλήματα πρόβλεψης. Εν τούτοις, επισημαίνοντας τους στατιστικούς περιορισμούς αυτών των τεχνικών, μπορούμε να μειώσουμε το διάστημα κατά το οποίο θα χρειαστεί να λειτουργήσει η υποκειμενική κρίση κι επομένως να βελτιώσουμε τη διαδικασία πρόβλεψης.

1.2.2 Η Φύση των Οικονομικών Χρονοσειρών

Οι διαχρονικές τιμές μιας οικονομικής χρονοσειράς σχετίζονται με έναν αριθμό παραγόντων που ονομάζονται συνιστώσες. Μελετάμε αυτές τις συνιστώσες για να εξηγήσουμε τις εξελίξεις και την κατεύθυνση των παρατηρήσεων της χρονοσειράς. Υπάρχουν τέσσερις γενικές συνιστώσες που εξηγούν τη συμπεριφορά των οικονομικών χρονοσειρών. Οι τρεις ονομάζονται συστηματικοί ή επαναλαμβανόμενοι συντελεστές και είναι: (α) ο συντελεστής εποχιακότητας, (β) η τάση και (γ) η κυκλική συνιστώσα. Ο τέταρτος συντελεστής είναι μη-επαναλαμβανόμενος και αναφέρεται ως τυχαία, μη-κανονική ή υπολειπόμενη συνιστώσα, δηλαδή σφάλμα.

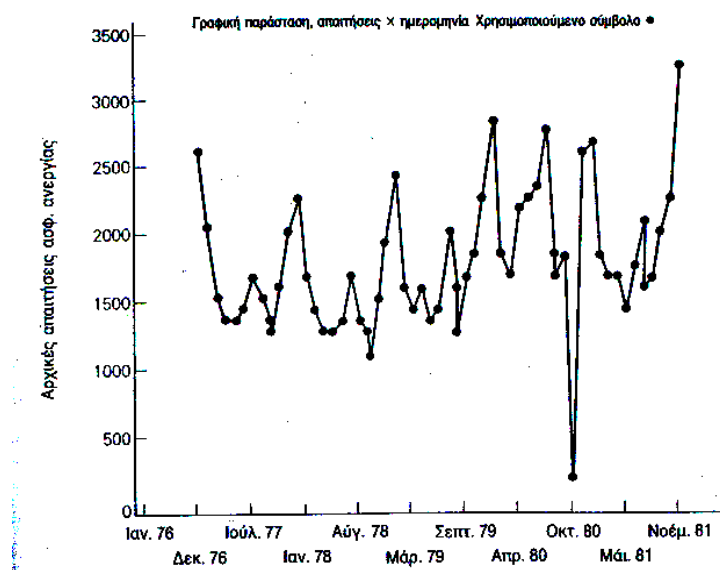
Οι υπεύθυνοι για την πρόβλεψη αναλύουν αυτές τις συνιστώσες για να καθορίσουν την επίδρασή τους πάνω στην εξέλιξη της χρονοσειράς. Οι τέσσερις συνιστώσες αλληλεπιδρούν για να εξηγήσουν τη διακύμανση στις τιμές των δεδομένων διαχρονικά. Οι επιδράσεις αυτές είναι υπεύθυνες για την συχνά ασταθή συμπεριφορά των οικονομικών χρονοσειρών.

Εποχιακότητα

Τα εποχικά πρότυπα παρατηρούνται σε δεδομένα που έχουν ταξινομηθεί ανά τρίμηνο, μήνα ή εβδομάδα. Λαμβάνουν χώρα ή επανακυκλώνονται μέσα στην περίοδο ενός χρόνου, επαναλαμβάνονται από τον ένα χρόνο στον άλλο και είναι προβλέψιμα.

Ένα παράδειγμα εποχικής διακύμανσης παρουσιάζεται στο σχήμα 1.3 όπου εμφανίζονται τα μηνιαία δεδομένα αρχικών απαιτήσεων των πολιτειακών ασφαλιστικών προγραμμάτων ανεργίας.

**Σχήμα 1.3** Αρχικές Απαιτήσεις Επιδομάτων Ανεργίας. Πολιτειακά Προγράμματα (εκατοντάδες)



Οι απαιτήσεις είναι υψηλές τον Ιανουάριο και πέφτουν σχεδόν μέχρι τον Μάιο, φθάνουν σε νέα



κορυφή συνήθως κατά τον Ιούλιο, πέφτουν πάλι κι αρχίζουν ν' ανεβαίνουν από τον Σεπτέμβριο έως τον Ιανουάριο του επόμενου έτους. Είναι φανερό ότι οι απαιτήσεις επιδομάτων ανεργίας είναι πολύ υψηλές τον Ιανουάριο, εφόσον πολλοί άνθρωποι αποδεσμεύονται από την εποχική απασχόληση κατά την περίοδο των διακοπών του Δεκεμβρίου. Οι απαιτήσεις μειώνονται μέχρι την περίοδο των θερμότερων μηνών. Την εποχή αυτή η προσφορά εργασίας αυξάνεται καθώς τα σχολεία και τα κολλέγια σταματούν την εκπαιδευτική τους δραστηριότητα. Οι τιμές που αφορούν την ανεργία συνεχίζουν να είναι υψηλές μέχρι τον Σεπτέμβριο, όπου τα σχολεία και τα κολλέγια αρχίζουν και πάλι την ακαδημαϊκή τους χρονιά.

Για άλλες χρονοσειρές η φύση και οι αιτίες των εποχικών διακυμάνσεων διαφέρουν. Οι αιτίες μπορεί να είναι Φυσικές ή αποτέλεσμα ανθρώπινης παρέμβασης. Αυτές που προέρχονται από ανθρώπινη παρέμβαση είναι εκείνες που σχετίζονται με: (α) τα έθιμα, όπως η προσφορά δώρων τα Χριστούγεννα, την ημέρα της μητέρας ή του πατέρα, (β) την αλλαγή του στυλ των ρούχων κατά τις αλλαγές των εποχών του έτους και (γ) τις ημερομηνίες εισαγωγής νέων μοντέλων αυτοκινήτων που επηρεάζουν τις πωλήσεις.

Στο σχήμα 1.3 απεικονίζεται μια μηνιαία εποχική κίνηση. Εάν είχαν καταγραφεί τα δεδομένα σε ετήσια βάση, οι μηνιαίες εποχικές διακυμάνσεις δεν θα είχαν αναδειχθεί. Ομοίως, μια οικονομική χρονοσειρά που δείχνει τη συμπεριφορά των καταθέσεων σε ένα χρηματοοικονομικό οργανισμό μπορεί να δείχνει ένα επανακυκλούμενο πρότυπο διάρκειας ενός μηνός. Τα δεδομένα της συμπεριφοράς των καταθέσεων σε εβδομαδιαία ή ημερήσια βάση θα μπορούσαν να δείξουν κάτι τέτοιο, εν τούτοις αυτό το εποχικό πρότυπο θα μπορούσε να μην καταγραφεί αν τα δεδομένα ήταν μόνο μηνιαία.

Όταν οι παρατηρήσεις καλύπτουν χρονική περίοδο μεγαλύτερη του έτους, ο διενεργών την πρόβλεψη μπορεί να εστιάσει το ενδιαφέρον του στις άλλες συνιστώσες.

### Τάση

Τάση είναι η συνεχής κίνηση της χρονοσειράς κατά μήκος μιας χρονικής περιόδου. Η τάση μπορεί να εμφανιστεί με γραμμική ή καμπυλόγραμμη μορφή. Συνήθως, οι δυνάμεις που επηρεάζουν ή παράγουν αυτές τις διαχρονικές κινήσεις περιλαμβάνουν μεταβολές στην τεχνολογία, στην παραγωγικότητα, στον εξωτερικό ανταγωνισμό και άλλα.

### Κυκλικές κινήσεις

Οι κυκλικές κινήσεις αναφέρονται σε κυματοειδείς διακυμάνσεις (συχνά μη-κανονικές) που καλύπτουν χρονικό διάστημα μεγαλύτερο του έτους και προέρχονται από μεταβολές της γενικής οικονομικής δραστηριότητας. Αποτελούν τη διαφορά μεταξύ της αναμενόμενης τιμής της χρονοσειράς που προέρχεται από την τάση και την πραγματική τιμή της χρονοσειράς. Επομένως, η κυκλική κίνηση μπορεί να θεωρηθεί ως υπολειπόμενη απόκλιση κυμαινόμενη γύρω από την τάση που σχετίζεται με τη μεταβαλλόμενη γενική οικονομική δραστηριότητα.

### Μη κανονικότητα

Η μη-συστηματική συνιστώσα, ή μη-κανονικές αποκλίσεις, προκαλείται από μη-περιοδικά ή απρόβλεπτα γεγονότα. Τα γεγονότα αυτά περιλαμβάνουν απεργίες, σεισμούς, πολέμους, καιρικές μεταβολές και αλλαγές νομοθεσίας που έχουν οικονομικές επιδράσεις. Τα γεγονότα αυτά δεν επαναλαμβάνονται όπως οι κύκλοι επιχειρηματικής δραστηριότητας που είναι επαναλαμβανόμενοι.

### 1.2.3 Αποσύνθεση Χρονοσειρών

Στην ανάλυση οικονομικών χρονοσειρών, υποτίθεται ότι οι τρεις συστηματικές συνιστώσες και η τυχαία συνιστώσα είναι αποτέλεσμα διαφορετικών αιτιών. Διαχωρίζοντας τις τρεις συστηματικές συνιστώσες, ρίχνουμε κάποιο φως στη φύση των παραγόντων που τις προκαλούν. Η γνώση αυτή μας επιτρέπει να προβάλουμε την επίδραση της συστηματικής συνιστώσας στο μέλλον. Στη συνέχεια, κάνοντας εκτιμήσεις ή υποθέσεις για τη μη-κανονική συνιστώσα μπορούμε να προχωρήσουμε στην πρόβλεψη.

Εάν το  $Y_t$  συμβολίζει την πραγματική παρατήρηση σε χρόνο  $t$ , εκφράζουμε το  $Y_t$  σε όρους της τάσης της σειράς,  $T_t$  σε χρόνο  $t$ , τον εποχικό παράγοντα,  $S_t$  σε χρόνο  $t$ , την κυκλική συνιστώσα,  $C_t$  και τους μη-κανονικούς παράγοντες με  $I_t$ . Στη συνέχεια υποθέτουμε ότι η σχέση παίρνει την παρακάτω μορφή:

$$Y_t = S_t * T_t * C_t * I_t \quad (1.23)$$

Η ισότητα αυτή υποδηλώνει ότι η τιμή της οικονομικής μεταβλητής σε χρόνο  $t$  είναι το γινόμενο των τεσσάρων συνιστωσών που αναφερθήκαν. Ακόμη, η εξίσωση δείχνει ότι τα  $T$ ,  $S$ ,  $C$  και  $I$  αλληλοσχετίζονται. Εναλλακτικά, η σχέση μπορεί να υποτεθεί ότι είναι προσθετική ως εξής:

$$Y_t = S_t + T_t + C_t + I_t \quad (1.24)$$

Σε αυτή την εξίσωση οι συνιστώσες δεν αλληλοσχετίζονται. Το μέγεθος της μιας συνιστώσας είναι ασυσχέτιστο με το μέγεθος οποιασδήποτε άλλης.

Από τα δυο μοντέλα διαχωρισμού, το προσθετικό είναι πιο δύσκολο για παραπέρα ανάλυση. Επίσης, το προσθετικό μοντέλο υποθέτει ότι οι συντελεστές μεταξύ τους είναι ανεξάρτητοι. Για παράδειγμα, αν χρησιμοποιήσουμε το προσθετικό μοντέλο, δεχόμαστε ότι η τάση δεν επηρεάζει τον εποχικό συντελεστή όσο υψηλός ή χαμηλός κι αν γίνεται. Κατά κανόνα, η παραδοχή αυτή δεν είναι ισχυρή, εκτός από την περίπτωση της πολύ βραχυχρόνιας πρόβλεψης. Επομένως, θα δώσουμε έμφαση στο πολλαπλασιαστικό μοντέλο, εφόσον στην πραγματικότητα όλοι οι χρήστες της διαχωριστικής (decomposition) ανάλυσης το βρίσκουν πιο ευλογοφανές.

### 1.2.4 Ανάλυση και Διαχωρισμός της Εποχιακότητας

Η γνώση των εποχικών προτύπων είναι ζωτικός παράγοντας για τον επιχειρηματικό σχεδιασμό. Για παράδειγμα, η πληροφορία ότι οι πωλήσεις λουλουδιών θα είναι υψηλές τον Μάιο στις ΗΠΑ εξαιτίας της ημέρας της Μητέρας, λέει στον ανθοπώλη ότι τα αποθέματά του σε λουλούδια και σε άλλες κατασκευές του είδους θα πρέπει να είναι αυξημένα, προκειμένου να ανταποκριθεί στη ζήτηση. Η γνώση ότι η παραγωγή θα είναι υψηλή κατά τους ζεστούς μήνες, βοηθάει τον κατασκευαστή παιχνιδιών να προγραμματίσει τις ανάγκες του σε κεφάλαιο, μηχανολογικό εξοπλισμό, σε εργατικά χέρια και σε πρώτες ύλες.

Εκτός από την κατάσταση ετήσιου επιχειρηματικού προγράμματος, η γνώση των εποχικών κινήσεων μας επιτρέπει να αναλύουμε τις επιδόσεις του παρελθόντος, να αλλάζουμε το διαχρονικό πρότυπο παραγωγής και διάθεσης κι ίσως ακόμα να μεταβάλλουμε και το ίδιο το εποχικό πρότυπο. Αν και υπάρχουν διάφορες μέθοδοι για τη μέτρηση της εποχικής διακύμανσης, θα εστιάσουμε την προσοχή μας σε μία, στη μέθοδο του λόγου της πραγματικής τιμής προς τον κινητό μέσο (*ratio-to-moving average*). Η λογική που βρίσκεται πίσω από αυτή και άλλες μεθόδους είναι η εξάλειψη της επίδρασης της τάσης, των κυκλικών και μη-κανονικών στοιχείων από τα αρχικά δεδομένα των χρονοσειρών, προκειμένου να διασφαλιστεί ότι παραμένει μόνο η επίδραση του εποχικού παράγοντα.

Με δεδομένα σε τρίμηνη βάση, κατασκευάζουμε έναν κινητό μέσο όρο τεσσάρων τριμήνων για να εκτιμήσουμε τους εποχικούς παράγοντες από τις παρατηρήσεις της χρονοσειράς. Αν χρησιμοποιηθούν μηνιαία δεδομένα, μπορεί να κατασκευαστεί ένας κινητός μέσος όρος 12 μηνών. Στόχος είναι η εξάλειψη των άλλων συνιστωσών έτσι ώστε να απομείνουν μόνο οι εποχικές διακυμάνσεις. Αυτοί οι κινητοί μέσοι όροι θα μπορούσαν να εξαλείψουν τις κυμάνσεις που οφείλονται σε εποχιακότητα και συνυπάρχουν στα δεδομένα. Η βασική παραδοχή που βρίσκεται πίσω από τη μέθοδο του κινητού μέσου όρου είναι ότι το εποχικό πρότυπο επανεμφανίζεται σε κανονική βάση κάθε επόμενο έτος. Μολονότι το εποχικό πρότυπο δεν είναι απόλυτα ακριβές κατά τα διαδοχικά έτη, ωστόσο υπάρχει μέσα σε λογικά όρια και ο κινητός μέσος όρος ελαχιστοποιεί την επίδραση σε έναν τουλάχιστον βαθμό.

Για να εξηγήσουμε τη θεωρία, θα υποθέσουμε ότι οι συνιστώσες της χρονοσειράς είναι πολλαπλασιαστικής μορφής [εξίσωση (1.23)]. Ο όρος *ratio-to-moving-average* χρησιμοποιείται επειδή παίρνουμε τη μεταβλητή  $Y_t$  της χρονοσειράς και τη διαιρούμε με τον κινητό μέσο όρο ως εξής:

$$\frac{Y_t}{MA} = \frac{S_t \times T_t \times C_t \times I_t}{T_t \times C_t \times I_t} = S_t \quad (1.25)$$

Η ποσότητα που βρίσκουμε, η  $S_t$ , είναι η καλύτερη δυνατή μέτρηση της εποχικής συνιστώσας στα δεδομένα της χρονοσειράς.

#### Παράδειγμα 1.6

Κατά το έτος 1980, ο Πρόεδρος του Συμβουλίου Οικονομικών Εμπειρογνομόνων (Council of Economic Advisors) μελέτησε τα στοιχεία απασχόλησης προσεκτικά ως δείκτες για την κατάσταση

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 1ο

της οικονομίας. Η μελέτη δημοσιεύτηκε στο Οικονομικό Δελτίο, για τον Πρόεδρο των ΗΠΑ το 1981. Οι πληροφορίες που χρησιμοποιήθηκαν σ' αυτή την έκθεση προέρχονταν από αρκετές πηγές, μία από τις οποίες είναι η συνολική απασχόληση στον μη-γεωργικό τομέα και αναφέρεται σε πολλές εκδόσεις, συμπεριλαμβανομένης και της *Monthly Labor Review of the US Department of Labor*. Τα δεδομένα για δύο μήνες του 1980 παρουσιάζονται στον Πίνακα 1.9.

**ΠΙΝΑΚΑΣ 1.9** Συνολική απασχόληση στον μη-γεωργικό τομέα (δεδομένα για δύο μήνες του 1980)

Μήνας	Αριθμός απασχολούμενων (χιλιάδες)	Εποχικός δείκτης	Αριθμός απασχολούμενων Εποχικά προσαρμοσμένος (χιλιάδες)
Μάρτιος	90316	99.092	91.144
Απρίλιος	90716	99.791	90.951
Μεταβολή	+400		-193

Στο παραπάνω παράδειγμα, η αύξηση στον αριθμό των απασχολούμενων μεταξύ Μαρτίου και Απριλίου 1980 ήταν 400 χιλιάδες. Εν τούτοις, για να προσδιορίσουμε σωστά τη μεταβολή, οι δύο στήλες πρέπει να προσαρμοστούν ως προς την εποχική διακύμανση. Όταν αυτό επιτευχθεί με τη διαίρεση κάθε αριθμού με τον κατάλληλο εποχικό δείκτη, η μεταβολή στην απασχόληση σε εποχικά προσαρμοσμένη βάση είναι μείον 193 χιλιάδες.

Στο παραπάνω παράδειγμα βλέπουμε τη σημασία της εποχικής προσαρμογής. Σ' αυτή την εφαρμογή κάνεις θα κατέληγε στο συμπέρασμα ότι ο αριθμός των απασχολούμενων μειώθηκε σε εποχικά προσαρμοσμένη βάση, δείχνοντας έτσι συρρίκνωση του αριθμού των απασχολούμενων κατά την περίοδο εκείνη. Εάν δεν γινόταν η εποχική προσαρμογή, το συμπέρασμα θα ήταν το αντίθετο και η ερμηνεία που θα δινόταν θα ήταν εσφαλμένη. Συμπερασματικά, μπορούμε να εξετάσουμε μόνο εποχικά προσαρμοσμένα δεδομένα προκειμένου να καθορίσουμε τις μεταβολές σε οικονομικές χρονοσειρές που εκφράζονται από μηνιαία (ή τριμηνιαία) δεδομένα.

Ένας αριθμός μεθόδων έχουν αναπτυχθεί για τη μέτρηση και την απομόνωση της εποχικής διακύμανσης. Θα περιγράψουμε μόνο τη μέθοδο του ratio-to-moving-average για την απαλοιφή της εποχιακότητας, εφόσον είναι η ευρύτερα εφαρμοζόμενη μέθοδος.

### Παράδειγμα 1.7

Τα δεδομένα που αφορούν δαπάνες για βιομηχανικές εγκαταστάσεις και εξοπλισμό (στο σύνολο του μη-αγροτικού τομέα) προέρχονται από το *Survey of Current Business*, διάφορα τεύχη (βλ. Πίνακα 1.10). Οι τριμηνιαίοι εποχικοί δείκτες θα υπολογιστούν με τη μέθοδο της σύγκρισης των πραγματικών τιμών ως προς έναν τριμηνιαίο κινητό μέσο όρο.

Κατά τη διαδικασία υπολογισμού των εποχικών δεικτών των δαπανών για βιομηχανικές εγκαταστάσεις και εξοπλισμό, θα ακολουθήσουμε έξι στάδια (steps).

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 1ο

**Στάδιο 1.** Υπολογίζουμε έναν συνολικό κινητό μέσο τεσσάρων τριμήνων, όπως εμφανίζεται στη στήλη 2. Η πρώτη εισροή σ' αυτή τη στήλη, 81.2, είναι το άθροισμα των τεσσάρων αριθμών που αντιστοιχούν στα τέσσερα τρίμηνα του 1971 και καταγράφεται στο μέσο της περιόδου ανάμεσα στο δεύτερο και το τρίτο τρίμηνο του 1971. Ο δεύτερος αριθμός της στήλης, 82.9, αποκτάται με την αφαίρεση του αριθμού του πρώτου τριμήνου του 1971 από τον 81.2 και την πρόσθεση του πρώτου τριμήνου του 1972: δηλαδή, 82.9 είναι το άθροισμα των τεσσάρων τριμήνων από το δεύτερο τρίμηνο του 1971 μέχρι το πρώτο τρίμηνο του 1972 και καταγράφεται στο μέσο αυτής της περιόδου. Όλες οι επόμενες εισροές στη στήλη βρίσκονται με τη συνεχή αφαίρεση και πρόσθεση τριμηνιαίων τιμών.

**ΠΙΝΑΚΑΣ 1.10** Υπολογισμοί σε στάδια 1 έως 4 για την εύρεση εποχικών δεικτών των νέων δαπανών για βιομηχανικές εγκαταστάσεις και εξοπλισμό - σύ- νολο μη-αγροτικών επιχειρήσεων (\$ δις)

Έτος	Τρίμηνο	(1)	(2)	(3)	(4)	(5)
		Πραγματ. δεδομένα	Κινητός μέσος 4 τριμήνων	κινητός μέσο κεντραρισμ. (συνολικός)	Κεντραρισμ. μέσος.	στήλη 1 ως ποσοστό %της στήλης 4
1971	1	17.7				
	2	20.6				
	3	20.1	81.2	164.1	20.5	98.0
	4	22.8	82.9	167.2	20.9	109.1
1972	1	19.4	84.3	170.4	21.3	91.1
	2	22.0	86.1	174.6	21.8	100.9
	3	21.9	88.5	179.1	22.4	97.8
	4	25.2	90.6	183.9	23.0	109.6
1973	1	21.5	93.3	189.7	23.7	90.7
	2	24.7	96.4	196.1	24.5	100.8
	3.	25.0	99.7	202.0	25.3	98.8
	4	28.5	102.3	208.1	26.0	109.6
1974	1	24.1	105.8	214.8	26.9	89.6
	2	28.2	109.0	221.4	27.7	101.8
	3	28.2	112.4	226.5	28.3	99.6
	4	31.9	114.1	228.4	28.6	111.5
1975	1	25.8	114.3	228.2	28.5	90.5
	2	28.4	113.9	226.6	28.3	110.4
	3	27.8	112.7	225.5	28.2	98.6
	4	30.7	112.8	226.9	28.4	108.1

**ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 1ο**

1976	1	25.9	114.1	230.8	28.9	89.6
	2	29.7	116.7	237.2	29.7	100.0
	3	30.4	120.5	244.3	30.5	99.7
	4	34.5	123.8	251.6	31.5	109.5
1977	1	29.2	127.8	260.0	32.5	89.8
	2	33.7	132.2	268.0	33.5	100.6
	3	34.8	135.8	290.2	36.3	95.9
	4	38.1	154.4	332.5	41.6	91.6
1978	1	47.8	178.1	379.9	47.5	110.6
	2	57.4	201.8	433.1	54.1	106.1
	3	58.5	231.3	472.1	59.0	99.2
	4	67.6	240.8	491.0	61.4	110.1
1979	1	57.3	250.2	510.3	63.8	89.8
	2	66.8	260.1	530.6	66.3	100.8
	3	68.4	270.5	548.8	68.6	99.7
	4	78.0	278.3	563.8	70.5	110.6
1980	1	65.1	285.5	576.7	72.1	90.3
	2	74.0	291.2	586.7	73.3	101.0
	3	74.1	295.5			
	4	82.3				

**Στάδιο 2.** Υπολογίζουμε το κεντραρισμένο άθροισμα των κινητών μέσων των μεγεθών της στήλης και αναγράφονται στη στήλη 3. Η πρώτη εισροή είναι το άθροισμα των δυο πρώτων τιμών της στήλης 2, η δεύτερη αποτελεί το άθροισμα της δεύτερης και τρίτης τιμής της στήλης 2, Κ.Ο.Κ. Οι εισροές της στήλης 3 Καταχωρούνται ανάμεσα στις εισροές της στήλης 2 και, επομένως, βρίσκονται σε συμφωνία ή είναι κεντραρισμένες με τα αρχικά δεδομένα.

**Στάδιο 3.** Κάθε μέγεθος στη στήλη 3 διαιρείται με το 8 (δύο Φορές τον αριθμό των τριμήνων) για να βρεθεί ο κεντραρισμένος κινητός μέσος. Οι τιμές αυτές του κινητού μέσου είναι οι εκτιμήσεις της τάσης και της κυκλικής διακύμανσης για την απαλοιφή των συνιστωσών T x C από την αρχική χρονοσειρά.

**Στάδιο 4.** Διαιρούμε τη στήλη 1 (τα δεδομένα  $S x T x C x I$  με την αντίστοιχη τιμή της στήλης 4 (εκτίμηση T x C) και πολλαπλασιάζουμε τους λόγους αυτούς επί 100. Με αυτό τον τρόπο φτάνουμε στις εκατοστιαίες τιμές των κινητών μέσων που εμφανίζονται στη στήλη 5. Εκείνο που απομένει είναι να εξαλείψουμε τη μη-κανονική συνιστώσα. Προς τούτο αναπαριστάνουμε τα δεδομένα όπως δίνονται στον Πίνακα 1.11

**Στάδιο 5.** Για κάθε τρίμηνο, βρίσκουμε την επί τοις εκατό τιμή του μέσου όρου για όλα τα έτη. Για το πρώτο τρίμηνο, η επί τοις εκατό τιμή του μέσου όρου είναι 91,4.

**ΠΙΝΑΚΑΣ 1.11.** Υπολογισμοί που αφορούν τα στάδια 5 και 6 προς εύρεση εποχικών δεικτών των δαπανών για βιομηχανικές εγκαταστάσεις και εξοπλισμό

Έτος	Τρίμηνο				
	(1)	(2)	(3)	(4)	
1971			98.0	109.1	
1972	91.9	100.9	97.8	109.6	
1973	90.7	100.8	98.8	109.6	
1974	89.6	101.8	99.6	111.5	
1975	90.5	100.4	98.6	108.1	
1976	89.6	100.0	99.7	109.5	
1977	89.8	100.6	95.9	91.6	
1978	100.6	106.1	99.2	110.1	
1979	89.8	100.8	99.7	110.6	
1980	90.3	101.0			
Σύνολο	822.8	912.4	887.3	969.7	Άθροισμα
Μέσος	91.4	101.4 στάδιο 598.6		107.7	399.1
Εποχικός δείκτης	91.6	101.6 στάδιο 6 98.8		108.0	400.0
$400/399.1 = 1.002\ 256$					

Ωστόσο, τόσο αυτή όσο και οι άλλες επί τοις εκατό τιμές του μέσου όρου δεν μπορούν να χρησιμοποιηθούν από μόνες τους σαν εποχικοί δείκτες. Και τούτο διότι οι τέσσερις επί τοις εκατό τιμές του μέσου όρου πρέπει να έχουν άθροισμα 400, ενώ οι τιμές αυτές αθροίζονται σε 399,1. Έτσι, είναι απαραίτητη μια παραπέρα διόρθωση ώστε να πάρουμε τον εποχικό δείκτη.

**Στάδιο 6.** Προσαρμόζουμε τις επί τοις εκατό τιμές του μέσου όρου ώστε να αθροίζονται στο 400 πολλαπλασιάζοντας κάθε εκατοστιαία τιμή με ένα συντελεστή προσαρμογής  $400/399.1$  ή με  $1.002,256$ . Το τελικό αποτέλεσμα είναι ένας νέος επί τοις εκατό αριθμός, ο οποίος διαφέρει από τη βάση 100 και αντιπροσωπεύει τον εποχικό δείκτη.

Τα δεδομένα μπορούν να απαλλαγούν από την εποχιακότητα ή να προσαρμοστούν για τις εποχικές επιδράσεις διαιρώντας τις αρχικές τιμές των τριμηνιαίων δεδομένων διά του εποχικού δείκτη. Τα απαλλαγμένα από εποχιακότητα δεδομένα αντιπροσωπεύουν μια μέση τιμή που θα μπορούσε να έχει συμβεί εάν ολόκληρο το έτος είχε λειτουργήσει στο ίδιο επίπεδο με το πρώτο τρίμηνο του 1971. Η απαλλαγμένη εποχιακότητας τιμή για το α' τρίμηνο του 1971, θα είναι:

$$\frac{\text{Πρώτο Τρίμηνο 1971}}{\text{Εποχικός Δείκτης}} = \frac{17,7}{91,6_t} \times 100 = 19,3 (\text{\$δισεκατομμύρια})$$

Για να απαλλάξουμε όλα τα δεδομένα της χρονοσειράς του παραδείγματος 1.7 από την εποχική συνιστώσα, διαιρούμε κάθε παρατήρηση με τον κατάλληλο εποχικό δείκτη και πολλαπλασιάζουμε επί 100. Η χρονοσειρά που προκύπτει θα περιλαμβάνει μόνο την τάση, την κυκλική και τη μη-κανονική συνιστώσα. Εξαλείφοντας την εποχική συνιστώσα, οι αναλυτές μπορούν στη συνέχεια να αναζητήσουν και να προσδιορίσουν τους άλλους παράγοντες που εξηγούν τις διακυμάνσεις στα δεδομένα της χρονοσειράς.

Σημειώστε ότι τα αποτελέσματα είναι όμοια, εκτός του ότι παρουσιάζονται οι εποχικοί συντελεστές. Ο εποχικός συντελεστής είναι ο δείκτης που έχει διαιρεθεί διά 100.

Τα στάδια που παρουσιάστηκαν πιο πάνω είναι παρόμοια με εκείνα που θα έπρεπε να ακολουθηθούν στον υπολογισμό των μηνιαίων εποχικών δεικτών. Οι μικρό-τροποποιήσεις που θα έπρεπε να γίνουν μπορεί να δειχθούν συγκρίνοντας τα παραπάνω στάδια με την ακόλουθη διαδικασία:

**Στάδιο 1.** Καθορίζουμε ένα δωδεκάμηνο κινητό σύνολο.

**Στάδιο 2.** Κεντράρουμε το κινητό σύνολο σε κάθε μήνα.

**Στάδιο 3.** Υπολογίζουμε τον κεντραρισμένο κινητό μέσο διαιρώντας το κεντραρισμένο κινητό σύνολο διά του μήκους (length) της εποχιακότητας πολλαπλασιασμένο επί 2(2x12=24).

**Στάδιο 4.** Διαιρούμε κάθε τιμή της στήλης 1 με την αντίστοιχη τιμή της στήλης 4, δηλαδή εκφράζουμε την πραγματική τιμή κάθε μήνα σαν εκατοστιαία αναλογία του κινητού μέσου στο στάδιο 3.

**Στάδιο 5.** Για κάθε μήνα (Ιανουάριος, Φεβρουάριος, ..., Δεκέμβριος) βρίσκουμε την ποσοστιαία τιμή του μέσου από το στάδιο 3 για όλα τα έτη.

**Στάδιο 6.** Προσαρμόζουμε τις ποσοστιαίες τιμές του μέσου ώστε να δίνουν άθροισμα 1200.

Στα προβλήματα που εξετάσαμε μέχρι τώρα, οι εποχικοί δείκτες στους οποίους αναφερθήκαμε ονομάζονται *εποχικές σταθερές* (constant seasonals). Εναλλακτικά διατυπωμένο αυτό σημαίνει ότι ο εποχικός δείκτης για κάθε μήνα υποτίθεται ότι είναι ο ίδιος από έτος σε έτος. Συχνά αυτή η παραδοχή δεν είναι ισχυρή, δεδομένου ότι οι εποχικοί δείκτες μπορεί να παρουσιάζουν διακυμάνσεις. Οι μέθοδοι αυτού του τμήματος με πρόσθετα στάδια μπορούν να εφαρμοστούν για τον προσδιορισμό ενός εποχικού κινητού μεγέθους.

Οι εποχικοί δείκτες που υπολογίστηκαν σ' αυτό το τμήμα χαρακτηρίζονται ως *σχετικοί* εποχικοί δείκτες, δεδομένου ότι αυτοί έχουν εκφραστεί ως ποσοστά επί τοις εκατό. Εδώ γίνεται η παραδοχή ότι η εποχική επίδραση είναι αναλογική προς το επίπεδο της σειράς, πράγμα που σημαίνει ότι σε υψηλότερο επίπεδο το εύρος των εποχικών διακυμάνσεων θα είναι μεγαλύτερο. Ο εποχικός δείκτης για το δεύτερο τρίμηνο στο παράδειγμα 1.7 ήταν 101.6. Επομένως, κατά μέσον όρο, οι πραγματικές τιμές είναι 1,6 τοις εκατό μεγαλύτερες ( $T_t \times C_t \times I_t$ ). Εναλλακτικά, οι εποχικές διακυμάνσεις δεν είναι



αναλογικές με το επίπεδο της χρονοσειράς, αλλά παραμένουν περίπου στα ίδια απόλυτα μεγέθη για όλα τα επίπεδα. Ο δείκτης που περιγράφει κινήσεις σαν αυτές ονομάζεται *απόλυτος εποχικός δείκτης*. Είναι φανερό πως η μέθοδος από όπου προέρχεται αυτός ο τελευταίος τύπος εποχικού δείκτη διαφέρει ελάχιστα από τη μέθοδο που περιγράψαμε παραπάνω. Επομένως, οι εποχικοί δείκτες μπορεί να είναι είτε σταθεροί είτε κινητοί και διατυπώνονται είτε σε σχετικούς είτε σε απόλυτους όρους.

### 1.2.5 Προσαρμογή και Ανάλυση της Τάσης

Μελετούμε τις τάσεις για να προσδιορίσουμε τη μακρόχρονη κατεύθυνση της χρονοσειράς. Γενικά, αυτές οι εξελίξεις προέρχονται από μια ποικιλία αιτιών συμπεριλαμβανομένων και μεταβολών της τεχνολογίας, των τιμών, του πληθυσμού και της παραγωγικότητας.

Το σύνολο των αποταμιευτικών καταθέσεων σε μια τράπεζα ταμειυτηρίου ή σε πιστωτικούς συνεταιρισμούς μπορεί να αυξάνεται από έτος σε έτος για ένα χρονικό διάστημα, επειδή ο αριθμός των καταθετών συνεχίζει να αυξάνεται.

Παραπέρα, ο κάθε καταθέτης τείνει να έχει μεγαλύτερη αποταμίευση σε δολάρια, πράγμα που σχετίζεται με μια γενική αύξηση στα εισοδήματα των καταθετών. Εν μέρει, αυτό προέρχεται από μια γενική αύξηση των τιμών που ονομάζεται πληθωρισμός. Ένας βιομήχανος laser μπορεί να παρατηρήσει μια αύξηση των πωλήσεων που οφείλεται σε τεχνολογικές μεταβολές. Η ανάπτυξη και η βελτίωση των laser συνοδευμένη από ευρύτερες εφαρμογές στη χρήση τους έχει αυξήσει τον αριθμό των πωλήσεών τους διαχρονικά.

Οι μεταβολές στην παραγωγικότητα που με τη σειρά τους μπορεί να προέρχονται από τεχνολογικές μεταβολές, δίνουν μια προς τα πάνω τάση σε πολλές οικονομικές χρονοσειρές.

Τέλος, το γενικό ανοδικό πρότυπο των εσόδων ενός νοσοκομείου σχετίζεται με τη γενική αύξηση των τιμών ή των τιμών των νοσοκομειακών υπηρεσιών. Είναι η αύξηση του μεγέθους και της χρήσης μιας πιο σύγχρονης ιατρικής τεχνολογίας, που έχει συμβάλει σε μια γενική αύξηση των τιμών που χρεώνουν τα νοσοκομεία για τις υπηρεσίες που προσφέρουν.

Ένα πρώτο βήμα στην ανάλυση της συνιστώσας της τάσης είναι ο καθορισμός του λόγου για τον οποίο γίνεται η ανάλυσή της. Μπορεί κάποιος να θέλει να προβάλει την τάση στο μέλλον ή να την εξαλείψει από την αρχική χρονοσειρά. Φυσικά, ίσως κάποιος να επιθυμεί να πράξει και τα δυο μαζί. Η γνώση του σκοπού βοηθάει τον ασχολούμενο με τις προβλέψεις στην επιλογή της μεθόδου και του εύρους της χρονοσειράς που θα αναλύσει. Εάν ο σκοπός είναι η προβολή της τάσης στο μέλλον, είναι σημαντικό να γνωρίζει το πόσο μακριά στ/μ μέλλον πρόκειται να γίνει αυτή η προβολή. Μια προβολή για 15 χρόνια κατά κανόνα απαιτεί τη χρησιμοποίηση χρονοσειράς μακρύτερης διάρκειας από ό,τι μια προβολή τριών μόνο χρόνων.

Το πρώτο βήμα για την ανάλυση της τάσης είναι η γραφική παράσταση της χρονοσειράς, τόσο σε αριθμητική όσο και σε ημι-λογαριθμική κλίμακα πριν την επιλογή της μεθόδου μέτρησης. Ειδικότερα, ας θεωρήσουμε το επόμενο παράδειγμα. Αν και υπάρχουν στατιστικά κριτήρια για τον καθορισμό της «σπουδαιότητας» της τάσης, μια εξέταση της γραφικής παράστασης ως προς τον χρόνο

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 1ο

θα είναι επαρκής.

### Παράδειγμα 1.8

Τα δεδομένα στον Πίνακα 1.12 αποτελούν μέσες εβδομαδιαίες αποδοχές που αφορούν εργάτες, οι οποίοι δεν ανήκουν στο εποπτικό προσωπικό του μη αγροτικού τομέα.

### **ΠΙΝΑΚΑΣ 1.12** Μέσες εβδομαδιαίες αποδοχές εργατών

Έτος	Μέσες Εβδομαδιαίες Αποδοχές
1950	53,13
1960	80,67
1970	119,83
1980	235,10

Κατά τη διάρκεια της περιόδου 1950-60, οι μέσες εβδομαδιαίες αποδοχές αυξήθηκαν κατά 27.54 \$. Για την περίοδο 1960-1970, η αύξηση ήταν 39.16 \$ και για την περίοδο 1970-80 η αύξηση ήταν 115.27 \$. Σημειώστε ότι η αριθμητική αύξηση στις μέσες εβδομαδιαίες αποδοχές ήταν περίπου η ίδια για τις δύο πρώτες δεκαετίες, αλλά αυξήθηκε κατά ένα πολύ μεγαλύτερο ποσό σε απόλυτους αριθμούς κατά την περίοδο της τελευταίας δεκαετίας. Οι σχετικές αυξήσεις και οι λόγοι από περίοδο σε περίοδο δίνονται στον Πίνακα 1.13.

### **ΠΙΝΑΚΑΣ 1.13** Σχετικές αυξήσεις εβδομαδιαίων αποδοχών

Έτος		Αύξηση(%)	Λόγος
Από	Έως		
1950	1960	51,84	1,52:1
1960	1970	48,54	1,49:1
1970	1980	96,19	1,96:1

Παρατηρώντας τα δεδομένα και στους δύο πίνακες βλέπουμε ότι οι σχετικές μεταβολές ήταν σχεδόν όμοιες κατά την πρώτη 20ετία, αλλά αυξήθηκαν ουσιαστικά κατά την τελευταία δεκαετία. Οι απόλυτες μεταβολές δεν ήταν ίδιες καθ' όλη τη διάρκεια και των τριών δεκαετιών.

Ο άξονας των  $X$  (οριζόντιος) καθορίζεται σύμφωνα με την αριθμητική κλίμακα και ο άξονας των  $Y$  (κάθετος) σύμφωνα με τη λογαριθμική κλίμακα. Η λογαριθμική κλίμακα καθορίζεται κατά τρόπο που η πραγματική γραφική παράσταση είναι ο λογάριθμος των δεδομένων (τιμές  $Y$ ) και όχι τα ίδια τα δεδομένα. Όταν τα δεδομένα παριστάνονται γραφικά σε τέτοια κλίμακα, οι ίσες κάθετες αποστάσεις παριστάνουν ίσες σχετικές μεταβολές παρά ίσες απόλυτες μεταβολές.

Στο σχήμα 1.4, τα δεδομένα μέσω εβδομαδιαίων αποδοχών για τη συγκεκριμένη χρονική περίοδο παρουσιάζονται και στις δύο κλίμακες, δηλαδή την αριθμητική και τη λογαριθμική. Σημειώστε ότι στην περίπτωση χρησιμοποίησης της αριθμητικής κλίμακας, οι αποστάσεις από το  $A$

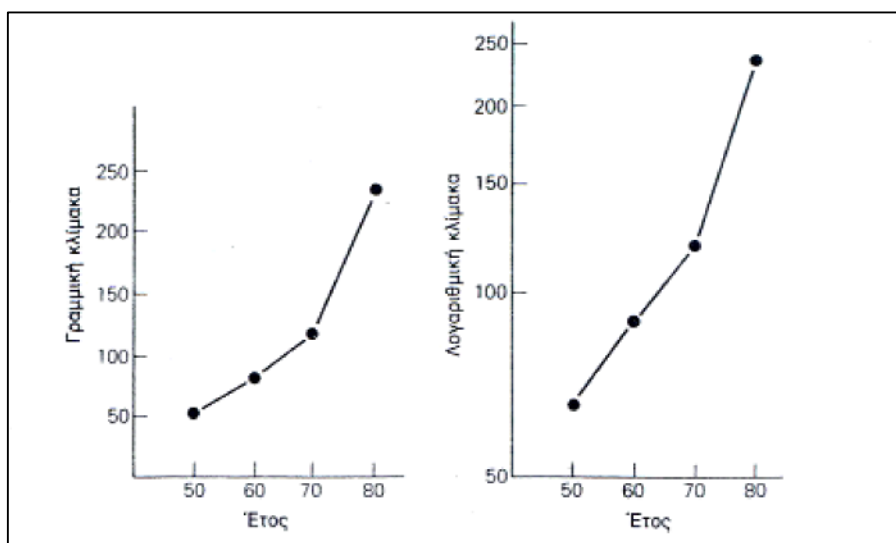
στο Β και από το Β στο C είναι μικρότερες της απόστασης από το C στο D. Αυτό, παριστάνει τις απόλυτες μεταβολές από το 1950 στο 1960, από το 1960 στο 1970 και από το 1970 στο 1980.

Στη γραφική παράσταση με ημι-λογαριθμική κλίμακα η κάθετη απόσταση από το Α στο Β είναι μεγαλύτερη από εκείνη ανάμεσα στο Β και στο C. Η κάθετη απόσταση από το C στο D είναι ακόμη μεγαλύτερη. Τα αποτελέσματα αυτά δείχνουν τις σχετικές μεταβολές που παρατηρήθηκαν στις τρεις χρονικές περιόδους. Παραπέρα, αυτές οι σχετικές μεταβολές δεν μπορούν να φανούν καθαρά στην αριθμητική γραφική παράσταση, αλλά φαίνονται ολοκάθαρα στην ημι-λογαριθμική κλίμακα.

Για να καταλάβουμε το γιατί έχουμε διαφορετική εικόνα σε ένα ημι-λογαριθμικό γράφημα και σε ένα αριθμητικό απαιτείται γνώση των λογαριθμικών κλιμάκων. Οι λογαριθμικές κλίμακες ποικίλλουν ανάλογα με τη σειρά. Η κλίμακα μπορεί να είναι μονής σειράς από το 1 έως το 10, διπλής σειράς 1 ως 10 ως 100, ή σειράς υψηλότερων επιπέδων, δηλαδή 1 ως 10 ως 100 ως 1000. Από τη στιγμή που ίσες αποστάσεις αναπαριστούν ίσες αναλογίες, η απόσταση από το 1 έως το 2 είναι ίδια με την απόσταση από το 2 έως το 4, από το 10 στο 20, από το 100 στο 200, Κ.Ο.Κ. Για τον ίδιο λόγο, η απόσταση από το 1 στο 2 (αναλογία 2:1) είναι μεγαλύτερη από την απόσταση από το 2 στο 3 (αναλογία 3:2) ή από το 3 στο 4 (αναλογία 4:3). Επομένως, όσο μεγαλύτερη είναι η σχετική μεταβολή, τόσο μεγαλύτερη είναι η κάθετη απόσταση.

Για τα δεδομένα του παραδείγματος 1.8, σημειώνουμε ότι ο λόγος των αποδοχών ανάμεσα σε κάθε δεκαετία ποικίλλει από 1,49:1 έως 1,96:1. Αυτό το σχετικά μικρό εύρος στις αναλογίες δείχνει ότι η ημι-λογαριθμική κλίμακα θα εμφανιζόταν περισσότερο σαν ευθεία γραμμή, εάν τα ίδια δεδομένα παριστάνονταν γραφικά πάνω σε αριθμητικό γράφημα. Επομένως, μπορούμε να καταλήξουμε ότι η φύση της οικονομικής χρονοσειράς καθορίζει ποια κλίμακα, δηλαδή η αριθμητική ή η λογαριθμική, είναι η πλέον κατάλληλη.

**Σχήμα 1.4 :** Οι εβδομαδιαίες αποδοχές σε γραμμική και λογαριθμική κλίμακα



Ακόμη, οι ημι-λογαριθμικές γραφικές παραστάσεις χρησιμοποιούνται όχι μόνο για να εκφράσουν σχετικές μεταβολές, αλλά και για να παραστήσουν δεδομένα που διαφέρουν σημαντικά σε

απόλυτα μεγέθη. Οι ενεργούντες προβλέψεις ενδέχεται να έχουν στη διάθεσή τους δύο ομάδες δεδομένων όπου το μέγεθος της μίας σειράς να εμφανίζει πολύ μεγαλύτερη διακύμανση σε σχέση με την άλλη. Τέλος, πρέπει να επιδείξουμε μεγάλη προσοχή όταν ερμηνεύουμε και εξηγούμε ότι ίσες κάθετες αποστάσεις πάνω σε ημι-λογαριθμική γραφική παράσταση αντιπροσωπεύουν ίσες αναλογίες, ή ίση σχετική μεταβολή.

### 1.2.6 Το Μοντέλο Γραμμικής Τάσης

Η αξιολογική κρίση (judgement) χρησιμοποιείται πολύ συχνά ως κριτήριο για τον προσδιορισμό του μοντέλου με την καλύτερη προσαρμογή. Μερικές φορές ο διενεργών προβλέψεις χαράζει απλώς μια γραμμή διαμέσου της σειράς των παρατηρήσεων, προκειμένου να αποκαλύψει το σχήμα και την κατεύθυνση της τάσης της χρονοσειράς. Ο προβλέπων, χαράζοντας τη γραμμή της τάσης πρέπει να έχει την ικανότητα να αναγνωρίζει την επαναλαμβανόμενη εποχική διακύμανση και τους κύκλους μέσω των οποίων διέρχεται η τάση. Με άλλα λόγια, θα πρέπει ο προβλέπων να έχει εξοικειωθεί σημαντικά με αυτές τις συγκεκριμένες σειρές, διαφορετικά η διαδικασία αυτή μπορεί να αποδειχθεί εξαιρετικά δύσκολη. Γι' αυτό οι διενεργούντες προβλέψεις συνήθως εμπιστεύονται την αντικειμενική μέθοδο των ελαχίστων τετραγώνων για να αποφύγουν την υποκειμενικότητα που υπεισέρχεται στη μέθοδο που βασίζεται στην αξιολογική κρίση.

Για να αρχίσουμε, θα ορίσουμε την εκτιμηθείσα γραμμική εξίσωση ως εξής:

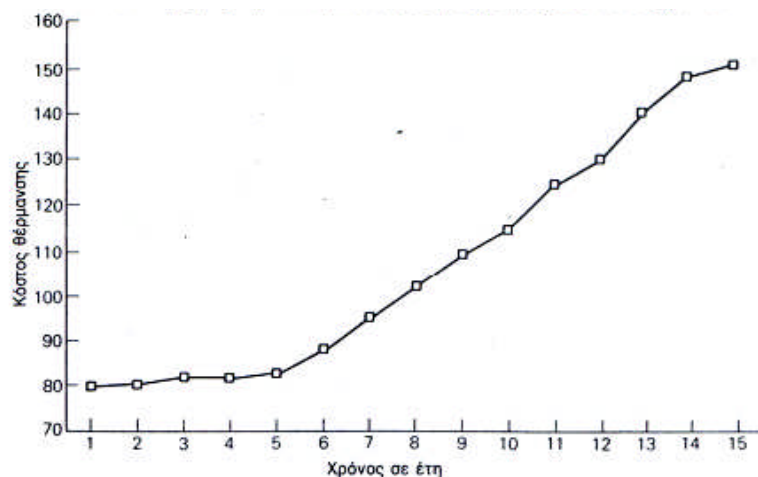
$$\hat{Y}_t = a + bX_t$$

όπου  $\hat{Y}_t$  είναι η προβλεφθείσα τιμή τάσης για τη μεταβλητή  $Y$  για μια επιλεγείσα κωδικοποιημένη χρονική περίοδο  $X$ .  $a$  είναι η τιμή της τάσης όταν  $X=0$  (η διαφορά ύψους).  $b$  είναι η κλίση της εξίσωσης (η μεταβολή στο  $\hat{Y}$  που σχετίζεται με μια μοναδιαία μεταβολή στη  $X$ ). Το  $X$  παριστάνει την παρατήρηση για την κωδικοποιημένη χρονική περίοδο.

#### Παράδειγμα 1.9

Τα πραγματικά δεδομένα του κόστους θέρμανσης ενός πανεπιστημίου, για μια χρονική περίοδο 15 χρόνων, παριστάνονται γραφικά στο σχήμα 3.2. Το γράμμα  $Y$  υποδηλώνει τις πραγματικές τιμές της χρονοσειράς. Αν και παρατηρείται κάποια απόκλιση στη χρονοσειρά, το κόστος θέρμανσης εμφανίζεται αυξανόμενο κατά ένα σταθερό ποσό καθ' όλη τη διάρκεια της χρονικής περιόδου. Επομένως, η γραμμική τάση είναι η κατάλληλη και μπορεί να υπολογιστεί με τη μέθοδο των ελαχίστων τετραγώνων (Πίνακας 1.14).

Σχήμα 1.5: Γραφική Παράσταση του κόστους θέρμανσης έναντι χρόνου



ΠΙΝΑΚΑΣ 1.14: Υπολογισμοί των ελαχίστων τετραγώνων

	Περίοδος X	Κόστος θέρμανσης \$ εκατοντάδες	X <sup>2</sup>	XY
	1	80	1	80
	2	81	4	102
	3	82	9	246
	4	82	16	328
	5	83	25	415
	6	88	36	528
	7	95	49	665
	8	102	64	816
	9	110	81	990
	10	115	100	1150
	11	125	121	1375
	12	130	144	1560
	13	140	169	1820
	14	148	196	2072
	15	151	225	2265
Σύνολο	120	1612	1240	14472
	ΣX	ΣY	ΣX <sup>2</sup>	ΣXY

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 1ο

Κατά την εκτίμηση της εξίσωσης της τάσης, οι τύποι και οι υπολογισμοί είναι οι ακόλουθοι:

1. Ο τύπος για την κλίση είναι

$$b = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2} \quad (1.26)$$

Με αντικατάσταση βρίσκουμε

$$\begin{aligned} b &= \frac{15(14472) - (120)(1612)}{15(1240) - (120)^2} \\ &= 5.6286 \end{aligned}$$

2. Αντικαθιστούμε τώρα την κλίση στον τύπο, για να βρούμε τη διαφορά ύψους

$$a = \frac{\sum Y}{n} - b \frac{\sum X}{n} = \frac{1612}{15} - 5.6286 \left( \frac{120}{15} \right) = 62.44 \quad (1.27)$$

Λαμβάνοντας προσεγγιστικά δύο δεκαδικά ψηφία, γράφουμε την εξίσωση ως εξής:

$$\hat{Y}_t = 62.44 + 5.63 X_t$$

(αρχή στο  $X=0$  μια μονάδα του  $X=1$  έτος).

Για να ερμηνεύσουμε την παραπάνω εξίσωση, λέμε ότι από έτος σε έτος (ή χρονική περίοδο) υπάρχει μια μεταβολή στον μέσο όρο της τάξης των 5.63 (\$ χιλιάδων) στο κόστος θέρμανσης. Στο έτος (χρονική περίοδος) μηδέν, το κόστος υπολογίζεται στα 62.44 (\$ χιλιάδες).

Μια εναλλακτική διαδικασία υπολογισμού των τιμών της τάσης είναι η κωδικοποίηση των χρονικών περιόδων (τιμές  $X$ ) με άλλο τρόπο. Σύμφωνα με αυτή τη δεύτερη μέθοδο, οι υπολογισμοί γίνονται ευκολότερα, γιατί οι χρονικές περίοδοι είναι κωδικοποιημένες έτσι που οι τιμές του  $X$  α-θροίζονται στο μηδέν. Όταν  $\sum X=0$ , ο τύπος της κλίσης γίνεται

$$b = \frac{n \sum XY}{n \sum XY^2} \quad (1.28)$$

Παραπέρα, ο τύπος της διαφοράς ύψους γίνεται:

$$a = \frac{\sum Y}{n} \quad (1.29)$$

Σημειωτέον πως η εξίσωση της γραμμής τάσης θα είναι:

$$\hat{Y}_t = 62.44 + 5.63 X$$

Εάν θέλουμε να προβλέψουμε μια περίοδο μπροστά, δηλαδή τη χρονική περίοδο 16, αντικαθιστούμε τον αριθμό αυτό στην εξίσωση τάσης

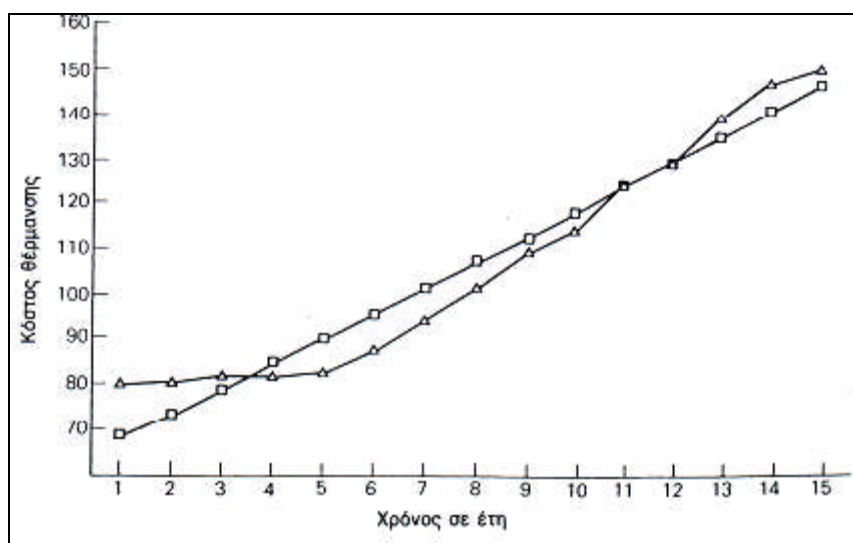
$$\hat{Y}_{16} = 62.44 + 5.63 (16) = 152.52$$

Κατά συνέπεια για την περίοδο 16, το κόστος θέρμανσης υπολογίζεται ότι είναι 152520 \$.

Υπολογίζοντας εν συνεχεία με όμοιο τρόπο τις εκτιμώμενες τιμές προκύπτει το διάγραμμα διασποράς εκτιμημένων (fitted) και πραγματικών τιμών του σχήμα 3.3, στο οποίο παρατηρείται μεγάλη προσέγγιση των εκτιμημένων ή θεωρητικών και των πραγματικών τιμών.

Αν και η εξίσωση της ευθύγραμμης τάσης εφαρμόζεται πιο συχνά στις χρονοσειρές από οποιοδήποτε άλλο μοντέλο για να περιγράψει μακροχρόνιες κινήσεις, η χρήση άλλων μοντέλων για την περιγραφή μεταβολών συχνά είναι πιο ενδεδειγμένη.

**Σχημα 1.6:** Παρατηρηθείσες και προβλεφθείσες τιμές. Οι προβλεφθείσες τιμές σχηματίζουν ευθεία γραμμή



### 1.2.7 Εκτίμηση της Τάσης με Μηνιαία ή Τριμηνιαία Δεδομένα

Στην παράγραφο αυτή, εκτιμήσεις τάσης με ετήσια δεδομένα θα μετασχηματιστούν σε εκτιμήσεις τάσης που θα αφορούν χρονική περίοδο μικρότερη του έτους.

Προηγουμένως, η τάση ορίστηκε σαν μακροχρόνια κίνηση (εξέλιξη) στη χρονοσειρά. Βασίζομενοι στον ορισμό αυτό θα αναπτύξουμε στην παράγραφο αυτή μια διαδικασία που θα μας επιτρέπει να παίρνουμε μια βραχυχρόνια εξίσωση τάσης από μια ήδη υπολογισθείσα ετήσια εξίσωση τάσης. Οι εκτιμήσεις αυτές γίνονται απλά με τον υπολογισμό των εκτιμήσεων των παραμέτρων σε ετήσιους όρους και στη συνέχεια διαιρούμε είτε με το 12 για να πάρουμε μηνιαίες εκτιμήσεις, είτε με το 4 για τριμηνιαίες.

#### Παράδειγμα 1.10

Δίνονται τα δεδομένα της American Hospital Supply, Inc. για τα έτη από 1965 έως 1980, για την εύρεση της συνιστώσας της τάσης της χρονοσειράς του καθαρού εισοδήματος (\$ εκατομμύρια). Βασισμένη στην πραγματική τιμή, η εξίσωση τάσης εκτιμάται ότι είναι  $\hat{Y}_t = -10.877 + 7.056X_t$  (αρχή το έτος 1964) και έχει υπολογιστεί με τη μέθοδο των ελαχίστων τετραγώνων. Σημειώστε ότι το  $X=0$

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 1ο

θα αναπαριστά την 1η Ιουλίου 1964, το μέσο του διαστήματος του έτους 1964.

Αλλάζοντας την κλίμακα της εξίσωσης από ετήσια σε τριμηνιαία, διαιρούμε την εξίσωση διά 4:

$$\hat{Y}_t = \frac{-10.877 + 7.056X_t}{4} = -2.71925 + 1.76400X_t \quad (\$ \text{ εκατομμύρια})$$

Εάν θέλουμε να μετατρέψουμε το καθαρό εισόδημα από εκατομμύρια δολάρια σε χιλιάδες \$, πολλαπλασιάζουμε την εξίσωση επί 1000 και έχουμε:

$$\hat{Y}_t = -2719.25 + 1764.0X_t \quad (\$ \text{ χιλιάδες})$$

έτη. Επομένως, το X μετατρέπεται για να αναπαραστήσει τρίμηνα μέσω της διαίρεσης διά 4, ως εξής:

$$\hat{Y}_t = \frac{-2719.25 + 1764.0X_t}{4} = -2719.25 + 441.0X_t$$

Εάν θέλουμε να μετατρέψουμε την αρχή από την 1η Ιουλίου 1964, στην 1η Ιανουαρίου 1965 (πρώτο τρίμηνο), κάνουμε την εξής αντικατάσταση:

$$\hat{Y}_t = -2719.25 + 441.0 \times (X_t + 2) = -1837.25 + 441.0 \times X_t$$

(αρχή στην 1η Ιανουαρίου 1965, μία μονάδα του X=1 τρίμηνο). Τέλος, για να κάνουμε πρόβλεψη για το πρώτο τρίμηνο του 1981 ( $X_t = 65$ ), θέτουμε

$$\hat{Y}_{65} = -1837.25 + 441.0 \times (65) = 26827.75 \quad (\$ \text{ χιλιάδες})$$

Η παραπάνω πρόβλεψη αποτελεί την πρόβλεψη της τάσης για το πρώτο τρίμηνο του 1981. Όμοια, οι μηνιαίες προβλέψεις προκύπτουν από εξισώσεις ετήσιας βάσης μετά από διαίρεση με τον αριθμό 12, ώστε να εκφραστούν σε μήνες, αντί για το 4 που χρησιμοποιήσαμε στο παράδειγμα προκειμένου για τα τρίμηνα.

### 1.2.8 Απομόνωση της Κυκλικότητας

Στη παράγραφο αυτή θα ασχοληθούμε με μια εφαρμογή της μεθόδου του κινητού μέσου όρου προκειμένου να απομονωθεί η κυκλική συνιστώσα. Αυτή η διαδικασία απομόνωσης είναι παρόμοια με τη θεωρία απομόνωσης της τάσης και της εποχικής συνιστώσας, με τις οποίες ασχοληθήκαμε νωρίτερα.

Η απομόνωση μιας κυκλικής συνιστώσας επιτυγχάνεται με την απόλεια της εποχιακότητας από τα αρχικά δεδομένα ως εξής:

$$\frac{Y_t}{\text{Εποχικός δείκτης}} = \frac{S_t \times T_t \times C_t \times I_t}{S_t} = T_t \times C_t \times I_t \quad (1.30)$$

Τα δεδομένα αυτά, τα οποία είναι απαλλαγμένα από την εποχιακότητα, μπορούν να απαλλαγούν από την τάση ως εξής:



$$\frac{T_t \times C_t \times I_t}{T_t} = C_t \times I_t \quad (1.31)$$

Οι σειρές που προκύπτουν περιλαμβάνουν μόνο την επίδραση που σχετίζεται με την κυκλική συνιστώσα και τη μη-κανονική διακύμανση.

Η μη-κανονική διακύμανση μπορεί να εξαλειφθεί από τη χρονοσειρά μέσω υπολογισμού του κινητού μέσου όρου, που εξομαλύνει αυτές τις μη-κανονικές διακυμάνσεις από τα δεδομένα. Για να περιγράψουμε αυτή τη διαδικασία, θεωρούμε το επόμενο παράδειγμα.

### Παράδειγμα 1.11

Τα δεδομένα που αφορούν δαπάνες για βιομηχανικές εγκαταστάσεις και νέο εξοπλισμό (σύνολο του μη-αγροτικού τομέα) έχουν συλλεγεί από το *Survey of Current Business* στο παράδειγμα 1.8.

Τα τριμηνιαία δεδομένα με τη σειρά τους απαλλάχθηκαν από την εποχιακότητα και το αποτέλεσμα αναγράφεται στη στήλη 1 του Πίνακα 1.15. Τα δεδομένα αυτά χρησιμοποιούνται για την εκτίμηση της συνιστώσας της τάσης με τη μέθοδο των ελαχίστων τετραγώνων. Η στήλη 2 του Πίνακα 1.15 περιέχει τη συνιστώσα της τάσης. Τα δεδομένα είναι τώρα έτοιμα για ανάλυση με την τεχνική του κινητού μέσου όρου.

Η στήλη 3 του Πίνακα 1.15 παριστάνει την κυκλική μη-κανονική συνιστώσα σαν ποσοστό επί τοις εκατό της τάσης που αναφέρεται στη στήλη 2. Σημειώστε ότι αυτή η συνιστώσα  $C_t \times I_t$  είναι τώρα έτοιμη για εξομάλυνση με τη διαδικασία της μεθόδου του κινητού μέσου όρου.

Η μεθοδολογία του κινητού μέσου όρου υποθέτει ότι το πρότυπο που ακολουθεί η χρονοσειρά μπορεί να αναπαρασταθεί με έναν μέσο όρο των παρατηρήσεων του παρελθόντος. Η παρουσία ή η απουσία των κυκλικών δυνάμεων στον κινητό μέσο όρο εξαρτάται από το εύρος του κινητού μέσου όρου, δηλαδή, τρεις, τέσσερις, πέντε ή περισσότερες προηγούμενες περιόδους.

**ΠΙΝΑΚΑΣ 1.15** Απομόνωση της κυκλικής συνιστώσας για τις δαπάνες σε βιομηχανικές εγκαταστάσεις και νέο εξοπλισμό (\$ δισεκατομμύρια)

Έτος	Τρίμηνο	(1)	(2)	(3)	(4)	(5)
		Τιμές απαλ- λαγμένες εποχιακότητας	Τάση $Y_t$	$C_t, X_t$ σαν ποσοστό επί της τάσης	Σταθμισμ. περίοδος μέσον, σύνολο	Κυκλική συνιστώσα
1971	1	19.3	9.6	201.0		
	2	20.3	11.1	182.9	727.9	182.0
	3	20.3	12.6	161.1	654.7	163.7
	4	21.1	14.1	149.6	596.2	149.1
1972	1	21.2	15.6	135.9	548.5	137.1
	2	21.6	17.0	127.1	510.1	127.5
	3	22.2	18.5	120.0	483.6	120.9
	4	23.3	20.0	116.5	462.3	115.6
1973	1	23.5	21.5	109.3	440.7	110.2
	2	24.3	23.0	105.6	423.8	106.0
	3	25.3	24.5	103.3	413.7	103.4
	4	26.4	26.0	101.5	401.9	100.5
1974	1	26.3	27.5	95.6	388.9	97.2
	2	27.8	28.9	96.2	381.8	95.5
	3	28.5	30.4	93.8	376.3	94.1
	4	29.5	31.9	92.5	363.2	90.8
1975	1	28.2	33.4	84.4	341.2	85.3
	2	27.9	34.9	79.9	321.4	80.4
	3	28.1	36.4	77.2	309.2	77.3
	4	38.4	37.9	74.9	298.8	74.7
1976	1	28.3	39.4	71.8	290.1	72.5
	2	29.2	40.8	71.6	286.6	71.7
	3	30.3	42.3	71.6	287.6	71.9
	4	31.9	43.8	72.8	287.6	71.9
1977	1	31.9	45.3	70.4	284.5	71.1
	2	33.2	46.8	70.9	285.1	71.3
	3	35.2	48.3	72.9	287.6	71.9
	4	35.3	49.8	70.9	316.5	79.1
1978	1	52.2	51.3	101.8	381.7	95.4
	2	56.5	52.7	107.2	425.4	106.4
	3	59.2	54.2	109.2	438.0	109.5
	4	62.6	55.7	112.4	443.4	110.9
1979	1	62.6	57.2	109.4	443.1	110.8
	2	65.7	58.7	111.9	448.2	112.1
	3	69.2	60.2	115.0	458.9	114.7
	4	72.2	61.7	117.0	461.5	115.4
1980	1	71.0	63.1	112.5	454.7	113.7
	2	72.8	64.6	112.7	451.4	112.9
	3	75.0	66.1	113.5	452.4	113.1
	4	76.2	67.6	112.7		

ΠΗΓΗ: Πραγματικά δεδομένα από το *Survey of Current Business*. διάφορα τεύχη.

Ένα απλό μοντέλο κινητού μέσου  $M_t$ , παρουσιάζεται στην ακόλουθη εξίσωση:

$$M_t = \frac{Z_t + Z_{t-1} + Z_{t-2} + \dots + Z_{t-n+1}}{n} \quad (1.32)$$

όπου  $M_t$  είναι ο κινητός μέσος όρος τον χρόνο  $t$ .  $Z_t$  είναι η τιμή στην περίοδο  $t$  στην εξίσωση (3.14), η απαλλαγμένη από την εποχιακότητα και την τάση, δηλαδή  $Z_t = C_t \times I_t$  και  $n$  είναι ο αριθμός των όρων στον κινητό μέσο όρο.

Για ευκολία στους υπολογισμούς, η εξίσωση μπορεί να ξαναγραφεί ως:

$$M_t = M_{t-1} + \frac{Z_t - Z_{t-n}}{n} \quad (1.33)$$

Επομένως, για τη χρονική περίοδο  $t$ , ο κινητός μέσος όρος είναι ο αριθμητικός μέσος των  $n$  πιο πρόσφατων παρατηρήσεων. Στις παραπάνω εξισώσεις, ορίζονται ίσοι συντελεστές στάθμισης σε κάθε μία από τις  $n$  πιο πρόσφατες παρατηρήσεις. Σε όλες τις άλλες ιστορικές παρατηρήσεις οι συντελεστές στάθμισης θεωρείται ότι είναι μηδενικοί. Σε κάθε παρατήρηση  $Z_t$  θα θεωρηθούν ως συντελεστές στάθμισης οι  $W_t$ . Οι τύποι ξαναγράφονται ως εξής:

$$M_t = \frac{W_t Z_t + W_{t-1} Z_{t-1} + W_{t-2} Z_{t-2} + \dots + W_{t-n+1} Z_{t-n+1}}{n} \quad (1.34)$$

και

$$M_t = M_{t-1} + \frac{W_t Z_t - W_{t-n} Z_{t-n}}{n} \quad (1.35)$$

όπου  $W_t$  είναι ο συντελεστής στάθμισης που αντιστοιχεί στην παρατήρηση  $Z_t$  σε χρόνο  $t$ .

Καθώς θα έχουμε στη διάθεσή μας μια νέα τιμή των δεδομένων, αυτή θα περιλαμβάνεται στον μέσο όρο και η τιμή που αντιστοιχεί στη η-οστή προηγούμενη περίοδο θα αγνοείται.

Επομένως,  $M_t$  είναι μια επικαιροποιημένη έκδοση της αμέσως προηγούμενης εκτίμησης. Η δεύτερη μορφή του  $M_t$  τονίζει αυτό το σημείο.

Οι κινητοί μέσοι όροι θα ανταποκρίνονται στις μεταβολές στο πρότυπο των δεδομένων σε βαθμό που εξαρτάται τόσο από τον αριθμό των περιόδων που περιλαμβάνονται στον κινητό μέσο όρο όσο και από τους συντελεστές στάθμισης, που αντιστοιχούν στη φόρμουλα. Σημειωτέον, όποτε οι συντελεστές στάθμισης θεωρηθούν οι ίδιοι, τότε οι τύποι για τους σταθμισμένους και τους μη σταθμισμένους κινητούς μέσους όρους είναι ισοδύναμοι. Εάν δίνονται μεγάλοι συντελεστές στάθμισης στις τρέχουσες περιόδους και πολύ μικροί στις παλαιότερες, τότε η πρόσφατη παρατήρηση θα επιδρά περισσότερο στην τιμή του κινητού μέσου όρου. Επίσης, όταν συμπεριλαμβάνονται περισσότεροι περίοδοι στον κινητό μέσο (μεγαλύτερες τιμές του  $n$ ), τόσο λιγότερο ευαίσθητος μπορεί να γίνει ο κινητός μέσος όρος στις μεταβολές του προτύπου της κυκλικής κίνησης. Αντιθέτως, μια μικρή τιμή του  $n$ , οδηγεί σε έναν κινητό μέσο που ανταποκρίνεται πολύ γρήγορα στις μεταβολές.

Τέλος, εάν κάποιος ήθελε να προβλέψει περιόδους πριν την τρέχουσα περίοδο  $t$ , η εξίσωση πρόβλεψης γίνεται

$$\hat{Z}_{t+1} = M_t = \frac{W_t Z_t + W_{t-1} Z_{t-1} + W_{t-2} Z_{t-2} + \dots + W_{t-n+1} Z_{t-n+1}}{W_t + W_{t-1} + W_{t-2} + \dots + W_{t-n+1}} \quad (1.36)$$

όπου  $\hat{Z}$  είναι η προβλεφθείσα τιμή του  $Z$  στο σημείο  $t+1$ .

Για να περιγράψουμε τη χρησιμότητα των κινητών μέσων όρων, θα επιστρέψουμε στο παράδειγμα 1.11 και στον Πίνακα 1.15, που το συνοδεύει. Στο παράδειγμα, η διαδικασία εξομάλυνσης θα είναι ένας σταθμισμένος κινητός μέσος με συντελεστές στάθμισης 1.2 και 1. Η τρέχουσα περίοδος έχει διπλό συντελεστή στάθμισης σε σύγκριση με την προηγούμενη και την περίοδο που ακολουθεί. Για παράδειγμα, για τη δεύτερη περίοδο του δεύτερου τριμήνου του 1971, ο σταθμισμένος συνολικός κινητός μέσος 727.9 βρίσκεται με την πρόσθεση των γινομένων των συντελεστών στάθμισης και των τιμών της στήλης (3). Με άλλα λόγια:

$$727.9 = (1)(201.0) + (2)(182.9) + (1)(161).$$

Με τη σειρά της, η κυκλική συνιστώσα υπολογίζεται από τη διαίρεση του σταθμισμένου συνολικού κινητού μέσου των τριών τριμήνων (στήλη 4) διά του αθροίσματος των συντελεστών στάθμισης (4). Μια κυκλική συνιστώσα 182.0 δείχνει ότι οι κυκλικές δυνάμεις αυξάνουν την οικονομική δραστηριότητα κατά έναν συντελεστή 82 τοις εκατό πάνω από το επίπεδο της τάσης. Εάν δεν ήταν παρούσα καμιά κυκλική συνιστώσα, το επίπεδο της τάσης θα ήταν μικρότερο. Από την άλλη, μια κυκλική συνιστώσα 72.5 (πρώτο τρίμηνο, (976) προϋποθέτει ότι τα πραγματικά έξοδα για βιομηχανικές εγκαταστάσεις και εξοπλισμό συμπίεστηκαν κατά 27.5 τοις εκατό, εξαιτίας της καθοδικής στροφής της οικονομικής δραστηριότητας.

Η γνώση της συνεισφοράς κάθε μιας από τις συστηματικές συνιστώσες, εποχική, τάση και κυκλικοί παράγοντες, που επιτυγχάνεται με την κλασική ανάλυση (classical decomposition) επιτρέπει στον προβλέποντα την επίτευξη του κυριότερου σκοπού της ανάλυσης χρονοσειρών. Ο σκοπός είναι η εξήγηση της διακύμανσης στα δεδομένα των παρατηρήσεων διαχρονικά.

### 1.2.9 Αποτελέσματα Ανάλυσης Χρονοσειρών και Πρόβλεψη

Η παράγραφος αυτή ασχολείται με μια μέθοδο για τη χρησιμοποίηση των αποτελεσμάτων της ανάλυσης στην πρόβλεψη των μελλοντικών τιμών της σειράς. Η έκταση της περιγραφής περιορίζεται στη χρήση των αποτελεσμάτων που αποκτήθηκαν κατά την ανάλυση του πολλαπλασιαστικού μοντέλου.

#### Παράδειγμα 1.12

Τον Ιανουάριο του 1986, στον υπεύθυνο κατάρτισης του προϋπολογισμού μιας συγκεκριμένης πολιτείας ανατέθηκε το έργο της πρόβλεψης των εισπράξεων, που θα προκύψουν από μια αύξηση του Φόρου επί των πωλήσεων στην πολιτεία για το 1987. Προκειμένου να κάνει την πρόβλεψη χρησιμοποίησε τριμηνιαία εξίσωση τάσης-κύκλου και εποχικούς δείκτες, που προμηθεύτηκε από την υπηρεσία του. Η εξίσωση και οι δείκτες δημιουργήθηκαν με τη χρήση δεδομένων από τη συλλογή

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 1ο

φόρων επί των πωλήσεων της κομητείας για τα προηγούμενα 15 χρόνια. Τα δεδομένα προσαρμόστηκαν ώστε να ελαχιστοποιηθούν οι διαφορές στην αναλογία του ρου επί των πωλήσεων για το χρονικό διάστημα των 15 ετών. Η εξίσωση ήταν:

$$\hat{Y}_t = \$14.5 + 1.5 X_t$$

όπου όλοι οι αριθμοί αντιπροσωπεύουν εκατομμύρια δολάρια και  $X=1$  που αντιστοιχεί στην πρώτη περίοδο του 1971. Εφόσον για το πρώτο τρίμηνο του 1987 είναι  $X=65$ , το δεύτερο τρίμηνο είναι  $X=66$  και ούτω καθ' εξής. Επίσης, οι δείκτες συλλογής των φόρων επί των πωλήσεων σε τριμηνιαία βάση ήταν:

Πρώτο τρίμηνο	85
Δεύτερο τρίμηνο	95
Τρίτο τρίμηνο	108
Τέταρτο τρίμηνο	112

Ο υπεύθυνος κατάρτισης του προϋπολογισμού για την ανάπτυξη της πρόβλεψης προχώρησε ως εξής:

$$\text{Πρόβλεψη} = T_t \times C_t \times \frac{S_t}{100}$$

Στη συνέχεια, συνδυάζει την παραπάνω έκφραση της πρόβλεψης με την εξίσωση τάσης -κύκλου που αναπτύχθηκε από το προσωπικό της διεύθυνσης προϋπολογισμού, ώστε να φτάσει στον παρακάτω τύπο

$$\hat{Y}_t = \frac{(14.5 + 1.5X_t)S}{100}$$

Τα προβλεφθέντα έσοδα από τους φόρους επί των πωλήσεων ανά τρίμηνο είναι:

$$\hat{Y}_{65} = \frac{(14.5 + 1.5 \times 65)85}{100} = 95200$$

$$\hat{Y}_{66} = \frac{(14.5 + 1.5 \times 66)95}{100} = 107825$$

$$\hat{Y}_{67} = \frac{(14.5 + 1.5 \times 67)108}{100} = 124200$$

$$\hat{Y}_{68} = \frac{(14.5 + 1.5 \times 68)112}{100} = 130480$$

Το άθροισμα των τριμηνιαίων προβλέψεων αποτελεί την ετήσια πρόβλεψη (\$ εκατομμύρια) της συλλογής των φόρων επί των πωλήσεων για το έτος 1987, που είναι \$ 457705 εκατομμύρια.

Η παραπάνω μέθοδος πρόβλεψης περιλαμβάνει τη χρησιμοποίηση μιας προβολής κύκλου-τάσης και εποχικών δεικτών. Η μέθοδος αυτή μπορεί να χρησιμοποιηθεί για προβλέψεις πολύ μακριά στο μέλλον. Εν τούτοις, τέτοιες προβολές γίνονται πιο αναξιόπιστες όσο πιο μακριά προχωρούμε στο μέλλον. Προβολές μεγαλύτερες των 5 μελλοντικών ετών μπορούν να γίνουν μόνο για εκείνες τις χρονοσειρές που εμφανίζουν υψηλή σταθερότητα και κανονικότητα. Όσο μεγαλύτερος είναι ο χρόνος ανάμεσα στην παρούσα και τη μελλοντική περίοδο τόσο λιγότερη εμπιστοσύνη πρέπει να έχουμε αναφορικά με την αξιопιστία μιας πρόβλεψης.

**ΚΕΦΑΛΑΙΟ 2<sup>ο</sup> :ΑΙΤΙΟΚΡΑΤΙΚΕΣ ΜΕΘΟΔΟΙ ΠΡΟΒΛΕΨΗΣ****2.1. ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ ΚΑΙ ΣΥΣΧΕΤΙΣΗ [7], [8], [14]****2.1.1 Εισαγωγή**

Υπάρχουν πολλοί λόγοι που συνηγορούν στη μελέτη των μεθόδων της παλινδρόμησης. Για παράδειγμα, ένας λόγος είναι για να μελετήσουμε τη φύση της σχέσης που υπάρχει ανάμεσα σε μια *εξαρτημένη* και σε μια *ανεξάρτητη* μεταβλητή. Έστω ότι η εξαρτημένη μεταβλητή  $Y$ , είναι η μεταβλητή την οποία θέλουμε να προβλέψουμε και θα μπορούσε να είναι η αξία σε δολάρια των ενυπόθηκων δανείων που χορηγούνται από μια κτηματική τράπεζα. Η ανεξάρτητη μεταβλητή,  $X$ , αυτή που χρησιμοποιείται για να μας βοηθήσει στην πρόβλεψη, θα μπορούσε να ήταν το μέσο ύψος του επιτοκίου για ενυπόθηκα δάνεια. Σκοπός μας είναι να προβλέψουμε τα ποσά των δανείων που θα ζητηθούν, βασιζόμενοι στη μετρηθείσα σχέση ανάμεσα σε αυτό το μέγεθος και το ύψος του επιτοκίου.

Σκοπός της γραμμικής συσχέτισης είναι να απαντήσει στην ερώτηση που αφορά το βαθμό της γραμμικής εξάρτησης ανάμεσα στις δυο μεταβλητές. Μολονότι μπορεί να υπάρχουν ευλογοφανείς αιτίες για μια τέτοια σχέση, η συσχέτιση δεν λέει τίποτα για αυτές τις αιτίες. Δείχνει μόνο το αν οι δύο μεταβλητές συμμεταβάλλονται. Αυτό, ίσως να συμβαίνει γιατί η μία μεταβλητή είναι η αιτία και η άλλη το αποτέλεσμα, ή ίσως και οι δυο μαζί μεταβάλλονται συγχρόνως ως αποτέλεσμα μιας τρίτης μεταβλητής που αποτελεί την αιτία. Η συσχέτιση ωστόσο δείχνει μόνο το βαθμό της εξάρτησης.

Όταν υπάρχουν περισσότερες από μία ανεξάρτητες μεταβλητές, τέτοιες σχέσεις ονομάζονται *πολλαπλές* συσχετίσεις. Σε αυτό το κεφάλαιο, περιορίζουμε τη μελέτη μας στις *απλές* σχέσεις που περιλαμβάνουν μόνο μια ανεξάρτητη μεταβλητή,  $X$ , και μια εξαρτημένη,  $Y$ . Ακόμη, σ' αυτό το κεφάλαιο θα μελετήσουμε μόνο τις σχέσεις που εκφράζονται από μια ευθεία γραμμή.

**2.1.2 Το Μοντέλο της Απλής Γραμμικής Παλινδρόμησης**

Το απλό παλινδρομικό μοντέλο εκφράζει τη σχέση ανάμεσα σε μια μεταβλητή  $Y$ , δηλαδή ενυπόθηκα δάνεια σε σταθερές τιμές και μια μεταβλητή  $X$ , δηλαδή το μέσο επίπεδο του επιτοκίου, σε μορφή ευθείας γραμμής. Η γραμμή παλινδρόμησης στον πληθυσμό της  $Y$  με τη  $X$  είναι η συσχέτιση των ενυπόθηκων δανείων με το εκάστοτε ύψος του επιτοκίου. Η παλινδρόμηση προϋποθέτει ότι ικανοποιείται το σύνολο των υποθέσεων στον πληθυσμό στον οποίο ανήκουν οι δύο μεταβλητές, προκειμένου να χρησιμοποιηθεί γραμμικό παλινδρομικό μοντέλο.

Η *πρώτη υπόθεση* είναι ότι η σχέση περιγράφεται από μια ευθεία γραμμή, δηλαδή είναι *γραμμική*. Από μαθηματική άποψη το γραμμικό παλινδρομικό μοντέλο μπορεί να εκφραστεί ως εξής:

$$Y = \beta_0 + \beta_1 X + e \quad (2.1)$$

Σε αυτή την εξίσωση το  $e$  είναι το τυχαίο σφάλμα που μετράει την κάθετη απόκλιση της κάθε τιμής της εξαρτημένης μεταβλητής από την πληθυσμιακή παλινδρομική γραμμή για την αντίστοιχη τιμή της ανεξάρτητης μεταβλητής.

Προσπαθώντας να προβλέψουμε το ύψος των ενυπόθηκων δανείων για ένα συγκεκριμένο επίπεδο τόκου, η πρόβλεψή μας υπόκειται σε σφάλμα. Αυτό το σφάλμα μπορεί να προέλθει από μία ή περισσότερες πιθανές αιτίες:

1. *Ατελής θεωρία*: Συγκεκριμένες μεταβλητές που σχετίζονται με τη συμπεριφορά της  $Y$  μπορεί να έχουν παραλειφθεί. Μια περισσότερο πλήρης θεωρία ίσως απαιτούσε να συμπεριληφθούν πληροφορίες αναφορικά με τα εισοδήματα των καταναλωτών ή τις συνθήκες λειτουργίας της αγοράς κατοικιών. Επομένως, θα μπορούσαν να προστεθούν περισσότερες ανεξάρτητες μεταβλητές στις παλινδρομικές σχέσεις.

2. *Ατελής Προσδιορισμός*: Πρόκειται για την περίπτωση που εκτιμάται μια γραμμική σχέση, ενώ η πραγματική σχέση ίσως δεν είναι γραμμική, αλλά εκφράζεται καλύτερα με την εξίσωση μιας παραβολής.

3. *Σφάλματα μέτρησης*: Ακόμη κι όταν η σχέση προετοιμάστηκε συστηματικά μπορούν να συμβούν υπολογιστικά σφάλματα κατά τη μέτρηση των μεταβλητών. Επομένως, το πείραμα της πρόβλεψης πρέπει να περιλαμβάνει ένα σύνολο «ελεγκτών» ώστε να περιορίζονται τα υπολογιστικά και τα μη-στατιστικά σφάλματα.

Βασικά, το μοντέλο παραδέχεται ότι αν οι αιτίες που δημιουργούν τα τυχαία σφάλματα είναι ελεγχόμενες, θα ήταν δυνατό να προβλέψουμε με ακρίβεια το ύψος του δανείου όταν είναι γνωστό το ύψος του επιτοκίου. Όταν όλοι οι παραπάνω παράγοντες συμβάλλουν στην εμφάνιση του σφάλματος, εμπλέκονται θεωρητικά με διαφορετικές επιπτώσεις στο καθορισμένο μοντέλο.

Η *δεύτερη υπόθεση* του μοντέλου είναι ότι θα εμφανιστεί η ίδια κατανομή των τιμών της  $X$  σε οποιαδήποτε επανάληψη του πειράματος. Όμως, η κατανομή των τιμών της  $Y$  για δεδομένες τιμές της  $X$  μπορεί να διαφέρει από πείραμα σε πείραμα εξαιτίας της επίδρασης του όρου του τυχαίου σφάλματος. Έτσι, για την απεικόνιση του σχήματος 4.1 είχαμε 10 παρατηρήσεις ποσών δανείου που κυμαίνονται από 8 έως 16 τοις εκατό. Εάν επαναλαμβανόταν το πείραμα, θα είχαμε και πάλι 10 παρατηρήσεις ενυπόθηκων δανείων από 8 έως 16 τοις εκατό. Η διασπορά των τιμών της  $Y$ , εντούτοις, θα μεταβαλλόταν από πείραμα σε πείραμα. Οι τιμές της ανεξάρτητης μεταβλητής,  $X$ , λέμε ότι έχουν προκαθοριστεί, ενώ η εξαρτημένη μεταβλητή,  $Y$ , είναι μια τυχαία μεταβλητή.

Οι τιμές προκαθορισμένων μεταβλητών καθορίστηκαν ανεξάρτητα από το τρέχον μοντέλο. Οι τιμές αυτές, όπως το εύρος των ποσοστών του επιτοκίου, μπορεί να καθοριστούν πριν την τρέχουσα χρονική περίοδο κατά την οποία διεξήχθη το πείραμα. Το πιθανό εύρος τιμών για τα δάνεια καθορίστηκε όταν σχεδιάστηκε το τεστ. Εξάλλου, τα δεδομένα των ενυπόθηκων δανείων θα μπορούσαν να καθοριστούν όταν διεξήχθη το πείραμα, ώστε τα δάνεια και το επιτόκιο να γίνουν αντικείμενο παρατήρησης σε ένα δείγμα οικονομικών οργανισμών. Η εξαρτημένη μεταβλητή,  $Y$ , σχετίζεται με τις τιμές της προκαθορισμένης, μεταβλητής, αλλά η προκαθορισμένη μεταβλητή,  $X$ , δεν σχετίζεται με τις τιμές της εξαρτημένης μεταβλητής,  $Y$ .

Η *τρίτη υπόθεση* είναι ότι η αναμενόμενη τιμή του τυχαίου σφάλματος είναι μηδέν και εκφράζεται ως εξής:

$$E(e)=0 \quad (2.2)$$

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

$$E(Y/X)=E(\beta_0+\beta_1X+e) \quad (2.3)$$

μπορεί να αναχθεί στο άθροισμα των αναμενόμενων τιμών:

$$E(Y/X) = E(\beta_0) + E(\beta_1X) + E(e) \quad (2.4)$$

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

$$E(Y/X)=\beta_0+\beta_1X \quad (2.5)$$

εφόσον η αναμενόμενη τιμή μιας σταθεράς είναι μια σταθερά και εφόσον  $E(e)$  είναι μηδέν. Αυτή η προσδοκία είναι ο *υπό συνθήκη μέσος* γιατί μετράει τον μέσο ή τη μέση τιμή της  $Y$  που σχετίζεται με μια συγκεκριμένη τιμή της  $X$ . Η τρίτη υπόθεση συνεπάγεται ότι για μια δεδομένη τιμή της  $X$ , ο μέσος των τιμών  $Y$  βρίσκεται πάνω στη γραμμή παλινδρόμησης.

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

$$VAR(e) = E[e - E(e)]^2 = E(e^2) = R \quad (2.6)$$

όπου  $R$  είναι ένας σταθερός όρος. Αλλά από το μοντέλο έχουμε

$$e = Y - (\beta_0 + \beta_1X) \quad (2.7)$$

Αυτή η υπόθεση, παραπέρα, εισάγει ότι  $E[Y - (\beta_0 + \beta_1X)]^2 = R$ , ή ότι η διακύμανση και η σταθερή απόκλιση είναι ίδιες για κάθε τιμή  $X$  στον πληθυσμό. Η  $VAR(e)$  μετράει τη μεταβλητότητα στην τιμή της εξαρτημένης μεταβλητής,  $Y$ , περί τη γραμμή παλινδρόμησης για δεδομένες τιμές της ανεξάρτητης μεταβλητής,  $X$ .

Η *πέμπτη υπόθεση* είναι ότι οι τιμές του  $e$  είναι ανεξάρτητες μεταξύ τους, δηλαδή  $E(e_i, e_j) = 0$ . Αυτό



σημαίνει ότι η τιμή του σφάλματος για οποιαδήποτε δεδομένη τιμή της ανεξάρτητης μεταβλητής δεν σχετίζεται με το σφάλμα για οποιαδήποτε άλλη τιμή της  $X$ . Όταν τα σφάλματα αυτά σχετίζονται, τότε έχουμε το πρόβλημα της αυτοσυσχέτισης.

Η *έκτη υπόθεση* είναι ότι ο όρος του σφάλματος για κάθε τιμή της  $X$  κατανέμεται κανονικά. Από τη στιγμή που οι τιμές της  $Y$  για δεδομένες τιμές της  $X$  ποικίλλουν από πείραμα σε πείραμα μόνον εξαιτίας του όρου του τυχαίου σφάλματος (οι τιμές της  $X$  μένουν ίδιες από πείραμα σε πείραμα), οι τιμές της  $Y$  πρέπει αναγκαστικά να κατανέμονται κανονικά.

Για μεγαλύτερη κατανόηση της σημασίας των παραπάνω υποθέσεων, ας θεωρήσουμε την πρόβλεψη αποταμιευτικών καταθέσεων για μια οικογένεια.

### Παράδειγμα 2.1

Ένα κοινό πρόβλημα της εφαρμογής της οικονομικής θεωρίας για την κατανόηση της αποταμιευτικής συμπεριφοράς των οικογενειών, είναι το να συσχετίσουμε τις αποταμιευτικές καταθέσεις με το οικογενειακό εισόδημα. Για μια χώρα ή περιοχή και για μια καθορισμένη χρονική περίοδο, το απλό γραμμικό παλινδρομικό μοντέλο που συσχετίζει το οικογενειακό εισόδημα για μια τετραμελή οικογένεια,  $X$ , με τη μηνιαία συνεισφορά στις αποταμιευτικές καταθέσεις,  $Y$ , εκφράζεται ως εξής:

$$\mu_Y = -100 + 0,07X$$

όπου  $\mu_Y$  είναι ο μέσος ή ο κατά συνθήκη μέσος των τιμών της  $Y$  για δοσμένο  $X$ .

Η παραπάνω γραμμική παλινδρόμηση της  $Y$  πάνω στη  $X$  είναι γραμμική με κανονικά κατανεμημένες τιμές  $Y$  για δεδομένες τιμές της  $X$  και με ανεξάρτητο τυχαίο όρο σφάλματος που έχει μηδενικό μέσο και σταθερά τυπική απόκλιση. Εάν το οικογενειακό εισόδημα είναι \$ 2000 ανά μήνα ( $X = \$ 2000$ ), η μηνιαία αύξηση στις αποταμιεύσεις θα μπορούσε να είναι \$ 40 κι αυτό εκφράζεται συμβολικά ως εξής:

$$\mu_Y = -100 + 0.07(2000) = \$ 40$$

Όμοια, εάν το οικογενειακό μηνιαίο εισόδημα είναι \$ 3000, τότε

$$\mu_Y = -100 + 0.07(3000) = \$ 110$$

Αυτή η εφαρμογή του παλινδρομικού μοντέλου παριστάνει μια ευθεία γραμμή με ιδιαίτερη γεωμετρική ονομασία για τις παραμέτρους  $\beta_0$  και  $\beta_1$ . Στην παραπάνω εφαρμογή το  $\beta_0$  ονομάζεται συνήθως η τεταγμένη  $Y$  της εξίσωσης, εάν  $\beta_1 X = 0$ ,  $\mu_Y$  ισούται με  $\beta_0$  (-100 σε αυτή την εφαρμογή). Αυτή είναι η μηνιαία συνεισφορά στις αποταμιεύσεις, ανεξάρτητα από το επίπεδο του μηνιαίου εισοδήματος.

Στο παράδειγμα,  $\beta_1$  είναι ίσο με 0.07. Αυτή είναι η κλίση της γραμμής παλινδρόμησης ή η μεταβολή στις αποταμιεύσεις που σχετίζεται με ενός δολαρίου (μίας μονάδας) μεταβολή στο προσωπικό εισόδημα. Δηλαδή, κατά μέσον όρο, κάθε μεταβολή 100 δραχμών στο προσωπικό εισόδημα σχετίζεται με μια κατά 7 δραχμές αύξηση στις αποταμιεύσεις. Η σχέση λέγεται ότι είναι προς την ίδια κατεύθυνση και ότι η εξίσωση έχει θετική κλίση. Από την άλλη, εάν μια αύξηση σε μια από τις μεταβλητές σχετίζεται με μια μείωση στην άλλη μεταβλητή, η σχέση λέγεται ότι έχει *αρνητική κλίση*.

Πριν συνεχίσουμε, θα πρέπει να σημειώσουμε ένα ενδιαφέρον θέμα στην εφαρμογή της παλινδρομικής θεωρίας. Εάν μια υπόθεση δεν είναι αληθής, πώς μπορεί κάποιος να αναγνωρίσει το πρόβλημα; Μερικές φορές οι μέθοδοι είναι ισχυρές (robust) δηλαδή μια μικρή απόκλιση από τις παραδοχές δεν επιδρά σοβαρά στη χρήση της παλινδρόμησης. Από την άλλη όμως, υπάρχουν καταστάσεις κατά τις οποίες εμείς ως υπεύθυνοι των προβλέψεων πρέπει να προσαρμόσουμε τα αποτελέσματα ή να χρησιμοποιούμε άλλες τεχνικές για την πρόβλεψη. Πολλά από αυτά τα προβλήματα εξετάζουμε στις άλλες παραγράφους του κεφαλαίου, καθώς και στο επόμενο κεφάλαιο.

### 2.1.3 Εκτιμώντας το Μοντέλο Απλής Γραμμικής Παλινδρόμησης

Παρατηρήσεις σε ένα δείγμα  $b_0$  τιμών του  $Y$  για ένα καθορισμένο δείγμα τιμών της  $X$  μας επιτρέπει να υπολογίσουμε την εξίσωση μιας γραμμής της ακόλουθης μορφής

$$\hat{Y} = b_0 + b_1X \quad (2.8)$$

όπου  $b_0$ ,  $b_1$  είναι εκτιμήσεις των παραμέτρων  $\beta_0$  και  $\beta_1$  της παλινδρόμησης.  $\hat{Y}$  είναι η εκτιμημένη τιμή του  $\mu_Y$ , ο υπό συνθήκη μέσος της  $Y$  για δοσμένο  $X$ . Για να βρούμε τα  $\beta_0$ ,  $\beta_1$ , για ένα δοσμένο σύνολο τιμών του δείγματος, μπορούμε να χρησιμοποιήσουμε τους παρακάτω τύπους:

$$b_1 = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{n\Sigma X^2 - (\Sigma X)^2} \quad (2.9)$$

και

$$b_0 = \frac{(\Sigma Y)(\Sigma X^2) - (\Sigma X)(\Sigma XY)}{n(\Sigma X^2) - (\Sigma X)^2} \quad (2.10)$$

Συνήθως χρησιμοποιείται ένας πιο απλός τύπος για το  $b_0$ . Ο τύπος αυτός είναι:

$$b_0 = \frac{\Sigma Y - b_1\Sigma X}{n} \quad (2.11)$$

Παράδειγμα 2.2

Τα δεδομένα του δείγματος του Πίνακα 2.1 αφορούν μια μελέτη της σχέσης των ενυπόθηκων δανείων  $Y$ , της Buckeye Savings and Loan Association, και του συμφωνηθέντος επιτοκίου για αγορά κατοικιών.

**ΠΙΝΑΚΑΣ 2.1.** Υπολογισμοί για την παλινδρόμηση

Επιτόκιο $X$	Ενυπόθηκα δάνεια (\$ εκατομμ) $Y$	$X^2$	$XY$	$Y$
7.60	7.50	57.760	57.000	56.250
7.45	7.44	55.502	55.428	55.354
7.78	7.93	60.528	61.695	62.885
8.71	8.41	75.864	73.251	70.728
8.75	9.21	76.563	80.587	84.824
8.76	12.03	76.738	105.383	144.721
8.80	14.33	77.440	126.104	205.349
9.30	16.25	86.490	151.125	264.063
10.48	17.99	109.830	188.535	323.640
12.25	21.73	150.063	266.192	472.193
89.88	122.82	826.788	1165.300	1740.007
$\Sigma X$	$\Sigma Y$	$\Sigma X^2$	$\Sigma XY$	$\Sigma Y^2$

Για να εκτιμήσουμε τη γραμμή παλινδρόμησης με αυτά τα δεδομένα, κάνουμε τους υπολογισμούς που εμφανίζονται στον πίνακα. Μετά από αντικατάσταση των σχετικών μεγεθών στους τύπους, βρίσκουμε:

$$b_1 = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{n\Sigma X^2 - (\Sigma X)^2} = \frac{10(1165.300) - (89.88)(122.82)}{10(826.778) - (89.88)^2} = 3.242$$

Αυτός ο συντελεστής παλινδρόμησης υποδηλώνει ότι κατά μέσον όρο αύξηση μιας εκατοστιαίας μονάδας στο επιτόκιο σχετίζεται με μια αύξηση \$ 3242 εκατομμυρίων των αποταμιευτικών καταθέσεων της Buckeye Savings and Loan. Επίσης, ο συντελεστής παλινδρόμησης για την τεταγμένη επί την αρχική  $Y$ , βρίσκεται από:

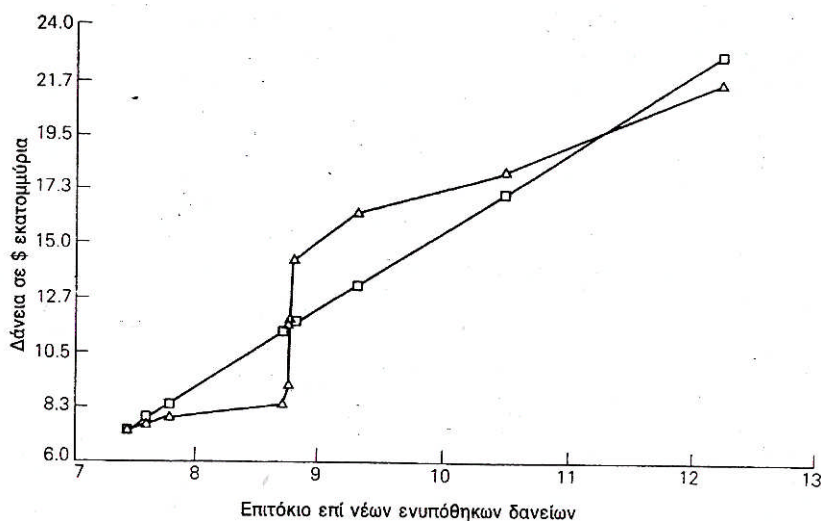
$$b_0 = \frac{\Sigma Y - b_1 \Sigma X}{n} = \frac{122.82 - 3.242(89.88)}{10} = -16.859$$

Η εκτιμηθείσα εξίσωση παλινδρομικής είναι:

$$\hat{Y} = -16.859 + 3.242X$$

Δεδομένου ότι οι τιμές της  $X$  κυμαίνονται μεταξύ 7.45 και 12.25 στο δείγμα, η σταθερά δεν αντιπροσωπεύει αναγκαστικά τις αποταμιευτικές καταθέσεις όταν το ύψος του επιτοκίου ευρίσκεται έξω από το παραπάνω διάστημα. Δεν έχουμε κανένα δείγμα παρατηρήσεων  $Y$  για την προκαθορισμένη τιμή  $X=0$  και, επομένως, δεν μπορούμε να συμπεράνουμε ότι η παλινδρομική εξίσωση ισχύει σ' αυτό το σημείο. Εάν εξετάσουμε το σχήμα 2.1, παρατηρούμε ότι η γραμμή παλινδρόμησης είναι η καλύτερη δυνατή εκτίμηση που μπορεί να γίνει για το μέγεθος των ενυπόθηκων δανείων, βασισμένη στις παραδοχές του απλού γραμμικού παλινδρομικού μοντέλου.

**Σχήμα 2.1** Ενυπόθηκα δάνεια και ύψος επιτοκίου (παρατηρηθέντα και προβλεφθέντα). Οι προβλεφθείσες τιμές σχηματίζουν μια ευθεία γραμμή



Σε αυτή την παράγραφο, ασχοληθήκαμε μόνο με το πρόβλημα της προσαρμογής μιας ευθείας γραμμής στα ζεύγη δεδομένων. Οι μέθοδοι της ανάλυσης της παλινδρόμησης μπορούν επίσης να χρησιμοποιηθούν για την εκτίμηση άλλων ειδών εξισώσεων, καθώς και για να εκτιμηθούν εξισώσεις πρόβλεψης με περισσότερους από δύο αγνώστους.

#### 2.1.4 Τυπικό Σφάλμα Παλινδρόμησης

Το τυπικό σφάλμα παλινδρόμησης μας δίνει μια εκτίμηση της ακρίβειας της εξίσωσης πρόβλεψης. Παρατηρούμε από το σχήμα 4.1 πως οι πραγματικές τιμές της  $Y$  βρίσκονται περί τη γραμμή

πρόβλεψης. Όσο μεγαλύτερη είναι η διασπορά των τιμών της  $Y$  γύρω από τη γραμμή παλινδρόμησης, τόσο μεγαλύτερη είναι η πιθανότητα σφάλματος κατά τη διάρκεια προβλέψεων με αυτή την εξίσωση.

Αν θεωρήσουμε ένα απλό γραμμικό παλινδρομικό μοντέλο σημειώνουμε ότι:

$$Y = \beta_0 + \beta_1 X + e \quad (2.12)$$

και

$$\mu_Y = \beta_0 + \beta_1 X \quad (2.13)$$

Η διαφορά ανάμεσα στην πραγματική τιμή του  $Y$  και τον υπό συνθήκη μέσο  $Y$  είναι ο όρος του σφάλματος  $e$ . Το τυπικό σφάλμα της παλινδρόμησης που μετράει τη διασπορά των παρατηρηθεισών τιμών της  $Y$  γύρω από τη γραμμή παλινδρόμησης ορίζεται από:

$$S_e = \left( \frac{\sum e^2}{n-2} \right)^{1/2} \quad (2.14)$$

Το άθροισμα των τετραγώνων των αποκλίσεων διαιρείται δια  $(n-2)$  διότι αυτός ο διαιρέτης καθιστά το  $S_e$  αμερόληπτο εκτιμητή της τυπικής απόκλισης γύρω από την πραγματική πληθυσμιακή γραμμή παλινδρόμησης. Γενικά, ο παρονομαστής είναι  $(n-k)$  όπου  $k$  είναι ο αριθμός των προς εκτίμηση συντελεστών (παραμέτρων) στην παλινδρομική εξίσωση. Τέλος, σημειώστε ότι αυτός ο τύπος είναι παρόμοιος με τον τύπο του μέσου τετραγωνικού σφάλματος (MSE) της παρ 1.2.

Μια κατάλληλη μορφή ένας απλούστερος τύπος για τον υπολογισμό του σφάλματος της παλινδρόμησης είναι:

$$S_e = \left( \frac{\sum Y^2 - b_0 \sum Y - b_1 \sum XY}{n-2} \right)^{1/2} \quad (2.15)$$

Όλες οι ποσότητες που απαιτούνται για την εφαρμογή του τύπου υπολογίστηκαν για το παράδειγμα 2.2. Επομένως:

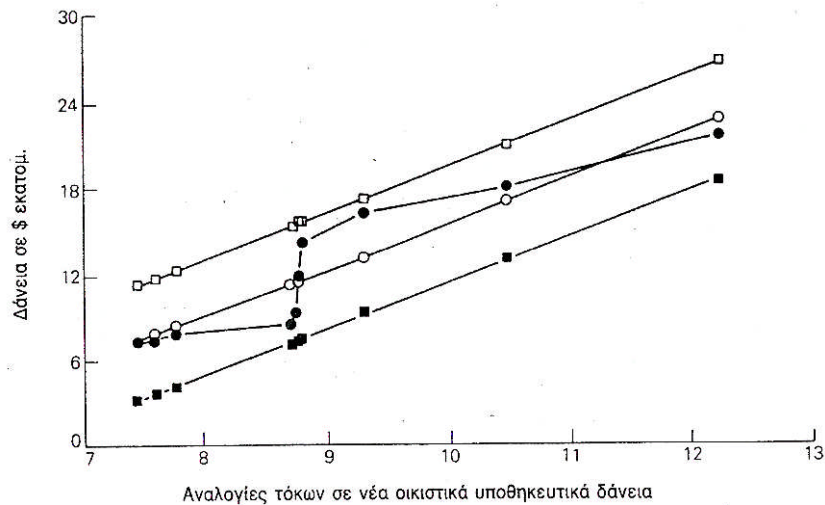
$$S_e = \left( \frac{1740.007 - (-16.859)(122.82) - (3.242)(1165.300)}{8} \right)^{1/2} = 2.015$$

Όπως συνάγεται από το όνομά του, το  $S_e$  είναι ερμηνεύσιμο ως ο βαθμός τυπικού σφάλματος πάνω σε μια κανονική καμπύλη. Κατά συνέπεια, ερμηνεύεται σε όρους τιμών πιθανοτήτων. Εναλλακτικά, θα μπορούσαμε να αναμένουμε ότι το 95 τοις εκατό των τιμών της  $Y$  θα πέσει εντός του διαστήματος, που προσδιορίζεται από δύο συν ή πλην τυπικά σφάλματα της παλινδρομικής γραμμής.

Ένας εύκολος τρόπος να κατανοήσουμε καλύτερα το τυπικό σφάλμα της παλινδρόμησης είναι να το παραστήσουμε γραφικά με μια παλινδρομική γραμμή.

Σχήμα 2.2 Παρατηρηθέντα, προβλεφθέντα και δύο όρια τυπικών σφαλμάτων.

Προβλεφθείσες τιμές και όρια σχηματίζουν ευθείες γραμμές.



### 2.1.5 Στατιστική Σημαντικότητα της Παλινδρόμησης

Υπάρχει ένδειξη γραμμικής συσχέτισης ανάμεσα στην εξαρτημένη μεταβλητή,  $Y$ , και την ανεξάρτητη μεταβλητή,  $X$ ; Εάν δεν υπάρχει, τότε η τιμή του συντελεστή κλίσης,  $\beta$ , θα είναι μηδέν. Εναλλακτικά, το ερώτημα μπορεί να τεθεί στο εάν υπάρχει η ελάχιστη απόδειξη ώστε να εναντιωθεί κανείς στην υπόθεση ότι, δεν υφίσταται συσχέτιση μεταξύ των  $X$  και  $Y$ , οπότε λέμε ότι οι μεταβλητές δεν σχετίζονται μεταξύ τους. Προς τούτο, θεωρούμε το παρακάτω παράδειγμα.

#### Παράδειγμα 2.3

Προκειμένου να προσδιοριστεί εάν η κλίση της γραμμής της παλινδρόμησης των ενυπόθηκων δανείων και του επιπέδου του επιτοκίου είναι στατιστικά σημαντική, πρέπει να καθορίσουμε το εάν η κλίση είναι μηδέν ή όχι. Εάν παραδεχθούμε ότι υπάρχει ρίσκο 5 τοις εκατό ( $\alpha=0.05$ ), οι κριτικές τιμές του  $t$  για οκτώ ( $n-2$ ) βαθμούς ελευθερίας είναι  $+2.306$  ή  $-2.306$ . Επομένως ο κανόνας απόφασης είναι:

Απορρίπτεται η υπόθεση  $B=0$  εάν το  $t$  που βασίζεται σε στοιχεία δείγματος είναι μεγαλύτερο από  $-2.306$  ή μικρότερο από  $+2.306$ , διαφορετικά το δεχόμαστε.

Η τυπική απόκλιση του συντελεστή κλίσης είναι:

$$S_b = \frac{S_e}{\{\sum X^2 - [(\sum X)^2 / n]\}^{1/2}} = \frac{2.015}{\{(826.778) - [(89.88)^2 / 10]\}^{1/2}} = 0.463$$

Εφόσον η υπόθεση που θα ερευνηθεί είναι:

$$H_0: \beta_1 = 0$$

$$H_A: \beta_1 \neq 0$$

έχουμε

$$t = \frac{b_1 - \beta_1}{S_{b_1}} = \frac{3.242 - 0}{0.463} = 7.00$$

Εφόσον η υπολογισθείσα τιμή του  $t$  είναι 2.306, η μηδενική υπόθεση απορρίπτεται στο επίπεδο σημαντικότητας 0.05. Επομένως, το  $\beta_1$  δεν ισούται με μηδέν. Κατά συνέπεια, υπάρχει μια σημαντική σχέση εξάρτησης ανάμεσα στα ενυπόθηκα δάνεια και το ύψος του επιτοκίου στις υποθήκες νέων κατοικιών.

Υπάρχει μια εναλλακτική στρατηγική για τον έλεγχο της στατιστικής σημαντικότητας μιας απλής παλινδρομικής ανάλυσης. Το εναλλακτικό κριτήριο είναι η τιμή του F-τεστ που ονομάζεται ανάλυση της διακύμανσης για την παλινδρόμηση. Ο λόγος  $F$  αποκαλύπτει πως η διακύμανση σε μια ομάδα μπορεί να είναι η ίδια ή διαφορετική από τη διακύμανση μιας άλλης ομάδας. Με δοσμένη την υπολογισμένη τιμή του  $F$  και τους βαθμούς ελευθερίας που σχετίζονται με αυτή, μπορούμε να εκτιμήσουμε την πιθανότητα όπου η  $F$  εμφανίζεται τυχαία. Επομένως, μπορούμε να ελέγξουμε την υπόθεση με αυτή την  $F$ .

Στην παλινδρομική ανάλυση, η  $F$  ορίζεται:

$$\frac{\text{Άθροισμα τετραγώνου που σχετίζονται με την παλινδρόμηση} \div \text{βαθμοί ελευθερίας}}{\text{Άθροισμα τετραγώνου που δεν σχετίζονται με την παλινδρόμηση} \div \text{βαθμοί ελευθερίας}}$$

Αν η διακύμανση που σχετίζεται με την παλινδρόμηση υπερτερεί της διακύμανσης που δεν σχετίζεται με την παλινδρόμηση, το λιγότερο που θα περιμέναμε θα ήταν η τιμή του  $F$  να έχει προκύψει κατά τύχη.

Όπως σε όλα τα F-τεστ, οι συνιστώσες της  $F$  προκύπτουν από άθροισμα τετραγώνων. Ο αριθμητής είναι το άθροισμα των τετραγώνων των αποκλίσεων της προβλεφθείσας μεταβλητής γύρω από τον μέσο της, δηλαδή το άθροισμα των τετραγώνων που οφείλονται στην παλινδρόμηση. Ο παρονομαστής είναι το άθροισμα των τετραγώνων των αποκλίσεων των διαφορών ανάμεσα στις πραγματικές και τις προβλεφθείσες τιμές, δηλαδή το άθροισμα των τετραγώνων του όρου του σφάλματος.

Η υπόθεση μας τίθεται ως εξής:

$H_0$ : Δεν υπάρχει ένδειξη μιας σημαντικής γραμμικής σχέσης (παλινδρόμησης)

$H_A$ : Υπάρχει ένδειξη μιας σημαντικής γραμμικής σχέσης (παλινδρόμησης)

Εάν θέλουμε με πιθανότητα 5 τοις εκατό να απορρίψουμε την αληθινή  $H_0$  χωρίς καμία ένδειξη παλινδρόμησης, η κρίσιμη τιμή της  $F$  είναι 5.32 (από Πίνακα τιμών του F). Τώρα μπορούμε να υπολογίσουμε την  $F$  για το πρόβλημα του ενυπόθηκου δανείου:

$$F = \frac{\sum (\hat{Y}_i - \bar{Y})^2 / (m-1)}{\sum (Y_i - \hat{Y})^2 / (n-m)} \quad m = \text{αριθμός παραμέτρων} \quad (2.16)$$

Σημειώστε ότι ο αριθμητής της εξίσωσης είναι το άθροισμα των τετραγώνων των αποκλίσεων της γραμμής παλινδρόμησης από τον μέσο όρο της τιμής  $Y$ , και ο παρονομαστής είναι το άθροισμα των τετραγώνων των αποκλίσεων των πραγματικών τιμών  $Y$  από την αντίστοιχη θέση στη γραμμή παλινδρόμησης. Για να το θέσουμε διαφορετικά το θέμα, ολόκληρος ο παρονομαστής είναι το τετράγωνο του σφάλματος της παλινδρόμησης. Η  $F$ -αναλογία μπορεί να ξαναγραφεί ως εξής:

$$F = \frac{b_1^2 (\sum X_i^2 - n\bar{X}^2)}{S_e^2} \quad (2.17)$$

Για τα δεδομένα του παραδείγματος 2.3, οι αντικαταστάσεις που γίνονται στον παραπάνω τύπο είναι οι εξής:

$$F = \frac{(3.242)^2 (826.795 - (10)(8.988)^2)}{(2.015)^2} = 49.03$$

Εφόσον η  $F$  που υπολογίστηκε από τα δεδομένα είναι μεγαλύτερη από 5.32, την κριτική τιμή της  $F$ , η ανάλυση δείχνει ότι η  $H_0$  δεν πρέπει να γίνει αποδεκτή και, κατά συνέπεια, υπάρχει ένδειξη ότι οι δύο μεταβλητές συσχετίζονται.

Τα  $F$ -τεστ των παραπάνω τύπων είναι ιδιαίτερα χρήσιμα στα μοντέλα παλινδρόμησης που έχουν περισσότερες από μία ανεξάρτητες μεταβλητές. Για παράδειγμα, με τη βοήθεια των  $F$ -τεστ μπορεί να εξακριβωθεί εάν οι επιπρόσθετες ανεξάρτητες μεταβλητές συνεισφέρουν σημαντικά στην ερμηνεία της συμπεριφοράς της  $Y$ .

Τέλος, σημειώνουμε ότι τα τεστ  $t$  και  $F$ , όπως συζητήθηκαν σε αυτή την παράγραφο, αποτελούν δύο τρόπους αντιμετώπισης του ίδιου πράγματος. Προϋποθέτουν τη διαίρεση της διακύμανσης σε δύο συνιστώσες, που αντιπροσωπεύουν τη διακύμανση που σχετίζεται με την παλινδρόμηση και τη διακύμανση που δεν σχετίζεται με την παλινδρόμηση. Παραπέρα, μας επιτρέπουν να εκτιμήσουμε την πιθανότητα ότι η διακύμανση που σχετίζεται με την παλινδρόμηση είναι μεγαλύτερη από εκείνη που θα προέκυπτε μόνο από τύχη. Από μαθηματική πλευρά, η  $F$ -στατιστική είναι ίση με το τετράγωνο του  $t$  και, επομένως, η παρατήρησή μας όσον αφορά την ομοιότητα στη χρήση τους έχει ισχυρή υποστήριξη.



### 2.1.6 Σημαντικότητα της Σταθεράς Παλινδρόμησης

Στην παράγραφο αυτή, θα ασχοληθούμε με τη σημαντικότητα της εκτίμησης  $\beta_0$  που υποδηλώνεται με  $b_0$  στην εξίσωση πρόβλεψης. Υπενθυμίζουμε ότι  $b_0 = -16859$ , όπως είδαμε στο παράδειγμα 2.3, που βασίζεται σε δειγματοληπτικά δεδομένα. Αν κατασκευάζαμε ένα εναλλακτικό δείγμα δέκα παρατηρήσεων, οι μέθοδοι παλινδρόμησης θα μπορούσαν ενδεχομένως να οδηγήσουν σε μια διαφορετική τιμή του  $b_0$ .

Στην καλύτερη περίπτωση η σταθερά  $Y$  (ή σταθερά της παλινδρόμησης) δίνει τη μέση τιμή  $Y$  σε προσδιορισμένη τιμή του  $X$ .  $X=0$ . Εν τούτοις, συχνά η  $\beta_0$  δεν έχει κάποια σαφή έννοια ή θέση μέσα στο πρόβλημα. Σε μερικά προβλήματα, εάν  $X=0$ , είναι λογικό και  $Y=0$ . Εάν ο  $X$  είναι ο αριθμός των μονάδων που πουλήθηκαν και ο  $Y$  παριστά τα έσοδα από τις πωλήσεις (τιμή επί πωληθείσες μονάδες), τότε είναι λογικό ότι το  $Y$  είναι μηδέν. Εξαιτίας της μεταβλητότητας του δείγματος ή του τρόπου προσδιορισμού του σφάλματος, η προβλεφθείσα γραμμή παλινδρόμησης θα μπορούσε να δείξει μια σταθερή διαφορετική από το μηδέν. Εναλλακτικά, σε μερικές σχέσεις, όπου ο συντελεστής της κλίσης είναι αρνητικός, θα περιμέναμε η σταθερά της παλινδρόμησης να μην είναι μηδέν. Για παράδειγμα, εάν τα επίπεδα του επιτοκίου για ενυπόθηκα δάνεια ήταν μηδέν, θα υπήρχε μια μη-μηδενική ζήτηση για δάνεια αυτής της κατηγορίας. Και στις δύο περιπτώσεις η υπόθεση μηδέν διατυπώνεται με τον ίδιο τρόπο, δηλαδή:

$$H_0: \beta_0 = 0$$

η εναλλακτική υπόθεση είναι

$$H_A: \beta_0 \neq 0$$

το στατιστικό  $t$  για το test καθορίζεται ως

$$t = \frac{(b_0 - \beta_0)}{\left\{ S_e \left[ \frac{1}{n} + \frac{n\bar{X}^2}{n(\sum X_i^2) - (\sum X_i)^2} \right]^{1/2} \right\}} \quad (2.18)$$

με  $n-2$  βαθμούς ελευθερίας. Η τιμή  $t$  (από Πίνακα Τιμών του  $t$ ) στο  $\alpha=0,01$  είναι 3355. Επομένως, εάν η υπολογισθείσα τιμή του στατιστικού  $t$  σε μια τιμή μεγαλύτερη από 3355 και μικρότερη από -3355, θα μπορούσε να μας οδηγήσει στην απόρριψη της υπόθεσης μηδέν και θα συμπεραίναμε ότι το  $\beta_0$  είναι διαφορετικό από το μηδέν.

Με βάση τις αριθμητικές τιμές του παραδείγματος 4.2 και 4.3, βρίσκουμε

$$t = \frac{(-16.859 - 0)}{\left\{ 2.015 \left[ \frac{1}{10} + \frac{10(8.988)}{10(826.778) - (89.88)^2} \right]^{1/2} \right\}} = -4.00$$

και συμπεραίνουμε ότι το  $\beta_0$  δεν είναι ίσο με το μηδέν. Ένα τέτοιο αποτέλεσμα προσδίδει ισχύ στο παλινδρομικό μοντέλο πρόβλεψης.

### 2.1.7 Βαθμός Συσχέτισης

Η συσχέτιση χαρακτηρίζει την ύπαρξη εξάρτησης ανάμεσα στις μεταβλητές. Αν και μπορεί να υπάρχει ένας μεγάλος αριθμός αιτιών για την ύπαρξη εξάρτησης, ο συντελεστής συσχέτισης δεν δίνει καμία πληροφορία για αυτές τις αιτίες. Ο συντελεστής αυτός το μόνο που μας πληροφορεί είναι το εάν δύο ή περισσότερες μεταβλητές συµμεταβάλλονται, είτε θετικά είτε αρνητικά. Επομένως, ο βαθμός συσχέτισης δεν μας δίνει καμία πληροφορία αναφορικά με την αιτιότητα της σχέσης. Η μόνη πληροφορία που δίνει είναι ότι ανάμεσα σε δύο μεταβλητές υπάρχει σχέση εξάρτησης.

Για να αρχίσουμε, ορίζουμε *το συντελεστή προσδιορισμού*,  $\rho^2$  σαν το μέτρο του ποσοστού της διακύμανσης στην εξαρτημένη μεταβλητή  $Y$  που εξηγείται από τη γραμμική παλινδρομική σχέση ανάμεσα στα  $Y$  και  $X$ . Ο συντελεστής δείχνει το πόσο καλά προσαρμόζεται η γραμμή παλινδρόμησης ελαχίστων τετραγώνων στα δεδομένα του διαγράμματος διασποράς. Ο συντελεστής προσδιορισμού  $r^2$  προκύπτει από τα δεδομένα του δείγματος και χρησιμοποιείται για την εκτίμηση του συντελεστή προσδιορισμού  $\rho^2$  του πληθυσμού.

Ο συντελεστής προσδιορισμού,  $r^2$ , αναλύεται σε όρους της διακύμανσης της εξαρτημένης μεταβλητής,  $Y$ . Δείχνει την αναλογία αυτής της διακύμανσης που εξηγείται από την παλινδρομική σχέση ανάμεσα στα  $X$  και  $Y$ . Για ένα μόνο δειγματικό σημείο, εάν θεωρήσουμε το  $\bar{Y}$  σαν μέση τιμή της εξαρτημένης μεταβλητής, το  $(Y - \bar{Y})$  θα δείχνει την συνολική απόκλιση του  $Y$ . Η απόκλιση μπορεί να διαχωριστεί σε δύο μέρη ως εξής:

$$(Y - \bar{Y}) = (\hat{Y} - \bar{Y}) + (Y - \hat{Y}) \quad (2.19)$$

Ένα τμήμα της συνολικής απόκλισης εξηγείται από τη γραμμή παλινδρόμησης ελαχίστων τετραγώνων. Το  $(\hat{Y} - \bar{Y})$  παριστά την απόκλιση που ερμηνεύεται. Το απομένον  $(Y - \hat{Y})$  σχετίζεται με το τυχαίο σφάλμα και δεν μπορεί να εξηγηθεί.

Η παραπάνω ανάλυση μπορεί επίσης να επεκταθεί ώστε να εφαρμοστεί στη συνολική διακύμανση της εξαρτημένης μεταβλητής που μετράται από τον όρο  $\Sigma(Y - \bar{Y})^2$ . Η διακύμανση μπορεί να διαχωριστεί σε δυο μεγέθη. Πρώτον στη διακύμανση των εκτιμηθεισών τιμών της παλινδρομικής γραμμής,  $\bar{Y}$ , γύρω από τον μέσο,  $Y$ , που μετράται ως εξής:

$$\Sigma(\hat{Y} - \bar{Y})^2$$

Το μέγεθος αυτό ονομάζεται *εξηγηθείσα διακύμανση* ή το *ερμηνευόμενο άθροισμα τετραγώνων*. Το δεύτερο στοιχείο είναι το υπολειπόμενο ή το *άθροισμα των τετραγώνων των σφαλμάτων*. Αυτό αντανακλά τη διακύμανση των πραγματικών τιμών της εξαρτημένης μεταβλητής  $Y$ , γύρω από τις

εκτιμηθείσες τιμές της γραμμής παλινδρόμησης,  $Y_e$  και μετράται από το  $(Y - \hat{Y})^2$ . Αυτό αντιπροσωπεύει την ανεξηγήτη διακύμανση. Επομένως, έχουμε:

Συνολική Διακύμανση = εξηγηθείσα + μη εξηγηθείσα διακύμανση:

$$\Sigma(Y_i - \bar{Y})^2 = \Sigma(\hat{Y}_i - \bar{Y})^2 + \Sigma(Y_i - \hat{Y}_i)^2 \quad (2.20)$$

Ο συντελεστής προσδιορισμού ορίζεται ως:

$$r^2 = \frac{\Sigma(\hat{Y}_i - \bar{Y})^2}{\Sigma(Y_i - \bar{Y})^2} \quad (2.21)$$

Ο συντελεστής προσδιορισμού λαμβάνει τιμές μεταξύ του μηδενός και του ένα και δείχνει την αναλογία διακύμανσης στην εξαρτημένη μεταβλητή που εξηγήσαμε από την παλινδρόμηση του  $Y$  πάνω στο  $X$ .

Όταν ο  $r^2$  ισούται με μηδέν, σημαίνει ότι δεν υπάρχει γραμμική εξάρτηση ανάμεσα στα  $X$  και  $Y$ . Σύμφωνα με την ορολογία μας για το  $r^2$ , αυτό θα σήμαινε ότι η εκτιμηθείσα γραμμή παλινδρόμησης συμπίπτει με τον μέσο της εξαρτημένης μεταβλητής  $Y$ . Επομένως,  $\Sigma(\hat{Y} - \bar{Y})^2$  ισούται με μηδέν και  $r^2=0$ .

Όταν  $r^2$  ισούται με τη μονάδα, υπάρχει πλήρης γραμμική σχέση ανάμεσα στα  $X$  και  $Y$ . Πράγμα που σημαίνει ότι όλα τα σημεία των δεδομένων του δείγματος βρίσκονται πάνω στην προβλεφθείσα γραμμή.

Για να περιγράψουμε καλύτερα τα παραπάνω, επιστρέφουμε στο παράδειγμα 2.3, και ερωτάμε ποια αναλογία της συνολικής διακύμανσης των ενυπόθηκων δανείων,  $Y$ , ερμηνεύεται από τη διακύμανση του επιτοκίου,  $X$ . Για καθαρά υπολογιστικούς λόγους, εφαρμόζουμε τον παρακάτω τύπο για τον συντελεστή προσδιορισμού:

$$r^2 = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{[n\Sigma X^2 - (\Sigma X)^2][n\Sigma Y^2 - (\Sigma Y)^2]}$$

Αντικαθιστώντας, βρίσκουμε:

$$r^2 = \frac{10(1165.30) - (89.88)(122.82)}{[10(826.778) - (89.88)^2][10(1740.007) - (122.82)^2]} = 0.860$$

Επομένως, το 86 τοις εκατό της συνολικής διακύμανσης στα δεδομένα των ενυπόθηκων δανείων ερμηνεύεται από την υφιστάμενη σχέση τους με το ύψος του επιτοκίου.

Ένα άλλο μέτρο της σχέσης ανάμεσα σε δυο μεταβλητές είναι ο *συντελεστής συσχέτισης*  $\rho$ . Είναι η τετραγωνική ρίζα του *συντελεστή προσδιορισμού*.

$$\rho = \pm(r^2)^{1/2}$$

όπου το πρόσημο δεν είναι απροσδιόριστο. Ο εκτιμητής του συντελεστή συσχέτισης του πληθυσμού είναι  $r$ , ο συντελεστής συσχέτισης του δείγματος. Ο συντελεστής συσχέτισης παίρνει τιμές στο διάστημα μεταξύ μείον ένα και συν ένα. Όταν το  $\rho$  ισούται με μηδέν, δεν υπάρχει συσχέτιση ανάμεσα στα  $X$  και  $Y$ . Επομένως, δεν υπάρχει γραμμική σχέση ανάμεσα στα  $X$  και  $Y$ . Μια τιμή του  $\rho$  που ισούται με συν ένα σημαίνει ότι υπάρχει μια τέλεια γραμμική σχέση ανάμεσα στα  $X$  και  $Y$ . Όμοια, όταν το  $\rho$  ισούται με μείον ένα, σημαίνει ότι υπάρχει μια τέλεια αντίστροφη γραμμική σχέση ανάμεσα στα  $X$  και  $Y$ .

Εφόσον  $r^2$  είναι 0.860, τότε  $r$  είναι 0.927. Σημειωτέον ότι το πρόσημο του συντελεστή συσχέτισης είναι πάντοτε ίδιο με τον συντελεστή παλινδρόμησης (κλίση) της ανεξάρτητης μεταβλητής  $X$ .

Μπορούμε άραγε να ελέγξουμε τη σημαντικότητα της συσχέτισης; Η απάντηση είναι καταφατική, αλλά αξίζει να κρατήσουμε μια επιφύλαξη. Ο έλεγχος για τη σημαντικότητα του συντελεστή συσχέτισης είναι ανάλογος με τον έλεγχο για τη σημαντικότητα της παλινδρόμησης. Επομένως, τα αποτελέσματα είναι και πρέπει να είναι τα ίδια.

Για την Buckeye Savings and Loan Association, στα παραδείγματα 2.2 και 2.3, θέλουμε να ελέγξουμε την υπόθεση μηδέν,  $\rho=0$ , ότι δηλαδή δεν υπάρχει συσχέτιση έναντι του εναλλακτικού  $\rho \neq 0$ . Συμβολικά, γράφουμε:

$$H_0: \rho=0$$

$$H_A: \rho \neq 0$$

Το  $t$ -στατιστικό γι' αυτό το πρόβλημα δίνεται από:

$$t = \frac{r}{\left[ \frac{(1-r^2)}{(n-2)} \right]^{1/2}} \quad (2.22)$$

με  $n-2$  βαθμούς ελευθερίας. Για τα δεδομένα των παραδειγμάτων 2.2 και 2.3, η κριτική τιμή του  $t$  στο  $\alpha=0.05$  για οκτώ βαθμούς ελευθερίας είναι +2.306 και -2.306. Επομένως, ο κανόνας είναι:

δεχόμαστε την υπόθεση  $\rho=0$  εάν το  $t$  που βασίζεται στα δεδομένα του δείγματος είναι μεγαλύτερο από -2.306 ή μικρότερο του +2.306, διαφορετικά το απορρίπτουμε.

Με αντικατάσταση, βρίσκουμε:

$$t = \frac{0.927}{\left[ \frac{(1-0.860)}{(10-2)} \right]^{1/2}} = 7.00$$

Επομένως, απορρίπτουμε την υπόθεση μηδέν και καταλήγουμε ότι υπάρχει σοβαρή ένδειξη συσχέτισης. Είναι το ίδιο με αυτό που υπολογίστηκε στο παράδειγμα 2.3.

### 2.1.8 Διαδοχική Συσχέτιση (Αυτοσυσχέτιση)

Στο παλινδρομικό μοντέλο

$$Y = \beta_0 + \beta_1 X + e \quad (2.1)$$

υποθέτουμε ότι το  $e$  είναι μια μεταβλητή με  $E(e) = 0$ , και η διακύμανση της  $e$  είναι σταθερή. Εάν το  $e$  δεν εξακολουθεί να είναι εντελώς τυχαίο, και οι διαδοχικές τιμές του  $e$  θεωρούνται ότι δεν είναι στατιστικά ανεξάρτητες, τα σφάλματα λέγεται ότι είναι *αυτοσυσχετιζόμενα*.

Όταν υπάρχει αυτοσυσχέτιση οι διαδικασίες της παλινδρόμησης των ελαχίστων τετραγώνων εξακολουθούν να δίνουν εκτιμητές αμερόληπτους και συνεπείς. Εν τούτοις, η προσέγγιση δεν είναι αποτελεσματική. Ο τύπος του τυπικού σφάλματος και τα τεστ σημαντικότητας δεν είναι κατάλληλα. Συχνά η εκτιμημένη γραμμή παλινδρόμησης προσαρμόζεται στις τιμές των δεδομένων του δείγματος πιο πολύ από ό,τι προσαρμόζεται η πληθυσμιακή παλινδρομική γραμμή. Έτσι προκύπτει ένας πλασματικά υψηλός συντελεστής προσδιορισμού, που δείχνει ότι το εκτιμηθέν παλινδρομικό μοντέλο είναι πιο αξιόπιστο από ό,τι είναι στην πραγματικότητα.

Υπάρχουν διάφορες διαδικασίες για τη μέτρηση του μεγέθους της αυτοσυσχέτισης και ελέγχων για την παρουσία της. Το περισσότερο αποδεκτό κριτήριο ελέγχου δίνεται από το τεστ *Durbin-Watson*. Το τεστ αυτό εμπεριέχει τον προσδιορισμό τού αν ή όχι η συσχέτιση ανάμεσα στα κατάλοιπα είναι ή δεν είναι μηδενική. Θεωρούμε:

$$e_t = \rho e_{t-1} + d \quad (2.23)$$

όπου  $\rho$  είναι η παράμετρος αυτοσυσχέτισης που μετράει τη συσχέτιση ανάμεσα στις τιμές των καταλοίπων, και  $d_t$  είναι ανεξάρτητες τυχαίες μεταβλητές (διαταρακτικοί όροι). Κάθε όρος σφάλματος περιλαμβάνει μια αναλογία του προηγούμενου όρου σφάλματος, όταν  $\rho$  είναι μεγαλύτερο από το μηδέν, και ένα νέο διαταρακτικό όρο (σφάλμα) που υποδηλώνεται με  $d_t$ . Εάν η παράμετρος  $\rho$  ισούται με 0,  $e_t$  ισούται με  $d_t$ . Αυτό σημαίνει ότι οι διαδοχικοί όροι σφάλματος είναι ανεξάρτητοι, εφόσον οι διαταρακτικοί όροι είναι ανεξάρτητοι.

Για το τεστ Durbin-Watson, η μηδενική και εναλλακτική υπόθεση είναι:

$$H_0: \rho = 0$$

$$H_A: \rho > 0$$

Σημειώστε ότι η εναλλακτική υπόθεση είναι μονόπλευρη εφόσον τα κατάλοιπα στη χρονοσειρά τείνουν να δείχνουν μόνο θετική συσχέτιση.

Για να υπολογίσουμε το στατιστικό Durbin-Watson, χρησιμοποιούμε τον παρακάτω τύπο (Durbin

και Watson):

$$DW = \frac{\sum (e_t - e_{t-1})^2}{\sum e_t^2} \quad (2.24)$$

όπου  $e_t$  είναι το σφάλμα ή η διαφορά ανάμεσα στις πραγματικές και τις εκτιμημένες τιμές τον χρόνο  $t$  και  $e_{t-1}$  είναι το σφάλμα ή η διαφορά ανάμεσα στις πραγματικές και τις εκτιμημένες τιμές κατά τη διάρκεια της προηγούμενης χρονικής περιόδου ( $t-1$ ).

Η διαδικασία Durbin-Watson δίνει ανώτερες ( $U$ ) και κατώτερες ( $L$ ) κριτικές τιμές ως κριτήρια ελέγχου ανάμεσα στη μηδενική και εναλλακτική υπόθεση. Η απόφαση διατυπώνεται ως εξής:

1. Αν  $DW > U$ , αποδεχόμαστε την  $H_0$  και άρα δεν υπάρχει αυτοσυσχέτιση.
2. Αν  $DW < L$ , αποδεχόμαστε την  $H_A$  και άρα υπάρχει αυτοσυσχέτιση.
3. Αν  $L < DW < U$ , το τεστ δεν καταλήγει σε συμπέρασμα.

Οι κριτικές τιμές για τα  $L$  και  $U$  περιλαμβάνονται στους Πίνακες του Παραρτήματος στο τέλος του βιβλίου. Για να εντοπίσουμε τις κατάλληλες τιμές  $L$  και  $U$ , πρέπει να ξέρουμε το μέγεθος του δείγματος, το επίπεδο σημαντικότητας και τον αριθμό των ανεξάρτητων μεταβλητών. Για να αναλύσουμε ένα πραγματικό σύνολο παρατηρήσεων, ας θεωρήσουμε το επόμενο παράδειγμα.

#### Παράδειγμα 2.4

Παρατίθενται τα δεδομένα της American Hospital Supply, Inc., για τα έτη 1965 έως 1980. Να βρεθεί η σχέση μεταξύ καθαρού εισοδήματος (\$ εκατομμύρια) και χρόνου. Τα δεδομένα γι' αυτή την παλινδρόμηση αναλύονται με τη μέθοδο των ελαχίστων τετραγώνων.

$$Y = -10877 + 7056X$$

(εκκίνηση το έτος 1964 μια μονάδα του  $X$  = ένα έτος) και το καθαρό εισόδημα εκφράζεται σε \$ εκατομμύρια.

Οι πραγματικές, οι εκτιμημένες τιμές και τα κατάλοιπα για αυτή την παλινδρόμηση παρουσιάζονται στον Πίνακα 2.2. Οι τιμές αυτές είναι απαραίτητες για τον υπολογισμό του στατιστικού DW. Πρώτον, υπολογίζουμε το τετράγωνο κάθε μιας από τις διαφορές ανάμεσα στο κατάλοιπο της περιόδου  $t$  και στο κατάλοιπο της προηγούμενης χρονικής περιόδου. Για τη χρονική περίοδο 2, έχουμε:

$$(e_2 - e_1) = (9.25 - 13.67) = -4.42$$

και

$$(e_2 - e_1)^2 = (-4.42)^2 = 19.5364$$

Η τιμή στον αριθμητή του τύπου για το στατιστικό DW είναι:

$$\sum (e_t - e_{t-1})^2 = 331.7922$$

2.1.9 Διόρθωση για Αυτοσυσχέτιση

Ειδικότερα, η αυτοσυσχέτιση δείχνει ότι υπάρχει μια σημαντική αναλογία της διακύμανσης στην εξαρτημένη μεταβλητή στο καθαρό εισόδημα της American Hospital Supply που δεν εξηγείται από την επίδραση της τάσης. Επομένως, η καλύτερη λύση είναι να ψάξουμε για εκείνες τις ερμηνευτικές μεταβλητές που εξηγούν την παραμένουσα αναλογία της διακύμανσης στην εξαρτημένη μεταβλητή. Στην περίπτωση του καθαρού εισοδήματος της American Hospital Supply, Inc. η προσθήκη μιας μεταβλητής για τη ζήτηση των υπηρεσιών της, όπως για παράδειγμα ο αριθμός των ημερών ασθένειας ή μια μεταβλητή για την τιμή των νοσοκομειακών εξόδων ή για την τιμή της εισροής νοσοκομειακών υπηρεσιών, μπορεί να ελαττώσουν την αυτοσυσχέτιση έτσι ώστε να καταστεί μη-σημαντική.

**ΠΙΝΑΚΑΣ 2.2.** Πραγματικές, εκτιμημένες και τιμές καταλοίπων από την παλινδρόμηση του καθαρού εισοδήματος πάνω στο χρόνο της American Hospital Supply, Inc. για την περίοδο 1965 ως 1980

Χρονική περίοδος	(1)	(2)	(3)
	Πραγματ. Υ	Προσαρμ. (\$ εκατομ.) (Υ)	Κατάλοιπα (e)
1	9.85	-3.82	13.67
2	12.48	3.23	9.25
3	15.19	10.29	4.90
4	20.92	17.35	3.57
5	25.00	24.40	0.60
6	25.65	31.46	-4.81
7	29.16	38.51	-9.35
8	35.60	48.57	-9.97
9	40.81	52.62	-11.81
10	46.33	59.68	-13.35
11	55.18	66.74	-11.56
12	66.27	73.79	-7.52
13	77.92	80.85	-2.93
14	92.60	87.90	4.70
15	109.40	94.96	14.44
16	122.20	102.02	20.18

Όταν δεν είναι εύκολη η ανεύρεση πρόσθετων μεταβλητών για να μειώσουν την αυτοσυσχέτιση σε κάποιο μη-σημαντικό επίπεδο, υπάρχουν διαθέσιμες μαθηματικές διαδικασίες που μπορούν να μειώσουν την αυτοσυσχέτιση έως ένα μη-ενοχλητικό επίπεδο. Μερικές τεχνικές για την αντιμετώπιση της αυτοσυσχέτισης είναι: (α) να εκτιμηθεί ένα μοντέλο ελαχίστων τετραγώνων στις από έτος σε έτος ποσοστιαίες μεταβολές

(β) να συσχετίσουμε τα απόλυτα μεγέθη μεταβολής από έτος σε έτος (γ) να γίνει υστέρηση της εξαρτημένης μεταβλητής κατά μία περίοδο και να συμπεριληφθεί σαν ανεξάρτητη μεταβλητή και (δ) να γίνει χρήση της διορθωτικής τεχνικής «ατο». Από την υπολογιστική πλευρά, αυτή η τελευταία τεχνική είναι η απλούστερη και μ' αυτή θα ασχοληθούμε παρακάτω.

Η τεχνική διόρθωσης  $rho$  είναι ένας παράγοντας ίσος με το μισό της διαφοράς ανάμεσα στο στατιστικό DW και στο 2.0. Για τα δεδομένα του παραδείγματος 2.4, έχουμε:

$$rho = \frac{2 - DW}{2} = \frac{2 - 0.198}{2} = 0.901$$

Για να γίνει προσαρμογή για αυτοσυσχέτιση, το κατάλοιπο της περιόδου  $t-1$  πολλαπλασιάζεται με τον παράγοντα διόρθωσης  $rho$  και προστίθεται στην εκτίμηση της περιόδου  $t$ . Εάν πρόκειται να εκτιμηθεί το 1981 ( $X=17$ ) βρίσκουμε:

$$\hat{Y} = -10.877 + 7.056(17) = 109.075$$

και

$$\begin{aligned} rho \times C_{1980} &= \text{διορθωτικός παράγοντας} \\ (0.901)(20.183) &= 18.185 \end{aligned}$$

$$\begin{aligned} \text{Πρόβλεψη (1981)} &= \hat{Y} + \text{διορθωτικός παράγοντας} \\ &= 109.075 + 18.185 \\ &= 127.260 \end{aligned}$$

Επομένως, η πρόβλεψη για το 1981, όσον αφορά το καθαρό εισόδημα της American Hospital Supply, Inc. είναι 127.260 (\$ εκατομμύρια). Η διαδικασία αυτή θα τείνει να μειώνει την αυτοσυσχέτιση και να βελτιώνει τις προβλέψεις οποτεδήποτε ο ενεργών την πρόβλεψη δεν θα μπορεί εύκολα να βρει μια νέα μεταβλητή που θα συμπεριληφθεί στην εξίσωση πρόβλεψης για την τάση.

### 2.1.10 Το Απλό Γραμμικό Παλινδρομικό Μοντέλο και Πρόβλεψη

Είδαμε πως για μια δοσμένη τιμή του  $X$ , χαρακτηριζόμενη  $X_0$ , το εκτιμημένο παλινδρομικό μοντέλο γίνεται

$$\hat{Y} = b_0 + b_1 X$$

Εφόσον τα  $b_0$  και  $b_1$  είναι τυχαίες μεταβλητές που έχουν τιμές οι οποίες μεταβάλλονται από δείγμα σε δείγμα, έχοντας μια από κοινού κατανομή πιθανότητας, το τυπικό σφάλμα της εκτίμησης  $Y_0$ , μπορεί να υπολογιστεί από

$$SE(\hat{Y}_0) = S_e \left[ \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right]^{1/2} \quad (2.25)$$

Παραπέρα, αυτό το τυπικό σφάλμα αφορά τη μέση πρόβλεψη για δοσμένη τιμή του  $X$ , που υποδηλώνεται με  $X_0$ . Παριστάνουμε το τυπικό σφάλμα με SE (μέσο). Στην εξίσωση 2.25, το μόνο μέγεθος που αλλάζει στη δεξιά πλευρά είναι το  $X_0$ , η νέα τιμή  $X$ . Εάν το  $X_0$  ισούται με τον μέσο των  $n$



γνωστών τιμών του  $X$ , τότε η εξίσωση 2.25 δίνει τη μικρότερη δυνατή τιμή για το τυπικό σφάλμα της μέσης πρόβλεψης. Καθώς το  $X_0$  απομακρύνεται από τον μέσο των τιμών της  $X$ ,  $\bar{X}$ , το τυπικό σφάλμα αυξάνεται. Αυτό φαίνεται εύκολα εάν παρατηρήσουμε ότι ο όρος  $(X_0 - \bar{X})$ , ο αριθμητής, αναφέρεται στο πόσο διαφέρει το  $X_0$  από τον μέσο των  $X$ . Η τιμή του παρονομαστή

$$\Sigma(X_i - \bar{X})^2$$

αναφέρεται στις  $n$  γνωστές τιμές του  $X$ .

Η παραπάνω εξίσωση θεωρεί την πρόβλεψη  $\hat{Y}_0$  ως εάν ήταν μια μέση ή αναμενόμενη τιμή. Συχνά, το ενδιαφέρον μπορεί να επικεντρωθεί στο πόσο πολύ μπορεί να διαφέρει μια συγκεκριμένη (ατομική) πρόβλεψη από την παρατηρηθείσα τιμή. Η έκταση των διαφορών των επιμέρους προβλέψεων από τις παρατηρηθείσες τιμές βρίσκεται δια του συνδυασμού των αποτελεσμάτων της από κοινού κατανομής των  $b_0$  και  $b_1$  και του παρατηρηθέντος προτύπου διασποράς γύρω από την εκτιμηθείσα παλινδρόμηση, το τυπικό σφάλμα της παλινδρόμησης  $\hat{\sigma}_e$ . Το τυπικό σφάλμα μιας ατομικής (ιδιαίτερης) πρόβλεψης δίνεται από:

$$SE(\text{ατομική}) = S_e \left[ \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\Sigma(X_i - \bar{X})^2} \right]^{1/2} \quad (2.26)$$

Όπως στην εξίσωση (2.25), εάν  $X_0 = \bar{X}$ , η εξίσωση (2.26) αποκτά τη μικρότερη τιμή για το τυπικό σφάλμα της ατομικής πρόβλεψης. Παραπέρα, αν  $X_0 = \bar{X}$ , η εξίσωση (2.25) και η (2.26) γίνονται:

$$SE(\text{μέσο}) = S_e \left( \frac{1}{n} \right)^{1/2} \quad (2.27)$$

$$SE(\text{ατομική}) = S_e \left( 1 + \frac{1}{n} \right)^{1/2} \quad (2.28)$$

Με τη σειρά της η αναλογία του SE (ατομικό) προς SE (μέσο), που είναι  $(n+1)/n$  δείχνει ότι το τυπικό σφάλμα των ατομικών προβλέψεων είναι μεγαλύτερο από το τυπικό σφάλμα των μέσων προβλέψεων. Για να παρουσιάσουμε καλύτερα τα σημεία που αναφέραμε, θεωρούμε το παρακάτω παράδειγμα.

Παράδειγμα 2.5

Στην παραπάνω μελέτη της σχέσης των ενυπόθηκων δανείων,  $Y$ , στην Buckeye Savings and Loan Association (παράδειγμα 2.2), η εκτιμηθείσα εξίσωση παλινδρόμησης βρέθηκε ότι είναι

$$\hat{Y} = -16.859 + 3.242X$$

όπου τα ενυπόθηκα δάνεια εκφράζονται σε εκατομμύρια δολάρια. Το  $F$ -στατιστικό ( $F=49.03$  με 1 και 8 βαθμούς ελευθερίας) ήταν σημαντικό σε επίπεδο 0.05.

Ο λαμβάνων αποφάσεις επιθυμεί να προβλέψει την τιμή των ενυπόθηκων δανείων που σχετίζονται με επιτόκιο 10, 11 ή 12 τοις εκατό. Για να λύσουμε το πρόβλημα αυτό αντικαθιστούμε τις τιμές αυτές στην εξίσωση παλινδρόμησης και παίρνουμε τα παρακάτω αποτελέσματα (σε \$ εκατομμύρια):

$$\hat{Y} = -16,859 + 3,242(10) = 15.561 \text{ όταν το επιτόκιο είναι 10 τοις εκατό}$$

$$\hat{Y} = -16,859 + 3,242(11) = 18.803 \text{ όταν το επιτόκιο είναι 11 τοις εκατό}$$

$$\hat{Y} = -16,859 + 3,242(12) = 22.045 \text{ όταν το επιτόκιο είναι 12 τοις εκατό}$$

Εάν το ενδιαφέρον επικεντρώνεται στην εκτίμηση ενυπόθηκων δανείων κατά μέσο όρο, θα μπορούσαμε να χρησιμοποιήσουμε την εξίσωση (2.27) και να εκτιμήσουμε το τυπικό σφάλμα της μέσης πρόβλεψης:

$$SE(\text{μέσο}) = S_e \left[ \frac{1}{10} + \frac{(X_0 - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right]^{1/2}$$

Η μόνη μη-προσδιορισμένη τιμή στη δεξιά πλευρά της εξίσωσης είναι η δοσμένη τιμή για τη  $X$ , η  $X_0$ , το ύψος του επιτοκίου. Για τα τρία επίπεδα επιτοκίου τα τυπικά σφάλματα έχουν ως εξής:

$$SE(\text{μέσος})=0.392 \text{ όταν } X = 10 \text{ τοις εκατό}$$

$$SE(\text{μέσος})=0.454 \text{ όταν } X = 11 \text{ τοις εκατό}$$

$$SE(\text{μέσος})=0.509 \text{ όταν } X = 12 \text{ τοις εκατό}$$

Τώρα, αν θέλουμε να είμαστε σίγουροι ότι τα 95 από τα 100 διαστήματα εμπιστοσύνης περιέχουν την πραγματική μέση πρόβλεψη, προσφεύγουμε στους  $t$ -πίνακες με 8 βαθμούς ελευθερίας, και  $t=2.306$  υπάρχει 0.025 της περιοχής στη δεξιά πλευρά της  $t$ -κατανομής. Επομένως, το  $-2.306 \leq t \leq 2.306$  παριστάνει το 0.90 της περιοχής της κατανομής  $t$  (ένα 95 τοις εκατό διάστημα εμπιστοσύνης). Για να υπολογίσουμε το 95 τοις εκατό του διαστήματος εμπιστοσύνης για τη μέση πρόβλεψη ενυπόθηκων δανείων, το κατάλληλο τυπικό σφάλμα πολλαπλασιάζεται με  $\pm 2.306$ . Το διάστημα, επομένως, είναι:

$$\hat{Y}_0 - 2.306 SE(\text{μέσος}) < \mu_{Y_0} < \hat{Y}_0 + 2.306 SE(\text{μέσος}) \quad (2.29)$$

Αντικαθιστώντας, βρίσκουμε

$$X = 10: 14.657 < \mu_{Y_0} < 16.465$$

$$X = 11:17.750 < \mu_{Y_0} < 19.850$$

$$X = 12:20.871 < \mu_{Y_0} < 23.219$$

Κατά ανάλογο τρόπο, χρησιμοποιούμε την εξίσωση (2.26) για να υπολογίσουμε τα τυπικά σφάλματα πρόβλεψης για συγκεκριμένα επίπεδα αποταμειωτικών καταθέσεων σε ένα δεδομένο επίπεδο επιτοκίου. Η γενική μορφή αυτής της εξίσωσης είναι:

$$SE(\text{ατομική}) = \left[ 1 + \frac{1}{10} + \frac{(X_0 - \bar{X})^2}{\sum (X - \bar{X})^2} \right]^{1/2}$$

όπου η τιμή  $X_0$  είναι το παραδεκτό επίπεδο επιτοκίου για την πρόβλεψη. Για τα τρία επίπεδα, τα τυπικά σφάλματα είναι:

$$SE(\text{ατομική}) = 1.074 \text{ όταν το ύψος του επιτοκίου είναι 10 τοις εκατό}$$

$$SE(\text{ατομική}) = 1.764 \text{ όταν το ύψος του επιτοκίου είναι 11 τοις εκατό}$$

$$SE(\text{ατομική}) = 2.028 \text{ όταν το ύψος του επιτοκίου είναι 12 τοις εκατό}$$

Βεβαίως, οι εκτιμήσεις σημείου για τα ενυπόθηκα δάνεια στα τρία επίπεδα επιτοκίου είναι οι ίδιες όπως και στους προηγούμενους υπολογισμούς. Όμως οι εκτιμήσεις του διαστήματος εμπιστοσύνης διαφέρουν. Το 95 τοις εκατό των εκτιμήσεων του διαστήματος εμπιστοσύνης για τις ατομικές προβλέψεις είναι:

$$X = 10:13.084 < Y < 18.038$$

$$X = 11:16.326 < Y < 21.280$$

$$X = 12:17.368 < Y < 26.722$$

Σημειώστε ότι οι εκτιμήσεις των διαστημάτων εμπιστοσύνης για τις ατομικές προβλέψεις είναι αρκετά διευρυμένες. Επίσης, όπως ισχύει για το διάστημα εμπιστοσύνης ή το διάστημα πρόβλεψης για τις ατομικές προβλέψεις, όσο μακρύτερα βρίσκεται η τιμή  $X_0$  από τη  $X$ , τόσο ευρύτερο είναι το διάστημα πρόβλεψης.

Ένας αριθμός παραγόντων επιδρά στο μέγεθος του τυπικού σφάλματος παλινδρόμησης και επομένως του τυπικού σφάλματος πρόβλεψης και του διαστήματος εμπιστοσύνης. Όσο μεγαλύτερο είναι το μέγεθος του δείγματος,  $n$ , τόσο μικρότερα είναι τα τυπικά σφάλματα και με τη σειρά τους, τα διαστήματα εμπιστοσύνης είναι στενότερα. Κάτι τέτοιο είναι σύμφωνο με την παρατήρηση ότι μεγαλύτερα μεγέθη δείγματος παρέχουν μεγαλύτερη ακρίβεια στις προβλέψεις.

Δεύτερον, όπως σημειώσαμε παραπάνω, όσο μεγαλύτερες είναι οι αποκλίσεις του  $X_0$  από τη  $\bar{X}$  τόσο ευρύτερα είναι τα διαστήματα εμπιστοσύνης. Οι εκτιμήσεις των διαστημάτων εμπιστοσύνης μπορεί να είναι από πολύ μικρές ως πολύ μεγάλες για τιμές του  $X$  που απέχουν από το μέσο, σε σύγκριση με ανάλογα διαστήματα για τις τιμές του  $X$  που βρίσκονται πολύ κοντά στη μέση τιμή,  $X$ .

Τρίτον, όσο λιγότερο ομοιόμορφα κατανέμονται τα δεδομένα, τόσο μεγαλύτερο είναι το τυπικό σφάλμα της παλινδρόμησης και τόσο ευρύτερα είναι τα διαστήματα εμπιστοσύνης. Οι προβλέψεις θα

είναι λιγότερο ακριβείς.

Τέλος, όσο μεγαλύτερη μεταβλητότητα εμφανίζουν οι τιμές του  $X$  στο δείγμα (στο παράδειγμά μας, τα επίπεδα του επιτοκίου) τόσο μεγαλύτερος είναι ο παρονομαστής

$$\Sigma(X_i - \bar{X})^2$$

Επομένως, όσο πιο μικρά είναι τα τυπικά σφάλματα τόσο πιο ακριβή είναι τα διαστήματα εμπιστοσύνης. Αυτό είναι συνεπές με την άποψη ότι εάν η διακύμανση των επιτοκίων ενός πιστωτικού ιδρύματος είναι μεγάλη, τότε θα έχουμε τη δυνατότητα να προβλέψουμε τα ενυπόθηκα δάνεια καλύτερα από ό,τι στην περίπτωση που η διακύμανση των επιτοκίων θα ήταν πολύ μικρή.

Η χρησιμότητα των διαστημάτων εμπιστοσύνης εξαρτάται από το σκοπό για τον οποίο πρόκειται να χρησιμοποιηθούν. Για μακροχρόνιες προβλέψεις, σχετικά ευρέα όρια μπορεί να είναι κατάλληλα και χρήσιμα. Παράλληλα, για αποφάσεις βραχυχρόνιων σχεδιασμών συνήθως απαιτούνται στενά και ακριβή διαστήματα. Αν μια παλινδρομική ανάλυση δύο μεταβλητών δίνει μεγάλα τυπικά σφάλματα που εμποδίζουν την ύπαρξη στενών διαστημάτων, η παλινδρομική αυτή ανάλυση ίσως να μην είναι χρήσιμη. Σε μια τέτοια περίπτωση, η εισαγωγή περισσότερων ανεξάρτητων μεταβλητών είναι απαραίτητη για να έχουμε μεγαλύτερη ακρίβεια στην εκτίμηση της πρόβλεψης.

**2.2 ΜΕΘΟΔΟΙ ΠΟΛΛΑΠΛΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ [9], [10], [14]****2.2.1 Εισαγωγή**

Στην απλή γραμμική παλινδρόμηση ερευνάται η σχέση μεταξύ μιας ανεξάρτητης και μιας εξαρτημένης μεταβλητής για την εξυπηρέτηση των σκοπών της πρόβλεψης. Η απλή γραμμική παλινδρόμηση μπορεί συχνά να οδηγήσει σε ακριβείς προβλέψεις. Υπάρχει όμως και το δεδομένο ότι, συνήθως, είναι απαραίτητη η χρησιμοποίηση περισσότερων της μιας ανεξάρτητων μεταβλητών ώστε να μπορέσουμε να προβλέψουμε τις τιμές μιας εξαρτημένης μεταβλητής με ικανοποιητική ακρίβεια. Τα προβλήματα παλινδρόμησης στα οποία περιλαμβάνονται περισσότερες της μιας ανεξάρτητες μεταβλητές τα ονομάζουμε προβλήματα πολλαπλής παλινδρόμησης.

**2.2.2 Προϋποθέσεις του Μοντέλου Πολλαπλής Γραμμικής Παλινδρόμησης**

Η πρώτη προϋπόθεση του μοντέλου πολλαπλής γραμμικής παλινδρόμησης είναι ότι η εξίσωση της σχέσης είναι γραμμικής μορφής και γράφεται

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m + e \quad (2.30)$$

όπου  $Y$  είναι η εξαρτημένη μεταβλητή,  $X_1, X_2, \dots, X_m$  είναι οι ανεξάρτητες μεταβλητές,  $e$  είναι ο όρος του τυχαίου σφάλματος,  $\beta_0$  είναι η σταθερά και  $\beta_1, \beta_2, \dots, \beta_m$  είναι οι συντελεστές παλινδρόμησης που εξηγούν το βαθμό εξάρτησης μεταξύ των ανεξάρτητων μεταβλητών και της εξαρτημένης. Το τυχαίο σφάλμα υπάρχει για τον ίδιο λόγο που υπάρχει και στην απλή γραμμική παλινδρόμηση, δηλαδή λόγω ημιτελούς θεωρίας, ατελούς εξειδίκευσης και υπολογιστικών σφαλμάτων.

Στην παλινδρομική ανάλυση μπορούμε να θέσουμε το ερώτημα αν μπορούμε ή όχι να βελτιώσουμε σημαντικά την πρόβλεψή μας αυξάνοντας την πολυπλοκότητα του μοντέλου που εφαρμόζουμε. Ειδικότερα, τίθεται το ερώτημα αν ένα μοντέλο της μορφής

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2^2 + \dots + e \quad (2.31)$$

δηλαδή ένα *πολυωνυμικό* μοντέλο (ή ένα καμπυλόγραμμο μοντέλο), έχει καλύτερη ή όχι προσαρμογή από ένα *πολυμεταβλητό γραμμικό* μοντέλο. Ή, ακόμη, μπορεί να τεθεί το ερώτημα για το αν ένα μοντέλο, για παράδειγμα, της μορφής

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3^{0.5} + \dots + e \quad (2.32)$$

αποτελεί καλύτερο μοντέλο για τη διενέργεια προβλέψεων. Η εξίσωση (2.32) είναι ένα παράδειγμα πολλαπλού *μη-γραμμικού* παλινδρομικού μοντέλου. Ο όρος του τυχαίου σφάλματος,  $e$ , είναι εν μέρει αποτέλεσμα της επιλογής ενός πολλαπλού γραμμικού μοντέλου αντί ενός μοντέλου όμοιου με εκείνο των εξισώσεων (2.31) ή (2.32). Επομένως, στην πολλαπλή γραμμική παλινδρόμηση υποθέτουμε ότι

το καλύτερο μοντέλο (δηλαδή το μοντέλο αναφοράς) είναι το γραμμικό.

Η *δεύτερη προϋπόθεση* είναι ότι εμφανίζεται η ίδια κατανομή τιμών για τις ανεξάρτητες μεταβλητές κάθε φορά που εκτελείται το πείραμα. Όμως, η κατανομή των τιμών της  $Y$ , για δεδομένες τιμές των ανεξάρτητων μεταβλητών, μπορεί να ποικίλλει για κάθε πείραμα εξαιτίας της επίδρασης του τυχαίου σφάλματος. Με άλλα λόγια, οι τιμές των ανεξάρτητων μεταβλητών έχουν προκαθοριστεί, ενώ η εξαρτημένη μεταβλητή είναι μια τυχαία μεταβλητή.

Η *τρίτη προϋπόθεση* είναι ότι η αναμενόμενη τιμή του όρου του τυχαίου σφάλματος είναι μηδέν.

$$E(e) = 0 \tag{2.33}$$

Δηλαδή, κατά μέσον όρο, η τιμή που παίρνει ο όρος του τυχαίου σφάλματος είναι ίση με μηδέν. Αν λάβουμε την αναμενόμενη τιμή μιας γραμμικής πολλαπλής παλινδρόμησης, για δεδομένες τιμές των ανεξάρτητων μεταβλητών, ας πούμε  $X_1, X_2, \dots, X_n$ , έχουμε:

$$E(Y|X_1, X_2, \dots, X_n) = E(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + e) \\ E(\beta_0) + E(\beta_1 X_1) + E(\beta_2 X_2) + \dots + E(\beta_n X_n) + E(e) \tag{2.34}$$

και

$$\mu_Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \tag{2.35}$$

εφόσον  $E(e) = 0$ . Το παραπάνω παλινδρομικό μοντέλο απλά λέει ότι για δεδομένες τιμές των  $X_1, X_2, \dots, X_n$ , ο μέσος των τιμών  $Y$ ,  $\hat{Y}$  είναι ένα σημείο που ορίζεται από το μοντέλο πολλαπλής παλινδρόμησης.

Η *τέταρτη προϋπόθεση* δέχεται ότι η διακύμανση του όρου του τυχαίου σφάλματος,  $\sigma_e^2$ , είναι σταθερή για κάθε σύνολο τιμών των ανεξάρτητων μεταβλητών. Η  $\sigma_e^2$ , μετρά τη μεταβλητικότητα των τιμών της εξαρτημένης μεταβλητής περί την επιφάνεια παλινδρόμησης για δεδομένες τιμές των ανεξάρτητων μεταβλητών.

Η *πέμπτη προϋπόθεση* δέχεται ότι στο πολλαπλό μοντέλο παλινδρόμησης οι τιμές του όρου του σφάλματος είναι ανεξάρτητες μεταξύ τους.

Η *έκτη προϋπόθεση* δέχεται ότι οι τιμές του όρου του σφάλματος,  $e$ , για κάθε σύνολο τιμών των ανεξάρτητων μεταβλητών κατανέμονται κανονικά και ανεξάρτητα<sup>1</sup>.

#### Παράδειγμα 2.4

Σε μελέτη ανάληψης κινδύνου από την επένδυση σε ένα κερδοσκοπικού χαρακτήρα Νοσηλευτικό Ίδρυμα, ένας αναλυτής ερευνά εκείνους τους παράγοντες που καθορίζουν το μέτρο του σχετικού κινδύνου, βήτα, μιας εγγύησης.

Στην εξίσωση πρόβλεψης για το βήτα, με το  $Y$  ως εξαρτημένη μεταβλητή περιλαμβάνει ορισμένους χρηματοοικονομικούς λόγους ως ανεξάρτητες μεταβλητές. Κι αυτοί είναι:

$X_1$  = ποσοστό των τόκων επί του χρέους ως προς το συνολικό εισόδημα

$X_2$  = μέση ετήσια μεταβολή των καθαρών εισπράξεων

$X_3$  = συντελεστής διακύμανσης των καθαρών εισπράξεων

$X_4$  = περιθώρια κέρδους.

Η εξίσωση πρόβλεψης είναι:

$$\hat{Y} = 0.759 + 0.580X_1 - 0.12X_2 + 0.034X_3 + 7.244X_4$$

Για τους σκοπούς της πρόβλεψης, αντικαθιστούμε κάθε ανεξάρτητη μεταβλητή ως εξής:

$$\hat{Y} = 0.759 + 0.580(0.309) - 0.012(10.967) + 0.034(2.859) - 7.244(0.013) = 0.810$$

---

Επομένως, κατά μέσον όρο, ένα Νοσηλευτικό Ίδρυμα που έχει οικονομικά χαρακτηριστικά παρόμοια με τις τιμές που αντικαταστάθηκαν στην εξίσωση πρόβλεψης, θα έχει ένα βήτα ίσο με 0.810. Από την εξίσωση πρόβλεψης παρατηρούμε ότι αυξήσεις στις τιμές των  $X_1$  και  $X_3$  αυξάνουν την εκτίμηση βήτα, ενώ μια αύξηση στις τιμές των  $X_2$  και  $X_4$  μειώνει την εκτίμηση βήτα.

Σημειώστε ότι οι συντελεστές παλινδρόμησης μετρούν την καθαρή μεταβολή στο βήτα που σχετίζεται με κατά μια μονάδα μεταβολή στην ανεξάρτητη μεταβλητή. Τέλος, ο αριθμός των βαθμών ελευθερίας του μοντέλου είναι ίσος με τον αριθμό των παρατηρήσεων μείον τον αριθμό των παραμέτρων που υπολογίζονται.

### 2.2.3 Εκτίμηση Πολλαπλής Γραμμικής Παλινδρόμησης

Όπως στο απλό γραμμικό παλινδρομικό μοντέλο, η εκτίμηση στηρίζεται στις παρατηρήσεις τυχαίων δειγμάτων της εξαρτημένης μεταβλητής για ορισμένες τιμές της ανεξάρτητης μεταβλητής. Η εξαρτημένη μεταβλητή είναι τυχαία μεταβλητή ενώ οι ανεξάρτητες μεταβλητές είναι προκαθορισμένες.

Η μέθοδος των ελαχίστων τετραγώνων είναι η μαθηματική διαδικασία για την εφαρμογή της εξίσωσης πολλαπλής παλινδρόμησης στα δεδομένα του δείγματος. Εκτιμούμε τις παραμέτρους της παλινδρόμησης ( $\beta_0, \beta_1, \beta_2, \dots, \beta_m$ ) ώστε η στατιστική σχέση που θα προκύψει να εκφράζει την καλύτερη δυνατή προσαρμογή των δεδομένων. Η διαδικασία έγκειται στην ελαχιστοποίηση του αθροίσματος των κάθετων τετραγωνικών αποκλίσεων ανάμεσα στις πραγματικές τιμές δείγματος της εξαρτημένης μεταβλητής,  $Y$ , και τις υπολογισθείσες τιμές,  $\hat{Y}$ . Η διαδικασία αποτελεί σχεδόν μια επέκταση της τεχνικής των ελαχίστων τετραγώνων για τις απλές γραμμικές παλινδρομικές σχέσεις.

Το ακόλουθο παράδειγμα περιγράφει τη μέθοδο των ελαχίστων τετραγώνων για ένα πρόβλημα πολλαπλής παλινδρόμησης της μορφής

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 \tag{2.36}$$

όπου  $b_0$  είναι η υπολογισθείσα σταθερά και  $b_1$  και  $b_2$  είναι οι υπολογισθέντες συντελεστές παλινδρόμησης.

Παράδειγμα 2.5

Η Buckeye Savings and Loan επιθυμούσε να βελτιώσει τις προβλέψεις της για τα ενυπόθηκα δάνεια σε δολάρια, προσθέτοντας μια επιπλέον μεταβλητή πρόβλεψης (ατομικό εισόδημα του Ohio σε δισ. δολάρια) στη σχετική εξίσωση. Σε προηγούμενα παραδείγματα συμπεριλαμβανόταν μόνο το επιτόκιο επί ενυπόθηκων δανείων στην εξίσωση πρόβλεψης. Οι υπολογισμοί για την πολλαπλή γραμμική παλινδρόμηση δίνονται στον Πίνακα 5.1.

**ΠΙΝΑΚΑΣ 2.3.** Προκαταρκτικός πίνακας για πολλαπλή παλινδρόμηση

Ενυπόθ δάνεια. (\$ εκατ) Y	Προσωπικό Εισόδημα στο Ohio(\$ δις) X <sub>1</sub>	Επιτόκιο X <sub>2</sub>	X <sub>1</sub> Y	X <sub>2</sub> Y	X <sub>1</sub> X <sub>2</sub>	X <sub>1</sub> <sup>2</sup>	X <sub>2</sub> <sup>2</sup>	Y <sup>2</sup>
7.50	44.80	7.60	336.00	57.000	340.48	2007.0	57.760	56.250
7.44	49.00	7.45	364.56	55.428	365.05	2401.0	55.502	55.354
7.93	54.47	7.78	431.95	61.695	423.78	2967.0	60.528	62.885
8.41	58.90	8.71	495.35	73.251	513.02	3469.2	75.864	70.728
9.21	62.75	8.75	577.93	80.587	549.06	3937.6	76.563	84.824
12.03	68.37	8.76	822.49	105.383	598.92	4674.5	76.738	144.721
14.33	75.87	8.80	1087.22	126.104	667.66	5756.3	77.440	205.349
16.25	84.00	9.30	1365.00	151.125	781.20	7056.0	86.490	264.063
17.99	93.50	10.48	1682.06	188.535	979.88	8742.3	109.830	323.640
21.73	101.24	12.25	2199.95	266.192	1240.19	10249.5	150.063	472.193
122.82	692.90	89.88	9362,51	1165,300	6459,24	51260,4	826,778	1740,007

Οι τρεις εξισώσεις που πρέπει να λυθούν για να καθοριστούν οι τιμές των  $b_0$ ,  $b_1$  και  $b_2$  είναι 3

$$\Sigma Y = nb_0 + b_1 \Sigma X_1 + b_2 \Sigma X_2$$

$$\Sigma X_1 Y = b_0 \Sigma X_1 + b_1 \Sigma X_1^2 + b_2 \Sigma X_1 X_2$$

$$\Sigma X_2 Y = b_0 \Sigma X_2 + b_1 \Sigma X_1 X_2 + b_2 \Sigma X_2^2$$

Οι υπολογισμοί δίνονται στον προκαταρκτικό πίνακα του παραδείγματος 2.5.

Η αντικατάσταση στις παραπάνω εξισώσεις δίνει

$$122.82 = 10b_0 + 692.90b_1 + 89.88b_2$$

$$9362.51 = 692.90 b_0 + 51.260.4 b_1 + 6459.24b_2$$



$$1165.300 = 89.88 b_0 + 64.59.2 b_1 + 826.78 b_2$$

Λύνοντας τις τρεις εξισώσεις ταυτόχρονα παίρνουμε

$$b_0 = -7.025$$

$$b_1 = 0.242$$

$$b_2 = 0.279$$

Στο παράδειγμα που αφορά την Buckeye Savings and Loan, η τιμή  $b_1$ : 0.242 δείχνει ότι κάθε αύξηση του ατομικού εισοδήματος στο Ohio της τάξης του ενός δισεκατομμυρίου δολαρίων, όταν τα επιτόκια κρατούνται σταθερά, συνεπάγεται μια αύξηση 0.242 εκατομμυρίων δολαρίων των ενυπόθηκων δανείων. Όμοια, η τιμή  $b_2$ : 0.279 δείχνει ότι κάθε αύξηση επιτοκίου κατά 1 τοις εκατό, όταν το ατομικό εισόδημα στο Ohio κρατείται σταθερό, συνεπάγεται μια αύξηση στα ενυπόθηκα δάνεια της τάξης των 0.279 εκατομμυρίων δολαρίων κατά μέσον όρο.

#### 2.2.4 Λύσεις Ελαχίστων Τετραγώνων και Έλεγχος Σημαντικότητας

Εφόσον τώρα οι λύσεις των ελαχίστων τετραγώνων στην πολλαπλή παλινδρόμηση δίνονται με τη βοήθεια του ηλεκτρονικού υπολογιστή, δεν αξίζει να γράψουμε τις ακριβείς αλγεβρικές εξισώσεις για τα  $b_0$ ,  $b_1$ ,  $b_2$ ... Ωστόσο, μεγάλη έμφαση θα πρέπει να δοθεί στη μελέτη μερικών ιδιοτήτων της παλινδρόμησης ελαχίστων τετραγώνων. Η εξίσωση (2.30) είναι ο μοναδικός γραμμικός συνδυασμός των ανεξάρτητων μεταβλητών  $X_1, X_2, \dots, X_m$  που έχει τη μέγιστη δυνατή συσχέτιση με την εξαρτημένη μεταβλητή. Για να το θέσουμε διαφορετικά, από όλες τις εκδοχές γραμμικών συνδυασμών της μορφής της εξίσωσης (5.1), ο γραμμικός εκτιμητής,  $\hat{Y}$ , είναι τέτοιος που ο συντελεστής προσδιορισμού είναι:

$$R_{\hat{Y}, Y}^2 = \frac{\sum (\hat{Y} - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} \quad (2.37)$$

Ο αριθμητής είναι το άθροισμα των τετραγώνων των αποκλίσεων που ερμηνεύεται από την παλινδρόμηση και ο παρονομαστής είναι το ολικό άθροισμα των τετραγώνων όπου  $\hat{Y}_i$  είναι η προβλεφθείσα τιμή του  $Y$  για την  $i$ -στή παρατήρηση.  $\bar{Y}_i$  είναι ο μέσος των προβλεφθεισών τιμών και είναι ίσος με  $\bar{Y}$ , τον μέσο των παρατηρηθεισών τιμών. Για τα δεδομένα του παραδείγματος 2.5, υπολογίζουμε τον  $R^2$  (R - SQUARE) ως εξής:

$$R^2 = \left( \begin{array}{c} \text{άθροισμα τετραγώνων} \\ \text{που οφείλεται στην} \\ \text{παλινδρόμηση} \end{array} \right) / \left( \begin{array}{c} \text{συνολικό άθροισμα} \\ \text{τετραγώνων} \end{array} \right)$$

$$= \frac{223.7570}{231.5307} = 0.9664$$

Επομένως, 96.64 τοις εκατό της συνολικής μεταβλητικότητας των ενυπόθηκων δανείων της Buckeye Savings and Loan σχετίζεται με τη γραμμική παλινδρόμηση του προσωπικού εισοδήματος του Ohio,  $X_1$ , και τα επιτόκια για δάνεια νέων οικιών,  $X_2$ . Θα έπρεπε να σημειώσουμε ότι ο συντελεστής πολλαπλής συσχέτισης είναι η θετική τετραγωνική ρίζα του  $R^2$  ( $R=0.983$ ). Ο συντελεστής πολλαπλής συσχέτισης είναι ένα μέτρο της ισχύος της γραμμικής σχέσης ανάμεσα στα  $Y$  και τον καλύτερο γραμμικό συνδυασμό των ανεξάρτητων μεταβλητών στην εξίσωση (2.30). Παραπέρα, μπορούμε να δείξουμε ότι ο συντελεστής πολλαπλής συσχέτισης είναι πάντοτε μη αρνητικός και αποτελεί γενίκευση του συντελεστή απλής συσχέτισης στην περίπτωση των δύο ή περισσότερων ανεξάρτητων μεταβλητών.

Όπως και στο μοντέλο της απλής γραμμικής παλινδρόμησης, η  $F$ -Κατανομή και ο σχετικός έλεγχος μπορούν να χρησιμοποιηθούν για να δώσουν μια συνολική περιεκτική εικόνα της ανάλυσης της πολλαπλής παλινδρόμησης. Οι έλεγχοι αυτοί δίνονται συνοπτικά σε πίνακα (βλ. Πίνακα 2.4) που ονομάζεται πίνακας ανάλυσης της διακύμανσης (analysis of variance: ANOVA) και παίρνουν τον γενικό τίτλο που δίνεται σε αυτή την κατηγορία διαγνωστικών ελέγχων.

**ΠΙΝΑΚΑΣ 2.4.** Η ANOVA ενυπόθηκων δανείων που έχουν παλινδρομηθεί πάνω σε δύο ανεξάρτητες μεταβλητές

Πηγή	DF	SS	MS	F
Μοντέλο	2	223.7570	111.8785	100.74
Σφάλμα	7	7.7738	1.1105	
	9	231.5308		

$F$  είναι σημαντικό σε επίπεδο μικρότερο του 0.01. R-SQUARE = 0.9664

Άθροισμα τετραγώνων μοντέλου =  $SSY - SSE = 223.7570$

Άθροισμα τετραγώνων σφαλμάτων =  $SSE = 7.7738$

Συνολικό άθροισμα τετραγώνων =  $SSY = 231.5308$

Θεωρούμε το  $SSY$  για να παραστήσουμε το συνολικό άθροισμα των τετραγώνων των αποκλίσεων:

$$SSY = \sum (Y_i - \bar{Y})^2 \tag{2.38}$$

που είναι η συνολική μεταβλητότητα στις παρατηρήσεις  $Y$  πριν λογαριάσουμε την από κοινού επίδραση των ανεξάρτητων μεταβλητών,  $X_1, X_2$ . Ο όρος  $SSE$ , το σφάλμα (ή το κατάλοιπο) του αθροίσματος τετραγώνων, ορίζεται ως

$$SSE = \sum (Y_i - \hat{Y}_i)^2 \quad (2.39)$$

Αυτό μετρά το ύψος της μεταβλητικότητας στην εξαρτημένη μεταβλητή  $Y$  που δεν εξηγείται από τις ανεξάρτητες μεταβλητές κατά την εκτίμηση του μοντέλου παλινδρόμησης για την πρόβλεψη του  $Y$ . Τέλος,

$$SSY - SSE = \sum (\hat{Y}_i - \bar{Y})^2 \quad (2.40)$$

ονομάζεται άθροισμα των τετραγώνων της παλινδρόμησης και μετράει το ποσοστό μεταβλητικότητας στις τιμές του  $Y$  που οφείλεται στις ανεξάρτητες μεταβλητές της εξίσωσης. Επομένως, τώρα έχουμε τη γνωστή έκφραση:

συνολικό άθροισμα τετραγώνων = ερμηνευόμενο άθροισμα τετραγώνων + άθροισμα τετραγώνων των σφαλμάτων

ή

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum (Y_i - \hat{Y}_i)^2 \quad (2.41)$$

Για να ελέγξουμε τη συνολική παλινδρόμηση, θέτουμε την ακόλουθη μηδενική υπόθεση:

$H_0$ : Όλες οι ανεξάρτητες μεταβλητές θεωρούμενες από κοινού δεν εξηγούν σημαντικό ποσοστό της μεταβλητικότητας στην  $Y$ .

$H_A$ : Υπάρχει επαρκής ένδειξη συσχέτισης (παλινδρόμηση).

Η παραπάνω υπόθεση ελέγχεται με την ακόλουθη στατιστική.

$$F = \frac{\text{άθροισμα τετραγώνων που σχετίζεται με την παλινδρόμηση/αριθμός ανεξάρτητων μεταβλητών}}{\text{άθροισμα τετραγώνων σφαλμάτων/συνολικός αριθμός συντελεστών στην παλινδρόμηση}} \quad (2.42)$$

$$= \frac{(SSY - SSE) / m}{SSE / (n - m - 1)}$$

όπου  $m$  είναι ο αριθμός των ανεξάρτητων μεταβλητών.

Η παραπάνω  $F$ -αναλογία συγκρίνεται με την κρίσιμη  $F$ -αναλογία με  $m$  και  $(n-m-1)$  βαθμούς ελευθερίας σε επίπεδο σημαντικότητας  $\alpha$ . Μια ισοδύναμη έκφραση για την  $F$  είναι

$$F = \frac{R^2 / m}{(1 - R^2) / (n - m - 1)} \quad (2.43)$$

Για το παραπάνω παράδειγμα,

$$F = \frac{0.9664 / 2}{(1 - 0.9664) / (10 - 2 - 1)} = 100.74$$

Συγκρίνουμε αυτό με την κριτική  $F$ , με δύο βαθμούς ελευθερίας στον αριθμητή και επτά βαθμούς ελευθερίας στον παρονομαστή στο  $\alpha=0.01$ , που είναι 9.55. Επομένως, στο επίπεδο 0.01 μπορούμε να απορρίψουμε την  $H_0$  και να συμπεράνουμε ότι υπάρχει ικανοποιητική ένδειξη συσχέτισης. Σημειώνεται ότι η υπολογισθείσα τιμή του  $F$  είναι η ίδια με εκείνη που υπολογίστηκε στον Πίνακα 2.4.

Εκτός από τον έλεγχο για τη συνολική σημαντικότητα του μοντέλου, υπάρχουν και άλλα ενδιαφέροντα ερωτήματα σε σχέση με αυτό. Πρώτον, μπορεί κάθε μια από τις ανεξάρτητες μεταβλητές ξεχωριστά να εξηγήσει σημαντικό ποσοστό της μεταβλητικότητας στο  $Y$ ; Δεύτερον, η προσθήκη μιας επιπλέον μεταβλητής συνεισφέρει αισθητά στην πρόβλεψη του  $Y$  μετά τον υπολογισμό της συμμετοχής των ήδη συμπεριληφθεισών ανεξάρτητων μεταβλητών; Τρίτον, έχει έννοια το εκτιμημένο μοντέλο παλινδρόμησης από την άποψη του τι πιστεύουμε ως αληθινό, δηλαδή, μπορεί να εξηγήσει σωστά τη συμπεριφορά;

Αυτές και άλλες ερωτήσεις μεγάλης σπουδαιότητας απαιτούν πολύ περισσότερους στατιστικούς υπολογισμούς. Τέτοιοι υπολογισμοί γίνονται συνήθως με τη χρήση κατάλληλου προγράμματος στον υπολογιστή. Στην επόμενη παράγραφο θα θεωρήσουμε τα αποτελέσματα ενός ιδιαίτερα χρήσιμου προγράμματος υπολογιστή για την εκτίμηση της πολλαπλής παλινδρόμησης.

Θα πρέπει να σημειώσουμε ότι το  $R^2$  αυξάνεται με τον αριθμό των ανεξάρτητων μεταβλητών, οι οποίες μειώνουν τον αριθμό των βαθμών ελευθερίας. Το στατιστικό  $\bar{R}^2$  λαμβάνει υπόψη τον αριθμό των βαθμών ελευθερίας. Αυτό υπολογίζεται ως εξής:

$$\bar{R}^2 = 1 - \left( \frac{n-1}{n-m} \right) (1 - R^2) \quad (2.44)$$

και ονομάζεται προσαρμοσμένος συντελεστής πολλαπλού προσδιορισμού.

Ο συντελεστής αυτός μπορεί να γίνει μικρότερος όταν μια νέα ανεξάρτητη μεταβλητή εισάγεται στο μοντέλο, γιατί η αύξηση στο  $R^2$  μπορεί να είναι μεγαλύτερη από το αντιστάθμισμα λόγω της απώλειας ενός βαθμού ελευθερίας στον παρονομαστή  $(n-m)$ .

Για τα δεδομένα του παραδείγματος 2.4 βρίσκουμε

$$\bar{R}^2 = 1 - \left( \frac{10-1}{10-2} \right) (1 - 0.9664) = 0.9622$$

Επομένως, 96.22 τοις εκατό της συνολικής διακύμανσης στα ενυπόθηκα δάνεια σχετίζεται με τη γραμμική παλινδρόμηση στις ανεξάρτητες μεταβλητές, το προσωπικό εισόδημα στο Ohio και τα επιτόκια, προσαρμοσμένο με τους βαθμούς ελευθερίας. Η μείωση από το  $R^2$  στο  $\bar{R}^2$  είναι πολύ μικρή.

**2.2.5 Πολυσυγγραμμικότητα**

Πολυσυγγραμμικότητα έχουμε όταν οι ανεξάρτητες μεταβλητές συσχετίζονται μεταξύ τους και συμβάλλουν έτσι στη δημιουργία πλασματικών πληροφοριών. Τα αποτελέσματα της ύπαρξής της μπορούν να ιδωθούν με πολλούς τρόπους:

1. Ένας συντελεστής με θετικό πρόσημο σε μοντέλο παλινδρόμησης δύο μεταβλητών μπορεί να μεταβάλει το πρόσημό του σε αρνητικό όταν χρησιμοποιηθεί μοντέλο πολλαπλής παλινδρόμησης με τρεις ή περισσότερες μεταβλητές. Η κατεύθυνση της μεταβολής του προσήμου μπορεί να είναι και η αντίστροφη.
  2. Μικρή αλλαγή στο δείγμα μπορεί να οδηγήσει σε σημαντικές αλλαγές.
  3. Εφόσον η πολλαπλή παλινδρόμηση έχει τον πρόσθετο σκοπό να εξηγήσει και να ερμηνεύσει τη συμπεριφορά, όπως επίσης και να την προβλέψει, οι σημαντικά αλληλοσχετιζόμενες μεταβλητές δεν επιτρέπουν το διαχωρισμό των ατομικών επιδράσεων των ανεξάρτητων μεταβλητών.
- Όλες αυτές οι επιδράσεις είναι αποτέλεσμα των μεγάλων, διακυμάνσεων του συντελεστή συσχέτισης ανάμεσα σε ζευγάρια των μεταβλητών.

Σε αυτό το κεφάλαιο αναπτύσσουμε δύο μοντέλα πολλαπλής παλινδρόμησης για να προβλέψουμε τα ενυπόθηκα δάνεια στην περίπτωση της Buckeye Savings and Loan Association. Το πρώτο μοντέλο δίνει την ακόλουθη εξίσωση:

$$\hat{Y} = -7.025 + 0.242X_1 + 0.279X_2$$

όπου  $X_1$  αφορά το προσωπικό εισόδημα στο Ohio, και  $X_2$  τα επιτόκια στις υποθήκες νέων κατοικιών. Με όμοιο τρόπο, ένα δεύτερο μοντέλο πολλαπλής παλινδρόμησης, που περιέχει μια τρίτη ανεξάρτητη μεταβλητή,  $X_3$ , για τον χρόνο (σε έτη), εκτιμήθηκε δίνοντας το ακόλουθο αποτέλεσμα

$$\hat{Y} = 2254.619 + 0.485X_1 - 0.246X_2 - 1.302X_3$$

Σημειώστε ότι το πρόσημο του συντελεστή για τη μεταβλητή του επιτοκίου,  $X_2$ , μεταβλήθηκε από το ένα μοντέλο στο άλλο. Το φαινόμενο αυτό δείχνει την παρουσία πολυσυγγραμμικότητας. Εάν υπάρχει, οι συντελεστές παλινδρόμησης  $b_1$ ,  $b_2$  και  $b_3$  δεν αποτελούν αξιόπιστες μετρήσεις των ατομικών επιδράσεων των σχετιζόμενων ανεξάρτητων μεταβλητών. Μπορεί να υπάρχουν διαθέσιμοι καλύτεροι δείκτες.

Μια μέθοδος, με τη βοήθεια της οποίας μπορούμε να διευκρινίσουμε την ύπαρξη ή όχι πολυσυγγραμμικότητας, είναι η δημιουργία του πίνακα συσχέτισης (correlation matrix), που δεν είναι τίποτε άλλο από καταγραφή των ατομικών συντελεστών γραμμικής συσχέτισης μεταξύ των ζευγών των μεταβλητών. Ο πίνακας συσχέτισης δημιουργείται εύκολα, με τη βοήθεια γλωσσών software, (SPSS για παράδειγμα),

Η ύπαρξη της πολυσυγγραμμικότητας υποδηλώνει ότι το μοντέλο παλινδρόμησης είναι ανίκανο να διαχωρίσει τη συγκεκριμένη σχέση ανάμεσα σε κάθε ανεξάρτητη μεταβλητή και στην εξαρτημένη. Αν και το πρόβλημα αυτό δεν έχει απλή λύση, το μοντέλο μπορεί να δείχνει μια σημαντική σχέση

ανάμεσα στην εξαρτημένη και τις εξαρτημένες μεταβλητές ως σύνολο. Παράλληλα, ίσως κάποιος να σκοπεύει να απαλείψει από το μοντέλο μια από τις ταυτόχρονα συσχετιζόμενες μεταβλητές. Αν υπάρχει τέτοιο θέμα, η εξίσωση πρόβλεψης που απομένει μπορεί να χρησιμοποιηθεί για πρόβλεψη, αλλά ίσως όχι για σκοπούς εξήγησης της συμπεριφοράς της εξαρτημένης μεταβλητής.

### 2.2.6 Ψευδομεταβλητές (dummy variables) στα Μοντέλα Παλινδρόμησης

Μεταβλητές κατηγοριοποίησης (σε κατονομαζόμενη κλίμακα) συμπεριλαμβάνονται σε μοντέλα παλινδρόμησης με τη χρήση *ψευδομεταβλητών*, όπου οι τάξεις είναι κωδικοποιημένες συχνά με τη χρήση του «0» ή «1» ή «-1» και «1». Η προσέγγιση αυτή είναι χρήσιμη στις περιπτώσεις που θέλουμε να ταξινομήσουμε τη σχέση μιας ομάδας ή μιας μη μετρήσιμης μεταβλητής σε μια ανεξάρτητη μεταβλητή.

#### Παράδειγμα 2.6

Ένας βιομήχανος επιθυμεί να μετρήσει το πόσο καλά προβλέπει ένα συγκεκριμένο τεστ ικανότητας την επίδοση σε μια εργασία. Δέκα γυναίκες και δέκα άντρες υποβάλλονται στο τεστ που μετράει την επιτηδειότητα στα χέρια. Η επίδοση στην εκτέλεση της δουλειάς,  $Y$ , και ο βαθμός του τεστ επιδεξιότητας,  $X$ , είναι συνεχείς μεταβλητές. Κάθε παρατήρηση δίνει ένα βαθμό «0» για τους άνδρες και «1» για τις γυναίκες. Αυτή η κωδικοποίηση, δημιουργεί την ψευδομεταβλητή μεταβλητή  $X_1$ .

Η παραπάνω προσέγγιση επιτρέπει στον προβλέποντα να προβεί σε μια συνολική πρόβλεψη αναφορικά με το τεστ ικανότητας εκτέλεσής της δουλειάς με βάση τα δεδομένα που αφορούν το φύλο των εργαζομένων. Παραπέρα, κάποιος μπορεί να ελέγξει τη σημαντικότητα της μεταβλητής κατηγοριοποίησης και να εκτιμήσει τη σχετική συνεισφορά που έχει η μεταβλητή του φύλου των εργαζομένων στην ακρίβεια της πρόβλεψης.

Για να περιγράψουμε τη χρήση των ψευδομεταβλητών, θεωρούμε το 0 για τους άνδρες και το 1 για τις γυναίκες και η πολλαπλή παλινδρόμηση γράφεται:

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2$$

όπου

$$X_1 = \text{ικανότητα εκτέλεσης της δουλειάς}$$

$$X_2 = \begin{cases} 0 & \text{για τους άνδρες} \\ 1 & \text{για τις γυναίκες} \end{cases}$$

Η ενιαία εξίσωση είναι ανάλογη με τις ακόλουθες δύο εξισώσεις:

$$\hat{Y} = b_0 + b_1 X_1 \quad \text{για τους άνδρες}$$

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 \quad \text{για τις γυναίκες}$$

Σημειώτεον ότι το  $b_2$  παριστάνει την ικανότητα μιας γυναίκας στην εκτέλεση της δουλειάς και  $b_1$  είναι οι διαφορές στην εκτίμηση εκτέλεσης της δουλειάς (η τιμή  $b_1$  υποτίθεται ότι είναι ίδια για

γυναίκες και άνδρες). Εκείνο που έχει σημασία είναι ότι μια πολλαπλή παλινδρόμηση του παραπάνω τύπου οδηγεί στην εκτίμηση δυο γραμμών, μία για τους άνδρες και μία για τις γυναίκες. Ίσως κάποιος αντιμετωπίσει τη  $X_2$  σαν μεταβλητή *μετατόπισης* και λειτουργεί μόνον όταν η παρατήρηση αφορά γυναίκα, ενώ δεν λειτουργεί όταν αφορά άνδρα.

Μια πολύ χρήσιμη εφαρμογή της χρησιμοποίησης ψευδομεταβλητών στην παλινδρόμηση είναι η εκτίμηση της εποχικής διακύμανσης σε μια χρονοσειρά. Προηγουμένως παρατηρήσαμε ότι οι τεχνικές πολλαπλής παλινδρόμησης χωρίζουν τη συνολική διακύμανση της εξαρτημένης μεταβλητής σε ένα μη-ερμηνεύσιμο τμήμα και σε ένα τμήμα που εξηγείται από την εξίσωση παλινδρόμησης. Αυτό υποδεικνύει τη δυνατότητα εκτίμησης του συντελεστή εποχιακότητας της χρονοσειράς. Η εξαρτημένη μεταβλητή θα είναι οι παρατηρήσεις της αρχικής χρονοσειράς και οι ανεξάρτητες μεταβλητές θα είναι ως εξής:

$X_1$  χρόνος, όπου η πρώτη παρατήρηση παίρνει την τιμή 1, η δεύτερη παρατήρηση την τιμή 2, και ούτω καθεξής

$X_2$  όλες οι παρατηρήσεις του δευτέρου τριμήνου παίρνουν την τιμή 1 και 0 για κάθε άλλο τρίμηνο (ψευδομεταβλητή)

$X_3$  όλες οι παρατηρήσεις του τρίτου τριμήνου παίρνουν την τιμή 1 και 0 σε κάθε άλλη περίπτωση,

$X_4$  όλες οι παρατηρήσεις του τετάρτου τριμήνου παίρνουν την τιμή 1 και 0 για τα άλλα τρίμηνα.

Σημειωτέον ότι οι παρατηρήσεις του πρώτου τριμήνου παίρνουν την τιμή 0. Επομένως υπάρχουν τέσσερις ανεξάρτητες μεταβλητές στην εξίσωση και η συναρτησιακή μορφή έχει ως εξής:

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4$$

### **2.2.7 Η «Άριστη» Εξίσωση Παλινδρόμησης**

Στην πρόβλεψη η επίτευξη υψηλού βαθμού προβλεπτικότητας, δηλαδή ενός συντελεστή προσδιορισμού κοντά στο ένα, δεν αποτελεί και το τέλος της όλης διαδικασίας. Ενδιαφερόμαστε να επιλέξουμε ένα πλήρες σύνολο δυνητικών ανεξάρτητων μεταβλητών που όλες προσθέτουν στην προβλεπτική ακρίβεια του μοντέλου. Θα πρέπει να συμπεριλαμβάνουμε μόνο εκείνες τις μεταβλητές που μπορούν να συμβάλουν στη βελτίωση της προβλεπτικής ικανότητας του μοντέλου. Παραπέρα βρισκόμαστε αντιμέτωποι με το δίλημμα του να έχουμε την πιο ακριβή πρόβλεψη με το μικρότερο δυνατό κόστος. Εφόσον χρειάζονται χρήματα για την απόκτηση πληροφοριών που αφορούν ένα μεγάλο αριθμό ανεξάρτητων μεταβλητών, στο μοντέλο θα πρέπει να συμπεριληφθεί όσο το δυνατό μικρότερος αριθμός σημαντικών ανεξάρτητων μεταβλητών. Εξισώσεις που έχουν μεγάλο αριθμό ανεξάρτητων μεταβλητών απαιτούν σχετικά μεγάλο μέγεθος δείγματος. Δείγματα μεγάλου μεγέθους

προϋποθέτουν αυξημένα έξοδα για την ακριβή πρόβλεψη της εξαρτημένης μεταβλητής.

Η επιλογή του «άριστου» μοντέλου παλινδρόμησης, γενικά, προϋποθέτει συμβιβασμό μεταξύ της επίτευξης ενός υψηλότερου συντελεστή προσδιορισμού και της επιλογής ενός ελάχιστου αριθμού μεταβλητών πρόβλεψης. Δεν υπάρχει μια μοναδική στατιστική διαδικασία για την επίτευξη αυτού του συμβιβασμού. Όμως μπορούν να γίνουν διάφορες προτάσεις.

Σαν πρώτο βήμα, ο διενεργών μια πρόβλεψη θα πρέπει να καταγράψει εκείνες τις δυνητικές ανεξάρτητες μεταβλητές που κρίνονται χρήσιμες στην πρόβλεψη. Ακολουθώντας αυτή την καταγραφή, ένα δεύτερο βήμα είναι η απόρριψη εκείνων των μεταβλητών πρόβλεψης που δεν φαίνονται κατάλληλες. Μερικά από τα κριτήρια που χρησιμοποιούνται για την απόρριψη αυτών των μεταβλητών είναι τα ακόλουθα:

1. Εξαιρείται η μεταβλητή που δεν εμφανίζει ευλογοφανή αιτιολογία για την ύπαρξή της.
2. Απορρίπτουμε τη μεταβλητή που μπορεί να είναι πανομοιότυπη με άλλες μεταβλητές (συγγραμμικότητα).
3. Απορρίπτουμε τη μεταβλητή που μπορεί να υπόκειται σε μεγάλα σφάλματα μετρήσεων ή αν είναι αδύνατο να μετρηθεί.

Το τελικό βήμα είναι να μειώσουμε τον κατάλογο των δυνητικών ανεξάρτητων μεταβλητών προκειμένου να έχουμε ένα λογικά καλό μοντέλο πρόβλεψης, επιτυγχάνοντας κάποιο συμβιβασμό μεταξύ των υψηλών συντελεστών προσδιορισμού και του κόστους. Αν και η προσωπική αξιολογική κρίση διαδραματίζει σπουδαίο ρόλο, υπάρχουν στη διάθεσή μας αρκετές τεχνικές που μπορούν μόνο να μας βοηθήσουν στον περιορισμό του καταλόγου των δυνητικών ανεξάρτητων μεταβλητών. Οι δύο στατιστικές προσεγγίσεις ονομάζονται: (α) όλες οι δυνατές παλινδρομήσεις και (β) κατά στάδια παλινδρόμηση. Και οι δύο αυτές προσεγγίσεις συνήθως διενεργούνται με καλό στατιστικό software (όπως SPSS).



**2.3 ΟΙΚΟΝΟΜΕΤΡΙΚΑ ΜΟΝΤΕΛΑ [11], [12], [13], [14]****2.3.1 Εισαγωγή**

Κατά γράμμα, η λέξη *οικονομετρία* σημαίνει οικονομική μέτρηση. Η μέτρηση αποτελεί ένα πολύ σημαντικό στοιχείο της οικονομετρίας. Όμως, δεν είναι οικονομετρία όλες οι οικονομικές μετρήσεις. Η μέτρηση του εθνικού εισοδήματος ή η κατασκευή του δείκτη τιμών καταναλωτή αποτελούν πολύ σημαντικά προβλήματα οικονομικής μέτρησης που όμως, δεν είναι οικονομετρικά προβλήματα.

Η Οικονομετρία είναι ο επιστημονικός κλάδος που ασχολείται με τη θεμελίωση ποσοτικών σχέσεων μεταξύ οικονομικών μεταβλητών με τη βοήθεια της στατιστικής θεωρίας. Στα δύο προηγούμενα κεφάλαια, είδαμε πως η θεωρία απλής παλινδρόμησης είναι μια ειδική περίπτωση της πολλαπλής παλινδρόμησης. Αντικείμενο μας είναι να αποδείξουμε ότι η πολλαπλή παλινδρόμηση είναι μια σημαντική ειδική περίπτωση της οικονομετρίας. Ένα οικονομετρικό μοντέλο συνήθως περιλαμβάνει αρκετές σχέσεις με μία ή περισσότερες μεταβλητές, οι οποίες εμφανίζονται σε περισσότερες από μία εξισώσεις του μοντέλου.

Προβλήματα ανακύπτουν όταν υπολογίζουμε μια εξίσωση κάθε φορά και, στη συνέχεια, υπολογίζουμε τις παραμέτρους που εμφανίζονται σ' αυτή με μεθόδους παλινδρόμησης. Η απλή μηχανιστική εφαρμογή αυτής της μεθόδου μπορεί να μας οδηγήσει σε αποτελέσματα χωρίς έννοια. Για παράδειγμα, οι μέθοδοι παλινδρόμησης μπορεί να οδηγήσουν στην εκτίμηση συντελεστών παλινδρόμησης που δεν είναι καλοί εκτιμητές των αληθών τιμών των παραμέτρων των οικονομετρικών εξισώσεων. Επίσης, η πολλαπλή παλινδρόμηση μπορεί να έχει σαν αποτέλεσμα την εκτίμηση της παραμέτρου για ένα συντελεστή μιας ανεξάρτητης μεταβλητής σε μια εξίσωση του μοντέλου. Όμως, η μεταβλητή εμφανίζεται σε περισσότερες της μιας εξισώσεις και ο εκτιμηθείς συντελεστής μπορεί να μην έχει κανένα νόημα στη δεύτερη εξίσωση.

Σε ένα σύνολο εξισώσεων που συνθέτουν ένα οικονομετρικό μοντέλο, επανεμφανίζονται ορισμένες από τις ερμηνευτικές (ανεξάρτητες) μεταβλητές. Μερικές φορές, όπως στις σχέσεις ζήτησης για έναν αριθμό διαφορετικών προϊόντων, δημιουργείται πρόβλημα ταυτοποίησης ή ταυτόχρονης εκτίμησης. Στις περιπτώσεις αυτές οι οικονομετρικές εξισώσεις μπορούν να αντιμετωπιστούν η μία μετά την άλλη.

Ενώ στις προηγούμενες μεθόδους η διάκριση ανάμεσα στις εξαρτημένες και τις ανεξάρτητες μεταβλητές ήταν πολύ σημαντική, για ένα σύστημά δύο ή περισσότερων εξισώσεων που έχουν αλληλοεξαρτώμενες μεταβλητές, η διάκριση αυτή δεν ισχύει. Αντί γι' αυτό, η σημαντική διάκριση βρίσκεται μεταξύ ταυτόχρονα εξαρτημένων και προκαθορισμένων μεταβλητών. Οι τελευταίες ασκούν μια επίδραση στις ταυτόχρονα (jointly) εξαρτημένες μεταβλητές, αλλά χωρίς να επηρεάζονται από αυτές.

Μια μεταβλητή με χρονική υστέρηση, που αναφέρεται σε προηγούμενη περίοδο, είναι μεταβλητή που τη χειριζόμαστε σαν προκαθορισμένη. Ωστόσο, οι τρέχουσες τιμές ορισμένων

μεταβλητών μπορούν επίσης να είναι προκαθορισμένες. Τέτοιες μεταβλητές ονομάζονται *εξωγενείς* εφόσον καθορίζονται έξω από το σύστημα εξισώσεων. Μεταβλητές που καθορίζονται μέσα στο σύστημα ονομάζονται *ενδογενείς*. Οι τρέχουσες τιμές των ενδογενών μεταβλητών αποτελούν τις συνδυασμένα εξαρτημένες μεταβλητές και ο όρος *ενδογενείς* μπορεί, με τη σειρά του, να χρησιμοποιηθεί αντί του όρου *ταυτόχρονα εξαρτημένες*.

Το θεωρητικό σχήμα του οικονομετρικού μοντέλου καθορίζει αν μια μεταβλητή θα ταξινομηθεί ως εξωγενής ή ως ενδογενής. Μια μη-οικονομική μεταβλητή συνήθως θεωρείται εξωγενής. Όμως, οικονομικές μεταβλητές όπως οι τιμές ή το κόστος παραγωγής, τις μεταχειριζόμαστε άλλοτε σαν ενδογενείς, άλλοτε σαν εξωγενείς.

Ένα πλήρες οικονομετρικό μοντέλο που αποτελείται από ένα σύστημα ταυτόχρονα προσδιοριζόμενων εξισώσεων θα περιέχει τόσες εξισώσεις, όσες και ενδογενείς μεταβλητές. Όμως, μπορεί να υπάρχει οποιοσδήποτε αριθμός εξωγενών μεταβλητών και ίσως να είναι περισσότερες ή λιγότερες στον αριθμό από τις ενδογενείς μεταβλητές.

### 2.3.2 Εξισώσεις Οικονομετρικών Μοντέλων

Ας θεωρήσουμε, αρχίζοντας, από ένα απλό γραμμικό σύστημα προσφοράς-ζήτησης για την τιμή  $p$ , και την ποσότητα  $q$ , που καθορίζονται από το ακόλουθο σύστημα δύο ταυτόχρονα προσδιοριζόμενων εξισώσεων

$$q = b_{01} + b_1 p + e \quad (2.45)$$

τη σχέση ζήτησης και

$$q = b_{02} + b_2 p + u \quad (2.46)$$

τη σχέση προσφοράς.

Οι όροι  $e$  και  $u$  είναι οι όροι του σφάλματος ή διαταραχής, τα  $b_{01}$  είναι σταθερές και τα  $b_1$  είναι συντελεστές παλινδρόμησης για τις μεταβλητές της δεξιάς πλευράς (right-hand side: RHS). Κάθε εξίσωση είναι συνάρτηση της τιμής και μπορούμε να θεωρήσουμε τα  $p$  και  $q$  ως ενδογενείς μεταβλητές. Όταν υπάρχουν δύο ενδογενείς μεταβλητές οι τιμές των συντελεστών τους καθορίζονται από δύο ταυτόχρονα προσδιοριζόμενες εξισώσεις.

Οι παραπάνω εξισώσεις ονομάζονται *διαρθρωτικές εξισώσεις* του μοντέλου. Δείχνουν τις σχέσεις μεταξύ μεταβλητών στα πλαίσια μιας ήδη ισχύουσας ή νέας οικονομικής θεωρίας που εξηγεί τη συμπεριφορά των οικονομικών φαινομένων. Οι συντελεστές παλινδρόμησης αναφέρονται ως *διαρθρωτικοί συντελεστές* του μοντέλου.

Για να εξετάσουμε παραπέρα ένα οικονομετρικό μοντέλο προσδιορισμού τιμών σε μία μόνο (single) αγορά, ας θεωρήσουμε το ακόλουθο σύστημα εξισώσεων

$$q_d = b_{01} + b_1 p + c_1 I + e \quad (2.47)$$

$$q_d = b_{02} + b_2 p + c_2 R + u \quad (2.48)$$

και

$$q_d = q_s \quad (2.49)$$

όπου  $q_d$  είναι η τριμηνιαία ζήτηση ενός αγαθού.  $q_s$  είναι η προσφερόμενη ποσότητα.  $p$  είναι η τιμή.  $I$  είναι το εισόδημα.  $R$  είναι δείκτης βροχόπτωσης και  $e$ ,  $u$  είναι οι όροι του σφάλματος. Οι ενδογενείς μεταβλητές είναι οι  $q_d$ ,  $q_s$  και  $p$ , ενώ οι εξωγενείς είναι οι  $I$  και  $R$ . Το τυχαίο σφάλμα, οι όροι  $e$  και  $u$ , συμπεριλαμβάνεται στην εξίσωση για έναν ή περισσότερους από τους παρακάτω λόγους:

1. Μεταβλητές που μπορεί να επηρεάσουν τη ζήτηση ή την προσφορά έχουν παραλειφθεί από τις ταυτόχρονα προσδιοριζόμενες εξισώσεις.

2. Σφάλματα που μπορεί να οφείλονται στη μη-ικανοποιητική εξειδίκευση του μοντέλου με την έννοια ότι οι συγκεκριμένες συναρτησιακές μορφές που έχουν επιλεγεί μπορεί να είναι μη-γραμμικές, ενώ έχουν χρησιμοποιηθεί ως γραμμικές.

3. Μεταβλητές που συμπεριλαμβάνονται μπορεί να μην έχουν μετρηθεί με ακρίβεια και ίσως υπάρχει κάποια βασική τυχαιότητα στη συμπεριφορά από μέρους τόσο των καταναλωτών όσο και των προμηθευτών.

Έξι ρητές παράμετροι συμπεριλαμβάνονται στο μοντέλο. Τα  $b_{01}$  και  $b_{02}$  είναι σταθερές. Τα  $b_1$  και  $b_2$  πολλαπλασιάζουν το  $p$  (τιμή). το  $c_1$  πολλαπλασιάζει το  $I$  (εισόδημα) και το  $c_2$  πολλαπλασιάζει το  $R$  (ασφάλιστρα έναντι βροχών). Το μοντέλο περιλαμβάνει επίσης μη-ρητές υποθέσεις, που αφορούν την κατανομή πιθανότητας του όρου του σφάλματος κάθε εξίσωσης.

Η εξίσωση της ζήτησης δείχνει ότι η ζήτηση είναι μια γραμμική συνάρτηση της τιμής, του εισοδήματος και του σφάλματος. Η παράμετρος  $b_1$  θα είναι αρνητική (δείχνοντας μία προς τα κάτω κλίση της καμπύλης της ζήτησης) και η παράμετρος  $c_1$  είναι γενικά θετική (το επιπλέον εισόδημα αυξάνει τη ζήτηση).

Για την εξίσωση προσφοράς, η  $b_2$  είναι θετική (η προσφερόμενη ποσότητα είναι θετική συνάρτηση της τιμής). η  $C_2$  είναι γενικά θετική δείχνοντας ότι, αύξηση του δείκτη βροχόπτωσης αυξάνει την προσφορά εμπορεύσιμων αγαθών.

Η δεξιά πλευρά των εξισώσεων (6.3) και (6.4) περιλαμβάνει δύο κατηγορίες τυχαίων μεταβλητών για κάθε εξίσωση. Η πρώτη είναι οι όροι του σφάλματος ( $e$  και  $u$ ) και η δεύτερη είναι η  $P$  που είναι ενδογενής, επομένως, δέχεται επιδράσεις και από τους δύο όρους του σφάλματος. Συνεπώς, οι ποσότητες της αριστεράς πλευράς,  $q_d$  και  $q_s$  είναι επίσης τυχαίες. Μια υπόθεση του όρου του σφάλματος είναι ότι

$$E(e)=0 \quad (2.50)$$

και

$$E(u)=0$$

Επομένως, οι αναμενόμενες τιμές των εξισώσεων (2.47) και (2.48) είναι:

$$E(q_d) = b_{01} + b_1 E(p) + c_1 I \quad (2.51)$$

και

$$E(q_d) = b_{02} + b_2 E(p) + c_2 R \quad (2.52)$$

Η εξίσωση (2.49) εκφράζει τη συνθήκη ισορροπίας, όπου η ζήτηση ισούται με την προσφορά. Η εξίσωση αυτή δεν περιέχει όρο σφάλματος (είναι μια ταυτότητα). Επομένως, η λύση ισορροπίας μπορεί να ευρεθεί με την εξίσωση των δύο ταυτόχρονα προσδιοριζόμενων εξισώσεων και με την έκφραση των ενδογενών μεταβλητών σαν συνάρτηση όλων των εξωγενών μεταβλητών και των όρων του σφάλματος. Η μέθοδος αυτή, που ονομάζεται μέθοδος της ανηγμένης μορφής, επιτρέπει τη λύση του συστήματος.

Για να μελετήσουμε το πώς μια εταιρία μπορεί να χρησιμοποιήσει στην πρόβλεψη της ένα σύστημα εξισώσεων, θεωρούμε το παρακάτω παράδειγμα.

### Παράδειγμα 2.7

Το ακόλουθο σύστημα πέντε εξισώσεων είναι μια απλή μέθοδος εφαρμογής οικονομικού μοντέλου για την πρόβλεψη πωλήσεων μιας εταιρίας που λειτουργεί κάτω από γενικές επιχειρηματικές συνθήκες. Αν και πολλά παρόμοια μοντέλα πρόβλεψης μπορεί να είναι περισσότερο πολύπλοκα, αυτό το μοντέλο χρησιμοποιείται ως σημείο εκκίνησης.

$$C_t = b_{01} + b_1 Y_t + e_{1t} \quad (2.53)$$

$$I_t = b_{02} + b_2 P_{t-1} + e_{2t} \quad (2.54)$$

$$Y_t = C_t + Y_t + G_t \quad (2.55)$$

$$(DY)_t = Y_t - A_t - T_t \quad (2.56)$$

$$S = b_{03} + b_3 (DY)_t \quad (2.57)$$

όπου  $C$  είναι οι προσωπικές δαπάνες κατανάλωσης  $I$  είναι οι επενδύσεις παγίου κεφαλαίου (καθαρό).  $P$  είναι τα κέρδη.  $G$  είναι οι κυβερνητικές δαπάνες για αγαθά και υπηρεσίες  $Y$  είναι το ακαθάριστο εθνικό προϊόν (gross national product=GNP).  $DY$  είναι το διαθέσιμο προσωπικό εισόδημα.  $A$  είναι το επιτρεπόμενο ύψος απόσβεσης κεφαλαίου.  $T$  είναι οι φόροι και οι μεταβιβαστικές πληρωμές  $S$  είναι οι πωλήσεις του προϊόντος  $X$  και  $e_1, e_2$  οι όροι του σφάλματος.

Οι εξισώσεις (2.52), (2.53) και (2.54) εκφράζουν υποθέσεις συμπεριφοράς. Η πρώτη υποδηλώνει ότι η κατανάλωση της τρέχουσας περιόδου είναι μια συνάρτηση του τρέχοντος επιπέδου του

ακαθάριστου εθνικού προϊόντος. Η δεύτερη δείχνει ότι η τρέχουσα επένδυση κεφαλαίου σχετίζεται με τα κέρδη της προηγούμενης περιόδου. Τέλος, η (2.57) υποδηλώνει ότι οι πωλήσεις του προϊόντος  $X$  σχετίζονται με το διαθέσιμο προσωπικό εισόδημα. Η ταυτότητα (2.56) δείχνει ότι το GNP είναι ίσο με την κατανάλωση, την επένδυση και τις κυβερνητικές δαπάνες, και η ταυτότητα (2.55) δηλώνει ότι το διαθέσιμο προσωπικό εισόδημα ισούται με το GNP μείον τις αποσβέσεις κεφαλαίου, τους Φόρους, και τις μεταβιβαστικές πληρωμές.

Η βασική σύλληψη του οικονομετρικού μοντέλου πρόβλεψης είναι ότι οι πωλήσεις μιας εταιρίας (ή προϊόντος) σχετίζονται με τις γενικές οικονομικές και επιχειρηματικές συνθήκες, που εκφράζονται στο μοντέλο πρόβλεψης των πολλών εξισώσεων. Αυτή η αλληλεξάρτηση ασκεί μεγάλη πίεση στις υπηρεσίες πρόβλεψης για τη συλλογή δεδομένων, τη διενέργεια στατιστικών υπολογισμών και την εκτίμηση των παραμέτρων του μοντέλου πρόβλεψης. Βεβαίως, η πρόβλεψη στο τέλος δεν μπορεί να είναι καλύτερη από το χρησιμοποιούμενο οικονομετρικό μοντέλο, που αποτελεί τον κεντρικό πυρήνα του συστήματος πρόβλεψης (απόφαση).

Τα θέματα που ανακύπτουν κατά την κατασκευή του οικονομετρικού μοντέλου είναι παρόμοια με εκείνα της πολλαπλής παλινδρόμησης, με την πρόσθετη δυσκολία ότι αυτά έχουν επεκταθεί στην περίπτωση των ταυτόχρονα προσδιοριζόμενων εξισώσεων. Οι σχετικές ενέργειες της πρόβλεψης περιλαμβάνουν: (α) καθορισμό των μεταβλητών που ενσωματώνονται σε κάθε εξίσωση' (β) προσδιορισμό της συναρτησιακής μορφής των εξισώσεων που θα εκτιμηθούν' (γ) εφαρμογή διαδικασίας ταυτόχρονης εκτίμησης για τη μέτρηση των συντελεστών παλινδρόμησης (δ) εξέταση της ισχύος των υποθέσεων αναφορικά με τους όρους του σφάλματος και επανεκτίμηση όπου κρίνεται απαραίτητο και (ε) διαγνωστικό έλεγχο των αποτελεσμάτων με στατιστικές τεχνικές. Οι ενέργειες (α) και (γ) είναι νέες, ενώ οι (β), (δ) και (ε) εξετάστηκαν σε προηγούμενες παραγράφους. Πρέπει, ωστόσο, να σημειώσουμε, ότι, οι ενέργειες αυτές σχετίζονται όλες μεταξύ τους και δεν μπορούν να διαχωριστούν πλήρως. Τέλος, ο ρόλος της οικονομικής θεωρίας εμφανίζεται ενισχυμένος.

### 2.3.3 Οφέλη και Περιορισμοί των Οικονομετρικών Μοντέλων στις Προβλέψεις

Ένα από τα πιο σημαντικά οφέλη των οικονομετρικών μοντέλων είναι ότι παρέχουν στον προβλέποντα ένα πλήρες και ρητά εκφρασμένο σύστημα, με το οποίο μπορεί να αξιοποιήσει τις οικονομικές πληροφορίες με συστηματικό και κατανοητό τρόπο. Η συγκέντρωση, συλλογή και παρουσίαση των δεδομένων επιτρέπει στον αναλυτή να επικοινωνεί με τα ευρήματα εύκολα και να προσφέρει, στους χρήστες όλες τις απαιτούμενες πληροφορίες. Επίσης, η επίδραση ή το βάρος της πληροφόρησης όπως μετράται από τους συντελεστές παλινδρόμησης εκφράζεται συστηματικά (formally) στο μοντέλο. Έτσι, εάν ο χρήστης της πρόβλεψης δεν συμφωνεί με τους εκτιμηθέντες συντελεστές μπορεί να τους αντικαταστήσει με διαφορετικές τιμές και να εξετάσει τις συνέπειες. Με αυτό τον τρόπο, ένας μπορεί να εξετάσει διάφορα σενάρια κάνοντας προσομοίωση των οικονομετρικών μοντέλων.

Οι οικονομικές σχέσεις σε ένα μοντέλο είναι ρητά εκφρασμένες και αντικειμενικές. Οι

διενεργούντες την πρόβλεψη μπορούν να χρησιμοποιήσουν εναλλακτικές υποθέσεις που αφορούν τις εξωγενείς μεταβλητές του μοντέλου. Στο παράδειγμα 2.7, μπορούμε να εξετάσουμε τις επιδράσεις μιας προτεινόμενης αλλαγής του φόρου στο προσωπικό εισόδημα πάνω στις πωλήσεις. Αν και η πρόβλεψη μπορεί να μην είναι ακριβής, οι επιδράσεις μιας μεταβολής σε μια εξωγενή μεταβλητή μπορούν να εκτιμηθούν.

Τα οικονομετρικά μοντέλα εφοδιάζουν τον προβλέποντα με μια ποσότητα λεπτομερών προβλέψεων. Είναι δυνατό να παρακολουθήσουμε και να αναπαραγάγουμε τις αιτίες τόσο των επιτυχών όσο και των ανακριβών προβλέψεων, όταν γνωρίζουμε τις υποθέσεις που αφορούν τις εξωγενείς μεταβλητές και τη λύση του μοντέλου. Όταν διερευνώνται τα σφάλματα, οι βελτιώσεις στις μελλοντικές προβλέψεις μπορούν να γίνουν με τις κατάλληλες προσαρμογές στο μοντέλο. Εν τούτοις, χωρίς τη μελέτη των λαθών του παρελθόντος, ο προβλέπων, έχοντας κάνει αμφισβητήσιμες υποθέσεις δεν είναι σε θέση να αλλάξει και να βελτιώσει τα μοντέλα.

Η πρόβλεψη δεν είναι η μοναδική εφαρμογή των οικονομετρικών μοντέλων. Η σωστή χρήση τέτοιων μοντέλων μπορεί να μας βοηθήσει να κατανοήσουμε καλύτερα τις οικονομικές σχέσεις και αλληλεξαρτήσεις.

Μας επιτρέπουν να ελέγξουμε και να ξεκαθαρίσουμε ελεγχόμενες υποθέσεις που αφορούν τη θεωρία της επιχειρηματικής συμπεριφοράς. Οι έλεγχοι αυτοί μπορεί να επιτρέψουν στον προ βλέποντα να βρει νέες και πιο βελτιωμένες υποθέσεις και ίσως τον οδηγήσουν σε μια καλύτερη κατανόηση του οικονομικού συστήματος και των οικονομικών διακυμάνσεων και έτσι να εξυπηρετηθούν οι σκοποί της πρόβλεψης.

Αν και τα οικονομετρικά μοντέλα μεγάλης κλίμακας θα μας απασχολούν συνεχώς, ο διενεργών την πρόβλεψη θα πρέπει να γνωρίζει τα όριά τους. Το βασικό πρόβλημα είναι ότι οι παραδοσιακές μέθοδοι παλινδρόμησης απαιτούν να χρησιμοποιηθούν εξαρτημένες μεταβλητές από ορισμένες εξισώσεις σαν ανεξάρτητες μεταβλητές σε άλλα τμήματα ενός πολυδιάστατου συστήματος εξισώσεων. Η παλινδρομική ανάλυση των πειραματικών δεδομένων δείχνει ότι θα μπορούσε να προκληθεί μεροληπτικό σφάλμα, εάν χρησιμοποιούσαμε την εξαρτημένη μεταβλητή από την πειραματική παλινδρόμηση σαν ανεξάρτητη μεταβλητή σε μια δεύτερη παλινδρόμηση. Τα τυχαία σφάλματα από την εξαρτημένη μεταβλητή της πρώτης πειραματικής παλινδρόμησης θα την ανάγκαζαν να συμπεριφερθεί με διαφορετικό τρόπο στη δεύτερη παλινδρόμηση από ό,τι θα συμπεριφερόταν μια ελεγχόμενη μεταβλητή. Μια ελεγχόμενη μεταβλητή θα μπορούσε να έχει τιμές καθορισμένες από τον πειραματιζόμενο και έτσι δεν θα περιέχει τυχαία στοιχεία.

Η πολυσυγγραμμικότητα, ένα σοβαρό πρόβλημα σε μια απλή εξίσωση χρονοσειράς, γίνεται περισσότερο οξύ όταν εξετάζουμε ένα οικονομετρικό μοντέλο. Όταν σε ένα μοντέλο υπολογίζονται πολλοί διαρθρωτικοί συντελεστές μεταξύ συμπεριληφθεισών μεταβλητών, τα προβλήματα αυτά πολλαπλασιάζονται και επεκτείνονται.

Συχνά, μεγάλης κλίμακας και έντονα αθροιστικά μοντέλα δεν ενσωματώνουν άμεσα λεπτομερείς πληροφορίες, που αφορούν εξελίξεις σε επιμέρους τομείς και τμήματα της οικονομίας. Για παράδειγμα, αξιολογοί οικονομικοί στατιστικολόγοι σε μια υπηρεσία, μπορεί να έχουν στη διάθεσή

τους ένα λεπτομερές δίκτυο οικονομικών και στατιστικών πληροφοριών που αφορά ένα συγκεκριμένο τομέα της οικονομίας. Όμως, το οικονομετρικό μοντέλο μπορεί να περιλαμβάνει μια μόνο απλή εξίσωση που αναπαριστά αυτό τον τομέα. Επομένως, το μοντέλο αγνοεί ένα μεγάλο μέγεθος σημαντικών στοιχείων που αφορούν αυτό τον τομέα της οικονομίας.

Τέλος, τα οικονομετρικά μοντέλα δεν μπορούν μόνο τους να λειτουργούν χωρίς συνεχή παρακολούθηση της εξέλιξης των αποτελεσμάτων. Οι περιοδικές αλλαγές είναι αναγκαίες και προϋποθέτουν πεπειραμένους ειδικούς με γνώσεις και εμπειρία στις εφαρμογές της οικονομετρίας. Εταιρίες μεσαίου και μικρού μεγέθους δεν διαθέτουν τους ανθρώπους που έχουν την ικανότητα να κατανοούν και να εφαρμόζουν οικονομικές προβλέψεις και μοντέλα μεγάλης κλίμακας. Επομένως, δημιουργήθηκαν εταιρίες που έχουν ειδικευθεί για να παρέχουν τέτοιου είδους βοήθεια σε επιχειρήσεις για την οικονομετρική τους πρόβλεψη.

### 2.3.4 Πρόβλημα Οικονομετρικής Εκτίμησης

Ένας από τους στόχους της εκτίμησης είναι η απόκτηση αμερόληπτων εκτιμήσεων για τις παραμέτρους του πληθυσμού. Παραπέρα, για όλο και μεγαλύτερα μεγέθη δείγματος, η ακρίβεια των εκτιμήσεων πρέπει επίσης να αυξάνεται. Εάν το μέγεθος του δείγματος είναι ίσο με το σύνολο των στοιχείων του πληθυσμού, οι εκτιμητές θα πρέπει να είναι ίδιοι με την τιμή των παραμέτρων του πληθυσμού.

Οι εκτιμητές των παραμέτρων,  $b_0, b_1, b_2, \dots, b_m$  μιας πολλαπλής παλινδρόμησης μπορεί να δεχτεί ότι είναι αμερόληπτοι κάτω από κατάλληλες συνθήκες. Σε οικονομετρικά μοντέλα πολλών εξισώσεων, αυτό δεν συμβαίνει. Το σημείο αυτό μπορεί να διευκρινιστεί με την επανεξέταση των εξισώσεων (2.53) και (2.55). Υπενθυμίζουμε ότι

$$C_t = b_{01} + b_1 Y_t + e_t$$

και

$$Y_t = C_t + I_t + G_t$$

όπου όλοι οι όροι έχουν οριστεί στο παράδειγμα 2.7.

Η ανεξάρτητη μεταβλητή,  $Y_t$  της εξίσωσης 2.53 προσδιορίζεται εν μέρει από το επίπεδο της κατανάλωσης,  $C_t$  μια ενδογενή μεταβλητή, και εν μέρει από τις εξωγενείς μεταβλητές  $I_t$  και  $G_t$ . Οι μεταβλητές αυτές μπορούν τώρα να ταξινομηθούν σε ένα σύστημα εξισώσεων είτε σαν ενδογενείς ( $C_t, Y_t$ ) είτε σαν εξωγενείς ( $I_t, G_t$ ). Ένα πρόβλημα δημιουργείται γιατί οι ενδογενείς μεταβλητές στην εξίσωση (2.53) σχετίζονται μεταξύ τους. Επομένως, η σχέση δημιουργεί εξάρτηση μεταξύ της εξαρτημένης μεταβλητής,  $C_t$ , και του όρου του σφάλματος,  $e_t$  που εμφανίζεται ως εξάρτηση μεταξύ διαδοχικών τιμών του  $e_{it}$ . Έτσι, η υπόθεση της ανεξαρτησίας του όρου του σφάλματος παραβιάζεται.

Για να εξετάσουμε καλύτερα το πρόβλημα, αντικαθιστούμε την εξίσωση (2.55) στην (2.53) ως ακολούθως:

$$C_t = b_0 + b_1(C_t + I_t + G_t) + e_{1t}$$

και

$$C_t - b_1 C_t = C_t(1 - b_1) = a_1 + b_1(I_t + G_t) + e_{1t} \quad (2.58)$$

και

$$C_t = \frac{b_0}{1 - b_1} + \frac{b_1 I_t}{1 - b_1} + \frac{b_1 G_t}{1 - b_1} + \frac{e_{1t}}{1 - b_1}$$

Με αντικατάσταση έχουμε

$$C_t = \frac{b_0}{1 - b_1} \text{ και } b_1 = \frac{I}{I - b_1}$$

έχουμε

$$Y_t = b_0 + b(I_t + G_t) + \frac{e_{1t}}{1 - b_1} \quad (2.59)$$

Η εξίσωση (2.59) υποδηλώνει την ύπαρξη μιας σχέσης μεταξύ της εξαρτημένης μεταβλητής C, και του όρου του σφάλματος  $e_{1t}$ . Το αποτέλεσμα είναι μια μεροληπτική εκτίμηση των συντελεστών  $b_0$  και  $b$  στην (2.59), καθώς και μεροληπτικοί συντελεστές για τα  $b_0$  και  $b$  της εξίσωσης (2.53). Τόσο για τα μικρά όσο και για τα μεγάλα δείγματα, η μεροληψία υπάρχει και μπορεί να προβλεφθεί εάν κάποιος είναι πρόθυμος να υποθέσει ότι η διακύμανση του πληθυσμού είναι γνωστή. Επομένως, η μέθοδος των ελαχίστων τετραγώνων (ordinary least squares, OLS) δεν μπορεί γενικά να χρησιμοποιηθεί αξιόπιστα για την πρόβλεψη με ένα σύστημα ταυτόχρονων εξισώσεων.

### 2.3.5 Διαδικασίες Οικονομετρικής Εκτίμησης

Για να αποφύγουμε ή να μειώσουμε το μεροληπτικό σφάλμα των ταυτόχρονα προσδιοριζόμενων εξισώσεων, έχει αναπτυχθεί ένας αριθμός διαφόρων στατιστικών τεχνικών ή μεθόδων. Τα σχετικά πλεονεκτήματα και μειονεκτήματα αυτών των μεθόδων εξαρτώνται από το μελετούμενο δείγμα και τη χρονική περίοδο που καλύπτει. Ο στόχος μας σε αυτή την παράγραφο είναι να δώσουμε μια σύντομη περιγραφή μερικών από τα πιο σημαντικά. Ακόμη, μία ή περισσότερες από αυτές της τεχνικές υπάρχουν στα περισσότερα προγράμματα software των υπολογιστών, όπως το SPSS, καθώς και σε άλλα που προορίζονται για ειδικές καταστάσεις.

Η μέθοδος της μέγιστης πιθανότητας με πλήρη πληροφόρηση (full information maximum likelihood=FIML) χρησιμοποιεί μια έννοια που ονομάζεται μέγιστη πιθανότητα. Αυτή η έννοια, που είναι μια σύνθετη μαθηματική διαδικασία, οδηγεί σε εκτιμητές που προκύπτουν μετά την ελαχιστοποίηση της ορίζουσας της μήτρας συνδιακυμάνσεων, που σχετίζεται με τα κατάλοιπα της ανηγμένης μορφής του συστήματος εξισώσεων. Το σύστημα αναφέρεται επίσης και ως LGRV και είναι ιδιαίτερα πολύπλοκο από θεωρητική και από υπολογιστική άποψη. Δεν θα ασχοληθούμε άλλο με την εξέταση αυτής της μεθόδου.



Παρόμοια με την FIML είναι η μέθοδος *μέγιστης πιθανότητας με περιορισμένη πληροφόρηση* (limited information likelihood=LIML). Η μέθοδος αυτή αναγνωρίζει μόνο ένα ποσοστό της αλληλεξάρτησης εκτιμώντας τις παραμέτρους κάθε εξίσωσης μία-μία αλλά μόνο σε σχέση με τις εξωγενείς μεταβλητές. Με διαδοχική εκτίμηση, οι εκτιμηθείσες τιμές αντικαθίστανται στην εξίσωση που εκτιμάται. Επομένως, με έναν περιορισμένο τρόπο, η LIML μετρά την ύπαρξη εξάρτησης μεταξύ των μεταβλητών.

Η *έμμεσος μέθοδος ελαχίστων τετραγώνων* (indirect least squares=ILS) βρίσκεται στο άλλο αντίθετο άκρο της υπολογιστικής δυσκολίας και της θεωρητικής αυστηρότητας. Αν και παρόμοια με τις OLS, η ILS εφαρμόζεται στην ανηγμένη μορφή του συστήματος εξισώσεων. Η ανηγμένη μορφή ενός οικονομετρικού μοντέλου αποκτάται μετά από διαδοχικές αντικαταστάσεις στις αρχικές εξισώσεις μέχρις ότου όλες οι ενδογενείς μεταβλητές εκφραστούν ως συναρτήσεις μόνο των εξωγενών μεταβλητών. Για να περιγράψουμε αυτή την τεχνική, ας επιστρέψουμε στις εξισώσεις (2.53 και (2.55) του παραδείγματος 2.7. Η εξίσωση (2.53) μεταβλήθηκε στην εξίσωση (2.58) που είναι η ανηγμένη μορφή

$$C_t = \frac{b_{01}}{1-b_1} + \frac{b_1}{1-b_1} I_t + \frac{b_1}{1-b_1} G_t + \frac{e_{1t}}{1-b_1}$$

Αφού η  $C_t$ , η εξαρτημένη μεταβλητή εκφράζεται σε όρους των εξωγενών μεταβλητών  $I_t$ ,  $G_t$ . Η εξίσωση (2.55) μπορεί επίσης να εκφραστεί σε ανηγμένη μορφή με αντικατάσταση της (2.53) στην (2.55):

$$Y_t = b_{01} + b_1 Y_t + I_t + G_t + e_{1t}$$

και

$$Y_t(1-b_1) = b_{01} + I_t + G_t + e_{1t}$$

και

$$Y_t = \frac{b_{01}}{1-b_1} + \frac{b_1}{1-b_1} (I_t + G_t) + \frac{e_{1t}}{1-b_1} \quad (2.60)$$

Οι εξισώσεις (2.58) και (2.60), είναι οι ανηγμένες μορφές των εξισώσεων (2.53) και (2.55). Με αντικατάσταση των

$$b_0 = \frac{b_{01}}{1-b_1}, b = \frac{b_1}{1-b_1}, c = \frac{1}{1-b_1}$$

στις (6.14) και (6.16) παίρνουμε τα ακόλουθα:

$$C_t = b_0 + b(I_t + G_t) + \frac{e_{1t}}{1-b_1} \quad (2.61)$$

$$C_t = b_0 + c(I_t + G_t) + \frac{e_{1t}}{1-b_1} \quad (2.62)$$

Οι εξισώσεις (2.61) και (2.62) μπορούν τώρα να εκτιμηθούν με τη μέθοδο ελαχίστων

τετραγώνων (OLS), αφού οι μεταβλητές  $I_t$  και  $G_t$  είναι εξωγενείς. Οι εκτιμητές των συντελεστών  $b$  και  $c$  είναι αμερόληπτοι και συνεπείς καθώς αυξάνεται το μέγεθος του δείγματος. Εν τούτοις, οι τιμές  $b_{01}$  και  $b_1$  που είναι συναρτήσεις των  $b_0$  και  $b$  είναι μεροληπτικές αλλά συνεπείς. Αν και η μέθοδος της ILS είναι σχετικά απλή, η υπολογιστική της συνθετότητα οδήγησε πολλούς οικονομόμετρους και ασχολούμενους με την πρόβλεψη στη χρήση άλλων διαδικασιών.

Η μέθοδος των ελαχίστων τετραγώνων σε δύο στάδια (the two-stage least squares=2SLS) συνδυάζει μερικά από τα πλεονεκτήματα και τους περιορισμούς των μεθόδων IML και IIS. Η σχετική διαδικασία είναι πρακτική, οι εκτιμητές είναι συνεπείς και η μεροληψία είναι μικρή για μεγαλύτερα μεγέθη δείγματος. Το πρώτο βήμα είναι η επιλογή μιας από τις ενδογενείς μεταβλητές ως της ανεξάρτητης μεταβλητής. Στο προηγούμενο μοντέλο,  $Y$  είναι η ανεξάρτητη μεταβλητή και θα έπρεπε να επιδιώξουμε την ελαχιστοποίηση της εξάρτησης του  $C$  από το  $e$ . Κι αυτό επιτυγχάνεται με την εφαρμογή της OLS στην ανηγμένη μορφή της εξίσωσης (2.62) που χρησιμοποιείται για την εκτίμηση των συντελεστών  $b_0$  και  $c$ . Τα αποτελέσματα αντικαθίστανται με τη σειρά τους στην αρχική εξίσωση (2.53) ως εξής:

$$C_t = b_{01} + b_1 Y_t + e_{1t} \quad (2.63)$$

$$Y_t = b_0 + c(I_t + G_t) \quad (2.64)$$

Με αντικατάσταση της (6.20) στην (6.19) παίρνουμε

$$C_t = b_{01} + b_1 \alpha + b_1 c(I_t + G_t) + e_{1t} \quad (2.65)$$

ή

$$C_t = b_{03} + b_3(I_t + G_t) + e_t \quad (2.66)$$

όπου

$$b_{03} = b_{01} + b_1 b_0$$

$$b_3 = b_1 c$$

Η εξίσωση (2.66) περιλαμβάνει εξωγενείς μεταβλητές και το  $G_t$  δεν σχετίζεται με τον όρο του σφάλματος  $e_t$ . Έτσι, η διαδικασία αυτή οδηγεί σε αμερόληπτους και σε συνεπείς εκτιμητές. Ένα σημαντικό μειονέκτημα της 2SLS είναι ότι αγνοεί την πλήρη έκταση της αλληλεξάρτησης μεταξύ των διαφόρων εξισώσεων. Κι αυτό συμβαίνει γιατί η 2SLS είναι μια διαδικασία διαδοχικής εκτίμησης που έχει ως αποτέλεσμα την απώλεια μέρους της εξάρτησης.

Επέκταση της μεθόδου 2SLS είναι η μέθοδος ελαχίστων τετραγώνων σε τρία στάδια (three-stage least squares=3SLS) που στις περισσότερες περιπτώσεις έχει ως αποτέλεσμα την εύρεση πιο αποτελεσματικών εκτιμητών. Αν και η μέθοδος πραγματεύεται την αλληλεξάρτηση με μεγαλύτερη πληρότητα και για μεγαλύτερα δείγματα μας δίνει ασυμπτωτικά πιο αποτελεσματικούς εκτιμητές, δεν χρησιμοποιείται τόσο πολύ όσο η 2SLS. Η μέθοδος 3SLS από υπολογιστική άποψη είναι πολύ πιο δύσκολη γιατί απαιτεί περισσότερο χώρο στα προγράμματα του υπολογιστή και μεγαλύτερη χρήση της μνήμης του υπολογιστή, πράγμα που συνεπάγεται αύξηση του κόστους.

**2.3.6 Το Πρόβλημα της Εξειδίκευσης**

Ένα από τα περιοριστικά προβλήματα που σχετίζονται με όλες τις διαδικασίες, οι οποίες βασίζονται στην παλινδρόμηση, είναι ότι η παράλειψη μιας σημαντικής ερμηνευτικής μεταβλητής θα δημιουργήσει μεγαλύτερο σφάλμα πρόβλεψης. Σε ένα οικονομετρικό μοντέλο προκύπτει μη-τυχαίο σφάλμα από την εσφαλμένη εξειδίκευση μιας εξίσωσης. Το σφάλμα γίνεται μεγαλύτερο στα οικονομετρικά μοντέλα γιατί επηρεάζονται όλες οι εξισώσεις όταν από μια εξίσωση παραλείπεται μια σημαντική μεταβλητή. Παραπέρα, έχουμε αυτοσυσχέτιση των όρων του σφάλματος και δεν υπάρχει επαρκής ένδειξη της πηγής αυτού του προβλήματος. Τέλος, εφόσον το οικονομετρικό μοντέλο περιλαμβάνει πολλές εξισώσεις, υπάρχει μια μεγαλύτερη πιθανότητα ότι θα προκύψει πρόβλημα εσφαλμένης εξειδίκευσης.

Οι εξωγενείς μεταβλητές χρειάζονται στις παραπάνω διαδικασίες εκτίμησης και η επιλογή τέτοιων μεταβλητών δεν είναι απαλλαγμένη από προβλήματα. Οι διενεργούντες την πρόβλεψη θα επιλέξουν εκείνες τις εξωγενείς μεταβλητές που δεν προσδιορίζονται μέσα στο σύστημα και αυτές οι επιλογές μπορεί συχνά να είναι αυθαίρετες. Για παράδειγμα, στο οικονομετρικό μοντέλο του παραδείγματος 2.7, οι κυβερνητικές δαπάνες για αγαθά,  $G_t$  ήταν μια εξωγενής μεταβλητή. Σε ένα πιο λεπτομερές μοντέλο της οικονομίας, το επίπεδο του  $G$ , μπορεί να επηρεάζεται από μεταβολές στις συνολικές οικονομικές συνθήκες, όπως μετρώνται από το GNP. Επομένως, η μεταβλητή που συμβολίζει τις κυβερνητικές δαπάνες μπορεί να χρησιμοποιηθεί ως ενδογενής και άλλες μεταβλητές, όπως το ποσοστό των φόρων ή η νομισματική "πολιτική, θα πρέπει να χρησιμοποιηθούν ως εξωγενείς.

Ακόμη, μια απαίτηση της οικονομετρίας είναι ότι είναι αναγκαία μια εξίσωση για κάθε ενδογενή μεταβλητή που υπάρχει στο μοντέλο. Εάν δεν συμβαίνει αυτό, το μοντέλο είναι υπο-εξειδικευμένο και τουλάχιστον μια ενδογενής μεταβλητή πρέπει να θεωρηθεί ως εξωγενής προκειμένου να διευκολυνθεί η διαδικασία της εκτίμησης. Από την άλλη, όταν υπάρχουν περισσότερες ενδογενείς μεταβλητές από ό,τι εξισώσεις στο σύστημα, το μοντέλο λέγεται ότι είναι υπερ-εξειδικευμένο. Τέτοια υπερ-εξειδίκευση μπορεί να οδηγήσει σε άλλα σφάλματα εκτίμησης και να μειώσει τη σημαντικότητα των παραμέτρων που εκτιμήθηκαν στα πλαίσια ενός συστήματος ταυτόχρονα προσδιοριζόμενων εξισώσεων.

**2.3.7 Το Πρόβλημα της Ταυτοποίησης**

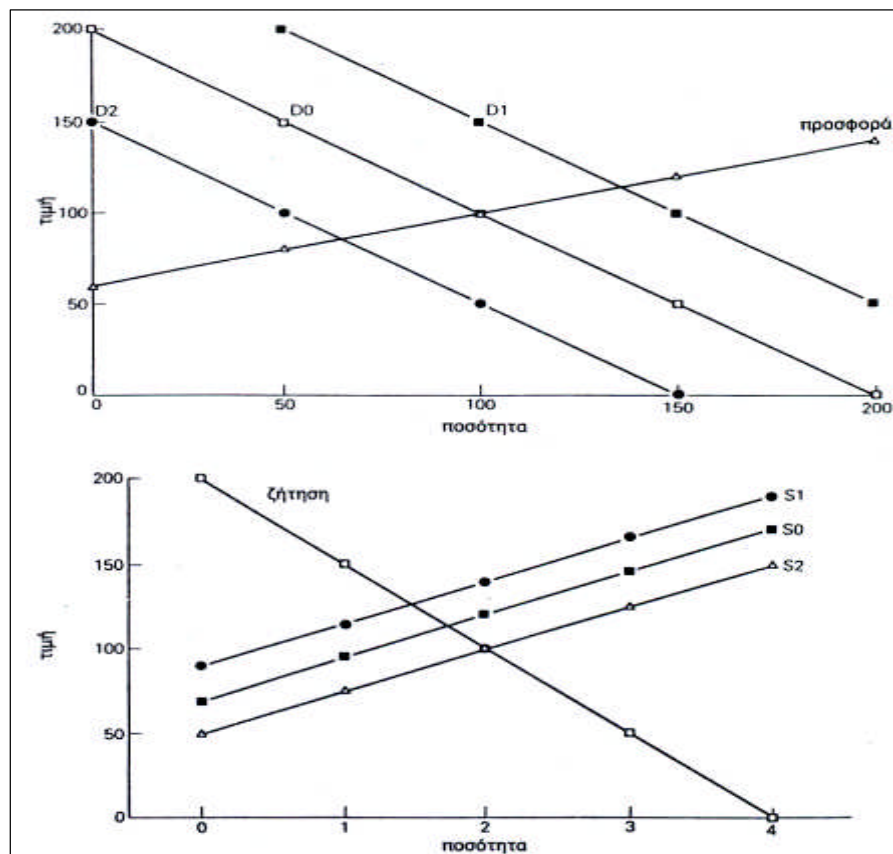
Σχετικό με το πρόβλημα της εξειδίκευσης και της επιλογής των εξωγενών και των ενδογενών μεταβλητών στο χώρο των μοντέλων πολλών εξισώσεων είναι και το πρόβλημα της ταυτοποίησης. Για να δούμε την ουσία αυτού του προβλήματος, ας επιστρέψουμε στο μοντέλο των εξισώσεων (2.45) και (2.46):

$$q = b_{01} + b_1 p + e \quad (2.45)$$

$$q = b_{02} + b_2 p + u \quad (2.46)$$

Η εξίσωση (2.45) εκφράζει τη σχέση ζήτησης και (2.46) είναι η σχέση της προσφοράς, όπου  $q$  είναι η ποσότητα που ζητείται ή προσφέρεται,  $p$  είναι η τιμή, και  $e$  και  $u$  είναι οι όροι του σφάλματος. Όταν τα δεδομένα συλλέγονται από ένα δείγμα, οι παρατηρηθείσες τιμές είναι τα συνδυασμένα αποτελέσματα και των δύο εξισώσεων και δεν αναπαριστούν παρατηρήσεις ξεχωριστά για κάθε εξίσωση. Τα ζεύγη των τιμών  $(p, q)$  είναι τα σημεία τομής (ή ισορροπίας) των καμπυλών προσφοράς-ζήτησης από τη στιγμή που είναι τα μόνα καταγεγραμμένα σημεία. Στο σχήμα 2.3(α) η καμπύλη ζήτησης είναι καθορισμένη και η καμπύλη προσφοράς μεταθετή. Επομένως, πρέπει να αναπτυχθούν εναλλακτικές μέθοδοι χωριστής ταυτοποίησης των σχέσεων προσφοράς και ζήτησης.

Σχήμα 2.3 : (α) Κίνηση κατά μήκος της καμπύλης προσφοράς.  
(β) Κίνηση κατά μήκος της καμπύλης ζήτησης



Μια τέτοια μέθοδος είναι η ταυτοποίηση μέσω πειράματος στην αγορά, όπου υπάρχουν μία ή περισσότερες ερμηνευτικές μεταβλητές. Σε ένα επαναλαμβανόμενο πείραμα, η τιμή ενός προϊόντος,  $p$ , θα μπορούσε να ποικίλλει σε έναν αριθμό καταστημάτων σύμφωνα με ένα προσεκτικά σχεδιασμένο πειραματικό πλάνο. Τα αποτελέσματα των διαφορών στα εισοδήματα των καταναλωτών για κάθε επαναλαμβανόμενο εγχείρημα θα πρέπει επίσης να ληφθούν υπόψη. Θα μπορούσαμε να πάρουμε εκτιμήσεις της συνάρτησης ζήτησης από μια παλινδρόμηση κατά κεφαλή ποσοτήτων που αγοράστηκαν από τους καταναλωτές και των προκαθορισμένων λιανικών τιμών. Η σχέση προσφοράς των παραγωγών (ή πωλητών) απλά δεν θα είναι σχετική με αυτό τον τύπο ελεγχόμενου πειράματος.

Για να εκτιμήσουμε τη συνάρτηση προσφοράς θα πρέπει να εκτελεστεί ένα άλλο ελεγχόμενο πείραμα.

Το δεύτερο πείραμα αφορά αποφάσεις παραγωγών. Αυτές οι αποφάσεις θα μπορούσαν να αναφέρονται σε μεταβολές στα επίπεδα παραγωγής για διαφορετικές τιμές.

Και στα δύο πειράματα οι συναρτήσεις προσφοράς και ζήτησης θα πρέπει να εκτιμηθούν η μία ανεξάρτητα από την άλλη. Δεν θα υπάρχει τρόπος στους αγοραστές να επηρεάσουν τα δεδομένα που διαμορφώθηκαν στο πείραμα της προσφοράς, ούτε οι παραγωγοί θα μπορούσαν να επιδράσουν πάνω στις παρατηρήσεις που έγιναν στο πείραμα της ζήτησης.

Μια εναλλακτική λύση έναντι των ελεγχόμενων πειραμάτων είναι η μέθοδος των *ανεπιβεβαιωμένων μορφών*. Για να εξετάσουμε αυτή τη μέθοδο θεωρούμε το παρακάτω παράδειγμα.

### Παράδειγμα 2.8

Σε μια μελέτη προκειμένου να δείχτεί ότι οι κοινές μετοχές δεν συμπεριφέρονται καλά, όπως οι ομολογίες, κατά τη διάρκεια περιόδων πληθωρισμού, κατασκευάζεται ένα μοντέλο ταυτόχρονα προσδιοριζόμενων εξισώσεων. Οι διαθρωπτικές εξισώσεις περιλαμβάνουν τα  $R_b$  και  $R_s$  τις αποδόσεις των ομολογιών και τις αποδόσεις των κοινών μετοχών. Αυτές είναι ενδογενείς μεταβλητές. Οι δύο διαθρωπτικές σχέσεις έχουν ως εξής:

$$R_b = b_{01} + b_1 R_s + b_2 I + e_1$$

και

$$R_s = b_{02} + b_3 R_b + b_4 E + e_2$$

όπου  $I$  είναι τα επιτόκια και  $E$  είναι οι αμοιβές.

Μπορούμε να λύσουμε τις παραπάνω εξισώσεις για να εκτιμήσουμε ταυτόχρονα την αλληλεξάρτηση των  $R_b$  και  $R_s$ . και να βρούμε

$$R_b = b_{03} + b_5 E + b_6 I + e_3$$

και

$$R_s = b_{04} + b_7 E + b_8 I + e_4$$

Οι εξισώσεις αυτές ονομάζονται εξισώσεις ανηγμένης μορφής του μοντέλου. Οι εξισώσεις αυτές δείχνουν το πώς οι ενδογενείς μεταβλητές είναι συνδυασμένα εξαρτημένες από τις εξωγενείς μεταβλητές (E,I) και τους όρους του σφάλματος του συστήματος. Εφαρμόζοντας τη μέθοδο 2SLS, οι ταυτόχρονα προσδιοριζόμενες εξισώσεις μπορούν να λυθούν ελαχιστοποιώντας το μεροληπτικό σφάλμα ταυτόχρονων εξισώσεων. Όπως και με το πρόβλημα της εξειδίκευσης, το πρόβλημα της ταυτοποίησης είναι και καλά τεκμηριωμένο και επιλύσιμο. Κανένα από τα προβλήματα αυτά δεν είναι τόσο σοβαρό όσο το πρόβλημα της εκτίμησης. Όμως, η λύση τους ενδέχεται να έχει κάποιο κόστος. Τα προβλήματα αυτά μαζί με εκείνα της απόκτησης αξιόπιστων δεδομένων δείγματος αποτελούν τα μεγαλύτερα εμπόδια στην ανάπτυξη και την εφαρμογή οικονομετρικών μοντέλων πολλών εξισώσεων.

### 2.3.8 Μια Εφαρμογή Οικονομετρικών Μοντέλων

Μια από τις μεγαλύτερες δυσκολίες στην κρίση της καταλληλότητας των ειδικών οικονομετρικών μοντέλων για την πρόβλεψη, είναι ότι αυτά αποτελούν ιδιοκτησία κάποιου και οι λεπτομέρειες του μοντέλου δεν είναι δημόσια διαθέσιμες. Ωστόσο, υπάρχουν κάποιες ελεύθερα διαθέσιμες εφαρμογές (γενικά υποστηριζόμενες από πόρους για έρευνα υπηρεσιών της κυβέρνησης των ΗΠΑ). Ένα πολύ καλό παράδειγμα αναπτύχθηκε από τη Διοίκηση Κοινωνικής Ασφάλειας των ΗΠΑ (Hambor, 1979).

#### Παράδειγμα 2.9

Αναπτύχθηκε ένα οικονομετρικό μοντέλο για να προσφέρει πλαίσιο ανάλυσης και πρόβλεψης της βραχυχρόνιας συμπεριφοράς του συστήματος OASDI (Old Age, Survivors, and Dependents Insurance –“Social Security Insurance”=Ηλικιωμένων, Επιζώντων, Εξαρτημένων από Ασφάλεια). Χρησιμοποιούνται ετήσια δεδομένα χρονοσειράς για την εξειδίκευση και την εκτίμηση των παραμέτρων ενός μοντέλου ταυτόχρονα προσδιοριζόμενων εξισώσεων για τις συνιστώσες των επιδομάτων και των εισφορών του συστήματος OASDI. ΟΙ προτεινόμενες διαρθρωτικές εξειδικεύσεις επιβεβαιώνονται από τα παρατηρηθέντα δεδομένα της χρονοσειράς. Ένα ιδιαίτερο αποτέλεσμα είναι ότι υπάρχουν ενδείξεις ότι «οι συνθήκες της αγοράς εργασίας και οι πληρωμές επιδομάτων σε σχέση με τα ημερομίσθια είναι και οι δύο σημαντικοί παράγοντες στον καθορισμό της αποχώρησης από την εργασία και του δικαιώματος αποζημίωσης λόγω ανικανότητας».

Η γενική μορφή του μοντέλου αποτελείται από τρεις εξισώσεις προς εκτίμηση με εξαρτημένες μεταβλητές: (α) τον αριθμό των αποζημιώσεων σε νέους δικαιούχους ασφαλιστικού επιδόματος, (β) το μέσο ασφαλιστικό επίδομα και (γ) τις συνολικές πληρωμές επιδομάτων από δεδομένο πιστωτικό ίδρυμα. Τα δεδομένα συγκεντρώθηκαν για μια χρονική περίοδο οκτώ χρόνων και υποβλήθηκαν σε έλεγχο μιας περιόδου ενός έτους μπροστά, προκειμένου να αξιολογηθεί η προβλεπτική ακρίβεια του μοντέλου. Αν και η πρόβλεψη μιας περιόδου δεν είναι ικανοποιητική για να αξιολογηθεί η προβλεπτική ακρίβεια του μοντέλου, ωστόσο παρέχει κάποια ένδειξη για το πώς θα πρέπει να γίνεται η επικαιροποίηση του μοντέλου στο μέλλον.

Ο Πίνακας 2.5 περιέχει δεδομένα για 19 από τις ενδογενείς μεταβλητές του μοντέλου OASDI. Αν και οι μεταβλητές δεν ταυτοποιούνται εδώ, αυτές εκφράζουν σημαντικές μετρήσεις της επίδρασης των οικονομικών μεταβλητών στον τομέα του OASDI.

ΟΙ στήλες 2 και 3 αναφέρονται στις πραγματικές και τις προβλεφθείσες τιμές για τις 19 ενδογενείς μεταβλητές. Η στήλη 4 αναφέρεται στο σφάλμα ή στη διαφορά ανάμεσα στις στήλες 2 και 3. Η στήλη 5 δίνει το σφάλμα πρόβλεψης,  $E_n$ , όταν οι προβλεφθείσες τιμές αποκτώνται από μια λύση που θέτει όλες τις ανεξάρτητες μεταβλητές ίσες με τις πραγματικές τιμές τους. Η λύση αυτή ονομάζεται «ουδέτερη» λύση.

Οι τιμές της «ουδέτερης» λύσης είναι ίσες με τις τιμές στη στήλη 3 εάν δεν υπάρχουν ενδογενείς ανεξάρτητες μεταβλητές που να καθορίζουν αυτή τη συγκεκριμένη μεταβλητή. Όταν η στήλη 3 και η ουδέτερη λύση είναι ίδιες, τότε και οι στήλες 4 και 5 είναι ίδιες.

Η στήλη 6 αναφέρεται στο εκτιμηθέν τυπικό σφάλμα παλινδρόμησης,  $\sigma_R$ , όπου είναι σχετικό. Τέλος, η στήλη 7 δείχνει το ποσοστιαίο σφάλμα,  $\%E$ , το σφάλμα στη στήλη 4, ως απόλυτη ποσοστιαία τιμή των πραγματικών τιμών της στήλης 2.

Μία μέθοδος για την αξιολόγηση της ακριβείας της πρόβλεψης είναι η σύγκριση του ποσοστιαίου σφάλματος με το τυπικό σφάλμα παλινδρόμησης. Δεκατρείς από τις δεκαεννιά τιμές του  $\%E$  είναι μικρότερες του δύο τοις εκατό και εννιά είναι μικρότερες του ένα τοις εκατό. Δεκατρείς απόλυτες τιμές είναι μικρότερες από το διπλάσιο του τυπικού σφάλματος της παλινδρόμησης ( $2\sigma_R$ ) και έξι τιμές είναι μικρότερες του  $\sigma_R$ . Δεκαέξι από τις δεκαεννιά τιμές στη στήλη 5,  $E_n$ , είναι μικρότερες του  $2\sigma_R$  και επτά είναι μικρότερες από το ένα τυπικό σφάλμα της παλινδρόμησης.

**ΠΙΝΑΚΑΣ 2.5.** Πραγματικά και προβλεφθέντα αποτελέσματα, σφάλματα, τυπικά σφάλματα και ποσοστιαίο σφάλμα για το οικονομικό μοντέλο

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Μεταβλητές	A	P	E=A-P	$E_n$	$\sigma_R$	%E
DB62 (εκατ.)	0.092	0.152	- 0.060	- 0.049	0.027	65.2
DB65 (εκατ.)	0.537	0.529	0.008	-0.008	0.030	1.5
TR (εκατ.)	0.932	0.986	-0.054	-0.054	0.030	5.8
CR (εκατ.)	0.149	0.149	0	0	0.003	0
ABAR(\$/mo)	205.9	203.7	2.20	2.20	1.57	1.1
VOAS (\$ δις)	58.51	58.88	-0.370	0.040	0.207	0.6
U65 (0/0)	5.3	4.9	0.4	0.4	0.26	7.5
U604(0/0 )	4.7	5.15	- 0.45	-0.45	0.40	9.6
ND (εκατ.)	0.592	0.587	0.005	0	0.014	0.8
ND (εκατ.)	0.592	0.566	0.026	0.023	0.012	4.4
ABAD(\$/mo)	233.6	234.6	-1.00	-1.00	1.36	0.4
VD (\$ δις)	8.414	8.330	0.084	0.084	0.049	1.0
VD (\$ δις)	8.414	8.246	0.168	0.073	0.049	2.0
U25 (0/0)	6.1	5.8	0.3	0.3	0.060	4.9
WS (\$ δις)	801.6	800.9	0.7	0.6	0.460	0.1
TESL (\$ δις)	31.61	31.87	-0.260	-0.210	0.097	0.8
TSL (\$ δις)	4.294	4.235	0.059	0.059	0.012	1.2
TLSE(\$ δις)	3.405	3.414	- 0.009	- 0.009	0.018	0.3
TSE (\$ δις)	3.335	3.335	0	0.001	-0.021	0

ΠΗΓΗ: Hambor (1979, σ. 16).

Αν και η παραπάνω ανάλυση δεν οδηγεί σε οριστικά συμπεράσματα, Φαίνεται ότι για τις περισσότερες μεταβλητές οι προβλέψεις είναι σχετικά ακριβείς. Διατυπωμένο διαφορετικά, αυτό σημαίνει ότι, εάν θέλουμε να δεχτούμε προβλέψεις που βρίσκονται εντός δυο τυπικών σφαλμάτων της πραγματικής τιμής, το οικονομετρικό μοντέλο θα προβλέπει τις περισσότερες Φορές με ακρίβεια. Η αποδοχή προβλέψεων στο διάστημα δυο τυπικών σφαλμάτων δείχνει ότι είμαστε πρόθυμοι ν' αποδεχτούμε με πιθανότητα 5%, ότι η πρόβλεψή μας μπορεί να είναι λάθος.

Πέρα από την ικανότητά τους για πρόβλεψη, τα οικονομετρικά μοντέλα μπορούν να χρησιμοποιηθούν για τη διενέργεια εκτενούς ανάλυσης, σχετικής με θέματα του τομέα λήψης επιχειρηματικών αποφάσεων.

Για παράδειγμα, στην περίπτωση του μοντέλου OASDI, μπορεί να υπάρχουν εξελίξεις στους περιβαλλοντικούς οικονομικούς παράγοντες ή μεταβολές στους νόμους της κοινωνικής ασφάλισης



που θα επηρεάσουν το καθεστώς των επιδομάτων ή το μέγεθός τους που χορηγείται σε δικαιούχους κατά το παρόν και το μέλλον. Τέτοιες μεταβολές μπορούν να ενσωματωθούν στο μοντέλο και να ελεγχθούν σε όρους ευαισθησίας του μοντέλου ως προς αυτές τις μεταβολές. Αυτή η τελευταία χρήση αξιάνει αισθητά τα οφέλη που σχετίζονται με την κατασκευή οικονομετρικών μοντέλων.

### 2.3.9 Προσαρμογή Οικονομετρικού Μοντέλου στην Πρόβλεψη

Στην προηγούμενη παράγραφο, περιγράψαμε ένα μικρο-οικονομετρικό μοντέλο. Το μοντέλο του τομέα OASDI επιτρέπει την πρόβλεψη ενός μεγάλου αριθμού ενδογενών μεταβλητών, συμπεριλαμβανομένου του καθεστώτος των ασφαλιστικών επιδομάτων και της μέσης τιμής επιδόματος ανά δικαιούχο.

Εν τούτοις, η ανάπτυξη ειδικών οικονομετρικών μοντέλων για έναν τομέα ή για μια εταιρία, έχει αποδειχθεί πολύ σύνθετη και πολύ δαπανηρή για όλες, εκτός από τις μεγάλες επιχειρήσεις και οργανισμούς.

Στο παράδειγμα 6.1 προτάθηκε μια μέθοδος ενσωμάτωσης ενός οικονομετρικού μοντέλου στην πρόβλεψη μιας επιχείρησης. Η πρόταση αυτή βασίστηκε στην αντίληψη ότι οι διευθυντές επιχειρήσεων είναι καλύτερα καταρτισμένοι σε σχέση με το παρελθόν. Ειδικότερα, αυτό μπορεί να αποδοθεί στην αύξηση του αριθμού του προσωπικού που έχει εκπαιδευτεί στη στατιστική και στην αυξανόμενη δύναμη και εξάπλωση των υπολογιστών. Σήμερα, οι περισσότερες οικονομετρικές εργασίες για την κυβέρνηση και τα μεγάλα πανεπιστήμια διενεργούνται από οικονομέτρες και στατιστικούς. Επίσης, παρουσιάστηκε ραγδαία ανάπτυξη του αριθμού των επιχειρήσεων που παρέχουν συμβουλές σε οικονομετρικά προβλήματα. Αυτοί οι οικονομετρικοί σύμβουλοι δίνουν τη δυνατότητα σε μια επιχείρηση να συνδέσει τη λήψη αποφάσεων και την πρόβλεψη με τα μοντέλα ταυτόχρονα προσδιοριζόμενων εξισώσεων που έχουν κατασκευαστεί από τον σύμβουλο. Επομένως, η οικονομετρική πρόβλεψη μπορεί να είναι μια διαδικασία με την οποία μια επιχείρηση συνδέει τη δική της πρόβλεψη, που βασίζεται στην παλινδρόμηση, με τις ερμηνευτικές μεταβλητές που εκτιμήθηκαν από το οικονομετρικό μοντέλο του συμβούλου.

Με αυτό τον τρόπο, το οικονομετρικό μοντέλο εκφράζει τη σχέση ανάμεσα στο τι μια επιχείρηση επιθυμεί να προβλέψει και σε μια ποικιλία οικονομικών δεικτών που εκτιμήθηκαν από ένα σύμβουλο. Αυτοί οι δείκτες μπορούν να περιλαμβάνουν τέτοιες μεταβλητές, όπως διαθέσιμο προσωπικό εισόδημα, δαπάνες για επενδύσεις, επιτόκια και προσφορά χρήματος, μεταξύ άλλων. Η σύνδεση εκφράζεται συχνά από ένα δεύτερο μοντέλο στο συγκεκριμένο επίπεδο της επιχείρησης, το οποίο μπορεί να δώσει εκτιμήσεις που έχουν εξομαλυνθεί και αφορούν μηνιαίες πωλήσεις ανά προϊόν ή τμήμα και να προβλέψει, αν είναι επιθυμητό, εποχικές συνιστώσες. Τέλος, θα μπορούσε να αναπτυχθεί ένα τρίτο επίπεδο πρόβλεψης βασισμένο σε ένα μοντέλο που εκφράζει σχέση μεταξύ τέτοιων μεταβλητών, όπως πωλήσεις της εταιρίας, διαφήμιση και πολιτικής προαγωγής του προσωπικού. Έτσι, η επίδραση της διαφήμισης και της πολιτικής προώθησης προσωπικού θα μπορούσε να αξιολογηθεί με επιστημονικό τρόπο.

**3.1 ΑΥΤΟΣΥΣΧΕΤΙΣΗ, ΑΥΤΟΠΑΛΙΝΔΡΟΜΙΚΑ ΜΟΝΤΕΛΑ ΚΑΙ ΑΝΑΛΥΣΗ ΧΡΟΝΟΣΕΙΡΩΝ [15], [17]**

**3.1.1 Εισαγωγή**

Στο κεφάλαιο αυτό θα επικεντρώσουμε την προσοχή μας στην ταυτοποίηση και επιλογή ενός μοντέλου πρόβλεψης βασισμένου στην ανάλυση δεδομένων χρονοσειράς. Το μοντέλο που τελικά θα επιλεγεί για να ικανοποιήσει τους σκοπούς της πρόβλεψης θα είναι εκείνο που προσαρμόζεται καλύτερα στη χρονοσειρά. Επομένως, ο σκοπός αυτού του κεφαλαίου είναι να γίνει μια εισαγωγή σε χρήσιμες έννοιες της ανάλυσης χρονοσειρών και στην ταυτοποίηση μοντέλων που εξυπηρετούν τους σκοπούς της πρόβλεψης. Καθώς και να παρουσιαστεί μια γενική τάξη μοντέλων για την πρόβλεψη χρονοσειρών.

Στην παράγραφο 1.1, τα μοντέλα που χρησιμοποιήθηκαν για την πρόβλεψη χρονοσειρών ήταν οι μέθοδοι του κινητού μέσου και της εκθετικής εξομάλυνσης. Δείξαμε, υπό κάποια έννοια, ότι ο κινητός μέσος αποτελεί ειδική μορφή εκθετικής εξομάλυνσης. Οι μέθοδοι της εκθετικής εξομάλυνσης μπορούν να θεωρηθούν σαν μια διαδικασία εύρεσης του μέσου όρου, εφόσον οι νέες προβλέψεις βασίζονται στο σταθμισμένο συνδυασμό των πραγματικών τιμών, των παρελθόντων σφαλμάτων και (για μερικά μοντέλα) των τάσεων του παρελθόντος. Για να κατανοήσουμε αυτό το σημείο ας θυμηθούμε ότι τα μοντέλα της απλής εκθετικής εξομάλυνσης έχουν την ακόλουθη μορφή:

$$Y_{t+1} = Y_t + a(X_t - Y_t) \tag{3.1}$$

όπου  $Y$  είναι η πρόβλεψη και  $X$  η πραγματική τιμή. Εάν αντικαταστήσουμε την πρόβλεψη της προηγούμενης περιόδου:

$$Y_t = Y_{t-1} + a(X_{t-1} - Y_{t-1}) \tag{3.2}$$

στην (7.1),

$$Y_{t+1} = Y_{t-1} + a(X_{t-1} - Y_{t-1}) + a(X_t - Y_t) \tag{3.3}$$

Η εξίσωση θα μπορούσε να επεκταθεί παραπέρα με αντικατάσταση της έκφρασης για  $Y_{t-1}$ ,  $Y_{t-2}$ ,... Η λογική της δημιουργίας τέτοιων μαθηματικών εκφράσεων, όπως η (3.1), είναι φανερή. Αφού δημιουργηθεί η αρχική πρόβλεψη, η νέα πρόβλεψη γίνεται ο σταθμισμένος συνδυασμός των σφαλμάτων του παρελθόντος και της αρχικής πρόβλεψης. Η προκύπτουσα πρόβλεψη,  $Y_{t+1}$ , βρίσκεται πλησίον της πραγματικής τάσης στα δεδομένα γιατί είναι ένας μέσος όρος των αρνητικών και των

θετικών σφαλμάτων.

Στην παράγραφο 1.2, οι μέθοδοι ανάλυσης χρονοσειρών σε συστηματικές και μη-συστηματικές (τυχαίες) συνιστώσες εξετάστηκαν με βάση την αρχή, ότι κάποιος θα μπορούσε να προβλέψει τα επαναλαμβανόμενα συστηματικά μοντέλα μιας χρονοσειράς. Αυτές οι συνιστώσες που ονομάζονται εποχικές, κυκλικές, τάση και μερικές φορές μοντέλα ημερών συναλλαγής, θα μπορούσαν να προβλεφθούν, και ο προβλέπων, με τη σειρά του, θα μπορούσε να ανασυνθέσει τις συνιστώσες για να φτάσει στην τελική πρόβλεψη. Όπως οι μέθοδοι της εξομάλυνσης, έτσι και οι κλασικές μέθοδοι διάσπασης προϋποθέτουν ότι οι προβλέψεις είναι συναρτήσεις μόνο του χρόνου.

Στην αρχή της παραγράφου 2.1, παρουσιάσαμε τις μεθόδους ανάπτυξης ερμηνευτικών ή αιτιοκρατικών μεθόδων πρόβλεψης. Παρουσιάσαμε επίσης τη μέθοδο συσχέτισης της διακύμανσης μιας εξαρτημένης μεταβλητής σε σχέση με μια ερμηνευτική ή ανεξάρτητη μεταβλητή. Συνεχίζοντας στην παράγραφο 2.2, ασχοληθήκαμε με τις μεθόδους συσχέτισης μιας εξαρτημένης μεταβλητής με περισσότερες από μία ανεξάρτητες μεταβλητές. Αυτές οι μέθοδοι παλινδρόμησης επιτρέπουν στον προβλέποντα να συσχετίσει την εξαρτημένη μεταβλητή που τον ενδιαφέρει με τους παράγοντες που μπορεί να σχετίζονται με τη διακύμανση στην εξαρτημένη μεταβλητή. Συμβολικά, η έκφραση γράφεται ως εξής:

$$Y=b_0+b_1X_1+b_2X_2+\dots+b_kX_k \quad (3.4)$$

Στην παράγραφο 2.3, ο σκοπός μας ήταν να παρουσιάσουμε την πλέον προχωρημένη προσέγγιση οικονομετρικών μοντέλων. Αυτά είναι μοντέλα πολλαπλών εξισώσεων που χρησιμοποιούν μεθόδους παλινδρόμησης για τη λύση προβλημάτων πρόβλεψης και εξηγούν τη διακύμανση σε μερικές ενδογενείς μεταβλητές που μας ενδιαφέρουν.

Ο σκοπός αυτής της παραγράφου επικεντρώνεται στις μεθόδους και τις αρχές παλινδρόμησης, όπως εφαρμόζονται απευθείας στη μελέτη των χρονοσειρών. Δηλαδή, μελετάμε ένα συνδυασμό των μεθοδολογιών και των εννοιών τους, δίνοντας έμφαση στη δημιουργία μοντέλων με την προσέγγιση ARIMA (autoregressive integrated moving average model).

Η συντομογραφία ARIMA εξηγείται ως: αυτοπαλινδρόμο μοντέλο ολοκληρωμένου κινητού μέσου. Το I (integrated) αναφέρεται στο διαφορισμό των δεδομένων που η εξήγησή του δίνεται σε αυτό το κεφάλαιο. Το AR (autoregression= αυτοπαλινδρόμηση), η διαφορίση (I) και ο κινητός μέσος (MA= moving average) αποτελούν τα γράμματα που συνθέτουν τον όρο ARIMA. Τέλος, τα μοντέλα MA δεν έχουν την ίδια έννοια με τον όρο «κινητός μέσος» που εξετάσαμε στο κεφάλαιο 2.

Η εξίσωση (7.4) παραπάνω, είναι μια παλινδρόμηση που συσχετίζει μια εξαρτημένη μεταβλητή όπως οι πωλήσεις μιας εταιρίας, φορτία εμπορευμάτων ή αποθέματα, με ένα σύνολο ερμηνευτικών (ανεξάρτητων) μεταβλητών. Εάν ορίσουμε

$$X_1=Y_{t-1}, X_2=Y_{t-2}, \dots, X_k=Y_{t-k}$$

τότε η εξίσωση (7.4) γίνεται

$$Y_t = b_0 + b_1 Y_{t-1} + b_2 Y_{t-2} + \dots + b_m Y_{t-m} + e_t \quad (7.5)$$

Μολονότι η εξίσωση (7.5) εξακολουθεί να είναι παλινδρόμηση, οι μεταβλητές της δεξιάς πλευράς (RHS= right-hand side) είναι απλώς οι τιμές της εξαρτημένης μεταβλητής σε προηγούμενες περιόδους. Οι μεταβλητές αυτές παριστάνουν χρονικές υστερήσεις της εξαρτημένης μεταβλητής και η παλινδρόμηση αυτής της μορφής ονομάζεται *αυτοπαλινδρόμηση* (AR = autoregression). Εάν επανεξετάσουμε τον ορισμό τού τύπου της εκθετικής εξομάλυνσης

$$Y_{t+1} = aX_t + a(1-a)X_{t-1} + a(1-a)^2 X_{t-2} + \dots \quad (7.6)$$

παρατηρούμε ότι αυτός είναι παρόμοιος με τον τύπο της αυτοπαλινδρόμησης. Οι παρελθούσες τιμές των προβλέψεων σταθμίζονται από εκθετικούς σταθμικούς μέσους:  $a, a(1-a), a(1-a)^2$ , κλπ. Αυτοί οι σταθμικοί όροι παίρνουν τη θέση των συντελεστών παλινδρόμησης  $b_1, b_2, \dots$  ενός αυτοπαλινδρομού σχήματος.

Τα μοντέλα αυτοπαλινδρόμησης διαφέρουν από άλλα μοντέλα παλινδρόμησης ως προς την υπόθεση της ανεξαρτησίας του όρου του σφάλματος. Εφόσον οι ανεξάρτητες μεταβλητές εκφράζουν τιμές χρονικών υστερήσεων της εξαρτημένης μεταβλητής, η υπόθεση της ανεξαρτησίας των όρων του σφάλματος παραβιάζεται εύκολα. Αν και η υπόθεση των ανεξάρτητων όρων του σφάλματος παραβιάζεται τόσο εύκολα, το πρόβλημα του καθορισμού του κατάλληλου αριθμού των παρελθόντων όρων του  $Y$  που θα ενσωματωθούν στην (3.5), δεν είναι εύκολο να λυθεί.

### 3.1.2 Χαρακτηριστικά Χρονοσειρών

Στην παράγραφο 1.1, εξετάσαμε έναν αριθμό μοντέλων που μπορούν να χρησιμοποιηθούν για την πρόβλεψη μελλοντικών τιμών μιας χρονοσειράς με βάση τη συμπεριφορά της κατά το παρελθόν. Τα μοντέλα αυτά ήταν προσδιοριστικά ως προς το ότι δεν γινόταν καμία αναφορά στις πηγές ή στη φύση του θεμελιώδους ζητήματος της τυχαιότητας της χρονοσειράς. Στην ουσία, τα μοντέλα αυτά αναφέρονται σε τεχνικές προεκτάσεων που χρησιμοποιούνταν κατά κανόνα στην επιχειρηματική και οικονομική πρόβλεψη. Όπως θα δούμε, οι τεχνικές αυτές δεν έχουν τόση προβλεπτική ακρίβεια όση έχουν τα μοντέρνα μοντέλα στοχαστικών χρονοσειρών. Επομένως, ένα σημαντικό ερώτημα που τίθεται είναι «αν η χρονοσειρά είναι τυχαία».

Ένα απλό τυχαίο μοντέλο είναι εκείνο όπου η παρατήρηση  $Y_t$  αποτελείται από δύο μέρη, που είναι: ένας συνολικός αριθμητικός μέσος  $\mu$  και μια συνιστώσα τυχαίου σφάλματος,  $e_t$  η οποία είναι ανεξάρτητη από περίοδο σε περίοδο (δεν υπάρχει αυτοσυσχέτιση).

$$Y_t = \mu + e_t$$

$$[\text{ARIMA}(0,0,0)] \quad (3.7)$$

Το μοντέλο αυτό ονομάζεται ARIMA (0,0,0). Δεν υπάρχει αναλογία AR (το  $Y_t$  δεν εξαρτάται από το  $Y_{t-1}$ ), δεν υπάρχει διαφοράση ούτε αναλογία. MA (το  $Y_t$  δεν εξαρτάται από το  $e_{t-1}$ ). Είναι μια απολύτως τυχαία σταθερή (stationary) χρονοσειρά.

Σε μια εντελώς τυχαία χρονοσειρά, ο συντελεστής αυτοσυσχέτισης θα είναι μηδέν. Κάθε τρέχουσα τιμή δεν σχετίζεται με την προηγούμενή της. Μια τυχαία χρονοσειρά που κατανέμεται κανονικά ονομάζεται σειρά *white noise* και λέγεται ότι είναι καθαρά τυχαία (όλα τα  $\rho_k=0$ ). Από την άλλη, όταν ο συντελεστής αυτοσυσχέτισης είναι πολύ μεγάλος, υπάρχει ένας υψηλός βαθμός συσχέτισης ανάμεσα στις διαδοχικές τιμές. Με τη μέτρηση και εξέταση της αυτοσυσχέτισης για χρονικές υστερήσεις μεγαλύτερες της μιας περιόδου, θα μπορούμε να έχουμε ενδείξεις σε ποιο βαθμό και με ποιο τρόπο σχετίζονται οι τιμές δεδομένης χρονοσειράς.

Ένα μοντέλο που είναι παρόμοιο με τη διαδικασία AR, επειδή  $Y_t$  είναι μια συνάρτηση του  $Y_{t-1}$ , αλλά ο συντελεστής του  $Y_{t-1}$  είναι ίσος με ένα, είναι το εξής:

$$Y_t = \mu + Y_{t-1} + e_t \quad (3.8)$$

$$[\text{ARIMA}(0,1,0)]$$

Η εξίσωση μπορεί για ευκολία να γραφεί ως εξής

$$Y_t - Y_{t-1} = \mu + e_t \quad (3.9)$$

που είναι η χρονοσειρά των πρώτων διαφορών.

Η σειρά  $(Y_t - Y_{t-1})$  είναι σταθερή, ενώ η χρονοσειρά για  $Y_t$  που εκφράζεται με την εξίσωση (7.7) είναι μη-σταθερή. Μια σειρά λέγεται στάσιμη εάν δεν υπάρχει μεταβολή στον μέσο διαχρονικά. Η σειρά ARIMA (0,1,0) έχει μέσο που μεταβάλλεται ως προς τον χρόνο. Η σειρά (0,0,0) είναι σειρά που είναι σταθερή στον μέσο. Θα δείξουμε ότι μια χρονοσειρά μπορεί να παρέχει ενδείξεις για την ύπαρξη ή όχι μεταβολής στη διακύμανση. Στις παραγράφους 2.1 και 2.2, χαρακτηρίσαμε την ιδιότητα της σταθερής διακύμανσης μιας χρονοσειράς ως *ομοσκεδαστικότητα* και την ιδιότητα της μεταβαλλόμενης διακύμανσης ως *επεροσκεδαστικότητα*. Όπως θα δούμε, η γνώση αυτών των ιδιοτήτων θα μας επιτρέψει να διαμορφώσουμε καλύτερα μοντέλα για τα δεδομένα χρονοσειράς και σαν συνέπεια θα έχουμε καλύτερες προβλέψεις.

Καθώς αναπτύσσουμε καλύτερα μοντέλα χρονοσειρών, επιθυμούμε να γνωρίζουμε εάν η εξεταζόμενη στοχαστική διαδικασία είναι *αμετάβλητη* ως προς τον χρόνο ή όχι. Εάν τα χαρακτηριστικά της στοχαστικής διαδικασίας αλλάζουν με τον χρόνο, τότε θα είναι δύσκολο να αναπαραστήσουμε μια χρονοσειρά ως προς τα διαστήματα του παρελθόντος και του μέλλοντος μέσω ενός απλού αλγεβρικού μοντέλου. Εναλλακτικά, αν η στοχαστική διαδικασία είναι σταθερή ή

καθορισμένη, μπορούμε εύκολα να διαμορφώσουμε τη χρονοσειρά μέσω μιας εξίσωσης με καθορισμένους συντελεστές που προέρχονται από εκτιμήσεις ιστορικών παρατηρήσεων. Ένα καλό παράδειγμα είναι το μοντέλο μιας παλινδρομικής εξίσωσης, στο οποίο μια οικονομική μεταβλητή σχετίζεται με άλλες οικονομικές μεταβλητές με συντελεστές που έχουν εκτιμηθεί κάτω από την υπόθεση μιας σταθερής σχέσης. Επομένως, ένα άλλο ερώτημα που προκύπτει για τον ορισμό μιας χρονοσειράς είναι αν η χρονοσειρά είναι σταθερή ή όχι.

Μεταξύ των διαφόρων τρόπων που μια χρονοσειρά μπορεί να συναντάται ως μη-σταθερή σε επιχειρηματικές και οικονομικές καταστάσεις, είναι και όταν ο μέσος όρος ή το επίπεδο της χρονοσειράς δεν είναι μια σταθερά. Η κατάσταση αυτή είναι γνωστή ως *τάση στον μέσο όρο*. Όπως σημειώσαμε προηγουμένως, εάν μπορεί να ευρεθεί μια κατάλληλη τάση ή εάν η τάση στον μέσο όρο μπορεί να απαλειφθεί αποτελεσματικά με τη διαφορίση της χρονοσειράς, τότε αυτή η συγκεκριμένη μορφή μη-σταθερότητας μπορεί να εξαλειφθεί. Είναι επίσης συνηθισμένη η εύρεση χρονοσειρών των οποίων η μεταβλητότητα αλλάζει διαχρονικά. Αυτό, ονομάζεται *τάση στη διακύμανση*. Μια τέτοια χρονοσειρά μπορεί συχνά να γίνει σταθερή, ή σχεδόν σταθερή, με τη λήψη των λογαρίθμων της χρονοσειράς. Ένας τέτοιος μετασχηματισμός συνήθως μεταφέρει την τάση στη διακύμανση σε τάση στον μέσο. Μπορούν να παρουσιαστούν ακόμη πιο σύνθετοι τύποι ύπαρξης μη-σταθερότητας, που απαιτούν τεχνικές μεγάλης συνθετότητας και μαθηματικής ωστηρότητας στο χειρισμό τους, γι' αυτό δεν θα τις μελετήσουμε εδώ.

Εάν η διαρθρωτική σχέση ενός μοντέλου μεταβάλλεται διαχρονικά, μια μη-σταθερή χρονοσειρά, δεν θα μπορούσαμε να χρησιμοποιήσουμε τις μεθόδους των παραγράφων 2.2 και 2.3 για την πρόβλεψη μελλοντικών παρατηρήσεων. Ένα ερώτημα που δημιουργείται ωστόσο, είναι σε ποιο σημείο μια χρονοσειρά γίνεται σταθερή. Η μελέτη των αυτοσυσχετίσεων συνήθως αποδεικνύει ότι πολύ λίγες χρονοσειρές είναι σταθερές. Όμως, όπως αναφέρθηκε προηγουμένως, πολλές από τις μη-σταθερές χρονοσειρές που συναντάμε έχουν την επιθυμητή ιδιότητα να μετασχηματίζονται σε σταθερές εάν διαφοριστούν μία ή περισσότερες φορές. Αργότερα, θα παρουσιάσουμε και τα σχετικά παραδείγματα. Τέτοιες μη-σταθερές χρονοσειρές ονομάζονται ομογενείς. Ο αριθμός των διαφορίσεων της αρχικής χρονοσειράς για να γίνει σταθερή, αναφέρεται ως *τάξη ομογένειας*.

Η βασική μορφή ενός σταθερού μοντέλου AR (1) = ARIMA (1,0,0) εκφράζεται ως:

$$(Y_t - \mu) = \phi_1 (Y_{t-1} - \mu) + e_t \quad (3.10)$$

όπου  $\mu$  είναι ο μέσος της σειράς  $Y$ . Για απλοποίηση γράφουμε:

$$Y_t = (\mu - \phi_1 \mu) + \phi_1 Y_{t-1} + e_t \quad (3.11)$$

και με αντικατάσταση για  $\mu' = \mu - \phi_1 \mu$  έχουμε:

$$Y_t = \mu' + \phi_1 Y_{t-1} + e_t \quad (3.12)$$

[ARIMA (1,0,0)]

Συνήθως, τα μοντέλα AR ενέχουν έναν όρο της μορφής ( $\mu = \phi_1 \mu - \phi_2 \mu - \dots$ ) ο οποίος γενικά παρουσιάζεται ως  $\mu'$ . Επίσης, σημειώστε ότι για το μοντέλο (1,0,0) η παρατήρηση  $Y_t$  σχετίζεται με την  $Y_{t-1}$ , δηλαδή την παρατήρηση της προηγούμενης περιόδου. Η παράμετρος παλινδρόμησης  $\phi_1$  υπόκειται στον περιορισμό να βρίσκεται μεταξύ -1 και +1. Ο λόγος για αυτό τον περιορισμό θα αναφερθεί αργότερα.

Εάν το μοντέλο AR(I) είναι μη-σταθερό, ίσως επιθυμούμε να μετατρέψουμε την εξίσωση του μοντέλου σε σταθερή παίρνοντας τις πρώτες διαφορές. Σε αυτή την περίπτωση, το μοντέλο γίνεται

$$Y_t - Y_{t-1} = \phi_1 (Y_{t-1} - Y_{t-2}) + \mu' - \mu' + e_t \quad (3.13)$$

και

$$\nabla Y_t = \phi_1 (\nabla Y_{t-1}) + e_t \quad (3.14)$$

[ARIMA (1,1,0)]

όπου το  $\nabla$  δείχνει τη διαφορά μεταξύ του  $Y$  στην περίοδο  $t$  και στην προηγούμενη περίοδο ( $t-1$ ). Θα επικεντρώσουμε το ενδιαφέρον μας στην εξέταση μοντέλων όλων των παραπάνω τύπων και σε ακόμη περισσότερα.

Στην παράγραφο 1.2, εξετάσαμε την εποχιακότητα μιας χρονοσειράς σαν την επαναλαμβανόμενη συμπεριφορά που συμβαίνει σε κανονική ημερολογιακή βάση, συνήθως ενός χρόνου. Για παράδειγμα, εάν μια μηνιαία χρονοσειρά  $Y_t$  παρουσιάζει ετήσια εποχιακότητα, τότε τα σημεία των δεδομένων στη χρονοσειρά θα πρέπει να δείχνουν κάποια συσχέτιση με τα αντίστοιχα σημεία των δεδομένων που προηγούνται ή υστερούν κατά 12 μήνες. Με άλλα λόγια, η αυτοσυσχέτιση του  $Y_t$ ,  $Y_{t+12}$  και  $Y_{t-24}$ , για παράδειγμα, θα πρέπει να είναι υψηλή. Αυτές οι συσχετίσεις θα δείχνουν τα μέγιστα σημεία  $k=12, 24, 36, 48, 60$  κ.ο.κ. Αξίζει να σημειώσουμε, ότι η εποχιακότητα είναι μια άλλη μορφή απόκλισης από τη σταθερότητα. Ωστόσο, για να αποφύγουμε τη σύγχυση, θα χρησιμοποιήσουμε τον όρο εποχιακότητα όταν η χρονοσειρά είναι μη-σταθερή εξαιτίας της εποχικής διακύμανσης και τον όρο μη-σταθερή όταν η χρονοσειρά δεν είναι σταθερή για λόγους διάφορους της εποχιακότητας. Τέλος, μπορούμε να παρατηρήσουμε και να εξακριβώσουμε την εποχιακότητα σημειώνοντας τα κανονικά μέγιστα και κατώτατα σημεία της αυτοσυσχέτισης μέσα στον χρόνο. Επομένως, επιθυμούμε να εξακριβώσουμε για κάθε χρονοσειρά εάν υπάρχει ή όχι εποχιακότητα. Εάν πράγματι υπάρχει, «ποια είναι η έκταση της εποχιακότητας;»

Συνοψίζοντας, για κάθε χρονοσειρά, θέλουμε να προσδιορίσουμε εάν:

1. τα δεδομένα είναι τυχαία
2. τα δεδομένα είναι σταθερά ή μη-σταθερά, σε ποιο επίπεδο είναι μεταβλητά στον χρόνο
3. τα δεδομένα είναι εποχικά, κι αν όντως είναι, ποια είναι η έκταση της εποχιακότητας.

Όλα αυτά τα χαρακτηριστικά προσδιορίζονται με τη μελέτη της *συνάρτησης αυτοσυσχέτισης* μιας χρονοσειράς.

### 3.1.3 Σημείωση για τα Μοντέλα ARIMA

Οι περισσότερες χρονοσειρές που συναντώνται σε μερικές εφαρμογές της πρόβλεψης δεν είναι σταθερές. Εάν μια μη-σταθερή χρονοσειρά μπορεί να μετατραπεί σε σταθερή με  $d$  διαφορές (συνήθως,  $d=1$  ή  $2$ ), το αποτέλεσμα είναι ένα μοντέλο για τη διαφορισμένη χρονοσειρά. Η αρχική χρονοσειρά αναφέρεται σαν μοντέλο ARIMA. Το  $I$  αφορά την ολοκληρωμένη (integrated) και χρησιμοποιείται για να δείξει ότι δεν έχει «διαφοριστεί».

Για ένα μοντέλο ARIMA, η τάξη δίνεται από τα τρία γράμματα  $p$ ,  $d$  και  $q$ . Η τάξη του συντελεστή αυτοπαλινδρόμησης είναι  $p$ , η τάξη της διαφορίσης που χρειάζεται για την επίτευξη σταθερότητας είναι  $d$  και η τάξη του μέρους του κινητού μέσου είναι  $q$ .

Το μοντέλο ARIMA ( $p, d, q$ ) είναι το πλέον γενικό μοντέλο από όσα εξετάζουμε εδώ, καθώς και το περισσότερο χρησιμοποιημένο. Συχνά, είναι απαραίτητο να γίνονται διαφορίσεις για την επίτευξη σταθερότητας, αλλά οι σειρές που προκύπτουν μπορεί να απαιτούν μόνο αυτοπαλινδρόμηση,  $p$ , ή έναν κινητό μέσο,  $q$ . Τα μοντέλα αυτά ονομάζονται αυτοπαλινδρομικά ολοκληρωμένα (autoregressive integrated = ARI) ή μοντέλα ολοκληρωμένου κινητού μέσου (integrated moving average= IMA). Ο όρος «ολοκληρωμένος» χρησιμοποιείται όταν ο διαφορισμός γίνεται για την απόκτηση σταθερότητας, εφόσον η σταθερή χρονοσειρά πρέπει να ολοκληρωθεί (μη-διαφορισμένη) για να αναπαραγάγει τα αρχικά δεδομένα.

Για τους σκοπούς της πρόβλεψης πρέπει πρώτα να ταυτοποιήσουμε το κατάλληλο μοντέλο ARIMA και να μετασχηματίσουμε ή και να διαφορίσουμε τα δεδομένα προς απόκτηση σταθερότητας. Έτσι, το μοντέλο πλέον ανάγεται σε μοντέλο της κατηγορίας ARMA.

### 3.1.4 Η Συνάρτηση Αυτοσυσχέτισης

Αν και είναι γενικά αδύνατο να έχουμε μια πλήρη περιγραφή μιας στοχαστικής διαδικασίας, η συνάρτηση αυτοσυσχέτισης είναι εξαιρετικά χρήσιμη για την απόκτηση μιας μερικής περιγραφής της διαδικασίας ανάπτυξης ενός μοντέλου πρόβλεψης. Ακόμη, η συνάρτηση αυτοσυσχέτισης μετράει το βαθμό συσχέτισης ανάμεσα στις γειτονικές παρατηρήσεις της χρονοσειράς. Αρχίζουμε ορίζοντας την αυτοσυσχέτιση με υστέρηση  $k$  ως εξής:

$$\rho_k = \frac{E(Y_t - \mu_y)(Y_{t+k} - \mu_y)}{\left[ E(Y_t - \mu_y)^2 E(Y_{t+k} - \mu_y)^2 \right]^{1/2}} \quad (3.15)$$

$$\rho_k = \frac{COV(Y_t, Y_{t+k})}{\sigma_{Y_t} \sigma_{Y_{t+k}}} \quad (3.16)$$

Για μια σταθερή διαδικασία, η διακύμανση στον χρόνο  $t$  στον παρονομαστή της εξίσωσης (3.16) είναι



ίδια με τη διακύμανση στον χρόνο  $t+k$ . Επομένως, ο παρανομαστής είναι ίδιος με τη διακύμανση της στοχαστικής διαδικασίας, και

$$\rho_k = \frac{COV(Y_t, Y_{t+k})}{\sigma_{Y_t} \sigma_{Y_{t+k}}} \quad (3.17)$$

Ο αριθμητής και ο παρονομαστής είναι επομένως ίσοι, και

$$\rho_0 = 1$$

Η συνάρτηση αυτοσυσχέτισης εκτιμάται από παρατηρήσεις δείγματος ως

$$r_k = \frac{\sum_1^{n-k} (Y_t - \bar{Y})(Y_{t-k} - \bar{Y}_{t-k})}{\sum_1^n (Y_t - \bar{Y})^2} \quad (3.18)$$

Το πώς υπολογίζεται ο συντελεστής αυτοσυσχέτισης δείχνεται στο παράδειγμα που ακολουθεί.

### Παράδειγμα 3.1

Δεδομένα για τα κέρδη ανά μετοχή έχουν συλλεγεί για την Gates Learjet Corporation για τα έτη 1970 έως 1980. Τα δεδομένα του παραδείγματος 7.1 που παρουσιάζονται στον Πίνακα 7.1 εκφράζονται από το ακόλουθο μοντέλο αυτοπαλινδρόμησης (AR) [ένα ARIMA (2,0,0)]:

$$Y_t = \mu + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + e_t \quad (3.19)$$

Η σειρά AR της εξίσωσης (7.19) εκφράζει το  $Y_t$  σαν γραμμικό συνδυασμό των δύο προηγούμενων τιμών του. Οι μεταβλητές της χρονοσειράς  $Y$  στους χρόνους  $t-1$  και  $t-2$  δημιουργήθηκαν εύκολα με την προς τα εμπρός μετακίνηση των τιμών κατά μία και δύο περιόδους αντίστοιχα.

**ΠΙΝΑΚΑΣ 3.1:** Χρονοσειρές κερδών ανά μετοχή ( $Y_t$ ) της Gates Learjet Corporation

(1)	(2)	(3)	(4)
		<i>Υστέρηση</i>	<i>Υστέρηση</i>
	<i>Αρχική</i>	<i>μας περιόδου</i>	<i>δύο περιόδων</i>
	$Y_t$	$Y_{t-1}$	$Y_{t-2}$
1970	0.93	0.78	0.83
1971	0.78	0.83	1.84
1972	0.83	1.84	1.22
1973	1.84	1.22	2.21
1974	1.22	2.21	1.64
1975	2.21	1.64	1.72
1976	1.64	1.72	2.22
1977	1.72	2.22	2.32
1978	2.22	2.32	3.05
1979	2.32	3.05	-
1980	3.05	-	-

1971 έως 1980, τα δεδομένα υποδηλώνουν χρόνο που παρουσιάζεται στη στήλη 1. Οι στήλες 2 και 4 παρουσιάζουν την ίδια χρονοσειρά με υστέρηση ενός και δύο χρόνων.

ΠΗΓΗ: *Value Line Selection and Opinion*, σ. 861, Ιανουάριος 9, 1981, Ν. Υόρκη: Arnold Bernhard and Co. Inc.

Επομένως η χρονοσειρά για  $Y_{t-1}$  χάνει μια παρατήρηση και για τη χρονοσειρά  $Y_{t-2}$  μπορεί να υπολογιστεί μάλλον εύκολα. Οι αυτοσυσχετίσεις αυτές δίνουν ενδείξεις για το πώς σχετίζονται μεταξύ τους διαδοχικές τιμές της ίδιας μεταβλητής. Η αυτοσυσχέτιση των  $Y_t$  και  $Y_{t-1}$  ονομάζεται *αυτοσυσχέτιση πρώτης τάξης* και η σχέση των  $Y_t$  και  $Y_{t-2}$  ονομάζεται *αυτοσυσχέτιση δεύτερης τάξης*.

Για υπολογιστικούς λόγους, μπορούμε να ξαναγράψουμε την εξίσωση που αναφέρθηκε παραπάνω, όσον αφορά τον συντελεστή αυτοσυσχέτισης τάξης 1, ως ακολούθως:

$$r_1 = \frac{\sum_1^{n-1} (Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{\sum_1^n (Y_t - \bar{Y})^2} \quad (320)$$

Η εξίσωση (3.20) ερμηνεύεται με τον ίδιο τρόπο που ερμηνεύεται και ο συντελεστής συσχέτισης μεταξύ δύο μεταβλητών δηλαδή, εάν ο συντελεστής αυτοσυσχέτισης υψωθεί στο τετράγωνο αποτελεί το λόγο της εξηγηθείσας διακύμανσης προς τη συνολική διακύμανση στο  $Y$ . Το γεγονός αυτό μας δείχνει αν η γραμμική αυτοπαλινδρόμησης μεταξύ  $Y_t$  και  $Y_{t-1}$  προβλέπει καλύτερα ή όχι από τη μέση

γραμμή του  $Y_t$ . Όσο πιο κοντά βρίσκεται το  $r_t$ , στο ένα ή το μείον ένα, τόσο καλύτερη είναι η αυτοπαλινδρόμηση ως μέσο πρόβλεψης (predictor).

Εξετάζουμε πώς χρησιμοποιείται η εξίσωση (3.20) για τον υπολογισμό του, από τα αρχικά δεδομένα. Για τα δεδομένα του παραδείγματος 3.1, έχουμε

$$r_1 = \frac{(0.93 - 1.71)(0.78 - 1.71) + (0.78 - 1.71)(1.84 - 1.71) + \dots + (2.32 - 1.71)(3.05 - 1.71)}{(0.93 - 1.71)^2 + (0.78 - 1.71)^2 + \dots + (3.05 - 1.71)^2} = 0.427$$

Όμοια, μπορούμε να υπολογίσουμε το  $r_2$ , ως εξής

$$r_2 = \frac{(0.93 - 1.71)(0.83 - 1.71) + (0.78 - 1.71)(1.84 - 1.71) + \dots + (2.22 - 1.71)(3.05 - 1.71)}{(0.93 - 1.71)^2 + (0.78 - 1.71)^2 + \dots + (3.05 - 1.71)^2} = 0.338$$

Με ανάλογο τρόπο μπορούμε να υπολογίσουμε:

$$r_3 = 0.024$$

$$r_4 = -0.011$$

$$r_5 = 0.020$$

και ούτω καθ' εξής.

Τέλος, η εξίσωση (7.20) περιέχει ιδιότητες που την καθορίζουν ευκολότερα εάν η χρονοσειρά είναι μη-σταθερή.

### 3.1.5 Κατανομή Δειγματοληψίας Αυτοσυσχετίσεων

Για να περιγράψουμε την έννοια μιας κατανομής δειγματοληψίας για μια σταθερή χρονοσειρά χρησιμοποιούμε το επόμενο παράδειγμα.

#### Παράδειγμα 3.2

Για την περιγραφή μιας στατικής χρονοσειράς, θεωρούμε ένα δείγμα τριών αριθμών που έχουν επιλεγεί από έναν πίνακα τυχαίων αριθμών. Οι αριθμοί είναι στο διάστημα από 000 έως 999 και έχουν μέσο 499.5. Επίσης, η πιθανότητα επιλογής οποιουδήποτε τριψήφιου αριθμού είναι ίδια με οποιουδήποτε άλλου και ισούται με 1/1000. Ένα δείγμα 40 αριθμών από έναν κοινό πίνακα τυχαίων αριθμών επιλέγεται για να αντιπροσωπεύσει παρατηρήσεις για διάστημα 40 περιόδων. Τα δεδομένα παρουσιάζονται στον Πίνακα 3.2.

**ΠΙΝΑΚΑΣ 3.2.** Χρονοσειρά 40 τυχαίων παρατηρήσεων

περίοδος	Παρατήρηση	Χρονική περίοδος	Παρατήρηση
1	457	21	793
2	351	22	239
3	791	23	369
4	558	24	186
5	429	25	523
6	412	26	788
7	795	27	534
8	592	28	120
9	659	29	797
10	703	30	309
11	830	31	500
12	138	32	792
13	695	33	858
14	245	34	615
15	423	35	807
16	682	36	959
17	438	37	349
18	822	38	046
19	614	39	932
20	982	40	725

ΠΗΓΗ: Δεδομένα τυχαίων αριθμών που συγκεντρώθηκαν από την A Million Random Digits With One Hundred Thousand Normal Deviates, , The Rand Corporation.

Στη θεωρία, αν μια χρονοσειρά είναι ένα σύνολο εντελώς τυχαίων και επομένως ανεξάρτητων παρατηρήσεων, τότε οι συντελεστές αυτοσυσχέτισης για όλες τις τάξεις πρέπει να είναι μηδέν. Ωστόσο, για ένα τυχαίο δείγμα 40 παρατηρήσεων, ο εκτιμηθείς συντελεστής αυτοσυσχέτισης (δείγματος) για οποιαδήποτε τάξη, 'θα μπορούσε να διαφέρει του μηδενός. Εφόσον ο αριθμός των τυχαίων αριθμών που μπορούν να επιλεγούν δεν είναι πεπερασμένος, τότε έχουμε έναν άπειρο αριθμό δειγμάτων που μπορεί να επιλεγεί από αυτόν τον πληθυσμό. Κάθε τυχαίο δείγμα 40 παρατηρήσεων έχει, πιθανόν, ένα διαφορετικό συντελεστή αυτοσυσχέτισης για οποιαδήποτε τάξη. Επιπλέον, αν επιλεγεί ένας αριθμός δειγμάτων από 40 τριψήφιους τυχαίους αριθμούς και είχαν προσδιοριστεί οι μέσοι των συντελεστών αυτοσυσχέτισης τους της τάξης 1,2,3,4,...,16, οι τιμές που θα προκύψουν θα βρίσκονται όλες γύρω από το μηδέν. Επομένως, εάν  $k$  συμβολίζει την αυτοσυσχέτιση για τον πληθυσμό, τότε οι αυτοσυσχετίσεις για διαφορετικά δείγματα παρατηρήσεων θα σχημάτιζαν μια κατανομή τιμών γύρω από το  $k$ . Η κατανομή αυτή ονομάζεται *κατανομή δειγματοληψίας των αυτοσυσχετίσεων*.

Η στατιστική θεωρία διδάσκει ότι η κατανομή δειγματοληψίας συντελεστών αυτοσυσχέτισης

τυχαίων δεδομένων είναι κανονική με

$$\begin{aligned}\mu_{\rho_k} &= 0 \\ \sigma_{\rho_k} &= 1/\sqrt{n}\end{aligned}$$

Εάν  $r_k$  είναι εκτίμηση του  $\rho_k$ , τότε μπορούμε να χρησιμοποιήσουμε τις γνώσεις μας για τις κανονικές κατανομές, προκειμένου να δημιουργήσουμε διαστήματα μέσα στα οποία θα βρίσκεται ο συντελεστής αυτοσυσχέτισης  $\rho_k$  του πληθυσμού. Εφόσον  $n$ , ο αριθμός των παρατηρήσεων στα τυχαία δεδομένα χρονοσειράς του παραδείγματος 7.2, είναι 40, το τυπικό σφάλμα είναι  $1/\sqrt{40} = 0.158$ . Εφαρμόζοντας τις γνώσεις μας για την κανονική κατανομή, μπορούμε να πούμε ότι το 99.73 τοις εκατό των συντελεστών αυτοσυσχέτισης του δείγματος, βρίσκονται μέσα στο διάστημα που καθορίζεται από τον μέσο όρο συν ή πλην τρία τυπικά σφάλματα του μέσου. Όμοια, το 95.45 τοις εκατό όλων των αυτοσυσχετίσεων του δείγματος βρίσκονται στο διάστημα που καθορίζεται από τον μέσο όρο συν ή πλην δύο τυπικά σφάλματα του μέσου όρου και το 68.3 τοις εκατό των συντελεστών συσχέτισης του δείγματος βρίσκονται με ένα τυπικό σφάλμα του μέσου όρου. Το ερώτημα παραμένει: σε ποιο επίπεδο ανάλυσης κινδύνου θέλουμε να συναγάγουμε ότι τα δεδομένα δεν είναι τυχαία, όταν στην πραγματικότητα αυτά είναι τυχαία; Με άλλα λόγια, εάν θέλουμε να δεχτούμε όρια σημαντικότητας των δύο τυπικών σφαλμάτων (95.45 τοις εκατό), μπορούμε να συμπεράνουμε ότι τα δεδομένα είναι τυχαία εάν οι συντελεστές αυτοσυσχέτισης του δείγματος βρίσκονται στο διάστημα:

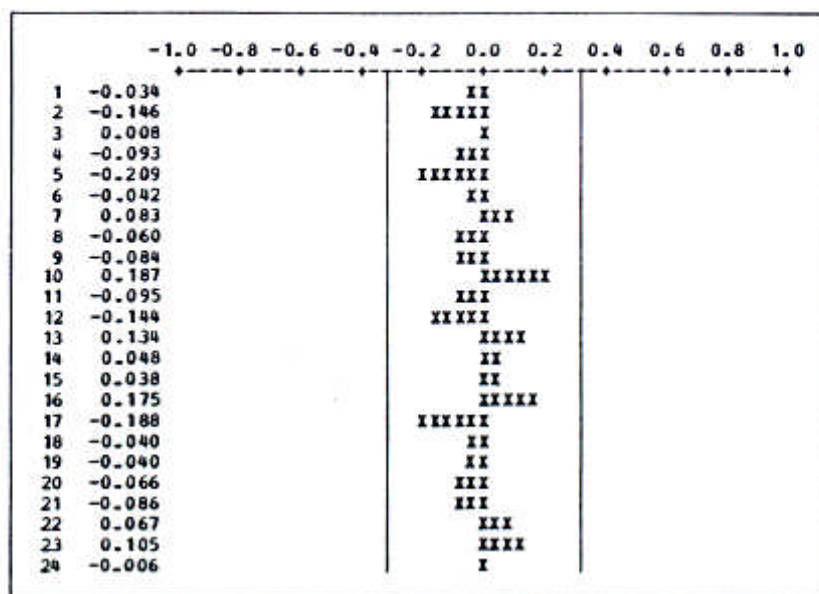
$$\begin{aligned}-2\sigma_{\rho} &\leq r_k \leq +2\sigma_{\rho} \\ -2(0.158) &\leq r_k \leq +2(0.158) \\ -0.316 &\leq r_k \leq 0.316\end{aligned}$$

Το σχήμα 3.1 δείχνει τους συντελεστές αυτοσυσχέτισης για τα δεδομένα του παραδείγματος 3.2 με χρονική υστέρηση 1,2,3,...,24. Οι δύο γραμμές χαραγμένες κάθετα στην οριζόντια, είναι τα όρια (συν και πλην δύο τυπικών σφαλμάτων) για μια τυχαία χρονοσειρά (-0.316, +0.316). Και οι 16 συντελεστές αυτοσυσχέτισης όντως βρίσκονται μέσα στα όρια του διαστήματος που επιβεβαιώνει την υπόθεση ότι η χρονοσειρά είναι τυχαία.

Θα μπορούσαμε να είχαμε ελέγξει κάθε χρονική υστέρηση,  $k$ , ξεχωριστά με την ακόλουθη υπόθεση:

$$\begin{aligned}H_0 &: \rho_k = 0 \\ H_A &: \rho_k \neq 0\end{aligned}$$

Σχήμα 3.1 Συντελεστές Αυτοσυσχέτισης ARIMA (1,0,0) (αποτελέσματα σύμφωνα με το πρόγραμμα Minitab)



Ωστόσο, η προσέγγιση του ορίου (ή ορίου εμπιστοσύνης) είναι απλή και δίνει τα ίδια αποτελέσματα. Χρησιμοποιώντας αυτή την προσέγγιση, έχουμε τη δυνατότητα να καθορίσουμε το αν οποιαδήποτε δεδομένα ή χρονοσειρά είναι τυχαία ή όχι.

Τέλος, θα πρέπει να σημειώσουμε ότι η έννοια της κατανομής δειγματοληψίας είναι τόσο κρίσιμα σημαντική στην ανάλυση χρονοσειρών όσο και σε άλλες εφαρμογές της στατιστικής θεωρίας. Ο συντελεστής αυτοσυσχέτισης της τάξης 5 είναι -0.209. Εφόσον το αποτέλεσμα αυτό βασίζεται σε πληροφορίες δείγματος, μπορούμε να συμπεράνουμε ότι η τιμή αυτή διαφέρει από το μηδέν εξαιτίας του σφάλματος δειγματοληψίας. Η κατανομή δειγματοληψίας μάς εφοδιάζει με μια μέθοδο για να συναγάγουμε εάν η διαφορά του δείγματος οφείλεται σε σφάλμα δειγματοληψίας ή σε κάποια υφιστάμενη σημαντική σχέση. Η τιμή -0.209, όπως και όλες οι άλλες τιμές, δεν οφείλονται στην ύπαρξη κάποιας σημαντικής σχέσης.

Εάν το μέγεθος του δείγματος στο παράδειγμα 3.2 ήταν 400, αντί 40, το τυπικό σφάλμα θα ήταν 0.05. Ακολουθώντας τις ίδιες μεθόδους, τα όρια των δύο τυπικών σφαλμάτων θα βρίσκονται τώρα στο 0.10 του μέσου όρου. Σε αυτή την περίπτωση, οι συντελεστές αυτοσυσχέτισης της τάξης 2,5,10,12,13 και 16 έχουν τιμές που βρίσκονται έξω από τα αποδεκτά όρια. Φυσικά, εάν το μέγεθος του δείγματος ήταν 400, η αυτοσυσχέτιση του δείγματος θα ήταν επίσης διαφορετική. Τέλος, θα πρέπει να σημειώσουμε ότι η επιτυχία στην ανάλυση χρονοσειρών οφείλεται σε μεγάλο βαθμό στην ερμηνεία των αποτελεσμάτων από την ανάλυση των συντελεστών αυτοσυσχέτισης για να ξεχωριστεί το πρότυπο στη χρονοσειρά από την τυχαιότητα.

### 3.1.6 Είναι τα Δεδομένα Τυχαία;

Οι διαδικασίες της προηγούμενης παραγράφου έδειξαν μια μέθοδο καθορισμού της τυχαιότητας ή μη, δεδομένης χρονοσειράς. Οι συντελεστές αυτοσυσχέτισης δύο ή περισσότερων χρονικών υστερήσεων μετρήθηκαν και εξετάστηκαν για να δούμε αν κάποιος απ' αυτούς διαφέρει σημαντικά από το μηδέν. Τοποθετώντας τους συντελεστές σε γράφημα αυτοσυσχέτισης, όπως στο σχήμα 3.1, μπορούμε να προσδιορίσουμε την ύπαρξη ή όχι πρότυπου. Δεν υπήρξε πρότυπο στο σχήμα 3.1.

Στην παράγραφο 2,1, εξετάσαμε κατάλοιπα από μοντέλα πρόβλεψης κάνοντας τη γραφική παράσταση της χρονοσειράς τους. Εάν υπήρχε πρότυπο στα δεδομένα, η γραφική αποτύπωση των καταλοίπων θα αναδείκνυε αυτό το πρότυπο. Σε αυτή την παράγραφο προχωρούμε ένα ακόμη βήμα στην ανάλυση, υπολογίζοντας αυτοσυσχετίσεις για τις σειρές καταλοίπων των σφαλμάτων ώστε να καθορίσουμε εάν είναι τυχαίες ή όχι.

Για να βρούμε πώς μπορούν να εξεταστούν εκτιμημένα (fitted) μοντέλα ώστε να δούμε εάν τα σφάλματα από την πρόβλεψη είναι τυχαία ή όχι, χρησιμοποιούμε το ακόλουθο παράδειγμα.

#### Παράδειγμα 3.3

Στο παράδειγμα 3.2 αυτού του κεφαλαίου εξετάσαμε τα δεδομένα σε δολάρια για Δάνεια Ακίνητης Περιουσίας στις ΗΠΑ. Οι προβλέψεις προετοιμάστηκαν με απλή (γραμμική) εκθετική εξομάλυνση και με το μοντέλο μας παραμέτρου του Brown με προσαρμογή για την τάση. Σε αυτό το παράδειγμα, θέλουμε να εξετάσουμε τα κατάλοιπα από την πρόβλεψη για να δούμε εάν τα δεδομένα είναι τυχαία ή όχι.

Ο Πίνακας 3.3 περιέχει τα κατάλοιπα από την πρόβλεψη για την παρατηρηθείσα χρονοσειρά. Η στήλη 1 του Πίνακα 3.3 είναι η διαφορά ανάμεσα στη στήλη 1, τις παρατηρηθείσες τιμές και τις προβλεφθείσες τιμές (στήλη 2) από την απλή εκθετική εξομάλυνση του Πίνακα 3.2. Η στήλη 2 του Πίνακα 3.3 είναι η διαφορά της στήλης 1 και στήλης 6, οι προβλεφθείσες τιμές με τις μεθόδους του Brown, του Πίνακα 3.2. Δεν έγιναν καθόλου προβλέψεις για τις περιόδους 1 και 2 με τις μεθόδους Brown και επομένως δεν υπάρχουν κατάλοιπα για αυτές τις περιόδους.

Οι συντελεστές αυτοσυσχέτισης μπορούν να υπολογιστούν από τα κατάλοιπα του Πίνακα 3.3 με την εξίσωση (3.15). Για τα δεδομένα της στήλης 2 (Πίνακας 3.3) μπορούμε να δούμε ότι υπάρχει ένα πρότυπο στα σφάλματα που δείχνει ότι η απλή εκθετική εξομάλυνση δεν ήταν το κατάλληλο μοντέλο πρόβλεψης. Αν και το πρότυπο στα σφάλματα είναι εμφανές σε αυτό το παράδειγμα, σε πολλές περιπτώσεις μπορεί να μην είναι. Επομένως, η προσέγγιση για τον καθορισμό της τυχαιότητας ή μη των σφαλμάτων βρίσκεται στον υπολογισμό των αυτοσυσχετίσεων των καταλοίπων από την πρόβλεψη (των σφαλμάτων).

Το σχήμα 3.2 αποτελεί τη γραφική παράσταση των αυτοσυσχετίσεων για τα κατάλοιπα του μοντέλου απλής εκθετικής εξομάλυνσης του παραδείγματος 3.2.

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

Εφόσον υπήρχαν 24 περίοδοι, τα όρια των δύο σίγμα είναι

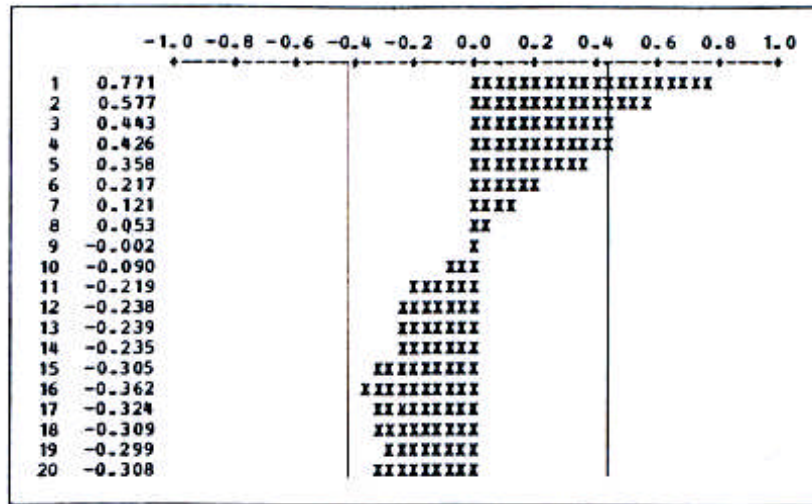
$$\begin{aligned} -2 \frac{I}{\sqrt{24}} &\leq r_k \leq 2 \frac{I}{\sqrt{24}} \\ -2(0.204) &\leq r_k \leq +2(0.204) \\ -0.408 &\leq r_k \leq 0.408 \end{aligned}$$

**ΠΙΝΑΚΑΣ 3.3.** Κατάλοιπα από την πρόβλεψη της χρονοσειράς δανείων ακίνητης περιουσίας (\$ δις)

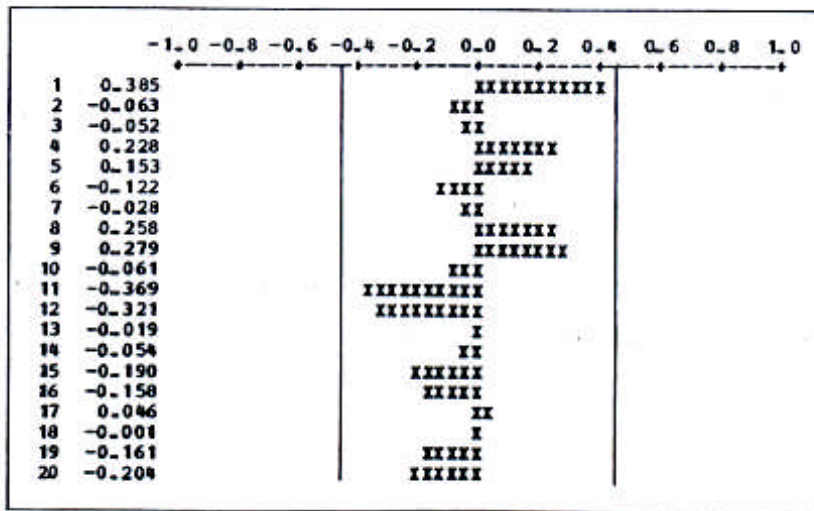
Περίοδος	(1) <i>Μοντέλο Απλής Εκθετικής Εξομάλυνσης</i>	(2) <i>Κατάλοιπα από το μοντέλο τάσης του Brown</i>
1	0.00	XXX
2	0.80	XXX
3	1.40	1.50
4	2.70	2.60
5	2.20	1.40
6	1.80	0.70
7	2.20	1.10
8	2.60	1.30
9	2.90	1.40
10	3.10	1.40
11	3.30	1.30
12	4.20	2.20
13	4.20	1.60
14	4.20	1.40
15	3.40	0.20
16	3.50	0.40
17	4.40	1.30
18	4.30	1.00
19	4.20	0.60
20	4.20	0.40
21	5.00	1.40
22	4.80	0.90
23	4.60	0.50
24	4.50	0.20



Σχημα 3.2: Οι συντελεστές αυτοσυσχέτισης των καταλοίπων από την πρόβλεψη δανείων ακίνητης περιουσίας με το μοντέλο απλής εκθετικής εξομάλυνσης (αποτελέσματα σύμφωνα με το πρόγραμμα Minitab)



Σχημα 3.3 Συντελεστές αυτοσυσχέτισης των καταλοίπων από πρόβλεψη των δανείων ακίνητης περιουσίας με το μοντέλο Brown για την τάση (αποτελέσματα σύμφωνα με το πρόγραμμα Minitab)



Από το σχήμα 3.2 η αυτοσυσχέτιση που σχετίζεται με χρονικές υστερήσεις 1,2,3 και 4 περιόδων είναι αντίστοιχα 0.771, 0.577, 0.443 και 0.426. Οι τιμές αυτές βρίσκονται έξω από τα όρια που δείχνουν ότι οι όροι του σφάλματος αυτοσυσχετίζονται στο προβλεπόμενο επίπεδο σημαντικότητας (περίπου 0.045). Το αποτέλεσμα ήταν αναμενόμενο εφόσον τα δεδομένα επί των Δανείων Ακίνητης Περιουσίας εμφανίζουν κάποια τάση και η απλή εκθετική εξομάλυνση δεν μπορεί να μεταχειριστεί αυτή την τάση σωστά.

Το σχήμα 3.3 δείχνει τους συντελεστές αυτοσυσχέτισης για τα κατάλοιπα του εκθετικού μοντέλου πρόβλεψης του Brown: Τα όρια των δύο σίγμα για αυτά τα δεδομένα είναι (θυμηθείτε πως εδώ έχουμε 22 περιόδους):

$$\begin{aligned}
 -2 \frac{1}{\sqrt{22}} &\leq r_k \leq 2 \frac{1}{\sqrt{22}} \\
 -2(0.213) &\leq r_k \leq +2(0.213) \\
 -0.426 &\leq r_k \leq 0.426
 \end{aligned}$$

Σημειώστε ότι δεν υπάρχουν συντελεστές αυτοσυσχέτισης που να βρίσκονται έξω από τα όρια του παραπάνω διαστήματος. Επομένως, δεν εμφανίζεται σημαντική αυτοσυσχέτιση και το μοντέλο του Brown εξάλειψε πλήρως την τάση και παρήγαγε ένα κατάλοιπο για το οποίο δεν ακολουθεί κανένα στατιστικά σημαντικό πρότυπο.

Οι αναλύσεις αυτές δείχνουν εάν μια χρονοσειρά δεδομένων (ή καταλοίπων πρόβλεψης) είναι τυχαία ή όχι. Η διαδικασία ολοκληρώνεται με τον υπολογισμό των συντελεστών αυτοσυσχέτισης και τη γραφική τους παράσταση για να δούμε αν διαφέρουν ή όχι στατιστικά από το μηδέν.

### 3.1.7 Είναι τα Δεδομένα Μη-Σταθερά (Nonstationary);

Μια μη-εποχική ομοσκεδαστική χρονοσειρά είναι *σταθερή* όταν δεν περιέχει στοιχεία ανάπτυξης ή απόκλισης, και *μη-σταθερή* όταν ακολουθεί κάποιο πρότυπο. Για μια μη-σταθερή χρονοσειρά οι συντελεστές αυτοσυσχέτισης είναι (τυπικά) στατιστικά διάφοροι του μηδενός για αρκετές από τις πρώτες χρονικές υστερήσεις, και μόνο σταδιακά πέφτουν στο μηδέν ή δείχνουν ένα πλασματικό πρότυπο, καθώς αυξάνεται ο αριθμός των χρονικών περιόδων. Το σχήμα 3.2 είναι ένα καλό παράδειγμα συντελεστών αυτοσυσχέτισης για μια μη-σταθερή χρονοσειρά. Στατιστικά σημαντικοί συντελεστές αυτοσυσχέτισης υπάρχουν για χρονικές υστερήσεις περισσότερες των 4 και το πρότυπο στα κατάλοιπα πέφτει στο μηδέν και στη συνέχεια γίνεται αρνητικό, αλλά όχι σημαντικό. Δια μέσου των καταλοίπων, μπορούμε να χαράξουμε μια ευθεία.

Όπως αποδείχτηκε παραπάνω, όταν υπάρχει μια τάση στα δεδομένα, οι διαδοχικές τιμές σχετίζονται μεταξύ τους. Η αυτοσυσχέτιση για μια χρονική υστέρηση είναι μεγάλη, αλλά η αυτοσυσχέτιση για μια υστέρηση δύο περιόδων δεν είναι τόσο μεγάλη. Αυτό συμβαίνει επειδή χρησιμοποιείται ένας όρος λιγότερος για τον υπολογισμό του αριθμητή του συντελεστή αυτοσυσχέτισης.

Εάν μια χρονοσειρά δείχνει στατιστικά σημαντικούς συντελεστές αυτοσυσχέτισης για δύο (ή περισσότερες) χρονικές υστερήσεις, αλλά όχι παραπάνω, η χρονοσειρά λέγεται ότι είναι μη-τυχαία αλλά σταθερή. Αν οι αυτοσυσχετίσεις είναι στατιστικά διάφορες του μηδενός για δύο μόνο χρονικές υστερήσεις, δεν παρουσιάζεται στοιχείο τάσης στα δεδομένα. Εντούτοις έχουμε ενδείξεις ύπαρξης ενός προτύπου διάφορου της τάσης

3.1.8 Εξάλειψη της Τάσης στις Χρονοσειρές

Οι πλασματικές συσχετίσεις λέγεται ότι κυριαρχούν στο πρότυπο της χρονοσειράς όταν υπάρχει τάση. Γι' αυτόν το λόγο, η τάση πρέπει να εξλειφθεί πριν προχωρήσουμε στην παραπέρα ανάλυση. Μια πολύ χρήσιμη μέθοδος είναι η μέθοδος των πρώτων διαφορών. Θεωρούμε την απλή σειρά 3,6,9,12,15,18, που αποτελείται από μία μη-τυχαία γραμμική τάση. Εάν αφαιρέσουμε διαδοχικές τιμές, 6-3, 9-6,12-9,15-12 και 18-15, το αποτέλεσμα είναι οι πρώτες διαφορές 3,3,3,3,3. Αυτή η σειρά είναι προφανώς σταθερή. Έτσι, δημιουργείται μια νέα σταθερή σειρά που αποτελείται από τις διαφορές μεταξύ διαδοχικών περιόδων:

$$\nabla Y_t' = Y_{t+1} - Y_t$$

Η νέα σειρά  $Y_t'$  περιλαμβάνει  $n-1$  τιμές και θα είναι σταθερή εάν η τάση στα αρχικά δεδομένα  $Y_t$  είναι γραμμική (της τάξης 1).

Για να εκφράσουμε και να κατανοήσουμε τα διαφορικά μοντέλα ARIMA, πρέπει να εισαγάγουμε την έννοια του τελεστή επιστροφής,  $B$ . Αυτός ο χειριστής δεν έχει μαθηματική έννοια άλλη από το να διευκολύνει τη γραφή των διαφόρων τύπων μοντέλων, που διαφορετικά θα ήταν πολύ δύσκολο να εκφραστούν. Ο τελεστής, επιστροφής ορίζεται ως  $B^m Y_t$ . Για παράδειγμα,

$$\begin{aligned} BY_t &= Y_{t-1} \quad \text{ή} \quad Be_t = e_{t-1} \\ B^2 Y_t &= Y_{t-2} \quad \text{ή} \quad B^2 e_t = e_{t-2} \\ B^3 Y_t &= Y_{t-3} \quad \text{ή} \quad B^3 e_t = e_{t-3} \\ &\dots\dots\dots \\ B^m Y_t &= Y_{t-m} \quad \text{ή} \quad B^m e_t = e_{t-m} \end{aligned}$$

Οποιοδήποτε μοντέλο ARIMA μπορεί να εκφραστεί με τελεστή επαναφοράς. Ένα μοντέλο AR(1), ARIMA(1,0,0) εκφράζεται ως

$$\begin{aligned} Y_t &= \phi_1 Y_{t-1} + e_t \\ \text{ή} & \\ Y_t - \phi_1 Y_{t-1} &= e_t \end{aligned} \tag{3.21}$$

Ωστόσο, εφόσον  $Y_{t-1} = BY_t$ , η εξίσωση (3.21) μπορεί να γραφεί ως

$$Y_t - \phi_1 BY_t = e_t$$

Για ένα μοντέλο AR(2),  $Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} = e_t$ , ARIMA(2,0,0), μπορούμε να γράψουμε

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

και

$$Y_{t-2} = B^2 Y_t$$

το οποίο μπορεί να ξαναγραφεί ως

$$Y_t - \phi_1 B Y_t - \phi_2 B^2 Y_t = e_t$$

ή ακόμη ως

$$(1 - \phi_1 B - \phi_2 B^2) Y_t = e_t \quad (3.22)$$

Εάν έχουμε μοντέλο AR(2), όπου έχουν χρησιμοποιηθεί οι πρώτες διαφορές, ARIMA(2,1,0), η έκφραση είναι:

$$(Y_t - Y_{t-1}) = \phi_1 (Y_{t-1} - Y_{t-2}) + \phi_2 (Y_{t-2} - Y_{t-3}) + e_t$$

Εφόσον  $Y_{t-1} = B Y_t$ ,  $Y_{t-2} = B Y_{t-1}$  και  $Y_{t-3} = B Y_{t-2}$ , η έκφραση μπορεί να ξαναγραφεί ως

$$(1 - B) Y_t = \phi_1 (1 - B) Y_{t-1} + \phi_2 (1 - B) Y_{t-2} + e_t \quad (3.23)$$

Ωστόσο, εφόσον  $B Y_t = Y_{t-1}$  και  $B^2 Y_t = Y_{t-2}$  η εξίσωση (7.20) γίνεται

$$\begin{aligned} (1 - B) Y_t &= \phi_1 (1 - B) B Y_t + \phi_2 (1 - B) B^2 Y_t + e_t \\ \text{ή} & \\ (1 - B) Y_t - \phi_1 (1 - B) B Y_t - \phi_2 (1 - B) B^2 Y_t &= e_t \end{aligned} \quad (3.24)$$

Τέλος, εάν  $(1 - B) Y_t$  χρησιμοποιηθεί σαν κοινός παράγοντας, τότε η εξίσωση (7.21) γίνεται:

$$\begin{aligned} Y_t (1 - B) (1 - \phi_1 B - \phi_2 B^2) &= e_t \\ \text{ή} & \\ (1 - B) (1 - \phi_1 B - \phi_2 B^2) Y_t &= e_t \end{aligned} \quad (3.25)$$

Η εξίσωση (3.22) είναι η ανάλογη της διαδικασίας AR(2) των πρώτων διαφορών.

Όπως θα δούμε στη συζήτησή μας για τις τεχνικές των χρονοσειρών, η χρήση του τελεστή επαναφοράς μάς επιτρέπει να εκφράσουμε πιο εύκολα ένα μοντέλο χρονοσειράς. Και η τάξη και ο τύπος της διαδικασίας εκφράζονται σε όρους του εκθέτη του  $\mathbf{B}$ , του τελεστή επαναφοράς, και του τύπου της παραμέτρου.

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

### Παράδειγμα 3.4

Στον Πίνακα 3.4 παρουσιάζονται τα δεδομένα που αφορούν αποδοχές από μερίσματα για τα έτη 1965 έως 1980 της Chase Manhattan Corporation.

Βρίσκουμε τις πρώτες διαφορές αφού αφαιρέσουμε την παρατήρηση της προηγούμενης περιόδου από την παρατήρηση στη δεδομένη περίοδο. Για παράδειγμα,

$$\nabla Y_{1966} = Y_{1966} - Y_{1965}$$

**ΠΙΝΑΚΑΣ 3.4.** Πρώτες διαφορές επί των δεδομένων για τα κέρδη ανά μέρισμα της Chase Manhattan Corporation

Έτος	(1) $Y_t$	(2) $BY_t$	(3) $\nabla Y_t = Y_t - BY_t$	(4) $\nabla^2 Y_t = \nabla Y_t - \nabla Y_{t-1}$
1965	2.93	—	—	—
1966	3.14	2.93	0.21	—
1967	3.35	3.14	0.21	0.00
1968	3.80	3.35	0.45	0.24
1969	3.59	3.80	-0.21	-0.66
1970	4.17	3.59	0.58	0.79
1971	4.63	4.17	0.46	-0.12
1972	4.65	4.63	0.02	-0.44
1973	5.15	4.65	0.50	0.48
1974	5.68	5.15	0.53	0.03
1975	4.89	5.68	-0.79	-1.32
1976	3.28	4.89	-1.61	-0.82
1977	3.71	3.28	0.43	2.04
1978	5.59	3.71	1.88	1.45
1979	9.07	5.59	3.48	1.60
1980	10.47	9.07	1.40	-2.08

$\nabla$  δείχνει τις πρώτες διαφορές

$\nabla^2$  δείχνει τις δεύτερες διαφορές

Βλέπουμε ότι, τα δεδομένα στη στήλη 3 του Πίνακα 3.4 είναι η διαφορά ανάμεσα στις στήλες 2 και 1. Αυτά τα δεδομένα είναι, επομένως, οι πρώτες διαφορές της αρχικής χρονοσειράς.

Για να καθορίσουμε αν η σειρά των πρώτων διαφορών δίνει μια σταθερή χρονοσειρά από μια αρχική που δείχνει ένα πρότυπο τάσης, παράγονται συντελεστές αυτοσυσχέτισης για τις πρώτες διαφορές των δεδομένων που αφορούν αποδοχές μερισμάτων της Chase Manhattan Corporation. Σημειώστε ότι μόνο ο πρώτος συντελεστής αυτοσυσχέτισης είναι σημαντικά διαφορετικός του μηδενός. Η ανάλυση δείχνει ότι τα δεδομένα πρώτων διαφορών της χρονοσειράς έχουν μετατραπεί σε σταθερή μορφή.

Εάν οι συντελεστές αυτοσυσχέτισης των δεδομένων των πρώτων διαφορών εξακολουθούν σταδιακά να πλησιάζουν το μηδέν, δεν έχουμε επιτύχει μια σταθερή κατάσταση. Για να λύσουμε το πρόβλημα, παίρνουμε τις δεύτερες διαφορές. Αυτό δεν είναι απαραίτητο για τα δεδομένα του παραδείγματος 3.4.

Για παράδειγμα, οι δεύτερες διαφορές για το έτος 1967 είναι ως εξής

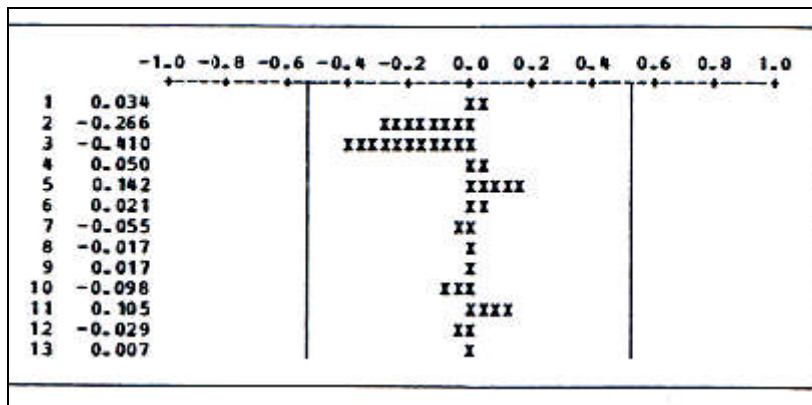
$$\begin{aligned}\nabla^2 Y_{1967} &= \nabla Y_{1967} - \nabla Y_{1966'} \\ &= 0.21 - 0.21 = 0.00\end{aligned}$$

Για το 1968, έχουμε

$$\begin{aligned}\nabla^2 Y_{1968} &= \nabla Y_{1968} - \nabla Y_{1967'} \\ &= 0.45 - 0.21 = 0.24\end{aligned}$$

Οι δεύτερες διαφορές για τα δεδομένα χρονοσειράς της Chase Manhattan δίνονται στη στήλη 4 του Πίνακα 3.4. Τα δεδομένα αυτά μπορούν να χρησιμοποιηθούν για να μετρήσουμε την αυτοσυσχέτιση της χρονοσειράς των δεύτερων διαφορών. Σημειώστε ότι στο σχήμα 3.4, παρουσιάζονται οι συντελεστές αυτοσυσχέτισης για τη χρονοσειρά των δεύτερων διαφορών.

**Σχημα 3.4** Συντελεστές αυτοσυσχέτισης για τη χρονοσειρά των δεύτερων διαφορών (αποτελέσματα σύμφωνα με το πρόγραμμα Minitab)



Εφόσον κανένας συντελεστής δεν βρίσκεται εκτός των ορίων των δυο σίγμα, μπορούμε να καταλήξουμε ότι η χρονοσειρά αυτή είναι σταθερή.

Επομένως, μπορούμε να συνοψίσουμε τη διαδικασία επίτευξης σταθερότητας σε μια χρονοσειρά. Πρώτον, υπολογίζουμε τις αυτοσυσχετίσεις για την αρχική χρονοσειρά και βλέπουμε εάν υπάρχει κάποιο πρότυπο. Εάν το πρότυπο δείχνει τάση, τότε αυτός ο ιδιαίτερος τύπος προτύπου υπαινίσσεται μη-σταθερότητα. Για να απαλλαγούμε από τη μη-σταθερότητα, η σειρά μετασχηματίζεται σε σειρά των πρώτων διαφορών και διενεργείται δεύτερη ανάλυση αυτοσυσχέτισης. Εάν επιτευχθεί σταθερότητα, η ανάλυση περατούται. Εάν δεν επιτευχθεί, υπολογίζουμε τις δεύτερες διαφορές και διενεργούμε νέα ανάλυση αυτοσυσχέτισης. Έτσι, η επίτευξη της σταθερότητας υποβαθμίζεται σε μια μάλλον μηχανική κίνηση λήψης διαδοχικών διαφορών, μέχρις ότου οι αυτοσυσχετίσεις πέσουν στο μηδέν με μία, δύο ή τρεις χρονικές υστερήσεις.

Συνήθως, δεν συνεχίζουμε πέρα από τις δεύτερες διαφορές, επειδή τα δεδομένα χρονοσειρών παρουσιάζουν μη στασιμότητα κυρίως στο πρώτο ή στο δεύτερο επίπεδο.

**3.1.9 Τα Δεδομένα είναι Εποχικά;**

Όπως είδαμε Στην παράγραφο 1.2, υπάρχει εποχιακότητα όταν ένα μοντέλο επαναλαμβάνει τον εαυτό του κατά τη διάρκεια μιας συγκεκριμένης περιόδου του χρόνου, συνήθως του έτους. Με την εξέταση των συντελεστών αυτοσυσχέτισης, μπορούμε να αναγνωρίσουμε ένα εποχικό μοντέλο.

Για παράδειγμα, η αυτοσυσχέτιση 12-μηνων υστερήσεων μπορεί να δώσει υψηλή θετική τιμή. Ακόμη, εάν οι συντελεστές ήταν σημαντικά διάφοροι του μηδενός, θα συμπεραίναμε ότι οι μήνες που απέχουν κατά ένα χρόνο μεταξύ τους, σχετίζονται ή είναι μη-τυχαίοι.

Η ανάλυση αυτοσυσχέτισης εφαρμόζεται στο πρόβλημα ταυτοποίησης της εποχιακότητας στα στατικά δεδομένα. Για να αναγνωρίσουμε την εποχιακότητα μετράμε περισσότερες από δύο ή τρεις χρονικές υστερήσεις ώστε να ορίσουμε εάν είναι στατιστικά διάφορες του μηδενός. Οι αυτοσυσχετίσεις που είναι σημαντικά διάφορες του μηδενός, δηλώνουν την ύπαρξη ενός προτύπου. Για την εύρεση της εποχιακότητας πρέπει να αναζητήσουμε σημαντικές αυτοσυσχετίσεις υψηλότερων τάξεων.

Παράδειγμα 3.5

Ο Πίνακας 3.5 περιέχει δεδομένα που αφορούν μεταφορές μιας εταιρίας χωρίς προσαρμογή για απόσβεση κεφαλαίου σε τριμηνιαία αθροίσματα που δεν είναι εποχικά προσαρμοσμένα (\$ χιλιάδες). Τα δεδομένα της χρονοσειράς δεν είναι εποχικά προσαρμοσμένα. Υπάρχει ένδειξη ύπαρξης σημαντικής εποχιακότητας στη χρονοσειρά;

**ΠΙΝΑΚΑΣ 3.5** Μεταφορές αποθεμάτων χωρίς προσαρμογή για απόσβεση κεφαλαίου, τριμηνιαία αθροίσματα μη-εποχικά προσαρμοσμένα (\$ χιλιάδες)

		Έτος					
Τρίμηνο	1	2	3	4	5	6	
1	42.6	46.0	45.9	52.6	54.3	55.0	
2	46.4	53.7	55.9	48.7	52.9	54.7	
3	48.0	53.3	52.2	50.2	52.6	51.8	
4	56.5	67.1	61.0	61.0	60.7	62.0	

Το σχήμα 3.5 δείχνει ένα σχεδιαγράφημα των συντελεστών αυτοσυσχέτισης για τα δεδομένα του Πίνακα 3.5. Σημειώστε ότι οι μεγαλύτεροι συντελεστές αυτοσυσχέτισης είναι οι  $r_4=0.592$  και  $r_8=0.422$ . Οι συντελεστές αυτοί είναι σημαντικοί εφόσον τα όρια είναι

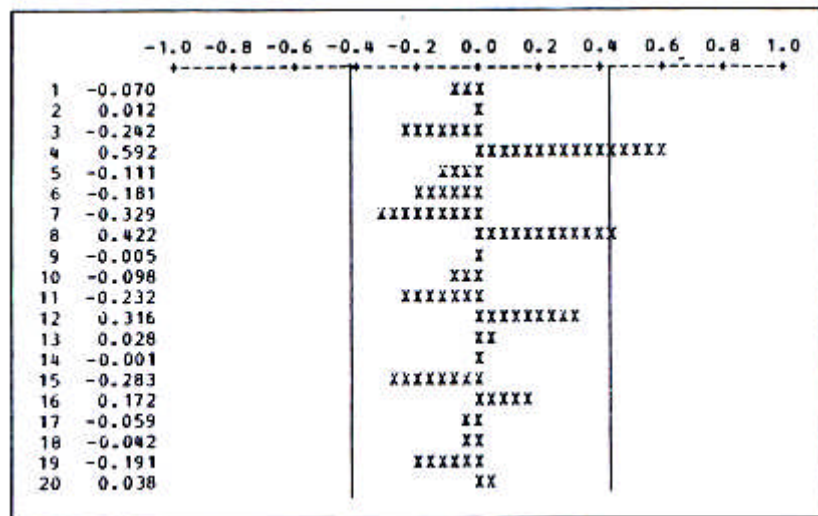
$$-2 \frac{1}{\sqrt{24}} \leq r_k \leq 2 \frac{1}{\sqrt{24}}$$

$$-0.408 \leq r_k \leq 0.408$$

στο επίπεδο των δύο σίγμα. Το σχήμα 7.5 δείχνει ότι αυτή η χρονοσειρά είναι εποχική με τη στατιστική έννοια. Εάν τα δεδομένα δεν ήταν εποχικά, τότε τα  $r_4, r_8$  κλπ., δεν θα είχαν τιμές εκτός των ορίων σημαντικότητας.

Η εποχιακότητα είναι συνήθως εύκολο να αναγνωριστεί όταν τα δεδομένα δεν έχουν ένα πρότυπο που να παράγει μη-σταθερότητα. Εάν υπάρχει το πρότυπο, είναι πιο δύσκολο να καθορίσουμε την παρουσία εποχιακότητας. Εφόσον υπάρχει πρότυπο, πρέπει πρώτα να μετασχηματιστούν τα δεδομένα σε σταθερή σειρά και στη συνέχεια να εξεταστούν για την παρουσία εποχιακότητας.

**Σχίμα 3.5:** Συντελεστές αυτοσυσχέτισης φορτίων για δεδομένα που αφορούν αποθέματα (αποτελέσματα σύμφωνα με το πρόγραμμα Minitab)



### 3.1.10 Συντελεστές Μερικής Αυτοσυσχέτισης

Κατά την ανάλυση ενός παλινδρομικού μοντέλου, μπορεί κανείς να ερωτήσει για την ερμηνευτική ισχύ μιας ανεξάρτητης μεταβλητής, εάν οι επιδράσεις των άλλων ανεξάρτητων μεταβλητών κρατούνταν σταθερές. Αυτό όμως, ίσως απαιτούσε την παλινδρόμηση του  $Y$  στο  $X_1$  (πχ. σε ένα παλινδρομικό μοντέλο με δύο ανεξάρτητες μεταβλητές), την εύρεση των σφαλμάτων και την παλινδρόμηση των καταλοίπων έναντι του  $X_2$ . Στην ανάλυση χρονοσειρών αυτού του κεφαλαίου υπάρχει μια παρόμοια έννοια.

Οι συντελεστές μερικής αυτοσυσχέτισης μετρούν το βαθμό της σχέσης μεταξύ των  $Y_t$  και  $Y_{t-k}$ , όταν το αποτέλεσμα των άλλων χρονικών υστερήσεων στο  $Y$ , κρατείται σταθερό. Όπως θα δούμε σε επόμενα κεφάλαια, ο σκοπός τους είναι να βοηθήσουν στην αναγνώριση του καλύτερου αυτοπαλινδρομικού μοντέλου-κινητού μέσου (ARMA) για την εξυπηρέτηση της πρόβλεψης.

Η μερική αυτοσυσχέτιση επιτρέπει στον προβλέποντα την αναγνώριση του βαθμού των σχέσεων μεταξύ τρεχουσών τιμών μιας μεταβλητής και προηγούμενων τιμών της ίδιας μεταβλητής, ενώ κρατούνται σταθερές οι επιδράσεις όλων των άλλων χρονικών υστερήσεων. Ειδικότερα, την υπολογίζουμε όταν δεν γνωρίζουμε τη σωστή τάξη της διαδικασίας αυτοπαλινδρόμησης που ταιριάζει στη χρονοσειρά. Το πρόβλημα αυτό είναι παρόμοιο με το πρόβλημα της παράγραφου 2.2, όπου



έπρεπε να αποφασίσουμε πόσες ανεξάρτητες μεταβλητές έπρεπε να συμπεριλάβουμε σε ένα μοντέλο πολλαπλής παλινδρόμησης.

Ο συντελεστής μερικής αυτοσυσχέτισης ορίζεται σε όρους του τελευταίου όρου αυτοπαλινδρόμησης ενός μοντέλου AR με  $m$  υστερήσεις. Με το ελληνικό γράμμα  $\phi$  συμβολίζουμε τη μερική αυτοσυσχέτιση, ενώ με  $\hat{\phi}$  την εκτιμηθείσα μερική αυτοσυσχέτιση. Επομένως, τα  $\phi_1, \phi_2, \dots, \phi_m$  είναι οι  $m$  μερικές αυτοσυσχετίσεις του μοντέλου AR ( $m$ ), όπως ορίζεται στις επόμενες εξισώσεις:

$$Y_t = \phi_1 Y_{t-1} + e_t \quad (3.26)$$

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t \quad (3.27)$$

·  
·  
·

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_{m-1} Y_{t-m-1} + e_t \quad (3.28)$$

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_{m-1} Y_{t-m-1} + \phi_m Y_{t-m} + e_t \quad (3.29)$$

Λύνοντας το παραπάνω σύνολο εξισώσεων για  $\phi_1, \phi_2, \dots, \phi_{m-1}$  μπορούμε να καθορίσουμε τις τιμές τους. Οι υπολογισμοί είναι σύνθετοι και χρονοβόροι. Μια πιο ικανοποιητική προσέγγιση είναι να πάρουμε εκτιμήσεις των  $\phi_1, \phi_2, \dots, \phi_m$  με βάση τους συντελεστές αυτοσυσχέτισής τους. Οι εκτιμήσεις αυτές γίνονται με την ακόλουθη μέθοδο που περιγράφεται στο παράρτημα Α αυτού του κεφαλαίου

$$\hat{\phi}_l = \hat{r}_l = r_l$$

Η μερική αυτοσυσχέτιση μιας χρονικής υστέρησης είναι  $\hat{\phi}_l$  και  $r_l$  είναι ο εκτιμητής της.

Για να χρησιμοποιήσουμε πληροφορίες σχετικές με τις μερικές συσχετίσεις πρέπει να ελέγξουμε πρώτα εάν είναι σημαντικά διάφορες του μηδενός. Η μηδενική και η εναλλακτική υπόθεση γράφονται ως εξής:

$$H_0 : \phi_k = 0$$

$$H_A : \phi_k \neq 0$$

Εάν η θεμελιώδης διαδικασία που δημιουργεί μια δεδομένη χρονοσειρά είναι ένα μοντέλο AR(1), θα πρέπει να κατανοήσουμε ότι μόνο το  $\phi_1$  θα είναι σημαντικά διάφορο του μηδενός, ενώ τα  $\phi_2, \phi_3, \dots, \phi_{m-1}, \phi_m$  δεν θα είναι σημαντικά διάφορα του μηδενός. Με παρόμοιο τρόπο, όταν η θεμελιώδης διαδικασία που δημιουργεί τη σειρά είναι AR(2), τότε, μόνο τα  $\phi_1$  και  $\phi_2$  θα είναι στατιστικά σημαντικά και όλες οι υπόλοιπες μερικές αυτοσυσχετίσεις δεν θα είναι σημαντικές. Με τον ίδιο τρόπο, η αρχή αυτή οδηγεί σε διαδικασία AR υψηλότερης τάξης. Τέλος, εάν υπάρχουν  $p$  σημαντικές μερικές αυτοσυσχετίσεις, τότε η τάξη είναι AR( $p$ ).

Για την επίτευξη της ταυτοποίησης του μοντέλου, εξετάζουμε τις μερικές αυτοσυσχετίσεις για να ορίσουμε την τάξη της διαδικασίας. Σημειώνεται, ότι η τάξη είναι ίση με τον αριθμό των στατιστικά σημαντικών μερικών αυτοσυσχετίσεων.

Η ανάλυση μερικής αυτοσυσχέτισης δεν λέει τίποτα για τη διαδικασία του κινητού μέσου (MA), που πρόκειται να συζητηθεί στο επόμενο κεφάλαιο. Οι μερικές αυτοσυσχετίσεις είναι κατασκευασμένες για να ταυριάζουν μόνο στη διαδικασία AR. Επίσης, εισάγουν μια εξάρτηση από τη μια υστέρηση στην επόμενη, που τις κάνει να συμπεριφέρονται με έναν τρόπο παρόμοιο με εκείνον των αυτοσυσχετίσεων στη διαδικασία AR. Ειδικότερα, οι μερικές αυτοσυσχετίσεις συγκλίνουν εκθετικά στο μηδέν. Όταν οι μερικές αυτοσυσχετίσεις δεν παίρνουν τυχαίες τιμές μετά από  $p$  χρονικές υστερήσεις συγκλίνουν εκθετικά στο μηδέν, και η πραγματική διαδικασία που δημιουργεί τη σειρά συμπεραίνεται ότι είναι η MA.

Τέλος, μια διαδικασία  $AR(p)$  έχει μόνο  $p$  μερικές αυτοσυσχετίσεις που είναι στατιστικά διάφορες του μηδενός. Επίσης, όταν οι μερικές αυτοσυσχετίσεις πέφτουν εκθετικά στο μηδέν, η διαδικασία λέγεται ότι είναι MA.

### 3.1.11 Ανάλυση Αυτοσυσχέτισης

Ανακεφαλαιώνοντας τα στάδια μελέτης της αυτοσυσχέτισης χρονοσειρών, μπορούμε να πούμε:

1. Για να υπολογίσουμε και να παραστήσουμε γραφικά τους συντελεστές αυτοσυσχέτισης μιας χρονοσειράς, εξετάζουμε εάν τείνει στο μηδέν μετά τη δεύτερη ή τρίτη υστέρηση. Αν πράγματι συμβαίνει έτσι, η χρονοσειρά ονομάζεται σταθερή στην αρχική της μορφή. Στη συνέχεια αναζητούμε την ύπαρξη κάποιου προτύπου. Αν η αυτοσυσχέτιση πλησιάζει το μηδέν (και παραμένει θετική), η σειρά είναι πιθανώς σταθερή, επομένως προχωρούμε στο τρίτο στάδιο. Αν η σειρά δεν είναι σταθερή, προχωρούμε στο στάδιο 2.
2. Για τις μη σταθερές χρονοσειρές, παίρνουμε τις πρώτες διαφορές και υπολογίζουμε τους συντελεστές αυτοσυσχέτισης. Εάν η νέα σειρά είναι σταθερή, προχωρούμε στο στάδιο 3. Αν η νέα σειρά είναι ακόμα μη σταθερή, παίρνουμε πάλι τις πρώτες διαφορές και υπολογίζουμε τις αυτοσυσχετίσεις. Συνήθως δεν χρειάζονται περισσότερες από δύο διαφορίσεις για το μετασχηματισμό των δεδομένων σε σταθερά. (Μια εναλλακτική διαδικασία για να επιτευχθεί σταθερότητα, είναι αυτή που περιγράφουμε στο Παράρτημα Γ' αυτού του Κεφαλαίου).
3. Για τις σταθερές χρονοσειρές μετρούμε και εξετάζουμε τις αυτοσυσχετίσεις τους. Υποδηλώνεται ένα τριμηνιαίο εποχικό πρότυπο εάν οι χρονικές υστερήσεις 4,8,12, κ.λ.π. είναι σημαντικές. Όμοια, εάν το  $r_k$  για χρονικές υστερήσεις 12,24,36 κ.λπ., είναι σημαντικό, τότε υπάρχει ένα μηνιαίο εποχικό πρότυπο. Τέλος, αν δεν υπάρχουν στατιστικά σημαντικές αυτοσυσχετίσεις, καταλήγουμε ότι τα δεδομένα είναι τυχαία.

**3.2 ΠΡΟΣΘΕΤΑ ΜΟΝΤΕΛΑ ΧΡΟΝΟΣΕΙΡΩΝ [16],[17]****3.2.1. Εισαγωγή**

Οι ιδιότητες των χρονοσειρών που εξετάσαμε στην τελευταία παράγραφο, ήταν: (α) η τυχαιότητα, (β) η σταθερότητα, (γ) εάν υπάρχει μη σταθερότητα, σε ποιο επίπεδο διαφορισμού θα γίνει σταθερή η χρονοσειρά, (δ) η εποχιακότητα και (ε) εάν η σειρά είναι εποχική, ποιο είναι το μήκος της εποχιακότητας. Η γνώση αυτών των ιδιοτήτων θα μας δώσει τη δυνατότητα επίτευξης καλύτερων προβλέψεων.

Σε αυτήν την παράγραφο, θα μελετήσουμε προσεγγίσεις στην πρόβλεψη χρονοσειρών, όπου η γνώση των ιδιοτήτων των δεδομένων της χρονοσειράς εξυπηρετεί καλύτερες προβλέψεις. Η παράγραφος 3.2 επικεντρώνεται στην εξέταση πρόσθετων μοντέλων χρονοσειρών και στη μελέτη των μεθόδων του αναπροσαρμοζόμενου Φιλτραρίσματος (adaptive filtering). Αυτές οι τελευταίες μέθοδοι πρωτοσυστήθηκαν με λεπτομέρειες από τους Μακριδάκη και Wheelwright σε μια σειρά άρθρων (1973a, 1973b, 1977a) και επαναπαρουσιάστηκαν στα βιβλία τους (Makridakis and Wheelwright, 1977b, Makridakis and Wheelwright, 1978). Αργότερα, οι Kahl και Ledolter παρουσίασαν αποδείξεις των περιορισμών της μεθόδου (1982, 1983). Η λογική που διέπει τη μέθοδο adaptive filtering στηρίζεται στο γεγονός ότι μια πρόβλεψη μπορεί να προσαρμοστεί στο πρότυπο των δεδομένων που μελετώνται.

Η βάση της μεθόδου είναι ότι οι σταθμικοί όροι που χρησιμοποιούνται στη συγκεκριμένη τεχνική αλλάζουν όταν αποκτηθούν νέες παρατηρήσεις. Αυτή η μέθοδος είναι παρόμοια με εκείνες που περιγράψαμε στην παράγραφο 1.1, δηλαδή, με την εκθετική εξομάλυνση, όπου οι σταθμίσεις είναι σταθερές. Ο βαθμός που οι συντελεστές στάθμισης στην εκθετική εξομάλυνση ελέγχονται ή μεταβάλλονται, εξαρτάται ολοκληρωτικά από την κρίση αυτού που κάνει τις προβλέψεις. Αντίθετα, η μέθοδος του αναπροσαρμοζόμενου φιλτραρίσματος συμπεριλαμβάνει μια εξομαλυντική διαδικασία που μετατρέπει τα σταθμικά, ώστε να ταιριάζουν με μεγαλύτερη ακρίβεια στη μεταβαλλόμενη Φύση της χρονοσειράς.

Σαν πρώτο βήμα θα ασχοληθούμε με αυτή την τεχνική, όπως εφαρμόζεται στις αυτοπαλινδρομες διαδικασίες.

**3.2.2 Αυτοπαλινδρομικά Μοντέλα**

Στην παράγραφο 3.1, ορίσαμε τα αυτοπαλινδρομικά (AR) ως εξής:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t \quad (3.30)$$

Η εξίσωση (3.30) διαφέρει από τα γενικά μοντέλα γραμμικής παλινδρόμησης στο ότι οι ανεξάρτητες

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

μεταβλητές είναι ιστορικές τιμές της εξαρτημένης μεταβλητής και στο ότι δεν υπάρχει σταθερά στην εξίσωση. Η εξάλειψη του σταθερού όρου επιτυγχάνεται με έναν από τους ακόλουθους δύο τρόπους:

1. Τα  $Y_t$  εκφράζονται ως αποκλίσεις από το μέσο, δηλαδή:

$$Y_t^* = Y_t - \bar{Y}$$

2. Επιτρέποντας στους συντελεστές να προσαρμοστούν ως προς τον σταθερό όρο.

Το γενικό μοντέλο AR(p) της εξίσωσης (3.30) μπορεί να λάβει αρκετές μορφές ανάλογα με την τάξη του  $p$ . Ειδικότερα, αν  $p=1$ , η τάξη του μοντέλου AR είναι ένα και σημειώνεται ως AR(1). Πριν να υπολογιστεί ένα μοντέλο AR, πρέπει να καθοριστεί η τάξη  $p$ . Όπως σημειώσαμε στην παράγραφο 3.1, η κατάλληλη τιμή για το  $p$ , που προσδιορίζει τον αριθμό των όρων που θα συμπεριληφθούν, καθορίζεται από την εξέταση των συντελεστών αυτοσυσχέτισης. Η φύση ενός μοντέλου AR(p) γίνεται καλύτερα κατανοητή παρατηρώντας τη μορφή του, ως ακολούθως:

$$Y_t = \phi_1 Y_{t-1} + e_t \quad (3.31)$$

Αν

$$Y_{t-1} = \phi_1 Y_{t-2} + e_{t-1} \quad (3.32)$$

αντικαθίσταται στην εξίσωση (8.2), οπότε

$$Y_t = \phi_1^2 Y_{t-2} + \phi_1 e_{t-1} + e_t \quad (3.33)$$

Με όμοιο τρόπο, όταν

$$Y_{t-2} = \phi_1 Y_{t-3} + e_{t-2} \quad (3.34)$$

αντικαθίσταται στην εξίσωση (8.4), και γίνεται:

$$Y_t = \phi_1^3 Y_{t-3} + \phi_1^2 e_{t-2} + \phi_1 e_{t-1} + e_t, \quad (3.35)$$

Συνεχίζοντας την αντικατάσταση των  $Y_{t-l}$ , η πρώτη παρατήρηση εκφράζεται ως:

$$Y_t = \phi_1^{n-1} Y_{t-n+1} + \phi_1^{n-2} Y_{t-n+2} + \dots + \phi_1^2 e_{t-2} + \phi_1 e_{t-1} + e_t \quad (3.36)$$

Η έκφραση (3.36) δείχνει ότι ένα μοντέλο AR(1) σταθμίζει σφάλματα του παρελθόντος με έναν εκθετικά φθίνοντα ρυθμό. Επομένως, αυτό το μοντέλο AR(1) είναι παρόμοιο με τη μέθοδο της απλής εκθετικής εξομάλυνσης για τη στάθμιση ιστορικών τιμών μιας χρονοσειράς με έναν εκθετικά φθίνοντα τρόπο.

Η έκφραση για το μοντέλο AR(2) είναι η εξής:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t \quad (3.37)$$

Η εξίσωση (3.37) μπορεί, ακόμη, να εκφραστεί σε όρους παρελθόντων σφαλμάτων και η στάθμιση θα είναι όμοια με την (3.36). Η νέα έκφραση, εν τούτοις, εμπεριέχει γινόμενα σφαλμάτων διαδοχικών περιόδων και εκφράζεται ως εξής:

$$Y_t = (\phi_1^2 + \phi_2) \phi_1 Y_{t-3} + (\phi_1^2 + \phi_2) \phi_2 Y_{t-4} + \phi_1^2 \phi_2 e_{t-2} + \phi_1 e_{t-1} + e_t \quad (3.38)$$

Επομένως, το γενικό μοντέλο AR(p) μας εφοδιάζει με μια ευφυή μέθοδο πρόβλεψης χρονοσειρών για το χειρισμό όλων σχεδόν των ειδών των δεδομένων, απλώς, μέσω του προσδιορισμού της τάξης p, του μοντέλου.

Για να περιγράψουμε τη χρήση των αυτοπαλίνδρομων διαδικασιών, θεωρούμε την ακόλουθη εφαρμογή ενός μοντέλου AR(2).

#### Παράδειγμα 3.6

Ο Πίνακας 3.6 περιέχει δεδομένα που αφορούν το καθαρό κέρδος ιδιωτικών εγχώριων επενδύσεων παραγωγών διαρκούς εξοπλισμού (\$ δις) στις ΗΠΑ για τα έτη από 1960 έως 1981 ( $Y_t$ ). Η δεύτερη και η τρίτη στήλη περιέχουν τα ίδια δεδομένα με υστέρηση μιας και δύο περιόδων ( $Y_{t-1}$  και  $Y_{t-2}$ ). Το ακόλουθο μοντέλο, ARIMA(2, 0, 0), εκτιμάται με τις μεθόδους πολλαπλής παλινδρόμησης.

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t$$

Αυτό είναι ένα μοντέλο AR(2) που δεν έχει σταθερό όρο και αποσκοπεί στην εκτίμηση των παραμέτρων  $\phi_1$  και  $\phi_2$ .

**ΠΙΝΑΚΑΣ 3.6** Καθαρό κέρδος ιδιωτικών εγχώριων επενδύσεων για έναν παραγωγό μη-οικιακού διαρκούς εξοπλισμούς (\$ δις)

Ετη	Περίοδος	(1) Παρατηρηθείσα τιμή του $Y_t$	(2) Χρον. Υστερ. 1 χρόνου $Y_{t-1}$	(3) Χρον. υστέρ. 2 ετών $Y_{t-2}$	(4) Προσαρμ. τιμές $Y$	(5) Κατάλοιπα $e_t$
1960	1	5.0	-	-	-	-
1961	2	3.6	5.0	-	-	-
1962	3	6.1	3.6	5.0	3.0	3.1
1963	4	7.5	6.1	3.6	6.8	0.7
1964	5	10.8	7.5	6.1	7.9	3.7
1965	6	16.0	10.6	7.5	11.3	4.7
1966	7	20.6	16.0	10.6	17.3	3.3
1967	8	18.2	20.6	16.0	21.5	-3.3
1968	9	19.1	18.2	20.6	16.8	2.3
1969	10	21.5	19.1	18.2	18.8	2.7
1970	11	17.6	21.5	19.1	21.6	-4.0
1971	12	15.5	17.6	21.5	15.7	-0.2
1972	13	20.4	15.5	17.6	14.3	6.1
1973	14	31.2	20.4	15.5	21.4	9.8
1974	15	30.3	31.2	20.4	33.9	-3.6
1975	16	17.3	30.3	31.2	29.0	-11.7
1976	17	20.2	17.3	30.3	12.3	7.9
1977	18	34.0	20.2	17.3	20.5	13.5
1978	19	44.8	34.0	20.2	37.6	7.2
1979	20	48.9	44.8	34.0	47.0	1.9
1980	21	34.2	48.9	44.8	48.7	-14.5
1981	22	28.9	34.2	48.9	28.0	0.9

ΠΗΓΗ: Υπουργείο Εμπορίου των ΗΠΑ. *Survey of Current Business*, διάφορα τεύχη.

Για τα δεδομένα του Πίνακα 3.6 το μοντέλο AR(2) εκτιμάται ότι είναι

$$\hat{Y}_t = 1.31Y_{t-1} - 0.343Y_{t-2}$$

με  $Se=7.008$  και  $MSE=49.12$ .

Ο πίνακας 3.6 περιέχει τα δεδομένα εισροών και εκροών. Οι στήλες 4 και 5 δίνουν τις προσαρμοσμένες τιμές των καταλοίπων της παραπάνω αυτοπαλινδρόμησης.

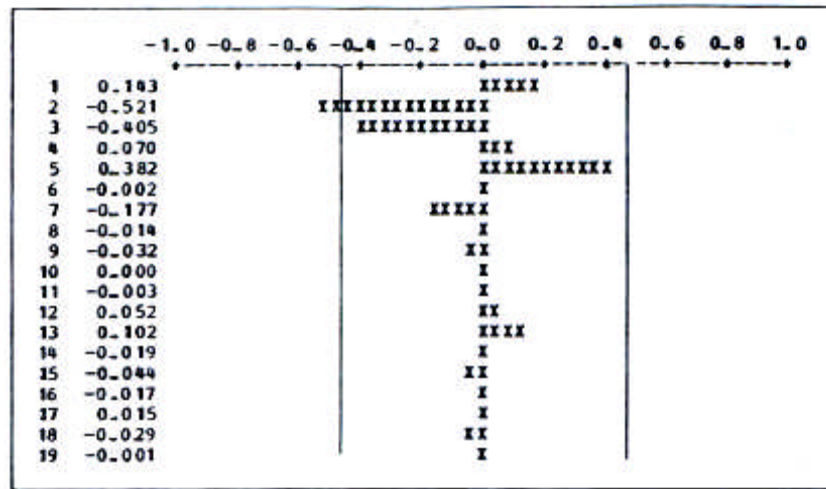
Για να ελέγξουμε την τυχαιότητα των καταλοίπων, υπολογίζουμε τις αυτοσυσχετίσεις και τις παριστάνουμε (με υπολογιστή) στο σχήμα 8.2. Σημειώνοντας το περίγραμμα του γραφήματος των συντελεστών αυτοσυσχετίσης, παίρνουμε την ένδειξη εάν τα κατάλοιπα είναι τυχαία ή όχι. Βασισμένοι στο γράφημα μπορούμε να συμπεράνουμε ότι μόνο μια υστέρηση 2 βρίσκεται εκτός των παραδεκτών ορίων των δύο σίγμα  $\pm 0.459^1$ . Παραπέρα, το MSE για το μοντέλο είναι 49.12. Ένας λόγος για την εκτίμηση του μοντέλου είναι η επίτευξη ενός προβλεπτικού μοντέλου που να έχει το μικρότερο MSE για ένα δοσμένο σύνολο δεδομένων.

Η προσέγγιση αυτή μας παρέχει επίσης αρκετή πληροφόρηση για να δημιουργήσουμε το καλύτερο μοντέλο ARIMA( $p,d,q$ ) (= autoregressive - integrated-moving average: ενοποιημένο

### ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

μοντέλο αυτοπαλινδρόμησης και κινητού μέσου όρου).

**Σχημα 3.6.** Συντελεστές αυτοσυσχέτισης των καταλοίπων από μοντέλο αυτοπαλινδρόμησης (αποτελέσματα σύμφωνα με το πρόγραμμα Minitab)



### 3.2.3. Εφαρμογή της Μεθόδου Adaptive Filtering

Η μέθοδος adaptive filtering εφαρμόζεται σε ένα μοντέλο  $AR(p)$  που έχει τη μορφή της εξίσωσης (3.30) μέσω μιας επαναλαμβανόμενης διαδικασίας. Η επαναληπτική μέθοδος «δοκιμή - και - σφάλμα», χρησιμοποιείται για τον καθορισμό των «άριστων» ή «καλύτερων» σταθμίσεων. Σαν πρώτο βήμα, χρησιμοποιούμε ένα δεδομένο σύνολο συντελεστών στάθμισης για να υπολογίσουμε μια πρόβλεψη. Ακολούθως, καθορίζονται τα σφάλματα ή κατάλοιπα από τις πραγματικές τιμές. Τα σφάλματα αυτά χρησιμοποιούνται για την προσαρμογή των συντελεστών στάθμισης έτσι ώστε να μειωθεί το σφάλμα πρόβλεψης έως ένα ελάχιστο επίπεδο. Αυτή η προσαρμοστική διαδικασία ονομάζεται αναπροσαρμοζόμενο φιλτράρισμα. Τον ορισμό αυτό τον συναντούμε στη θεωρία της τηλεπικοινωνίας όπου χρησιμοποιείται για να περιγράψει μια διαδικασία φιλτραρίσματος του θορύβου της μεταφοράς ενός μηνύματος. Αυτή η ίδια έννοια εφαρμόζεται στη δική μας προσπάθεια να διαχωρίσουμε τις τυχαίες διακυμάνσεις (white noise) από το πραγματικό πρότυπο στις τιμές μιας χρονοσειράς.

Αρκετά από τα χαρακτηριστικά της μεθόδου adaptive filtering την καθιστούν κατάλληλη για τους σκοπούς της πρόβλεψης. Πρώτον, οι πληροφορίες που περιέχουν οι παρελθούσες παρατηρήσεις επικαιροποιούνται συνεχώς και προσαρμόζονται αυτόματα στη μεταβολή του προτύπου της χρονοσειράς.

Η μέθοδος προσαρμογής των τιμών των συντελεστών στάθμισης χρησιμοποιεί μόνο την απλή (single) εξίσωση:

$$\phi'_i = \phi_i + 2ke_i^* Y_{t-i} \quad (3.39)$$

όπου  $\phi'_i$  είναι οι αναθεωρημένες σταθμίσεις·  $\phi_i$  συμβολίζει τις αρχικές σταθμίσεις  $k$  είναι η σταθερά μάθησης,  $e_i$  είναι το τυποποιημένο σφάλμα πρόβλεψης·  $Y_i$  είναι οι παρατηρηθείσες τιμές  $i = 1, 2, \dots, p$  και  $t = p+1, p+2, \dots, m$ . Η εξίσωση (3.39) δείχνει ότι οι αναθεωρημένοι συντελεστές στάθμισης θα πρέπει να είναι ίσοι με τις αρχικές σταθμίσεις, πλέον μια προσαρμογή για το σφάλμα πρόβλεψης.

Η προσαρμογή γίνεται για κάθε στάθμιση και βασίζεται στο σφάλμα πρόβλεψης, την παρατηρηθείσα τιμή και τη σταθερά μάθησης.

Η διαδικασία της προσαρμογής (adapting) ή δοκιμασίας των σταθμίσεων αποτελείται από μια σειρά σταδίων, που παρουσιάζονται στον Πίνακα 3.7.



**ΠΙΝΑΚΑΣ 3.7** Στάδια της adapting filtering για μια αυτοπαλίνδρομη διαδικασία

Αριθμός σταδίων	Περιγραφή
1	Καθορίζουμε τον αριθμό των σταθμίσεων με την εύρεση του βαθμού της διαδικασίας AR.
2	Υπολογίζουμε τις αρχικές σταθμίσεις.
3	Καθορίζουμε τη σταθερά αρχικής φάσης (μάθησης) σταθεροποιώντας τη χρονοσειρά και το σφάλμα πρόβλεψης.
4	Χρησιμοποιούμε τη σταθερά μάθησης για τον υπολογισμό νέων σταθμίσεων.
5	Επαναλαμβάνουμε τα στάδια 3 και 4 μέχρις ότου χρησιμοποιηθούν όλες οι παρατηρήσεις ή μέχρι να μην μπορούμε να μειώσουμε παραπέρα το MSE (μέσο σφάλμα τετραγώνου).

Εάν επανέλθουμε στο παράδειγμα 3.6, όπου μελετήσαμε τα δεδομένα που αφορούν καθαρά κέρδη ιδιωτικών εγχώριων επενδύσεων ενός παραγωγού μη-οικιακού διαρκούς εξοπλισμού, μπορούμε να αρχίσουμε να εφαρμόζουμε τη διαδικασία της γενικευμένης μεθόδου adaptive filtering (generalized adaptive filtering: GAF) στην αυτοπαλίνδρομη διαδικασία που εξετάσαμε προηγουμένως. Τα στάδια είναι τα ακόλουθα:

**Στάδιο 1:** Προηγουμένως, εξετάσαμε την αυτοσυσχέτιση της παραπάνω χρονοσειράς και καταλήξαμε ότι ακολουθούσε το μοντέλο AR(2). Με άλλα λόγια, ο βαθμός του μοντέλου θα πρέπει να είναι ίσος με τη χρονική υστέρηση του υψηλότερου συντελεστή αυτοσυσχέτισης. Επομένως, το

$$\hat{Y}_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t \quad (3.40)$$

είναι το μοντέλο που επιλέχθηκε.

**Στάδιο 2.** Για να υπολογίσουμε τις αρχικές σταθμίσεις,  $\phi_1$ ,  $\phi_2$ , αρχίζουμε καθορίζοντας ίσους σταθμικούς όρους στην εξίσωση (3.40), δηλαδή

$$\phi_1 = \phi_2 = 0,50$$

Μπορούμε τώρα να χρησιμοποιήσουμε την εξίσωση (3.39) για να τροποποιήσουμε αυτές τις σταθμίσεις μέχρι την ελαχιστοποίηση του MSE της πρόβλεψης:

$$\hat{Y}_3 = (0.50)(3.6) + (0.50)(5.0) = 4.3$$

και

$$e_3 = Y_3 - \hat{Y}_3 = 6.1 - 4.3 = 1.80$$

Με παρόμοιους υπολογισμούς μπορούν να ευρεθούν τα σφάλματα για τις περιόδους 4, 5, 6 κ.ο.κ. Το MSE για όλα τα κατάλοιπα (σφάλματα) θα είναι σχετικά υψηλό. Μια εναλλακτική διαδικασία που αναπτύχθηκε από τους Makridakis και Wheelwright (κεφ. 9) είναι η απόκτηση των τιμών των σταθμίσεων από μια πολύ γνωστή σειρά εξισώσεων, τις εξισώσεις Yule-Walker. Τα αποτελέσματα αυτών των διαδικασιών δίνονται από τους παρακάτω τύπους

$$\phi_1 = \frac{r_1(1-r_2)}{1-r_1^2} \quad (3.41)$$

και

$$\phi_2 = \frac{r_2 - r_1^2}{1-r_1^2} \quad (3.42)$$

Τώρα υπολογίζονται οι αυτοσυσχετίσεις και τα αποτελέσματα δείχνονται στο σχήμα 3.7. Με αντικατάσταση των αυτοσυσχετίσεων στις εξισώσεις (3.41) και (3.42), βρίσκουμε:

$$\phi_1 = \frac{(0.788)(1-0.462)}{1-0.788^2} \cong 1.12$$

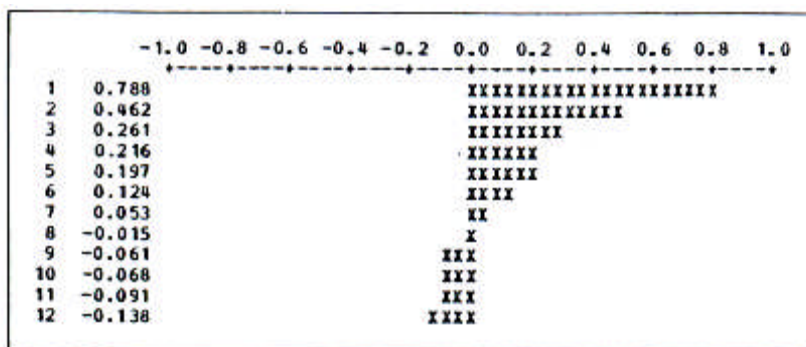
και

$$\phi_2 = \frac{0.462(0.788)^2}{1-0.788^2} \cong -0.42$$

Οι παραπάνω σταθμίσεις θα μπορούσαν να δώσουν ένα μικρότερο MSE, σε σχέση με το σφάλμα που μετρήθηκε από το πρώτο σύνολο σταθμίσεων.

**Στάδιο 3.** Εφόσον έχουμε δυο σταθμίσεις, χρειαζόμαστε μόνο δυο τιμές της χρονοσειράς, προκειμένου να προβλέψουμε την επόμενη τιμή. Εν τούτοις, είναι επιθυμητό να τυποποιήσουμε πρώτα τις χρησιμοποιούμενες τιμές του  $Y_t$ . Η τυποποίηση προσαρμόζει τις τιμές του  $Y_t$  ώστε να κυμαίνονται μεταξύ 0 και 1. Υπάρχουν δυο διαδικασίες διαθέσιμες. Πρώτον, μπορούμε να διαιρέσουμε όλα τα δεδομένα με τη μεγαλύτερη τιμή στη χρονοσειρά, ή, δεύτερον, μπορούμε να διαιρέσουμε κάθε σύνολο των  $p$  τιμών δια της τετραγωνικής ρίζας του αθροίσματος των τετραγώνων των τιμών. Και οι δυο διαδικασίες συγκλίνουν. Με την πρώτη μέθοδο μπορούμε να έχουμε αποτελέσματα όπως αυτά που παρατίθενται στον Πίνακα 8.3. Η στήλη 2 είναι η στήλη 1 διαιρεμένη με τη μεγαλύτερη τιμή (maximum)  $Y_{20} = 48.9$ .

Σχημα 3.7 Συντελεστές αυτοσυσχέτισης των καθαρών κερδών ιδιωτικών εγχώριων επενδύσεων μη οικιακού διαρκούς εξοπλισμού (αποτελέσματα σύμφωνα με το πρόγραμμα Minitab)



ΠΙΝΑΚΑΣ 3.8 Τυποποιημένες τιμές μετά από διαίρεση των τιμών με τη μεγαλύτερη τιμή

Περίοδος	(1) $Y_t$	(2) $Y_t - Y_{tmax}$
1	5.0	0.102
2	3.6	0.074
3	6.1	0.125
4	7.5	0.153
5	10.6	0.217
.	.	.
.	.	.
.	.	.
19	44.8	0.916
20	48.9	1.000
21	34.2	0.699
22	28.9	0.591

Η δεύτερη μέθοδος παρουσιάζεται στον Πίνακα 3.9. Η στήλη 2 είναι το τετράγωνο των εκ παρατηρήσεως τιμών του  $Y_t$ , στήλη I. Η στήλη 3 είναι το πηλίκο που σχετίζεται με τη διαίρεση του  $Y_t$  με την τετραγωνική ρίζα του αθροίσματος των τετραγώνων των τιμών του  $Y_t$  για  $p = 2$  περιόδους.

**ΠΙΝΑΚΑΣ 3.9** Μέθοδος τυποποίησης με διαίρεση κάθε ομάδας τιμών δια της τετραγωνικής ρίζας του αθροίσματος των τετραγώνων των τιμών

Περίοδος	(1) $Y_t$	(2) $Y_t^2$	(3) $Y_t(\sum Y_t^2)^{1/2}$
1	5.0	25.0	0.812
2	3.6	12.96	0.584
(α) Περιγραφή των περιόδων 1 και 2 $(\sum Y_t^2)^{1/2}=6.161$			
2	3.6	12.96	0.508
3	6.1	37.21	0.861
		50.17	
(β) Περιγραφή των περιόδων 2 και 3 $(\sum Y_t^2)^{1/2}=7.083$			

Για να τελειώνουμε με τον υπολογισμό της σταθεράς μάθησης,  $k$ , αντικαθιστούμε τις τυποποιημένες τιμές  $Y_t$  (για την περιγραφή, προτιμούμε τη χρήση της δεύτερης μεθόδου) στην εξίσωση (3.40) και παίρνουμε:

$$\hat{Y}_t = 1.12Y_{t-1} + (-0.42Y_{t-2})$$

$$\hat{Y}_3 = 1.12(0.812) - 0.42(0.584) = 0.245$$

το τυποποιημένο σφάλμα είναι:

$$e_t^* = \frac{Y_t}{(\sum Y_t^2)^{1/2}} - (0.245) \quad (3.43)$$

Για  $t = 3$

$$e_3^* = \frac{6.1}{6.161} - (0.245)$$

$$= 0.745 \text{ (το τυπικό σφάλμα)}^4$$

Η σταθερά μάθησης,  $k$ , πρέπει τώρα να επιλεγεί. Γενικά, καθώς αυξάνεται το μέγεθος της τυχαιότητας σε μια σειρά και καθώς ο αριθμός των σταθμίσεων μεγαλώνει, μειώνεται η άριστη τιμή για  $k$ . Εν τούτοις, συχνά ο κανόνας είναι να θέσουμε την  $k$  ίση με το αντίστροφο του αριθμού των σταθμίσεων (σε αυτή την περίπτωση: 2). Επομένως,  $k = 1/p = 1/2 = 0.50$ .

**Στάδιο 4.** Το τυποποιημένο σφάλμα 0.745 και η σταθερά μάθησης 0.50 μπορεί τώρα να αντικατασταθεί στην εξίσωση (3.39) ως εξής, για κάθε νέα στάθμιση:

$$\phi'_1 = 1.12 + 2(0.50)(0.745)(0.812) = 0.515$$

και

$$\phi'_2 = -0.42 + 2(0.50)(0.745)(0.812) = -0.015$$

**Στάδιο 5.** Η εξάσκηση με τους συντελεστές στάθμισης μπορεί να επαναληφθεί για τις περιόδους 2 και 3 και η πρόβλεψη μπορεί να συνεχιστεί για την περίοδο 4. Και πάλι βρίσκουμε τις τυποποιημένες τιμές  $Y_t$  από τον Πίνακα 3.9, μέρος (β). Το τυποποιημένο σφάλμα για την περίοδο 4 είναι:

$$\begin{aligned} e_4^* &= \frac{Y_4}{7.083} - \hat{Y}_4 \\ &= \frac{7.5}{7.083} - [0.515(0.861) - (-0.015)(0.508)] \\ &= 1.059 - 0.436 = 0.623 \end{aligned}$$

Το νέο σφάλμα χρησιμοποιείται τώρα για την ανάπτυξη νέων συντελεστών στάθμισης. Ακόμη, η διαδικασία επαναλαμβάνεται μέχρις ότου εξαντληθούν όλες οι παρατηρήσεις στη χρονοσειρά. Έτσι θα ολοκληρωθεί η επαναληπτική διαδικασία εξάσκησης μετά από την οποία μπορούμε να επαναλάβουμε την προσπάθεια από την έναρξη της χρονοσειράς και να συνεχίσουμε τη διαδικασία μέχρις ότου το MSE να μην μπορεί να μειωθεί πέρα από ένα αυθαίρετα καθορισμένο μικρό μέγεθος.

Είναι εμφανές ότι για να χρησιμοποιηθεί η μέθοδος του αναπροσαρμοζόμενου φιλτραρίσματος ως χρήσιμο εργαλείο πρόβλεψης, πρέπει να γίνει χρήση προγραμμάτων λογισμικού. Οι διαδικασίες είναι κατά πολύ ισχυρότερες από τις μεθόδους της εκθετικής εξομάλυνσης, αλλά αρκετά πιο πολύπλοκες. Η χρησιμοποίηση της adaptive filtering στο συζητούμενο πρόβλημα, μας δίνει ένα μέσο σφάλμα τετραγώνου (MSE) ίσο με 1.654. Μετά από τρεις επαναλήψεις του κύκλου εξάσκησης το MSE έχει μειωθεί σε 1.306 και δεν μπορεί να ελαττωθεί περισσότερο.

Η χρήση της μεθόδου adaptive filtering στην προετοιμασία μιας πρόβλεψης, αποτελείται από δύο διακεκριμένες φάσεις. Πρώτον, εφαρμόζουμε τη διαδικασία των πέντε σταδίων για τον προσδιορισμό των κατάλληλων σταθμίσεων με τα ιστορικά δεδομένα της χρονοσειράς. Δεύτερον, χρησιμοποιούμε τις τελικές σταθμίσεις για την πρόβλεψη της χρονοσειράς στο μέλλον. Για να εξετάσουμε το δεύτερο επίπεδο της διαδικασίας, θεωρούμε το επόμενο παράδειγμα.

### Παράδειγμα 3.7

Μια χρονοσειρά με σταθερό συντελεστή αύξησης για διάστημα έξι περιόδων δίνεται στον Πίνακα 3.10

Θα χρησιμοποιηθεί ένα μοντέλο AR(2) και θα αναθεωρηθεί με τη διαδικασία της adaptive filtering. Βρίσκουμε ότι οι σταθμίσεις είναι  $\phi_1 = 2$  και  $\phi_2 = -1$  μετά από όλες τις επαναλήψεις. Επομένως, η πρόβλεψη για την περίοδο 7 δίνεται από τον τύπο:

$$\hat{Y}_7 = \phi_1 Y_6 + \phi_2 Y_6$$

**ΠΙΝΑΚΑΣ 3.10** Χρονοσειρά με σταθερό συντελεστή αύξησης για διάστημα έξι περιόδων

Περίοδος	Παρατηρηθείσες τιμές	Αποκλίσεις Y
- 1	3	3
2	6	3
3	9	3
4	12	3
5	15	3
6	18	3

Υπενθυμίζουμε ότι οι τιμές Y είναι σε αποκλίσεις από τους μέσους. Με αντικατάσταση, η πρόβλεψη γίνεται:

$$\hat{Y}_7 = 2(3) - 1(3)$$

Πρέπει να προσθέσουμε την παρατηρηθείσα τιμή στην περίοδο 6, η οποία είναι 18, και έτσι παίρνουμε 21. Επομένως, η πρόβλεψη είναι πλήρης.

Σημειώστε ότι το παραπάνω αποτέλεσμα είναι ταυτόσημο με το αποτέλεσμα που θα είχαμε, αν στην πρόβλεψη χρησιμοποιούσαμε ένα γενικό γραμμικό παλινδρομικό μοντέλο της μορφής:

$$Y = \beta_0 + \beta_1 X$$

Παραπέρα, η παλινδρόμηση είναι πιο αποτελεσματική και λιγότερο δαπανηρή. Είναι φανερό πως το παράδειγμα 3.7 παρατέθηκε ώστε να περιγραφεί καλύτερα η διαδικασία.

Η μέθοδος της adaptive filtering θα πρέπει να εφαρμόζεται μόνον όταν οι απλούστερες και λιγότερο δαπανηρές μέθοδοι δεν δίδουν καλές προβλέψεις. Σε επόμενη παράγραφο αυτού του κεφαλαίου, θα επιστρέψουμε στα πλεονεκτήματα αυτών των μεθόδων.

### 3.2.4. Μοντέλα Κινητών Μέσων

Τα μοντέλα AR(p) δεν μπορούν να απομονώσουν ορισμένα πρότυπα δεδομένων, όταν το p είναι πολύ μικρό. Ένα άλλο μοντέλο, αυτό των κινητών μέσων (MA), ίσως απομονώσει το πρότυπο στις περιπτώσεις που τα μοντέλα AR(p) αποτύχουν. Ο Wold (1954) υπέδειξε ότι οποιαδήποτε διακεκριμένη (discrete) χρονοσειρά μπορεί να εκφραστεί σαν ένα μοντέλο AR ή σαν ένα μοντέλο MA ή σαν ο συνδυασμός και των δυο, οπότε ονομάζεται μοντέλο ARMA. Πρόσφατες εξελίξεις και στην τεχνολογία των υπολογιστών και του λογισμικού (software), επέτρεψαν στους ασχολούμενους με τις προβλέψεις και τους αναλυτές να χρησιμοποιούν τα μοντέλα ARMA με αποτελεσματικό τρόπο.

Τα μοντέλα MA δίνουν προβλέψεις του  $Y_t$  βασισμένες σ' ένα γραμμικό συνδυασμό παρελθόντων σφαλμάτων πρόβλεψης. Αντίθετα, τα μοντέλα AR εκφράζουν το  $Y_t$  σαν γραμμική συνάρτηση των  $p$  παρελθουσών πραγματικών τιμών του  $Y_t$ . Το γενικό μοντέλο MA είναι:

$$Y_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (3.44)$$

Εάν  $\mu = 0$ , τότε έχουμε

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (3.45)$$

Η παραπάνω εξίσωση εκφράζεται χωρίς ένας σταθερός όρος να είναι συνεπής με τον προηγούμενο ορισμό του μοντέλου AR (εξίσωση 8.1). Σημειώστε ότι ο όρος «κινητός μέσος» που χρησιμοποιείται στο παρόν πλαίσιο, δεν έχει καμία σχέση με τους κινητούς μέσους που εξετάσαμε στο Κεφάλαιο 2, ή εκείνους που χρησιμοποιήθηκαν στο Κεφάλαιο 3 κατά την ανάλυση χρονοσειρών. Επίσης, το άθροισμα  $\theta_1 + \theta_2 + \dots + \theta_q$  δεν είναι απαραίτητως ίσο με ένα, ούτε αποτελούν τις τιμές του  $\theta_i$  που αλλάζουν ή «κινούνται», καθώς παίρνονται νέες παρατηρήσεις. Αν και μπορεί να δημιουργηθεί σύγχυση στη χρήση του όρου, θα συνεχίσουμε να τον χρησιμοποιούμε σε αυτό το κεφάλαιο προκειμένου να είμαστε συνεπείς με τη βιβλιογραφία των χρονοσειρών.

Το γράμμα  $q$  στο  $MA(q)$  αναφέρεται στην τάξη του μοντέλου. Για παράδειγμα, ένα μοντέλο  $MA(1)$  είναι

$$Y_t = e_t - \theta_1 e_{t-1}, \text{ ARIMA}(0,0,1) \quad (3.46)$$

Σημειώστε ότι το  $Y_t$  είναι ένας γραμμικός συνδυασμός των τρεχόντων και παρελθόντων ασυσχέτιστων (white noise) καταλοίπων.

Ένα μοντέλο  $MA(2)$  (χωρίς σταθερό όρο) γράφεται ως εξής:

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}, \text{ ARIMA}(0,0,2) \quad (3.47)$$

Θα πρέπει να σημειώσουμε ότι αν και χρησιμοποιούμε τον όρο κινητός μέσος, η μέθοδος είναι όμοια με τα μοντέλα εκθετικής εξομάλυνσης της παραγράφου 1.1 και όχι με τους κινητούς μέσους του ίδιου κεφαλαίου.

Στην παράγραφο 3.1, εξετάσαμε την ιδιότητα της *σταθερότητας* ενός μοντέλου χρονοσειράς. Στη διαδικασία μοντελοποίησης MA απαιτείται μια άλλη ιδιότητα, που ονομάζεται *αναστρεψιμότητα*, για την περιγραφή και την πρόβλεψη μιας χρονοσειράς. Αν και δεν πρόκειται να αναπτύξουμε τον αυστηρό ορισμό της αναστρεψιμότητας, θα δείξουμε εκείνες τις συνθήκες που αφορούν τη σταθερότητα και την αναστρεψιμότητα ενός δεδομένου μοντέλου χρονοσειράς, που μπορούν να εκφραστούν σε όρους των παραμέτρων του μοντέλου.

Για να αρχίσουμε, θεωρούμε το μοντέλο  $ARIMA(0,0,1)$

$$Y_t = \mu + e_t - \theta_1 e_{t-1} \quad (3.48)$$

ή

$$Y_t = e_t - \theta_1 e_{t-1} \quad \text{όταν } \mu = 0 \quad (3.46)$$

Μπορούμε να δείξουμε ότι για οποιαδήποτε τιμή της παραμέτρου  $\theta_1$ , το μοντέλο ARIMA (0,0,1) περιγράφει τη συμπεριφορά μιας σταθερής (stationary) χρονοσειράς. Επομένως, δεν υπάρχουν συνθήκες που πρέπει να εφαρμοστούν στο  $\theta_1$  για να καταστήσουν σταθερό το μοντέλο που περιγράφεται από την εξίσωση (3.46). Εν τούτοις, ορισμένοι περιορισμοί απαιτούνται για τις παραμέτρους των μοντέλων MA προκειμένου να διασφαλιστεί η αναστρεψιμότητα. Η αναστρεψιμότητα αναφέρεται στην πιθανότητα αντιστροφής ενός μοντέλου MA και της έκφρασής του ως μοντέλο AR άπειρης τάξης. Ένα μοντέλο AR άπειρης τάξης, είναι αυτό που διαθέτει άπειρο αριθμό αυτοπαλίνδρομων συντελεστών. Ένα μοντέλο MA είναι αναστρέψιμο, για παράδειγμα, εάν η απόκλιση ( $Y_t - \mu$ ) δεν εξαρτάται κυριαρχικά από αποκλίσεις που συνέβησαν στο μακρινό παρελθόν. Το γεγονός αυτό αποτελεί εύλογη προϋπόθεση. Για ένα μοντέλο MA(1), η συνθήκη αναστρεψιμότητας είναι η ακόλουθη:

$$-1 < \theta_1 < 1$$

Η γνώση των συνθηκών της αναστρεψιμότητας και της σταθερότητας στις παραμέτρους ενός μοντέλου χρονοσειράς, είναι ύψιστης σημασίας για την εκτίμηση των παραμέτρων του μοντέλου.

Μπορούμε να δείξουμε ότι ο μέσος ενός μοντέλου χρονοσειράς, ARIMA (0,0,1), είναι  $\mu$  και ότι η θεωρητική συνάρτηση μερικής αυτοσυσχέτισης αυτού του μοντέλου φθίνει προς το μηδέν μέσω ενός προτύπου που χαρακτηρίζεται από εκθετική μείωση (decay). Παραπέρα, η θεωρητική συνάρτηση αυτοσυσχέτισης διακόπτεται μετά από χρονική υστέρηση 1. Επίσης, μπορούμε να δείξουμε ότι

$$\rho_k = \begin{cases} \frac{-\theta_1}{1 + \theta_1^2} & \text{για } k=1 \\ 0 & \text{για } k>1 \end{cases} \quad (3.49)$$

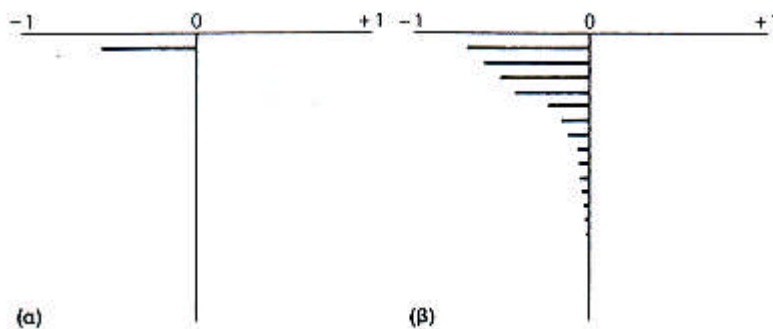
Η γνώση της σχέσης μεταξύ του  $\rho_k$  και των παραμέτρων ενός δεδομένου μοντέλου χρονοσειράς, είναι σημαντική για την εκτίμηση των παραμέτρων του μοντέλου. Θα πρέπει να σημειώσουμε ότι κατά την περιγραφή της συμπεριφοράς των μερικών συσχετίσεων ενός μοντέλου ARIMA (0,0,1), λέμε ότι «κατέρχεται με εκθετικά φθίνοντα τρόπο». Στην παράγραφο 3.3, θα περιγράψουμε τον τρόπο αυτό, και για τις συναρτήσεις αυτοσυσχέτισης και για τις συναρτήσεις μερικής αυτοσυσχέτισης, με λέξεις όπως εκθετικά φθίνων τρόπος ή ημιτονοειδής (ή ημιτονοειδώς) φθίνων τρόπος. Θα πρέπει να σημειώνεται τότε πρόκειται για τις συναρτήσεις αυτοσυσχέτισης και τότε για τις συναρτήσεις μερικής



αυτοσυσχέτισης που φθίνουν ή διακόπτονται. Τα πρότυπα αυτά θα μας επιτρέψουν να περιγράψουμε μια δεδομένη χρονοσειρά με ένα συγκεκριμένο μοντέλο χρονοσειράς.

Σαν παράδειγμα, θεωρούμε ένα μοντέλο MA(1), όπου  $\theta_1 > 0$ . Στο σχήμα 3.8 σημειώνουμε ότι η συνάρτηση αυτοσυσχέτισης (ACF) διακόπτεται μετά από χρονική υστέρηση 1. Αυτή είναι μια χαρακτηριστική ιδιότητα του μοντέλου MA(1) ή ενός ARIMA (0,0,1). Αυτό σημαίνει ότι η μνήμη ενός μοντέλου MA(1) έχει εύρος μιας μόνο χρονικής περιόδου. Με άλλα λόγια, το  $Y_t$  συσχετίζεται με το αμέσως προηγούμενο  $Y_{t-1}$  και με το αμέσως επόμενο  $Y_{t+1}$ , αλλά όχι με οποιουδήποτε άλλους όρους της ακολουθίας των μεταβλητών  $Y_1, Y_2, \dots$ . Η θεωρητική συνάρτηση μερικής αυτοσυσχέτισης (PACF) φθίνει προς το μηδέν, όπως δείχνεται στο σχήμα 3.8. Τα ακριβή πρότυπα της PACF εξαρτώνται από την τιμή της παραμέτρου  $\theta_1$ . Για  $\theta_1 > 0$ , το πρότυπο φθίνει με διαδοχικά μικρότερες αρνητικές μερικές συσχετίσεις.

**Σχήμα 3.8.** (α) ACF για μοντέλο MA(1) με  $\theta_1 > 0$ ,  
(β) PACF για μοντέλο MA(1) με  $\theta_1 < 0$



Τέλος, το μοντέλο MA(1) είναι μια απλή και χρήσιμη χρονοσειρά αν και δεν χρησιμοποιείται τόσο ευρέως, όπως τα μοντέλα AR(1) και AR(2). Το MA(1) χαρακτηρίζεται από το γεγονός ότι η ACF που αντιστοιχεί σ' αυτό, διακόπτεται μετά από χρονική υστέρηση 1 και η PACF του φθίνει προς το μηδέν.

### 3.2.5 Αυτοπαλινδρομικά Μοντέλα Κινητού Μέσου

Τα πρότυπα αυτοσυσχέτισης ίσως απαιτούν πιο πολύπλοκα μοντέλα. Ένα πιο γενικό μοντέλο είναι ένα μίγμα των AR(p) και MA(q) και ονομάζεται αυτοπαλινδρομο μοντέλο κινητού μέσου (ARMA):

$$Y_t = \mu + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (3.50)$$

Αν  $\mu = 0$ , τότε έχουμε

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

$$Y_t = \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} \quad (3.51)$$

Όπως το μοντέλο  $AR(p)$ , το μοντέλο  $ARMA(p,q)$  έχει αυτοσυσχετίσεις που εξαλείφονται, καθώς αυξάνεται η απόσταση μεταξύ των καταλοίπων. Εν τούτοις, τα πρότυπα στη χρονοσειρά που μπορούν να περιγραφούν με μια διαδικασία  $ARMA(p,q)$ , είναι πιο γενικά από τα μοντέλα  $AR(p)$  ή τα  $MA(q)$ .

Ένα μοντέλο  $ARMA(1,1)$  είναι

$$Y_t = \mu + \varphi_1 Y_{t-1} + e_t - \theta_1 e_{t-1}, \text{ ARIMA}(1,0,1) \quad (3.52)$$

και ένα  $ARMA(2,1)$  είναι

$$Y_t = \mu + \varphi_1 Y_{t-2} + \varphi_2 Y_{t-2} + e_t - \theta_1 e_{t-1}, \text{ ARIMA}(2,0,1) \quad (3.53)$$

Μπορεί να δειχθεί ότι το μικτό μοντέλο  $ARMA(1,1)$  είναι σταθερό, αν

$$-1 < \varphi_1 < 1$$

και είναι αναστρέψιμο, αν

$$-1 < \theta_1 < 1$$

Η σταθερά  $\mu$  του μοντέλου είναι  $\mu/(1-\varphi_1)$ .

Η θεωρητική συνάρτηση αυτοσυσχέτισης που αντιστοιχεί σ' αυτό και η συνάρτηση μερικής αυτοσυσχέτισης φθίνει προς το μηδέν με εκθετικά φθίνοντα τρόπο. Συγκεκριμένα, μπορούμε να δείξουμε ότι:

$$\rho_1 = \frac{(1 - \varphi_1 \theta_1)(\varphi_1 - \theta_1)}{1 - \theta_1^2 - 2\theta_1 \varphi_1} \quad (3.54)$$

$$\rho_2 = \varphi_1 \rho_1$$

$$\rho_k = \varphi_1 \rho_{k-1} \text{ για } k \geq 3$$

Οι σχέσεις αυτές χρησιμοποιούνται για να πάρουμε τις παραμέτρους  $\varphi_1$  και  $\theta_1$  σε όρους  $\rho_1$  και  $\rho_2$  που υπολογίζονται από τα  $r_1$  και  $r_2$ . Επομένως, αυτές οι εκτιμήσεις δίνουν ένα σύνολο προκαταρκτικών εκτιμήσεων  $\hat{\varphi}$  και  $\hat{\theta}_1$  που ικανοποιούν τις συνθήκες σταθερότητας και αναστρεψιμότητας.

$$-1 < \hat{\varphi} < 1$$

$$-1 < \hat{\theta}_1 < 1$$

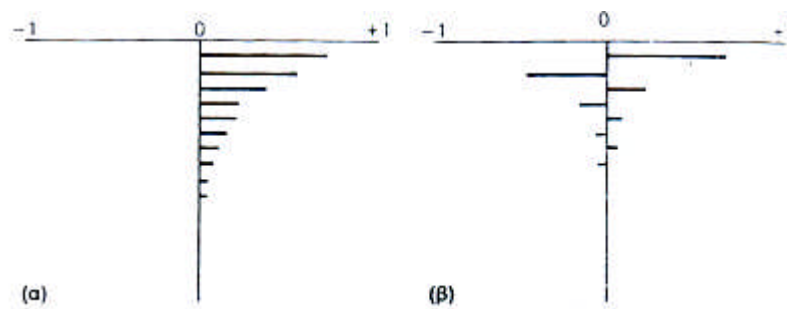
Οι μικτές διαδικασίες  $ARMA$  έχουν θεωρητικές  $ACF$  και με τα δυο χαρακτηριστικά  $AR$  και  $MA$ . Σαν παράδειγμα, το  $ACF$  του σχήματος 3.9 φθίνει προς το μηδέν με εκθετική ταχύτητα. Οι ημιτονοειδείς

καμπύλες που φθίνουν, θα μπορούσαν να χαρακτηρίζουν επίσης μια ACF ενός μοντέλου ARMA (1,1). Πρακτικά, τα  $p$  και  $q$  δεν είναι συνήθως μεγαλύτερα από δύο σε ένα μοντέλο ARMA για δεδομένα χωρίς εποχιακότητα. .

Η PACF για το ίδιο μοντέλο μάλλον φθίνει προς το μηδέν παρά διακόπτεται μετά από μία ή δύο χρονικές υστερήσεις, στο σχήμα 3.9. Κι αυτό, είναι ένα σημαντικό χαρακτηριστικό των μοντέλων ARMA. Ούτε η ACF, ούτε η PACF διακόπτονται στα μικτά μοντέλα ARMA.

Τέλος, έχει βρεθεί ότι οι περισσότερες από τις σταθερές (stationary) χρονοσειρές που προκύπτουν στην πράξη μπορούν να εκτιμηθούν με μοντέλα AR (1), AR (2), MA (1), MA (2), ARMA (1,1) ή από ένα μοντέλο white noise (τυχαίο). Με άλλα λόγια, αυτά τα έξι μοντέλα είναι τα μόνα μοντέλα χρονοσειρών που χρειάζονται συνήθως στην πράξη.

Σχήμα 3.9 (α) ACF για το μοντέλο ARMA (1,1)  
(β) PACF για το μοντέλο ARMA (1,1)



### 3.2.6 Σύνοψη για τη Μοντελοποίηση Μη-Εποχικών Μοντέλων Χρονοσειρών

Μέχρις αυτού του σημείου εξετάσαμε κοινά μοντέλα που χρησιμοποιήθηκαν στην πρόβλεψη χρονοσειρών με τις μεθόδους Box-Jenkins. Πριν συνεχίσουμε την παραπέρα εξέταση των μεθόδων πρόβλεψης, θα συνοψίσουμε τις γνώσεις μας για τα μη-εποχικά μοντέλα και τις στατιστικές τους ιδιότητες. Ο Πίνακας 3.11 συνοψίζει όσα αναφέρθηκαν ως τώρα για τα γενικά αυτοπαλινδρομικά μοντέλα κινητών μέσων και τα μικτά αυτοπαλινδρομικά μοντέλα κινητών μέσων. Στον ίδιο πίνακα, συνοψίζονται οι συνθήκες σταθερότητας και αναστρεψιμότητας που απαιτούνται για κάθε μοντέλο, καθώς και μια περιγραφή της συμπεριφοράς των συναρτήσεων θεωρητικής αυτοσυσχέτισης και των συναρτήσεων μερικής αυτοσυσχέτισης για κάθε μοντέλο.

Ο Πίνακας 3.12 συνοψίζει τις ίδιες συνθήκες σταθερότητας και αναστρεψιμότητας που αφορούν χρήσιμα και συγκεκριμένα μοντέλα χρονοσειρών. Συνοψίζεται επίσης, η περιγραφή της συμπεριφοράς των συναρτήσεων θεωρητικής και μερικής αυτοσυσχέτισης.

Ο πίνακας αυτός καθώς και ο προηγούμενος βοηθούν πολύ στην ταυτοποίηση των κατάλληλων

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

μοντέλων χρονοσειρών για την εκτίμηση Box-Jenkins. Για παράδειγμα, εάν για μια σταθερή χρονοσειρά καταλήγαμε στο συμπέρασμα ότι ένα μοντέλο MA (1) διαμορφώνει τη χρονοσειρά, επειδή ο συντελεστής αυτοσυσχέτισης διακόπτεται μετά από χρονική υστέρηση 1, και η μερική αυτοσυσχέτιση κυριαρχείται από εκθετικά φθίνοντα ρυθμό, θα μπορούσαμε να εκτιμήσουμε το μοντέλο με μια συγκεκριμένη μέθοδο Box-Jenkins. Θα πρέπει να σημειωθεί ότι αν δεν μπορούμε να ταυτοποιήσουμε το κατάλληλο μοντέλο, οι εκτιμήσεις μας ίσως να μην οδηγήσουν σε αποτελέσματα που να ικανοποιούν τους σκοπούς της πρόβλεψης.

**ΠΙΝΑΚΑΣ 3.11.** Ταυτοποίηση γενικών μοντέλων χρονοσειρών

(1)	(2)	(3)	(4)	(5)
<i>Μοντέλο</i>	<i>Συνθήκες Σταθερότητας</i>	<i>Συνθήκες αναστρεψιμότη τας</i>	<i>Συντελεστές αυτοσυσχέτισης</i>	<i>Συντελεστές Μερικής αυτοσυσχέτισης</i>
1 AR( $q$ )	Ναι	Όχι	Φθίνει προς τα κάτω	Διακόπτεται μετά από χρον. υστέρηση, $p$
2 MA( $q$ )	Όχι	Ναι	Διακόπτ. μετά από χρον. υστερήσεις $q$	Φθίνει προς τα κάτω
3 ARMA( $p,q$ )	Ναι	Ναι	Φθίνει προς τα κάτω	Φθίνει προς τα κάτω

*Μοντέλα χρονοσειρών*

$$1 \text{ AR}(p) \ Y_t = \mu + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + e_t$$

$$2 \text{ MA}(q) \ Y_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

$$3 \text{ ARMA}(p,q) \ Y_t = \mu + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} + e_t$$

**ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο**
**ΠΙΝΑΚΑΣ 3.12** Ειδικά μοντέλα χρονοσειρών

(1)	(2)	(3)	(4)	(5)
			<i>Θεωρητική Συνάρτηση</i>	
<i>Μοντέλο</i>	<i>Συνθήκες σταθερότητας</i>	<i>Συνθήκες αναστρεψιμ</i>	<i>Συντελεστές αυτοσυσχέτισης</i>	<i>Συντελεστές μερικής αυτοσυσχέτισης</i>
1 ARIMA(1,0,0)	$-1 < \varphi_1 < 1$	Καμία	Φθίνει προς τα κάτω	Διακόπτεται μετά από μια χρονική υστέρηση
ή				
AR (1)				
2 ARIMA(2,0,0)	$\varphi_1 + \varphi_2 < 1$	Καμία	Φθίνει προς τα κάτω	Διακόπτεται μετά από δύο χρονικές υστερήσεις
ή	$\varphi_1 - \varphi_2 < 1$			
AR(2)	$1 < \varphi_2 < 1$			
3 ARIMA(0,0, 1)	None	$\theta_1 < 1$	Διακόπτεται μετά μία χροιν. υστέρηση	Φθίνει προς τα κάτω
ή				
MA(1)				
4 ARIMA(0,0,2)	None	$\theta_1 + \theta_2 < 1$	Διακόπτεται μετά δύο χροιν. υστερήσεις	Φθίνει προς τα κάτω
		$\theta_2 - \theta_1 > 1$		
ή				
MA(2)		$\theta_2 < 1$		
5 ARIMA(1,0,1)	$-1 < \varphi_1 < 1$	$-1 < \theta_1 < 1$ Φθίνει προς τα κάτω		Φθίνει προς τα κάτω

Προσέξτε ότι στον Πίνακα 3.12 και η θεωρητική συνάρτηση αυτοσυσχέτισης μιας διαδικασίας AR (1) και η θεωρητική συνάρτηση αυτοσυσχέτισης μιας διαδικασίας AR (2) μπορεί να φθίνουν εκθετικά. Επομένως, είναι δύσκολο να ξεχωρίσουμε τα μοντέλα AR (1) και AR (2) χωρίς να εξετάσουμε τους συντελεστές μερικής αυτοσυσχέτισης του δείγματος. Για να διακρίνουμε τις διαφορές μεταξύ των μοντέλων AR (1) και AR (2), είναι σημαντικό να καθορίσουμε εάν οι συντελεστές μερικής αυτοσυσχέτισης του δείγματος διακόπτονται μετά από μία ή δύο χρονικές υστερήσεις

Με όμοιο τρόπο, δεν μπορεί κανείς να προσέξει μόνο τους συντελεστές μερικής αυτοσυσχέτισης στη διαδικασία των κινητών μέσων για να διαχωρίσει μια διαδικασία MA (1) από μια MA (2).

Μπορούμε να διαχωρίσουμε τις διαδικασίες MA (1) και MA (2) μόνο εφόσον εξετάσουμε τη συμπεριφορά των συντελεστών αυτοσυσχέτισης του δείγματος, για να διαπιστώσουμε πότε διακόπτονται. Γενικά, για να ταυτοποιήσουμε τη συγκεκριμένη διαδικασία AR( $p$ ), MA ( $q$ ) ή ARMA ( $p,q$ ) που διαμόρφωσε τη χρονοσειρά, πρέπει να εξετάσουμε τη συμπεριφορά των συντελεστών αυτοσυσχέτισης και των συντελεστών μερικής αυτοσυσχέτισης του δείγματος. Χρειάζεται προσπάθεια για να καθοριστεί εάν η συμπεριφορά είναι φθίνουσα ή διακόπτεται. Τέλος, ο αριθμός των χρονικών υστερήσεων των συντελεστών που εξετάζονται πρέπει να είναι αρκετά μεγάλος ώστε να αναγνωρίζεται εάν η κίνηση κυριαρχείται από ένα συγκεκριμένο πρότυπο στη χρονοσειρά. Γενικά όμως, το επιθυμητό επίπεδο χρονικών υστερήσεων είναι τουλάχιστον δέκα.

Τα μοντέλα που χρησιμοποιήθηκαν στη μεθοδολογία Box-Jenkins περιέχουν λίγες μόνο παραμέτρους. Κανένα από τα μοντέλα που συνοψίσαμε στον πίνακα 3.12 δεν περιέχει περισσότερες από τρεις παραμέτρους. Ένας από τους στόχους που θέτει η διαδικασία ανάπτυξης μοντέλων πρόβλεψης είναι η ανάπτυξη *φειδωλών* (parsimonious) μοντέλων. Με άλλα λόγια, επιθυμούμε την επίτευξη μοντέλων πρόβλεψης που περιγράφουν ικανοποιητικά τις χρονοσειρές και χρησιμοποιούν μόνο λίγες παραμέτρους. Ένας κανόνας που πρέπει να ακολουθηθεί είναι ότι εάν δύο διαφορετικά μοντέλα περιγράφουν εξίσου καλά μια χρονοσειρά, πρέπει να προτιμήσουμε το μοντέλο με τις λιγότερες παραμέτρους. Έχοντας να εκτιμήσουμε λιγότερες παραμέτρους, η διαδικασία της εκτίμησης γίνεται απλή, η ερμηνεία του μοντέλου απλούστερη και συχνά έχουμε λιγότερες πηγές πρόκλησης σφαλμάτων πρόβλεψης. Τέλος, με σχετικά λιγότερες παραμέτρους, είναι λιγότερο δύσκολο να βρούμε εκτιμήσεις των παραμέτρων που να είναι σημαντικές υπό στατιστική έννοια.

### 3.2.7 Περισσότερα για την Εκτίμηση με τη Μέθοδο Adaptive Filtering

Όπως ειπώθηκε νωρίτερα στην παράγραφο 3.2.3, η τεχνική adaptive filtering είναι παρόμοια με την εκθετική εξομάλυνση που βασίζει τις προβλέψεις της σε κάποιο είδος σταθμικού μέσου ιστορικών μετρήσεων. Το πρώτο στάδιο στην adaptive filtering είναι η ταυτοποίηση με τις μεθόδους της ανάλυσης αυτοσυσχέτισης του κατάλληλου μοντέλου χρονοσειράς (δηλαδή εξασφαλίζει σταθερότητα, αναστρεψιμότητα ή και τα δύο), που περιγράφει τη χρονοσειρά δείγματος. Δεύτερον, ακολουθώντας τα στάδια του Πίνακα 3.7, επαναλαμβάνουμε τις αρχικές σταθμίσεις (συνοψίζονται στο παράρτημα αυτού του κεφαλαίου) μέχρι να βρούμε τις παραμέτρους του μοντέλου που έχουν το χαμηλότερο δυνατό MSE.

Προηγουμένως, ο τύπος (εξίσωση 3.39) χρησιμοποιήθηκε για τον υπολογισμό των «άριστων» εκτιμήσεων των παραμέτρων ενός μοντέλου πρόβλεψης. Αντίθετα με τα μοντέλα AR που περιλαμβάνουν όρους  $Y_t$ , τα μοντέλα MA περιλαμβάνουν όρους  $e_t$ . Για να βελτιώσουμε τις παραμέτρους σε ένα μοντέλο MA, μπορούμε να χρησιμοποιήσουμε την εξίσωση (3.39) ως ακολούθως:

$$\theta'_i = \theta_i + 2ke_t^* (-e_{t-1}^*) \quad (3.55)$$

όπου  $e_t^*$  και  $e_{t-1}^*$  είναι οι τιμές του τυποποιημένου σφάλματος, που πήραμε από μια διαδικασία παρόμοια με εκείνη του Πίνακα 3.9. Εν συνεχεία, η εξίσωση (3.55) μπορεί να χρησιμοποιηθεί για να μας δώσει τις άριστες τιμές των παραμέτρων για ένα μοντέλο MA (1) ή MA (2).

Ένα μικτό ARMA υπδδειγμα αναπροσαρμοζόμενου φίλτρου μπορεί να μας επιτρέπει την αριστοποίηση των εκτιμήσεων των παραμέτρων ενός μοντέλου ARMA. Οι τύποι που δίνουν τις σταθμίσεις γι' αυτό το αναπροσαρμοζόμενο φιλτράρισμα, δίνονται από την εξίσωση (3.39) για τα μοντέλα AR και από την εξίσωση (8.26) για τα μοντέλα MA. Ειδικότερα,

$$\phi'_i = \phi_i + 2ke_t^* Y_{t-1} \quad (3.56)$$

$$\theta'_i = \theta_i + 2ke_t^* e_{t-1}^* \quad (3.57)$$

όπου  $e_t^*$  και  $Y_{t-1}$  είναι τυποποιημένες τιμές.

Οι εξισώσεις αυτές μπορούν να χρησιμοποιηθούν για την ανάπτυξη λύσεων αναπροσαρμοζόμενου φιλτραρίσματος στα μικτά μοντέλα ARMA. Μπορούμε να ξεχωρίσουμε τη χρήση του adaptive filtering για ένα μοντέλο ARMA ( $p, q$ ) από τη χρήση της στα απλά μοντέλα AR και MA, αναφερόντάς την ως γενικευμένη *adaptive filtering* (generalized adaptive filtering = GAF). . .

Οι μέθοδοι adaptive filtering και GAF απαιτούν να αρχίσουμε με τις αρχικές σταθμίσεις. Οι σταθμίσεις αυτές, καθώς και οι τύποι εύρεσής τους, συνοψίζονται στο Παράρτημα αυτού του κεφαλαίου και βασίζονται σε εργασίες που αναφέρονται από τους Makridakis και Wheelwright (1978) και στηρίζονται σε προγενέστερη μελέτη του Widrow (1966). Εφόσον η επανάληψη των σταθμίσεων χρειάζεται πολλούς στατιστικούς υπολογισμούς, συνιστάται η χρήση υπολογιστή.

Η ανάπτυξη καλών και ευκολοεύρετων προγραμμάτων software για υπολογιστές όσον αφορά τις μεθόδους GAF, έχει καθυστερήσει σημαντικά την ανάπτυξη προγραμμάτων υπολογιστών για άλλες τεχνικές εκτίμησης. Όπως θα δούμε στην παράγραφο 3.3, οι διαδικασίες Box-Jenkins δίνουν πολύ χρήσιμες μεθόδους για την εκτίμηση των μοντέλων και την πρόβλεψη.

Οι Kahl και Ledolter (1982, 1983) μελέτησαν τη μέθοδο του adaptive filtering για να ελέγξουν και να σχολιάσουν τους ισχυρισμούς των Makridakis και Wheelwright. Το έργο τους έδειξε ότι η επιλογή της σταθεράς μάθησης αν και είναι εύκολη, είναι πολύ κρίσιμη. Ακόμη, έδειξαν ότι η μέθοδος του adaptive filtering δεν οδηγεί απαραίτητα σε καλύτερη πρόβλεψη από ό,τι οι πλέον απλούστερες και αποτελεσματικές ως προς το κόστος μέθοδοι της σταθεράς τιμής παραμέτρου. Οι Kahl και Ledolter αναφέρουν ακόμη, ότι σε περιπτώσεις που οι συντελεστές είναι είτε σταθεροί είτε μέτρια μεταβαλλόμενοι ως προς τον χρόνο, η μέθοδος του adaptive filtering με μια σωστά επιλεγμένη σταθερά μάθησης οδηγεί σε μικρή μόνο αύξηση του MSE. Μόνο όταν υπάρχει σημαντική μεταβλητότητα της παραμέτρου, μπορεί η μέθοδος του adaptive filtering να οδηγήσει σε βελτιώσεις της ακριβείας της πρόβλεψης. Εν τούτοις, μόνο η χρήση της μεθόδου με διάφορες χρονοσειρές μπορεί να μας αποκαλύψει τις πιθανές αδυναμίες, αλλά και τα οφέλη τους.

Ένας άλλος σημαντικός λόγος για τη μελέτη της GAF είναι ότι μας επιτρέπει να εμβαθύνουμε

περισσότερο στις μεθόδους των Box-Jenkins (1976). Όπως θα δείξουμε στο επόμενο κεφάλαιο, η ανάλυση Box-Jenkins συνενώνει πολλές μεθόδους πρόβλεψης κάτω από ένα μόνο σύστημα, με ένα λεπτό τρόπο. Η συστηματική διαδικασία της ταυτοποίησης του μοντέλου της εκτίμησης των παραμέτρων και των ελέγχων για περισσότερες ή λιγότερες παραμέτρους, οδηγεί σε ένα συμπαγές μοντέλο πρόβλεψης που παρουσιάζει καλύτερα τη χρονοσειρά.

Επειδή οι διαδικασίες των Box-Jenkins για την εκτίμηση των μοντέλων ARMA δεν θέτουν προϋποθέσεις αναφορικά με το μοντέλο, ξεπερνούν τις αντιρρήσεις που εγείρονται για τα μοντέλα εκθετικής εξομάλυνσης και GAF. Ακόμη, όπως η GAF, η διαδικασία αυτή υπερπηδά τα προβλήματα που προκύπτουν από αυτοσυσχετιζόμενα σφάλματα, διασφαλίζοντας ότι τα λάθη  $e_t$  είναι τυχαία. Επίσης, με κατάλληλες διαφορίσεις και μετασχηματισμούς, εξαλείφονται τα προβλήματα της ετεροσκεδαστικότητας. Επομένως, η μελέτη της ανάλυσης αυτοσυσχέτισης και του adaptive filtering είναι χρήσιμα προκαταρκτικά στάδια στη μελέτη της ανάλυσης χρονοσειρών Box-Jenkins.

### 3.2.8 Τελεστές Backshift για τα Μοντέλα MA και ARMA

Στην παράγραφο 3.1 αναφέρθηκε ο τελεστής backshift που μας επιτρέπει να εκφράσουμε τα μοντέλα AR. Με όμοιο τρόπο, τα μοντέλα MA ή τα μικτά ARMA μπορούν να εκφραστούν σε όρους του τελεστή backshift,  $B$ . Για παράδειγμα, ένα μοντέλο MA (1) γράφεται:

$$Y_t = e_t - \theta_1 B e_t$$

ή

$$Y_t = (1 - \theta_1 B) e_t, \text{ ARIMA } (0,0,1) \quad (3.58)$$

Ένα μοντέλο MA (2) γράφεται:

$$Y_t = (1 - \theta_1 B - \theta_2 B^2) e_t, \text{ ARIMA } (0,0,2) \quad (3.59)$$

Ένα μοντέλο MA (2) για δεδομένα των πρώτων διαφορών γράφεται:

$$(1 - B) Y_t = (1 - \theta_1 B - \theta_2 B^2) e_t, \text{ ARIMA } (0,1,2) \quad (3.60)$$

Ένα μοντέλο ARIMA (1,0,1) γράφεται:

$$(1 - \phi_1 B) Y_t = (1 - \theta_1 B) e_t \quad (3.61)$$

Τέλος, ένα μοντέλο ARIMA (1,1,1) γράφεται:

$$(1 - B)(1 - \phi_1 B) Y_t = (1 - \theta_1 B) e_t \quad (3.62)$$



3.3 ΟΙ ΜΕΘΟΔΟΙ BOX- JENKINS [18], [19]

3.3.1 Εισαγωγή

Η μεθοδολογία των Box-Jenkins (1976) αποτελεί μια στατιστικά εξειδικευμένη προσέγγιση στην ανάλυση και την κατασκευή ενός μοντέλου πρόβλεψης, με στόχο την όσο το δυνατόν καλύτερη αναπαράσταση μιας χρονοσειράς. Η προσέγγιση αυτή έχει έναν αριθμό πλεονεκτημάτων έναντι άλλων μεθόδων ανάλυσης χρονοσειρών. Πρώτον, η μέθοδος είναι λογική και στατιστικά ακριβής. Δεύτερον, αποσπά ένα μεγάλο αριθμό πληροφοριών από τα ιστορικά δεδομένα της χρονοσειράς. Τέλος, επιδρά στην αύξηση της ακριβείας της πρόβλεψης, ενώ συγκρατεί τον αριθμό των παραμέτρων σε ένα ελάχιστο επίπεδο, σε σύγκριση με άλλες παρόμοιες διαδικασίες κατασκευής μοντέλων.

Με τον υπολογισμό των συντελεστών αυτοσυσχέτισης, για διάφορες χρονικές υστερήσεις μιας χρονοσειράς, μπορούμε να καθορίσουμε τις παρακάτω ιδιότητες ενός συνόλου δεδομένων:

1. Είναι τα δεδομένα τυχαία;
2. Είναι η χρονοσειρά μη-σταθερή;
3. Εάν είναι μη-σταθερή, ποιος είναι ο βαθμός της διαδικασίας στον οποίο γίνεται στάσιμη;
4. Είναι τα δεδομένα εποχικά;
5. Εάν είναι εποχικά, ποιο είναι το εποχικό πρότυπο;

Εάν η σειρά είναι τυχαία, ο βαθμός συσχέτισης ενός ζεύγους δεδομένων,  $Y_t$  και  $Y_{t-1}$ , είναι περίπου μηδέν. Ακόμη, στο Στην παράγραφο 3.1, είδαμε ότι συντελεστές αυτοσυσχέτισης που βρίσκονται κοντά στο μηδέν δείχνουν ότι οι διαδοχικές τιμές μιας χρονοσειράς δεν σχετίζονται γραμμικά μεταξύ τους. Οι συντελεστές αυτοσυσχέτισης για τις διαδοχικές τιμές μιας χρονοσειράς -3, 6, 9, 12, 15, 18, 21, 24 ... - θα ήταν πολύ υψηλοί. Με άλλα λόγια, η σχέση μεταξύ διαδοχικών τιμών θα ήταν ισχυρή. Με την εξέταση των συντελεστών αυτοσυσχέτισης για χρονικές υστερήσεις μεγαλύτερες της μιας περιόδου, μπορούμε να αντλήσουμε επιπρόσθετες πληροφορίες που αφορούν τη σχέση των δεδομένων σε μια δεδομένη χρονοσειρά.

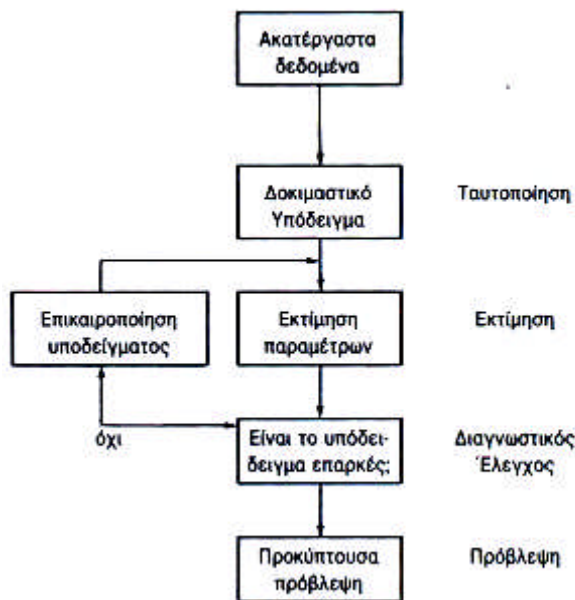
Στην παράγραφο 3.1, εξετάσαμε ένα αυτοπαλίνδρομο μοντέλο που είχε *τριμηνιαίο* εποχικό πρότυπο. Για να ταυτοποιήσουμε αυτό το πρότυπο, η μεθοδολογία της ανάλυσης αυτοσυσχέτισης προτείνει να καθορίσουμε τη συσχέτιση μεταξύ  $Y_t$  και  $Y_{t+4}$ . Εάν η αυτοσυσχέτιση βρίσκεται πλησίον του μηδενός, τότε μπορεί ίσως να λεχθεί ότι δεν υπάρχει τριμηνιαία εποχική σχέση. Με παρόμοιο τρόπο, για τις μηνιαίες παρατηρήσεις, μελετάται η αυτοσυσχέτιση των  $Y_t$  και  $Y_{t+12}$ .

Σκοπός των υπόλοιπων τμημάτων αυτού του κεφαλαίου, είναι να συμπληρώσει τις γνώσεις μας για την ανάλυση αυτοσυσχέτισης, παράλληλα με τις γνώσεις μας για τις διαδικασίες ARMA από την παράγραφο 3.2. Οι Box και Jenkins (1976) ενοποίησαν με επιτυχία τη μεθοδολογία που εκφράζει τις ανάγκες γνώσεων για την κατανόηση και χρησιμοποίηση απλών μοντέλων ARMA με μία μεταβλητή (univariate). Η βάση αυτής της προσέγγισης περιγράφεται στο διάγραμμα του σχήματος 3.10 και περιλαμβάνει τέσσερις διαφορετικές φάσεις.

Πρώτον, η μεθοδολογία υποθέτει ότι δεν υπάρχει κάποιο ιδιαίτερο πρότυπο στα ιστορικά δεδομένα της χρονοσειράς ώστε να προβλεφθεί. Με μια επαναλαμβανόμενη προσέγγιση, η διαδικασία ταυτοποιεί ένα πιθανό χρήσιμο μοντέλο από μια γενική τάξη μοντέλων ARMA. Μετά τον υπολογισμό, το επιλεγέν μοντέλο ελέγχεται διαγνωστικά έναντι της ιστορικής χρονοσειράς, προκειμένου να αποδειχθεί αν την περιγράφει με ακρίβεια.

Εάν τα κατάλοιπα μεταξύ της πρόβλεψης και της πραγματικής χρονοσειράς είναι μικρά, τυχαία κατανομημένα και ανεξάρτητα, το επιλεγμένο μοντέλο ARMA λέγεται ότι εμφανίζει καλή προσαρμογή. Αν όμως το επιλεγμένο μοντέλο δεν είναι ικανοποιητικό, η διαδικασία Box-Jenkins επαναλαμβάνεται χρησιμοποιώντας ένα άλλο μοντέλο, σχεδιασμένο για να βελτιώσει το αρχικό. Η διαδικασία αυτή συνεχίζεται μέχρις ότου βρεθεί ικανοποιητικό μοντέλο.

Σχήμα 3.10 Η διαδικασία Box-Jenkins



### 3.3.2. Ταυτοποίηση

Σκοπός της φάσης της ταυτοποίησης είναι η επιλογή ενός ιδιαίτερου μοντέλου ARMA από τη γενική τάξη των μοντέλων ARMA  $(p,q)$  που γράφεται ως εξής:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (3.63)$$

Η επιλογή των κατάλληλων τιμών των  $p$  και  $q$  προϋποθέτει την εξέταση των συντελεστών αυτοσυσχέτισης και μερικής αυτοσυσχέτισης που έχουν υπολογιστεί για τα δεδομένα.

Οι μερικές αυτοσυσχετίσεις αναφέρθηκαν σε προηγούμενο κεφάλαιο, και στο εξής θα παίρνουμε σαν βάση αυτές τις θέσεις πριν επιστρέψουμε στο ερώτημα του κατά πόσο ένα ειδικό μοντέλο ARMA μπορεί να επιλεγεί μεταξύ της γενικής τάξης μοντέλων ARMA  $(p, q)$  της εξίσωσης (3.63).

Εάν στην εξίσωση (3.63), θέσουμε διαδοχικά  $q=0$  και  $p = 0, 1, 2, 3, \dots, p$ , τότε η εξίσωση (3.63) γίνεται:

$$Y_t = e_t \quad (3.64)$$

$$Y_t = \phi_1 Y_{t-1} + e_t \quad (3.65)$$

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t \quad (3.66)$$

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + e_t \quad (3.67)$$

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \dots + \phi_p Y_{t-p} + e_t \quad (3.68)$$

Όταν η αληθής τάξη της εξίσωσης (3.65) είναι  $p=0$ , η παράμετρος  $\Phi_1$ , θα έχει μια τιμή που δεν είναι στατιστικά διάφορη του μηδενός. Επομένως, το αποτέλεσμα θα είναι μια διαδικασία AR(0) και  $\Phi_1=0$ . Εναλλακτικά, αν η πραγματική τάξη είναι  $p=1$ , το  $\Phi_2$  δεν θα είναι στατιστικά διάφορο του μηδενός. Τέλος, η  $p$ -οστή παράμετρος μιας διαδικασίας AR  $(p)$  μόνο τότε θα είναι στατιστικά διάφορη του μηδενός, όταν η διαδικασία AR είναι τουλάχιστον τάξης  $p$  ή υψηλότερης.

Η ταυτοποίηση της τάξης μιας διαδικασίας AR μπορεί να γίνει με την εξέταση των συντελεστών μερικής αυτοσυσχέτισης.

Η τάξη συμπίπτει με τον αριθμό των μερικών αυτοσυσχετίσεων που είναι στατιστικά διάφορες του μηδενός. Οι μερικές αυτοσυσχετίσεις που φθάνουν έως  $p$  χρονικές υστερήσεις, θα είναι στατιστικά σημαντικές, ενώ οι υπόλοιποι συντελεστές θα είναι σχεδόν ίσοι με το μηδέν. Η τιμή του  $p$  που προκύπτει θα αποτελεί την τάξη της διαδικασίας AR.

Οι διαδικασίες κινητών μέσων δεν είναι όπως στα μοντέλα (3.65) έως και (3.68) και η προσπάθεια υπολογισμού των  $\Phi_1, \Phi_2, \Phi_3, \dots, \Phi_p$  σε τέτοιες περιπτώσεις είναι ένα παράδειγμα εκτίμησης λάθους μοντέλου στα δεδομένα. Για μια διαδικασία MA, οι διαφορετικές τιμές  $Y_t$  είναι εξαρτημένες μεταξύ τους. Εφόσον η  $Y_t$  είναι συνάρτηση των  $Y_{t-1}$  και  $e_t$ , και η  $Y_{t-1}$  είναι συνάρτηση των  $Y_{t-2}$  και  $e_{t-1}$ , και ούτω καθ' εξής, θα μπορούσαμε να συνεχίσουμε τον υπολογισμό μιας νέας παραμέτρου  $\phi$  για κάθε εξίσωση. Επομένως, θα απαιτεί το άπειρο αριθμός  $\phi_i$  όρων για να εκτιμηθεί ένα μοντέλο AR με δεδομένα MA. Οι συντελεστές μερικής συσχέτισης για δεδομένα MA θα είναι αρχικά μεγάλοι και το μέγεθός τους θα μειώνεται εκθετικά προς το μηδέν. Σε αντίθεση με τις μερικές αυτοσυσχετίσεις για τις διαδικασίες AR, οι μερικές αυτοσυσχετίσεις μιας διαδικασίας MA δεν διακόπτονται μετά από  $P$  χρονικές υστερήσεις.

Με τη γνώση της συμπεριφοράς των αυτοσυσχετίσεων και των μερικών αυτοσυσχετίσεων, που αναπτύχθηκε και δόθηκε συνοπτικά στις παραγράφους 3.1 και 3.2, μπορούμε να αρχίσουμε την περιγραφή των πρώτων σταδίων της ανάλυσης Box-Jenkins.

Το πρώτο στάδιο ταυτοποίησης του μοντέλου είναι ο καθορισμός της ύπαρξης σταθερότητας ή

όχι στη χρονοσειρά. Αν η χρονοσειρά δεν είναι σταθερή, μπορεί να μετατραπεί σε σταθερή με τη μέθοδο των διαφορών. Όπως είδαμε στην παράγραφο 3.1, χρειαζόμαστε μόνο τις διαφορές στα δεδομένα είτε πρώτου είτε δευτέρου βαθμού για να μετατρέψουμε τη χρονοσειρά σε σταθερή. Για την ανάλυση Box-Jenkins, τα δεδομένα πρέπει να σταθεροποιηθούν και/ή να αναστραφούν και έπειτα να ακολουθήσουν οι υπολογισμοί που θα χρησιμοποιούν τα μετασχηματισμένα δεδομένα.

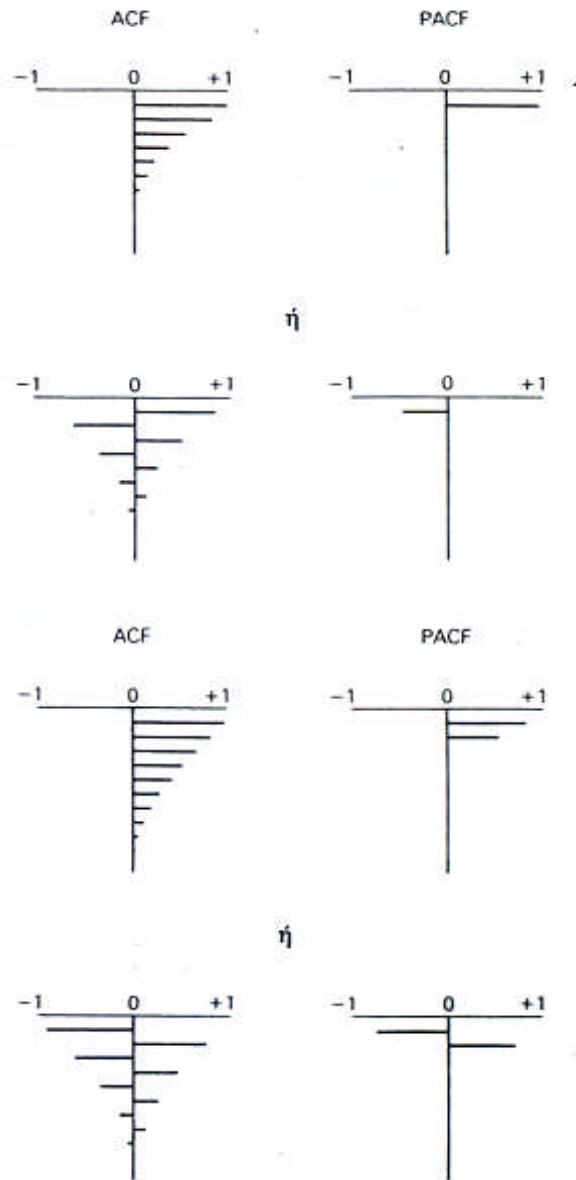
Για να πάρουμε μια σταθερή και/ή αναστρέψιμη χρονοσειρά, πρέπει να ταυτοποιήσουμε τη μορφή του μοντέλου που πρόκειται να χρησιμοποιηθεί. Η μορφή επιτυγχάνεται με τη σύγκριση των συντελεστών αυτοσυσχέτισης και μερικής αυτοσυσχέτισης των δεδομένων, ώστε να γίνει προσαρμογή με τις αντίστοιχες κατανομές των διαφόρων μοντέλων ARMA. Οι πιο χρήσιμες θεωρητικές κατανομές για τα μοντέλα ARMA, δείχνονται στα σχήματα 3.11, 3.12 και 3.13. Σημειώστε ότι το σχήμα 3.11 αφορά τα μοντέλα AR (1) και το AR (2). Το σχήμα 3.12 περιέχει τις αυτοσυσχετίσεις και τις μερικές αυτοσυσχετίσεις για τα μοντέλα MA (1) και MA (2). Τέλος, οι αυτοσυσχετίσεις και μερικές αυτοσυσχετίσεις για τα μοντέλα ARMA (1,1) αποτυπώνονται στο σχήμα 3.13. Τα σχήματα αυτά βασίζονται στις αναλύσεις που έγιναν στο στην παράγραφο 3.2.

Κάθε τύπος της διαδικασίας ARIMA έχει ένα μοναδικό σύνολο αυτοσυσχετίσεων και μερικών αυτοσυσχετίσεων και θα μπορούσαμε να εξομοιώσουμε τους αντίστοιχους συντελεστές των εκ παρατηρήσεως χρονοσειρών με εκείνους των θεωρητικών κατανομών.

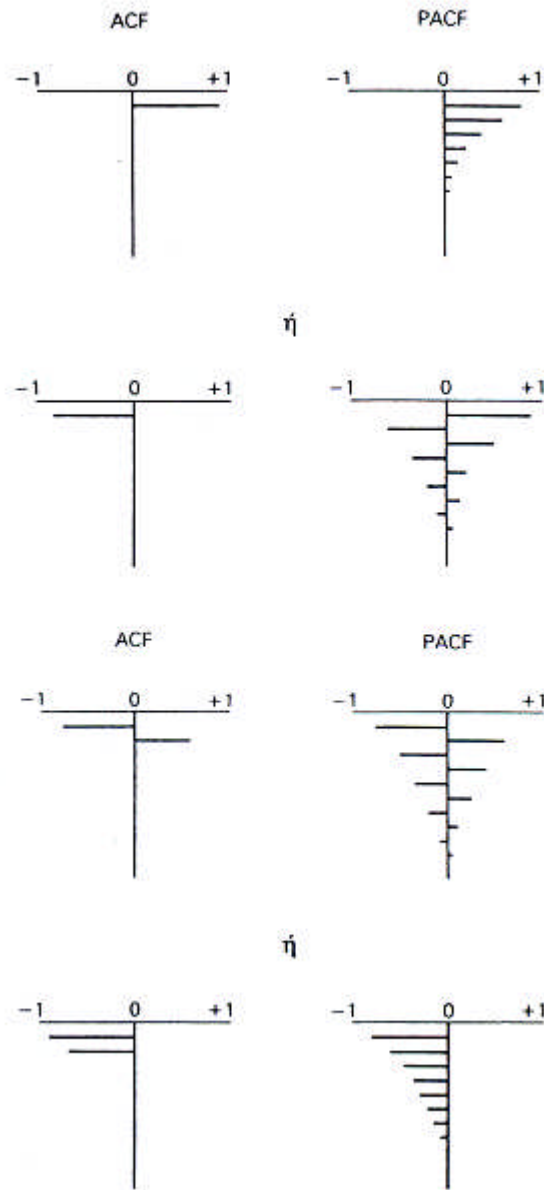
Δεν είναι δυνατό να εξομοιώσουμε τις παρατηρηθείσες χρονοσειρές επακριβώς με τα θεωρητικά μοντέλα, ωστόσο μπορούν να γίνουν διαγνωστικοί έλεγχοι σε ένα από τα επόμενα στάδια για τον καθορισμό της επάρκειας του μοντέλου. Εάν το προκαταρκτικό μοντέλο δεν είναι επαρκές, μπορούμε να εξειδικεύσουμε ένα εναλλακτικό.

Παρατηρήσεις: Πρώτον, εάν οι αυτοσυσχετίσεις φθίνουν εκθετικά προς το μηδέν, ενδείκνυται, ένα μοντέλο AR. Κατά παρόμοιο τρόπο, εάν οι μερικές αυτοσυσχετίσεις φθίνουν προς το μηδέν, τότε χρειάζεται ένα μοντέλο MA. Εάν και η αυτοσυσχέτιση και η μερική αυτοσυσχέτιση φθίνουν στο μηδέν, τότε το σωστό μοντέλο είναι ARMA. Τέλος, η τάξη του μοντέλου AR υποδηλώνεται από τον αριθμό των μερικών αυτοσυσχετίσεων, και η τάξη του μοντέλου MA από τον αριθμό των αυτοσυσχετίσεων, που είναι στατιστικά διαφορές του μηδενός.

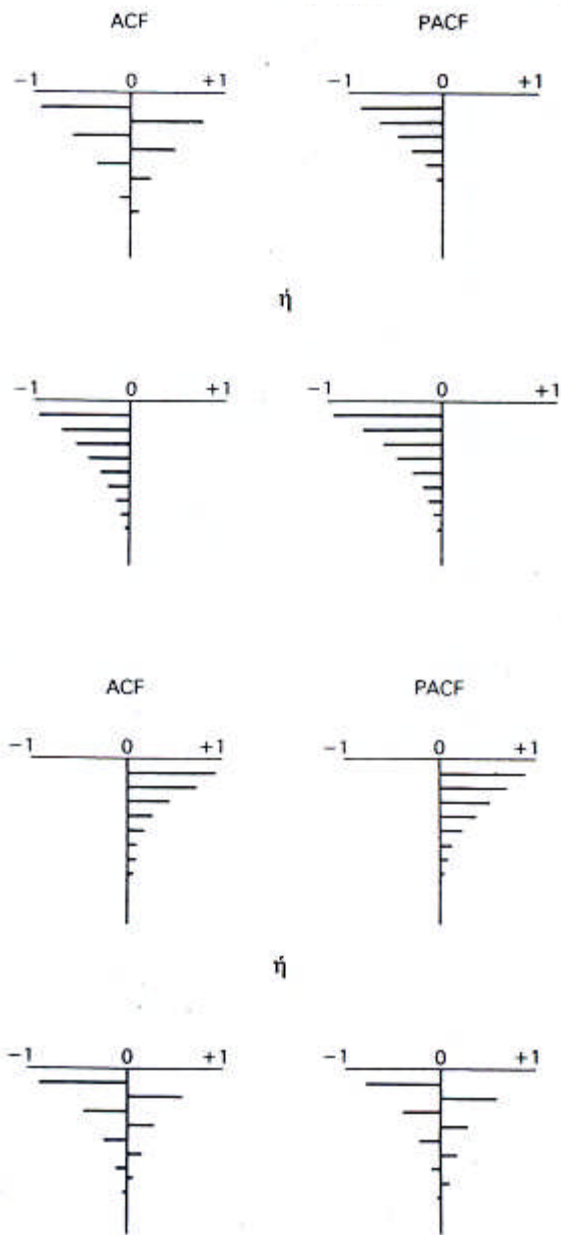
Σχήμα 3.11 (α) Μοντέλο AR(1),  
(β) Μοντέλο AR(2)



Σχήμα 3.12 (α) Μοντέλο MA(1),  
(β) Μοντέλο MA(2)



Σχήμα 3.13 (α) Μοντέλο ARMA(1,1), (β) Μοντέλο ARMA(1,1)



### 3.3.3 Εκτίμηση Μοντέλου και Διαγνωστικός Έλεγχος

Το δεύτερο στάδιο στην κατασκευή του μοντέλου είναι η εκτίμηση ή το στάδιο προσαρμογής. Τα μοντέλα ARIMA μπορούν να εκτιμηθούν με τη μέθοδο των ελαχίστων τετραγώνων. Μια επαναληπτική μη-γραμμική διαδικασία ελαχίστων τετραγώνων χρησιμοποιείται στις εκτιμήσεις της παραμέτρου ενός μοντέλου ARMA  $(p,q)$ . Οι εκτιμήσεις ελαχιστοποιούν το άθροισμα των τετραγώνων των σφαλμάτων που είναι  $(\sum e_t^2)$  με δεδομένη τη μορφή του μοντέλου και των στατιστικών στοιχείων. Εφόσον, γενικά, η διαδικασία είναι μη-γραμμική - που οφείλεται στους όρους του κινητού μέσου - η προσέγγιση των ελαχίστων τετραγώνων είναι και αυτή μη-γραμμική.

Οι Box και Jenkins (1976) προτείνουν μια ποικιλία διαδικασιών εκτίμησης των παραμέτρων, στις οποίες ο αναγνώστης που ενδιαφέρεται μπορεί να προστρέξει για λεπτομέρειες. Για τα περισσότερα μοντέλα ARMA χαμηλής τάξης υπάρχουν προγράμματα software για τον υπολογισμό τους και για τον έλεγχο των παραμέτρων του μοντέλου. Τα προγράμματα αυτά μας επιτρέπουν την επίτευξη ακριβείας των εκτιμήσεων των παραμέτρων που ονομάζονται φειδωλά μοντέλα πρόβλεψης.

Με τον όρο αυτό χαρακτηρίζουμε εκείνα τα μοντέλα που έχουν μάλλον λιγότερες παρά περισσότερες παραμέτρους στο μοντέλο. Εφόσον τα μοντέλα αυτά είναι πρακτικά και έχουν θεωρητική υπόσταση, οι ασχολούμενοι με τις προβλέψεις κατέληξαν ότι τα απλούστερα μοντέλα είναι και τα καλύτερα.

Πριν το μοντέλο χρησιμοποιηθεί για πρόβλεψη, θα πρέπει να ελεγχθεί για επάρκεια. Αυτός ο διαγνωστικός έλεγχος γίνεται με την εξέταση των όρων του σφάλματος  $e_t$  για να βεβαιωθούμε ότι είναι τυχαίοι. Εάν οι όροι του σφάλματος είναι στατιστικά διάφοροι του μηδενός, το μοντέλο δεν θεωρείται επαρκές. Εάν αρκετές από τις αυτοσυσχετίσεις είναι μεγάλες, θα πρέπει να επιστρέψουμε στο αρχικό στάδιο και να επιλέξουμε ένα εναλλακτικό μοντέλο και έπειτα να συνεχίσουμε την ανάλυση.

Για να ελέγξουμε την επάρκεια, εξετάζουμε διαγνωστικά τις αυτοσυσχετίσεις των καταλοίπων με τον υπολογισμό του στατιστικού  $X^2$ . Το κριτήριο ελέγχου είναι:

$$Q = n(n+2) \sum_i^k \frac{r_i^2}{n-k}$$

το οποίο κατανέμεται κατά προσέγγιση, σύμφωνα με την κατανομή  $X^2$  με  $k-p-q$  βαθμούς ελευθερίας. Σε αυτή την εξίσωση,  $n$  είναι το μήκος της χρονοσειράς  $k$  είναι οι πρώτες  $k$  αυτοσυσχετίσεις που έχουν ελεγχθεί,  $p$  είναι η τάξη της διαδικασίας AR,  $q$  είναι η τάξη της διαδικασίας MA, και  $r$  είναι ο εκτιμημένος συντελεστής αυτοσυσχετίσης του  $i$ -στού κατάλοιπου όρου. Εάν η υπολογισθείσα τιμή του  $Q$  είναι μεγαλύτερη από το  $X^2$  για  $k-p-q$  βαθμούς ελευθερίας, τότε το μοντέλο θα πρέπει να θεωρείται ανεπαρκές. Ο προβλέπων θα πρέπει να επιστρέψει στην επιλογή ενός εναλλακτικού μοντέλου και να συνεχίσει την ανάλυση Box-Jenkins μέχρι να βρεθεί ένα ικανοποιητικό μοντέλο.



Και οι δύο διαγνωστικές διαδικασίες βοηθούν στο να φθάσει η ανάλυση σε ένα τελικό μοντέλο πρόβλεψης. Όμως καμία διαδικασία δεν μπορεί να θεωρηθεί ως η τελική λέξη. Για παράδειγμα, εάν κάποιες μεγάλες αποκλίσεις από την πρόβλεψη μπορούν να εξηγηθούν ικανοποιητικά ως ασυνήθεις και μη επαναλαμβανόμενες καταστάσεις, αυτές οι αποκλίσεις μπορούν να αγνοηθούν.

Τέλος, αν δύο ή περισσότερα μοντέλα θεωρούνται περίπου ίδια, μολονότι κανένα μοντέλο δεν δείχνει ακριβή προσαρμογή, κατά την επιλογή πρέπει να επικρατήσει η ιδιότητα της φειδωλότητας.

### 3.3.4 Η Πρόβλεψη με Μοντέλα ARIMA

Μετά την ταυτοποίηση και τον έλεγχο της καταλληλότητας ενός μοντέλου, μπορούν να γίνουν προβλέψεις μιας ή περισσότερων περιόδων στο μέλλον. Καθώς η περίοδος πρόβλεψης γίνεται πιο μακρινή μέσα στο μέλλον, οι πιθανότητες σφάλματος πρόβλεψης μεγαλώνουν.

Μετά την απόκτηση νέων παρατηρήσεων για μια χρονοσειρά, το μοντέλο πρέπει να επανεξετάζεται και να ελέγχεται ως προς την ακρίβεια. Εάν η χρονοσειρά φαίνεται ότι αλλάζει διαχρονικά, οι παράμετροι του μοντέλου πρέπει να επαναυπολογιστούν ή θα πρέπει να αναπτυχθεί ένα εντελώς νέο μοντέλο. Όταν παρατηρηθούν μικρές μόνο διαφορές στο σφάλμα πρόβλεψης, είναι αρκετό να επαναυπολογίσουμε μόνο τις παραμέτρους του μοντέλου. Από την άλλη, αν παρατηρηθούν μεγάλες διαφορές στο μέγεθος του σφάλματος πρόβλεψης, αυτό δείχνει την ανάγκη για νέο μοντέλο πρόβλεψης. Από εκείνη τη στιγμή θα πρέπει να επιστρέψουμε στο πρώτο στάδιο της διαδικασίας Box-Jenkins.

Ο δείκτης βιομηχανικής παραγωγής (ΙΠ) που κατασκευάστηκε από το Federal Reserve Board δεν είναι σταθερός, εφόσον το επίπεδο και η κλίση της χρονοσειράς μεταβάλλονται διαχρονικά. Ο Πίνακας 3.13 περιέχει δεδομένα που αφορούν τον ΙΠ ανά μήνα για μια χρονική περίοδο 8 1/2 ετών. Μελετάμε τους αριθμούς του ΙΠ που έχουν προηγουμένως απαλλαγεί της εποχιακότητας, για να αποφύγουμε προβλήματα που σχετίζονται με τη συνιστώσα της εποχιακότητας της χρονοσειράς.

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

ΠΙΝΑΚΑΣ 3.13 Δείκτης Βιομηχανικής παραγωγής

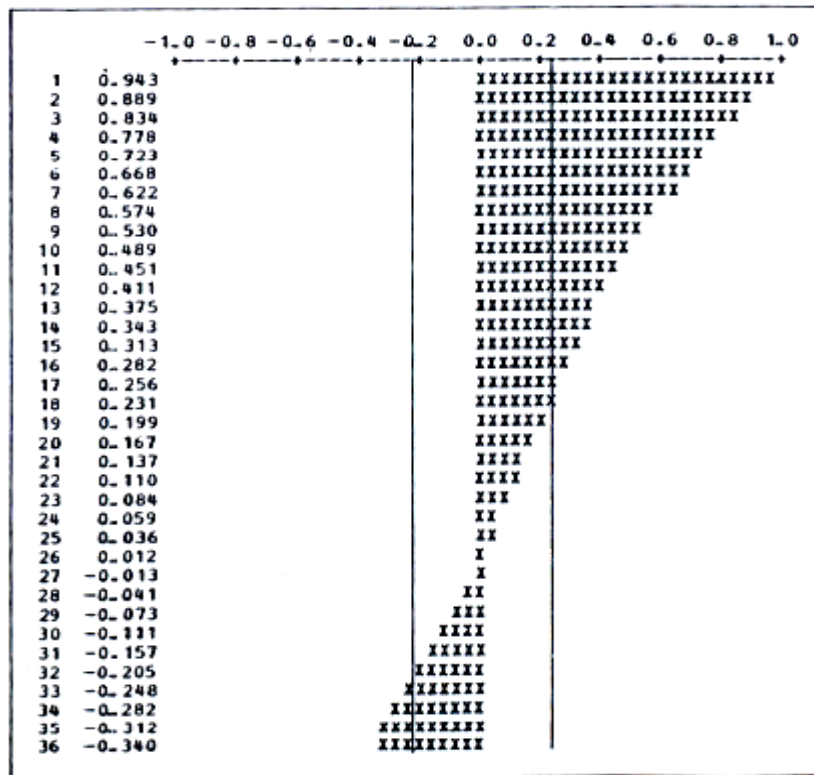
Μήνας	Έτος	Περίοδος	ΠΡ
Ιούλ.	1975	1	151.5
Αύγ.	1975	2	152
Σεπ.	1975	3	153
Οκτ.	1975	4	150.8
Νοέμ.	1975	5	152.4
.	.	.	.
.	.	.	.
.	.	.	.
Αύγ.	1982	86	140.5
Σεπ.	1982	87	141.2
Οκτ.	1982	88	138.5
Νοέμ.	1982	89	134.8
Δεκ.	1982	90	131.2
Ιαν.	1983	91	133.5
Φεβρ.	1983	92	138.1
Μάρ.	1983	93	140.5
Απρ.	1983	94	141.9
Μάιος	1983	95	143.7

ΠΗΓΗ: Επιτροπή Κυβερνητών του Ομοσπ. Συστήματος αποθεματικών των ΗΠΑ, Federal Reserve Bulletin και Υπουργείο Εμπορίου των ΗΠΑ. Survey of Current Business, διάφορα τεύχη.

Οι πρώτες 90 παρατηρήσεις χρησιμοποιούνται για την ταυτοποίηση, εκτίμηση και το διαγνωστικό έλεγχο ενός κατάλληλου μοντέλου. Οι παρατηρήσεις 91 έως 95 θα χρησιμοποιηθούν για τον έλεγχο της επίδοσης του μοντέλου πρόβλεψης.

**Στάδιο 1: Ταυτοποίηση.** Το πρώτο στάδιο στην ταυτοποίηση ενός προκαταρκτικού μοντέλου είναι να εξετάσουμε τις αυτοσυσχετίσεις και τις μερικές αυτοσυσχετίσεις για τα ακατέργαστα δεδομένα. Οι τιμές αυτές δείχνονται στα σχήματα 3.14 και 3.15

Σχήμα 3.14. Συντελεστές αυτοσυσχέτισης για το δείκτη βιομηχανικής παραγωγής χρονοσειράς (αποτελέσματα σύμφωνα με το πρόγραμμα Minitab)



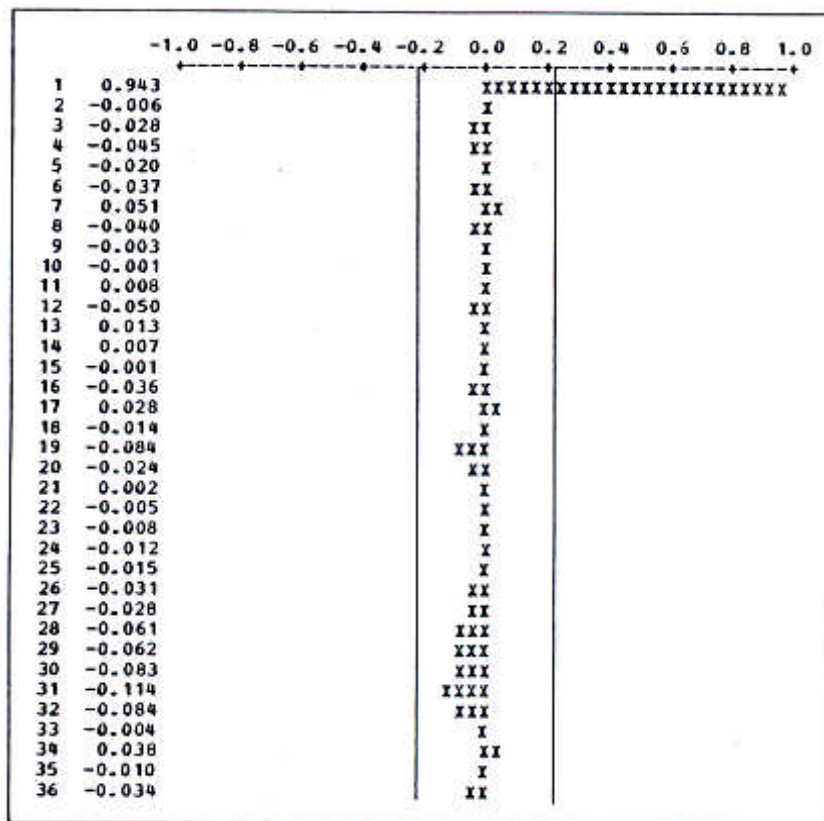
Στο σχήμα 3.14 αποτυπώνονται οι πρώτες 15 αυτοσυσχετίσεις. Οι αυτοσυσχετίσεις που είναι πολύ μεγάλες στην αρχή δεν φθίνουν γρήγορα προς το μηδέν. Εμφανίζονται να σχηματίζουν ένα πρότυπο ημιτονοειδούς καμπύλης, αλλά επειδή η φθίνουσα διαδικασία είναι πολύ αργή μπορούμε να συμπεράνουμε ότι η διαδικασία είναι μη-σταθερή. Σημειωτέον ότι οι πρώτοι 18 συντελεστές βρίσκονται εκτός των ορίων των δύο τυπικών σφαλμάτων.

Το σχήμα 3.15 περιέχει τις μερικές αυτοσυσχετίσεις. Για χρονική υστέρηση ένα, η μερική αυτοσυσχέτιση βρίσκεται εκτός των ορίων των δύο σίγμα. Αυτά τα γραφήματα (ονομάζονται coneIograms) δείχνουν ότι η χρονοσειρά είναι πιθανόν μια διαδικασία AR (1), αλλά είναι μη-σταθερή και συνιστάται διαφόριση.

Για να διαφορίσουμε μια διαδικασία ARIMA ώστε να προκύψει μια σταθερή διαδικασία ARMA για την πρόβλεψη, πρέπει να χαραχθούν νέα διαγράμματα των συντελεστών αυτοσυσχέτισης και των συντελεστών μερικής αυτοσυσχέτισης.

ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

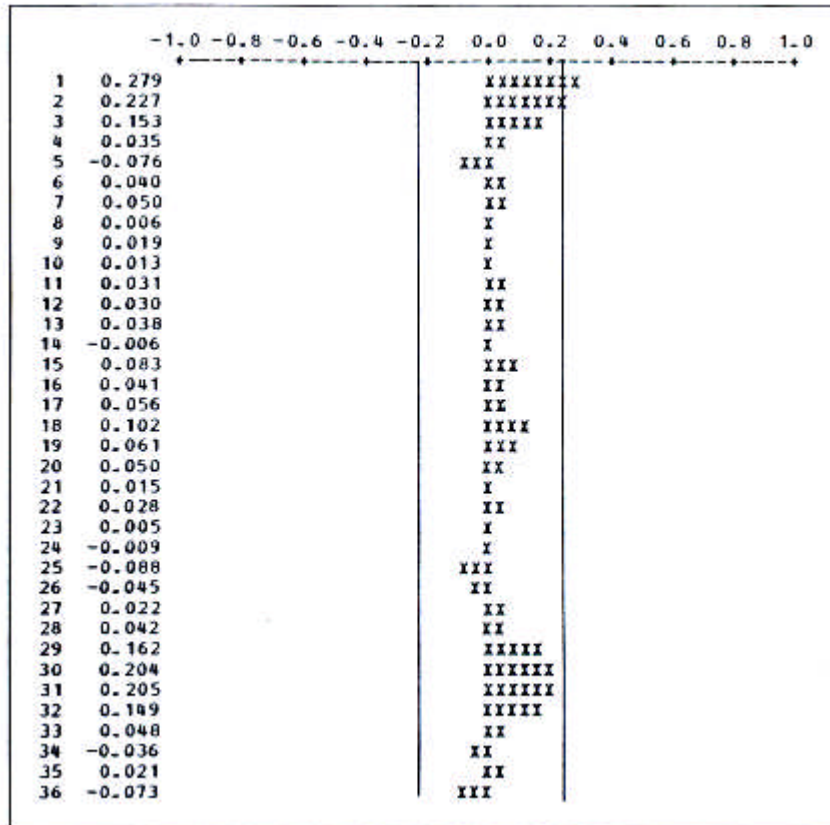
Σχήμα 3.15 Συντελεστές μερικής αυτοσυσχέτισης του δείκτη βιομηχανικής παραγωγής χρονοσειράς (αποτελέσματα σύμφωνα με το πρόγραμμα Minitab)



## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

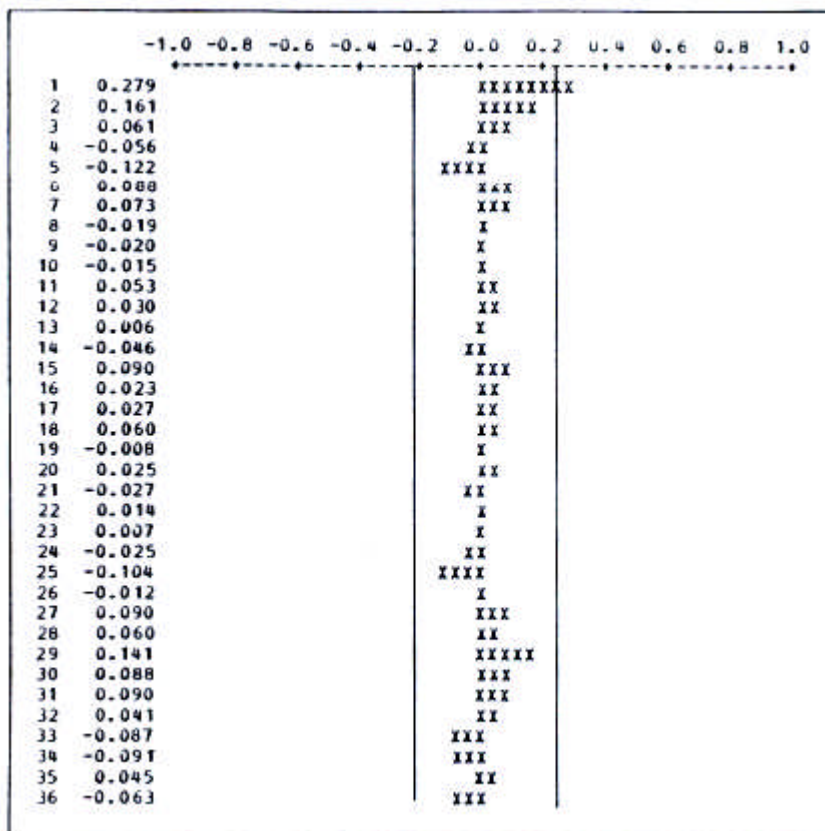
Το σχήμα 3.16 περιέχει την αυτοσυσχέτιση από τις πρώτες διαφορές.

**Σχήμα 3.16** Συντελεστές αυτοσυσχέτισης για χρονοσειρά πρώτων διαφορών (αποτελέσματα σύμφωνα με το πρόγραμμα Minitab)



Οι αυτοσυσχετίσεις αυτές φθίνουν προς το μηδέν. Οι μερικές αυτοσυσχετίσεις (σχήμα 3.17) διακόπτονται μετά από μια χρονική υστέρηση. Και τα δυο αυτά πρότυπα υποδηλώνουν ένα μοντέλο ARIMA (1,1,0). Ωστόσο, αυτή είναι μια δοκιμαστική επιλογή.

Σχήμα 3.17 Συντελεστές μερικής αυτοσυσχέτισης χρονοσειράς για χρονοσειρά πρώτων διαφορών (αποτελέσματα σύμφωνα με το πρόγραμμα Minitab)



**Στάδιο 2: Εκτίμηση:** Εφόσον επιλεγεί το προκαταρκτικό μοντέλο, αρχίζει το στάδιο της εκτίμησης. Οι εκτιμητές πρέπει να είναι όπως το παρακάτω μοντέλο:

$$(1 - B)(1 - \phi_1)Y_t = e_t \tag{3.70}$$

όπου  $(1-B)Y_t$  είναι οι πρώτες διαφορές των αρχικών τιμών του Πίνακα 3.13 εκφρασμένοι σε όρους αποκλίσεων. Ο σκοπός της εκτίμησης είναι η εύρεση των εκτιμητών της παραμέτρου που ελαχιστοποιούν το MSE. Η διαδικασία επαναλαμβάνεται όπου οι τελικές τιμές των εκτιμήσεων των παραμέτρων είναι σημαντικά διαφορετικές από τις αρχικές τιμές της διαδικασίας εκτίμησης. Όμως, συνήθως, οι εκτιμήσεις θα συγκλίνουν σε μια άριστη τιμή για τις παραμέτρους μετά από σχετικά μικρό αριθμό επαναλήψεων. Εάν ο αλγόριθμος αποτύχει να συγκλίνει μετά ένα σημαντικό αριθμό επαναλήψεων, τότε μια εξέταση των δοκιμαστικών τιμών θα δείξει την κατεύθυνση των αλλαγών που πρέπει να γίνουν, και επομένως μπορούν να γίνουν νέες αρχικές εκτιμήσεις.

Για τα δεδομένα του δείγματος του ΠΠ, ένα μοντέλο AR (1) επιλέγεται δοκιμαστικά και χρησιμοποιείται η αρχική υπόθεση  $\phi_1 = 0$ . Το πρόγραμμα συγκλίνει μετά από πέντε επαναλήψεις που δίνουν ένα δοκιμαστικό μοντέλο με  $\phi_1 = 0.3063$  (AR 1).

Η τελική εκτίμηση του μοντέλου είναι 0.3063, η τυπική απόκλιση είναι 0.1041 και η t-αναλογία

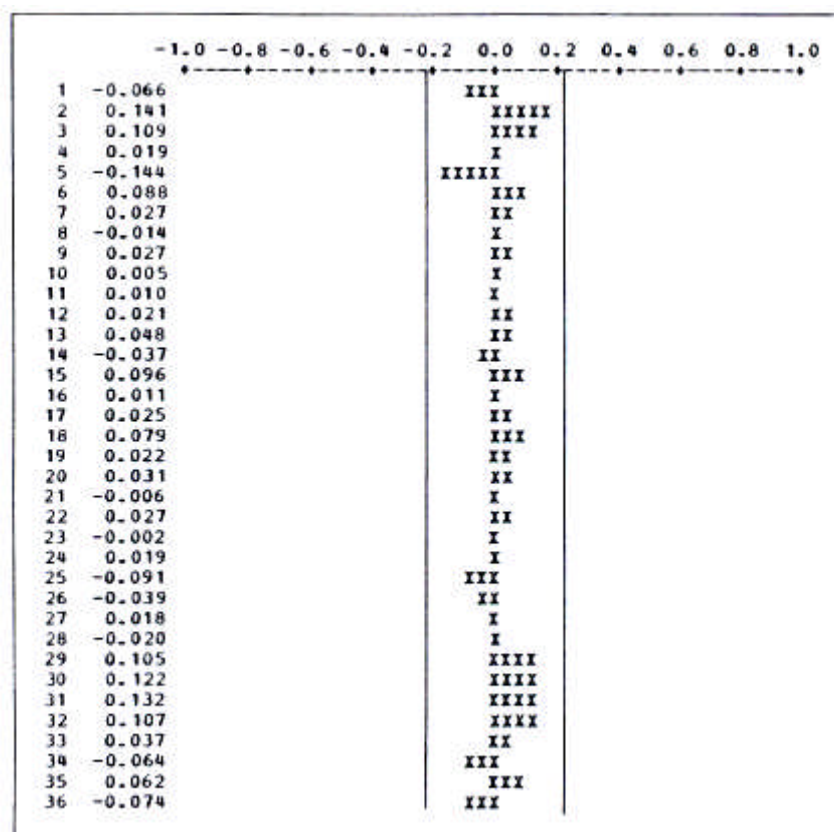
## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

(πρόκειται για το λόγο της εκτίμησης προς την τυπική απόκλιση) είναι 2.947. Το σφάλμα μέσου τετραγώνου, που συμβολίζεται ως MS, είναι 2.813. Τέλος, το αποτέλεσμα στον υπολογιστή δίνει προβλέψεις με 95 τοις εκατό όρια εμπιστοσύνης για τις περιόδους 91 έως 95.

**Στάδιο 3: Διαγνωστικός έλεγχος.** Για να ελέγξουμε την καταλληλότητα του μοντέλου, εκτιμούμε και κάνουμε το γράφημα των αυτοσυσχετίσεων των καταλοίπων, για να καθορίσουμε εάν είναι σημαντικά διάφορες του μηδενός.

Εξετάζουμε το σχήμα 3.18

**Σχήμα 3.18** Συντελεστές αυτοσυσχετίσης για τα κατάλοιπα του εκτιμημένου μοντέλου ARIMA (1,0,0) (αποτελέσματα σύμφωνα με το πρόγραμμα Minitab)



Τα όρια των διαστημάτων εμπιστοσύνης των δύο σίγμα είναι

$$-2 \frac{1}{\sqrt{89}} \leq r \leq 2 \frac{1}{\sqrt{89}}$$

$$-0.212 \leq r \leq +0.212$$

Επομένως, καμία από τις αυτοσυσχετίσεις δεν είναι σημαντικά διάφορη του μηδενός (στα όρια των δύο σίγμα), πράγμα που σημαίνει πως το μοντέλο είναι επαρκές. Τέλος, το εύρος αυτών των

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

αυτοσυσχετίσεων είναι από -0.144 ως +0.141 και δεν δείχνει την ύπαρξη κάποιου ιδιαίτερου προτύπου.

Ο δεύτερος έλεγχος για την επάρκεια του μοντέλου είναι ο έλεγχος Ljung-Box ( $Q$  test). (Εξίσωση 3.69).

Χρησιμοποιώντας τις πρώτες  $k = 30$  αυτοσυσχετίσεις για τον έλεγχο, η μηδενική και η εναλλακτική υπόθεση είναι:

$$H_0 = \text{Το μοντέλο είναι επαρκές}$$

$$H_A = \text{Το μοντέλο είναι ανεπαρκές}$$

Ο αριθμός των βαθμών ελευθερίας (degrees of freedom =  $d.f$ ) είναι:  $d.f = k - p - q = 30 - 1 = 29$

Από την τιμή του  $X^2$  στον πίνακα σε  $d.f = 29$ , βρίσκουμε  $X^2 = 42.557$  σε επίπεδο σημαντικότητας 0.05. Η υπολογισθείσα τιμή είναι

$$Q = 90(92) \left( \frac{(-0.066)^2}{89} + \frac{(+0.141)^2}{88} + \dots + \frac{(-0.122)^2}{60} \right) = 90(92) * (0.001 * 6750) = 13.8757$$

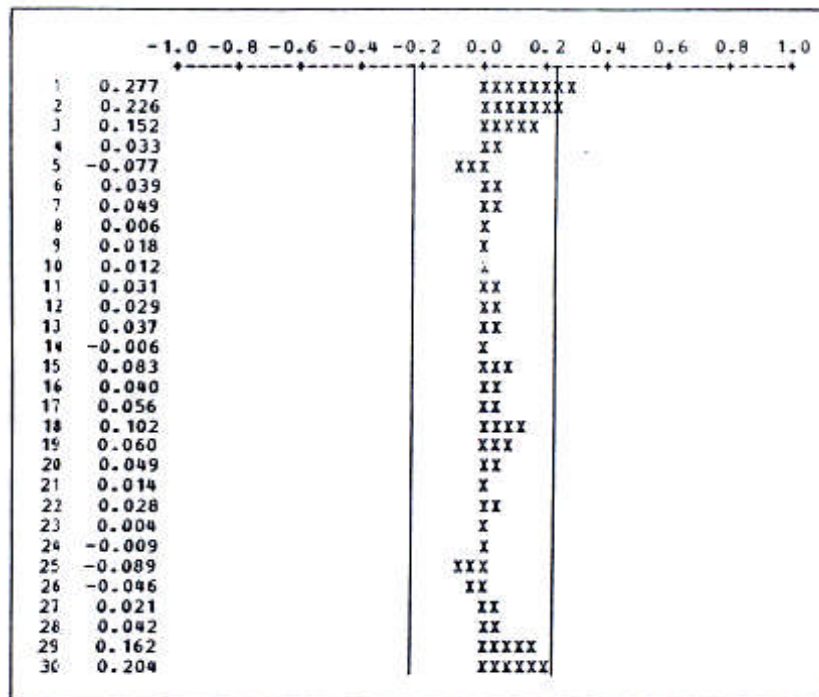
Εφόσον η υπολογισθείσα τιμή είναι μικρότερη από την τιμή του πίνακα, μπορούμε να δεχθούμε τη μηδενική υπόθεση στο επίπεδο 0.05 και να καταλήξουμε στο συμπέρασμα ότι το μοντέλο είναι επαρκές.

Για λόγους σύγκρισης, θα συγκρίνουμε το μοντέλο ARIMA (1,1,0) που υπολογίστηκε παραπάνω, με ένα μοντέλο ARIMA (1,0,0) (σχήμα 3.19). Με άλλα λόγια, θα συγκρίνουμε ένα μοντέλο AR (1), με διαφόριση βαθμού ένα, με ένα μοντέλο AR (1) χωρίς καμία διαφόριση. Ο λόγος είναι ότι πρέπει να καθορίσουμε ποιο από τα μοντέλα είναι το πλέον επαρκές.



ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

Σχήμα 3.19: Συντελεστές Συσχέτισης για τα κατάλοιπα εκτιμημένου μοντέλου ARIMA(1,0,0)  
(αποτελέσματα σύμφωνα με το πρόγραμμα Minitab)



**ΠΙΝΑΚΑΣ 3.14** Σύγκριση δύο μοντέλων AR(1)

		ARIMA(1.0.0) χωρίς διαφορίση	ARIMA(1.1.0) βαθμός διαφορίσης 1
Ljung-Box(Q)		29.8159	13.8757
d.f.		29	29
Σφάλματα Περίοδος	Περίοδος στο μέλλον		
91	1	-2.36	-3.40
92	2	-7.01	-8.34
93	3	-9.48	- 10.84
94	4	-10.93	- 12.28
95	5	-12.79	- 14.09
Διακύμανση καταλοίπων (MSE)		3.073	2.813
Παράμετρος AR		0.9996	0.3063
i -τιμή		767.26	2.94

Ο Πίνακας 3.14 συγκρίνει τα αποτελέσματα των δύο μοντέλων AR (1). Η τιμή του  $Q$  (το στατιστικό των Ljung-Box) είναι μικρότερη για το διαφορισμένο μοντέλο (1,1,0) από ό,τι για το μοντέλο που δεν έχει υποστεί διαφορίση. Το διαφορισμένο μοντέλο εμφανίζεται πιο κατάλληλο σύμφωνα με αυτό το κριτήριο. Εάν τα σφάλματα πρόβλεψης, για τις περιόδους από 1 έως 5 στο μέλλον, ελεγχθούν το μοντέλο ARIMA (1,0,0) έχει ελαφρά εγγύτερες εκτιμήσεις. Επίσης, η διακύμανση στα κατάλοιπα (MSE) είναι μικρότερη για το μοντέλο AR (1,1,0). Ωστόσο, το πιο σημαντικό είναι ότι ο συντελεστής για το μοντέλο ARIMA (1,0,0) είναι 0.9996. Η τιμή αυτή σχεδόν παραβιάζει τη συνθήκη σταθερότητας και, επομένως, το μοντέλο που δεν έχει διαφοριστεί δεν είναι αποδεκτό για πρόβλεψη.

Τέλος, ο ACF για τα κατάλοιπα από το μοντέλο (1,0,0) δείχνει ότι τα  $r_1$  και  $r_2$  βρίσκονται εκτός των δύο ορίων του τυπικού σφάλματος. Επομένως, το μοντέλο δεν παρέχει ενδείξεις σταθερότητας στις προβλέψεις.

**Στάδιο 4: Πρόβλεψη:** Προβλέπουμε για πέντε στάδια στο μέλλον, χρησιμοποιώντας το παραπάνω μοντέλο ARIMA (1,1,0) με τιμή έναρξης την παρατήρηση 31.

Αρχικά, επαναδιατυπώνουμε το εκτιμημένο μοντέλο πρόβλεψης:

$$(1 - B)(1 - \phi_1)Y_t = e_t \quad (3.71)$$

Σημειώστε τη χρήση του τελεστή τής προς τα πίσω μετατόπισης (backshift) για την περιγραφή των πρώτων διαφορών  $(1-B)$  και την AR-αναλογία του μοντέλου  $(1-\phi_1B)$ . Οι όροι μπορούν να πολλαπλασιαστούν και να επανατοποθετηθούν ως εξής:

$$\begin{aligned} (1 - \phi_1 B - B + \phi_1 B^2)Y_t &= e_t \\ Y_t &= Y_{t-1} + \phi_1(Y_{t-1} - Y_{t-2}) + e_t \end{aligned} \quad (3.72)$$

Μπορούμε τώρα να αντικαταστήσουμε το  $\phi_1 = 0.2178$  στην εξίσωση, καθώς και τα  $Y_{t-1}$  και  $Y_{t-2}$ . Η πρόβλεψη για την περίοδο 91, γράφεται:

$$\hat{Y}_{91} = Y_{90} + \phi_1(Y_{90} - Y_{89})$$

με  $e_t$  ίσο με μηδέν. Μετά από αντικατάσταση, παίρνουμε:

$$\hat{Y}_{91} = 131.2 + 0.3063(131.2 - 134.8) = 130.097$$

Με όμοιο τρόπο, μπορούμε να προβλέψουμε για αρκετές μελλοντικές περιόδους.

Το αποτέλεσμα αυτών των προβλέψεων για την πρώτη περίοδο είναι συνάρτηση της πραγματικής παρατήρησης της τελευταίας περιόδου και του σφάλματος της τελευταίας περιόδου. Όλες οι μελλοντικές προβλέψεις βασίζονται μόνο σε τιμές του  $Y_t$  που έχουν προβλεφθεί εφόσον οι μελλοντικές τιμές του  $e_t$  είναι άγνωστες.

Τώρα μπορούμε να υπολογίσουμε το σφάλμα πρόβλεψης και τη διακύμανσή του για το μοντέλο ARIMA (1,1,0) για να πάρουμε το διάστημα εμπιστοσύνης μιας πρόβλεψης. Όπως θα δούμε, το διάστημα εμπιστοσύνης για το  $Y_t$ , σχετίζεται με το διάστημα εμπιστοσύνης της πρόβλεψης για τη διαφορισμένη χρονοσειρά  $B Y_t$ .

Αρχικά, θεωρούμε το σφάλμα για την πρόβλεψη μιας περιόδου μπροστά  $Y_t(1)$ :

$$e_t(1) = Y_{t+1} - \hat{Y}_t(1) \quad (3.73)$$

όπου το σύμβολο  $\hat{\phantom{x}}$  δείχνει την εκτιμημένη τιμή. Με τη σειρά του, εάν

$$Z_t = Y_t - Y_{t-1}$$

τότε

$$e_t(1) = Y_t + Z_{t+1} - \hat{Z}_t(1) = e_{t+1} \quad (3.74)$$

το οποίο έχει διακύμανση  $\sigma_e^2$ . Η πρόβλεψη για δύο περιόδους στο μέλλον δίνεται από:

$$\begin{aligned}
 e_t(2) &= Y_{t+2} - \hat{Y}_t(2) = \\
 &= Y_t + Z_{t+1} + Z_{t+2} - Y_t - \hat{Z}_t(1) - \hat{Z}_t(2) = \\
 &= [Z_{t+1} - \hat{Z}_t(1)] + [Z_{t+2} - \hat{Z}_t(2)] = \\
 &= (1 + \phi_1)e_{t+1} + e_{t+2}
 \end{aligned} \tag{3.75}$$

και αυτό έχει διακύμανση

$$E[e_t^2(2)] = \sigma_e^2 [(1 + \phi_1)^2 + 1] \tag{3.76}$$

Σημειώστε ότι το σφάλμα πρόβλεψης και η διακύμανση του διαμορφώνονται σωρευτικά. Το σφάλμα πρόβλεψης είναι ίσο με το σφάλμα των δύο περιόδων για  $Z_t(2)$ , επιπλέον από το σφάλμα μιας περιόδου για  $\hat{Z}_t(1)$ . Επομένως, το σφάλμα στο  $\hat{Y}_t(2)$  είναι μια συσσώρευση των σφαλμάτων πρόβλεψης στο  $\hat{Z}_t(1)$  και  $\hat{Z}_t(2)$ .

$$e_t(n) = \sum_{i=1}^n e_{t+i}(n) \sum_{j=0}^{n-i} \phi_1^j \tag{3.77}$$

και αυτό έχει διακύμανση

$$E[e_t^2(n)] = \sigma_e^2 \sum_{i=1}^n \left( \sum_{j=0}^{n-i} \phi_1^j \right)^2 \tag{3.78}$$

Όπως αναμενόταν, το σφάλμα στην  $\hat{Y}_t(n)$  είναι μια συσσώρευση σφαλμάτων στις  $Z_t(1)$ ,  $Z_t(2)$ , ...,  $Z_t(n)$ . Επομένως, το διάστημα εμπιστοσύνης για την  $\hat{Y}_t(n)$  αυξάνει ραγδαία, δεδομένου ότι περιλαμβάνει τη συσσώρευση των σφαλμάτων πρόβλεψης στη διαφορισμένη χρονοσειρά. Εάν επιστρέψουμε στο σχήμα 3.18, θα δούμε ότι το εύρος του διαστήματος εμπιστοσύνης στο 95 τοις εκατό γίνεται μεγαλύτερο, καθώς αυξάνεται ο αριθμός των μελλοντικών περιόδων πρόβλεψης.

Συμπερασματικά, αν και εξετάσαμε ιδιότητες μόνο ενός απλού μοντέλου ARIMA, μερικά από τα συμπεράσματά μας μπορούν να εφαρμοστούν σε πιο σύνθετα (υψηλότερης τάξης) μοντέλα ARIMA. Ιδιαίτερα, ένα μοντέλο MA, τάξης  $q$ , έχει μνήμη μόνο για  $q$  περιόδους, εφόσον υπάρχουν μόνο  $q$  όροι σφάλματος. Επομένως, τα παρατηρηθέντα δεδομένα θα επιδράσουν στην πρόβλεψη μόνο εάν ο χρόνος προπόρευσης (lead time) είναι μικρότερος από  $q$ . Ένα μοντέλο AR έχει μνήμη απείρου μήκους, επομένως, όλες οι προηγούμενες παρατηρήσεις θα έχουν κάποια επίδραση στην πρόβλεψη κι αυτό είναι αληθές όταν το  $n$  είναι μεγάλο. Μόνο οι πρόσφατες παρατηρήσεις, ωστόσο, θα έχουν μεγάλη επίδραση. Συνεπώς, και στο μοντέλο AR και στο μικτό ARMA, οι προηγούμενες παρατηρήσεις έχουν μόνο ελάχιστες επιδράσεις στην πρόβλεψη εάν ο χρόνος προπόρευσης είναι πολύ μακρύς. Μπορούμε να καταλήξουμε ότι τα μοντέλα ARIMA είναι κατάλληλα μόνο για βραχυχρόνια πρόβλεψη, δηλαδή, όπου ο προπορευμένος χρόνος δεν είναι μεγαλύτερος από  $p+q$ .

Τέλος, οι προβλέποντες παρατηρούν ότι στην πράξη είναι πολύ χρήσιμο να συγκρίνουν προβλέψεις αρκετών συναγωνιστικών εκτιμημένων μοντέλων χρονοσειρών, όπως έγινε στον Πίνακα 3.14. Κι αυτό ισχύει κυρίως εάν ο προβλέπων συναντά δυσκολίες στην επιλογή ανάμεσα σε μια

ποικιλία πιθανών μοντέλων. Συχνά, η επιλογή δεν είναι κρίσιμη εάν οι προβλέψεις είναι παρόμοιες. Εάν δεν είναι, τότε είναι απαραίτητο να γίνει προσπάθεια προκειμένου να υπάρξει ομοιότητα. Ο καθορισμός αυτός θα βοηθήσει στην επιλογή μεταξύ των μοντέλων. Για παράδειγμα, θα επιλέγαμε το μοντέλο ARIMA (1,0,0) εφόσον ο συντελεστής  $\phi$  δεν ικανοποιεί το κριτήριο της σταθερότητας. Τουλάχιστον 50 ή και περισσότερες παρατηρήσεις απαιτούνται σε μια εκτίμηση των Box-Jenkins. Για ένα εποχικό μοντέλο, είναι επιθυμητές ακόμα περισσότερες παρατηρήσεις.

### 3.3.5 Πρόβλεψη Χρονοσειρών με Εποχιακότητα

Πολλά δεδομένα οικονομικών χρονοσειρών που είναι διαθέσιμα για τις τράπεζες δεδομένων υπολογιστών, καθώς και δημοσιεύσεις, είναι εποχικά προσαρμοσμένα. Αυτές οι χρονοσειρές χρησιμοποιούνται σε μοντέλα για την πρόβλεψη συνιστωσών του κύκλου και της τάσης. Οικονομετρικά μοντέλα παλινδρόμησης χρησιμοποιούνται επίσης εκεί όπου τα εποχικά πρότυπα άλλως θα κάλυπταν τις πληροφορίες που ενδιαφέρουν τον προβλέποντα. .

Μπορεί να θεωρείται επιθυμητή η πρόβλεψη εποχικών χρονοσειρών που δεν είναι προσαρμοσμένες για εποχιακότητα. Η εποχική τη τα ποικίλλει από χρόνο σε χρόνο δείχνοντας ότι μοντέλα που βασίζονται σε μη προσαρμοσμένα δεδομένα είναι πολύ πιθανό να είναι περισσότερο ευέλικτα ή χρήσιμα. Αν αναμένεται μεταβαλλόμενη εποχιακότητα, είναι συνήθως προτιμότερο να λαμβάνεται αυτό υπόψη μέσω της ανάπτυξης ενός καταλλήλως εξειδικευμένου μοντέλου ARIMA. Για χρονοσειρές οι οποίες περιέχουν μια εποχική συνιστώσα που επαναλαμβάνεται μετά από κάθε \$ παρατήρηση, μπορεί να χρησιμοποιηθεί ένα βοηθητικό μοντέλο στο μη-εποχικό μοντέλο ARIMA. Για μια εποχική χρονοσειρά με περίοδο  $s$ , το  $s=12$  αφορά μηνιαία δεδομένα και  $s=4$  αφορά τριμηνιαία δεδομένα. Τα εποχικά μοντέλα ARIMA  $(P,D,Q)^s$  μπορούν να αναπτυχθούν για εποχικά δεδομένα με τρόπο παρόμοιο της αρχικής διαδικασίας ARIMA.

Σε ένα πολλαπλασιαστικό μοντέλο ARIMA, οι κανονικές και οι εποχικές συνιστώσες αυτοπαλινδρόμησης, οι διαφορές και οι συνιστώσες κινητών μέσω των πολλαπλασιάζονται μαζί στο γενικό μοντέλο. Συχνά, οι περισσότερες από τις τιμές  $p, d, q$  είναι 0 στις πρακτικές απεικονίσεις και τα μοντέλα που λαμβάνονται είναι συχνά πολύ απλά ή φειδωλά.

Ένας πολύ χρήσιμος συμβολισμός που περιγράφει τις τάξεις των διαφόρων συνιστωσών στο πολλαπλασιαστικό μοντέλο, δίδεται από:

$$(p,d,q) \times (P,D,Q)^s$$

και ανταποκρίνεται στις τάξεις των κανονικών και των εποχικών παραγόντων αντίστοιχα.

Αναπαριστώντας μια χρονοσειρά σε όρους ενός πολλαπλασιαστικού μοντέλου, είναι συχνά δυνατό να μειώνεται ο αριθμός των παραμέτρων που πρόκειται να εκτιμηθούν. Επίσης, βοηθά στην ερμηνεία της δομής του μοντέλου.

Θα θεωρούμε ένα μηνιαίο εποχικό μοντέλο που έχει εποχική διαφύριση (σε μήνες) και επίσης,

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

κανονικές και εποχικές παραμέτρους αυτοπαλινδρόμησης πρώτης τάξης. Επομένως,

$$(p,d,q) \times (P,D,Q)^s = (1,0,0) \times (1,1,0)^{12}$$

Σε όρους του  $Y_t$ , έχουμε

$$Y_t = \phi Y_{t-1} + (1 + \phi_{12}) Y_{t-12} - \phi_1 (1 + \phi_{12}) Y_{t-13} - \phi_{12} Y_{t-24} + \phi_1 \phi_{12} Y_{t-25} + e_t \quad (3.80)$$

Υπενθυμίζεται ότι η χρήση του τελεστή της προς τα πίσω μετατόπισης (backshift) μας επιτρέπει να εκφράζουμε την τάξη και τον τύπο της διαδικασίας σε όρους του εκθέτη του  $B$  και του τύπου της παραμέτρου.

Εάν μελετάμε εποχικά δεδομένα, μια διαδικασία ARMA μπορεί να αποτελείται από δυο μέρη:

1. Το κανονικό μέρος (αναλογία).
2. Το εποχικό μέρος (αναλογία).

Για σκοπούς που εξυπηρετούν την ταυτοποίηση μιας εποχικής διαδικασίας ARMA, διαιρούμε τη διαδικασία σε δύο τμήματα. Για την ταυτοποίηση του εποχικού προτύπου αγνοούμε τη μη-εποχική διαδικασία που εξετάζουμε εάν η εποχιακότητα ορίζεται ή όχι από μια διαδικασία AR ή MA, συγκεντρώνοντας την προσοχή μας στους συντελεστές των εποχικών όρων. Αν υποθέσουμε ότι το μη-εποχικό τμήμα είναι ένα ARIMA (1,0,1) και ότι η χρονοσειρά δείχνει ένα τριμηνιαίο εποχικό πρότυπο, τότε το πλήρες μοντέλο γράφεται:

$$(1 - \phi_1 B)(1 - \phi_4 B^4) Y_t = (1 - \theta_1 B) e_t \quad (3.81)$$

εάν η εποχιακότητα βρίσκεται στο μέρος (αναλογία) AR, ή

$$(1 - \phi_1 B) Y_t = (1 - \theta_1 B)(1 - \theta_4 B^4) e_t \quad (3.82)$$

εάν η επόχικότητα βρίσκεται στο μέρος MA, όπου

$$(1 - \phi_4 B^4) Y_t = Y_t - \phi_4 B^4 Y_t \quad (3.83)$$

Με όμοιο τρόπο

$$(1 - \theta_4 B^4) e_t = e_t - \theta_4 B^4 e_t = e_t - \theta_4 e_{t-12} \quad (3.84)$$

Επομένως, τα παραπάνω μοντέλα περιλαμβάνουν εποχικές παραμέτρους που χρησιμοποιούν τα  $Y_t$  ή  $e_t$  των προηγούμενων τεσσάρων τριμήνων ώστε η εποχιακότητα να λαμβάνεται υπόψη.

Σε μια εποχική διαδικασία MA(2) με ένα επίπεδο διαφορίσης, έχουμε την εξής έκφραση:

$$(1 - B)(1 - B^4)Y_t = (1 - \theta_1 B - \theta_2 B^2)e_t \quad (3.85)$$

(διαφόριση ενός επιπέδου)\*(εποχικό πρότυπο)=(MA(2) μη εποχικό)

$$(1 - B)Y_t = (1 - \theta_1 B - \theta_2 B^2)(1 - \theta_4 B^4)e_t \quad (3.86)$$

(διαφόριση ενός επιπέδου)=(MA(2) μη εποχικό) \*(εποχικές παράμετροι στο MA)

**3.4 ΠΡΟΒΛΕΨΕΙΣ ΜΕ ΧΡΗΣΗ ΑΣΑΦΟΥΣ ΛΟΓΙΚΗΣ (FUZZY LOGIC) [20] – [30]****3.4.1 Μοντέλο Πρόβλεψης Βασισμένο σε Ασαφή Λογική [20]****3.4.1.1 Εισαγωγή**

Η ασαφής θεωρία συνόλων και η ασαφής λογική παρέχουν μια γενική μέθοδο διαχείρισης των αβέβαιων και ασαφών πληροφοριών, οι οποίες είναι δυστυχώς αναπόφευκτες σε πολλές διαδικασίες λήψης αποφάσεων. Συνήθως οι αβέβαιες και ασαφείς πληροφορίες σημαίνουν ότι η λήψη απόφασης πρέπει να γίνει με σχετικά μη εξακριβωμένες, ασυμβίβαστες, ακόμη και ξεπερασμένες πληροφορίες, χωρίς οποιαδήποτε επίσημη προσέγγιση.

Παραδείγματος χάρη, το κύριο πρόβλημα στην αγορά υλικών για την κατασκευή προϊόντων επικοινωνιών για δεδομένα και κυψελοειδή συστήματα εναπόκειται σε χρονοδιαγράμματα. Ο χρόνος που απαιτείται για να αποκτηθεί το υλικό για την κατασκευή (καθυστερήσεις παραγγελίας και παράδοσης) είναι πολύ πιο μεγάλος από τον αποδεκτό χρόνο μεταξύ της παραγγελιάς και της παράδοσης για το τελικό προϊόν. Αυτό είναι ένα κλασσικό πρόβλημα κατασκευής, αλλά δεν μπορούμε απλά να το λύσουμε αποθηκεύοντας προϊόντα τηλεπικοινωνιών. Και αυτό επειδή τα προϊόντα ανανεώνονται πολύ γρήγορα (ο μέσος κύκλος για τα νέα προϊόντα είναι λιγότερο από 12 μήνες και επίσης τα εξαρτήματα εξελίσσονται εξαιρετικά γρήγορα), και υπάρχουν πολύ μεγάλες διακυμάνσεις στις παραγγελίες.

Επομένως φαίνεται να είναι απαραίτητο να αποδοθεί το αίτημα των πελατών στις συγκεκριμένες περιοχές αγοράς, προκειμένου να προγραμματιστεί η διαδικασία αγοράς υλικών σωστά. Το μοντέλο πρόγνωσης ή πρόβλεψης χρησιμοποιούμενο για αυτόν το λόγο πρέπει να επεξεργαστεί τις ιδιαίτερα ασαφείς και αβέβαιες πληροφορίες.

Ο μόνος τρόπος να κατασκευασθεί αυτό το είδος μοντέλου είναι να χρησιμοποιηθεί μια τεχνική που από τη φύση της να έχει τις ικανότητες για την εκτέλεση αυτού του τύπου επεξεργασίας πληροφοριών. Η τεχνική ασαφούς λογικής είναι σε θέση να λάβει υπόψη την κατά προσέγγιση φύση τέτοιων πληροφοριών. Επιπλέον είναι ένα πολύτιμο εργαλείο για να λύσει τις δυσκολίες που προκύπτουν από τις ανεπαρκείς πληροφορίες.

Σε αυτή την ενότητα, θα εισάγουμε ένα προσδόκιμο συμβουλευτικό εργαλείο ασαφούς λογικής (fuzzy logic advisory tool -FLAT) . Ένα προσδόκιμο συμβουλευτικό εργαλείο ασαφούς λογικής διαφέρει από ένα παραδοσιακό ασαφές σύστημα στο ότι μόλις χρησιμοποιηθούν οι ασαφείς κανόνες για να παραγάγουν μια τιμή εξόδου για το σύστημα, μια διαδικασία πρόβλεψης καλείται να προβλέψει την επίδραση αυτής της τιμής στην έξοδο του συστήματος. Το FLAT είναι ένα εργαλείο λογισμικού που έχει αναπτυχθεί προκειμένου να προβλεφθεί και να ελεγχθεί η πολύ σύνθετη και ιδιαίτερα μη γραμμική σε συμπεριφορά αγορά υλικών, για έναν κατασκευαστή τηλεπικοινωνιών (περίπου 14.000 διαφορετικά εξαρτήματα με τα περισσότερα από αυτά από διαφορετικές πηγές, με διαφορετικούς χρόνους και δυνατότητες παράδοσης).



Η ανάπτυξη έχει γίνει με το συνδυασμό της παραδοσιακής ασαφούς θεωρίας με τη προσέγγιση γλωσσική εξίσωσης (Linguistic Equation-LE) που παρέχει μια μέθοδο για ανάπτυξη και συντονισμό προσαρμοσμένων ειδικών συστημάτων χωρίς προγραμματισμό βασισμένο σε κανόνες. Επίσης μελετήθηκαν οι παραδοσιακές μέθοδοι ανάλυσης χρονοσειρών, όπως οι ARMA και ARIMA, προκειμένου να αναλυθεί η χρονοσειρά.

### 3.4.1.2 Θεωρία Ασαφούς Συνόλου και Ασαφής λογική

Η θεωρία ασαφούς συνόλου παρουσιάστηκε αρχικά από τον Zadeh στο πρωτογενές έγγραφο του "Ασαφή Σύνολα" (Zadeh, 1965). Εντούτοις, μερικές από τις ιδέες-κλειδιά διατυπώθηκαν ήδη από το 1937 από τον Black. Η ασαφής λογική αναπτύχθηκε αργότερα από τη θεωρία ασαφούς συνόλου, πρωτίστως για να δικαιολογήσει αβέβαιες και ασαφείς πληροφορίες και δευτερευόντως για να αντιπροσωπεύσει τη γνώση μέσα από μια πιο δυναμικά λειτουργικά μορφή.

Στη θεωρία ασαφούς συνόλου, το όνομα ασαφή σύνολα χρησιμοποιείται για να τα διακρίνουμε από τα σαφή σύνολα της συμβατικής θεωρίας συνόλων. Η χαρακτηριστική λειτουργία ενός σαφούς συνόλου  $C$ ,  $\mu_C(u)$ , το οποίο ορίζει μια ιδιαίτερη τιμή (συνήθως 0 ή 1) σε κάθε στοιχείο  $u$  στο ολικό σύνολο  $U$ , δηλ., αυτό διακρίνει τα μέλη και τα μη μέλη του σαφούς συνόλου (έπειτα για κάθε στοιχείο  $u$  του συνόλου  $U$ , είτε  $u \in C$  ή  $u \notin C$ ).

Η χαρακτηριστική λειτουργία μπορεί να γενικευτεί στη θεωρία ασαφούς συνόλου έτσι ώστε οι τιμές που δίνονται στα στοιχεία  $u$  από το ολικό σύνολο  $U$  να ανήκουν μέσα σε μια προκαθορισμένη κλίμακα (συνήθως στο διάστημα μονάδων  $[0,1]$  δείχνοντας το βαθμό ιδιότητας μέλους αυτών των στοιχείων στο ασαφές σύνολο  $F$ . Κατόπιν δεν είναι απαραίτητο ότι για κάθε  $u$  ισχύει  $u \in F$  ή  $u \notin F$ . Η γενικευμένη σχέση που χρησιμοποιείται καλείται εξίσωση μέλους και το σύνολο που καθορίζεται με βάση αυτήν είναι αντίστοιχα, ένα ασαφές σύνολο. Η εξίσωση μέλους καθορίζει για κάθε  $u \in U$  μια τιμή από το διάστημα τιμών  $[0,1]$  αντί του διαστήματος  $\{0,1\}$ .

#### Τυποποιημένες διαδικασίες

Τυποποιημένες διαδικασίες της θεωρίας ασαφούς συνόλου για το χειρισμό των ασαφών συνόλων φαίνονται στις παρακάτω εξισώσεις

$$\mu_{\bar{F}}(u) = I - \mu_F(u) \quad (3.87)$$

$$\mu_{F_1 \dot{\cup} F_2}(u) = S^*(\mu_{F_1}(u), \mu_{F_2}(u)) \quad (3.88)$$

$$\mu_{F_1 \cap F_2}(u) = T^*(\mu_{F_1}(u), \mu_{F_2}(u)) \quad (3.89)$$

Παράδειγμα για την χρήση των παραπάνω εξισώσεων δίνεται στο τμήμα 4.

### 3.4.1.3 Συμπέρασμα του βαθμού ιδιότητας μέλους.

Η εξαγωγή συμπεράσματος για το βαθμό ιδιότητας μέλους μπορεί να γίνει είτε χρησιμοποιώντας το βασισμένο σε σύνθεση συμπέρασμα είτε το βασισμένο σε κάθε μέλος συμπέρασμα. Στο βασισμένο σε σύνθεση συμπέρασμα όλοι οι κανόνες συνδυάζονται σε μια ρητή σχέση και έπειτα τροφοδοτούνται με ασαφή δεδομένα εισαγωγής, ενώ το βασισμένο σε κάθε μέλος συμπέρασμα κάθε ένας κανόνας τροφοδοτείται χωριστά με την σαφή δεδομένα εισαγωγής τα οποία συνδυάζονται έπειτα σε ένα γενικό ασαφές σύνολο. Στην παρούσα μελέτη μας χρησιμοποιούμε την εξαγωγή συμπερασμάτων βασισμένη σε κάθε μέλος χωριστά. Ο κύριος λόγος επιλογής αυτής της μεθόδου είναι η πολύ απλή διαδικασία εφαρμογής του.

Στο βασισμένο σε κάθε μέλος συμπέρασμα, ο βαθμός ιδιότητας μέλους του κάθε αποτελέσματος μπορεί να διαμορφωθεί με τη λήψη της T-norm τιμής (παραδείγματος χάριν το ελάχιστο των ασαφών συνόλων) από τους βαθμούς ιδιότητας μέλους των δεδομένων εισαγωγής.

Ο καθορισμός του είναι βασισμένος στη σχέση  $R_c$  (c για την σύζευξη που καθορίζεται από την T-norm).

$$\mu_{R_c}(x, y) = \min(\mu_A(x), \mu_B(y)) \quad (3.90)$$

όπου το x και το y δείχνουν τις μεταβλητές εισαγωγής ενώ το A και το B είναι έννοιες του x και του y, αντίστοιχα. Η έννοια ολόκληρου του συνόλου κανόνων δίνεται με τη λήψη της τιμής S-norm (παραδείγματος χάριν το μέγιστο των ασαφών συνόλων) του βαθμού (-ών) ιδιότητας μέλους, από τους κανόνες με την ίδια τιμή εξόδου, για να διαμορφωθεί το αποτέλεσμα μόνο με γλωσσολογικά διαφορετικές τιμές.

### 3.4.1.4 Γλωσσικό πλαίσιο εξισώσεων

#### Οι γλωσσικές σχέσεις

Στα συστήματα, όπου η γνώση θα εκφραζόταν σε μια γλωσσική ή λεκτική μορφή, μια προσανατολισμένη προς τη γλώσσα προσέγγιση μπορεί να χρησιμοποιηθεί σε μια παραγωγή μοντέλου (γλωσσικό μοντέλο). Στη προσανατολισμένη προς τη γλώσσα προσέγγιση αντιμετωπίζουμε μια έννοια των γλωσσικών σχέσεων, η οποία περιγράφει τους βαθμούς σχέσεων μεταξύ των ασαφών συνόλων που δίνονται σε γλωσσική μορφή.

Γλωσσικές εξισώσεις

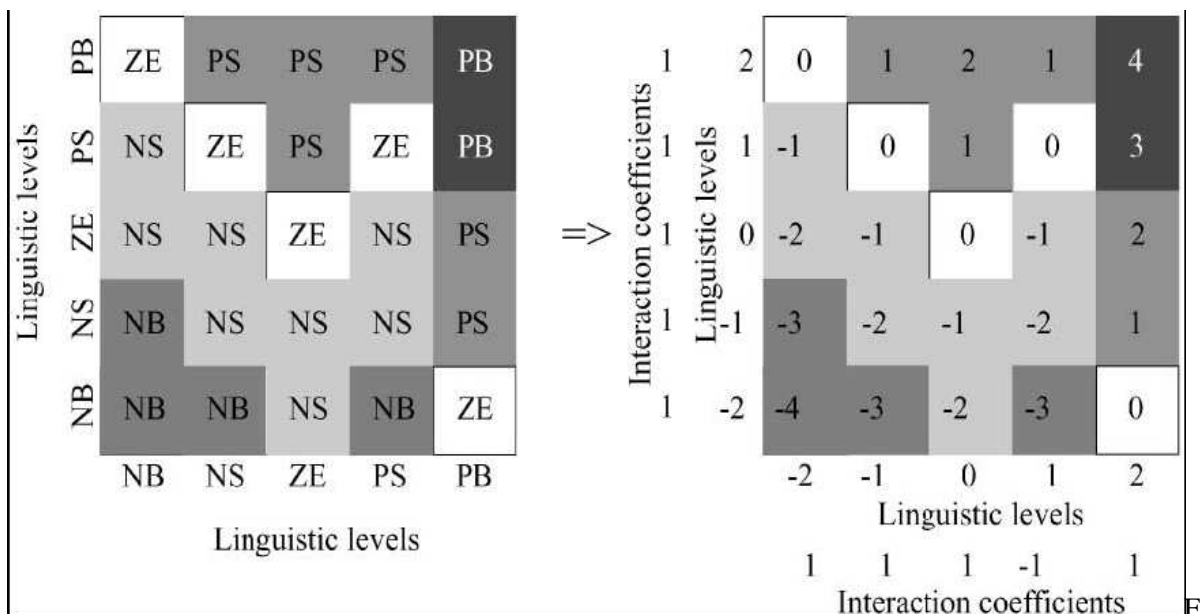
Στα πλαίσια των γλωσσικών εξισώσεων (LE) το γλωσσικό μοντέλο ενός συστήματος μπορεί να περιγραφεί από τις ομάδες γλωσσικών σχέσεων. Οι γλωσσικές σχέσεις διαμορφώνουν μια βάση κανόνων για το σύστημα, που μπορεί να μετατραπεί στις εξισώσεις μητρών. Υποθέτουμε, για παράδειγμα, ότι το  $X_j$  με  $j= 1, \dots, m$  (το  $m$  είναι περιττός αριθμός), είναι ένα γλωσσικό επίπεδο (π.χ. μεγάλος αρνητικός, μικρός αρνητικός, μηδέν, μικρός θετικός και μεγάλος θετικός) για μια μεταβλητή. Η γενίκευση για τα λεπτότερα ασαφή τμήματα εκτελείται εύκολα. Τα γλωσσικά επίπεδα (κάθε ένα από τα οποία περιγράφεται από ένα διαφορετικό ασαφές σύνολο) αντικαθίστανται από τους ακέραιους αριθμούς  $-(j-1)/2, \dots, -2, -1, 0, 1, 2, \dots, (j-1)/2$ .

Η κατεύθυνση της αλληλεπίδρασης μεταξύ των ασαφών συνόλων αντιπροσωπεύεται από τους συντελεστές  $A_{ij} \in \{-1, 0, 1\}$ ,  $i=1, \dots, m$ . Οι μεταβλητές των σχέσεων επιλέγονται κατά τέτοιο τρόπο ώστε οι κατευθύνσεις των αλλαγών να είναι ισορροπημένες. Αυτό σημαίνει ότι οι κατευθύνσεις των αλλαγών στη μείωση ή την αύξηση της μεταβλητής εξόδου, εξαρτάται από τις κατευθύνσεις των αλλαγών στις μεταβλητές εισαγωγής. Κατά συνέπεια μια εξίσωση είναι η

$$\sum_{i,j=1}^m A_{ij} X_j = 0 \tag{3.91}$$

Ο συλλογισμός είναι βασισμένος σε αυτές τις εξισώσεις ή στα ενοποιημένα σύνολα γλωσσικών σχέσεων που λαμβάνονται με την επίλυση των εξισώσεων. Αυτό βοηθά στην παρουσίαση και επίλυση των ογκωδών βάσεων κανόνα. Η απεικόνιση των γλωσσικών σχέσεων από τις γλωσσικές εξισώσεις περιγράφεται στο σχήμα 3.20. Με τη χρησιμοποίηση των γλωσσικών εξισώσεων (LE), μπορούν να συνδυαστούν οι διαφορετικές προσεγγίσεις και οι μέθοδοι διαμόρφωσης μοντέλων.

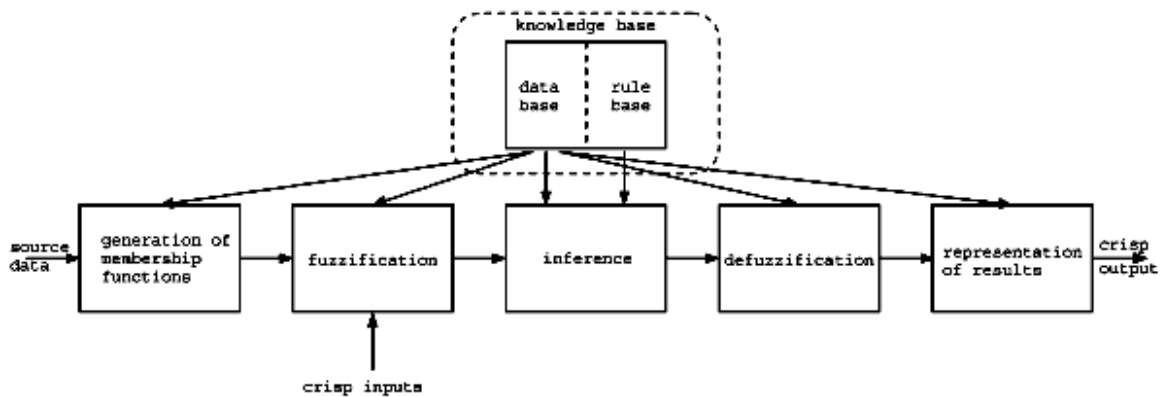
**Σχήμα 3.20** Η απεικόνιση των γλωσσικών σχέσεων από τις γλωσσικές εξισώσεις



3.4.15 Η δομή του συμβουλευτικού εργαλείου ασαφούς λογικής (Fuzzy Logic Advisory Tool-FLAT)

Το λογισμικό FLAT αποτελείται από τέσσερις βασικές ενότητες (βλ. σχήμα 3.21): ενότητα δημιουργίας λειτουργιών ιδιότητας μέλους (membership functions generation module-MGM), ενότητα "δημιουργίας ασάφειας" (fuzzification module-FM), ενότητα συμπεράσματος (inference module-IM) και ενότητα αναίρεσης ασάφειας (defuzzification module-DM).

Σχήμα 3.21 Συμβουλευτικό εργαλείο ασαφούς λογικής (fuzzy logic advisory tool-FLAT)



Οι τιμές πρόβλεψης των διαφορετικών προϊόντων προκύπτουν με την έναρξη από ένα σύνολο οκτώ εισαγομένων τιμών:

- (α) Ζήτηση που προβλέπεται από έναν πελάτη.
- (β) Ζήτηση που προβλέπεται από έναν εμπειρογνώμονα της αγοράς.
- (γ) Σταθμισμένη μέση τιμή της προηγούμενης ζήτησης.
- (δ) Τάση της προηγούμενης ζήτησης.
- (ε) Σταθμισμένη μέση τιμή των παραγγελιών.
- (στ) Τάση των παραγγελιών.
- (ζ) Σταθμισμένη μέση τιμή των προηγούμενων προβλέψεων.
- (η) Τάση των προηγούμενων προβλέψεων.

Ο πελάτης και ο εμπειρογνώμονας αγοράς στον εν λόγω τομέα αξιολογούν χωριστά τις ζήτησεις (εισαγωγές δεδομένων α και β) των διάφορων προϊόντων για τις επόμενες εβδομάδες σε εβδομαδιαία βάση. Υποβάλλονται στο μοντέλο από έναν χρήστη. Θα πρέπει να σημειωθεί ότι ακόμη και στο "μη εξελισσόμενο" περιβάλλον κατασκευής (δηλ. στο οποίο δεν επέρχεται καμία ταχεία τεχνολογική αλλαγή στο τελικό προϊόν) ο συνδυασμός προβλέψεων των εσωτερικών εμπειρογνομώνων και των προβλέψεων του πελάτη, αυξάνει σημαντικά τον αριθμό των πληροφοριών.

Η εισαγωγή δεδομένων ( $\gamma$ ) είναι σταθμισμένος μέσος όρος των προηγούμενων τιμών της ζήτησης ενώ η εισαγωγές δεδομένων ( $\epsilon$ ) και ( $\zeta$ ) είναι σταθμισμένοι μέσοι όροι των παραγγελιών και της πρόβλεψης. Οι τιμές είναι εκθετικά κατανεμημένες, έτσι ώστε οι πιο πρόσφατες παρατηρήσεις να έχουν μεγαλύτερο βάρος. Το χαρακτηριστικό σχήμα στάθμισης πρέπει να συντονίζεται από εμπειρογνώμονα και συνήθως τα αποτελέσματα αυτού αποτελούν εμπορικά μυστικά.

Τυπικά τα τελευταία 4-6 τέταρτα έχουν πάνω από 70% των συντελεστών βαρύτητας. Οι εισαγωγές δεδομένων ( $\delta$ ), ( $\sigma$ ) και ( $\zeta$ ) είναι αυτοπαλίνδρομοι κινούμενοι μέσοι όροι της ζήτησης, των παραγγελιών και των προβλέψεων, αντίστοιχα. Οι εισαγωγές δεδομένων ( $\gamma$ ) και ( $\zeta$ ) προσδιορίζονται αυτόματα από το μοντέλο από τις τιμές που υπάρχουν στις βάσεις δεδομένων κατασκευής του εκάστοτε εργοστασίου.

Κάθε τιμή εισαγωγής μετατρέπεται σε ασαφής από την ενότητα "fuzzification module". Παραδείγματος χάριν, η σταθμισμένη μέση τιμή της προηγούμενης ζήτησης του προϊόντος  $n=1000$  μονάδες θα μπορούσε να ερμηνευθεί σαν μικρός θετικός σε βαθμό ιδιότητας μέλους 0,71 και μεγάλος θετικός σε βαθμό ιδιότητας μέλους 0,29.

Έπειτα, τα ασαφή αποτελέσματα εισάγονται στην ενότητα συμπεράσματος σε τρεις διαδοχικές φάσεις. Σε κάθε φάση ο αριθμός μεταβλητών είναι χωρισμένος λόγω της ιεραρχικής δομής της ενότητας συμπεράσματος. Αυτή η μορφή δενδρικής δομής μειώνει τον αριθμό μεταβλητών έτσι ώστε τελικά να έχουμε μόνο την πρόβλεψη για την ζήτηση που μένει. Στην πρώτη φάση του ασαφούς συμπεράσματος οι τιμές εισαγωγής χρησιμοποιούνται ανά δύο για να περάσουν τέσσερις εκλεπτυσμένες τιμές εισαγωγής στη δεύτερη φάση συμπεράσματος. Στη δεύτερη φάση, οι τέσσερις εισηγμένες τιμές χρησιμοποιούνται ανά δύο στη διαδικασία συμπεράσματος για να παραγάγουν δύο τιμές που θα εισαχθούν στην τρίτη φάση. Η τρίτη φάση παράγει την τελική τιμή εξόδου που χρησιμοποιείται στη φάση "defuzzification". Μια βάση γνώσεων και μια βάση κανόνων διαφούνται αναλόγως σε τρία ιεραρχικά επίπεδα. Οι κανόνες αντιπροσωπεύονται από γλωσσικές σχέσεις που αλλάζουν σε μορφή εξισώσεων μητρών, προκειμένου να εφαρμοστεί το πλαίσιο γλωσσικής τεχνικής (LE) εξισώσεων. Ο προσδιορισμός των κανόνων γινόταν με ανάλυση της συμπεριφοράς και των ιδιοτήτων του συστήματος, συνδυάζοντας τη στατιστική ανάλυση με την ανθρώπινη νοημοσύνη των εμπειρογνομόνων μιας αγοράς.

Σαν παράδειγμα του συλλογισμού και της χρήσης των γλωσσικών εξισώσεων, υποθέτουμε ότι οι ακόλουθοι ασαφείς κανόνες εισάγονται στην πρώτη φάση (μέρος 1 στο σχήμα 3.21). Η αντίστοιχη γλωσσική εξίσωση παρουσιάζεται εντός παρενθέσεως (η σημείωση  $\lceil x \rceil$  δείχνει το μικρότερο ακέραιο αριθμό μεγαλύτερο ή ίσο του  $x$ ):

IF (εάν) η ζήτηση που προβλέπεται από έναν πελάτη IS (είναι) μεγάλος θετικός AND (και) η ζήτηση που προβλέπεται από έναν εμπειρογνώμονα αγοράς IS (είναι) μικρός θετικός THEN (τότε) η πρώτη έξοδος 1 της φάσης 1 IS (είναι) μεγάλος θετικός

$$((+I_1 x_2) + (+I_3 x_4) = \left\lceil \frac{3_s}{2_6} \right\rceil)$$

όπου

1. +1 δείχνει την κατεύθυνση της αλληλεπίδρασης (οι αλλαγές στις τιμές εισαγωγής έχουν επιπτώσεις προς την ίδια κατεύθυνση της εξόδου),
2. 2 δείχνει ασαφές σύνολο θετικά μεγάλο,
3. +1 δείχνει την κατεύθυνση της αλληλεπίδρασης (οι αλλαγές στις τιμές εισαγωγής έχουν επιπτώσεις προς την ίδια κατεύθυνση της εξόδου),
4. 1 δείχνει ασαφές σύνολο θετικά μικρό,
5. 3 δείχνει το άθροισμα του ασαφούς συνόλου, δηλ., 1+2, και
6. 2 είναι μια σταθερά εξαιτίας της διακύμανσης των αθροισμάτων στην κλίμακα των εισαγομένων (-2 → +2).

IF (εάν) η ζήτηση που προβλέπεται από έναν πελάτη IS (είναι) μικρός θετικός AND (και) η ζήτηση που προβλέπεται από έναν εμπειρογνώμονα αγοράς IS (είναι) μικρός θετικός THEN (τότε) η πρώτη έξοδος 2 της φάσης 1 IS (είναι) μικρός θετικός

$$((+1 \times 1) + (+1 \times 1)) = \left[ \frac{+2}{2} \right].$$

Ο βαθμός ιδιότητας μέλους για το μεταγενέστερο μέρος των κανόνων καθορίζεται με τη λήψη της T-norm, π.χ. ελάχιστο, των τιμών των αντίστοιχων προγενέστερων. Η τιμή της ιδιότητας μέλους για την κατηγορία δεδομένων εξόδου (π.χ. το μέρος n της φάσης 1) καθορίζεται με τη λήψη της S-norm, π.χ. μέγιστο, των τιμών του βαθμού ιδιότητας μέλους των μεταγενέστερων, εάν φυσικά είναι γλωσσολογικά ίδιοι. Διαφορετικά, υπάρχουν διαφορετικές τιμές εξόδου με το δικό τους βαθμό ιδιότητας μέλους.

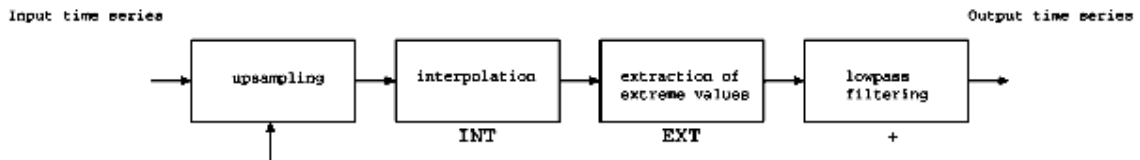
Οι τιμές εξόδου της φάσης n-1 χρησιμοποιούνται ως τιμές εισαγωγής στη φάση n. Τέλος, η τιμή εξόδου (εκτίμηση ζήτησης του προϊόντος) παράγεται από τις οκτώ αρχικές τιμές δεδομένων εισαγωγής. Στη φάση “defuzzification”, η ασαφής τιμή εξόδου, παραδείγματος χάριν μικρός θετικός σε βαθμό ιδιότητας μέλους 0,84 και μηδέν σε βαθμό ιδιότητας μέλους 0,16, μετασχηματίζεται στην σαφής τιμή με τη χρησιμοποίηση συγκεκριμένης μεθόδου “defuzzification”.

### **3.4.1.6 Η ενότητα δημιουργίας λειτουργιών ιδιότητας μέλους**

Η ενότητα υπολογίζει τις λειτουργίες ιδιότητας μέλους για τις διάφορες μεταβλητές εισαγωγής (η ζήτηση που προβλέφθηκε από έναν πελάτη, η ζήτηση που προβλέφθηκε από έναν εμπειρογνώμονα αγοράς, σταθμισμένη μέση τιμή των προηγούμενων τιμών της ζήτησης, τάση της προηγούμενης ζήτησης, σταθμισμένη μέση τιμή των παραγγελιών, τάση των παραγγελιών, σταθμισμένη μέση τιμή των προηγούμενων προβλέψεων, τάση των προηγούμενων προβλέψεων) στο μοντέλο. Η πρώτη φάση της διαδικασίας είναι η προ-επεξεργασία των ακατέργαστων στοιχείων (προηγούμενες τιμές των προβλέψεων, της ζήτησης και των παραγγελιών που αποθηκεύονται στη βάση δεδομένων του εκάστοτε εργοστασίου) στην κατάλληλη μορφή για την κατασκευή αξιόπιστων λειτουργιών ιδιότητας μέλους. Η προ-επεξεργασία των δεδομένων χρονοσειρών εκτελείται με εφαρμογή των μεθόδων, της άνω-δειγματοληψίας, της παρεμβολής, του φίλτραρίσματος και της κάτω-δειγματοληψίας

(σχήμα3.22). Στη δεύτερη φάση οι λειτουργίες ιδιότητας μέλους υπολογίζονται από την προ-επεξεργασμένη χρονοσειρά.

Σχήμα 3.22 Η προ-επεξεργασία των on-line δεδομένων.



Η άνω-δειγματοληψία των μηδενικών εκτελείται προκειμένου να επιτευχθεί η σωστή χρονική εξάρτηση των τιμών των δεδομένων, επειδή τα αρχικά δεδομένα περιλάμβαναν τιμές χρονοσειράς μόνο διαφορετικές από το μηδέν (οι περίοδοι μη-ζήτησης συμπεριλαμβάνονται στη χρονοσειρά σαν σαφείς μηδενικές τιμές της ζήτησης). Η παρεμβολή της χρονοσειράς εφαρμόζεται για να "παράγει" δεδομένα μεταξύ των σημείων δειγμάτων. Γενικά η εφαρμογή της μεθόδου παρεμβολής πρέπει να επιλεχτεί προσεκτικά. Η πολυωνυμική παρεμβολή είναι συνήθως επαρκής εφ' όσον δεν γίνεται για υπερβολικά μεγάλο χρονικό διάστημα.

Μετά από την παρεμβολή οι ακραίες τιμές αφαιρούνται. Το φιλτράρισμα ακραίων τιμών χρησιμοποιείται για να αποφευχθεί η πολύ ευρεία κατανομή του βαθμού ιδιότητας μέλους. Οι ακραίες τιμές, ελάχιστη και μέγιστη, αυξάνουν την κλίμακα της χρονοσειράς και επομένως και το πλάτος της αποκαλούμενης ασαφούς διανομής και των ασαφών λειτουργιών ιδιότητας μέλους, (βλ. σχήμα 3.23).

Κατόπιν, η χρονοσειρά τροφοδοτούνται σε ένα χαμηλής διέλευσης φίλτρο με συντελεστές  $1/2$  και  $1/2$ , προκειμένου να υπολογιστεί ο κινούμενος μέσος όρος των διαδοχικών τιμών. Το προκύπτον σήμα είναι μια ακριβής αναδημιουργία του "συνεχούς χρονικά" σήματος στο χρόνο δειγματοληψίας. Το χαμηλής διέλευσης φίλτρο παρεμβάλλει τη σωστή αναδημιουργία μεταξύ των δειγμάτων της χρονοσειράς.

Σε γενικές γραμμές, η ασαφής διανομή της αρχικής χρονοσειράς περιγράφει τη δυνατότητα των γεγονότων πιο σωστά σε μέγεθος, από ότι η ασαφής διανομή άνω- δειγματοληφθείσας χρονοσειράς. Γι' αυτό, η χρονική εξάρτηση των γεγονότων εξετάζεται με τη χρησιμοποίηση τιμών πιθανότητας που καθορίζονται από την αρχική χρονοσειρά. Οι τιμές πιθανότητας εφαρμόζονται στα τελικά αποτελέσματα, προκειμένου να καθοριστεί με βεβαιότητα το επίπεδο της ζήτησης. Εάν η πιθανότητα είναι πολύ χαμηλή ( $\leq 0.2$ ), η πρόβλεψη ζήτησης δεν εξετάζεται. Εντούτοις, η χαμηλή τιμή πιθανότητας λαμβάνεται υπόψη για τον προσδιορισμό της πιθανότητας πρόβλεψης της επόμενης εβδομάδας ως ρεαλιστική τιμή ζήτησης.

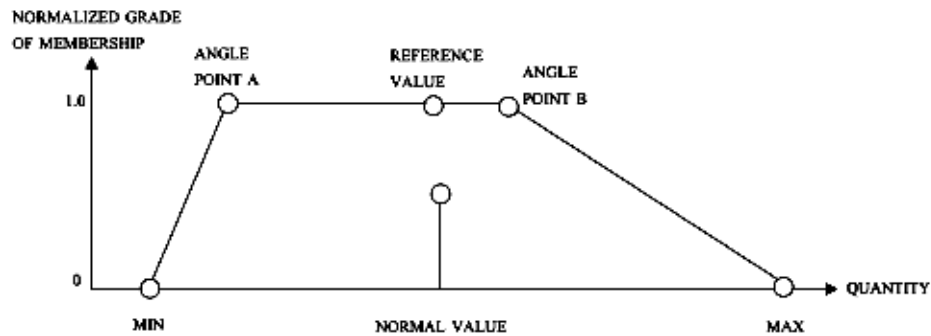
Επιπλέον, χωριστό πρόγραμμα μπορεί να χρησιμοποιηθεί για να υπολογίσει τη λειτουργία αυτοσυσχέτισης και τις περιόδους για τα στοιχεία. Αυτό είναι ένα χρήσιμο διαγνωστικό εργαλείο για

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

το χρήστη προκειμένου να είναι βέβαιος ότι κανένα περιοδικό ή έντονα χρονικά-συσχετισμένο γεγονός δεν περιλαμβάνεται στα δεδομένα.

Προκειμένου να καθοριστούν οι ασαφείς λειτουργίες ιδιότητας μέλους από την προεξεργασμένη χρονοσειρά (χρονοσειρά εξόδου στο σχήμα 3.22) μια ειδική κατανομή, μια ασαφής διανομή (σχήμα 3.23) δημιουργείται.

**Σχήμα 3.23** Ασαφής διανομή



Αποτελείται από τις ακόλουθες τιμές που υπολογίζονται από τη χρονοσειρά:

- ελάχιστη τιμή,
- μέσος όρος της ελάχιστης τιμής και της μέσης τιμής (= σημείο γωνίας A),
- μέσος όρος της μέσης τιμής και της μέγιστης τιμής (= σημείο γωνίας B),
- μέγιστη τιμή.

Ο προσδιορισμός των ασαφών λειτουργιών ιδιότητας μέλους από την προαναφερθείσα διανομή, έχει γίνει με τη διαίρεση του σε διάφορα μέρη ανάλογα με τον υπό εξέταση αριθμό γλωσσικών επιπέδων. Η κατανομή αρχίζει από τη μέση της διανομής. Τα σημεία κατανομής καθορίζονται συμμετρικά και στις δύο πλευρές της τιμής αναφοράς και ο αριθμός τους είναι:

αριθμός σημείων κατανομής σε κάθε πλευρά=

$$\frac{2 \times (\text{number of labels}) - 4}{2} \quad (3.92)$$

Τα σημεία κατανομής καθορίζονται, παραδείγματος χάριν στην περίπτωση πέντε ετικετών (έξι σημεία κατανομής), στην προκειμένη περίπτωση με τον ακόλουθο τρόπο (σχήμα 3.24):

(1) καθορίστε τη μέση τιμή των δειγμάτων μεταξύ του σημείου γωνίας A και της τιμής βάρους (τιμή αναφοράς στο σχήμα 3.23) =σημείο κατανομής 1.

(2) καθορίστε τη μέση τιμή των δειγμάτων μεταξύ του σημείου κατανομής 1 και της τιμής βάρους =σημείο κατανομής 2.

(3) καθορίστε τη μέση τιμή των δειγμάτων μεταξύ του σημείου κατανομής 2 και της τιμής βάρους =σημείο κατανομής 3.

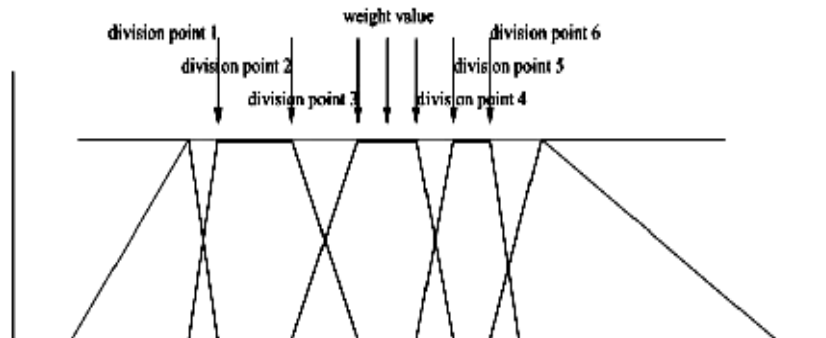


(4) καθορίστε τη μέση τιμή των δειγμάτων μεταξύ του σημείου γωνίας B και της τιμής βάρους =σημείο κατανομής 4.

(5) καθορίστε τη μέση τιμή των δειγμάτων μεταξύ του σημείου κατανομής 4 και του σημείου γωνίας B =σημείο κατανομής 5.

(6) καθορίστε τη μέση αξία των δειγμάτων μεταξύ του σημείου γωνίας B και του σημείου κατανομής 5 =σημείο κατανομής 6.

**Σχήμα 3.24** Κατανομή της ασαφούς διανομής



Όταν ο αριθμός κατανομών πλησιάζει τον αριθμό δειγμάτων τα σημεία κατανομής διαμορφώνουν μια αρχική διανομή ( όπως των δειγμάτων). Επομένως το κύριο τμήμα κατανομής της χρονοσειράς περιγράφει τις γλωσσικές περιοχές, εάν ο διαχωρισμός της κατανομής είναι αρκετά υψηλός. Λόγω της απλότητας και της μηχανοποίησης, η διανομή δεν μπορεί να ενταχθεί στις γνωστές στατιστικές κατανομές προκειμένου να βρεθεί αναλυτική λύση.

Το πλεονέκτημα της αυτόματης μεθόδου παραγωγής είναι ότι οι λειτουργίες ιδιότητας μέλους ενημερώνονται συνεχώς σύμφωνα με τις δυναμικές αλλαγές του συστήματος. Επομένως τα συστήματα μπορούν να προσαρμοστούν δυναμικά στις νέες περιστάσεις μέσω του συντονισμού των λειτουργιών ιδιότητας μέλους. Αυτό είναι απολύτως απαραίτητο στον τομέα των τηλεπικοινωνιών δεδομένου ότι η κατάσταση είναι ιδιαίτερα δυναμική για τους κατασκευαστές. Ως επιπλέον πλεονέκτημα μπορεί να αναφερθεί ότι μέσω της αυτόματης παραγωγής των λειτουργιών της ιδιότητας μέλους αποφεύγουμε την παραδοσιακή χρονοβόρα παραγωγή των λειτουργιών ιδιότητας μέλους για περισσότερα από 1000 διαφορετικά προϊόντα.

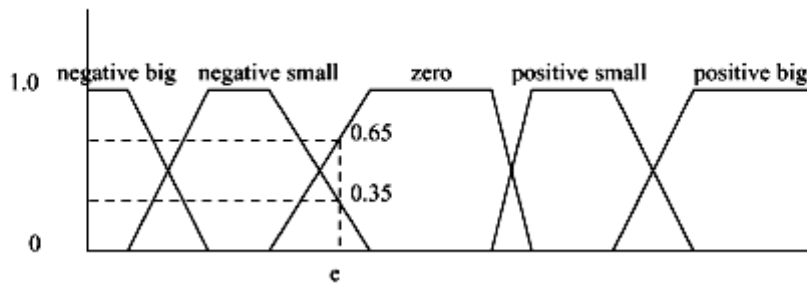
3.4.1.7 Ενότητα "δημιουργίας ασάφειας" (Fuzzification module)

Η ενότητα fuzzification μετατρέπει μια σαφή τιμή μιας μεταβλητής σε ένα ασαφές σύνολο. Τα ασαφή αποτελέσματα προκύπτουν από τις ιδιότητες μέλους ασαφών συνόλων των μεταβλητών που εισάγονται στην ενότητα συμπεράσματος, με την βοήθεια των βάσεων δεδομένων και των κανόνων. Η διαδικασία "fuzzification" δεν προσφέρει οποιαδήποτε ευελιξία ή βαθμό ελευθερίας στο σχεδιασμό των ασαφών μοντέλων. Η διαδικασία "fuzzification" με σαφή τιμή εισαγωγής  $e$  μπορεί να περιγραφεί με υψηλού επιπέδου μορφή ψευδοκώδικα (βλ. σχήμα 3.25)

```
Select-Fired-Labels(real crisp-input-value)
{IF  $e \geq$  minimum value of a label  $X$ 
AND  $e <$  maximum value of a label  $X$ 
THEN fired-label( $i$ ) :=  $X$ }
```

```
Calculate-Label-Value(integer fired-label (:))
{Grade-of-Membership( $i$ ) = Value-of-Label
(fired-label( $i$ ),  $e$ )}.
```

Σχήμα 3.25 Διαδικασία " Fuzzification " μιας μεταβλητής εισαγωγής με τιμή  $e$ .



Στο σχήμα 3.26, οι τροφοδοτούμενες ετικέτες είναι μικροί αρνητικοί σε βαθμό ιδιότητας μέλους 0,35 και μηδέν σε βαθμό ιδιότητας μέλους 0,65. Σε άλλες λέξεις, η σαφής τιμή εισαγωγής  $e$  μπορεί να ερμηνευθεί τώρα με γλωσσικούς όρους μικροί αρνητικοί και μηδέν με τον παρουσιαζόμενο βαθμό ιδιότητας μέλους.

3.4.1.8 Ενότητα συμπεράσματος

Στη φάση συμπεράσματος τα ασαφή αποτελέσματα προκύπτουν από τις ιδιότητες μέλους των ασαφών συνόλων με την βοήθεια της βάσης γνώσεων. Η βάση γνώσεων αποτελείται από μια βάση δεδομένων και μια βάση κανόνων. Η βάση δεδομένων έχει πληροφορίες για μια κατάλληλη λειτουργία των φάσεων κανονικοποίησης, "δημιουργίας ασάφειας-fuzzification", βάσης κανόνων, "ανάφρασης ασάφειας-defuzzification" και αποκανονικοποίησης. Οι πληροφορίες της περιλαμβάνουν

τις ασαφείς λειτουργίες ιδιότητας μέλους και τις φυσικές περιοχές, τον τύπο ασαφών σχέσεων, τον τύπο των αλληλεπιδράσεων χειρισμού, τον τύπο τροφοδότησης και τον αριθμό ιεραρχικών επιπέδων στο συλλογισμό. Η βάση δεδομένων κατασκευάστηκε κατά τη διάρκεια της ανάπτυξης του μοντέλου. Η ασαφής βάση κανόνων χρησιμοποιεί τις πληροφορίες της βάσης δεδομένων και αντιπροσωπεύει, σε γλωσσικό επίπεδο, τη λειτουργία του διαμορφωμένου συστήματος υπό μορφή κανόνων παραγωγής ή γλωσσικών εξισώσεων. Μπορεί γενικότερα να υποστηριχθεί ότι, ένας μηχανικός συστημάτων είναι συνήθως αρμόδιος για την παραγωγή βάσεων δεδομένων και οι εμπειρογνώμονες σε κάθε τομέα εφαρμογής είναι αρμόδιοι για τις πληροφορίες της βάσης κανόνων, ενώ η παραγωγή του ανήκει στο μηχανικό συστημάτων.

Η βάση κανόνων του FLAT διαιρείται σε διάφορα ιεραρχικά υποσυστήματα. Η βάση κανόνων κάθε υποσυστήματος αντιπροσωπεύεται από τις γλωσσικές σχέσεις, οι οποίες μπορούν να μετατραπούν σε εξισώσεις μητρών. Ο συλλογισμός είναι βασισμένος σε αυτές τις εξισώσεις. Η αλυσίδα συλλογισμού του FLAT μπορεί να περιγραφεί με τον ακόλουθο τρόπο (βλ. σχήμα 3.26):

- Οι γλωσσικές σχέσεις μετασχηματίζονται σε μορφή εξίσωσης σε κάθε ένα από τα ιεραρχικά υποσυστήματα (μέρη) στο πρώτο επίπεδο (φάση)
- Το συμπέρασμα ή ο συλλογισμός στο πρώτο επίπεδο του κάθε υποσυστήματος εκτελείται με την επίλυση των εξισώσεων
- Τα ασαφή συμπεράσματα που προκύπτουν ως αποτέλεσμα του συλλογισμού σε κάθε υποσύστημα στο επίπεδο 1, συνδυάζονται σύμφωνα με το σχήμα 3.26 και μετατρέπονται σε μορφή εξίσωσης στο επίπεδο 2. Ο συλλογισμός εκτελείται πάλι με την επίλυση αυτών των εξισώσεων.
- Οι λύσεις του επιπέδου 2 συνδυάζονται και μετασχηματίζονται άλλη μια φορά σε μορφή εξίσωσης στο επίπεδο 3, οι οποίες επιλύονται για να δώσουν το τελικό συμπέρασμα.

Σαν παράδειγμα θεωρούμε τον ακόλουθο συλλογισμό.

Γλωσσική σχέση: IF(εάν) η ζήτηση που προβλέφθηκε από έναν πελάτη IS(είναι) μεγάλος θετικός σε βαθμό ιδιότητας μέλους 0.65 AND (και) η ζήτηση που προβλέφθηκε από έναν εμπειρογνώμονα αγοράς IS(είναι) μικρός θετικός σε βαθμό ιδιότητας μέλους 1,0 THEN (τότε) η έξοδος της φάσης 1 IS(είναι) μεγάλος θετικός σε βαθμό ιδιότητας μέλους 0,65.

Γλωσσική σχέση με μορφή εξίσωσης:

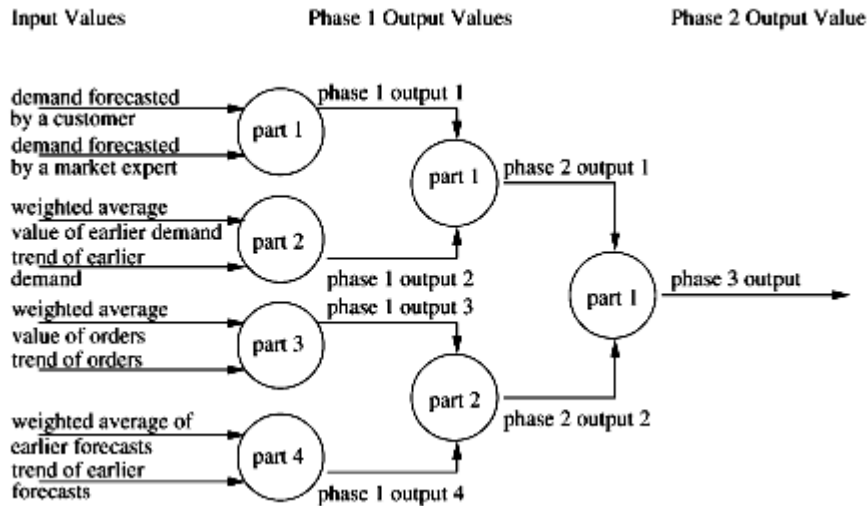
$$((+1 \times 2) + (+1 \times 1) = \left[ \frac{3}{2} \right])$$

Γλωσσική σχέση: IF(εάν) η ζήτηση που προβλέφθηκε από έναν πελάτη IS(είναι) μικρός θετικός σε βαθμό ιδιότητας μέλους 0.35 AND (και) η ζήτηση που προβλέφθηκε από έναν εμπειρογνώμονα αγοράς IS(είναι) μικρός θετικός σε βαθμό ιδιότητας μέλους 1,0 THEN (τότε) η έξοδος της φάσης 1 IS(είναι) μικρός θετικός σε βαθμό ιδιότητας μέλους 0,35.

Γλωσσική σχέση με μορφή εξίσωσης:

$$((+1 \ x \ 1) + (+1 \ x \ 1) = \left[ \begin{matrix} 2 \\ 2 \end{matrix} \right])$$

Σχήμα 3.26 Η λειτουργία εξαγωγής συμπεράσματος στο FLAT



Ο συλλογισμός στο πρώτο επίπεδο του μέρους 1 δίνει αποτέλεσμα μεγάλο θετικό σε βαθμό ιδιότητας μέλους 0,65 και μικρό θετικό σε βαθμό ιδιότητας μέλους 0,35. Ας υποθέσουμε ότι τα αποτελέσματα από το μέρος 2 του πρώτου επιπέδου είναι μικρός αρνητικός σε βαθμό ιδιότητας μέλους 0,70 και μηδέν σε βαθμό ιδιότητας μέλους 0,30. Τότε οι εισαγωγές στο μέρος 1 του επιπέδου 2 είναι μεγάλος θετικός, μικρός θετικός, μικρός αρνητικός και μηδέν με το δεδομένο βαθμό ιδιοτήτων μέλους.

Σε μορφή μητρών, οι κανόνες του μέρους 1 του επιπέδου 1 μπορούν να παρουσιαστούν ως εξής (χάρην απλότητας των απεικονίσεων χρησιμοποιούμε την ίδια μήτρα και στα άλλα μέρη και επίπεδα):

$$\begin{pmatrix} -2 & -1 & 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} +1 \\ +1 \\ +1 \\ +1 \\ +1 \end{pmatrix} + \begin{pmatrix} -2 & -1 & 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} +1 \\ +1 \\ +1 \\ -1 \\ +1 \end{pmatrix} \quad (3.93)$$

Στο δεύτερο επίπεδο τροφοδοτούμε τους ακόλουθους κανόνες.

IF(εάν) η εισαγωγή 1 IS(είναι) μεγάλος θετικός σε βαθμό ιδιότητας μέλους 0,65 AND (και) η εισαγωγή 2 IS(είναι) μικρός αρνητικός σε βαθμό ιδιότητας μέλους 0,70 THEN (τότε) η έξοδος 1 της φάσης 2 IS(είναι) μικρός θετικός σε βαθμό ιδιότητας μέλους 0,65.

$$\text{Σε μορφή εξίσωσης: } ((+1 \ x \ 2) + (+1 \ x \ -1) = \left[ \frac{1}{2} \right])$$

IF(εάν) η εισαγωγή 1 IS(είναι) μεγάλος θετικός σε βαθμό ιδιότητας μέλους 0,65 AND (και) η εισαγωγή 2 IS(είναι) μηδέν σε βαθμό ιδιότητας μέλους 0,30 THEN (τότε) η έξοδος 2 της φάσης 2 IS(είναι) μεγάλος θετικός σε βαθμό ιδιότητας μέλους 0,30.

$$\text{Σε μορφή εξίσωσης: } ((+1 \ x \ 2) + (+0 \ x \ 0) = \left[ \frac{2}{1} \right])$$

IF(εάν) η εισαγωγή 1 IS(είναι) μικρός θετικός σε βαθμό ιδιότητας μέλους 0,35 AND (και) η εισαγωγή 2 IS(είναι) μικρός αρνητικός σε βαθμό ιδιότητας μέλους 0,70 THEN (τότε) η έξοδος 3 της φάσης 2 IS(είναι) μηδέν σε βαθμό ιδιότητας μέλους 0,35.

$$\text{Σε μορφή εξίσωσης: } ((+1 \ x \ 1) + (+1 \ x \ -1) = \left[ \frac{0}{2} \right])$$

IF(εάν) η εισαγωγή 1 IS(είναι) μικρός θετικός σε βαθμό ιδιότητας μέλους 0,35 AND (και) η εισαγωγή 2 IS(είναι) μηδέν σε βαθμό ιδιότητας μέλους 0,30 THEN (τότε) η έξοδος 4 της φάσης 2 IS(είναι) μικρός θετικός σε βαθμό ιδιότητας μέλους 0,30.

$$\text{Σε μορφή εξίσωσης: } ((+1 \ x \ 1) + (+1 \ x \ 0) = \left[ \frac{1}{2} \right])$$

Ο συλλογισμός στο δεύτερο επίπεδο του μέρους 1 δίνει αποτέλεσμα μεγάλο θετικό σε βαθμό ιδιότητας μέλους 0,30, μικρό θετικό σε μέγιστο βαθμό ιδιότητας μέλους (0,65, 0,30)=0,65 και μηδέν σε βαθμό ιδιότητας μέλους 0,35. Εάν θέσουμε τιμή κατωφλίου στο επίπεδο του 0,40 συνεπάγεται ότι η έξοδος του επιπέδου 2 του μέρους 1 είναι μικρός θετικός σε βαθμό 0,65. Ας υποθέσουμε ότι η έξοδος από το μέρος 2 του δεύτερου επιπέδου είναι μικρός αρνητικός σε βαθμό ιδιότητας μέλους 0,70. Τότε οι εισαγωγές στο επίπεδο 3 είναι μικροί θετικοί και μικροί αρνητικοί με το δεδομένο βαθμό ιδιότητας μέλους.

Κατόπιν στο τρίτο επίπεδο τροφοδοτείται μόνο ένας κανόνας.

IF(εάν) η εισαγωγή 1 IS(είναι) μικρός θετικός σε βαθμό ιδιότητας μέλους 0,65 AND (και) η εισαγωγή 2 IS(είναι) μικρός αρνητικός σε βαθμό ιδιότητας μέλους 0,70 THEN (τότε) η έξοδος της φάσης 3 IS(είναι) μικρός θετικός σε βαθμό ιδιότητας μέλους 0,65.

Σε μορφή εξίσωσης:  $((+1 \ x \ 1) + (-1 \ x \ -1) = \frac{-2}{2})$

Το πλεονέκτημα σε αυτήν την ιεραρχική δομή είναι ότι ο αριθμός κανόνων θα αυξηθεί γραμμικά με τον αριθμό μεταβλητών του συστήματος και όχι εκθετικά όπως στα συμβατικά συστήματα. Εν τω παρόντι, ο αριθμός κανόνων στη βάση κανόνων του FLAT είναι συνολικά 175 κανόνες :100 στο πρώτο επίπεδο (25 κανόνες σε κάθε μέρος), 50 κανόνες στο δεύτερο επίπεδο και 25 κανόνες στο τρίτο επίπεδο.

Ο αριθμός κανόνων σε μια συμβατική βάση κανόνων, για σύστημα που προβλέπει την ζήτηση περισσότερων από 1000 διαφορετικών προϊόντων, θα ήταν φυσικά πάρα πολύ μεγάλος για να διαμορφωθεί σε μοντέλο και να ενημερωθεί.

#### **3.4.1.9 Ενότητα Αναίρεσης Ασάφειας (defuzzification module)**

Η αναίρεση ασάφειας των ασαφών συμπερασμάτων εκτελείται στην ενότητα αυτή. Ο τελεστής αναίρεσης ασάφειας καθορίζει τη χρησιμοποιημένη μέθοδο. Οι τελεστές που εξετάστηκαν στην εφαρμογή FLAT ήταν το κέντρο της περιοχής, το κέντρο των αθροισμάτων, κέντρο της μεγαλύτερης περιοχής, το πρώτο από τα μέγιστα, το τελευταίο από τα μέγιστα, το μέσο των μεγίστων και το ύψος της αναίρεσης ασάφειας. Δεν υπήρξαν σημαντικές διαφορές μεταξύ των αποτελεσμάτων που επιτεύχθηκαν με τη χρήση των παραπάνω τελεστών. Εντούτοις, επιλέχθηκε τελικά η μέθοδος του κέντρου της περιοχής, επειδή θεωρητικά φαίνεται να είναι σχεδόν τέλεια έναντι των άλλων μεθόδων.

Το μειονέκτημα της μεθόδου του κέντρου περιοχής είναι ότι είναι υπολογιστικά, αρκετά σύνθετη και αργή. Η πιο γρήγορη έκδοση είναι η μέθοδος του κέντρου της αθροισμάτων. Η διαφορά είναι ότι αντί να λαμβάνει την ένωση των τροφοδοτούμενων ασαφών συνόλων, παίρνει το άθροισμά τους. Κατά συνέπεια, οι πιθανές επικαλυπτόμενες περιοχές απεικονίζονται περισσότερες από μία φορά σε αυτήν την μέθοδο.

Το κέντρο της μεγαλύτερης περιοχής είναι τροποποιημένο κέντρο της μεθόδου περιοχής που απευθύνεται στη μεγαλύτερη ασαφή κυρτή περιοχή. Η απόδοσή της έναντι του κέντρου της μεθόδου περιοχής, εξαρτάται από το πόσο εύκολα βρίσκεται η μεγαλύτερη κυρτή περιοχή έναντι του καθορισμένου σημείου βάρους της συνολικής τροφοδοτούμενης περιοχής.

Οι μέθοδοι του πρώτου και του τελευταίου από τα μέγιστα, καθώς και του μέσου των μεγίστων είναι υπολογιστικά αποτελεσματικές. Αυτή η ακρίβεια εξαρτάται κατά πολύ από τη μορφή και τον τύπο λειτουργιών ιδιότητας μέλους που χρησιμοποιούνται, συμπεριλαμβανομένου του αριθμού ετικετών.

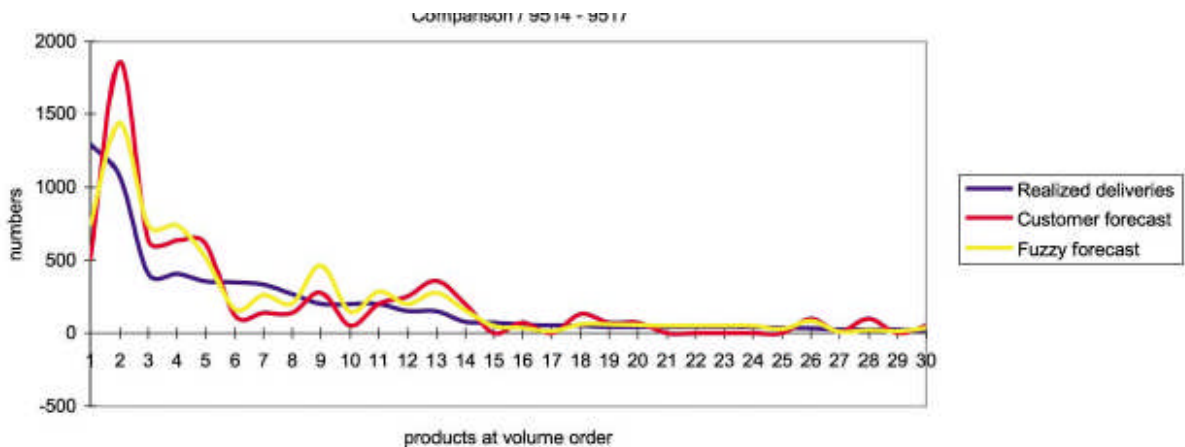
Η μέθοδος ύψους αναίρεσης της ασάφειας είναι μια "επιταχυνόμενη" έκδοση της μεθόδου του κέντρου αθροισμάτων. Το μειονέκτημά της είναι ότι δεν εξετάζει τη μορφή των ασαφών συνόλων.

3.4.1.10 Συμπεράσματα

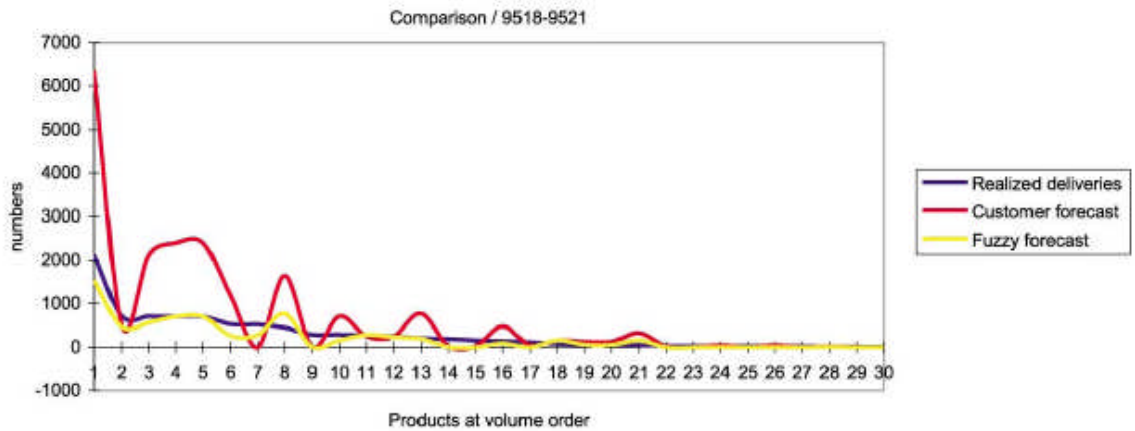
Το παρόν τμήμα της εργασίας εστίασε στα προβλήματα που αφορούν τη λήψη απόφασης για την αγορά υλικών σε ένα αβέβαιο και ασαφές περιβάλλον με την βοήθεια της θεωρίας ασαφών συνόλων. Το λογισμικό FLAT προβλέπει τη ζήτηση των πελατών για διάφορες συσκευές μετάδοσης σημάτων σε μια συγκεκριμένη περιοχή αγοράς. Οι προβλέψεις διαμορφώνονται πάνω σε μια βάση προϊόντων και σε μια βάση πελατών, δηλ. για κάθε προϊόν και κάθε πελάτη χωριστά.

Μια σύγκριση της πραγματικής ζήτησης, η πρόβλεψη του πελάτη, η πρόβλεψη του ειδικού και η πρόβλεψη που παρέχεται από το FLAT παρουσιάζονται στα σχήματα 3.27 και 3.28. Η πραγματική ζήτηση συγκεντρώθηκε από τις πραγματοποιημένες πωλήσεις κατά τη διάρκεια διαστήματος τεσσάρων εβδομάδων. Οι προβλέψεις πελατών παραλήφθηκαν από διαφορετικούς πελάτες (teleoperators) της επιχείρησης και συγκεντρώθηκαν από έναν εμπειρογνώμονα. Τα προϊόντα ταξινομούνται με βάση τον όγκο τους. Όπως μπορεί να φανεί στα σχήματα 3.27 και 3.28, οι καμπύλες του FLAT ακολουθούν την πραγματική ζήτηση ακριβέστερα από ότι η προβλεπόμενη καμπύλη, και το τετραγωνικό λάθος του FLAT είναι λιγότερο από αυτό που προβλεφθηκε.

**Σχήμα 3.27** Σύγκριση της πραγματικής ζήτησης, της ζήτησης που προβλέπεται από τον πελάτη και τον εμπειρογνώμονα μαζί και της ζήτησης που προβλέπεται από το FLAT κατά τη διάρκεια της πρώτης περιόδου



**Σχήμα 3.28** Σύγκριση της πραγματικής ζήτησης, της ζήτησης που προβλέπεται από τον πελάτη και τον εμπειρογνώμονα μαζί και της ζήτησης που προβλέπεται από το FLAT κατά τη διάρκεια της δεύτερης περιόδου



Κατά τη διάρκεια της πρώτης περιόδου η πρόβλεψη πελατών ήταν καλύτερη στο 12% των περιπτώσεων ενώ η πρόβλεψη του FLAT ήταν καλύτερη στο 60% των περιπτώσεων, τα αποτελέσματα είναι ίδια στο 28% των περιπτώσεων, ενώ οι αντίστοιχοι αριθμοί στη δεύτερη περίοδο ήταν στο 19% των περιπτώσεων, στο 56% των περιπτώσεων και στο 25% των περιπτώσεων. Οι τετραγωνικές ρίζες των τετραγωνικών σφαλμάτων των προβλέψεων των πελατών ήταν 505,4 και 1836,1, ενώ εκείνες από τις προβλέψεις του FLAT ήταν 271,8 και 356,2.

### 3.4.1.11 Αξιολόγηση της επιλεγμένης μεθόδου

Η χρησιμοποιημένη προσέγγιση είναι πολλά υποσχόμενη και χρήσιμη σε διάφορες περιοχές των εφαρμογών, π.χ. στις κατασκευές. Η μέθοδος μπορεί να προβλέψει μια μη γραμμική συμπεριφορά καλύτερα από πολλές παραδοσιακές μεθόδους χρονοσειρών οι οποίες είναι βασισμένες, π.χ., στις πολυωνμικές προσεγγίσεις. Έγιναν εκτενείς μελέτες με μη γραμμικές μεθόδους. Μέθοδοι βασισμένες σε χαρακτηριστικά όπως το πολυώνυμο Lyapunov, η κοντινότερη προσέγγιση ή η προσέγγιση Hurst-παραμέτρου παρείχαν ενδιαφέρουσες ιδέες στην ελλοχεύουσα διαδικασία. Αν και αυτές οι μέθοδοι απέδειξαν την έμφυτη μη γραμμικότητα της διαδικασίας, οι παραδοσιακές μέθοδοι πρόβλεψης χρονοσειρών δεν θα μπορούσαν να ανταγωνιστούν την προσέγγιση FLAT. Ειδικά αν θεωρηθεί το μέγιστο τετραγωνικό σφάλμα των εκτιμήσεων του FLAT. Μπορεί να υποστηριχτεί ότι αυτό συμβαίνει λόγω:

- (α) της προσαρμοστικότητας του αλγορίθμου ασαφούς λογικής και
- (β) του αυξανόμενου βαθμού προβλεψιμότητας που οφείλεται στη βάση κανόνων και στις ειδικές πληροφορίες στο σύνολο εισαγωγής.



### 3.4.1.12 Προσαρμοστικότητα μέσω του συντονισμού του ασαφούς μοντέλου

Ο συντονισμός σημαίνει τροποποίηση μιας ή περισσοτέρων από τις παραμέτρους σχεδιασμού. Οι μέθοδοι συντονισμού μπορούν να διακριθούν σε απευθείας σύνδεσης (on-line) και σε μη απευθείας σύνδεσης (off-line) τύπους, ανάλογα με το εάν οι παράμετροι σχεδιασμού συντονίζονται κατά τη λειτουργία του μοντέλου ή κατόπιν.

Το πρώτο βήμα είναι συνήθως να τροποποιήσουμε τις λειτουργίες ιδιότητας μέλους. Το δεύτερο βήμα είναι να επεξεργαστούμε τη βάση κανόνων, τα βάρη των κανόνων, ή να εισάγουμε νέες μεταβλητές. Επίσης, μπορούμε να αλλάξουμε ή να τροποποιήσουμε τη μέθοδο αναίρεσης ασάφειας (defuzzification).

Προκειμένου να αυξηθεί η απόδοση του ασαφούς μοντέλου, ο αλγόριθμος συντονισμού απαιτεί την προσεκτική ανάλυση των παραμέτρων. Στο συντονισμό των ασαφών λειτουργιών ιδιότητας μέλους, η μορφή και ο αριθμός τους καθώς και το μέγεθος της χρονοσειράς θεωρούνται σαν παράμετροι συντονισμού και αναλύονται κατά τη διάρκεια της ανάπτυξης του αλγορίθμου.

### 3.4.1.13 Η δημιουργία των λειτουργιών ιδιότητας μέλους

Η δημιουργία των λειτουργιών ιδιότητας μέλους εμφανίζεται είτε κατά τη διάρκεια της λειτουργίας του μοντέλου (on-line) είτε πριν τη λειτουργία του (off-line). Συνήθως (off-line). Η on-line εκτίμηση δεν συμπεριλαμβάνεται στα παραδοσιακά ασαφή μοντέλα. Στην off-line παραγωγή των ασαφών λειτουργιών ιδιότητας μέλους, τα αποτελέσματα εξαρτώνται συνήθως από την ανάλυση των δεδομένων, τη μέθοδο συντονισμού του μοντέλου και από τις δεξιότητες του χειριστή και του μηχανικού συστημάτων. Εντούτοις, υπάρχουν πολλές περιπτώσεις, όπου οι ασαφείς λειτουργίες ιδιότητας μέλους είναι δύσκολο ή αδύνατο να δημιουργηθούν ή και να ενημερωθούν και κατ' αυτό τον τρόπο μπορεί να είναι πιο κατάλληλη μια αυτόματη προσέγγιση. Αυτό εμφανίζεται, παραδείγματος χάριν, στις καταστάσεις όπου απαιτείται η γρήγορη προσαρμογή ενός μοντέλου σύμφωνα με τη δυναμική συστημάτων. Επομένως, θεωρούμε εδώ την on-line παραγωγή σαν χωριστό μέρος σχεδιασμού, όπου διαφορετικά είδη μεθόδων μπορούν να εφαρμοστούν στα δεδομένα, προκειμένου να υπολογιστούν οι λειτουργίες ιδιότητας μέλους.

Το πλεονέκτημα της αναπτυσσόμενης μεθόδου είναι, ότι οι λειτουργίες ιδιότητας μέλους ενημερώνονται συνεχώς σύμφωνα με τις δυναμικές αλλαγές του συστήματος (μέσω του συντονισμού των λειτουργιών ιδιότητας μέλους τα συστήματα μπορούν να προσαρμοστούν σε νέες περιστάσεις).

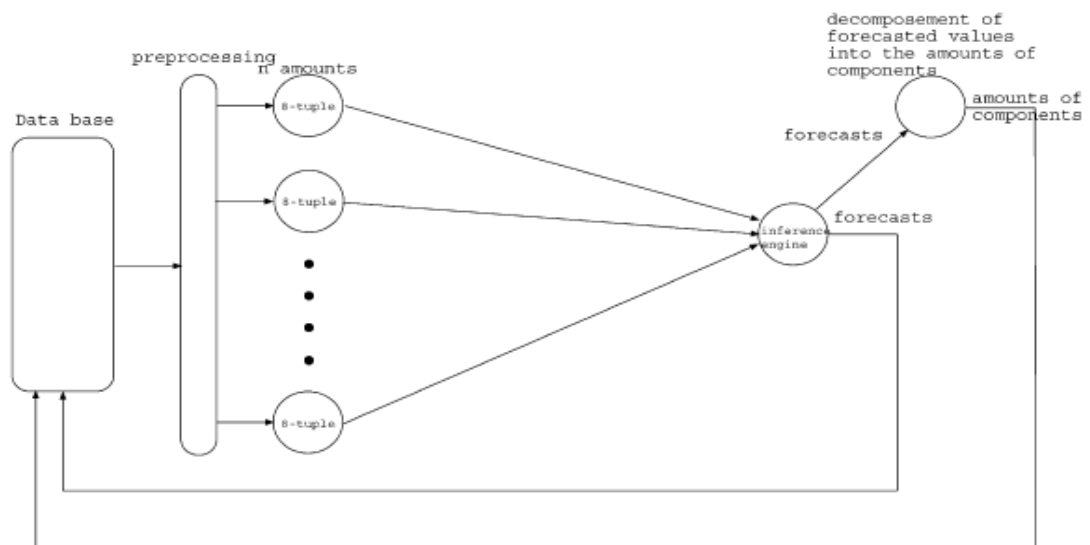
Ο τομέας επιτυχούς εφαρμογής φαίνεται να είναι τα γενικά συστήματα λήψης απόφασης με μεγάλο αριθμό μεταβλητών εισαγωγής. Οι μέθοδοι που χρησιμοποιούνται σε αυτό το πρόγραμμα είναι επίσης εφαρμόσιμες και πρακτικές, ειδικά για τις χρονικά κρίσιμες εφαρμογές, όπου απαιτείται η on-line προσαρμογή. Οι μέθοδοι αυτές έχουν εφαρμοστεί σε συστήματα όπως π.χ. στον έλεγχο ισχύος τερματικών στα κυψελοειδή τηλεφωνικά συστήματα, στην έρευνα συγχρονισμού των κυψελοειδών

τηλεφωνικών συστημάτων κ.α. Το ίδιο είδος προσέγγισης μπορεί να χρησιμοποιηθεί επίσης για διάφορα πολυδιάστατα προβλήματα αναγνώρισης μορφών, προκειμένου να βρεθούν αυτόματα οι βαθμοί ιδιότητας μέλους για ταξινομημένα δεδομένα.

**3.4.1.14 Ικανότητα επεξεργασίας των δεδομένων οκτώ διαστάσεων**

Ακόμα κι αν η έξοδος του συστήματος είναι μονοδιάστατο διάνυσμα  $1 \times n$  (μια τιμή για ένα προϊόν), το σύνολο εισαγωγής είναι διάνυσμα οκτώ-διαστάσεων για κάθε προϊόν, δηλ., η εισαγωγή είναι διάνυσμα  $8 \times n$  διαστάσεων, όπου το  $n$  είναι ο αριθμός προϊόντος. Το διάνυσμα εξόδου διασπάται αργότερα σε διάνυσμα  $1 \times m$ , όπου το  $m$  δείχνει το αριθμό των στοιχείων (υπολογισμός των αριθμό κάθε στοιχείου). Το μοντέλο FLAT μπορεί να δώσει τις τιμές πρόβλεψης για διαφορετικά προϊόντα για  $N$  χρονικές περιόδους στο μέλλον. Εντούτοις, λόγω του ορίου εμπιστοσύνης των προβλέψεων δεν εξάγουμε τιμές πρόβλεψης για περιόδους πάνω από 3 μήνες μπροστά. Η όλη διαδικασία περιγράφεται, στο υψηλότερο επίπεδο, στο σχήμα 3.29.

**Σχήμα 3.29** Διαδικασία αναφοράς σε μεγάλη κλίμακα.



### 3.4.2 Ασαφής πρόβλεψη εποχιακότητας [21]

#### 3.4.2.1 Εισαγωγή

Οι οργανισμοί, που είτε παράγουν αγαθά είτε παρέχουν υπηρεσίες, απαιτούν προβλέψεις των απαιτήσεων για τις αποφάσεις τους. Επομένως, ένας από τους κύριους σκοπούς (στόχους) των λειτουργιών του management περιλαμβάνει τεχνικές πρόβλεψης του τύπου των απαιτήσεων που πρέπει να είναι έγκαιρες, εύλογα ακριβείς και αξιόπιστες. Υπάρχουν δύο κατηγορίες ποσοτικών τεχνικών πρόβλεψης που μπορούν να χρησιμοποιηθούν: τα αιτιώδη πρότυπα και οι χρονοσειρές. Τα αιτιώδη πρότυπα στοχεύουν στη διευκρίνιση των παραγόντων που επηρεάζουν και στις σχέσεις τους με τα αποτελέσματα. Από την άλλη, τα πρότυπα των χρονοσειρών παράγουν προβλέψεις από μια σειρά στοιχείων που παρατηρούνται σε κανονικά (τακτά) χρονικά διαστήματα, βασισμένα σε μελλοντικά δρώμενα που προκύπτουν από το σχέδιο (πρότυπο) του παρελθόντος.

Από τη δημιουργία της από τον καθηγητή Zadeh, η θεωρία ασαφών συνόλων αναγνωρίζεται όλο και περισσότερο ως χρήσιμο εργαλείο στο χειρισμό της ασάφειας της ανθρώπινης γνώσης. Η ασαφής ανάλυση παλινδρόμησης έχει εισαχθεί και μπορεί να χρησιμοποιηθεί για την καθιέρωση των συγκεκριμένων (ασαφών) αιτιωδών προτύπων. Η ασαφής ανάλυση χρονοσειρών έχει εισαχθεί με την εισαγωγή των αποκαλούμενων "τεχνικών ασαφούς διαμόρφωσης". Σε δεδομένα χρονοσειρών, τυπικά πρότυπα όπως η τάση, η εποχιακότητα, και ακόμη και ο κύκλος μπορούν να υπάρξουν.

Σε αυτό το κεφάλαιο, αναπτύσσεται μια ασαφής τεχνική για τις τάσεις και την πρόβλεψη της εποχιακότητας. Μέσω της χρήσης της ασαφούς παλινδρόμησης, αναλύεται η ασαφής τάση των χρονοσειρών. Κατόπιν, η **ασαφής εποχιακότητα** καθορίζεται με τη χρησιμοποίηση των βαθμών ιδιότητας μέλους των εποχών στο πρότυπο παλινδρόμησης. Χρησιμοποιούνται τα ακόλουθα βήματα:

- (1) ανάλυση της ασαφούς τάσης με χρήση της ασαφούς ανάλυσης παλινδρόμησης
- (2) υπολογισμός των **συνόλων των δεικτών ασαφούς εποχιακότητας** για τις εποχές
- (3) χρήση των αποτελεσμάτων των βημάτων (1) και (2) για την παραγωγή συγκεκριμένων (ασαφών) ή/και σταθερών προβλέψεων
- (4) ανάλυση εποχιακής ασάφειας και εποχιακών τάσεων

#### 3.4.2.2 Καταδεικνύοντας την ασαφή τάση μέσω της ασαφούς ανάλυσης παλινδρόμησης

Κατ' αρχάς, θέτουμε υπό μελέτη το συγκεκριμένο πρότυπο παλινδρόμησης για τις χρονοσειρές. Η έννοια της ασαφούς ανάλυσης παλινδρόμησης (fuzzy regression analysis ,FRA) εισήχθη αρχικά από τον Tanaka et al. Κατόπιν, διάφορες μέθοδοι προτάθηκαν. Μια από αυτές τις μεθοδολογίες παρατίθεται παρακάτω.

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

Σύμφωνα με τους ορισμούς που καθορίζονται από τους Dubois και Prade, ένας ασαφής αριθμός μπορεί να οριστεί ως  $M = (\alpha, \beta, \gamma)_{LR}$ , όπου το  $\alpha$  δείχνει τον τρόπο, το  $\beta$  και το  $\gamma$  την προς τα αριστερά και δεξιά διασπορά (εξάπλωση), αντίστοιχα, και τα L και R τις αριστερές και δεξιές αναφορές (δηλ., λειτουργίες μορφής), αντίστοιχα, του M. Επιπλέον, ένας συμμετρικός ασαφής αριθμός δείχνεται ως  $(\alpha, \beta)_L$  με  $\beta = \gamma$  και  $L = R$  (δηλ., έχοντας την ίδια μορφή).

Για τη ασαφή ανάλυση παλινδρόμησης, θεωρούμε το ακόλουθο γενικό συγκεκριμένο γραμμικό πρότυπο (*general fuzzy linear model, GFLM*):

$$Y = Ax = A_1x_1 + \dots + A_nx_n, \quad (3.94)$$

όπου  $x = (x_1, \dots, x_n)^t$  είναι ένα διάνυσμα των μη-συγκεκριμένων (σταθερών) ανεξάρτητων μεταβλητών και  $A = (A_1, \dots, A_n)$  ένα διάνυσμα των συγκεκριμένων παραμέτρων. Η ασαφής παράμετρος ορίζεται ως ένας συμμετρικός ασαφής αριθμός,  $A_j = (a_j, c_j)_L$ , με την τριγωνική συνάρτηση ιδιότητας μέλους

$$\mu_{A_j}(a_j) = \begin{cases} 1 - |a_j - a_i|/c_i, & \alpha_j - c_j \leq a_j \leq \alpha_j + c_j \\ 0, & \text{αλλού} \end{cases} \quad (3.95)$$

Για ευκολία, οι συγκεκριμένοι παράμετροι  $A = (A_1, \dots, A_n)$  μπορούν να γραφούν σε μορφή διανύσματος  $A = (\alpha, c)$  με  $\alpha = (\alpha_1, \dots, \alpha_n)^t$  και  $c = (c_1, \dots, c_n)^t$

Για την έξοδο  $Y$  του GFLM, εξ. (3.94), η επέκταση της αρχής του Zadeh ή ισοδύναμα η ασαφής αριθμητική μπορούν να εφαρμοστούν ως εξής:

$$\begin{aligned} Y &= (\alpha_1, c_1)_L x_1 + \dots + (\alpha_n, c_n)_L x_n \\ &= \left( \sum_{j=1}^n \alpha_j x_j, \sum_{j=1}^n c_j |x_j| \right)_L \\ &= (\alpha^t x, c^t |x|)_L \end{aligned} \quad (3.96)$$

όπου  $|x| = (|x_1|, \dots, |x_n|)^t$ .

Το  $Y$  έχει τη συμμετρική τριγωνική συνάρτηση μέλους

$$\mu_Y(y) = \begin{cases} 1 - |y - \alpha^t x|/c^t |x|, & \alpha^t x - c^t |x| \leq y \leq \alpha^t x + c^t |x| \\ 0, & \text{αλλιώς} \end{cases}$$

Υποθέτοντας ότι υπάρχουν  $N$  ζευγάρια των σταθερής εισόδου και σταθερής εξόδου δεδομένων  $(y_i, x_i)$  ( $i = 1, \dots, N$ ) διαθέσιμα, όπου  $x_i = (x_{i1}, \dots, x_{in})$ . Για τον υπολογισμό των συγκεκριμένων παραμέτρων  $A$ , ο Tanaka et Al πρότεινε την ακόλουθη διατύπωση υπό μορφή γραμμικού προγραμματισμού (linear programming, LP):

$$\text{Min} \quad J = \sum_{i=1}^N c' |x_i| \quad (3.97)$$

$$\text{Υπό τον όρο ότι} \quad (1-h)c' |x_i| + \alpha' x_i \geq y_i \quad (3.98)$$

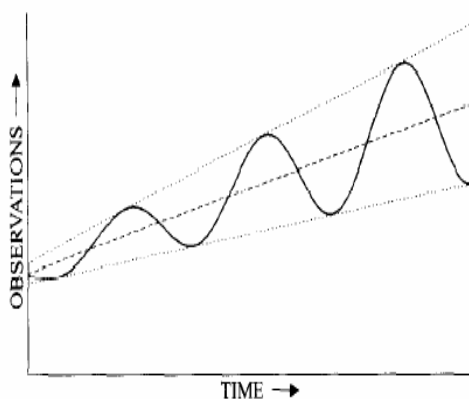
$$(1-h)c' |x_i| - \alpha' x_i \geq -y_i \\ \text{for } i = 1, \dots, N \quad (3.99)$$

Οι εξισώσεις (3.97) - (3.99) έχουν την παραδοχή ότι το ελάχιστο της ασάφειας του GFLM μπορεί να δειχθεί με την προϋπόθεση ότι  $\mu_{y_i}(y_i) \geq h$ , για κάθε  $i$ , όπου  $0 \leq h < 1$  καθορίζεται από αυτόν που παίρνει την απόφαση.

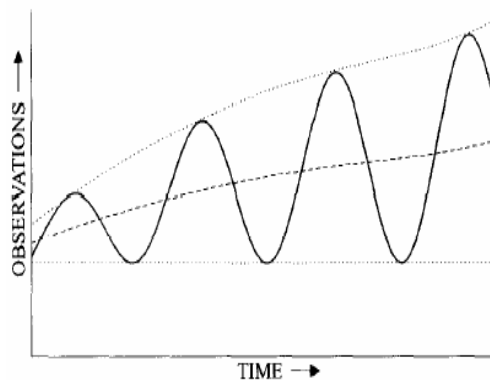
Ενώ η διατύπωση των εξισώσεων (3.95) - (3.97) μπορεί να παρέχει απλά αλλά ισχυρά μέσα στην ασαφή ανάλυση παλινδρόμησης, αυτό που ακολουθεί πρέπει επίσης να σημειωθεί. Οι εξ. (3.97) - (3.99) δεν παρέχουν ένα σύστημα με ένα σωστό πρότυπο (μοντέλο) εάν η υπό όρους τάση και η εκτεταμένη τάση του συστήματος είναι σε σύγκρουση ή όχι απολύτως σύμφωνες (συνεπείς). Επιπλέον, οι συγκεκριμένες αναλύσεις παλινδρόμησης για τις μη συμμετρικά συγκεκριμένες εκτιμήσεις δεν εξετάζονται. Διάφορες μεθοδολογίες για την εξέταση τέτοιων προβλημάτων έχουν προταθεί. Εντούτοις, δεν θα ασχοληθούμε περαιτέρω.

Στα σχήματα 3.30-3.34 φαίνονται μερικά χαρακτηριστικά παραδείγματα της ασαφούς τάσης χρονοσειρών που αναλύονται με την ασαφή ανάλυση παλινδρόμησης. Η συνεχής γραμμή αντιπροσωπεύει τα δεδομένα της χρονοσειράς, η με παύλες γραμμή αντιπροσωπεύει τους τρόπους των συγκεκριμένων εκτιμήσεων παλινδρόμησης, και οι δύο εστιγμένες γραμμές αντιπροσωπεύουν τα ανώτερα και τα χαμηλότερα όρια, αντίστοιχα. Αυτά τα χαρακτηριστικά πρότυπα μπορούν να παρέχουν σημαντικές πληροφορίες για τις πιο πρόσφατες αναλύσεις της εποχικότητάς τους. Παραδείγματος χάριν, στο σχήμα 3.32 παρουσιάζεται μια ασαφής τάση που έχει σταθερό το χαμηλότερο όριο, ενώ το ανώτερο αυξάνεται.

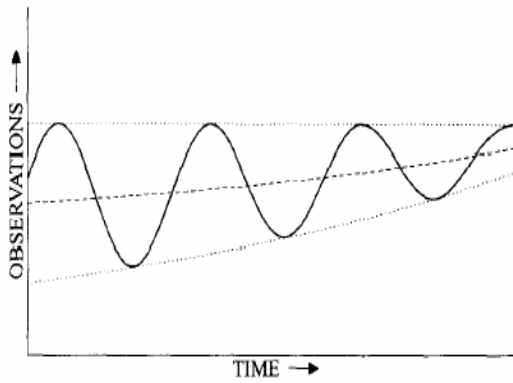
Σχήμα 3.30



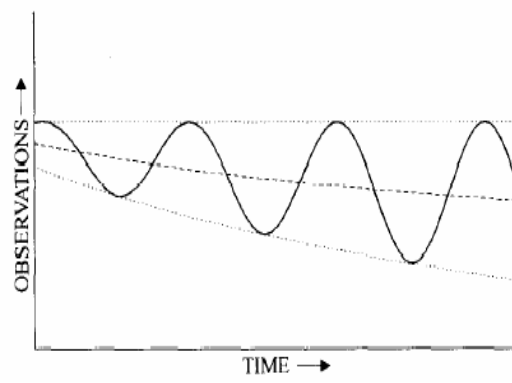
Σχήμα 3.31



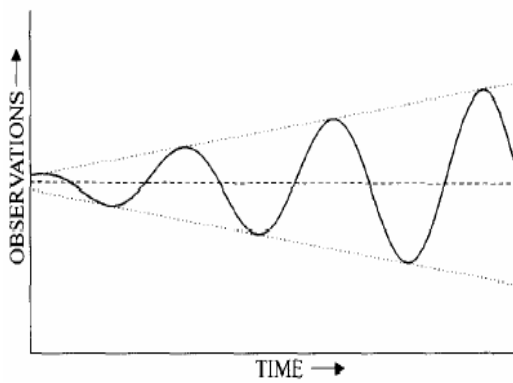
Σχήμα 3.32



Σχήμα 3.33



Σχήμα 3.34



### 3.4.2.3 Υπολογισμός των συνόλων των ασαφών δεικτών εποχιακότητας

Όπως φαίνεται στα σχ. 3.30-3.34, προφανώς το πρότυπο της εποχιακότητας πρέπει να ληφθεί υπόψη. Αυτές οι εποχιακές διακυμάνσεις (μεταβολές) μπορούν να εξηγηθούν άμεσα από το συγκεκριμένο πρότυπο παλινδρόμησης που δημιουργείται. Δύο περιπτώσεις της εποχιακότητας μπορούν να καθοριστούν.

Κατ' αρχάς, υποθέτουμε ότι υπάρχουν  $T + 1$  σύνολα διαθέσιμων στοιχείων (δεδομένων), και το κάθε ένα απ' αυτά τα σύνολα  $t$  ( $t = 0, 1, \dots, T$ ) έχει  $m$  εποχιακά στοιχεία. Από το καθιερωμένο μοντέλο ασαφούς παλινδρόμησης, για περίοδο  $k$  ( $k=1, \dots, m$ ), οι βαθμοί ιδιότητας μέλους των παρατηρήσεων της περιόδου  $k$  στα  $T + 1$  σύνολα στο συγκεκριμένο πρότυπο παλινδρόμησης μπορούν να γίνουν αντιληπτοί από την εξ. (3.100) ως

$$\mu_{Y_k}(y_k), \mu_{Y_{k+m}}(y_{k+m}), \mu_{Y_{k+2m}}(y_{k+2m}), \dots, \mu_{Y_{k+tm}}(y_{k+tm}), \dots, \mu_{Y_{k+T \times m}}(y_{k+T \times m}),$$

Εισάγοντας για κάθε  $k$  ( $k=1, \dots, m$ ) και  $t = 0, 1, \dots, T$

$$S_{k+tm} \equiv \begin{cases} 2 - \mu_{Y_{k+tm}}(y_{k+tm}), & \text{if } y_{k+tm} \geq \alpha^t x_{k+tm}, \\ \mu_{Y_{k+tm}}(y_{k+tm}) & \text{αλλιώς} \end{cases} \quad (3.100)$$

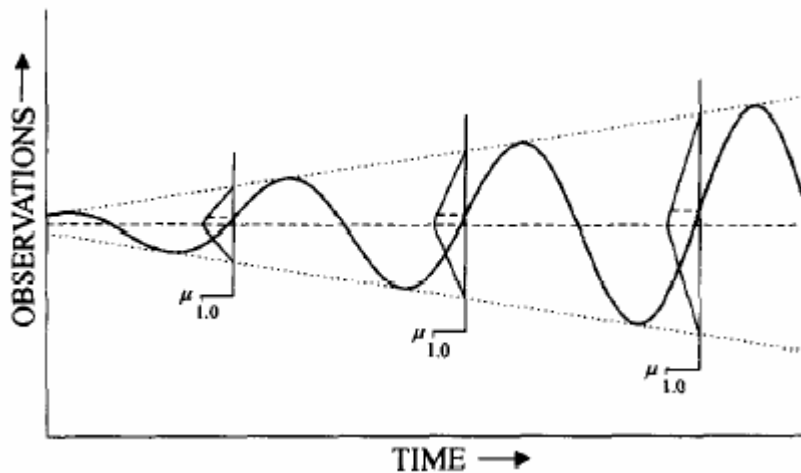
**Ορισμός 1.** Μια χρονοσειρά μπορεί να κληθεί τέλεια (σταθερή) χρονοσειρά ως προς την εποχιακότητα, εάν δεν υπάρχει καμία ασάφεια εμπλεκόμενη στην εποχιακότητα, δηλ.,

$$S_k^* \equiv S_k = S_{k+m} = \dots = S_{k+t \times m} = S_{k+T \times m} \quad \forall k = 1, \dots, m$$

όπου ο  $S_k^*$  καλείται *τέλειος δείκτης εποχιακότητας* της περιόδου  $k (k=1, \dots, m)$

Ο ορισμός 1 δείχνει ότι μόνο σε ιδανικές περιπτώσεις, οι εποχιακές διακυμάνσεις των χρονοσειρών μπορούν να παραμείνουν «σταθερές». Η διακύμανση μιας δεδομένης περιόδου μέσω του χρόνου προκαλείται μόνο από τη ασαφής τάση της χρονοσειράς. Το σχ.3.35 μπορεί να αποτελέσει ένα παράδειγμα που, ωστόσο, μπορεί να προσεγγίσει μόνο μια περίπτωση τέλειας-εποχιακότητας.

Σχήμα 3.35 «Σταθερή» εποχιακή διακύμανση



Επομένως, πρέπει να καθοριστεί η ασαφής εποχιακότητα που οδηγεί στον ακόλουθο ορισμό

**Ορισμός 2.** Μια χρονοσειρά καλείται *χρονοσειρά ασαφούς εποχιακότητας*, εάν υπάρχει ασάφεια που περιλαμβάνεται στην εποχιακότητα. Η ασαφής εποχιακότητα μπορεί να περιγραφεί ως  $m$  κοινά υποσύνολα των  $S_{k+t \times m}$  με τουλάχιστον δύο διαφορετικά-εκτιμώμενα στοιχεία μέχρι τώρα, δηλ.,

$$S_k^m = \{S_k, S_{k+m}, \dots, S_{k+t \times m}, \dots, S_{k+T \times m}\}, \quad k = 1, \dots, m$$

και το  $S_k^s$  καλείται *συγκεκριμένο σύνολο δεικτών εποχιακότητας* της περιόδου  $k (k=1, \dots, m)$

3.4.2.4. Κάνοντας προβλέψεις

Με τα σύνολα δεικτών ασαφούς εποχιακότητας και το ασαφές μοντέλο παλινδρόμησης, η διαδικασία διεξαγωγής προβλέψεων μπορεί να συζητηθεί στις δύο περιπτώσεις: τέλειας εποχιακότητας και ασαφούς εποχιακότητας, ως εξής.

1. Η περίπτωση της τέλειας εποχιακότητας

Λόγω της τέλειας εποχιακότητας, η πρόβλεψη είναι ξεκάθαρη (crisp). Σε αυτήν την περίπτωση, η διεξαγωγή της πρόβλεψης (δείχνεται ως  $f_{k+(T+v) \times m}$  για περίοδο  $k$  σε μελλοντικό χρόνο  $k+(T+v) \times m$  για  $v=1,2,\dots$ ) μπορεί να γίνει ως

$$\begin{aligned} f_{k+(T+v) \times m} &= \mu_{Y_{k+(T+v) \times m}}^{-1} (s_k^*) \\ &= \alpha^t x_{k+(T+v) \times m} + (s_k^* - 1) e^t \left| x_{k+(T+v) \times m} \right| \end{aligned} \quad (3.101)$$

2. Η περίπτωση της ασαφούς εποχιακότητας

Στην περίπτωση της ασαφούς εποχιακότητας ωστόσο, η πρόβλεψη μπορεί να γίνει είτε σαφής είτε ασαφής.

Οι σαφείς προβλέψεις

Το να γίνει σαφής πρόβλεψη σε αυτήν την περίπτωση απαιτεί την εύρεση ενός αντιπροσωπευτικού δείκτη (δείχνεται ως  $S_k^*$ ) από το σύνολο των δεικτών ασαφούς εποχιακότητας. Εδώ λοιπόν, οι ακόλουθοι τελεστές μπορεί να είναι κατάλληλοι.

*Περίπτωση 1. W-period min-operator ( $1 \leq W \leq T$ ):*

Είναι κατάλληλος για περίοδο  $k$  η οποία παρουσιάζει μια αυστηρά μειούμενη τάση στο σύνολο των δεικτών ασαφούς εποχιακότητας  $S_k^s$ , δηλαδή

$$\begin{aligned} \text{Αν } \mu_{Y_{k+(T-W+1) \times m}} \left( y_{k+(T-W+1) \times m} \right) \gg \mu_{Y_{k+(T-W+2) \times m}} \left( y_{k+(T-W+2) \times m} \right) \gg \dots \gg \mu_{Y_{k+T \times m}} \left( y_{k+T \times m} \right), \text{ τότε} \\ s_k^* \equiv s_i = \min \left( \mu_{Y_{k+(T-W+1) \times m}} \left( y_{k+(T-W+1) \times m} \right), \mu_{Y_{k+(T-W+2) \times m}} \left( y_{k+(T-W+2) \times m} \right), \dots, \mu_{Y_{k+T \times m}} \left( y_{k+T \times m} \right) \right). \end{aligned} \quad (3.102)$$

*Περίπτωση 2. W-period max-operator ( $1 \leq W \leq T$ ):*



Είναι κατάλληλος για περίοδο  $k$  η οποία παρουσιάζει μια αυστηρά αυξανόμενη τάση στους δικούς της  $S_k^S$

$$\text{Αν } \mu_{Y_{k+(T-W+1) \times m}}(y_{k+(T-W+1) \times m}) \langle \mu_{Y_{k+(T-W+2) \times m}}(y_{k+(T-W+2) \times m}) \rangle \dots \langle \mu_{Y_{k+T \times m}}(y_{k+T \times m}) \rangle, \text{ τότε}$$

$$s_k^* \equiv s_k^+ = \max \left( \mu_{Y_{k+(T-W+1) \times m}}(y_{k+(T-W+1) \times m}), \mu_{Y_{k+(T-W+2) \times m}}(y_{k+(T-W+2) \times m}), \dots, \mu_{Y_{k+T \times m}}(y_{k+T \times m}) \right). \quad (3.103)$$

*Περίπτωση 3. W-period smoothing operators ( $1 \leq W \leq T$ ):*

Μπορούν να χρησιμοποιηθούν όταν δεν είναι ούτε η περίπτωση 1 ούτε η περίπτωση 2. Τότε, δύο τελεστές μπορούν να επιλεγούν από

(α) *W-period geometric-mean operator (Γεωμετρικός-μέσος χειριστής):*

$$s_k^* \equiv s_i^G = \left( \prod_{w=0}^{W-1} \mu_{Y_{k+(T-w) \times m}}(y_{k+(T-w) \times m}) \right)^{1/W} \quad (3.104)$$

για  $\mu_{Y_{k+(T-w) \times m}}(y_{k+(T-w) \times m}) > 0$  και  $\forall w$

(β) *W-period arithmetic-mean operator: (Αριθμητικός-μέσος χειριστής)*

$$s_k^* \equiv s_i^D = \frac{\left( \sum_{w=0}^{W-1} \mu_{Y_{k+(T-w) \times m}}(y_{k+(T-w) \times m}) \right)}{W} \quad (3.105)$$

Εφόσον βρεθεί ο αντιπροσωπευτικός δείκτης  $S_k^*$ , τότε μια σταθερή πρόβλεψη για την περίοδο  $k$  σε έναν μελλοντικό χρόνο μπορεί να ληφθεί με τη χρήση της εξ.(3.101). Επιπλέον, με τις εξ. (3.102)-(3.105), οι προβλέψεις πρέπει να αναμένονται περισσότερο ευαίσθητες στις αλλαγές των πιο πρόσφατων δεδομένων, καθώς το  $W$  μειώνεται.

### Οι ασαφείς προβλέψεις

Σε αυτήν την περίπτωση, μια ασαφής πρόβλεψη θα έπρεπε να είναι μια σειρά από προβλέψεις (ένα εύρος προβλέψεων), αλλά με μια συνάρτηση ιδιότητας μέλους η οποία θα ερμηνεύει την αβεβαιότητα που υπάρχει στις χρονοσειρές. Τότε, η παραγωγή μιας ασαφούς πρόβλεψης απαιτεί πρώτα τον ορισμό ενός *δείκτη ασαφούς εποχιακότητας* (ένας ασαφής αριθμός που δείχνεται ως  $S_k^*$ ). Καθορίζουμε αυτόν τον δείκτη ασαφούς εποχιακότητας  $S_k^*$  από τα σύνολα δεικτών ασαφούς εποχιακότητας ως

$$S_k^* = (s_k^M, s_k^L, s_k^R)_{LR}, \quad (3.106)$$

$$\text{όπου } S_k^M = \begin{cases} s_k^- & \text{για την Περίπτωση 1} \\ s_k^+ & \text{για την Περίπτωση 2} \\ s_k^G \text{ ή } s_k^G & \text{για την Περίπτωση 3} \end{cases} \quad (3.107)$$

$$S_k^L = \begin{cases} 0 & \text{για την Περίπτωση 1} \\ s_k^M - s_k^- & \text{για τις Περίπτώσεις 2 και 3} \end{cases} \quad (3.108)$$

$$S_k^R = \begin{cases} s_k^M - s_k^- & \text{για τις Περίπτώσεις 1 και 3} \\ 0 & \text{για την Περίπτωση 2} \end{cases} \quad (3.109)$$

Το  $S_k^*$  ορίζεται ώστε να έχει μια μη συμμετρική τριγωνική συνάρτηση ιδιότητας μέλους .

Με αυτόν τον δείκτη ασαφούς εποχιακότητας  $S_k^*$  (ένας ασαφής αριθμός), μια ασαφής πρόβλεψη ( $F_{k+(T+\nu) \times m}$ ) για περίοδο  $k$  σε χρόνο  $k+(T+\nu) \times m$  για  $\nu=1, 2, \dots$  μπορεί να ληφθεί ως

$$F_{k+(T+\nu) \times m} = \left( f_{k+(T+\nu) \times m}^M, f_{k+(T+\nu) \times m}^L, f_{k+(T+\nu) \times m}^R \right)_{LR} \quad (3.110)$$

όπου

$$f_{k+(T+\nu) \times m}^M = \mu_{Y_{k+(T+\nu) \times m}}^{-1} \left( S_k^M \right) \quad (3.111)$$

$$f_{k+(T+\nu) \times m}^L = f_{k+(T+\nu) \times m}^M - \min \left[ \mu_{Y_{k+(T+\nu) \times m}}^{-1} \left( S_k^- \right), \mu_{Y_{k+(T+\nu) \times m}}^{-1} \left( S_k^+ \right) \right] \quad (3.112)$$

$$f_{k+(T+\nu) \times m}^R = f_{k+(T+\nu) \times m}^M - \max \left[ \mu_{Y_{k+(T+\nu) \times m}}^{-1} \left( S_k^+ \right), \mu_{Y_{k+(T+\nu) \times m}}^{-1} \left( S_k^- \right) \right] - f_{k+(T+\nu) \times m}^M \quad (3.113)$$

όπου  $\mu_{Y_{k+(T+\nu) \times m}}^{-1} \left( S_k^M \right), \mu_{Y_{k+(T+\nu) \times m}}^{-1} \left( S_k^- \right)$  και  $\mu_{Y_{k+(T+\nu) \times m}}^{-1} \left( S_k^+ \right)$  λαμβάνονται με χρήση της εξ. (3.101). Η ασαφής πρόβλεψη είναι ένας μη συμμετρικός τριγωνικός ασαφής αριθμός

### 3.4.2.5 Εποχιακή ασάφεια και τάσεις

Έτσι όπως ορίζεται, ο δείκτης ασαφούς εποχιακότητας αντιπροσωπεύει την εποχιακή ασάφεια μιας περιόδου σε μια χρονοσειρά. Ένας ευρύς, μεγάλος δείκτης ασαφούς εποχιακότητας αντιπροσωπεύει μια ευρύτερη εποχική διακύμανση και μια μεγαλύτερη αβεβαιότητα σε εκείνη την περίοδο. Επομένως, οι δείκτες ασαφούς εποχιακότητας μπορούν να παρέχουν μια κοινή βάση για τη σύγκριση της αβεβαιότητας των προβλέψεων των περιόδων.

Επιπλέον, οι εποχιακές τάσεις μπορούν να αναλυθούν. Ειδικότερα, εάν ενδιαφερόμαστε για τις υπό όρους τάσεις των εποχών, τότε, οι ακόλουθες αναλύσεις μπορούν να πραγματοποιηθούν:

(1) Σχεδιάζονται τα κατ' εκτίμηση στοιχεία

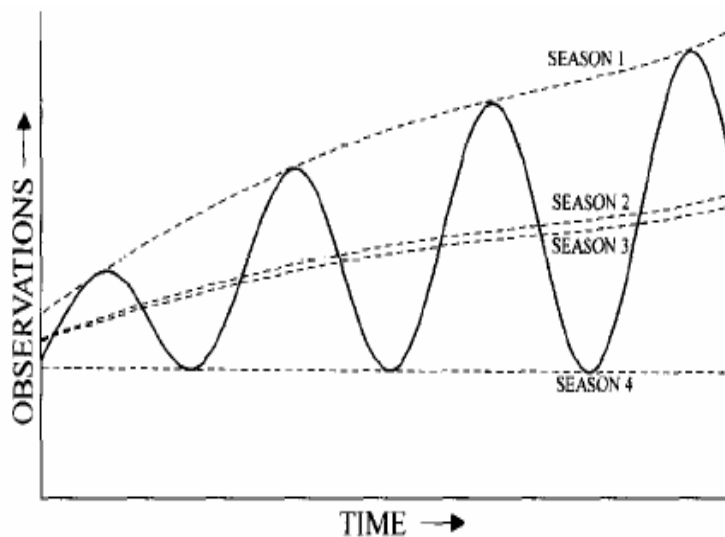
$$\mu_{y_i}^{-1}(s_k^M) = \alpha' x_i + (s_k^M - 1) c' |x_i| \text{ για } i = 1, 2, \dots, N, \quad (3.114)$$

για κάθε περίοδο  $k$  ( $k = 1, \dots, m$ ). Παραδείγματος χάριν, στο σχήμα 3.36 μετά από τη σχεδίαση της εξ. (3.114), η περίοδος 4 παρουσιάζει μια τάση που είναι σταθερή, ενώ οι περίοδοι 1-3 έχουν αυξητική τάση. Επιπλέον, η αύξηση της τάσης της περιόδου 3 εμφανίζεται να είναι πιο αργή από αυτή της περιόδου 2, η οποία είναι στη συνέχεια πιο αργή από αυτή της περιόδου 1.

(2) Συγκρίνεται η παράγωγος του δείκτη  $I_k$ ,

$$I_k = \left. \frac{d\mu_{y_i}^{-1}(s_k^M)}{dx} \right|_{k+t \times m} \quad k = 1, 2, \dots, m, \quad (3.115)$$

Σχήμα 3.36 Σχεδίαση των κατ' εκτίμηση στοιχείων



### 3.4.2.6. Εφαρμογή σε πρόβλημα πρόβλεψης πωλήσεων

Αυτή η μέθοδος εφαρμόζεται σε πρόβλημα πρόβλεψης πωλήσεων, κάποιας επιχείρησης διανομής τροφίμων στην Ταϊβάν. Τα σταθερά στοιχεία εισόδου-εξόδου των μηνιαίων πωλήσεων ενός γαλακτοκομικού προϊόντος (σε κιβώτια) αυτής της επιχείρησης κατά τη διάρκεια των ετών 1992-1994 παρατίθενται στον πίνακα 3.15. Το GFLM αποφασίζεται ως εξής:  $Y = A_0 + A_1x + A_2x^2$ , το οποίο είναι το πρότυπο που ταιριάζει καλύτερα στα δεδομένα στοιχεία. Οι ασαφείς παράμετροι που λαμβάνονται

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

με τη χρήση των εξ. (3.97)-(3.99) (με  $h = 0$ ) παρουσιάζονται στον πίνακα 3.16. Αυτά τα αποτελέσματα συνδέονται έπειτα με το GFLM και σχεδιάζονται στο σχ. 3.37

Βασισμένα σε αυτό το πρότυπο ασαφούς παλινδρόμησης, τα σύνολα των δεικτών ασαφούς εποχιακότητας και οι δείκτες ασαφούς εποχιακότητας για τις 12 περιόδους (μήνες) υπολογίζονται όπως φαίνονται στον πίνακα 3.15, στον οποίο το Min δείχνει τον W-period min-operator, το MAX τον W-period max-operator και το AVG τον W-period arithmetic-mean operator, όλα με  $W=3$ .

### ΠΙΝΑΚΑΣ 3.15

Μηνιαίες Πωλήσεις Γαλακτοκομικού Προϊόντος (σε κουτιά)  
κατά την περίοδο 1992 -1994

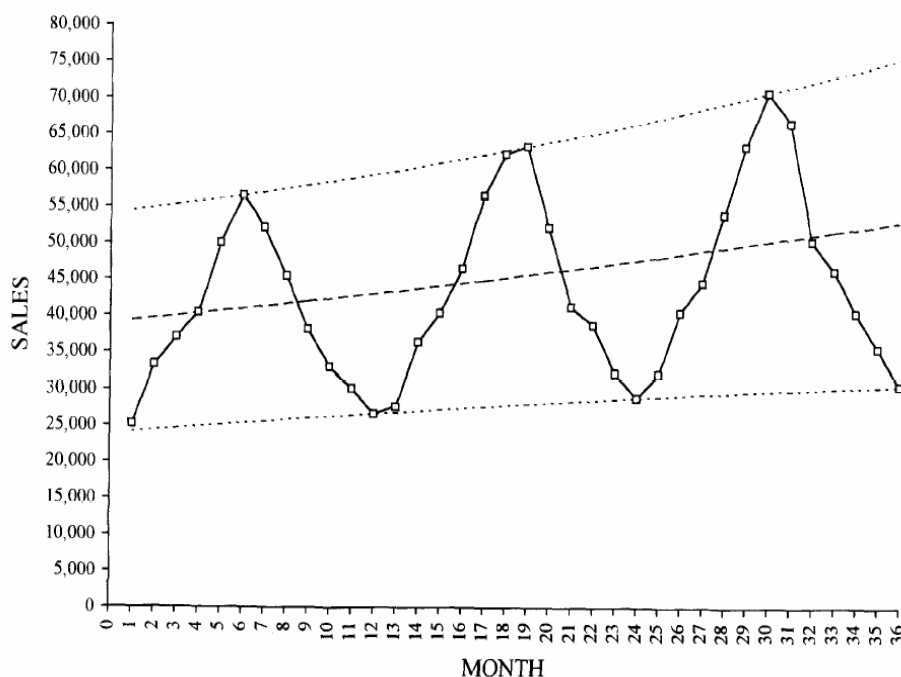
ΜΗΝΑΣ (k)	Μηνιαίες πωλήσεις το 1992	x	Μηνιαίες πωλήσεις το 1993	x	Μηνιαίες πωλήσεις το 1994	x
1	25200	1	27697	13	32113	25
2	33244	2	36410	14	40456	26
3	37008	3	40400	15	44562	27
4	40411	4	46528	16	53977	28
5	49997	5	56559	17	63277	29
6	56523	6	62226	18	70711	30
7	52111	7	63321	19	66543	31
8	45433	8	52177	20	50332	32
9	38115	9	41288	21	46233	33
10	33005	10	38785	22	40467	34
11	30117	11	32229	23	35612	35
12	26661	12	28884	24	30555	36

### ΠΙΝΑΚΑΣ 3.16 Ασαφείς παράμετροι $A=(A_0, A_1, A_2); (h=0)$

Fuzzy parameters  $A = (A_0, A_1, A_2); (h = 0)$

Fuzzy parameter	$A_0$	$A_1$	$A_2$
Mode $\alpha_j$	38998.36	310.412	2.1541
Spread $c_j$	15094.36	57.412	4.0499

Σχήμα 3.37 Αποτελέσματα της ανάλυσης ασαφούς παλινδρόμησης των πωλήσεων γαλακτοκομικού προϊόντος κατά τα έτη 1992-1994



**ΠΙΝΑΚΑΣ 3.17** Συγκεκριμένα σύνολα δεικτών εποχιακότητας και συγκεκριμένοι δείκτες εποχιακότητας

Fuzzy seasonality index sets and fuzzy seasonality indices

MONTH (k)	$\{s_k,$ (1992)	$s_{k+12},$ (1993)	$s_{k+12 \times 12}\}$ (1994)	Operator for $s_k^M$	$(s_k^M,$	$s_k^L,$	$s_k^R)_{LR}$	$s_k^L + s_k^R$
1	0.0689	0.0499	0.1610	AVG	0.0933	0.0434	0.0677	0.1111
2	0.5807	0.5593	0.5824	AVG	0.5741	0.0149	0.0083	0.0232
3	0.8078	0.7783	0.7761	MIN	0.7761	0	0.0317	0.0317
4	1.0089	1.1180	1.2312	MAX	1.2313	0.2225	0	0.2225
5	1.6067	1.6764	1.6677	AVG	1.6503	0.0436	0.02611	0.0697
6	2	1.9715	2	AVG	1.9905	0.0190	0.0095	0.0285
7	1.6903	2	1.7634	AVG	1.8179	0.1276	0.1821	0.3097
8	1.2412	1.3420	0.9618	AVG	1.1816	0.2199	0.1603	0.3802
9	0.7584	0.7136	0.7498	AVG	0.7406	0.0269	0.0178	0.0447
10	0.4206	0.5586	0.4673	AVG	0.4822	0.0616	0.0765	0.1380
11	0.2257	0.1891	0.2346	AVG	0.2164	0.0273	0.0181	0.0455
12	0	0	0		0	0	0	0

Κατόπιν, οι ασαφείς προβλέψεις για τις μηνιαίες πωλήσεις το 1995 υπολογίζονται όπως φαίνεται στον πίνακα 3.18 και στο σχέδιο 3.38. Η ασάφεια αυτών των προβλέψεων υπολογίζεται όπως φαίνεται στη στήλη  $(S_k^L + S_k^R)$  του πίνακα 3.17. Αποκαλύπτεται ότι ο ΜΗΝΑΣ 7 (Ιούλιος) έχει την ευρύτερη ασαφή διακύμανση μεταξύ όλων των ΜΗΝΩΝ και ο ΜΗΝΑΣ 12 (Δεκέμβριος) έχει τη μικρότερη ασάφεια.

Επιπλέον, οι τάσεις των 12 μηνών αναλύονται:

(1) με τη χάραξη των εκτιμώμενων καμπυλών (εξ. (3.114)) στο σχ 3.39

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

και

(2) με τον υπολογισμό του παράγωγου δείκτη  $I_k$  (εξ. (3.115)) στον πίνακα 3.19.

Είναι φανερό ότι οι τάσεις των 12 μηνών είναι όλες αυξανόμενες. Οι αυξανόμενες τάσεις των πρώτων έξι μηνών είναι στην αύξουσα σειρά: ΜΗΝΑΣ 1, ΜΗΝΑΣ 2, ΜΗΝΑΣ 3, ΜΗΝΑΣ 4, ΜΗΝΑΣ 5, ΜΗΝΑΣ 6. Αλλά, οι αυξανόμενες τάσεις των επόμενων έξι μηνών επιβραδύνουν με την κατιούσα σειρά: ΜΗΝΑΣ 7, ΜΗΝΑΣ 8, ΜΗΝΑΣ 9, ΜΗΝΑΣ 10, ΜΗΝΑΣ 11, ΜΗΝΑΣ 12. Επιπλέον, η αυξητική τάση του ΜΗΝΑ 1 (Ιανουάριος) είναι γρηγορότερη από αυτή του ΜΗΝΑ 12 (Δεκέμβριος). Η αυξητική τάση του ΜΗΝΑ 7 (Ιούλιος) είναι γρηγορότερη από αυτή του ΜΗΝΑ 5 (Μάιος).

**ΠΙΝΑΚΑΣ 3.18** Συγκεχυμένες προβλέψεις των μηνιαίων πωλήσεων για το 1995

### Fuzzy forecasts of monthly sales for 1995

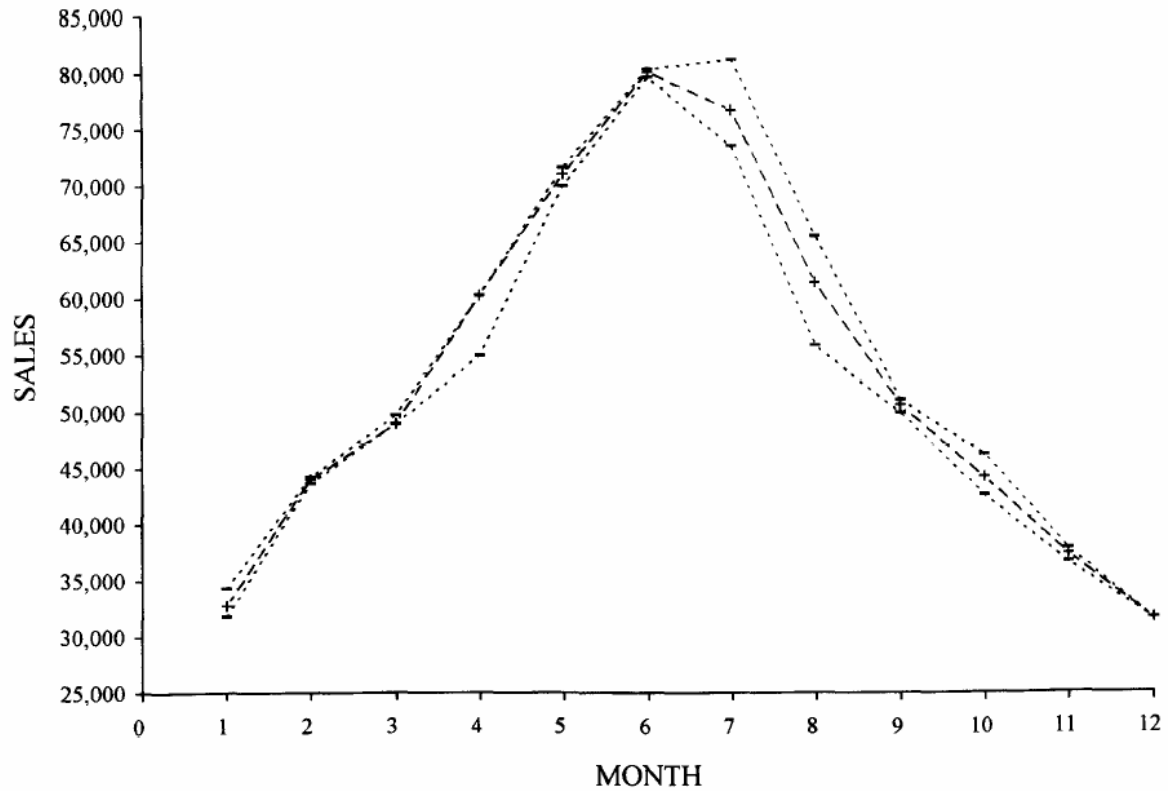
MONTH, $k$	$F_{k+3 \times 12} = (f_{k+3 \times 12}^M, f_{k+3 \times 12}^L, f_{k+3 \times 12}^R)_{LR}$		
1	32792.9	987.7	1541.7
2	44057.1	343.6	191.8
3	49120.5	0	745.2
4	60383.6	5310.4	0
5	71118.9	1057.4	633.4
6	80250.7	468.9	234.5
7	76818.1	3196.0	4562.2
8	61451.7	5597.7	4082.4
9	50615.7	697.3	459.8
10	44213.6	1619.2	2011.3
11	37394.4	731.3	484.5
12	31680.0	0	0

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

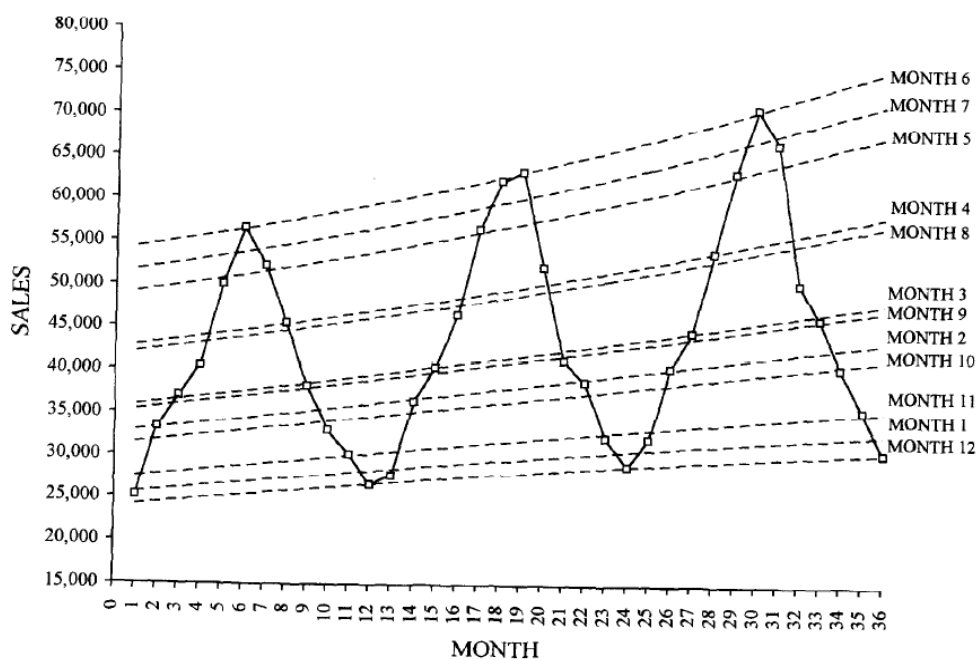
**ΠΙΝΑΚΑΣ 3.19** Δείκτες Παραγώγου  $I_k$

Derivative indices $I_k$												
MONTH, $k$	1	2	3	4	5	6	7	8	9	10	11	12
$I_k$	182.45	308.29	364.91	496.79	625.42	737.20	696.29	505.78	368.35	284.55	194.08	116.50

**Σχήμα 3.38** Ασαφείς προβλέψεις για τις πωλήσεις του γαλακτοκομικού προϊόντος για το έτος 1995



**Σχήμα 3.39** Modal Trend Analysis



### 3.4.2.7 Επίλογος -Συμπέρασμα

Σε αυτό το τμήμα αναπτύσσεται μια ασαφής μέθοδος πρόβλεψης εποχιακότητας. Καθορίζονται η ασαφής τάση και η ασαφής εποχιακότητα σε μια χρονοσειρά. Προφανώς, το δεύτερο εξαρτάται από τα πρώτο. Η μέθοδος ασαφούς πρόβλεψης για ένα πρότυπο χρονοσειράς, διατυπώνεται με το συσχετισμό της πληροφορίας της (γενικής) ασαφούς τάσης με αυτήν των μεμονωμένων περιόδων, με την εφαρμογή των δεικτών ασαφούς εποχιακότητας στο πρότυπο ασαφούς παλινδρόμησης.

Ασαφής προβλέψεις επομένως μπορούν να γίνουν. Η μέθοδος ασαφούς πρόβλεψης και η ανάλυση της ασάφειας των προβλέψεων είναι τα χαρακτηριστικά γνωρίσματα αυτής της μεθόδου που την κάνουν να διαφέρει από τις παραδοσιακές τεχνικές πρόβλεψης για την εποχιακότητα.

### 3.4.3 Διαχείριση προβλημάτων πρόβλεψης με χρήση ασαφών χρονοσειρών [22]

#### 3.4.3.1. Εισαγωγή

Οι δραστηριότητες πρόβλεψης διαδραματίζουν σημαντικό ρόλο στην καθημερινή μας ζωή. Προβλέπουμε συχνά τον καιρό, τους σεισμούς, το χρηματιστήριο, και οτιδήποτε που οι άνθρωποι θέλουν να προγνώσουν. Οι άνθρωποι μπορούν να προγραμματίσουν ή να αποτρέψουν κάτι εκ των προτέρων με τις δραστηριότητες πρόβλεψης. Είναι αδύνατο να γίνει μια εκατό τοις εκατό πρόβλεψη, αλλά μπορούμε να κάνουμε το καλύτερό δυνατό προκειμένου να αυξήσουμε την ακρίβεια των προβλέψεων. Οι παραδοσιακές μέθοδοι πρόβλεψης μπορούν να εξετάσουν πολλές περιπτώσεις πρόβλεψης, αλλά δεν μπορούν να λύσουν τα προβλήματα πρόβλεψης στα οποία τα ιστορικά στοιχεία είναι γλωσσικές τιμές. Οι Song και Chissom έχουν προτείνει ένα πρώτης τάξης χρονικά-αμετάβλητο ασαφές πρότυπο χρονοσειράς και ένα πρώτης τάξης χρονικά-μεταβλητό ασαφές πρότυπο χρονοσειράς για να λύσουν τα προβλήματα πρόβλεψης. Προβλέπουν τις εγγραφές του πανεπιστημίου της Αλαμπάμα με τη χρήση 20ετών ιστορικών στοιχείων εγγραφής. Το μέσο λάθος πρόβλεψης είναι 3,18% στο [26] και κυμαίνονται από 4,49% ως 3,15% στο [27] και φαίνεται ότι τα αποτελέσματα πρόβλεψης είναι καλύτερα από εκείνα που προκύπτουν με τη χρήση των παραδοσιακών προτύπων πρόβλεψης. Οι Song και Chissom [26,27,28] χρησιμοποίησαν την ακόλουθη μέθοδο για να προβλέψουν τις εγγραφές του πανεπιστημίου της Αλαμπάμα:

$$A_i = A_{i-1} \circ R, \quad (3.116)$$

όπου το  $A_{i-1}$  είναι η εγγραφή του έτους  $i-1$ ,  $A_i$  είναι η προβλεπόμενη εγγραφή του έτους  $i$  με όρους ασαφών συνόλων, το  $R$  είναι η ένωση των ασαφών σχέσεων και “ $\circ$ ” είναι ο μέγιστος -ελάχιστος τελεστής σύνθεσης.



Σε αυτό το κεφάλαιο, παρουσιάζουμε μια νέα μέθοδο βασισμένη στις χρονικά μεταβλητές ασαφής χρονοσειρές για να αντιμετωπίσουμε τα προβλήματα πρόβλεψης, όπου τα ιστορικά στοιχεία εγγραφής χρησιμοποιούνται για να επεξηγήσουν τη διαδικασία πρόβλεψης.

Η ιδέα της προτεινόμενης μεθόδου είναι ότι η διακύμανση της εγγραφής αυτού του έτους συσχετίζεται με την τάση των εγγραφών των προηγούμενων ετών. Παραδείγματος χάριν, εάν η τάση των εγγραφών των προηγούμενων ετών αυξάνεται, τότε ο αριθμός των εγγραφών αυτού του έτους πιθανόν να αυξηθεί. Για να καθορίσουμε το βαθμό των διακυμάνσεων, εκτελούμε τους συστηματικούς υπολογισμούς για να υπολογίσουμε τη σχέση μεταξύ των διακυμάνσεων του τελευταίου χρόνου και των διακυμάνσεων των προηγούμενων ετών. Κατόπιν, μπορούμε να πάρουμε τις προβλεπόμενες εγγραφές από την παραγόμενη σχέση. Η διακύμανση των εγγραφών του περασμένου χρόνου είναι ένα κριτήριο για να προβλέψει τις εγγραφές του επόμενου έτους εξαιτίας του γεγονότος ότι θεωρούμε ότι η διακύμανση αυτού του έτους είναι η «πιο παρόμοια», η πιο κοντινή με την μεταβολή - διακύμανση του περασμένου χρόνου. Τα μέσα λάθη πρόβλεψης και η χρονική πολυπλοκότητα των διάφορων μεθόδων πρόβλεψης συγκρίνονται. Η προτεινόμενη μέθοδος είναι αποδοτικότερη από αυτές που παρουσιάζονται μέσα [26,27,] εξαιτίας του γεγονότος ότι απλοποιεί την αριθμητική διαδικασία. Επιπλέον, το μέσο λάθος πρόβλεψης της προτεινόμενης μεθόδου είναι μικρότερο από αυτές που παρουσιάζονται στα [26,27].

### 3.4.3.2 Έννοιες ασαφών χρονοσειρών

Σε αυτό το τμήμα, αναθεωρούμε την έννοια των ασαφών χρονοσειρών και των μεθόδων πρόβλεψης. Έστω  $U$  το σύνολο του λόγου (the universe of discourse), όπου  $U = \{u_1, u_2, \dots, u_n\}$ . Ένα ασαφές σύνολο  $A$  του  $U$  καθορίζεται ως εξής:

$$A = \mu_A(u_1)/u_1 + \mu_A(u_2)/u_2 + \dots + \mu_A(u_n)/u_n, \quad (3.117)$$

όπου  $\mu_A$  είναι η συνάρτηση ιδιότητας μέλους (MEMBERSHIP FUNCTION) του  $A$ ,  $\mu_A : U \rightarrow [0,1]$ ,  $\mu_A(u_i)$  δείχνει το βαθμό ιδιότητας μέλους του  $u_i$  στο  $A$ ,  $\mu_A(u_i) \in [0,1]$ . Το σύμβολο “/” χωρίζει τους βαθμούς ιδιότητας μέλους από τα στοιχεία του συνόλου του λόγου  $U$ , και το σύμβολο “+” σημαίνει “την ένωση” και όχι το συνήθως χρησιμοποιημένο αλγεβρικό σύμβολο του αθροίσματος.

Έστω  $Y(t) \ t=(\dots, 0,1,2,\dots)$ , ένα υποσύνολο του  $R^1$ , είναι το σύνολο του λόγου στο οποίο ορίζεται το ασαφές σύνολο  $\mu_i(t)(i=1,2,\dots)$  και έστω  $F(t)$  είναι μια συλλογή από  $\mu_i(t)(i=1,2,\dots)$ . Κατόπιν, το  $F(t)$  καλείται ασαφής χρονοσειρά στο  $Y(t) \ t=(\dots, 0,1,2,\dots)$ . Είναι προφανές ότι η  $F(t)$  μπορεί να θεωρηθεί ως γλωσσική μεταβλητή, και τα  $\mu_i(t)$  μπορούν να θεωρηθούν ως πιθανές γλωσσικές τιμές της  $F(t)$ , όπου τα  $\mu_i(t)(i=1,2,\dots)$  αντιπροσωπεύονται από ασαφή σύνολα. Επιπλέον, μπορούμε επίσης να δούμε ότι η  $F(t)$  είναι μια λειτουργία του χρόνου  $t$ , δηλ., οι τιμές της  $F(t)$  μπορεί να είναι διαφορετικές σε διαφορετικές χρονικές στιγμές. Εάν η  $F(t)$  προκαλείται από την  $F(t-1)$  μόνο, τότε αυτή η σχέση αναπαρίσταται ως

$$F(t-1) \rightarrow F(t).$$

Έστω η  $F(t)$  είναι μια ασαφής χρονοσειρά. Εάν για οποτεδήποτε χρονική στιγμή  $t$ ,  $F(t) = F(t-1)$  και η  $F(t)$  έχει μόνο πεπερασμένα στοιχεία, τότε η  $F(t)$  καλείται χρονικά-αμετάβλητη ασαφής χρονοσειρά. Διαφορετικά, καλείται χρονικά-μεταβλητή ασαφής χρονοσειρά.

Οι Song και Chissom πρότειναν το μοντέλο χρονικά-μεταβλητής ασαφούς χρονοσειράς και πρόβλεψαν τις εγγραφές του πανεπιστημίου της Αλαμπάμα βασισμένοι σ' αυτές. Η μέθοδος των Song - Chissom για τις εγγραφές του πανεπιστημίου της Αλαμπάμα περιγράφεται εν συντομία ως εξής:

*Βήμα 1:* Καθορισμός του συνόλου του λόγου  $U$  μέσα στο οποίο τα ασαφή σύνολα θα οριστούν.

*Βήμα 2:* Χωρισμός του συνόλου του λόγου  $U$  σε διάφορα ίσου μήκους διαστήματα.

*Βήμα 3:* Καθορισμός μερικών γλωσσικών τιμών.

*Βήμα 4:* «Ασαφοποίηση» των ιστορικών στοιχείων εγγραφής.

*Βήμα 5:* Επιλογή μια κατάλληλης παραμέτρου  $w$  (όπου  $w > 1$ ), υπολογισμός της ασαφούς λειτουργίας  $R^w(t,t-1)$  και πρόβλεψη της εγγραφής. Η μέθοδος της χρονικά-μεταβλητής ασαφούς χρονοσειράς μπορεί να εκφραστεί ως

$$F(t) = F(t-1) \circ R^w(t,t-1), \quad (3.118)$$

όπου

$$R^w(t,t-1) = F^T(t-2) \times F(t-1) \cup F^T(t-3) \times F(t-2) \cup \dots \cup F^T(t-w) \times F(t-w+1) \quad (3.109)$$

και όπου το  $w$  καλείται πρότυπη βάση η οποία δείχνει, υποδηλώνει τον αριθμό των ετών πριν από το  $t$ ,  $F(t)$  είναι η τιμή της ασαφούς χρονοσειράς, το " $\times$ " είναι ο καρτεσιανός τελεστής του γινομένου, και το  $T$  είναι τελεστής αναστροφής.

*Βήμα 6:* «Αποσαφήνιση» των αποτελεσμάτων πρόβλεψης. Στο [27], χρησιμοποιείται η μέθοδος νευρωνικών δικτύων για την «Αποσαφήνιση» των αποτελεσμάτων πρόβλεψης. Στο [29], οι Sullivan και Woodall χρησιμοποιούν το μαρκοβιανό πρότυπο για να προβλέψουν την εγγραφή των

σπουδαστών της Αλαμπάμα. Χρησιμοποιούν το ακόλουθο πρότυπο για να εξετάσουν τα προβλήματα πρόβλεψης

$$P'_{t+1} = P'_t * R_m \quad (3.119)$$

όπου  $P_t$  είναι το διάνυσμα των πιθανών καταστάσεων στο χρόνο  $t$ ,  $P_{t+1}$  είναι το διάνυσμα των πιθανών καταστάσεων στο χρόνο  $(t + 1)$ , και  $R_m$  είναι η μήτρα μετάβασης. Ο πολλαπλασιασμός στον τύπο (3.119) είναι ένας συμβατικός πολλαπλασιασμός μητρών. Επειδή το  $R_m$  στον τύπο (3.119) δεν αλλάζει με το χρόνο, πρόκειται για το μοντέλο χρονικά-αμεταβλητής ασαφούς χρονοσειράς. Μια παραλλαγή του μαρκοβιανού προτύπου είναι η ακόλουθη:

$$P'_{t+1} = P'_t * R_m^k, \quad k = 1, 2, \dots, \quad (3.120)$$

όπου το  $R_m^k$  ποικίλλει με το χρόνο, και είναι ένα μοντέλο χρονικά-μεταβλητής ασαφούς χρονοσειράς.

### 3.4.3.3 Μια νέα μέθοδος για το χειρισμό προβλημάτων πρόβλεψης βασισμένα σε ασαφείς χρονοσειρές

Σε αυτό το τμήμα, παρουσιάζουμε μια νέα μέθοδο για την αντιμετώπιση των προβλημάτων πρόβλεψης. Υποθέτουμε ότι η εγγραφή του έτους  $t$  είναι  $x$  και ότι η εγγραφή του έτους  $t - 1$  είναι  $y$ , τότε η μεταβολή των εγγραφών μεταξύ του έτους  $t$  και του έτους  $t - 1$  είναι ίση με το  $x - y$ . Αρχικά, περιγράφουμε μερικούς εμπειρικούς κανόνες που είναι παρόμοιοι με την ανθρώπινη σκέψη:

*Κανόνας 1:* Η μεταβολή-διακύμανση των εγγραφών μεταξύ του τρέχοντος έτους και του προηγούμενου σχετίζεται με τις μεταβολές-διακυμάνσεις των εγγραφών μεταξύ του τρέχοντος έτους και των προηγούμενων ετών, και η σχέση των εγγραφών μεταξύ του τρέχοντος έτους και του προηγούμενου είναι πιο στενή από αυτή μεταξύ του τρέχοντος έτους και των υπολοίπων προηγούμενων ετών.

*Κανόνας 2:* Εάν η τάση του αριθμού των εγγραφών των προηγούμενων ετών αυξάνεται, τότε ο αριθμός εγγραφών αυτού του έτους αυξάνεται. Εάν η τάση του αριθμού εγγραφών των προηγούμενων ετών μειώνεται, τότε ο αριθμός εγγραφών αυτού του έτους μειώνεται.

Από τους κανόνες 1 και 2, μπορεί να προκύψουν δύο προβλήματα. Αφενός, εάν η τάση των μεταβολών των εγγραφών των προηγούμενων ετών δεν είναι τόσο εμφανής, πώς μπορούμε να ξέρουμε την τάση της μεταβολής της εγγραφής φέτος; Αφετέρου, πώς να καθοριστεί ο βαθμός μεταβολής - διακύμανσης αυτού του έτους; Οι λύσεις αυτών των δύο προβλημάτων περιγράφονται από τον ακόλουθο εμπειρικό κανόνα:

Κανόνας 3: Έστω ότι η μεταβολή - διακύμανση του περασμένου χρόνου είναι ένα κριτήριο. Υπολογίζουμε τις ασαφείς σχέσεις μεταξύ του περασμένου χρόνου και των άλλων προηγούμενων ετών βασισμένοι στις μεταβολές - διακυμάνσεις των δεδομένων. Από τις παραγόμενες ασαφείς σχέσεις, μπορούμε να ξέρουμε τους βαθμούς των σχέσεων μεταξύ της μεταβολής - διακύμανσης του περασμένου χρόνου και των μεταβολών - διακυμάνσεων άλλων προηγούμενων ετών. Η μεταβολή - διακύμανση αυτού του έτους μπορεί να ληφθεί από τις παραγόμενες ασαφείς σχέσεις.

Με βάση αυτούς τους εμπειρικούς κανόνες, αρχικά μπορούμε να «ασαφοποιήσουμε» τα ιστορικά στοιχεία των εγγραφών. Σε αυτό το κεφάλαιο, χρησιμοποιούμε την «ασαφοποιημένη» μεταβολή - διακύμανση των ιστορικών εγγραφών και των γλωσσικών τιμών-μεταβλητών (μεγάλη μείωση), (μείωση), (καμία αλλαγή), (αύξηση), (μεγάλη αύξηση), (πάρα πολύ μεγάλη αύξηση) για την πρόβλεψη των πανεπιστημιακών εγγραφών. Η μεταβολή - διακύμανση των ιστορικών εγγραφών μεταξύ του έτους  $t$  και του έτους  $t - 1$  μπορεί να περιγραφεί ως εξής:

$$F(t) = u_1 / (\text{μεγάλη μείωση}) + u_2 / (\text{μείωση}) + \dots + u_i / (L) + \dots + u_m / (\text{πολύ μεγάλη μείωση}), \quad (3.12)$$

1)

όπου η  $F(t)$  δείχνει την ασαφοποιημένη μεταβολή - διακύμανση των εγγραφών μεταξύ του έτους  $t$  και του έτους  $t - 1$ ,  $u_i$  είναι ο βαθμός ιδιότητας μέλους στη γλωσσική τιμή-μεταβλητή  $L$ , το  $m$  είναι ο αριθμός των στοιχείων στο σύνολο του λόγου, και  $1 \leq i \leq m$ .

Για να προβλέψουμε τις εγγραφές του έτους  $t$ , πρέπει να αποφασίσουμε πόσα έτη από τα στοιχεία-δεδομένα των εγγραφών θα χρησιμοποιηθούν, όπου ο αριθμός ετών των στοιχείων των εγγραφών που χρησιμοποιήσαμε καλείται *βάση παραθύρου (window basis)*. Υποθέτοντας ότι θέτουμε μια βάση παραθύρου στα  $w$  έτη, τότε η μεταβολή - διακύμανση του περασμένου χρόνου χρησιμοποιείται ως κριτήριο και οι άλλες μεταβολές - διακυμάνσεις των  $w$  προηγούμενων ετών χρησιμοποιούνται για να διαμορφώσουν μια μήτρα που καλείται μήτρα λειτουργίας. Η μήτρα κριτηρίου  $C(t)$  και η μήτρα λειτουργίας  $O^w(t)$  στο έτος  $t$  εκφράζονται ως εξής:

$$C(t) = F(t-1) = \begin{bmatrix} (\text{μεγάλη μείωση}) & (\text{μείωση}) & \dots & (\text{πολύμεγάλη μείωση}) \\ C_1 & C_2 & \dots & C_m \end{bmatrix}. \quad (3.122)$$

$$O^w(t) = \begin{bmatrix} F(t-2) \\ F(t-3) \\ \vdots \\ F(t-w-1) \end{bmatrix} = \begin{bmatrix} (\text{μεγάλη μείωση}) & (\text{μείωση}) & \dots & (\text{πολύ μεγάλη μείωση}) \\ O_{11} & O_{12} & \dots & O_{1m} \\ O_{21} & O_{22} & \dots & O_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ O_{w1} & O_{w2} & \dots & O_{wm} \end{bmatrix}. \quad (3.123)$$

Μπορούμε να υπολογίσουμε τη σχέση μεταξύ της μήτρας λειτουργίας  $O^w(t)$  και της μήτρας κριτηρίου  $C(t)$ , και μπορούμε να πάρουμε μια μήτρα σχέσης  $R(t)[w, m]$  με την εκτέλεση του

$$R(t) = O^w(t) \otimes C(t), \text{ όπου}$$

$$R(t) = \begin{bmatrix} O_{11} \times C_1 & O_{12} \times C_2 & \dots & O_{1m} \times C_m \\ O_{21} \times C_1 & O_{22} \times C_2 & \dots & O_{2m} \times C_m \\ \vdots & \vdots & \vdots & \vdots \\ O_{w1} \times C_1 & O_{w2} \times C_2 & \dots & O_{wm} \times C_m \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & \dots & R_{1m} \\ R_{21} & R_{22} & \dots & R_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ R_{w1} & R_{w2} & \dots & R_{wm} \end{bmatrix}. \quad (3.124)$$

όπου  $R_{ij} = O_{ij} \times c_{ij}$ ,  $1 \leq i \leq w$ ,  $1 \leq j \leq m$ , και "×" είναι η λειτουργία πολλαπλασιασμού. Από τη μήτρα σχέσης  $R(t)$ , μπορούμε να ξέρουμε το βαθμό σχέσεων μεταξύ του περασμένου χρόνου και των άλλων προηγούμενων ετών στις μεταβολές - διακυμάνσεις δεδομένων.

Κατόπιν, μπορούμε να πάρουμε την μεταβολή - διακύμανση της πρόβλεψης των εγγραφών του έτους  $t$ , όπου

$$F(t) = [ \text{Max}(R_{11}, R_{21}, \dots, R_{w1}) \text{Max}(R_{12}, R_{22}, \dots, R_{w2}) \dots \text{Max}(R_{1m}, R_{2m}, \dots, R_{wm}, ) ]. \quad (3.125)$$

Η προτεινόμενη μέθοδος παρουσιάζεται τώρα ως εξής:

*Βήμα 1:* Από τα ιστορικά στοιχεία εγγραφής, υπολογίζουμε τις μεταβολές - διακυμάνσεις των εγγραφών μεταξύ οποιωνδήποτε δύο συνεχών ετών. Η μεταβολή - διακύμανση αυτού του έτους είναι η εγγραφή αυτού του έτους μείον την εγγραφή του περασμένου χρόνου. Παραδείγματος χάριν, εάν η εγγραφή του 1972 είναι 13.563, και η εγγραφή του 1971 είναι 13.055, κτότε η μεταβολή - διακύμανση του έτους 1972 = 13.563 - 13.055 = 508. Βασισμένοι στα ιστορικά στοιχεία εγγραφής που παρουσιάζονται, μπορούμε να λάβουμε τις μεταβολές - διακυμάνσεις των εγγραφών μεταξύ οποιωνδήποτε δύο συνεχών ετών όπως φαίνεται στον πίνακα 1. Μπορούμε να βρούμε την ελάχιστη αύξηση  $D_{\min}$  και τη μέγιστη αύξηση  $D_{\max}$ . Κατόπιν καθορίζουμε το σύνολο του λόγου  $U$ ,  $U = [D_{\min} - D_1, D_{\max} + D_2]$ , όπου  $D_1$  και  $D_2$  είναι κατάλληλοι θετικοί αριθμοί.

Σε αυτό το κεφάλαιο, θέτουμε  $D_{\min} = -955$ ,  $D_{\max} = 1291$ ,  $D_1 = 45$ ,  $D_2 = 109$ , έτσι το  $U$  μπορεί να αντιπροσωπευθεί ως  $U = [- 1000 \dots 1400 ]$

**ΠΙΝΑΚΑΣ 3.20**

Πραγματικές εγγραφές και διακυμάνσεις των ιστορικών στοιχείων

Actual enrollments and variations of historical data		
Years	Actual enrollments	Variations
1971	13 055	
1972	13 563	+ 508
1973	13 867	+ 304
1974	14 696	+ 829
1975	15 460	+ 764
1976	15 311	- 149
1977	15 603	+ 292
1978	15 861	+ 258
1979	16 807	+ 946
1980	16 919	+ 112
1981	16 388	- 531
1982	15 433	- 955
1983	15 497	+ 64
1984	15 145	- 352
1985	15 163	+ 18
1986	15 984	+ 821
1987	16 859	+ 875
1988	18 150	+ 1291
1989	18 970	+ 820
1990	19 328	+ 358
1991	19 337	+ 9
1992	18 876	- 461

*Βήμα 2:* Χωρίζουμε το σύνολο του λόγου  $U$  σε διάφορα ομαλά διαστήματα μήκους  $u_1, u_2, \dots, u_m$ . Εδώ χωρίζουμε το  $U$  σε έξι διαστήματα,  $u_1 = [-1000, -600]$ ,  $u_2 = [-600, -200]$ ,  $u_3 = [-200, 200]$ ,  $u_4 = [200, 600]$ ,  $u_5 = [600, 1000]$ , και  $u_6 = [1000, 1400]$ .

*Βήμα 3:* Καθορίζουμε ασαφή σύνολα στο σύνολο του λόγου  $U$ . Αρχικά, καθορίζουμε μερικές γλωσσικές τιμές, που αντιπροσωπεύονται από ασαφή σύνολα, για να περιγράψουμε το βαθμό μεταβολής - διακύμανσης μεταξύ δύο συνεχών ετών. Εξετάζουμε έξι συγκεκριμένα σύνολα που είναι τα  $A_1$  = (μεγάλη μείωση),  $A_2$  = (μείωση),  $A_3$  = (καμία αλλαγή),  $A_4$  = (αύξηση),  $A_5$  = (μεγάλη αύξηση),  $A_6$  = (πολύ μεγάλη αύξηση). Κατόπιν, καθορίζουμε τα συγκεκριμένα σύνολα  $A_1, A_2, \dots, A_6$  στο σύνολο του λόγου  $U$  όπως ακολουθεί:

$$\begin{aligned}
 A_1 &= 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + 0/u_5 + 0/u_6, \\
 A_2 &= 0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + 0/u_5 + 0/u_6, \\
 A_3 &= 0/u_1 + 0.5/u_2 + 1/u_3 + 0.5/u_4 + 0/u_5 + 0/u_6, \\
 A_4 &= 0/u_1 + 0/u_2 + 0.5/u_3 + 1/u_4 + 0.5/u_5 + 0/u_6, \\
 A_5 &= 0/u_1 + 0/u_2 + 0/u_3 + 0.5/u_4 + 1/u_5 + 0.5/u_6, \\
 A_6 &= 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + 0.5/u_5 + 1/u_6,
 \end{aligned}
 \tag{3.126}$$

Βήμα 4: «Ασαφοποιούμε» τις τιμές των ιστορικών δεδομένων. Εάν ο αριθμός μεταβολής - διακύμανσης της εγγραφής του έτους  $i$  είναι  $p$ , όπου  $p \in u_j$ , και εάν υπάρχει μια τιμή που αντιπροσωπεύεται από ασαφές σύνολο  $A_k$  στο οποίο η μέγιστη τιμή ιδιότητας μέλους εμφανίζεται στο  $u_j$ , τότε το  $p$  μεταφράζεται σε  $A_k$ . Οι μεταβολές - διακυμάνσεις των στοιχείων εγγραφής παρουσιάζονται στον πίνακα 3.21.

**ΠΙΝΑΚΑΣ 3.21** Ασαφοποιημένες ιστορικές εγγραφές

Fuzzified historical enrollments		
Years	Variations	Fuzzified variations
1971		
1972	+ 508	$A_4$
1973	+ 304	$A_4$
1974	+ 829	$A_5$
1975	+ 764	$A_5$
1976	- 149	$A_3$
1977	+ 292	$A_4$
1978	+ 258	$A_4$
1979	+ 946	$A_5$
1980	+ 112	$A_3$
1981	- 531	$A_2$
1982	- 955	$A_1$
1983	+ 64	$A_3$
1984	- 352	$A_2$
1985	+ 18	$A_3$
1986	+ 821	$A_5$
1987	+ 875	$A_5$
1988	+ 1291	$A_6$
1989	+ 820	$A_5$
1990	+ 358	$A_4$
1991	+ 9	$A_3$
1992	- 461	$A_2$

Βήμα 5: Επιλέγουμε μια κατάλληλη βάση  $w$  παραθύρου, και υπολογίζουμε την έξοδο από τη μήτρα λειτουργίας  $O^w(t)$  και τη μήτρα κριτηρίου  $C(t)$ , όπου το  $t$  είναι το έτος για το οποίο θέλουμε να προβλέψουμε τις εγγραφές. Παραδείγματος χάριν, εάν θέσουμε  $w = 5$ , τότε μπορούμε να θέσουμε την  $4 \times 6$  μήτρα λειτουργίας  $O^5(t)$  και μια  $1 \times 6$  μήτρα κριτηρίου  $C(t)$ . Επειδή  $w = 5$ , πρέπει να χρησιμοποιήσουμε δεδομένα εγγραφών των έξι προηγούμενων ετών, και έτσι αρχίζουμε να προβλέπουμε για το 1977. Σε αυτήν την περίπτωση, η μήτρα λειτουργίας  $O^5(t)$  και η μήτρα κριτηρίου  $C(t)$  είναι οι

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

ακόλουθες:

$$O^5(1977) = \begin{matrix} \begin{matrix} \text{ασαφής διακύμανση των εγγραφών του 1975} \\ \text{ασαφής διακύμανση των εγγραφών του 1974} \\ \text{ασαφής διακύμανση των εγγραφών του 1973} \\ \text{ασαφής διακύμανση των εγγραφών του 1972} \end{matrix} & = & \begin{matrix} A_5 \\ A_5 \\ A_4 \\ A_4 \end{matrix} \\ \begin{matrix} \text{(μεγάλη μείωση)} & \text{(μείωση)} & \text{(καμιά αλλαγή)} & \text{(αύξηση)} & \text{(μεγάλη αύξηση)} & \text{(πολύ μεγάλη αύξηση)} \end{matrix} & & \begin{matrix} 0 & 0 & 0 & 0,5 & 1 & 0,5 \\ 0 & 0 & 0 & 0,5 & 1 & 0,5 \\ 0 & 0 & 0,5 & 1 & 0,5 & 0 \\ 0 & 0 & 0,5 & 1 & 0,5 & 0 \end{matrix} \end{matrix}$$

$$C(1977) = \text{ασαφής διακύμανση των εγγραφών του 1976} = [A_3]$$

$$\begin{matrix} \text{(μεγάλη μείωση)} & \text{(μείωση)} & \text{(καμιά αλλαγή)} & \text{(αύξηση)} & \text{(μεγάλη αύξηση)} & \text{(πολύ μεγάλη αύξηση)} \\ = [ & 0 & 0,5 & 1 & 0,5 & 0 & 0 & ] \end{matrix}$$

Υπολογίζουμε τη μήτρα σχέσης  $R(t)$  από τη σχέση

$$R(t)[i, j] = O^w(t)[i, j] \times C(t)[j], \text{ όπου } 1 \leq i \leq 4, 1 \leq j \leq 6.$$

Κατόπιν, βασισμένοι στον τύπο (10), μπορούμε να πάρουμε

$$R(1997) = \begin{matrix} \text{(μεγάλη μείωση)} & \text{(μείωση)} & \text{(καμιά αλλαγή)} & \text{(αύξηση)} & \text{(μεγάλη αύξηση)} & \text{(πολύ μεγάλη αύξηση)} \\ \begin{matrix} 0 & 0 & 0 & 0,25 & 0 & 0 \\ 0 & 0 & 0 & 0,25 & 0 & 0 \\ 0 & 0 & 0,5 & 0,5 & 0 & 0 \\ 0 & 0 & 0,5 & 0,5 & 0 & 0 \end{matrix} \end{matrix}$$

Με βάση προηγούμενο τύπο, μπορούμε να πάρουμε την «ασαφοποιημένη» μεταβολή - διακύμανση πρόβλεψης  $F(1977)$  του έτους 1977 ως εξής:

$$F(1977) = \begin{matrix} \text{(μεγάλη μείωση)} & \text{(μείωση)} & \text{(καμιά αλλαγή)} & \text{(αύξηση)} & \text{(μεγάλη αύξηση)} & \text{(πολύ μεγάλη αύξηση)} \\ [ & 0 & 0 & 0,5 & 0,5 & 0 & 0 & ] \end{matrix}$$

Οι «ασαφοποιημένες» προβλεφθείσες μεταβολές - διακυμάνσεις για τα υπόλοιπα εναπομείναντα έτη μπορούν να υπολογιστούν με τον ίδιο τρόπο και όλα τα αποτελέσματα παρατίθενται στον πίνακα 3.22.



**ΠΙΝΑΚΑΣ 3.22** Προβλεπόμενες διακυμάνσεις με τη βάση παραθύρου  $w = 5$

Forecasted variations with the window basis  $w = 5$

Years	Membership functions of forecasted variations					
	$u_1$	$u_2$	$u_3$	$u_4$	$u_5$	$u_6$
1977	0	0	0.5	0.5	0	0
1978	0	0	0.5	1	0.5	0
1979	0	0	0.5	1	0.5	0
1980	0	0	0	0.5	1	0.25
1981	0	0.25	1	0.5	0	0
1982	0	0.5	0.5	0	0	0
1983	0.5	0.5	0	0	0	0
1984	0	0.5	1	0.25	0	0
1985	0.5	1	0.5	0	0	0
1986	0	0.5	1	0.25	0	0
1987	0	0	0	0.25	0	0
1988	0	0	0	0.25	1	0.25
1989	0	0	0	0	0.5	0.5
1990	0	0	0	0.25	1	0.5
1991	0	0	0	0.5	0.5	0
1992	0	0	0.5	0.5	0	0

*Βήμα 6:* Αποσαφηνίζουμε τις ασαφείς προβλεφθείσες μεταβολές - διακυμάνσεις που παρήχθησαν στο βήμα 5. Σε αυτό το κεφάλαιο, χρησιμοποιούμε τις ακόλουθες αρχές για την αποσαφήνιση των προβλεφθεισών ασαφοποιημένων μεταβολών - διακυμάνσεων:

(1) Εάν οι βαθμοί ιδιότητας μέλους της προβλεφθείσας ασαφούς μεταβολής - διακύμανσης έχουν μόνο ένα μέγιστο  $u_i$ , και η μέση τιμή τάξεως του  $u_i$  είναι το  $m_i$ , τότε η προβλεφθείσα μεταβολή - διακύμανση είναι  $m_i$ . Εάν οι βαθμοί ιδιότητας μέλους της προβλεφθείσας ασαφοποιημένης μεταβολής - διακύμανσης έχουν περισσότερα από ένα μέγιστο  $u_1, u_2, \dots, u_k$ , και οι μέσες τιμές τάξεώς τους είναι  $m_1, m_2, \dots, m_k$ , αντίστοιχα, τότε η προβλεφθείσα μεταβολή - διακύμανση είναι  $(m_1 + m_2 + \dots + m_k)/k$ . Παραδείγματος χάριν, από τον πίνακα 3, μπορούμε να δούμε ότι η μέγιστη τιμή ιδιότητας μέλους του  $F(1977)$  είναι 0,5 η οποία εμφανίζεται στα  $u_3$  και  $u_4$ , όπου η μέση τιμή τάξεως του  $u_3$  είναι 0 και του  $u_4$  είναι 400. Η προβλεφθείσα μεταβολή - διακύμανση για το 1977 είναι  $(0 + 400)/2 = 200$ ,

(2) Εάν οι βαθμοί ιδιότητας μέλους της προβλεφθείσας ασαφοποιημένης μεταβολής - διακύμανσης είναι όλοι 0, τότε θέτουμε την προβλεφθείσα μεταβολή - διακύμανση 0.

*Βήμα 7:* Υπολογίζουμε τις προβλεπόμενες εγγραφές. Η προβλεπόμενη εγγραφή είναι προβλεφθείσα μεταβολή - διακύμανση συν τον αριθμό πραγματικής εγγραφής του περασμένου χρόνου. Παραδείγματος χάριν, εάν η προβλεφθείσα μεταβολή - διακύμανση το 1977 είναι 200, και οι πραγματικές έγγραφες το 1976 είναι 15.311, τότε οι προβλεπόμενες εγγραφές για το 1977 είναι

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

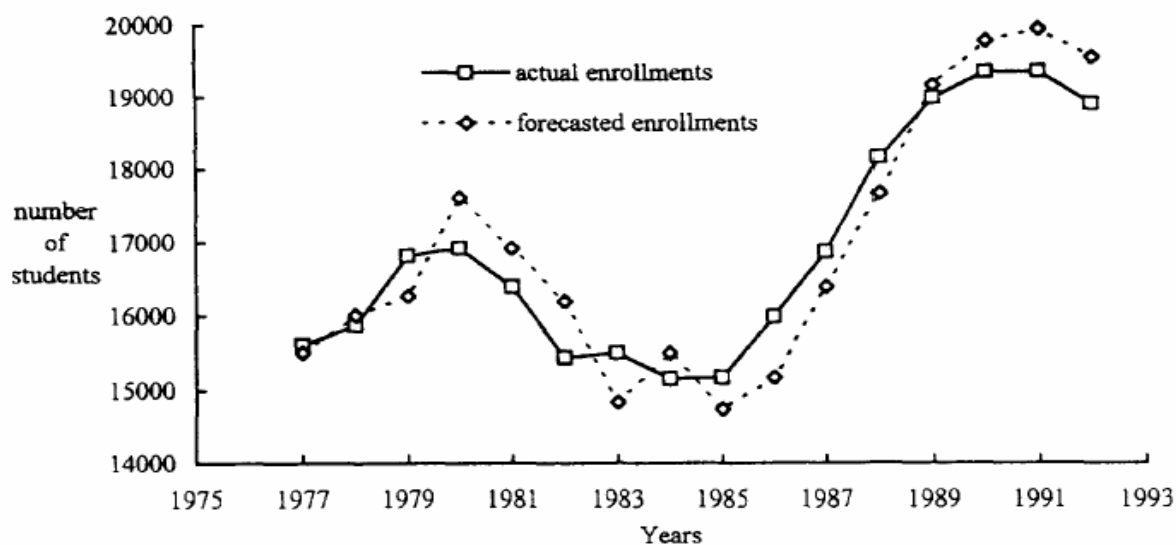
$15.311 + 200 = 15.511$ . Τα αποτελέσματα των προβλεπόμενων εγγραφών του πανεπιστημίου της Αλαμπάμα φαίνονται στον πίνακα 3.23. Το συνεπακόλουθο σφάλμα κάθε έτους με την προτεινόμενη μέθοδο κάτω από τη βάση παραθύρου  $w = 5$  παρουσιάζεται επίσης στον πίνακα 3.23. Οι καμπύλες των πραγματικών εγγραφών και των προβλεπόμενων παρουσιάζονται στο Σχ.3.40, όπου η βάση παραθύρου είναι 5.

**ΠΙΝΑΚΑΣ 3.23** Προβλεπόμενα αποτελέσματα με χρήση της προτεινόμενης ασαφούς μεθόδου χρονοσειρών με βάση παραθύρου  $w = 5$

Forecasted results using the proposed fuzzy time-series method with the window basis  $w = 5$

Years	Actual enrollments	Forecasted enrollments	Errors
1977	15 603	15 511	0.59%
1978	15 861	16 003	0.90%
1979	16 807	16 261	3.25%
1980	16 919	17 607	4.07%
1981	16 388	16 919	3.24%
1982	15 433	16 188	4.89%
1983	15 497	14 833	4.28%
1984	15 145	15 497	2.32%
1985	15 163	14 745	2.76%
1986	15 984	15 163	5.14%
1987	16 859	16 384	2.82%
1988	18 150	17 659	2.71%
1989	18 970	19 150	0.95%
1990	19 328	19 770	2.29%
1991	19 337	19 928	3.06%
1992	18 876	19 537	3.50%

Σχήμα 3.40 Καμπύλες των πραγματικών και των προβλεπόμενων εγγραφών



3.4.3.4 Ανάλυση απόδοσης

Σε προηγούμενη υποπαράγραφο, έχουμε χρησιμοποιήσει την ακόλουθη μέθοδο για να προβλέψουμε τις εγγραφές του πανεπιστημίου της Αλαμπάμα:

$$R(t) = O^w(t) \otimes C(t) = \begin{bmatrix} R_{11} & R_{12} & \dots & R_{1m} \\ R_{21} & R_{22} & \dots & R_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ R_{w1} & R_{w2} & \dots & R_{wm} \end{bmatrix}$$

$$F(t) = [ \text{Max}(R_{11}, R_{21}, \dots, R_{w1}) \text{Max}(R_{12}, R_{22}, \dots, R_{w2}) \dots \text{Max}(R_{1m}, R_{2m}, \dots, R_{wm}, ) ] \\ = [Q_1 \ Q_2 \ \dots \ Q_m]$$

όπου το  $F(t)$  είναι η προβλεπόμενη μεταβολή - διακύμανση του έτους  $t$  που αντιπροσωπεύεται από ένα ασαφές σύνολο,  $O^w(t)$  είναι η μήτρα λειτουργίας,  $R(t)$  είναι μια ασαφής σχέση που δείχνει την ασαφή συγγένεια μεταξύ των εγγραφών του περασμένου χρόνου και των υπόλοιπων προηγούμενων ετών στις μεταβολές - διακυμάνσεις των δεδομένων και "x" είναι η πράξη του πολλαπλασιασμού. Η χρονική πολυπλοκότητα για τον υπολογισμό της ασαφούς σχέσης  $R(t)$  είναι  $O(n)$ , και η χρονική πολυπλοκότητα για τον υπολογισμό της προβλεπόμενης εγγραφής  $F(t)$  είναι  $O(wn)$ , όπου το  $w$  είναι η βάση παραθύρου και το  $n$  ο αριθμός στοιχείων στο σύνολο του λόγου. Στα [26,27], ο Song χρησιμοποίησε τον τύπο  $A_i = A_{i-1} \circ R$  για να προβλέψει τις εγγραφές του πανεπιστημίου της Αλαμπάμα, όπου  $A_{i-1}$  είναι η εγγραφή του έτους  $i-1$  αντιπροσωπευόμενη από ένα ασαφές σύνολο,  $A_i$  είναι η προβλεπόμενη εγγραφή του έτους  $i$  που αντιπροσωπεύεται από ένα ασαφές σύνολο, το  $R$  είναι η ένωση των ασαφών σχέσεων, και "o" είναι ο ανώτατος-ελάχιστος τελεστής σύνθεσης. Πρέπει να πάρει τη χρονική πολυπλοκότητα του  $O(kn^2)$  για να υπολογίσει τις ανώτατος-ελάχιστες διαδικασίες σύνθεσης, όπου το  $k$  είναι ο αριθμός ασαφών λογικών σχέσεων και  $n$  είναι ο αριθμός των στοιχείων στο σύνολο του λόγου. Επειδή  $O(kn^2) \gg O(wn)$ , η προτεινόμενη μέθοδος είναι αποδοτικότερη από αυτές που παρουσιάζονται στα [26,27].

Ο πίνακας 3.24 δείχνει αποτελέσματα πρόβλεψης διαφορετικών βάσεων παραθύρου  $w$  που κυμαίνονται από 2 έως 9. Από τον πίνακα 3.25 μπορούμε να δούμε ότι τα μέσα λάθη πρόβλεψης για τις διαφορετικές βάσεις παραθύρου κυμαίνονται από 3,12% ως 2,79, τα οποία είναι καλύτερα από το 3,18% του [26] και καλύτερα από αυτά που παρουσιάζονται στο [27] και κυμαίνονται από 4,49% ως 3,15%.

**ΠΙΝΑΚΑΣ 3.24** Προβλεπόμενες εγγραφές με διαφορετικές βάσεις παραθύρου  
Forecasting enrollments with different window bases

Years	Actual enrollments	Forecasted enrollments							
		Window bases							
		w = 2	w = 3	w = 4	w = 5	w = 6	w = 7	w = 8	w = 9
1974	14696	14267							
1975	15460	15296	15296						
1976	15311	16260	16260	16260					
1977	15603	15711	15711	15511	15511				
1978	15861	15803	16003	16003	16003	16003			
1979	16807	16261	16261	16261	16261	16261	16261		
1980	16919	17407	17407	17407	17607	17607	17607	17607	
1981	16388	17319	17119	17119	16919	16919	16919	16919	16919
1982	15433	16188	16188	16188	16188	16188	16188	16188	16188
1983	15497	14833	14833	14833	14833	14833	14833	14833	14833
1984	15145	15097	15297	15497	15497	15497	15497	15497	15497
1985	15163	14945	14745	14745	14745	14745	14745	14745	14745
1986	15984	14963	15163	15163	15163	15163	15163	15163	15163
1987	16859	16384	16384	16384	16384	16384	16384	16784	16784
1988	18150	17659	17659	17659	17659	17659	17659	17659	17659
1989	18970	19150	19150	19150	19150	19150	19150	19150	19150
1990	19328	19970	19770	19770	19770	19770	19770	19770	19770
1991	19337	19928	19928	19928	19928	19928	19928	19928	19928
1992	18876	19537	19537	19537	19537	19537	19337	19337	19337

**ΠΙΝΑΚΑΣ 3.25** Λάθη πρόβλεψης με διαφορετικές βάσεις παραθύρου

Forecasting errors with different window bases

	Window bases							
	w = 2	w = 3	w = 4	w = 5	w = 6	w = 7	w = 8	w = 9
Average forecasting errors	2.99%	2.94%	3.12%	2.92%	3.01%	3.08%	2.89%	2.79%

Φαίνεται ότι η προτεινόμενη μέθοδος παίρνει καλύτερα αποτελέσματα πρόβλεψης από αυτά που παρουσιάζονται στα [26,27]. Στον πίνακα 3.26, συγκρίνουμε τις προβλεπόμενες εγγραφές του πανεπιστημίου της Αλαμπάμα κάτω από τις διαφορετικές μεθόδους πρόβλεψης. Στον πίνακα 3.27, συγκρίνουμε τα μέσα λάθη πρόβλεψης των διαφορετικών μεθόδων πρόβλεψης με την προτεινόμενη μέθοδο κάτω από τη βάση παραθύρου  $w = 4$ . Ο καθορισμός της βάσης παραθύρου  $w$  στην προτεινόμενη μέθοδο είναι σαν τον καθορισμό της πρότυπης βάσης στο [27], και όλοι χρησιμοποιούν τα προηγούμενα  $w$  έτη των δεδομένων για να προβλέψουν την εγγραφή αυτού του έτους, με τη διαφορά ότι η προτεινόμενη μέθοδος χρησιμοποιεί πραγματική μεταβολή - διακύμανση για να προβλέψει ενώ η μέθοδος στο [27] χρησιμοποιεί πραγματικές εγγραφές. Στον πίνακα 3.28, συγκρίνουμε τα πραγματικά λάθη πρόβλεψης αυτών των δύο μεθόδων. Από τον πίνακα 3.27, μπορούμε να δούμε ότι η προτεινόμενη μέθοδος είναι πιο αποτελεσματική από αυτές που παρουσιάζονται στα [27,28,29] αναφορικά με το γεγονός ότι απλοποιεί την αριθμητική διαδικασία λειτουργίας. Επιπλέον, μπορούμε επίσης να δούμε ότι το μέσο λάθος πρόβλεψης της προτεινόμενης μεθόδου είναι μικρότερο από αυτά που παρουσιάζονται στα [26,27].

**ΠΙΝΑΚΕΣ 3.26 & 3.27**

Τα αποτελέσματα πρόβλεψης διαφορετικών μεθόδων πρόβλεψης (3.26)

Μια σύγκριση των μέσων λαθών πρόβλεψης των διαφορετικών μεθόδων πρόβλεψης (3.27)

The forecasting results of different forecasting methods

Years	Actual enrollments	Song–Chissom method [7]	Song–Chissom method [8] (under model basis $w = 4$ and using neural net method)	Chen’s method [2]	Markov method [10] (time-invariant)	The proposed method (under window basis $w = 4$ )
1972	13 563	14 000		14 000	13 563	
1973	13 867	14 000		14 000	13 867	
1974	14 696	14 000		14 000	14 696	
1975	15 460	15 500	14 700	15 500	15 460	
1976	15 311	16 000	14 800	16 000	15 311	16 260
1977	15 603	16 000	15 400	16 000	15 603	15 511
1978	15 861	16 000	15 500	16 000	15 861	16 003
1979	16 807	16 000	15 500	16 000	16 807	16 261
1980	16 919	16 813	16 800	16 833	16 919	17 407
1981	16 388	16 813	16 200	16 833	16 388	17 119
1982	15 433	16 789	16 400	16 833	15 433	16 188
1983	15 497	16 000	16 800	16 000	15 497	14 833
1984	15 145	16 000	16 400	16 000	15 145	15 497
1985	15 163	16 000	15 500	16 000	15 163	14 745
1986	15 984	16 000	15 500	16 000	15 984	15 163
1987	16 859	16 000	15 500	16 000	16 859	16 384
1988	18 150	16 813	16 800	16 833	18 150	17 659
1989	18 970	19 000	19 300	19 000	18 970	19 150
1990	19 328	19 000	17 800	19 000	19 328	19 770
1991	19 337	19 000	19 300	19 000	19 337	19 928
1992	18 876	not forecasted	19 600	19 000	not forecasted	19 537

Table 8

A comparison of the average forecasting errors of different forecasting methods (Note:  $k$  denotes the number of fuzzy logical relationships,  $n$  denotes the number of elements in the universe of discourse,  $p$  denotes the number of fuzzy logical relationship groups,  $c$  denotes the number of transitions in the historical data, and  $w$  denotes the window basis)

	Song–Chissom method [7]	Song–Chissom method [8] (under model basis $w = 4$ and using neural net method)	Chen’s method [2]	Markov method [10] (time-invariant)	The proposed method (under window basis ( $w = 4$ ))
Style	Time-invariant	Time-variant	Time-invariant	Time-invariant	Time-variant
Time complexity	$O(kn^2)$	$O(kn^2)$	$O(p)$	$O(cn^2)$	$O(wn)$
Average forecasting errors	3.2%	4.37%	3.22%	2.6%	3.12%

**ΠΙΝΑΚΑΣ 3.28**

Μια σύγκριση των μέσων λαθών πρόβλεψης της προτεινόμενης μεθόδου και της μεθόδου Song Chissom

A comparison of average forecasting errors of the proposed method and Song-Chissom method

	Window bases (or Model bases)							
	w = 2	w = 3	w = 4	w = 5	w = 6	w = 7	w = 8	w = 9
Forecasting errors of the Song-Chissom method [8] (using neural net method)	3.15%	3.89%	4.37%	4.41%	4.49%	4.35%	4.45%	4.23%
Forecasting errors of the proposed method	2.99%	2.94%	3.12%	2.92%	3.01%	3.08%	2.89%	2.79%

**3.4.3.5 Συμπεράσματα**

Σε αυτό το τμήμα, έχουμε προτείνει μια νέα μέθοδο για τα προβλήματα πρόβλεψης βασισμένα σε ασαφείς χρονοσειρές, όπου τα ιστορικά δεδομένα των εγγραφών του πανεπιστημίου της Αλαμπάμα που παρουσιάζεται στα [26,27] υιοθετούνται για να επεξηγήσουν τη διαδικασία πρόβλεψης. Επίσης έχουμε δείξει ότι η προτεινόμενη μέθοδος είναι αποδοτικότερη από αυτές που παρουσιάζονται στα [26,27,29]. Επιπλέον, το μέσο λάθος πρόβλεψης της προτεινόμενης μεθόδου είναι μικρότερο από αυτά των μεθόδων που παρουσιάζονται στα [26,27]. Μπορούμε να δούμε ότι όσο μεγαλύτερη είναι η βάση παραθύρου  $w$ , τόσο περισσότερος χρόνος απαιτείται για να εκτελεστούν οι υπολογισμοί. Από τον πίνακα 3.25, μπορούμε να δούμε ότι το μεγαλύτερο μέσο λάθος πρόβλεψης εμφανίστηκε για  $w = 4$ , και το μικρότερο λάθος πρόβλεψης εμφανίστηκε για  $w = 9$ . Είναι δύσκολο να βρεθούν οι σχέσεις μεταξύ της βάσης παραθύρου και του μέσου λάθους πρόβλεψης, αλλά υπάρχει ένας αποδοτικός τρόπος που χρησιμοποιεί τους γενετικούς αλγορίθμους ώστε να βρει την καλύτερη βάση παραθύρου για να προβλέψει τις εγγραφές. Από τον πίνακα 3.27, μπορούμε να δούμε ότι η προτεινόμενη μέθοδος έχει μια καλύτερη μέση ακρίβεια πρόβλεψης από τις άλλες μεθόδους που παρουσιάζονται στα [26,27]. Επιπλέον, από τον πίνακα 3.27, μπορούμε επίσης να δούμε ότι η χρονική πολυπλοκότητα της προτεινόμενης μεθόδου είναι καλύτερη από αυτές που παρουσιάζονται στα [26,27,29].

**3.5 ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ [31], [32], [33]****3.5.1 Εισαγωγή**

Οι χρονοσειρές είναι μια ειδική μορφή στοιχείων όπου οι προηγούμενες τιμές στη σειρά μπορούν να επηρεάσουν μελλοντικές τιμές, εξαρτώμενες από τη παρουσία υποκείμενων αιτιοκρατικών δυνάμεων. Αυτές οι δυνάμεις μπορούν να χαρακτηριστούν από τις τάσεις, τους κύκλους και τη μη στατική συμπεριφορά στη χρονοσειρές και προφητικά μοντέλα προσπαθούν να αναγνωρίσουν τα επαναλαμβανόμενα σχήματα και τις μη γραμμικές σχέσεις. Ενώ τα γραμμικά μοντέλα, όπως εκείνα που είναι βασισμένα στις τεχνικές παλινδρόμησης, ήταν η βάση των παραδοσιακών στατιστικών μεθόδων πρόβλεψης, τα μειονεκτήματά τους έχουν οδηγήσει στην αυξανόμενη δραστηριότητα στη μη γραμμική διαμόρφωση. Τα νευρωνικά δίκτυα είναι μη γραμμικά μοντέλα που μπορούν να εκπαιδευθούν για να χαρτογραφήσουν τις προηγούμενες και μελλοντικές τιμές μιας χρονοσειράς, και με αυτόν τον τρόπο εξάγουν την κρυμμένη δομή και τις σχέσεις που διέπουν τα δεδομένα.

Η πρόσφατη αυξητική τάση στην ερευνητική δραστηριότητα των νευρωνικών δικτύων είναι μια συνέπεια της διαπίστωσης ότι τα νευρωνικά δίκτυα έχουν μεγάλες δυνατότητες αναγνώρισης σχημάτων, που ξεπερνούν εκείνες άλλων τεχνικών σε πολλές εφαρμογές - και για την ακρίβεια ή/και για την ταχύτητα υπολογισμού.

Αν και οι περιορισμένες μη γραμμικές τεχνικές παλινδρόμησης αναπτύσσονται, οι απεριόριστες μη γραμμικές ιδιότητες του νευρωνικού δικτύου το κάνουν προτιμότερο υποψήφιο στο τομέα της πρόβλεψης χρονοσειρών. Για να τοποθετήσουμε τα νευρωνικά δίκτυα στα πλαίσια των παραδοσιακών στατιστικών μεθόδων, πρέπει να θεωρηθούν ως "τεχνική εξαγωγής συμπεράσματος, μη γραμμική, μη παραμετρική και πολλών μεταβλητών". Το "πολλών μεταβλητών" αναφέρεται στο ότι τα δεδομένα εισαγωγής του νευρωνικού δικτύου έχουν πολλές διαφορετικές μεταβλητές των οποίων οι αλληλεξαρτήσεις και αλληλεπιδράσεις χρησιμοποιούνται για τη πρόβλεψη της μελλοντικής συμπεριφορά μιας χρονικής ακολουθίας. Το "μη παραμετρικό", είναι συνέπεια της έλλειψης οποιωνδήποτε υποθέσεων σχετικά με τη σχέση μεταξύ των μεταβλητών εισόδου και των παρεκτάσεων στο μέλλον. Μάλλον το δίκτυο «εκπαιδεύεται» από την προσαρμογή των ελεύθερων παραμέτρων για να ανακαλύψει οποιεσδήποτε πιθανές σχέσεις, απαλλαγμένο από του περιορισμούς των μοντέλων, καθώς διαμορφώνεται και Σχηματίζεται απλώς από τα δεδομένα εισόδου.

Οι ελεύθεροι παράμετροι είναι βάρη που συνδέονται με τις γραμμές επικοινωνίας σημάτων μεταξύ των νευρώνων, και οι οποίοι μειώνουν τα σήματα που διέρχονται, ή τα δεδομένα.

Μία σημαντική πτυχή των νευρωνικών δικτύων είναι ικανότητάς τους να κατασκευάζουν μη γραμμικές σχέσεις μεταξύ των δεδομένων εισόδου και στόχου στην έξοδο. Η γραμμικότητα,  $(p_t = \alpha_1 p_{t-1} + \alpha_2 \dots)$ , για μια χρονοσειρά τιμών  $p_t$  με σταθερές  $\alpha_i$  περιγράφει την απλούστερη σχέση που μπορεί να εμφανιστεί σε μια χρονοσειρά.

Κατά συνέπεια τα νευρωνικά δίκτυα καλύπτουν πολλά επιθυμητά χαρακτηριστικά γνώρισμα, ως εργαλείο ανάλυσης στοιχείων μέσα σε ένα σχετικά αποδοτικό σχέδιο εφαρμογής-από την άποψη των απαιτήσεων μνήμης και ταχύτητας υπολογισμών.

Το νευρωνικό δίκτυο μπορεί να θεωρηθεί ως τεχνική επεξεργασίας δεδομένων που καθοδηγεί, ή συσχετίζει, κάποιο τύπο ρεύματος εισαγωγής πληροφοριών με ένα ρεύμα εξαγωγής δεδομένων. Παραδείγματος χάριν, η είσοδος μπορεί να είναι υπό μορφή διδιάστατης εικόνας με ελλείποντα ή διαστρεβλωμένα τμήματα, και η επεξεργασμένη έξοδος μπορεί να παραγάγει μια διορθωμένη εικόνα. Στις εφαρμογές χρονοσειρών που ενδιαφέρουν εδώ, η εισαγωγή μπορούν να είναι χρόνος μιας διάστασης, και η έξοδος η παραγωγή της καλύτερης εκτίμησης του επόμενου στοιχείου στη σειρά. Γενικά οι στόχοι των νευρωνικών δικτύων μπορούν να διαιρεθούν σε τέσσερις τύπους ευδιάκριτων εφαρμογών:

1. **Κατάταξη:** Απόφαση σε ποια κατηγορία περιέρχεται ένα σχέδιο εισαγωγής.
2. **Συσχέτιση:** Ενεργεί ως προσπελάσιμη μνήμη που ανακαλεί ένα δεδομένο εξόδου δίδοντάς του κάποιο μέρος από αυτό, ως είσοδο.
3. **Κωδικοποίηση:** Η κωδικοποίηση συμπιέζει ένα δεδομένο εισόδου με την παραγωγή ενός δεδομένου εξόδου με μειωμένη διάσταση. Ο αντίθετος στόχος, που αποκωδικοποιεί, μπορεί επίσης να είναι ενδιαφέρον.
4. **Προσομοίωση:** Η δημιουργία ενός δεδομένου εξόδου για ένα δεδομένο εισόδου που ενεργεί ως ερέθισμα, όταν το δίκτυο έχει εκτεθεί σε ένα δείγμα των πιθανών ερεθισμάτων.

Οποιοσδήποτε ιδιαίτερος στόχος μπορεί να περιλάβει διάφορες ανωτέρω λειτουργίες.

Τα νευρωνικά δίκτυα μπορούν να εφαρμοστούν σε καθένα υλικό λογισμικού (software) ή υλικού (hardware).

Μερικοί στόχοι επεξεργασίας που εκτελούνται από τα νευρωνικά δίκτυα μπορούν επίσης να εκπληρωθούν από τις παραδοσιακές τεχνικές, όπως η ανάλυση βάσεων δεδομένων ή η ανάλυση παλινδρόμησης. Εντούτοις, κανένας δεν μοιράζεται τις μοναδικές δυνατότητες των νευρωνικών δικτύων, τα πλεονεκτήματα των οποίων μπορούν να συνοψιστούν ως εξής:

**Γενίκευση:** Τα νευρωνικά δίκτυα καθιερώνονται αρχικά μέσω μιας φάσης "κατάρτισης", στην οποία παρουσιάζονται εισαγωγές παραδειγμάτων και το δίκτυο εκπαιδεύεται για να εξαγάγει τις σχετικές πληροφορίες από αυτά τα πρότυπα. Στη συνέχεια το δίκτυο έχει την ικανότητα να γενικοποιήσει, έτσι ώστε μια έως τώρα απαρατήρητη εισαγωγή προτύπου να μπορεί επίσης υποβληθεί σε επεξεργασία.

**Ευελιξία:** Η σειρά των στόχων στην οποία το νευρωνικό δίκτυο μπορεί να εφαρμοστεί, υπερβαίνει κατά πολύ οποιαδήποτε παραδοσιακή τεχνική.

**Μη γραμμική "διαμόρφωση":** Η διαδικασία χαρτογράφησης δικτύων περιλαμβάνει τις μη γραμμικές λειτουργίες που μπορούν συνεπώς να καλύψουν μια μεγαλύτερη σειρά της πολυπλοκότητας του



προβλήματος. Αν και υπάρχουν και άλλες μη γραμμικές τεχνικές, το νευρωνικό δίκτυο είναι ανώτερο στη γενικότητα και την πρακτική ευκολία του στην εφαρμογή.

Υπάρχουν πολυάριθμα νευρωνικά σχέδια δικτυακής αρχιτεκτονικής, αποκαλούμενα επίσης νευρωνικά *παραδείγματα*, εντούτοις μπορούν όλα να διαιρεθούν σε τρεις ευρείες κατηγορίες, βάσει της τεχνικής που χρησιμοποιείται για να εκπαιδεύσει επαναληπτικά τις ελεύθερες παραμέτρους στο δίκτυο, γνωστές ως βάρη. Αυτές οι κατηγορίες εκμάθησης είναι οι ακόλουθες:

**Εποπτευόμενη εκμάθηση:** Οι στόχοι των δεδομένων εξόδου των δικτύων είναι γνωστοί κατά τη διάρκεια της κατάρτισης, έτσι ώστε το λάθος, η διαφορά μεταξύ του επιθυμητού στόχου και της πραγματικής εξόδου, μπορεί να ανατροφοδοτηθεί στο δίκτυο για τη βελτίωση της απόδοσης.

**Ενισχυόμενη Εκμάθηση:** Οι στόχοι για κάθε ένα από τα παραδείγματα που εισάγονται στο δίκτυο δεν είναι γνωστοί (ή δεν χρησιμοποιούνται). Μόνο ένα κριτήριο, σχετικά με το εάν η απόδοση δικτύων είναι σωστή ή όχι, χρησιμοποιείται στην ανατροφοδότηση. Καμιά πληροφορία σχετικά με την πραγματική-σωστή έξοδο δεν είναι διαθέσιμη για την κατάρτιση. Σε μερικές περιπτώσεις μόνο η κρίση σχετικά με την εκτίμηση είναι διαθέσιμη, αφού έχει παρουσιαστεί μια ακολουθία εισαγωγών, η οποία δημιουργεί έναν περαιτέρω χωρισμό μεταξύ της κάθε εισαγωγής και της επιθυμητής απόδοσης των δικτύων.

**Εκμάθηση χωρίς Επίβλεψη:** Υπάρχει συνολική έλλειψη ανατροφοδότησης σχετικά με το εάν ο σκοπός εκμάθησης επιτεύχθηκε κατά το σωστό τρόπο ή όχι. Μάλλον, αυτοί οι τύποι δικτύων είναι αυτο-οργανωμένοι, χρησιμοποιώντας σχέσεις μεταξύ των προτύπων εισαγωγής για κατηγοριοποιηθούν σε αυτό-ανακαλυπτόμενες κατηγορίες.

Το όνομα "νευρωνικό δίκτυο" προέρχεται από τη νευρωνική δομή στον εγκέφαλο, και ορίστηκε έτσι από τους βιολόγους και τους φυσιολόγους που προσπάθησαν να μιμηθούν και να τυποποιήσουν τους νευρώνες στον εγκέφαλο. Το θέμα είναι σήμερα διεπιστημονικό, κυμαινόμενο από τους βιολόγους και τους ψυχολόγους, στόχος των οποίων είναι να γίνει κατανοητή καλύτερα η πραγματική διαδικασία υπολογισμού στον εγκέφαλο, από τους φυσικούς που έχουν διατυπώσει επιτυχώς τα στατιστικά μηχανικά πρότυπα των δικτύων, από τους μηχανικούς που ενδιαφέρονται για τα παράλληλα σχέδια φίλτρων, από τους μαθηματικούς, τους επιστήμονες υπολογιστών και τους στατιστικούς για τους οποίους, η λογική της αναγνώρισης του συλλογισμού και της ομιλίας είναι ενεργοί τομείς της έρευνας. Κάθε ένας από αυτούς τους επιστημονικούς κλάδους αντιμετωπίζει το πρόβλημα της τεχνητής νοημοσύνης από τη μοναδική οπτική του γωνία και ταυτόχρονα εμπνέεται από τις νευροφυσιολογικές εξελίξεις.

### 3.5.2 Υποκειμενική – Κριτική Επέμβαση στις Προβλέψεις και τα Ιστορικά Δεδομένα Χρονοσειρών (Judgmental Adjustment) με Χρήση Νευρωνικών Δικτύων

#### Βασικές Έννοιες για τα Ιστορικά Δεδομένα και τη Κριτική Επέμβαση

Τα μοντέλα χρονοσειρών χρησιμοποιούν τα ιστορικά δεδομένα για την παραγωγή μελλοντικών προβλέψεων αναγνωρίζοντας παρελθοντικά πρότυπα χωρίς να μπορούν να ενσωματώσουν την επίδραση εξωτερικών μελλοντικών γεγονότων ακόμα και αν αυτά έχουν ανακοινωθεί πριν συμβούν. Είναι σύνηθες λοιπόν οι χρήστες των στατιστικών προβλέψεων να τις προσαρμόζουν σύμφωνα με την πείρα τους όταν ανακοινώνεται ένα εξωτερικό γεγονός ή όταν διαισθάνονται κάποιες επερχόμενες αλλαγές στην αγορά που τους ενδιαφέρει. Αυτή η διαδικασία οδηγεί σε βελτίωση των προβλέψεων και ονομάζεται “Υποκειμενική – Κριτική Επέμβαση στις Προβλέψεις Στατιστικών Μοντέλων” είναι όμως ιδιαίτερα χρονοβόρα και επίσης πολλές φορές προκατελιημμένη ιδιαίτερα όταν οι προβλέψεις αναφέρονται σε προϊόντα της εταιρείας τους. Η αυτοματοποίηση αυτής της διαδικασίας μέσω Νευρωνικών Δικτύων οδηγεί σε πολύ καλά αποτελέσματα.

*Υποκειμενικός – Κριτικός Παράγων (Judgmental Factor)* ορίζεται ένας παράγοντας του υπό εξέταση μεγέθους (για το οποίο ζητούνται μελλοντικές προβλέψεις) του οποίου η επίδραση στα ιστορικά δεδομένα δεν μπορεί να αναγνωριστεί σαφώς και να οριστεί αντικειμενικά.

Κοινά χαρακτηριστικά των υποκειμενικών παραγόντων είναι τα ακόλουθα:

- Συνήθως συμβαίνουν απροειδοποίητα και οπωσδήποτε σε μη περιοδικά χρονικά διαστήματα. Το δείγμα είναι πολύ μικρό για στατιστική μοντελοποίηση.
- Η επίδραση τους είναι πολύ σημαντική για να αγνοηθεί.
- Η επίδραση τους είναι παροδική
- Είναι δύσκολο να προσδιοριστεί το μέγεθος της επίδρασης τους στο υπό μελέτη μέγεθος ακόμα και να προσδιοριστεί το ακριβές χρονικό διάστημα που αναμένεται να συμβούν.

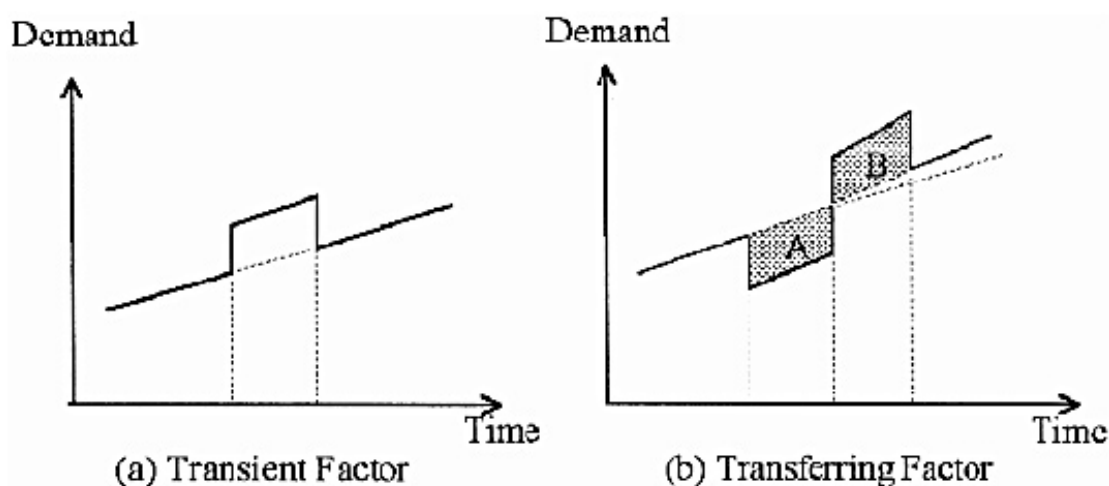
Στο παράδειγμα που θα εξετάσουμε, που αφορά τις πωλήσεις των κρατικών Διυλιστηρίων της Κορέας, μπορούν να αναγνωριστούν τέσσερις τυπικοί υποκειμενικοί παράγοντες που επηρεάζουν την ζήτηση Βενζίνης στην Κορεατική αγορά.

1. Κυβερνητικές διατάξεις εξοικονόμησης ενέργειας
2. Απεργίες βιομηχανιών που καταναλώνουν πολύ ενέργεια (πετρέλαιο)
3. Ειδικές αργίες – διακοπές
4. Ανακοινώσεις μελλοντικών αλλαγών στην τιμή των καυσίμων

Οι τέσσερις αυτοί παράγοντες αποτελούν την *Βάση Υποκειμενικών Παραγόντων* για το μέγεθος που εξετάζουμε.

Η επίδραση των τριών πρώτων παραγόντων είναι παροδική (transient) ενώ του τελευταίου μεταφερόμενη (transferring) πριν και μετά την αναγγελθείσα ημερομηνία. Η διαφορά των δύο επιδράσεων φαίνεται στο σχήμα που ακολουθεί:

Σχήμα 3.41 Παροδική (transient) και μεταφερόμενη (transferring) επίδραση



Η μοντελοποίηση και τα χαρακτηριστικά των τεσσάρων υποκειμενικών παραγόντων ορίζονται ακολούθως:

*Κυβερνητικές διατάξεις εξοικονόμησης ενέργειας*

Συμβολισμός:  $R_m(k, i, d)$ , (R: Regulation), όπου

k : Τύπος διάταξης

i : Αυστηρότητα της διάταξης (υποχρεωτική, συμβουλευτική κ.λ.π.)

d : Διάρκεια επίδρασης (σε ημέρες)

*Απεργίες βιομηχανιών που καταναλώνουν πολύ ενέργεια (πετρέλαιο)*

Συμβολισμός:  $S_m(k, i, d)$ , (S: Strike), όπου

k : Τύπος απεργίας

i : Ένταση της απεργίας

d : Διάρκεια επίδρασης (σε ημέρες)

*Ειδικές αργίες – διακοπές*

Συμβολισμός:  $H_m(k, d)$ , (H: Holiday), όπου

k : Κατηγορία διακοπών

d : Διάρκεια (σε ημέρες)

*Ανακοινώσεις μελλοντικών αλλαγών στην τιμή των καυσίμων*

Συμβολισμός:  $A_m(p, t_a, t_e)$ , (A: Advanced announcement), όπου

$p$  : Ποσοστό αλλαγής της τιμής

$t_a$  : Ημερομηνία ανακοίνωσης

$t_e$  : Ημερομηνία εφαρμογής

Η εμφάνιση ενός υποκειμενικού παράγοντα σε κάποια χρονική στιγμή ονομάζεται *Υποκειμενικό – Κριτικό Γεγονός (Judgmental Event)* και η αναγνώριση αυτών των εμφανίσεων στις ιστορικές παρατηρήσεις οδηγεί στην κατασκευή της *Βάσης Ιστορικών Υποκειμενικών Γεγονότων*.

Ένα υποκειμενικό γεγονός και η επίδραση αυτού στην τιμή κάποιων παρατηρήσεων της χρονοσειράς ονομάζεται *Υποκειμενικό – Κριτικό Σενάριο (Judgmental Case)* και η αναγνώριση αυτών των επιδράσεων στις ιστορικές παρατηρήσεις αλλά και η κατασκευή μελλοντικών πιθανολογούμενων εμφανίσεων υποκειμενικών γεγονότων μαζί με τις επιδράσεις τους οδηγεί στην κατασκευή της *Βάσης Υποκειμενικών Σεναρίων*.

### 3.5.3 Τεχνητά Νευρωνικά Δίκτυα

#### 3.5.3.1 Γενικά

Τα Τεχνητά Νευρωνικά Δίκτυα ανήκουν στο γνωστικό πεδίο της Υπολογιστικής Νοημοσύνης και η χρήση τους έγκειται σε επίλυση προβλημάτων Τεχνητής Νοημοσύνης, χρησιμοποιώντας αριθμητικά μοντέλα.

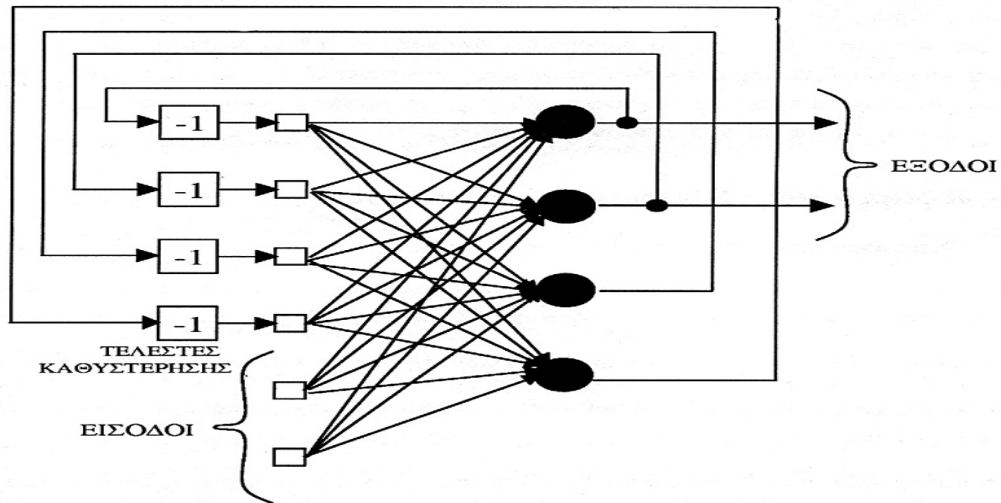
Παλαιότερα, οι υπολογισμοί στον προγραμματισμό ήταν ακολουθιακοί αλγόριθμοι. Με την εξέλιξη, όμως της Τεχνητής Νοημοσύνης έγιναν βήματα στην προσομοίωση της διαδικασίας υπολογισμών του ανθρώπινου εγκέφαλου.

Οι υπολογισμοί αυτοί: α) είναι καταναμημένοι και εκτελούνται παράλληλα, β) δεν περιέχουν όλη την πληροφορία για την άρτια λειτουργία του προγράμματος, η «μάθηση» είναι που παρέχει αυτή την δυνατότητα.

*Τεχνητό Νευρωνικό Δίκτυο* είναι μια αρχιτεκτονική δομή αποτελούμενη από έναν αριθμό τεχνητών νευρώνων. Κάθε νευρώνας χαρακτηρίζεται από εισόδους και εξόδους και υλοποιεί έναν απλό υπολογισμό. Κάθε σύνδεση μεταξύ δύο μονάδων χαρακτηρίζεται από μία τιμή βάρους. Οι τιμές των βαρών των συνδέσεων αποτελούν την αποθηκευμένη γνώση στο δίκτυο και καθορίζουν πλήρως την λειτουργικότητά του. Η έξοδος κάθε νευρώνα καθορίζεται από τον τύπο του νευρώνα, τη διασύνδεση με τους υπόλοιπους και πιθανώς κάποιες εξωτερικές εισόδους.

Στο σχήμα που ακολουθεί στην επόμενη σελίδα, φαίνεται μία δομή ενός νευρωνικού δικτύου με πρόσθια τροφοδότηση (Επαναληπτικό Δίκτυο).

Σχήμα 3.42 Επαναληπτικό Δίκτυο με ένα κρυμμένο επίπεδο



### 3.5.3.2 Βιολογικοί Νευρώνες – Τεχνητοί Νευρώνες

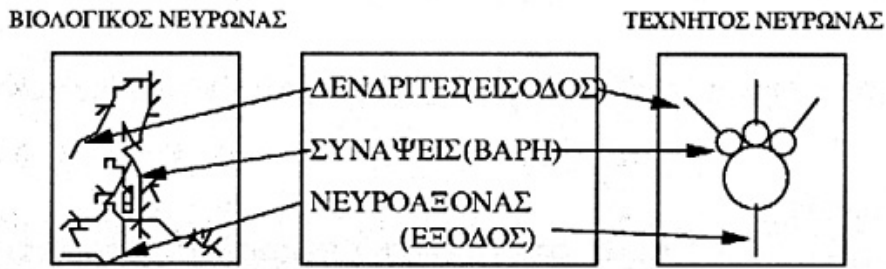
Είδαμε παραπάνω ότι τα Τ.Ν.Δ. προήλθαν από την προσπάθεια μίμησης των διεργασιών του ανθρώπινου μυαλού. Ας εστιάσουμε περισσότερο στις ομοιότητες αυτές:

Ο εγκέφαλος έχει ιδιότητες όπως προσαρμοστικότητα, ικανότητα αναγνώρισης από τα συμφοραζόμενα, ανοχή στα λάθη, μεγάλη χωρητικότητα μνήμης, ικανότητα επεξεργασίας πληροφοριών σε πραγματικό χρόνο. Αυτές οι ιδιότητες προκύπτουν κυρίως από την τοπολογία του εγκεφάλου.

Ο εγκέφαλος αποτελείται από νευρικά κύτταρα συνδεδεμένα μεταξύ τους, οι λεγόμενοι νευρώνες. Τα κύτταρα αυτά κάνουν συνάψεις με άλλους γειτονικούς νευρώνες. Κάθε νευρώνας εκτελεί απλούς ακολουθιακούς υπολογισμούς στα σήματα που δέχεται από τους άλλους νευρώνες. Τα φιλτράρει, τα ενισχύει κατάλληλα και παράγει τελικά ένα σήμα εξόδου το οποίο μεταδίδει μέσω των συνάψεών του στους άλλους νευρώνες. Η επίδραση ενός σήματος σε έναν νευρώνα μπορεί να είναι θετική ή αρνητική. Αυτή η απλή διεργασία εκτελείται παράλληλα σε κάθε νευρώνα, έτσι έχουμε ένα πανίσχυρο, υπολογιστικά, μοντέλο.

Σε πλήρη αντιστοιχία με το βιολογικό αυτό μοντέλο, ο τεχνητός νευρώνας αποτελείται από εισόδους που εφαρμόζονται με κάποια βάρη και από εξόδους. Το σχήμα 3.43. περιγράφει την αντιστοιχία:

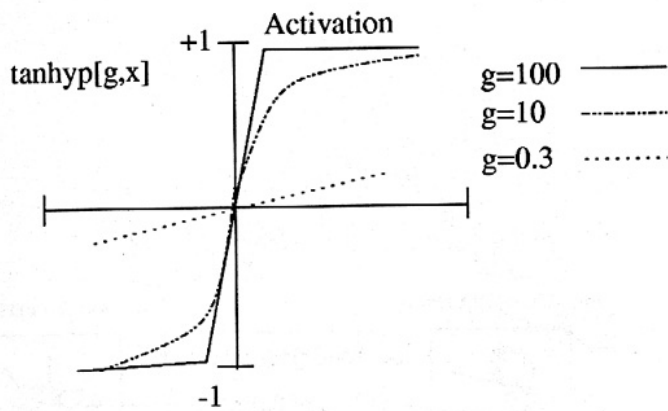
Σχήμα 3.43 Αντιστοιχία βιολογικού-τεχνητού νευρώνα.



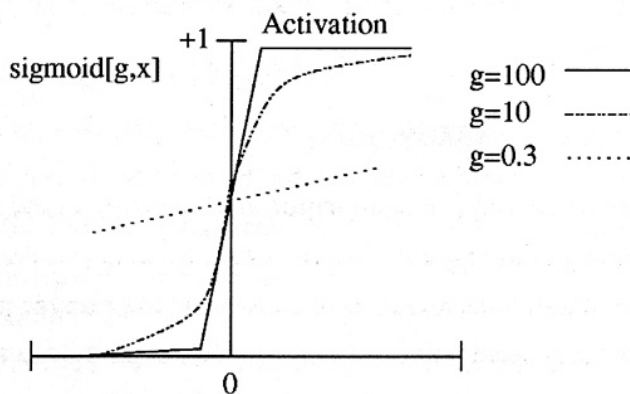
Έστω τεχνητός νευρώνας  $d$  με συνδέσεις εισόδου  $x_1, \dots, x_d$  και αντίστοιχες τιμές βαρών  $w_1, \dots, w_d$ . Ο υπολογισμός που εκτελεί ένας τέτοιος νευρώνας χωρίζεται σε δύο στάδια α) υπολογισμός της ενεργοποίησης  $u = \sum w_i x_i + \theta$ , όπου  $\theta$  η πόλωση του νευρώνα, β) υπολογισμός της εξόδου του νευρώνα περνώντας την ενεργοποίηση  $u$  μέσα από μια συνάρτηση ενεργοποίησης  $f: y = f(u)$ .

Η συνάρτηση ενεργοποίησης είναι μη γραμμική (εισαγάγει τη «μη γραμμικότητα» στο νευρώνα) και τις περισσότερες περιπτώσεις σιγμοειδής. Βασικοί τύποι σιγμοειδών συναρτήσεων είναι η λογιστική συνάρτηση και η υπερβολική εφαπτομένη. Στα σχήματα 3.44 και 3.45 φαίνονται αντίστοιχα οι συναρτήσεις αυτές.

Σχήμα 3.44 Λογιστική συνάρτηση  $f(u) = 1 / (1 + \exp(-gu))$



Σχήμα 3.45 Υπερβολική Εφαπτομένη  $f(u) = \tanh(gu)$



Η λογιστική συνάρτηση έχει πεδίο τιμών από 0 έως +1, ενώ η υπερβολική εφαπτομένη από -1 έως +1.

Μερικές από τις βασικές ικανότητες του ανθρωπίνου εγκεφάλου που οφείλονται στις πολλές μονάδες επεξεργασίας (Νευρώνες) και στην πολυπλοκότητα των μεταξύ τους διασυνδέσεων είναι οι ακόλουθες:

1. Αποθήκευση εμπειριών όπως η κατηγοριοποίηση ή συσχέτιση των δεδομένων εισόδου. Αυτό-οργάνωση των εμπειριών αυτών.
2. Απόκριση σε νέες εμπειρίες μέσω της συσχέτισής τους με τις αποθηκευμένες.
3. Εκτέλεση προβλέψεων για νέες καταστάσεις σύμφωνα με τις αποθηκευμένες εμπειρίες. Ικανότητα γενίκευσης.
4. Ανοχή στην παραμόρφωση, διαταραχή ή ατέλεια των δεδομένων εισόδου.
5. Ανοχή στις βλάβες. Ακόμα και η απώλεια μερικών νευρώνων αντιμετωπίζεται με κατάλληλη προσαρμογή αυτών που μένουν και πρόσθετη εκπαίδευση.
5. Ο εγκέφαλος φαίνεται να έχει διαθέσιμους νευρώνες, πιθανών αχρησιμοποίητους, έτοιμους προς χρήση. Άρα έχει τη δυνατότητα διαρκώς να μαθαίνει.
7. Η εξέταση του εγκεφάλου δεν παρέχει αρκετές πληροφορίες για την λειτουργία του. Υπάρχει μία αδιαφάνεια στην λειτουργία του.

Πολλές από τις ιδιότητες αυτές υπάρχουν και στα Τ.Ν.Δ. και έτσι τα καθιστούν την αιχμή του δόρατος στις εφαρμογές τεχνητής νοημοσύνης και εξομοίωσης της ανθρώπινης συμπεριφοράς όπως: επεξεργασία εικόνας/σήματος, μηχανική όραση, αναγνώριση προτύπων, ιατρικές εφαρμογές, αμυντικά συστήματα, εφαρμογές οικονομίας, έμπειρα συστήματα, πρόβλεψη χρονοσειρών, επικοινωνία ανθρώπου-υπολογιστή.

### 3.5.4 Μοντέλο Πρόβλεψης

Είναι πλέον σαφές ότι η χρήση των Τ.Ν.Δ. στις στατιστικές προβλέψεις και τις κριτικές επεμβάσεις γίνεται γιατί αυτά είναι πιο κοντά στον τρόπο σκέψης το ανθρώπου. Βέλτιστο μοντέλο προβλέψεων είναι το ανθρωπομορφικότερο.

Το προτεινόμενο μοντέλο αποτελείται από τα ακόλουθα βήματα:

- α. Κατασκευή Βάσης Υποκειμενικών Παραγόντων (Judgmental Factor Base)
- β. Κατασκευή Βάσης Ιστορικών Υποκειμενικών Γεγονότων (Judgmental Event Base)
- γ. Φιλτράρισμα της χρονοσειράς για απαλοιφή των ιστορικών υποκειμενικών γεγονότων (Monthly Judgmental Instances)
- δ. Κατασκευή Βάσης Υποκειμενικών Σεναρίων (Judgmental Case Base)
- ε. Κατασκευή Μοντέλου Νευρωνικού Δικτύου για Επεξεργασία Υποκειμενικών Σεναρίων (Judgmental Adjustment)
- στ. Κατασκευή Μοντέλου Νευρωνικού Δικτύου Παραγωγής Βασικής Τάσης (Main Trend Forecasting)

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

---

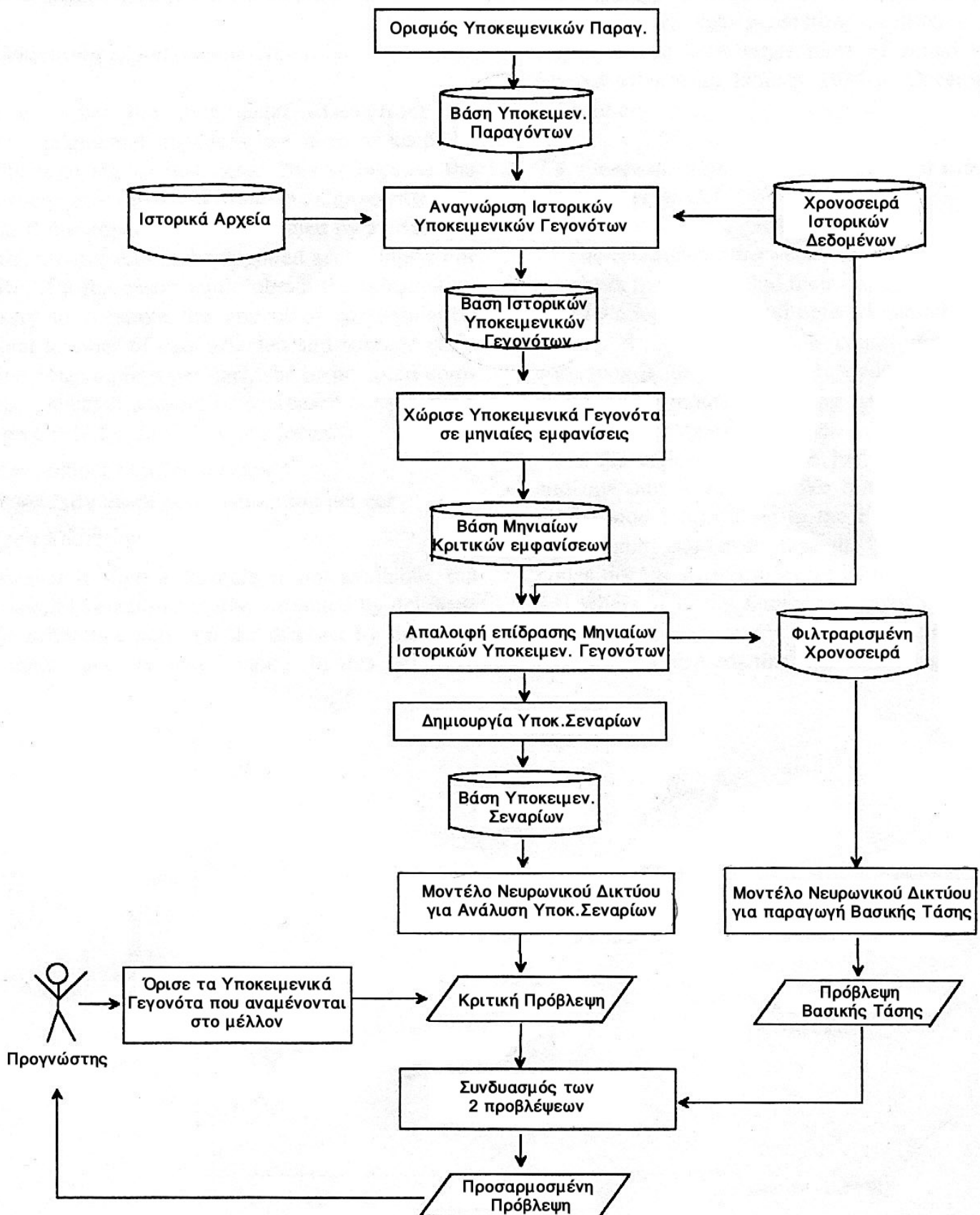
ζ. Ο Προγνώστης, ορίζει τις εισόδους στο Νευρωνικό Δίκτυο, δηλαδή τα υποκειμενικά σενάρια προς επεξεργασία (Judgmental Occurences)

η. Παραγωγή προβλέψεων από κάθε Μοντέλο Νευρωνικών Δικτύων και συνδυασμός αυτών (Additive Adjustment of Main Trend Forecast)

Σχηματικά το μοντέλο παρουσιάζεται στην επόμενη σελίδα με τις επεξηγήσεις που παραθέσαμε σε κάθε βήμα.



Σχήμα 3.46 Μοντέλο Πρόβλεψης κατά Jae Kyu Lee και Chang Seon Yum.



Αναλυτική περιγραφή των βασικών βημάτων του αλγορίθμου:

### **A. Κατασκευή Βάσης Υποκειμενικών Παραγόντων (Judgmental Factor Base)**

Η κατασκευή της Βάσης Υποκειμενικών παραγόντων γίνεται ανάλογα την εφαρμογή. Για παράδειγμα, στη περίπτωση της ζήτησης πετρελαίου έχουμε τέσσερις παράγοντες.

Η Βάση περιέχει τους παράγοντες με τις επιμέρους παραμέτρους του κάθε ενός.

### **B. Κατασκευή Βάσης Ιστορικών Υποκειμενικών Γεγονότων (Judgmental Event Base)**

Η Βάση Ιστορικών Υποκειμενικών Γεγονότων γίνεται σε δεύτερη φάση, αφού έχουν οριστεί οι Υποκειμενικοί Παράγοντες.

Τότε, με είσοδο μία Βάση Ιστορικών Γεγονότων, κατασκευάζουμε την Βάση Ιστορικών Υποκειμενικών Γεγονότων τις οποίες οι καταχωρίσεις είναι τα Ιστορικά Υποκειμενικά Γεγονότα που συνέβησαν σε μία χρονική στιγμή (π.χ. 1 μήνας) με επίδρασή στο υπό-μελέτη μέγεθος (π.χ. Οκτ.1998 |  $H_{98}(1,4)$ ,  $\alpha(22.7,12)$  ).

### **Γ. Φιλτράρισμα της χρονοσειράς για απαλοιφή των ιστορικών υποκειμενικών γεγονότων (Monthly Judgmental Instances)**

Από τα δεδομένα της χρονοσειράς αφαιρούμε την επίδραση των Υποκειμενικών Γεγονότων για την εκάστοτε περίοδο όπου αυτά συμβαίνουν. Το αποτέλεσμα είναι η φιλτραρισμένη χρονοσειρά που περιέχει μόνο την βασική τάση.

Υπάρχουν 2 μέθοδοι φιλτραρίσματος: α) με χρήση εξωτερικής φόρμουλας υπολογισμού της επίδρασης των Υποκειμενικών Γεγονότων π.χ. η επίδραση των «διατάξεων εξοικονόμησης ενέργειας» στο μέγεθος «ζήτηση πετρελαίου» μπορεί να υπολογιστεί από την παρακάτω σχέση, αν είναι διαθέσιμα τα απαιτούμενα στοιχεία:

**Επίδραση = (αριθμός αυτοκίνητων που αναφέρεται η διάταξη)\*(μέση ημερήσια κατανάλωση ανά αυτοκίνητο)\*(διάρκεια δράσης της διάταξης)**

Από την χρονοσειρά, αφαιρούμε την υπολογισμένη επίδραση, κατασκευάζοντας τελικά την καμπύλη της Βασικής Τάσης, καθώς και μία βάση με τα Ιστορικά Υποκειμενικά Γεγονότα και την επίδρασή τους σε συγκεκριμένη χρονική περίοδο.

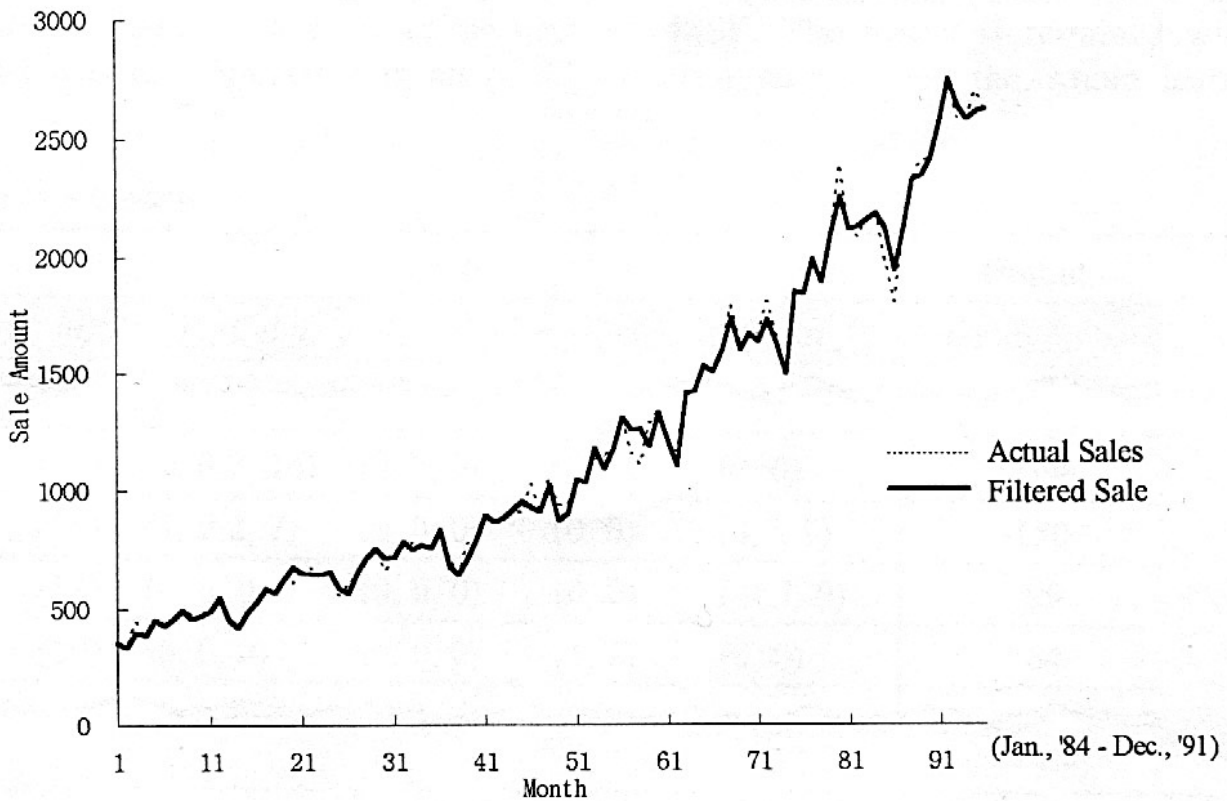
β) με την εκτίμηση που κάνει ο προγνώστης. Έχοντας στην διάθεσή του τα ιστορικά στοιχεία, ο προγνώστης, ορίζει για κάθε Υποκειμενικό γεγονός που συνέβη, πόση ήταν η επίδρασή του. Όπως και παραπάνω, αφαιρείται η εκάστοτε επίδραση από την χρονοσειρά και κατασκευάζουμε την βασική τάση και την Βάση με τα Ιστορικά Υποκειμενικά Γεγονότα και την επίδρασή τους.

Παραθέτουμε ένα παράδειγμα χρονοσειράς, η οποία έχει φιλτραριστεί από τα Υποκειμενικά Γεγονότα στο παρακάτω σχήμα 3.47. Παρατηρούμε ότι η χρονοσειρά έχει «εξομαλυνθεί» αφού έχουν

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

αφαιρεθεί τα Υπ.Γεγ. Τα Γεγονότα αυτά, μαζί με την επίδρασή τους, θα δημιουργήσουν την Βάση Δεδομένων που εξετάζουμε στο Δ.

**Σχήμα 3.47** Χρονοσειρά Φιλτραρισμένη από τα Υποκειμενικά Γεγονότα.



### Δ. Κατασκευή Βάσης Υποκειμενικών Σεναρίων (Judgmental Case Base)

Με την διεργασία Γ. παράγεται, όπως είδαμε, και μία βάση με τα Υποκειμενικά Γεγονότα που συνέβησαν μία χρονική περίοδο, μαζί με την επίδρασή τους στην χρονοσειρά. Αυτή η Βάση Δεδομένων, ονομάζεται *Βάση Υποκειμενικών Σεναρίων* και η μορφή των εγγραφών της είναι: π.χ. Σενάριο 34: Οκτ.1998 |  $H_{98}(1,4)$ ,  $\alpha(22.7,12)$  | -150.

Στο παρακάτω σχήμα 3.48 φαίνεται η δομή μίας τέτοιας βάσης δεδομένων. Η Μηνιαία επίδραση (Monthly Impact) είναι η επίδραση που έχει ένας συνδυασμός Υποκειμενικών Γεγονότων αναλυμένων στους επιμέρους παράγοντές τους, σε συγκεκριμένη χρονική περίοδο (m).

Σχήμα 3.48 Τμήμα Βάσης Δεδομένων Υποκειμενικών Σεναρίων

Judgmental Cases	Inputs					Output
	$m$	$r_m(k, i, d_m)$	$s_m(k, i, d_m)$	$h_m(k, d_m)$	$a_m(p_m, d_m)$	Monthly Impact
Case 20	(57)	(1, 0.2, 16)	(0, 0, 0)	(2, 1)	(0, 0)	-106
Case 21	(58)	(1, 0.2, 7)	(0, 0, 0)	(0, 0)	(4.7, 9)	-150
Case 22	(59)	(0, 0, 0)	(0, 0, 0)	(0, 0)	(-4.7, 9)	99
Case 23	(60)	(0, 0, 0)	(0, 0, 0)	(1, 2)	(0, 0)	64

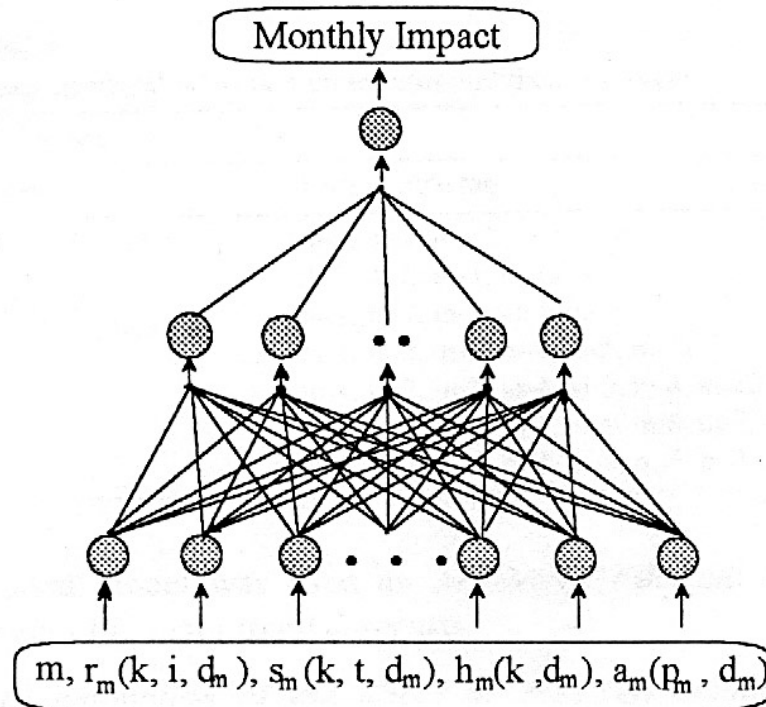
**Ε. Κατασκευή Μοντέλου Νευρωνικού Δικτύου για Επεξεργασία Υποκειμενικών Σεναρίων (Judgmental Adjustment)**

Η Βάση Δεδομένων με τα Υποκειμενικά Σενάρια είναι η είσοδος σε ένα Νευρωνικό Δίκτυο, κατάλληλα επιλεγμένο, το οποίο αφού εκπαιδευτεί κατάλληλα, θα μπορεί να προβλέπει την μελλοντική επίδραση που θα έχει ένας ή συνδυασμός υποκειμενικών παραγόντων που θα συμβούν μία χρονική περίοδο, στην χρονοσειρά.

Το Μοντέλο Νευρωνικού Δικτύου θα εκπαιδευτεί σαν «προγνώστης επίδρασης Υποκειμενικών Σεναρίων» χρησιμοποιώντας τη Βάση Δεδομένων. Η εκπαίδευσή του θα το εξειδικεύσει στην υπό-μελέτη χρονοσειρά με τους Υποκειμενικούς παράγοντες μόνο, που υπάρχουν στην Β.Δ.

Το σχήμα 3.49 δείχνει ένα τέτοιο Νευρωνικό Δίκτυο με εισόδους  $m$ =μήνας και τα Υποκειμενικά Γεγονότα αναλυμένα στις παραμέτρους των. Το “Monthly Impact” είναι η επίδραση των Υποκειμενικών Γεγονότων την εκάστοτε χρονική περίοδο (μήνας).

Σχήμα 3.49. Νευρωνικό Δίκτυο επεξεργασίας Υποκειμενικών Σεναρίων



**ΣΤ. Κατασκευή Μοντέλου Νευρωνικού Δικτύου Παραγωγής Βασικής Τάσης (Main Trend Forecasting)**

Η φιλτραρισμένη χρονοσειρά από το βήμα Γ. θα χρησιμοποιηθεί σαν είσοδος σε ένα δεύτερο μοντέλο Νευρωνικού Δικτύου το οποίο θα παράγει την Βασική Τάση (χρονοσειρά χωρίς την επίδραση των Υποκειμενικών Γεγονότων).

Το Μοντέλο αυτό θα πρέπει να επιλεγεί κατάλληλα και να εκπαιδευτεί ειδικευμένα σαν «προγνώστης Βασικής Τάσης της Χρονοσειράς» και ακολούθως θα προλέγει τις μελλοντικές τιμές της με βάση τις παρελθοντικές.

Στο σχήμα 3.50 φαίνονται οι είσοδοι (inputs) που χρειάζεται ένα τέτοιο μοντέλο νευρωνικού, προκειμένου να παράγει την πρόβλεψη (output). Παρατηρούμε ότι οι είσοδοι είναι προσεκτικά επιλεγμένοι ώστε να περιλαμβάνουν τις γειτονικές τιμές της χρονοσειράς ( $m-1$ ,  $m-2$ , ...) καθώς όμως και τις απομακρυσμένες. Έτσι εξασφαλίζουμε ότι η πρόβλεψη Βασικής Τάσης ( $m$ ) θα είναι το δυνατόν ακριβέστερη. Η Βάση Δεδομένων αυτή, αποτελεί την Βάση Δεδομένων των Σειρών Εκπαίδευσης με τα οποία θα «μάθει» το Νευρωνικό να προβλέπει.

**Σχήμα 3.50** Είσοδοι και έξοδοι Μοντέλου Νευρωνικού Δικτύου για την πρόβλεψη της Βασικής Τάσης. Τμήμα της Βάσης Δεδομένων Σετ Εκπαίδευσης

Patterns	Inputs	Output
	Filtered Data Set at $(m-12, m-6, m-5, m-4, m-3, m-2, m-1)$	Filtered Data at $(m)$
Pattern 55	(885, 872, 902, 1039, 1031, 1176, 1088)	1,174
Pattern 56	(917, 902, 1039, 1031, 1176, 1088, 1174)	1,306
Pattern 57	(953, 1039, 1031, 1176, 1088, 1174, 1306)	1,259
Pattern 58	(928, 1031, 1176, 1088, 1174, 1306, 1259)	1,257
Pattern 59	(908, 1176, 1088, 1174, 1306, 1259, 1257)	1,182
Pattern 60	(1018, 1088, 1174, 1306, 1259, 1257, 1182)	1,326
Pattern 61	(872, 1174, 1306, 1259, 1257, 1182, 1326)	1,204
Pattern 62	(902, 1306, 1259, 1257, 1182, 1326, 1204)	1,097
Pattern 63	(1039, 1259, 1257, 1182, 1326, 1204, 1097)	1,408
Pattern 64	(1301, 1257, 1182, 1326, 1204, 1097, 1408)	1,419
Pattern 65	(1176, 1182, 1326, 1204, 1097, 1408, 1419)	1,529
Pattern 66	(1088, 1326, 1204, 1097, 1408, 1419, 1529)	1,499

**Z. Ο Προγνώστης, ορίζει τις εισόδους στο Νευρωνικό Δίκτυο, δηλαδή τα υποκειμενικά σενάρια προς επεξεργασία (Judgmental Occurrences)**

Στο Μοντέλο πρόβλεψης που παραθέτουμε, απαραίτητη είναι και η ύπαρξη ενός εξωτερικού προγνώστη, ο οποίος ορίζει τα Υποκειμενικά Γεγονότα που πρόκειται να συμβούν στο μέλλον.

Ο εξωτερικός προγνώστης δεν είναι απαραίτητο να είναι πρόσωπο, μπορεί να είναι και ένας συνδυασμός αλγορίθμων όπου ο ένας π.χ. υπολογίζει τις αργίες που πρόκειται να συμβούν με βάση το ημερολόγιο, ο άλλος μαζεύει τις ανακοινώσεις του υπουργείου συγκοινωνιών για τα αυτοκίνητα τον τύπο, τις κατατάσσει, τις αξιολογεί και υπολογίζει τις μελλοντικές εξαγγελίες της κυβέρνησης κ.α.

### **Η. Παραγωγή προβλέψεων από κάθε Μοντέλο Νευρωνικών Δικτύων και συνδυασμός αυτών (Additive Adjustment of Main Trend Forecast)**

Από τα Ε. (με κατάλληλη είσοδο για τα Υποκειμενικά Γεγονότα που θα συμβούν Ζ.) και ΣΤ. (με είσοδο την παρελθοντική Βασική Τάση) έχουμε δύο δεδομένα για το μέλλον: α) την επίδραση που θα έχουν οι μελλοντικοί Υποκειμενικοί παράγοντες και β) την τιμή της Βασικής Τάσης στο μέλλον. Με την άθροιση αυτών, έχουμε την ολική πρόβλεψη για την υπό-μελέτη χρονοσειρά.

### **3.5.5. Πληροφοριακό σύστημα κριτικών και στατιστικών προβλέψεων με χρήση τ.ν.δ. σε στατιστικές προβλέψεις**

#### **3.5.5.1 Νευρωνικό Δίκτυο για την Κριτική και Στατιστική Πρόβλεψη.**

Σκοπός μας είναι να κατασκευάσουμε ένα νευρωνικό δίκτυο που θα μπορεί να προβλέπει την επίδραση ενός ή συνδυασμού κριτικών γεγονότων που θα συμβούν στο μέλλον..

Το μοντέλο αυτό θα πρέπει:

A. Να επιλεγεί κατάλληλα η αρχιτεκτονική του.

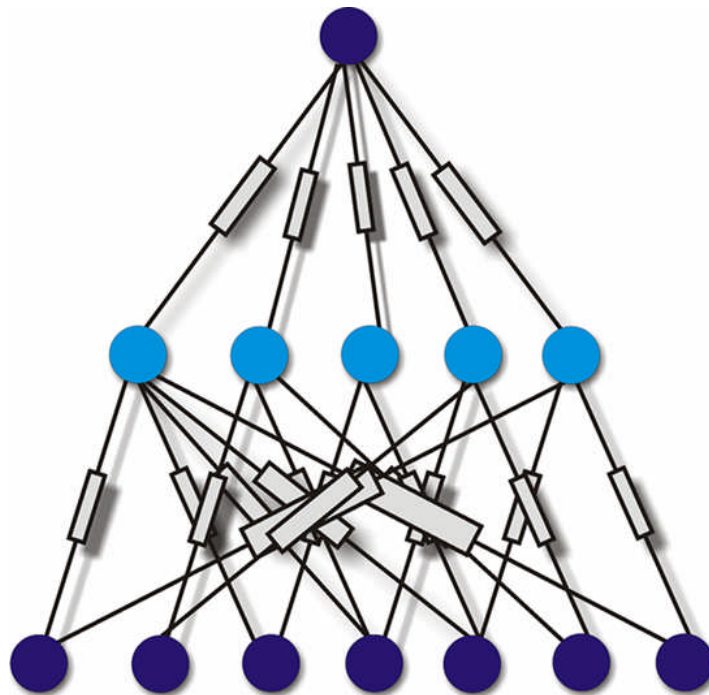
B. Να επιλεγούν τα Σει Εκπαίδευσής του και να καθοριστούν οι εποχές εκπαίδευσης..

Γ. Να εκπαιδευτεί κατάλληλα.

Δ. Με δοσμένα μελλοντικά κριτικά γεγονότα, να κάνει πρόβλεψη για την επίδρασή τους στην Βασική Τάση.

Με βάση την μελέτη των Jea Kyu Lee και Chang Seon Yum, για την παραπάνω εφαρμογή, ένα νευρωνικό δίκτυο με ένα κρυμμένο επίπεδο και μία έξοδο είναι το κατάλληλο. Σχήμα 3.51 στην επόμενη σελίδα.

Σχήμα 3.51: Νευρωνικό δίκτυο για την εφαρμογή πρόβλεψης της επίδρασης Κριτικών Γεγονότων



παρακάτω εξετάζουμε τα βήματα του πληροφοριακού μας συστήματος όπως τα θέσαμε ήδη.

#### **A. Επιλογή κατάλληλου Τ.Ν.Δ.**

Το Τ.Ν.Δ. θα πρέπει να επιλέγεται ανάλογα με τις εισόδους (παράμετροι των κριτικών γεγονότων). Έτσι, αν έχουμε 3 κριτικά γεγονότα με 3 παραμέτρους το κάθε ένα, θα προκύψουν 9 εισοδοί, άρα το Νευρωνικό θα είναι της μορφής  $9-\chi-1$  όπου  $\chi$  οι νευρώνες του κρυμμένου επιπέδου.

Το κρυμμένο επίπεδο μπορεί να έχει από 1 έως  $2*(\text{είσοδοι})+1$  νευρώνες σύμφωνα με το θεώρημα των Hecht-Nielsen, άρα στην περίπτωση μας οι δυνατοί νευρώνες είναι 1-19.

Προκειμένου να καθορίσουμε το βέλτιστο νευρωνικό δίκτυο, θα δοκιμάσουμε όλες τις δυνατές αρχιτεκτονικές (19 στην περίπτωση των 9 εισόδων) και αυτή με το μικρότερο σφάλμα θα επιλεγεί. Έτσι, στον αλγόριθμό μας, επιλέγουμε τα 2/3 των Σετ Εκπαίδευσης και εκπαιδεύουμε κάθε δυνατή αρχιτεκτονική για 50 εποχές (1 εποχή=1 πέρασμα το Σετ Εκπαίδευσης). Στο τέλος των 50 εποχών βλέπουμε το μέσο τετραγωνικό σφάλμα.

Αφού περάσουμε όλες τις αρχιτεκτονικές, συγκρίνουμε τα σφάλματα και επιλέγουμε την αρχιτεκτονική εκείνη με το μικρότερο.



**Β. Επιλογή σετ εκπαίδευσης και εποχών εκπαίδευσης.**

Τα Σετ εκπαίδευσης πρέπει να είναι χαρακτηριστικά της χρονοσειράς και να καλύπτουν όλες τις πιθανές περιπτώσεις κριτικών γεγονότων. Είναι πολύ βασική η επιλογή των Σετ Εκπαίδευσης, αφού από αυτά εξαρτάται η άρτια λειτουργία του Τ.Ν.Δ.

Οι εποχές εκπαίδευσης πρέπει να επιλεγούν και αυτές με βάση το μέσο τετραγωνικό σφάλμα. Έτσι για την αρχιτεκτονική-νικητής που έχουμε επιλέξει, δοκιμάζουμε να τρέξουμε τα τελευταία 2/3 των Σετ Εκπαίδευσης για εποχές από 30 έως 1000. Στο τέλος των υπό-έλεγχο εποχών κρατάμε το μέσο τετραγωνικό σφάλμα.

Τέλος, συγκρίνουμε τα σφάλματα και επιλέγουμε για εποχή-νικητή αυτή με το μικρότερο.

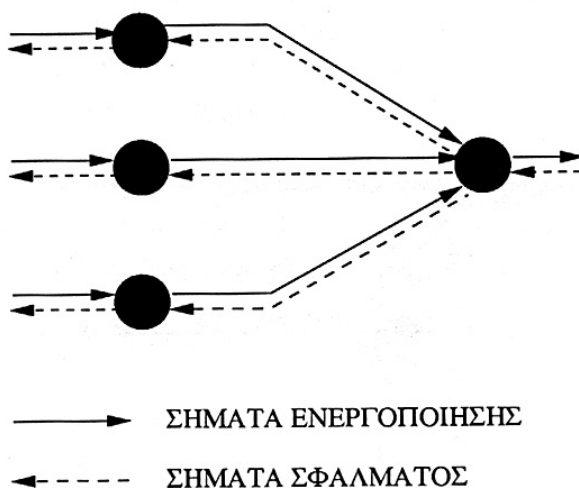
**Γ. Εκπαίδευση Τ.Ν.Δ.**

Η εκπαίδευση του Τ.Ν.Δ. που έχει επιλεγεί γίνεται σε όλα τα Σετ Εκπαίδευσης για τόσες εποχές όσες βρήκαμε στον διαγωνισμό εποχών.

Ο αλγόριθμος που χρησιμοποιούμε είναι ο BackPropagation, που αποτελείται από τα σήματα ενεργοποίησης και τα σήματα σφάλματος. Στα σήματα ενεργοποίησης υπολογίζεται η έξοδος και στα σήματα σφάλματος αναδιοργανώνονται τα βάρη προκειμένου να μειωθεί το σφάλμα.

Στο σχήμα 3.52. βλέπουμε τα σήματα αυτά.

**Σχήμα 3.52** Ροή σημάτων σε ένα τμήμα Τ.Ν.Δ. που τρέχει BackPropagation.

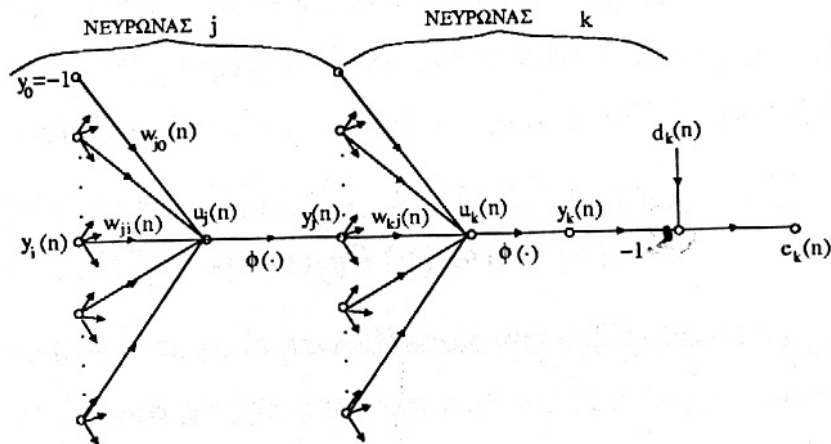


Αναλυτικότερα, οι υπολογισμοί σε ευθεία φορά έχουν σαν αποτέλεσμα τις εξόδους κάθε νευρώνα και τελικά την τελική έξοδο.

Γνωρίζοντας, όμως, την αναμενόμενη έξοδο, εξάγουμε το ολικό σφάλμα, το οποίο «μοιράζουμε» σε κάθε νευρώνα και με κατάλληλους υπολογισμούς διορθώνουμε τα βάρη των εισόδων του, έτσι ώστε στο επόμενο πέρασμα το ολικό λάθος να είναι μικρότερο.

Η Ροή των υπολογισμών στους νευρώνες φαίνεται και στο παρακάτω σχήμα:

Σχήμα 3.53 Γραφική αναπαράσταση της ροής των υπολογισμών.



### BackPropagation

1. Αρχικοποίηση. Αρχικοποίηση των βαρών των συνδέσεων με μικρές τυχαίες τιμές ομοιόμορφα κατανεμημένες (συνήθως στο (-1,+1)).
2. Παρουσίαση των προτύπων εκπαίδευσης. Για κάθε πρότυπο του συνόλου εκτέλεση των σταδίων 3 και 4 που ακολουθούν.
3. Υπολογισμοί ευθείας φοράς. Θεωρούμε το πρότυπο εκπαίδευσης  $|x(n), d(n)|$ , όπου  $x(n)$  είναι το διάνυσμα που εφαρμόζεται στην είσοδο του δικτύου και  $d(n)$  είναι το επιθυμητό διάνυσμα εξόδου του δικτύου. Υπολογισμός των σημάτων που προκύπτουν ανά επίπεδο. Η στάθμη ενεργοποίησης του νευρώνα  $j$  του επιπέδου  $l$  είναι

Όπου  $y^{(l-1)}(n)$  είναι το σήμα εισόδου του νευρώνα  $l$  του προηγούμενου επιπέδου  $l-1$ , κατά την επανάληψη  $n$ , και  $w_{ji}^{(l)}(n)$  είναι το βάρος της σύναψης του νευρώνα  $j$  του επιπέδου  $l$  που

$$u_j^{(l)}(n) = \sum_{i=0}^p w_{ji}^{(l)}(n) y_i^{(l-1)}(n)$$

τροφοδοτείται από το νευρώνα  $l$  του επιπέδου  $l-1$ . Για  $l=0$  έχουμε

$y_0 = -1$  και  $w_{j0} = \theta_j$  είναι η τιμή πόλωσης. Χρησιμοποιώντας την υπερβολική εφαπτομένη σαν συνάρτηση ενεργοποίησης, η έξοδος για το νευρώνα  $j$  του επιπέδου  $l$  είναι

$$y_j^{(l)}(n) = \frac{1}{1 + \exp(-u_j^{(l)}(n))}$$

Αν ο νευρώνας  $j$  ανήκει στο πρώτο κρυμμένο επίπεδο ( $l=1$ ), θέτουμε

$$y_j^{(0)}(n) = x_j(n)$$

Όπου  $x_j(n)$  είναι το  $j$ -οστό στοιχείο του διανύσματος εισόδου  $x(n)$ . Αν ο νευρώνας ανήκει στο επίπεδο εξόδου ( $l=L$ ), θέτουμε

$$y_j^{(L)}(n) = o_j(n)$$

Έτσι υπολογίζεται το σήμα σφάλματος

$$e_j(n) = d_j(n) - o_j(n)$$

όπου  $d_j(n)$  είναι το  $j$ -οστό στοιχείο του επιθυμητού διανύσματος εξόδου  $d(n)$ .

4. Υπολογισμοί αναστροφής φοράς. Υπολογισμός των τιμών  $\delta$  (τοπικές κλίσεις) προχωρώντας αντίστροφα στο δίκτυο ανά επίπεδο:

$$\delta_j^{(L)}(n) = e_j^{(L)}(n) o_j(n) [1 - o_j(n)]$$

$$\delta_j^{(l)}(n) = y_j^{(l)}(n) [1 - y_j^{(l)}(n)] \sum_k \delta_k^{(l+1)}(n) w_{kj}^{(l+1)}(n)$$

για  $j$  στο επίπεδο εξόδου, και

για  $j$  σε κρυμμένο επίπεδο.

Προσαρμογή των βαρών του επιπέδου  $l$ , σύμφωνα με τον κανόνα δέλτα:

$$w_{ji}^{(l)}(n+1) = w_{ji}^{(l)}(n) + \alpha [w_{ji}^{(l)}(n) - w_{ji}^{(l)}(n-1)] + \eta \delta_j^{(l)}(n) y_i^{(l-1)}(n)$$

Όπου  $\eta$  είναι ο ρυθμός μάθησης και  $\alpha$  είναι η σταθερά momentum. Η σταθερά momentum αυξάνει τον ρυθμό μάθησης και λέγεται και όρος ορμής.

5. Επανάληψη. Οι υπολογισμοί επαναλαμβάνονται για τα λοιπά Σει εκπαίδευσης, όσες φορές, όσο είναι οι εποχή-νικητής είτε εάν ξεπεραστεί ένα minimum σφάλμα

#### Δ. Πρόβλεψη με την εφαρμογή των μελλοντικών γεγονότων

Σε αυτή τη φάση το T.N.Δ. είναι εκπαιδευμένο και έχει ειδικευτεί στο να μπορεί να προβλέπει την επίδραση που έχει ένας συνδυασμός κριτικών Γεγονότων στην υπό-μελέτη χρονοσειρά.

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

Ο προγνώστης, δίνει τώρα τα κριτικά γεγονότα που θα γίνουν στο άμεσο μέλλον και τα Τ.Ν.Δ. επιστρέφει την πρόβλεψη για την επίδρασή τους στην Βασική Τάση.

Όπως είναι φυσικό όσο μακρύτερα ζητείται να γίνει πρόβλεψη, τόσο θα μεγαλώνει το σφάλμα. Ενδεικτικά παραθέτουμε το παρακάτω παράδειγμα που στηρίζει αυτή την άποψη.

**ΠΙΝΑΚΑΣ 3.29:** *Η πρόβλεψη για 1 μήνα μετά έχει μικρότερο σφάλμα από αυτή για 3 μήνες μετά.*

Effect of judgmental model				
Forecasting time	Criteria	<i>MAIN(NN Filtered)</i>	<i>MAIN(NN Filtered) + JUDGMENTAL(NN)</i>	Effect of <i>JUDGMENTAL(NN)</i>
One-month ahead	MSE	9315	7418	-1897 (-20.3%)
	MAE	85.8	76.9	-8.9 (-10.4%)
	MAPE	2.66	2.41	-0.25 (-9.4%)
Three-months ahead	MSE	12144	9527	-2617 (-21.5%)
	MAE	100.1	89.1	-11.0 (-10.9%)
	MAPE	3.17	2.82	-0.35 (-11.0%)

MSE: mean squared errors, MAE: mean absolute errors, MAPE: mean absolute percentage errors.

Το παράδειγμα αναφέρεται σε ένα πείραμα Κριτικής Πρόβλεψης, το οποίο εφαρμόστηκε σε χρονοσειρά για την ζήτηση πετρελαίου. Χρησιμοποιήθηκε η μέθοδος που περιγράφεται στο σχήμα 11.14 και εκτός από την επίδραση του Νευρωνικού για τις Υποκειμενικές Προβλέψεις (Judgmental NN) στα αποτελέσματα του Νευρωνικού για την Βασική Τάση (MainNN/Filtered), φαίνονται και οι μετρήσεις σφάλματος (σφάλμα=απόκλιση της εξόδου από την αναμενόμενη τιμή).

Το μέσο τετραγωνικό σφάλμα (MSE) για προβλέψεις σε βάθος ενός μήνα είναι μικρότερο από αυτό των προβλέψεων σε βάθος τριών μηνών.

**3.5.6 Χρήση των νευρωνικών δικτύων για προβλέψεις σε συστήματα εμπορικών συναλλαγών**

Τα νευρωνικά δίκτυα παρέχουν προβλέψεις των κινήσεων της αγοράς, οι οποίες μπορούν έπειτα να αποτελέσουν τη βάση για την μετατροπή της αγοράς σε ένα αυτοματοποιημένο σύστημα. Ένα προδιαμορφωμένο δίκτυο είναι η φυσική επιλογή για εμπόριο σε πραγματικό χρόνο, το οποίο έχει καταρτισθεί κατά τη νύκτα, κατά το προηγούμενο Σαββατοκύριακο, ή πολύ πιο πίσω στο παρελθόν. Σε οποιοδήποτε από τα παραπάνω, η ταχύτητα υπολογισμού του προδιαμορφωμένου δικτύου είναι αρκετά γρήγορη, ώστε να καταστήσει ένα τέτοιο σύστημα εφαρμόσιμο για στιγμιαίο «κατέβασμα» δεδομένων αγοράς σε πραγματικό χρόνο και ικανό να παραγάγει μια άμεση απάντηση.

Η εφαρμογή των προβλέψεων απαιτεί μια στρατηγική για την αντιμετώπιση αντίθετων κινήσεων της αγοράς. Το ερώτημα για το πότε μπαίνεις ή βγαίνεις από αγορά εναπόκειται κατά ένα μεγάλο μέρος από τις προβλέψεις. *Τερματισμοί και όρια* παρέχουν εγκαταστάσεις ασφάλειας για τον έλεγχο του κινδύνου. Οι καλύτερες αξίες που θα χρησιμοποιηθούν μπορούν να βρεθούν μέσω μιας διαδικασίας βελτιστοποίησης πάνω σε μια ιστορική περίοδο κατάρτισης.

Ο σκοπός ενός τερματισμού είναι να προετοιμάζει μια έξοδο από την αγορά, αφήνοντας εντολή να εκτελεστεί η έξοδος μόλις η τιμή αγγίξει ένα ορισμένο επίπεδο.

Το όριο είναι παρόμοιο με τον τερματισμό, αλλά εφαρμόζεται στην είσοδο παρά στην έξοδο. Κατά συνέπεια αν μια εντολή μπει στην αγορά, αυτό μπορεί να γίνει με την εξασφάλιση ότι εάν η τιμή έχει αλλάξει από την προβλεπόμενη θέση, λόγω της ολίσθησης, κατά ένα ποσό μεγαλύτερο από το όριο, τότε η είσοδος στην αγορά καθυστερείται μέχρι (και εάν) αντιστραφεί η αγορά.

Η αξιολόγηση του γενικού συστήματος, προβλέψεις συν το τερματισμό/τη στρατηγική ορίου, πρέπει να εκτελεστεί κατά τη διάρκεια μιας δοκιμαστικής περιόδου. Το σύστημα πρέπει να παραγάγει τουλάχιστον 30 συναλλαγές κατά τη διάρκεια της δοκιμαστικής περιόδου προκειμένου να παρέχει ένα ελάχιστο δείγμα μεγέθους για να μετρήσουμε την απόδοση.

Τα μέτρα απόδοσης είναι σημαντικά στην αξιολόγηση σχετικά με το εάν οι στόχοι έχουν εκπληρωθεί και επιπλέον για τη σύγκριση των διάφορων εναλλακτικών στρατηγικών σε ένα σύστημα εμπορικών συναλλαγών. Τα κριτήρια στον καθορισμό μιας στρατηγικής εμπορικών συναλλαγών θα ποικίλουν σύμφωνα με τον επιτρεπόμενο βαθμό έκθεσης σε κίνδυνο, έτσι ώστε τα κριτήρια αξιολόγησης που επιλέγονται, βασιζόμενα σε προτεραιότητες, να είναι θέμα επιλογής.

Ο ακόλουθος κατάλογος περιγράφει τα κατάλληλα μέτρα απόδοσης:

**Υπόλοιπο:** Το υπόλοιπο ενός λογαριασμού στο τέλος της δοκιμαστικής περιόδου: διακρινόμενο σε συνολικό κέρδος, συνολική απώλεια, και υπόλοιπο κέρδους και απώλειας.

**Αριθμός Συναλλαγών και Κόστος:** Ο αριθμός κλειστών συναλλαγών συν τον τόκο και το κόστος συναλλαγών. Το κόστος πρέπει να περιληφθεί στο μέρος απώλειας κατά τη δήλωση υπολοίπου. Επίσης παρέχει το χωριστό αριθμό κερδοφόρων και ζημιογόνων συναλλαγών, χρήσιμο όταν δηλώνεται ως ποσοστό.

**Μέσοi όροι:** Ο τόκος είναι πιθανόν καλύτερο να μείνει εκτός των υπολογισμών. Το μέσο κέρδος ανά συναλλαγή (συμπεριλαμβανομένου του κόστους) - ένας χαμηλός αριθμός δείχνει ένα στενό περιθώριο λάθους. Το μέσο κέρδος μόνο σε κερδοφόρες συναλλαγές - ένας υψηλός αριθμός δείχνει την καλή εκμετάλλευση των κερδοφόρων τάσεων.

**Μέγιστα:** Η μέγιστη ελάττωση, με και χωρίς το τόκο - και οι δύο αριθμοί παρέχουν χρήσιμες πληροφορίες. Η μέγιστη διαδοχική απώλεια και το μέγιστο διαδοχικό κέρδος, παρέχουν χρήσιμα όργανα ελέγχου, καθώς επίσης και το μέγιστο κέρδος και τη μέγιστη απώλεια σε μια και μόνο συναλλαγή.

**Αναλογίες Επιστροφής:** Αυτές οι αναλογίες μπορούν να αξιολογηθούν κατά τη διάρκεια της δοκιμαστικής περιόδου για τη σύγκριση μεταξύ τους, παραδείγματος χάριν όπου οι παράμετροι στο σύστημα εμπορικών συναλλαγών είναι ποικίλες. Στη σύγκριση έναντι σε άλλους δημοσιευμένους αριθμούς απόδοσης, τα ποσοστά επιστροφής πρέπει να υπολογιστούν με τη διαίρεση της επιστροφής με την περίοδο που καλύπτεται (στις ίδιες χρονικές μονάδες που χρησιμοποιούνται για να παράγουν τον αριθμό επιστροφής). Το ποσοστό επιστροφής μπορεί έπειτα να χρησιμοποιηθεί για να παρέχει μια επιστροφή για οποιαδήποτε περίοδο T, απαιτείται, με κατάλληλο πολλαπλασιασμό με T (που παράγει μια απλή επιστροφή ενδιαφέροντος).

Η επιστροφή περιόδου εκμετάλλευσης R , εκφρασμένη ως ποσοστό, καθορίζεται από τη σχέση

$$R = 100 \frac{I + B - M}{C + M} \quad (3.127)$$

όπου I είναι ο τόκος που κερδίζεται από καταθέσεις μικτού κέρδους, B είναι το συνολικό υπόλοιπο συμπεριλαμβανομένων του κόστους, C είναι το κόστος και M είναι το συνολικό ποσό του μικτού κέρδους που χάνεται κατά τη διάρκεια της τελευταίας δοκιμαστικής περιόδου. Αυτό μπορεί επίσης να κληθεί επιστροφή στο μικτό κέρδος.

Δύο λόγοι που περιλαμβάνουν ένα μέτρο του κινδύνου, μέσω της σταθερής απόκλισης, είναι ο λόγος Sharpe και μια παραλλαγή που αποδίδεται στον Kaufman: Η αναλογία Sharpe καθορίζεται από τη σχέση

$$R_s = \frac{R - R_f}{\sigma} \quad (3.128)$$

όπου σ είναι η σταθερή απόκλιση του υπολοίπου μεταξύ κέρδους και απώλειας κατά τη διάρκεια της δοκιμαστικής περιόδου, το R είναι η επιστροφή της περιόδου εκμετάλλευσης και R<sub>f</sub> το επιτόκιο των κρατικών ομολόγων, ή άλλη ασφάλεια απαλλαγμένη κινδύνου, προσαρμοσμένη στη διάρκεια της περιόδου εκμετάλλευσης. Ο ετήσιος λόγος Sharpe μπορεί να υπολογιστεί με την αξιολόγηση του λόγου κατά τη διάρκεια των εβδομαδιαίων περιόδων και τον υπολογισμό μέσου

όρου αναλόγως. Ο χαμηλός κίνδυνος συνεπάγεται τη μικρή σταθερή απόκλιση, έτσι ο λόγος Sharpe θα είναι σχετικά υψηλός για την ίδια επιστροφή.

Ο Kaufman εισήγαγε το λόγο μείωσης αξίας μετοχής:  $EDR = R/\sigma_d$ , όπου  $\sigma_d$  είναι η σταθερή απόκλιση μόνο της μείωσης αξίας μετοχής, όπου η μείωση της αξίας μιας μετοχής είναι το υπόλοιπο μεταξύ κέρδους και απώλειας που μετριέται με βάση όλα τα προηγούμενα σημεία μέγιστου και ελαχίστου.

Ο Schwager ανέπτυξε το μέσο μέγιστο λόγο επανασχεδίασης, ως εξής :

$$AMR = \frac{1}{N} \sum_{i=1}^N pos(MCE_i - TE_i) \quad (3.129)$$

όπου η λειτουργία  $pos(\dots)$  είναι ίση με το περιεχόμενο (...) όταν αυτό είναι θετικό, και είναι μηδέν ειδάλλως. MCE είναι τιμή κλεισίματος της μετοχής πριν από την ημέρα  $i$ , TE είναι η συνολική τιμή της μετοχής (κλεισίματος και τόκος) μόνο για την ημέρα  $i$ .

Ο λόγος ανταμοιβής-κινδύνου ορίζεται ως το καθαρό κέρδος κατά τη διάρκεια της δοκιμαστικής περιόδου διαιρεμένου με τη μέγιστη ελάττωση. Ο λόγος απόδοσης ορίζεται ως ο λόγος του μέσου κέρδους ανά συναλλαγή διαιρεμένου με τη σταθερή απόκλιση του μέσου κέρδους.

Ο βαθμός φορητότητας πρέπει επίσης να υπολογιστεί, όπου τα ικανοποιητικά δεδομένα ιστορικού το επιτρέπουν, μετρώντας την απόδοση κατά τη διάρκεια περισσότερων από μιας δοκιμαστικών περιόδων.

Η τελική εκτίμηση αφορά την ποιότητα του συγχρονισμού. Τα συστήματα συναλλαγών που είναι βασισμένα σε κανόνες τεχνικής ανάλυσης είναι ή συντονισμένα για να είναι κερδοφόρα κατά τη διάρκεια των περιόδων τάσης ή κατά τη διάρκεια των περιόδων που οι τιμές μετακινούνται διαφορετικά, δεδομένου ότι ένα σύστημα βασισμένο σε ένα κανόνα κανόνες δεν μπορεί να είναι κερδοφόρο και στις δύο περιπτώσεις. Αυτό φέρνει στο φως την ιδιότητα της ποιότητας συγχρονισμού, αφού μερικές καλές συναλλαγές, που παράγουν μεγάλα κέρδη, είναι σημαντικότερες από πολλές μικρές απώλειες. Κατά συνέπεια, η ποσοστιαία ακρίβεια των συναλλαγών δεν είναι χρήσιμο μέτρο της απόδοσης όταν αυτή λαμβάνεται μεμονωμένα, (εκτός αν, φυσικά, είναι πάρα πολύ υψηλότερη από 50%).

Η πτυχή του συγχρονισμού έχει επίσης σχέση στα συστήματα που είναι βασισμένα στα νευρωνικά δίκτυα. Ενώ οι πολύ υψηλές ακρίβειες θα οδηγήσουν αναμφισβήτητα σε κερδοφόρα αποτελέσματα, δεν υπάρχει κανένας λόγος να υποθέσουμε ότι μια πρόβλεψη μικρής ακρίβειας είναι απαραίτητως ασύμφορη, αρκεί οι προβλέψεις να εφαρμόζονται σε ένα σύστημα συναλλαγών και η ικανότητα του να εκμεταλλεύεται τις μεγάλες κινήσεις να μετράται.

3.5.6.1 Διαχείριση κίνδυνου - Κίνδυνος χρεοκοπίας

Τα κύρια στοιχεία που ενδιαφέρουν εδώ είναι τα ακόλουθα:

1. Κίνδυνος χρεοκοπίας.
2. Διάθεση του κεφαλαίου με βέλτιστο τρόπο.
3. Διαφοροποίηση: εφαρμογή της σύγχρονης θεωρίας χαρτοφυλακίων χρηματιστηρίου στο μέλλον.

**Κίνδυνος χρεοκοπίας**

Οι δύο παράμετροι που διαδραματίζουν κεντρικό ρόλο στους υπολογισμούς κινδύνου είναι η πιθανότητα κέρδους,  $p$  (και η αντίστοιχη πιθανότητα της απώλειας  $q=1-p$ ), και η ο λόγος εξόφλησης,  $R_p$ . Αντί της άγνωστης μελλοντικής απόδοσης ενός συστήματος συναλλαγών, υπολογίζεται η αναμενόμενη πιθανότητα κέρδους από την ιστορική απόδοση όπως είναι ο λόγος εξόφλησης, ο οποίος καθορίζεται από τη σχέση

$$R_p = \frac{\bar{W}}{\bar{L}} \tag{3.130}$$

όπου  $\bar{W}$  είναι ο μέσος όρος μόνο των ποσών κέρδους, και ομοίως  $\bar{L}$  είναι το μέγεθος του μέσου όρου απωλειών.

Το ρίσκο του υπολογισμού χρεοκοπίας και η σχετική διάρκεια (σε αριθμό των συναλλαγών που οδηγούν είτε στη καταστροφή είτε στην επιτυχία), παρέχει μια εκτίμηση σχετικά με το εάν οι δεδομένες αρχικές παράμετροι ενός συστήματος συναλλαγών θα οδηγήσουν, μετά από πολλές συναλλαγές, σε μια ευνοϊκή έκβαση. Επιπρόσθετα με τα  $p$  και  $R_p$ , αυτές οι παράμετροι είναι το αρχικό κεφάλαιο  $z$ , το μέγεθος του κεφαλαίου  $a$  στο οποίο ο έμπορος θα εγκαταλείψει συνολικά, το ποσό  $k$  που διακυβεύεται σε κάθε συναλλαγή, υποτιθέμενο σταθερό σε μια δοκιμή, και με κάθε συναλλαγή κλεισίματος είτε με κέρδος  $kR_p$ , είτε με απώλεια ποσού  $k$ .

Το ρίσκο του προβλήματος χρεοκοπίας για λόγο εξόφλησης  $R_p=1$  είναι επίσης γνωστός ως *κλασικό πρόβλημα χρεοκοπίας* και μελετήθηκε από τον Feller. Υποθέτοντας ότι το κόστος είναι αμελητέο, τότε ο τύπος για την πιθανότητα της τελευταίας καταστροφής  $q_z$ , αναμενόμενου κέρδους  $E[G]$ , και διάρκεια παιχνιδιού  $D_z$  δίνονται ως εξής για  $p \neq q$  :



$$q_z = \frac{(q/p)^{a/k} - (q/p)^{z/k}}{(q/p)^{a/k} - 1} \quad (3.131)$$

$$E[G] = a(1 - q_z) - z \quad (3.132)$$

$$.D_z = \frac{1}{k(q-p)} \times \left[ z - a \frac{1 - (q/p)^{z/k}}{1 - (q/p)^{a/k}} \right] \quad (3.133)$$

όπου οι συναλλαγές παύουν είτε με την απώλεια του  $z$  είτε με κέρδος  $a-z$ . Με  $p=q=0,5$  οι τύποι παύουν να ισχύουν και αντικαθίσταται από :

$$q_z = 1 - \frac{z}{a} \quad (3.134)$$

$$E[G] = 0 \quad (3.135)$$

$$.D_z = \frac{z(a-z)}{k^2} \quad (3.136)$$

σημειώστε ότι το  $q_z$  είναι τώρα ανεξάρτητο από την διακύβευση , αλλά η διάρκεια εξαρτάται πάντα από το  $k$ .

Εάν ο λόγος  $a/k$  είναι πολύ μεγάλος(π.χ.  $a \rightarrow \infty$ ) τότε οι τύποι μπορούν να απλοποιηθούν ως κάτωθι (το αναμενόμενο κέρδος δεν έχει νόημα τώρα)

$$\begin{aligned} q_z &= 1 \text{ _if_ } -p \leq q \\ &= \left(\frac{q}{p}\right)^{z/k} \text{ _if_ } -p > q \end{aligned}$$

$$\begin{aligned} .D_z &= \frac{z}{(q-p)k} \text{ _if_ } -p < q \\ &= \infty \text{ _if_ } -p \geq q \end{aligned}$$

Η μέθοδος του Feller για το  $R_p=1$  μπορεί επίσης να επεκταθεί για  $R_p=2$ , όπου μια κερδοφόρα συναλλαγή κερδίζει  $2k$  και μια ζημιογόνα συναλλαγή χάνει ποσό  $k$ . Οι τύποι είναι τώρα, για  $2p \neq q$

$$q_z = \frac{(\lambda)^{a/k} - (\lambda)^{z/k}}{(\lambda)^{a/k} - 1} \quad (3.137)$$

$$E[G] = a(1 - q_z) - z \quad (3.138)$$

$$.D_z = \frac{1}{k(q - 2p)} \times [z - a \frac{(\lambda)^{z/k} - 1}{(\lambda)^{a/k} - 1}] \quad (3.139)$$

όπου

$$\lambda = \sqrt{\frac{1}{4} + \frac{q}{p}} - \frac{1}{2} \quad (3.140)$$

Όταν  $\lambda = 1$  οι παραπάνω εκφράσεις δεν ισχύουν, το οποίο συμβαίνει για  $2p = q$ , ( $p=1/3$  και  $q=2/3$ ), με αποτέλεσμα την ανάγκη αντικατάστασης των τύπων με :

$$q_z = 1 - \frac{z}{a} \quad (3.141)$$

$$E[G] = 0 \quad (3.142)$$

Σημείωση: δεν υπάρχει απλός αναλυτικός τύπος για το  $D_z$  σε αυτήν την περίπτωση. Η περιοριστική περίπτωση  $a/k \rightarrow \infty$  δίνεται από:

$$q_z = 1 \text{ if } 2p \leq q \\ = (\lambda)^{z/k} \text{ if } 2p > q$$

$$.D_z = \frac{z}{(q - 2p)k} \text{ if } 2p < q \\ = \infty \text{ if } 2p \geq q$$

όπου αυστηρά, η διάρκεια δεν έχει νόημα για  $2p \geq q$ , στην επίτευξη ενός άπειρου στόχου.

Η στρατηγική για την βελτίωση της πιθανότητας επιτυχίας, εξαρτάται από τις πιθανότητες  $p$  και  $q$ . Οι πίνακες 3.30 και 3.31 επεξηγούν τα χαρακτηριστικά του προβλήματος χρεοκοπίας με την επιλογή βασικών παραδειγμάτων, για ένα αρχικό κεφάλαιο  $z=5000$ .

Τα χαρακτηριστικά σχήματα που προκύπτουν από τα παραδείγματα στους πίνακες, αποκαλύπτουν ότι όταν  $p > q$  για  $R_p=1$  και  $2p > q$  για  $R_p=2$ , η καλύτερη στρατηγική είναι η διακύβευση μικρών ποσών. Όταν  $p < q$  για  $R_p=1$  και  $2p < q$  για  $R_p=2$ , ο στόχος επιτυγχάνεται καλύτερα από την επένδυση μεγάλων ποσών ανά συναλλαγή, στοχεύοντας για ένα μέτριο κέρδος. Η λόγος εξόφλησης  $R_p=2$  επιτρέπει σε

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

μια επιτυχή στρατηγική για να παιχτεί για τιμές του  $p$ , κάτω έως λίγο επάνω από  $1/3$ , με θετικά αναμενόμενα κέρδη.

Οι τελευταίες καταχωρήσεις σε κάθε πίνακα παρουσιάζουν παραδείγματα όπου ο κίνδυνος χρεοκοπίας είναι κάτω του μέσου όρου, με το αναμενόμενο κέρδος όμως να είναι αρνητικό. Ο λόγος για αυτό είναι ότι, ενώ σε έναν αριθμό προσπαθειών μεγαλύτερου του μέσου όρου επιτυγχάνεται μέτριο κέρδος, όταν εμφανίζεται μια χρεοκοπία αυτή περιλαμβάνει μια πολύ μεγαλύτερη απώλεια, ως εκ τούτου η γενική προσδοκία είναι αρνητική.

**ΠΙΝΑΚΑΣ 3.30:** Επιλεγμένες περιπτώσεις για πρόβλημα χρεοκοπίας με  $R_p=1$  και αρχικό κεφάλαιο  $z=5000$ . Οι συναλλαγές παύουν με απώλεια  $z$  (χρεοκοπία) ή κέρδος  $\alpha-z$  (κεφάλαιο  $\alpha$ )

Πιθανότητες		Στόχος $\alpha$	Διακρίβωση Ποσό $k$	Πιθανότητες		Προσδοκίες	
Κέρδος $p$	Απώλεια $q$			Χρεοκοπία	Επιτυχία	Κέρδος	Διάρκεια
0.55	0.45	$10^6$	100	0.000	1.000	106	$10^5$
			500	0.134	0.866	$8.6 \cdot 10^5$	$1.7 \cdot 10^4$
			1000	0.367	0.633	$6.3 \cdot 10^5$	6284
			2500	0.669	0.331	$3.3 \cdot 10^5$	1302
0.50	0.50	5500	1	0.091	0.909	0	$2.5 \cdot 10^6$
		5500	100	0.091	0.909		250
		$10^4$	1000	0.500	0.500		25
		$10^6$	1000	0.995	0.005		4975
0.45	0.55	$10^4$	100	1.000	0.000	-5000	500.0
			1000	0.732	0.268	-2317	23.2
			2500	-0.599	0.401	-990	4.0
		7500	2500	0.402	0.598	-515	2.1

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

**ΠΙΝΑΚΑΣ 3.31:** Παραδείγματα για πρόβλημα χρεοκοπίας με λόγο εξόφλησης  $R_p=2$  και αρχικό κεφάλαιο  $z=5000$ . Οι συναλλαγές παύουν με απώλεια  $z$  (χρεοκοπία) ή κέρδος  $a-z$  (κεφάλαιο  $a$ )

Πιθανότητες		Στόχος $a$	Διακύβευση Ποσού $k$	Πιθανότητες		Προσδοκίες	
Κέρδος $p$	Απώλεια $q$			Χρεοκοπία	Επιτυχία	Κέρδος	Διάρκεια
0.50	0.50	$10^6$	1000	0.090	0.910	$9.1 \cdot 10^5$	1809.7
0.40	0.60	$10^6$	10	0.000	1.000	$10^5$	$5.0 \cdot 10^5$
			500	0.142	0.858	$8.5 \cdot 10^5$	8526.6
			1000	0.377	0.623	$6.2 \cdot 10^5$	3088.6
			2500	0.677	0.323	$3.210^5$	635.8
1/3	2/3	5500	1	0.091	0.909	0	
		5500	100	0.091	0.909		
		$10^4$	1000	0.500	0.500		
		$10^6$	1000	0.995	0.005		
0.28	0.72	7500	100	0.984	0.016	-4879	305.0
			500	0.614	0.386	-2104	26.3
			1000	0.476	0.524	-1073	6.7
			2500	0.390	0.610	-423	1.1

### 3.5.6.2 Καταμερισμός Κεφαλαίου

Δεδομένου ότι κάθε συναλλαγή περιλαμβάνει έναν κίνδυνο απώλειας, το πρόβλημα για το πώς θα μεγιστοποιήσουμε τα κέρδη, δεδομένου ενός τελικού κεφαλαίου, αφορά τον υπολογισμό του λόγου  $f$  του κεφαλαίου που διακυβεύεται σε κάθε συναλλαγή. Εάν απλοποιήσουμε τις υποθέσεις τότε μπορούμε να διατυπώσουμε έναν αναλυτικό τύπο για το  $f$ , το οποίο είναι γνωστό ως βέλτιστη αξία  $f$ .

Όπως στο κλασικό πρόβλημα χρεοκοπίας, υποθέτουμε ότι όλα τα κέρδη συμπεριλαμβάνονται στο σταθερό ποσό  $kR_p$  και όλες τις απώλειες στο ποσό  $k$ . Το σύστημα του Thorp ή το σύστημα Kelly εκτιμά το λόγο  $f$ , σαν τη διαφορά μεταξύ του πιθανού αναμενόμενου ποσού που θα κερδηθεί,  $pR_p$  και το αναμενόμενο ποσό που θα χαθεί  $(1-p)k$ , διαιρεμένο από το πρώτο, δίνοντας:

$$f = \frac{(1 + R_p)p - 1}{R_p} \quad (3.143)$$

το οποίο είναι ανεξάρτητο από το  $k$ .

Η εμπειρική μέθοδος ταιριάζει στα συστήματα συναλλαγών, δεδομένου ότι  $w$   $f$  μπορεί να είναι συντονισμένο με τα χαρακτηριστικά του ιδιαίτερου συστήματος. Η μέθοδος είναι βασισμένη στη βελτιστοποίηση μια σύνθετης, ή γεωμετρικής σειράς, σταθμισμένων σχετικών επιστροφών,  $R_w$ , που καθορίζεται από:

$$R_{wi} = 1 + f \frac{R_i}{\hat{R}_L} \quad (3.144)$$

όπου  $R_i$  είναι η επιστροφή για την  $i^{th}$  συναλλαγή, και  $IR_L$  είναι η απόλυτη αξία της επιστροφής για τη μέγιστη ζημιογόνα συναλλαγή (η δοκιμαστική περίοδος πρέπει να περιέχει τουλάχιστον ζημιογόνα συναλλαγή για να ισχύει αυτή η μέθοδος). Η γεωμετρική σειρά  $R_{wi}$  για  $n$  προηγούμενες συναλλαγές δίνει ένα σύνθετο κέρδος

$$G_c = \prod_{i=1}^{\pi} R_{wi} \quad (3.145)$$

( $\Pi$  είναι σημείωση για τον πολλαπλασιασμό) το οποίο βελτιστοποιείται έπειτα με αριθμητική επίλυση ως προς το βέλτιστο  $f$ , που μεγιστοποιεί  $G_c$ . Η απλή μέθοδος αναζήτησης μπορεί να χρησιμοποιηθεί δεδομένου ότι το  $f$  βρίσκεται στο διάστημα  $[0, 1]$ .

Ο αριθμός συμβάσεων  $N_c$ , για συναλλαγές σε προϊόντα είναι τώρα ακριβώς θέμα απόφασης σχετικά με ένα μέγιστο επιτρεπόμενο κίνδυνο για τη δεδομένη συναλλαγή, όπως τίθεται με τον προσδιορισμό θέσης του επιπέδου τερματισμού –θεωρώντας αυτό το ποσό  $S$ , τότε:

$$N_c = \frac{fT}{S} \quad (3.146)$$

όπου  $T$  είναι το συνολικό διαθέσιμο κεφάλαιο.

### 3.5.6.3 Διαφοροποίηση

Η θεμελιώδης έννοια της θεωρίας επενδύσεων είναι ότι η μείωση του κινδύνου μπορεί να επιτευχθεί μέσω διαφοροποίησης της επένδυσης τίτλων, μετοχών και προϊόντων, οι μεμονωμένοι παράγοντες κινδύνου των οποίων είναι ανεξάρτητοι μεταξύ τους ή είναι ιδιαίτερα ασυσχέτιστοι. Ο κίνδυνος

## ΜΕΡΟΣ 1ο – ΚΕΦΑΛΑΙΟ 3ο

μετρίεται από τη σταθερή απόκλιση ή από το τετράγωνό της μεταβλητής, των αναμενόμενων επιστροφών.

Ας θεωρήσουμε το παράδειγμα δύο προϊόντων Α και Β, που συνδυάζονται σε ποσότητες  $N_A$  και  $N_B$  μέσα σε ένα επενδυτικό πακέτο P, έτσι ώστε ένα σύνολο  $N=N_A+N_B$  μελλοντικών συμβολαίων να ενεργεί ταυτόχρονα. Τα ποσοστά των Α και Β στο επενδυτικό πακέτο μπορούν να γραφτούν,  $W_A=N_A/N$  και  $W_B=N_B/N$ . Θεωρούμε τις προηγούμενες συναλλαγές κάθε ποσότητας χωριστά: έστω ότι υπάρχουν  $n$  δεδομένα επιστροφής σε περίοδο ενός χρόνου (ή άλλη κατάλληλη περίοδο). Τότε η μεταβλητή των επιστροφών θα δίνεται από τη σχέση

$$V_A = \frac{1}{n-1} \sum_{i=1}^n (R_{Ai} - \bar{R}_A)^2 \quad (3.147)$$

και ομοίως για τη ποσότητα Β, η μεταβλητή του επενδυτικού πακέτου μπορεί να γραφτεί (σημείωση:  $R_p$  είναι η επιστροφή του επενδυτικού πακέτου)

$$V_B = \frac{1}{n-1} \sum_{i=1}^n (R_{Bi} - \bar{R}_B)^2 \quad (3.148)$$

Και χρησιμοποιώντας το  $R_p = W_A R_A + W_B R_B$  and  $\bar{R}_p = W_A \bar{R}_A + W_B \bar{R}_B$ , έχουμε

$$V_p = \frac{1}{n-1} \sum_{i=1}^n W_A (R_{Ai} - \bar{R}_A) + W_B (R_{Bi} - \bar{R}_B)^2 \quad (3.149)$$

$$= W_A^2 S_A^2 + W_B^2 S_B^2 + 2W_A W_B S_A S_B C_{AB} \quad (3.150)$$

όπου η σταθερή απόκλιση των επιστροφών για τα μεμονωμένα προϊόντα είναι  $S_A$  και  $S_B$ , ενώ  $C_{AB}$  είναι ο γραμμικός συντελεστής συσχέτισής τους, καθορισμένος από τη σχέση

$$C_{AB} = \frac{\sum_{i=1}^n (R_{Ai} - \bar{R}_A)(R_{Bi} - \bar{R}_B)}{S_A S_B} \quad (3.151)$$

Ο ρόλος του γραμμικού συντελεστή συσχέτισης στη μείωση της γενικής μεταβολής είναι τώρα σαφής: εάν  $C_{AB}=1$  (υψηλή θετική συσχέτιση) έπειτα η μεταβολή, και ως εκ τούτου ο κίνδυνος, θα ήταν μέγιστος. Εάν  $C_{AB}=0$  (ελάχιστη συσχέτιση, ή μη-συσχέτιση) τότε κάποια μείωση του κινδύνου θα ακολουθούσε. Παραδείγματος χάριν, εάν οι μεμονωμένες σταθερές αποκλίσεις και οι αναλογίες που συναλλάχθηκαν ήταν ίδιες, τότε η σταθερή απόκλιση του επενδυτικού πακέτου θα μειωνόταν ως προς την τετραγωνική ρίζα του αριθμού συμβάσεων που συναλλάχθηκαν για κάθε προϊόν:

$$S_p = \frac{S}{\sqrt{N}} \quad (3.152)$$

Στη περίπτωση υψηλής μη-συσχέτισης  $C_{AB} = -1$  θα είχαμε ιδανική μείωση του κινδύνου και τότε η σταθερή απόκλιση επενδυτικού πακέτου θα μπορούσε να γραφτεί σαν

$$S_p = |W_A S_A - W_B S_B| \quad (3.153)$$

και οι αναλογίες που θα συναλλαγούν θα μπορούσαν έπειτα να υπολογιστούν έτσι ώστε να δίνουν  $S_p = 0$ .

Στη πράξη, συνδυάζοντας προϊόντα που ενέχουν κίνδυνο μπορεί στην καλύτερη περίπτωση να παραγάγει αμοιβαίους συντελεστές συσχέτισης μεταξύ 0 και 0,5 εξαιτίας των παρόμοιων δυνάμεων που επηρεάζουν τέτοια προϊόντα. Εντούτοις αυτή είναι ακόμα καλύτερη στρατηγική από να ανταλλάσσεται μόνο ένα προϊόν που ενέχει κίνδυνο. Οι αγορές υψηλού κινδύνου ευνοούνται συνήθως από τους εμπόρους που ρισκάρουν, αφού η αστάθεια εξισώνεται με τις ευκαιρίες για κέρδη, έτσι μια προσέγγιση επενδυτικού πακέτου παρέχει έναν ορισμένο βαθμό ενσωματωμένης ασφάλειας.

Η βέλτιστη μέθοδος  $f$  που εφαρμόζεται στον υπολογισμό της κυρίας κατανομής για κάθε μεμονωμένο προϊόν, μπορεί επίσης να εφαρμοστεί στη βελτιστοποίηση της αναλογίας των μετοχών σε ένα επενδυτικό πακέτο.

Η αύξηση του αριθμού διαφορετικών προϊόντων σε ένα επενδυτικό πακέτο πάσχει από έναν νόμο που μειώνει το κέρδος. Η μεταβολή του επενδυτικού πακέτου για τα  $M$  διαφορετικά προϊόντα δίνεται από τη σχέση

$$V_p = \sum_{j=1}^M W_j^2 S_j^2 + \sum_{\substack{j=1 \\ j \neq k}}^M \sum_{k=1}^M W_j W_k S_j S_k C_{jk} \quad (3.154)$$

Τώρα όσο το  $M$  αυξάνεται ο δεύτερος όρος τείνει να επικρατήσει, και στο όριο θα είναι ίσος με τη μεταβολή της αγοράς συνολικά, καθιερώνοντας κατά συνέπεια ένα κατώτατο επίπεδο κινδύνου. Η μεταβολή του επενδυτικού πακέτου τείνει να πέφτει γρήγορα δεδομένου ότι το χαρτοφυλάκιο είναι σύνθετο, και φεύγει από την τιμή κορυφής με περίπου 10 προϊόντα.

## ***Βιβλιογραφία 1ου Μέρους***

- [1] Jeffrey Jarrett, *Business Forecasting Methods (Second Edition)*, Basil Blackwell, 1991, Vol.2, p.p 24-41, 48-52
- [2] John E. Hanke, Arthur G. Reich, *Business Forecasting (Sixth Edition)*, Prentice Hall, 1998, Vol.5, p.p 136-165
- [3] Jeffrey Jarrett, *Business Forecasting Methods (Second Edition)*, Basil Blackwell, 1991, Vol.3, p.p 93-108, 112-120
- [4] John E. Hanke, Arthur G. Reich, *Business Forecasting (Sixth Edition)*, Prentice Hall, 1998, Vol.8, p.p 303-333
- [5] Edwin Mansfield, *Managerial Economics (Fourth Edition)*, W.W Norton & Company, Vol.6, p.p 179-191
- [6] Β. Ασημακόπουλος, *Μέθοδοι Προβλέψεων*, ΕΜΠ, 1994, Vol.2, 3
- [7] Jeffrey Jarrett, *Business Forecasting Methods (Second Edition)*, Basil Blackwell, 1991, Vol.4, p.p 143-154, 158-170
- [8] John E. Hanke, Arthur G. Reich, *Business Forecasting (Sixth Edition)*, Prentice Hall, 1998, Vol.6, p.p 195-218
- [9] Jeffrey Jarrett, *Business Forecasting Methods (Second Edition)*, Basil Blackwell, 1991, Vol.5, p.p 182-188, 193-198
- [10] John E. Hanke, Arthur G. Reich, *Business Forecasting (Sixth Edition)*, Prentice Hall, 1998, Vol.7, p.p 243-252
- [11] Jeffrey Jarrett, *Business Forecasting Methods (Second Edition)*, Basil Blackwell, 1991, Vol.6, p.p 223-235
- [12] Mark Hirschey, James L. Pappas, *Managerial Economics (Seventh Edition)*, The Drynen Press, Vol.72, p.p 296-309
- [13] Milton H. Spencer, Colin G. Clark, Peter W. Hoquet, *Business and Economic Forecasting, An Econometric Approach*, Richard D. Irwin Inc, Vol.10, p.p 361-381
- [14] Β. Ασημακόπουλος, *Μέθοδοι Προβλέψεων*, ΕΜΠ, 1994, Vol.4, 5, 6
- [15] Jeffrey Jarrett, *Business Forecasting Methods (Second Edition)*, Basil Blackwell, 1991, Vol.7, p.p 254-272
- [16] Jeffrey Jarrett, *Business Forecasting Methods (Second Edition)*, Basil Blackwell, 1991, Vol.8, p.p 289-307
- [17] Β. Ασημακόπουλος, *Μέθοδοι Προβλέψεων*, ΕΜΠ, 1994, Vol.7
- [18] Jeffrey Jarrett, *Business Forecasting Methods (Second Edition)*, Basil Blackwell, 1991, Vol.9, p.p 319-337



- [19] John E. Hanke, Arthur G. Reich, *Business Forecasting (Sixth Edition)*, Prentice Hall, 1998, Vol.10, p.p 407-454
- [20] Fuzzy logic-based forecasting model, Tapio Frantti,, Petri Mahonen, *Engineering Applications of Artificial Intelligence* 14 (2001) 189–201, March 2000
- [21] Fuzzy Seasonality Forecasting, Ping-Teng Chang, *Fuzzy Sets and Systems* 90 (1997) 1-10, April 1996
- [22] Handling forecasting problems using fuzzy time series, Jeng-Ren Hwang, Shyi-Ming Chen, Chia-Hoang Lee, *Fuzzy Sets and Systems* 100 (1998) 217 - 228, March 1997
- [23] S.M. Chela, Forecasting enrollments based on fuzzy time series, *Fuzzy Sets and Systems* 81 (1996), 311-319.
- [24] M.Y. Day, Research of applying genetic algorithms to fuzzy forecasting focus on sales forecasting, M.S. Thesis, Tamkang University, Taipei, Taiwan, ROC, 1995.
- [25] J.R. Hwang, S.M. Chen, C.H. Lee, A new method for handling forecasting problems based on fuzzy time series, *Proc. 7th Internat.Conf. on Information Management*, Chungli, Taoyuan, Taiwan, ROC, 1996 pp. 312-321.
- [26] Q. Song, B.S. Chissom, Fuzzy time series and its models, *Fuzzy Sets and Systems* 54 (1993) 269-277.
- [27] Q. Song, B.S. Chissom, Forecasting enrollments with fuzzy time series - part I, *Fuzzy Sets and Systems* 54 (1993) 1-9.
- [28] Q. Song, B.S. Chissom, Forecasting enrollments with fuzzy time series part II, *Fuzzy Sets and Systems* 62 (1994) 1-8.
- [29] J. Sullivan, W.H. Woodall, A comparison of fuzzy forecasting and Markov modeling, *Fuzzy Sets and Systems* 64 (1994) 279-293.
- [30] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning, part I-III, *Inform. Sci.* 8 (1975) 199-249; 8 (1975) 301-357; 9 (1975) 43-80.
- [31] Eitan Michael Azoff, *Neural Networks Time Series Forecasting of Financial Markets*, Willey, Vol 1, p.p 1-5, Vol 2, p.p125-140
- [32] Judgmental adjustment in time series forecasting using neural networks, Jae Kyu Lee, Chang Seon Yum, *Decision Support Systems* 22 (1998) 135-154, 1998
- [33] An intelligent sales forecasting system through integration of artificial neural network and fuzzy neural network, R.J.Kuo, K.C. Xue, *Computers in industry* 37 (1998) 1-15, February 1998.

**ΜΕΡΟΣ ΔΕΥΤΕΡΟ**

**ΚΑΤΑΓΡΑΦΗ ΚΑΙ ΕΚΤΙΜΗΣΗ ΛΟΓΙΣΜΙΚΩΝ ΠΟΥ  
ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΓΙΑ ΤΗΝ ΠΑΡΑΓΩΓΗ  
ΠΡΟΒΛΕΨΕΩΝ**

**ΚΕΦΑΛΑΙΟ 1ο : ΚΑΤΑΓΡΑΦΗ ΤΩΝ ΛΟΓΙΣΜΙΚΩΝ ΚΑΙ ΣΥΛΛΟΓΗ  
ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ**

**1.1 Εισαγωγή**

Στο δεύτερο μέρος της εργασίας επιχειρείται μια ενημέρωση και επικαιροποίηση (update) παλαιότερων δημοσιεύσεων, σχετικές με την εκτίμηση των λογισμικών για προβλέψεις (forecasting software evaluation) ([1], [2], [3], [4], [5]). Οι δημοσιεύσεις αυτές ήταν βασισμένες σε συγκεκριμένα χαρακτηριστικά. Ο τρόπος με τον οποίο αρχικά συλλέξαμε τα στοιχεία και στη συνέχεια ο τρόπος που τα αξιολογήσαμε περιγράφεται παρακάτω:

**1.2 Επισκόπηση βιβλιογραφίας-δημοσιεύσεων**

Αναζητήθηκαν στις μηχανές αναζήτησης του διαδικτύου ([6],[7]) καθώς και σε σελίδες επιστημονικού περιεχομένου ([8]-[13]), δημοσιεύσεις σχετικές με λογισμικά προβλέψεων και ειδικότερα δημοσιεύσεις που αφορούσαν την εκτίμηση και αξιολόγηση αυτών των λογισμικών. Συγκεντρώθηκαν αρκετές, εκ των οποίων αξιολογήθηκαν και μελετήθηκαν εκείνες που επικεντρώνονταν στην συγκριτική παράθεση λογισμικών βάσει συγκεκριμένων στοιχείων.

**1.3 Καταγραφή λογισμικών προβλέψεων**

Αφού έγινε επισκόπηση των δημοσιεύσεων, αναζητήθηκε εκ νέου στις μηχανές αναζήτησης του διαδικτύου ([6], [7]), κάθε είδους λογισμικό (είτε αυτό αποτελεί στατιστικό ή μαθηματικό εργαλείο ή καθαρά εργαλείο προβλέψεων) που εκτελεί προβλέψεις. Απώτερος στόχος μας ήταν η ενημέρωσή μας για τυχόν νεότερες εκδόσεις παλαιότερων εφαρμογών ή κυκλοφορία νέων προϊόντων. Συγκεντρωθήκαν γενικά στοιχεία και πληροφορίες για 82 λογισμικά πακέτα και εφαρμογές, που σχετίζονται καθ' οιονδήποτε τρόπο με τις διαδικασίες προβλέψεων. Τα λογισμικά αυτά φαίνονται στο πίνακα που ακολουθεί:

**ΠΙΝΑΚΑΣ 1.1: ΚΑΤΑΛΟΓΟΣ ΛΟΓΙΣΜΙΚΩΝ ΠΟΥ ΕΚΤΕΛΟΥΝ ΠΡΟΒΛΕΨΕΙΣ**

A/A	ΛΟΓΙΣΜΙΚΟ	A/A	ΛΟΓΙΣΜΙΚΟ	A/A	ΛΟΓΙΣΜΙΚΟ
1	Modler	30	SCA Forecasting & Modeling Package	59	Aremos
2	Modler Blue	31	B34S Econometric Package	60	BMDP
3	ABStat	32	Shazam	61	Prophix Enterprize
4	WEBSURV	33	Smart Forecasts for Windows	62	G@RCH
5	GAUSS Applications	34	SPSS for Windows	63	STAMP
6	Autobox	35	SPSS Trends for Windows	64	Financial Genome
7	Forecast Pro XE	36	Statistica for Windows	65	Neuroshell Predictor
8	Forecast Pro Standard Edition	37	WINKS Professional	66	Alea Forecast
9	Forecast Pro Unlimited Edition	38	TSP	67	MKtSIM
10	WinX11	39	SAP SCM	68	REMI Policy Insight
11	PeerPlanner	40	Microsoft Forecaster	69	REMI Consulting
12	Peerforecaster	41	JD Edwards EnterpriseOne Demand Forecasting	70	Turbo Spring Stat
13	LIMDEP	42	PYTHIA	71	CB Predictor
14	Nlogit	43	@RISK	72	Data Desk
15	RATS	44	4CAST/2	73	Data Plore
16	EXPO	45	Sibyl/Runner	74	DELPHY
17	EXPO/Econometrics Add-in Module	46	ySTAT	75	Decision Time
18	Statgraphics Centurion standard	47	tsMetrix	76	DSTS
19	Statgraphics Centurion professional	48	Soritec	77	GiveWin
20	S-PLUS 7 for Windows	49	PcGive Professional	78	ITSM Professional
21	MATLAB	50	PcGets	79	X-12 ARIMA
22	MATLAB Financial Toolbox	51	Comtrack	80	SsfPack
23	MATLAB Statistics Toolbox	52	mathStatica	81	UNISTAT
24	MATLAB Systems Identification Toolbox	53	Microfit	82	RETINA
25	MATLAB Neural Net Toolbox	54	LISREL		
26	Minitab	55	TKSolver		
27	NCSS 2004	56	OxGauss		
28	Eviews	57	Stata		
29	SAS Software	58	Amos		

**1. 4 Αρχική αξιολόγηση των λογισμικών**

Από το σύνολο των εφαρμογών που συγκεντρώθηκαν επιλέχθηκαν 42 λογισμικά πακέτα, για τα οποία υπήρχε η δυνατότητα συλλογής μεγάλου αριθμού στοιχείων από το διαδίκτυο. Επισημαίνεται, ότι πολλές εταιρείες δεν παρέχουν πληροφορίες για τα χαρακτηριστικά των προϊόντων τους, με αποτέλεσμα να καθίσταται αδύνατη η αξιολόγηση τους. Τα 42 επιλεγθέντα λογισμικά φαίνονται στο πίνακα που ακολουθεί:

**ΠΙΝΑΚΑΣ 1. 2: ΚΑΤΑΛΟΓΟΣ ΕΠΙΛΕΓΜΕΝΩΝ ΛΟΓΙΣΜΙΚΩΝ**

<b>A/A</b>	<b>ΛΟΓΙΣΜΙΚΟ</b>	<b>A/A</b>	<b>ΛΟΓΙΣΜΙΚΟ</b>	<b>A/A</b>	<b>ΛΟΓΙΣΜΙΚΟ</b>
1	Modler	15	RATS	29	SAS Software
2	Modler Blue	16	EXPO	30	SCA Forecasting & Modeling Package
3	ABStat	17	EXPO/Econometrics Add-in Module	31	B34S Econometric Package
4	WEBSURV	18	Statgraphics Centurion standard	32	Shazam
5	GAUSS Applications	19	Statgraphics Centurion professional	33	Smart Forecasts for Windows
6	Autobox	20	S-PLUS 7 for Windows	34	SPSS for Windows
7	Forecast Pro XE	21	MATLAB	35	SPSS Trends for Windows
8	Forecast Pro Standard Edition	22	MATLAB Financial Toolbox	36	Statistica for Windows
9	Forecast Pro Unlimited Edition	23	MATLAB Statistics Toolbox	37	WINKS Professional
10	WinX11	24	MATLAB Systems Identification Toolbox	38	TSP
11	PeerPlanner	25	MATLAB Neural Net Toolbox	39	SAP SCM
12	Peerforecaster	26	Minitab	40	Microsoft Forecaster
13	LIMDEP	27	NCSS 2004	41	JD Edwards EnterpriseOne Demand Forecasting
14	Nlogit	28	Eviews	42	PYTHIA

### ΚΕΦΑΛΑΙΟ 2ο: ΤΑΞΙΝΟΜΗΣΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ

#### 2.1 Εισαγωγή

Σε συνεργασία με τον επιβλέποντα την εργασία και το προσωπικό του εργαστηρίου Μονάδας Συστημάτων Πρόβλεψης και Προοπτικής καταρτίστηκε λίστα με συγκεκριμένα χαρακτηριστικά, προκειμένου να αποτελέσουν τη κοινή βάση αξιολόγησης των λογισμικών που συγκεντρώθηκαν. Η λίστα αυτή συντάχθηκε κατά τέτοιο τρόπο ώστε να επιτρέψει, αφενός μεν την ανάδειξη των επικρατέστερων λογισμικών (σύμφωνα πάντα με τις απαιτήσεις που καθορίστηκαν από το εργαστήριο) αφετέρου δε την τελική αξιολόγηση και σύγκριση αυτών (των επικρατέστερων λογισμικών που θα προκύψουν από την αναζήτηση), με τα χαρακτηριστικά του λογισμικού ΡΥΤΗΙΑ (λογισμικό που δημιουργήθηκε από τη Μονάδα Συστημάτων Πρόβλεψης και Προοπτικής).

#### 2.2 Αναζήτηση και συλλογή ειδικών χαρακτηριστικών

Η συγκέντρωση και καταγραφή των χαρακτηριστικών έγινε με τους παρακάτω τρόπους:

1. Κατόπιν ευρύτερης αναζήτησεως στο διαδίκτυο ([6], [7], [8]).
2. Κατόπιν αναζήτησεως στις ιστοσελίδες των εταιρειών ([6]-[36]).
3. Κατόπιν επισκέψεως στα γραφεία αντιπροσωπειών των εταιρειών στην Ελλάδα και αναζήτησεως των χαρακτηριστικών στα τεχνικά εγχειρίδια (Technical Manual) των λογισμικών.
4. Με αποστολή της επιστολής του Παραρτήματος “Α”, επισυνάπτοντας τη λίστα των επιλεγμένων χαρακτηριστικών του Παραρτήματος “Β”, μέσω ηλεκτρονικού ταχυδρομείου στα τμήματα τεχνικής υποστήριξης και πωλήσεων των 26 εταιρειών που παράγουν τα 42 λογισμικά πακέτα. Στην επιστολή που εστάλη δεν ανταποκρίθηκαν όλες οι εταιρείες.

#### 2.3 Κατάρτιση Συγκριτικών Πινάκων

Με βάση όσα από τα χαρακτηριστικά καταφέραμε να συγκεντρώσουμε καταρτίστηκαν πίνακες, οι οποίοι μπορούν να ταξινομηθούν σε τέσσερις κύριες κατηγορίες :

1. Πληροφορίες για την εταιρεία παραγωγής του λογισμικού.
2. Απαιτήσεις σε υλικό/λογισμικό (Hardawre/Software).
3. Μέθοδοι προβλέψεων που χρησιμοποιεί κάθε λογισμικό.
4. Διάφορες δυνατότητες των λογισμικών.

Η μορφή των πινάκων βασίστηκε στους αντίστοιχους πίνακες σχετικής παλαιότερης δημοσίευσης ([1]). Επίσης, οι πίνακες που αφορούν τις χρησιμοποιούμενες τεχνικές προβλέψεων κατηγοριοποιήθηκαν με βάση την ταξινόμησή τους στο 1<sup>ο</sup> μέρος της εργασίας, σε παραδοσιακές (μοντέλα κινητών μέσων όρων, εκθετική εξομάλυνση, απλή και πολλαπλή παλινδρόμηση, αποσύνθεση) και προηγμένες (Box-Jenkins-ARIMA, οικονομετρικά μοντέλα και τεχνητά νευρωνικά δίκτυα).

## ΜΕΡΟΣ 2ο – ΚΕΦΑΛΑΙΟ 2ο

### 2.3.1 Στοιχεία εταιρειών λογισμικού(Company Information)

Οι πρόσφατες εκδόσεις των λογισμικών που επιλέχθηκαν, καθώς και γενικές πληροφορίες για τις εταιρείες παραγωγής φαίνονται στον παρακάτω πίνακα:

**ΠΙΝΑΚΑΣ 2.1: ΣΤΟΙΧΕΙΑ ΕΤΑΙΡΕΙΩΝ ΛΟΓΙΣΜΙΚΟΥ (COMPANY INFORMATION)**

A/A	ΛΟΓΙΣΜΙΚΟ	ΕΚΔΟΣΗ	ΕΤΑΙΡΕΙΑ
1	Modler	10.6	Alphametrics Limited
2	Modler Blue	10.6	Heath House Princes Mews Royston, Herts SG8 9RT United Kingdom Tel: +44 (0)1763 242 277 Fax: +44 (0)1763 243 988 Sales: EuroSales@modler.com Support: EuroSupport@modler.com Sales: EuroSales@modler.com <a href="http://www.modler.com">http://www.modler.com</a>
3	ABStat	1.93	Anderson Bell Corp
4	WEBSURV		PO Box 745160 Arvada, CO 80006 USA ph: 800-745-9751 fax: 303-940-0595 e-mail: abell@andersonbell.com <a href="http://www.andersonbell.com">http://www.andersonbell.com</a> .
5	GAUSS Applications		Aptech Systems, Inc. PO Box 250 Black Diamond, WA 98010 USA Phone: (425) 432-7855 Fax: (425) 432-7832 Email: info@Aptech.com URL: www.Aptech.com  Φορμίωνος 119-121, Αθήνα TK 16 121 2107600955 Email info@mpassociates.gr <a href="http://www.mpassociates.gr">www.mpassociates.gr</a>
6	Autobox	5.00	Automatic Forecasting Systems PO Box 563 Hatboro, PA 19040 USA ph: 215-675-0652 fax: 215-672-2534 e-mail: sales@autobox.com techsupport@autobox.com <a href="http://www.autobox.com">http://www.autobox.com</a>
7	Forecast Pro XE	4.3.1	Business Forecast Systems, Inc.
8	Forecast Pro Standard Edition	4.3.1	68 Leonard St. Belmont, MA 02178 USA ph: 617-484-5050 fax: 617-484-9219 e-mail: sales@forecastpro.com <a href="http://www.forecastpro.com">http://www.forecastpro.com</a>
9	Forecast Pro Unlimited Edition	4.4	
10	WinX11		Delphus, Inc.
11	PeerPlanner		103 Washington St., Suite 348 Morristown, NJ 07960 USA ph: 973-267-9269 fax: 973-285-9248 e-mail: info@delphus.com hlevelbach@delphus.com <a href="http://www.delphus.com">http://www.delphus.com</a>
12	Peerforecaster		
13	LIMDEP	7.00	Econometric Software, Inc. 15 Gloria Place Plainview, NY 11803 USA ph: 516-938-5254 fax: 516-938-2441 e-mail: EconSoft@worldnet.aft.net <a href="http://www.limdep.com">http://www.limdep.com</a>
14	Nlogit	2.00	
15	RATS	6.2	Estima 1800 Sherman Ave., #301 Evanston, IL 60201 USA ph: 847-864-8772 fax: 847-864-6221 e-mail: estima@estima.com <a href="http://www.estima.com">http://www.estima.com</a>
16	EXPO	3.20	Leading Market Technologies One Kendall Square Cambridge, MA 02139 USA ph: 617-494-4747 fax: 617-494-4788 e-mail: LMT@WORLD.STD.COM <a href="http://www.lmt-expo.com">http://www.lmt-expo.com</a>
17	EXPO/Econometrics Add-in Module		

**ΜΕΡΟΣ 2ο – ΚΕΦΑΛΑΙΟ 2ο**

A/A	ΛΟΓΙΣΜΙΚΟ	ΕΚΔΟΣΗ	ΕΤΑΙΡΕΙΑ
18	Statgraphics Centurion standard	XV	StatPoint, Inc. 2325 Dulles Corner Boulevard, Suite 500 Herndon, Virginia 20171 Phone: 1-800-232-7828 or 1-540-364-0420 Fax: 1-540-364-0421 E-mail: info@statgraphics.com Web site: www.statgraphics.com
19	Statgraphics Centurion professional	XV	Φορμίωνος 119-121, Αθήνα TK 16121 2107600955 Email info@mpassociates.gr www.mpassociates.gr
20	S-PLUS 7 for Windows		Insightful UK 5th Floor Network House Basing View Basingstoke, Hampshire RG21 4HG Tel: +44 (0) 1256 339800 Fax: +44 (0) 1256 339839 mail:shelp@insightful.com info.uk@insightful.com www.insightful.com
21	MATLAB	7.2	The MathWorks, Inc.
22	MATLAB Financial Toolbox	3.0	24 Prime Park Way Natick, MA 01760-1500 USA ph: 508-647-7000 fax: 508-647-7013
23	MATLAB Statistics Toolbox	5.2	
24	MATLAB Systems Identification Toolbox	6.1.3	e-mail: finance@mathworks.com info@mathworks.com
25	MATLAB Neural Net Toolbox	5.0	http://www.mathworks.com/finprod
26	Minitab	14.2	Minitab Inc.3081 Enterprise Dr. State College, PA 16801 USA ph: 800-488-3555 fax: 814-238-4383 e-mail: sales@minitab.com http://www.minitab.com  Φορμίωνος 119-121, Αθήνα TK 16121 2107600955 Email info@mpassociates.gr www.mpassociates.gr
27	NCSS 2004		NCSS 329 N. 1000E. Kaysville, UT 84037 USA ph: 801-546-0445 fax: 801-546-3907 e-mail: sales@ncss.com http://www.ncss.com
28	Eviews	5.1	4521 Campus Drive, #336 Irvine, CA 92612-2621 USA (hours) M-F, 8 a.m. - 5 p.m. (Pacific Time) (ph) 949-856-3368 (fax) 949-856-2044 e-mail: sales@eviews.com http://www.eviews.com  Φορμίωνος 119-121, Αθήνα TK 16121 2107600955 Email info@mpassociates.gr www.mpassociates.gr



**ΜΕΡΟΣ 2ο – ΚΕΦΑΛΑΙΟ 2ο**

<b>A/A</b>	<b>ΛΟΓΙΣΜΙΚΟ</b>	<b>ΕΚΔΟΣΗ</b>	<b>ΕΤΑΙΡΕΙΑ</b>
29	SAS Software	6.12	SAS Institute, Inc.  SAS Campus Drive Cary, NC 27513 USA ph: 919-677-8000 fax: 919-677-4444 e-mail: software@sas.com <a href="http://www.sas.com">http://www.sas.com</a>
30	SCA Forecasting & Modeling Package	7.2	Scientific Computing Associates Corp. 913 W. Van Buren St. Suite 3H Chicago, IL 60607-3528 USA ph: 312-455-0222 fax: 312-455-1652 e-mail: sca@scausa.com support@scausa.com <a href="http://www.scausa.com">http://www.scausa.com</a> Bill Lattyak Scientific Computing Associates Corp. N. Harlem Avenue, Suite F River Forest, Illinois 60305-1262 U.S.A.  Telephone: 708-771-4567 Facsimile: 708-366-0849 e-mail: blattyak@scausa.com
31	B34S Econometric Package		
32	Shazam	10.00	Shazam SHAZAM Order Information info@econometrics.com Fax: 1-707-317-5364 (Best method is Email)
33	Smart Forecasts for Windows	7.00	Smart Software, Inc. 4 Hill Road Belmont, MA 02178 USA ph: 800-762-7899 fax: 617-489-2748 email: info@smartcorp.com <a href="http://www.smartcorp.com">http://www.smartcorp.com</a>
34	SPSS for Windows	14.0	SPSS Inc. 233 S. Wacker Drive 11th Floor Chicago, IL 60606 Tel: 1.312.651.3000 8:00 a.m. – 5:00 p.m. CST e-mail: sales@spss.com <a href="http://www.spss.com">http://www.spss.com</a>
35	SPSS Trends for Windows	14.0	
36	Statistica for Windows	5.10	StatSoft, Inc. 2300 E. 14th St. Tulsa, OK 74104 USA ph: 918-749-1119 e-mail: info@statsoft.com <a href="http://www.statsoft.com">http://www.statsoft.com</a>  Φορμίωνος 119-121, Αθήνα TK 16121 2107600955 Email info@mpassociates.gr <a href="http://www.mpassociates.gr">www.mpassociates.gr</a>
37	WINKS Professional	4.8	TexaSoft PO Box 1169 Cedar Hill, TX 75106 USA ph: 972-291-2115 fax: 972-291-3400 e-mail: info@texasoft.com <a href="http://www.texasoft.com">http://www.texasoft.com</a>

**ΜΕΡΟΣ 2ο – ΚΕΦΑΛΑΙΟ 2ο**

A/A	ΛΟΓΙΣΜΙΚΟ	ΕΚΔΟΣΗ	ΕΤΑΙΡΕΙΑ
38	TSP	5.0	TSP International PO Box 61015 Palo Alto, CA 94306 USA ph: 650-326-1927 fax: 650-328-4163 e-mail: <a href="mailto:sales@tspintl.com">sales@tspintl.com</a> <a href="http://www.tspintl.com">http://www.tspintl.com</a>
39	SAP SCM		<b>SAP America Inc.</b> Strategic Planning & Support Office 3999 West Chester Pike Newtown Square, PA 19073 USA Phone: +1-610-661-1000 U.S. Products and Services Phone: +1-800-872-1727 SAP Business One Response Center Phone: +1-888-227-1727 <a href="http://www.sap.com">http://www.sap.com</a>  <b>SAP Hellas S.A.</b> 20 Ellinidon str. P. Faliro 17564 Athens, Greece Phone: +30/210/947 38 00 ,Fax: +30/210/947 39 00
40	Microsoft Forecaster	6.7	FRx Software Corporation - a Microsoft subsidiary Corporate Headquarters address and phone: 4700 South Syracuse Parkway, Suite 700 Denver, Colorado 80237 +1 303-741-8000- Main 800-379-8733- Toll-free Sales Division- covering Europe, Middle East and Africa address and phone: FRx International Microsoft House 10 Great Pulteney Street London W1R 3DG +44 (118) 909 3759 <a href="http://www.frxsoftware.com">http://www.frxsoftware.com</a> Training and Consulting <a href="mailto:frxtrain@microsoft.com">frxtrain@microsoft.com</a> Technical Support-FRx <a href="mailto:frxhelp@microsoft.com">frxhelp@microsoft.com</a> FRx Software Product Sales <a href="mailto:frxsales@microsoft.com">frxsales@microsoft.com</a> FRx Software product information <a href="mailto:frxinfo@microsoft.com">frxinfo@microsoft.com</a> Company Web site <a href="mailto:frxwbmst@microsoft.com">frxwbmst@microsoft.com</a> FRx Software Documentation <a href="mailto:frxdoco@microsoft.com">frxdoco@microsoft.com</a>
41	JD Edwards EnterpriseOne Demand Forecasting		265, Mesogion Ave. 154 51 Neo Psychiko Athens  Tel.: +30 210-67 89 200 Fax: +30 210-67 75 500 E-mail: <a href="mailto:mailto:%20mailbox_gr@oracle.com">mailto:%20mailbox_gr@oracle.com</a> <a href="http://www.oracle.com">www.oracle.com</a>
42	PYTHIA		National Technical University of Athens <u>Department of Electrical and Computer Engineering</u> Electrical Power Division, Forecasting Systems unit 9, Iroon Polytechniou Str. 15773 Zografou Athens, Greece Tel.: +3210 7723637, 7723738 - Fax: +3210 7723740 <a href="http://fsu.ece.ntua.gr">http://fsu.ece.ntua.gr</a>

## ΜΕΡΟΣ 2ο – ΚΕΦΑΛΑΙΟ 2ο

### 2.3.2 Απαιτήσεις σε Hardware/Software (Hardware/Software Requirements)

Οι απαιτήσεις των εφαρμογών σε Υλικό/Λογισμικό (Hardware/Software Requirements) φαίνονται στον πίνακα που ακολουθεί:

**ΠΙΝΑΚΑΣ 2.2: ΑΠΑΙΤΗΣΕΙΣ ΣΕ ΥΛΙΚΟ/ΛΟΓΙΣΜΙΚΟ (HARDWARE/SOFTWARE REQUIREMENTS)**

Α/Α	Λογισμικό	Απαιτήσεις Υλικού		Άλλα Λειτουργικά Συστήματα				
		Ελάχιστη απαιτούμενη RAM	Χώρος Σκληρού Δίσκου για Εγκατάσταση	DOS	Windows	OS/2	UNIX	MacInt och
1	Modler	2 MB	3 MB	Y	Y		Y	
2	Modler Blue	2 MB	3 MB	Y	Y		Y	
3	ABStat	4 MB	1.5 MB	Y	Y			
4	WEBSURV	16 MB	4 MB	Y				
5	GAUSS Applications	4 MB	10 MB	Y	Y	Y	Y	
6	Autobox	512 MB	2 MB	Y	Y			
7	Forecast Pro XE	16 MB	15 MB		Y			
8	Forecast Pro Standard Edition	16 MB	15 MB	Y	Y			
9	Forecast Pro Unlimited Edition	*	*	*	*	*	*	*
10	WinX11	256 MB	1 MB		Y			
11	Peer Planner for Windows	256 MB	1 MB		Y			
12	PeerForecaster	256 MB	1 MB		Y			
13	LIMDEP for Windows	16 MB	6 MB	Y	Y		Y	
14	Nlogit	16 MB	6 MB		Y			
15	RATS	256 (depends largely on the size of data sets being used)	90Mb (including all examples, etc.)		Y		Y	Y
16	EXPO	32 MB	20 MB		Y		Y	
17	EXPO/Econometrics Add-in Module	16 MB	20 MB		Y		Y	
18	Statgraphics Centurion	64 MB	65 MB	Y	Y			
19	Statgraphics Centurion Professional	*	*		Y			
20	S-PLUS 7 for Windows	96 MB	125 MB	Y	Y		Y	
21	MATLAB	460MB	512 MB				Y	Y
22	MATLAB Financial Toolbox	460MB	512 MB				Y	Y
23	MATLAB Statistics Toolbox	460MB	512 MB				Y	Y
24	MATLAB Systems Identification Toolbox	460MB	512 MB				Y	Y
25	MATLAB Neural Net Toolbox	460MB	512 MB				Y	Y
26	Minitab	64 MB	85 MB	Y	Y		Y	Y

**ΜΕΡΟΣ 2ο – ΚΕΦΑΛΑΙΟ 2ο**

Α/Α	Λογισμικό	Απαιτήσεις Υλικού		Άλλα Λειτουργικά Συστήματα				
		Ελάχιστη απαιτούμενη RAM	Χώρος Σκληρού Δίσκου για Εγκατάσταση	DOS	Windows	OS/2	UNIX	MacIntoch
27	NCSS 2004	32 MB	25 MB	Y	Y			
28	EViews	64 MB	60.5 MB		Y			
29	SAS Software	32 MB	110 MB	Y	Y	Y	Y	Y
30	SCA Forecasting & Modeling Package	32 MB	10 MB	Y	Y		Y	
31	B34S Econometric Package	32 MB	10 MB	Y	Y		Y	
32	Shazam	8 MB	4 MB		Y			
33	Smart Forecasts for Windows	12 MB	20 MB		Y			
34	SPSS for Windows	16 MB	55 MB		Y		Y	Y
35	SPSS Trends	16 MB	55 MB		Y		Y	Y
36	Statistica for Windows	8 MB	30 MB	Y	Y			Y
37	WINKS Professional	16 MB	4 MB	Y	Y			
38	TSP	4 MB	2 MB	Y	Y		Y	Y
39	SAP SCM	*	*		Y			
40	Microsoft Forecaster	128 MB	100 MB		Y			
41	JD Edwards EnterpriseOne Demand Forecasting	*	*	*	*	*	*	*
42	PYTHIA	128 MB	100 MB		Y			

(Y) Χαρακτηριστικά που πληρεί το λογισμικό

(\*) Στοιχεία που δεν αναφέρονται στις ιστοσελίδες των εταιρειών ή δεν απαντήθηκαν στην επιστολή του Παρτήματος "Α"

**2.3.3 Μέθοδοι προβλέψεων των λογισμικών (Selected Forecasting techniques)**

Σε αυτή τη παράγραφο καταβάλλεται προσπάθεια να καταγραφούν όσο το δυνατόν περισσότερες τεχνικές προβλέψεων χρησιμοποιεί κάθε λογισμικό. Κάτι τέτοιο αποδείχθηκε, κατά τη διεξαγωγή της έρευνας, αρκετά δύσκολο καθώς οι περισσότερες εταιρείες δεν αποκαλύπτουν λεπτομέρειες για τον τρόπο λειτουργίας και τη διαδικασία υπολογισμών του λογισμικού τους.

**ΠΙΝΑΚΑΣ 2.3: ΚΙΝΗΤΟΙ ΜΕΣΟΙ ΟΡΟΙ-ΕΚΘΕΤΙΚΗ ΕΞΟΜΑΛΥΝΣΗ (MOVING AVERAGES-EXPONENTIAL SMOOTHING)**

Α/Α	Λογισμικό	Μοντέλα Μέσου Όρου	Εκθετική Εξομάλυνση				
			Single	Holt's	Winter's	Damped	Other
1	Modler	*	*	*	*	*	*
2	Modler Blue	*	*	*	*	*	*
3	ABStat	*	*	*	*	*	*
4	WEBSURV	*	*	*	*	*	*
5	GAUSS Applications	*	*	*	*	*	*
6	Autobox	Y	Y	Y	Y	Y	
7	Forecast Pro XE	Y	Y	Y	Y		
8	Forecast Pro Standard Edition	Y	Y	Y	Y		
9	Forecast Pro Unlimited Edition	Y	Y	Y	Y		
10	WinX11	Y					
11	PeerPlanner	Y	Y	Y	Y	Y	Additive & Multiplicative Seasonality options
12	Peerforecaster	Y	Y	Y	Y	Y	Additive & Multiplicative Seasonality options Additive & Multiplicative error structure
13	LIMDEP	*	*	*	*	*	*
14	Nlogit	*	*	*	*	*	*
15	RATS		Y	Y	Y		
16	EXPO	Y	Y				
17	EXPO/Econometrics Add-in Module	Y	Y				
18	Statgraphics Centurion standard	Y	Y	Y	Y		Brown's, Quadrantic
19	Statgraphics Centurion Professional	Y	Y	Y	Y		Brown's, Quadrantic
20	S-PLUS 7 for Windows	Y					
21	MATLAB	*	*	*	*	*	*
22	MATLAB Financial Toolbox	Y	Y				
23	MATLAB Statistics Toolbox	*	*	*	*	*	*
24	MATLAB Systems Identification Toolbox	*	*	*	*	*	*

**ΜΕΡΟΣ 2ο – ΚΕΦΑΛΑΙΟ 2ο**

A/A	Λογισμικό	Μοντέλα Μέσου Όρου	Εκθετική Εξομάλυνση				
			Single	Holt's	Winter's	Damped	Other
25	MATLAB Neural Net Toolbox	*	*	*	*	*	*
26	Minitab	Y	Y		Y		Double
27	NCSS 2004		Y	Y	Y		Brown's, Double
28	Eviews	Y	Y	Y	Y		
29	SAS Software	Y	Y	Y	Y		Brown's, Double, triple, Holt two-parameter Support for up to three levels of seasonality for the Holt-Winters method: time-of year, day-of-week, or time-of day
30	SCA Forecasting & Modeling Package	Y	Y	Y	Y		Brown's Simple, Double, Holt's, Winters' additive, Winters' Multiplicative, Seasonal indicator, and Harmonic smoothing method
31	B34S Econometric Package						Simple, Double, Holt's, Winters' additive, Winters' Multiplicative, Seasonal indicator, and Harmonic smoothing methods
32	Shazam	Y					
33	Smart Forecasts for Windows	Y	Y		Y		Brown's
34	SPSS for Windows						
35	SPSS Trends for Windows		Y	Y	Y		Brown's
36	Statistica for Windows		Y	Y	Y		Brown's , Triple
37	WINKS Professional	*	*	*	*	*	*
38	TSP	Y					
39	SAP SCM	*	*	*	*	*	*
40	Microsoft Forecaster	*	*	*	*	*	*
41	JD Edwards EnterpriseOne Demand Forecasting		Y				Croston's method
42	PYTHIA		Y	Y	Y	Y	

(Y) Χαρακτηριστικά που πληρεί το λογισμικό

(\*) Στοιχεία που δεν αναφέρονται στις ιστοσελίδες των εταιρειών ή δεν απαντήθηκαν στην επιστολή του Παρτήματος "Α"

## ΜΕΡΟΣ 2ο – ΚΕΦΑΛΑΙΟ 2ο

Οι δυνατότητες του κάθε λογισμικού στη χρησιμοποίηση των μεθόδων αποσύνθεσης φαίνονται στο παρακάτω πίνακα:

**ΠΙΝΑΚΑΣ 2.4: ΜΕΘΟΔΟΙ ΑΠΟΣΥΝΘΕΣΗΣ (DECOMPOSITION METHODS)**

Α/Α	Λογισμικό	Μέθοδοι Αποσύνθεσης		
		Classical	Census X-11	Other
1	Modler	Y		
2	Modler Blue	Y		
3	ABStat	*	*	*
4	WEBSURV	*	*	*
5	GAUSS Applications	*	*	*
6	Autobox	Y		
7	Forecast Pro XE		Y	
8	Forecast Pro Standard Edition	*	*	*
9	Forecast Pro Unlimited Edition		Y	
10	WinX11		Y	
11	PeerPlanner	Y		
12	Peerforecaster	Y		
13	LIMDEP	*	*	*
14	Nlogit	*	*	*
15	RATS	Y	Y	Y
16	EXPO	*	*	*
17	EXPO/Econometrics Add-in Module	*	*	*
18	Statgraphics Centurion standard	Y		
19	Statgraphics Centurion professional	Y		
20	S-PLUS 7 for Windows	Y		Seasonal decompositions
21	MATLAB	*	*	*
22	MATLAB Financial Toolbox	*	*	*
23	MATLAB Statistics Toolbox	*	*	*
24	MATLAB Systems Identification Toolbox	*	*	*
25	MATLAB Neural Net Toolbox	*	*	*
26	Minitab	Y		Trend Analysis
27	NCSS 2004	Y	Y	Seasonal Analysis Trend Analysis Cycle Analysis
28	Eviews		Y	

Α/Α	Λογισμικό	Μέθοδοι Αποσύνθεσης		
		Classical	Census X-11	Other
29	SAS Software	Y	Y	Decomposes raw and transformed series variables and displays the seasonally adjusted series, the trend-cycle component, the seasonal component, or the irregular component
30	SCA Forecasting & Modeling Package	*	*	*
31	B34S Econometric Package	*	*	*
32	Shazam	*	*	*
33	Smart Forecasts for Windows	Y		
34	SPSS for Windows	*	*	*
35	SPSS Trends for Windows		Y	
36	Statistica for Windows	Y	Y	
37	WINKS Professional			
38	TSP	Y		
39	SAP SCM	*	*	*
40	Microsoft Forecaster	*	*	*
41	JD Edwards EnterpriseOne Demand Forecasting	*	*	*
42	PYTHIA	Y		

(Y) Χαρακτηριστικά που πληρεί το λογισμικό

(\*) Στοιχεία που δεν αναφέρονται στις ιστοσελίδες των εταιρειών ή δεν απαντήθηκαν στην επιστολή του Παρτήματος “Α”



## ΜΕΡΟΣ 2ο – ΚΕΦΑΛΑΙΟ 2ο

Οι δυνατότητες του κάθε λογισμικού στη χρησιμοποίηση των προηγμένων μεθόδων πρόβλεψης φαίνονται στο παρακάτω πίνακα:

**ΠΙΝΑΚΑΣ 2.5: ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ-ΟΙΚΟΝΟΜΕΤΡΙΚΑ ΜΟΝΤΕΛΑ (NEURAL NETWORKS-ECONOMETRIC MODELS)**

Α/Α	Λογισμικό	Τεχνητά Νευρωνικά Δίκτυα	Box-Jenkins (ARIMA) (number of integrations)	Οικονομετρικά Μοντέλα	
				Single Equation	Multi-Equation
1	Modler			Y	Y
2	Modler Blue			Y	Y
3	ABStat	*	*	*	*
4	WEBSURV	*	*	*	*
5	GAUSS Applications			Y	Y
6	Autobox		Y	Y	
7	Forecast Pro XE			Y	
8	Forecast Pro Standard Edition			Y	
9	Forecast Pro Unlimited Edition			Y	
10	WinX11		Y		
11	PeerPlanner	*	*	*	*
12	Peerforecaster		Y		
13	LIMDEP			Y	Y
14	Nlogit	*	*	*	*
15	RATS	Y	Y (any)	Y	Y
16	EXPO	*	*	*	*
17	EXPO/Econometrics Add-in Module		Y	Y	
18	Statgraphics Centurion standard		Y		
19	Statgraphics Centurion professional		Y		
20	S-PLUS 7 for Windows		Y		
21	MATLAB	*	*	*	*
22	MATLAB Financial Toolbox	*	*	*	*
23	MATLAB Statistics Toolbox	*	*	*	*
24	MATLAB Systems Identification Toolbox	Y	Y		
25	MATLAB Neural Net Toolbox	Y			
26	Minitab		Y		
27	NCSS 2004		Y		
28	Eviews			Y	Y
29	SAS Software	Y	Y	Y	Y
30	SCA Forecasting & Modeling Package		Y	Y	Y

Α/Α	Λογισμικό	Τεχνητά Νευρωνικά Δίκτυα	Box-Jenkins (ARIMA) ( number of integrations)	Οικονομικά Μοντέλα	
				Single Equation	Multi- Equation
31	B34S Econometric Package		Y	Y	Y
32	Shazam		Y		
33	Smart Forecasts for Windows			Y	
34	SPSS for Windows				
35	SPSS Trends for Windows	Y		Y	Y
36	Statistica for Windows	Y			
37	WINKS Professional				
38	TSP			Y	Y
39	SAP SCM	*	*	*	*
40	Microsoft Forecaster	*	*	*	*
41	JD Edwards EnterpriseOne Demand Forecasting		Y		
42	PYTHIA	*	*	*	*

(Y) Χαρακτηριστικά που πληρεί το λογισμικό

(\* ) Στοιχεία που δεν αναφέρονται στις ιστοσελίδες των εταιρειών ή δεν απαντήθηκαν στην επιστολή του Παρτήματος “Α”

## ΜΕΡΟΣ 2ο – ΚΕΦΑΛΑΙΟ 2ο

Οι δυνατότητες του κάθε λογισμικού στη χρησιμοποίηση των μεθόδων παλινδρόμησης φαίνονται στο παρακάτω πίνακα:

**ΠΙΝΑΚΑΣ 2.6 : ΜΕΘΟΔΟΙ ΠΑΛΙΝΔΡΟΜΗΣΗΣ (REGRESSION METHODS)**

Α/Α	Λογισμικό	Μέθοδοι Παλινδρόμησης		
		Απλή	Πολλαπλή	Άλλη
1	Modler			Ridge Adaptive
2	Modler Blue			Ridge
3	ABStat	*	*	*
4	WEBSURV	*	*	*
5	GAUSS Applications	*	*	*
6	Autobox	Y	Y	Y
7	Forecast Pro XE	*	*	*
8	Forecast Pro Standard Edition	*	*	*
9	Forecast Pro Unlimited Edition	*	*	*
10	WinX11	Y		
11	PeerPlanner	*	*	*
12	Peerforecaster	*	*	*
13	LIMDEP		Y	Ridge Poisson Truncated Switching Stepwise Unobservable
14	Nlogit	*	*	*
15	RATS	Y	Y	Y Truncated Stepwise
16	EXPO	*	*	*
17	EXPO/Econometrics Add-in Module		Y	Ridge Adaptive Stepwise
18	Statgraphics Centurion standard		Y	Ridge Adaptive Poisson Switching Stepwise
19	Statgraphics Centurion professional		Y	Ridge Adaptive Poisson Switching Stepwise
20	S-PLUS 7 for Windows	Y	Y	Robust model, Loess smooth model Basic linear regression Polynomial regression Parametric spline models Constrained regression Logistic regression Generalized linear models Nonlinear regression

Α/Α	Λογισμικό	Μέθοδοι Παλινδρόμησης		
		Απλή	Πολλαπλή	Άλλη
21	MATLAB	*	*	*
22	MATLAB Financial Toolbox	*	*	*
23	MATLAB Statistics Toolbox	*	*	*
24	MATLAB Systems Identification Toolbox	*	*	*
25	MATLAB Neural Net Toolbox	*	*	*
26	Minitab	Y	Y	Stepwise ANOVA GLM Nested ANOVA MANOVA Best Subset Partial least square Logistic regression
27	NCSS 2004		Y	Ridge Stepwise Partial least square Two-stage least squares Nonlinear least squares Nonlinear regression
28	Eviews			Poisson Truncated
29	SAS Software		Y	Ridge Poisson Stepwise Dynamic models Stepwise autoregressive models with constant, linear, or quadratic trend and autoregressive errors to any order
30	SCA Forecasting & Modeling Package		Y	Ridge Stepwise Adaptive Lagged (dynamic) regression Regression with autocorrelated errors
31	B34S Econometric Package		Y	Stepwise Lagged (dynamic) regression Regression with autocorrelated errors
32	Shazam		Y	
33	Smart Forecasts for Windows	*	*	*
34	SPSS for Windows			Ridge Poisson
35	SPSS Trends for Windows	*	*	*
36	Statistica for Windows		Y	Ridge Switching Stepwise

Α/Α	Λογισμικό	Μέθοδοι Παλινδρόμησης		
		Απλή	Πολλαπλή	Άλλη
37	WINKS Professional	*	*	*
38	TSP			Adaptive Poisson Truncated Switching
39	SAP SCM	*	*	*
40	Microsoft Forecaster	*	*	*
41	JD Edwards EnterpriseOne Demand Forecasting	*	*	*
42	PYTHIA	Υ	Υ	

(Υ) Χαρακτηριστικά που πληρεί το λογισμικό

(\*) Στοιχεία που δεν αναφέρονται στις ιστοσελίδες των εταιρειών ή δεν απαντήθηκαν στην επιστολή του Παρτήματος “Α”

## ΜΕΡΟΣ 2ο – ΚΕΦΑΛΑΙΟ 2ο

### 2.3.4 Δυνατότητες των Λογισμικών (Miscellaneous Capabilities)

Αρχικά, αναφέρονται γενικά στοιχεία που αφορούν τη τιμή πώλησης και την εξυπηρέτησης-βοήθεια του χρήστη κατά τη διάρκεια λειτουργίας του λογισμικού και εν συνεχεία γίνεται αναφορά και επεξήγηση σε κάποιες ειδικές λειτουργίες που εφαρμόζονται από τα λογισμικά κατά τη διαδικασία της πρόβλεψης.

Τα παραπάνω περιγράφονται στους πίνακες που ακολουθούν :

**ΠΙΝΑΚΑΣ 2.7 ΔΥΝΑΤΟΤΗΤΕΣ ΛΟΓΙΣΜΙΚΩΝ (MISCELLANEOUS CAPABILITIES)**

A/A	Λογισμικό	Τιμή (\$ US)	Site license	Εκπαιδευτική Έκδοση Διαθέσιμη	Τιμή (\$ US)	On-line Βοήθεια	On-line Εκμάθηση	Τεχνική Υποστήριξη
1	Modler	call	Y	Y	call	Y	Y	Y
2	Modler Blue	call	Y		call	Y	Y	Y
3	ABStat	\$449	Y	Y	\$49.95	Y	Y	Y
4	WEBSURV	\$15000	Y				Y	Y
5	GAUSS Applications	call	Y	Y	call	Y		Y
6	Autobox	\$ 695and up	negotiable	Y		Y		
7	Forecast Pro XE	\$995	Y	Y	call	Y		Y
8	Forecast Pro Standard Edition	\$595	Y	Y	call	Y		Y
9	Forecast Pro Unlimited Edition	\$4995	Y					Y
10	WinX11	\$895	Contact Vendor			Y		Y
11	Peer Planner for Windows	\$ 19,000 +	Contact Vendor			Y	Y	Y
12	PeerForecaster	\$425	Contact Vendor			Y	Y	Y
13	LIMDEP for Windows	\$595	Y	Y		Y		Y
14	Nlogit	\$795						
15	RATS	\$500	Depends on number of users	Y	\$25-\$60	Y		Y
16	EXPO	call	Y	Y	free	Y		Y
17	EXPO/Econometrics Add-in Module	call	Y	Y	free	Y		Y
18	Statgraphics Centurion	\$695	Y	Y	\$295	Y		Y
19	Statgraphics Centurion Professional	\$1295	Y	Y	\$495	Y		Y
20	S-PLUS 7 for Windows	call	Y	Y	call	Y	Y	Y
21	MATLAB	call	Y	Y	call	Y	Y	Y
22	MATLAB Financial Toolbox	call	Y	Y	call	Y	Y	Y
23	MATLAB Statistics Toolbox	call	Y	Y	call	Y	Y	Y

**ΜΕΡΟΣ 2ο – ΚΕΦΑΛΑΙΟ 2ο**

A/A	Λογισμικό	Τιμή (\$ US)	Site license	Εκπαιδευτική Έκδοση Διαθέσιμη	Τιμή (\$ US)	On-line Βοήθεια	On-line Εκμάθηση	Τεχνική Υποστήριξη
24	MATLAB Systems Identification Toolbox	call	Y	Y	call	Y	Y	Y
25	MATLAB Neural Net Toolbox	call	Y	Y	call	Y	Y	Y
26	Minitab	\$786 per year Annual Use Licence (1 copy, 1-5 users)	Y	Y	\$99.99	Y	Y	Y
27	NCSS 2004	\$499.95	Y	Y		Y	Y	Y
28	EViews	\$950	Y	Y	\$39.95			
29	SAS Software	call	Y	Y	call	Y	Y	Y
30	SCA Forecasting & Modeling Package	call	Y	Y	call	Y	Y	Y
31	B34S Econometric Package	call	Y	Y	call	Y		Y
32	Shazam	\$395	Y			Y	Y	Y
33	Smart Forecasts for Windows	call	Y	Y	call	Y	Y	Y
34	SPSS for Windows	\$1599	Y	Y	\$499	Y	Y	Y
35	SPSS Trends	\$799	Y	Y	\$499			Y
36	Statistica for Windows	call	Y	Y		Y	Y	Y
37	WINKS Professional	\$229	Y	Y	\$99	Y	Y	Y
38	TSP	call	Y	Y		Y		Y
39	SAP SCM	call				Y	Y	Y
40	Microsoft Forecaster	Base Application-single user \$8500 Additional User \$5000 Five Users \$1100 Fifty Users \$27500 Unlimited Users \$55000				Y	Y	Y
41	JD Edwards EnterpriseOne Demand Forecasting	*	*	*	*	*	*	*
42	PYTHIA	Δεν έχει προσδιοριστεί ακόμη		Y	Δεν έχει προσδιοριστεί ακόμη	Y	Y	Y

(Y) Χαρακτηριστικά που πληρεί το λογισμικό

(\*) Στοιχεία που δεν αναφέρονται στις ιστοσελίδες των εταιρειών ή δεν απαντήθηκαν στην επιστολή του Παρτήματος “Α”

**ΠΙΝΑΚΑΣ 2.8 ΕΙΔΙΚΕΣ ΛΕΙΤΟΥΡΓΙΕΣ ΛΟΓΙΣΜΙΚΩΝ (SPECIAL FEATURES)**

A/ A	Λογισμικό (1)	Ειδικά Γεγονότα και Συμβάντα (Special Events And Actions) (2)		Λειτουργία με χρήση της διαδικασίας bottom-up και top-down (3)	Περβήλλον Πολλαπλών Χρηστών (Multi-user interface) (4)	Σύγκριση των αποτελε- σμάτων διαφόρων χρηστών (5)	Ανατροφοδ ότηση των αποτελεσμά των και παράκολού θηση (Feedback of the results and monitoring) (6)	Λειτουργία λογισμικού μέσω διαδικτύου (7)
		(a)	(b)					
1	Modler	*	*	*	*	*	*	*
2	Modler Blue	*	*	*	*	*	*	*
3	ABStat	*	*	*	*	*	*	*
4	WEBSURV	*	*	*	*	*	*	*
5	GAUSS Applications	*	*	*	*	*	*	*
6	Autobox	Y	Y			Y		Y
7	Forecast Pro XE	*	*	*	*	*	*	*
8	Forecast Pro Standard Edition	*	*	*	*	*	*	*
9	Forecast Pro Unlimited Edition	*	*	*	*	*	*	*
10	WinX11	Y	Y			Y	Y	Y
11	PeerPlanner	Y	Y	Y	Y	Y	Y	Y
12	PeerForecaster	Y				Y	Y	Y
13	LIMDEP	*	*	*	*	*	*	*
14	Nlogit	*	*	*	*	*	*	*
15	RATS	*	*	*	*	*	*	*
16	EXPO	*	*	*	*	*	*	*
17	EXPO/Economet rics Add-in Module	*	*	*	*	*	*	*
18	Statgraphics Centurion standard	*	*	*	*	*	*	*
19	Statgraphics Centurion professional	*	*	*	*	*	*	*
20	S-PLUS 7 for Windows	*	*	*	*	*	*	*
21	MATLAB	*	*	*	*	*	*	*
22	MATLAB Financial Toolbox	*	*	*	*	*	*	*
23	MATLAB Statistics Toolbox	*	*	*	*	*	*	*



**ΜΕΡΟΣ 2ο – ΚΕΦΑΛΑΙΟ 2ο**

A/ A	Λογισμικό (1)	Ειδικά Γεγονότα και Συμβάντα (Special Events And Actions) (2)		Λειτουργία με χρήση της διαδικασίας bottom-up και top-down (3)	Περβήλλον Πολλαπλών Χρηστών (Multi-user interface) (4)	Σύγκριση των αποτελεσμά- των διαφόρων χρηστών (5)	Ανατροφο- δότηση των αποτελεσμά- των και παράκολου- θηση (Feedback of the results and monitoring) (6)	Λειτουργία λογισμικού μέσω διαδικτύου (7)
		(α)	(β)					
24	MATLAB Systems Identification Toolbox	*	*	*	*	*	*	*
25	MATLAB Neural Net Toolbox	*	*	*	*	*	*	*
26	Minitab	*	*	*	*	*	*	*
27	NCSS 2004	*	*	*	*	*	*	*
28	Eviews	*	*	*	*	*	*	*
29	SAS Software	Y	Y	Y	Y	Y		Y
30	SCA Forecasting & Modeling Package	*	*	*	*	*	*	*
31	B34S Econometric Package	*	*	*	*	*	*	*
32	Shazam							Y
33	Smart Forecasts for Windows	*	*	*	*	*	*	*
34	SPSS for Windows	*	*	*	*	*	*	*
35	SPSS Trends for Windows	*	*	*	*	*	*	*
36	Statistica for Windows	*	*	*	*	*	*	*
37	WINKS Professional	*	*	*	*	*	*	*
38	TSP	*	*	*	*	*	*	*
39	SAP SCM	*	*	*	*	*	*	*
40	Microsoft Forecaster			Y	Y		Y	Y
41	JD Edwards EnterpriseOne Demand Forecasting	Y						
42	PYTHIA	Y	Y	Y	Y	Y	Y	Y

(Y) Χαρακτηριστικά που πληρεί το λογισμικό

(\*) Στοιχεία που δεν αναφέρονται στις ιστοσελίδες των εταιρειών ή δεν απαντήθηκαν στην επιστολή του Παρτήματος "Α"

### Επεξηγήσεις

Τα διάφορα στοιχεία που φαίνονται στο παραπάνω πίνακα επεξηγούνται όπως παρακάτω:

(1) Ονομασία προϊόντος.

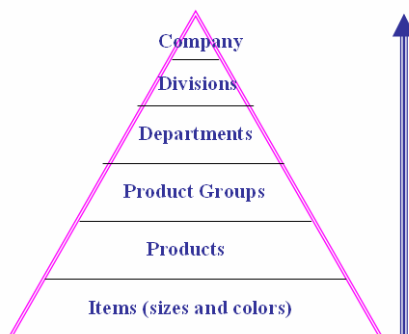
(2) Special Event ή Special Action (SEA) νοείται όταν υπάρχει διαφορά επιπέδου στα δεδομένα, βάση κάποιου αρχικού επιπέδου ή βάση του μέσου όρου των δεδομένων. Τα SEA, αφού ανιχνευθούν, ομαδοποιούνται ώστε να δώσουν περιόδους στις οποίες θεωρείται πως έγινε κάποιο συμβάν. Κατόπιν ο χρήστης επαληθεύει ή απορρίπτει τα αποτελέσματα. Εφόσον κάποιο αποτέλεσμα επαληθευτεί, τότε ο χρήστης καλείται να το επεξεργαστεί, δηλαδή να ορίσει τι τύπου είναι και να δώσει το ύψος του προϋπολογισμού για αυτό. Ο χρήστης μπορεί ακόμα να αλλάξει την χρονική τους περίοδο (έναρξη και λήξη). Σαν SEA μπορούν να οριστούν τα παρακάτω γεγονότα:

- Προωθήσεις (ποσό/τύπος)
  - Διαφήμιση (ποσό/μέσα)
  - Αυξήσεις τιμών
  - Μειώσεις τιμών
  - Προϊόντα υποκατάστατων: Αυξήσεις τιμών
  - Προϊόντα υποκατάστατων: Μειώσεις τιμών
  - Νέα προϊόντα (αναλογία απόδοσης/τιμή)
  - Ανεπαρκείς κατάλογοι
  - Ανεπαρκείς κατάλογοι των ανταγωνιστών
  - Ανταγωνιστικές προωθήσεις (ποσό/τύπος)
  - Ανταγωνιστική διαφήμιση (ποσό/μέσα)
  - Ανταγωνιστικές αυξήσεις τιμών
  - Ανταγωνιστικές μειώσεις τιμών
  - Ανταγωνιστικά προϊόντα υποκατάστατων: Αυξήσεις τιμών
  - Ανταγωνιστικά προϊόντα υποκατάστατων: Μειώσεις τιμών
  - Νέα προϊόντα που εισάγονται από τους ανταγωνιστές (αναλογία απόδοσης/τιμή)
  - Ασυνήθιστες καιρικές συνθήκες
  - Απεργίες
  - Απεργίες στους ανταγωνιστές
- (α) Εάν παρέχει το λογισμικό τη δυνατότητα αναγνώρισης των SEA.
- (β) Εάν παρέχει το λογισμικό τη δυνατότητα αποθήκευσης των SEA και της επίδρασής τους και εάν χρησιμοποιεί τα SEA στη διαδικασία εξομάλυνσης.

### (3) Διαδικασία από “κάτω προς τα πάνω” (Bottom – Up procedure)

Οι χρήστες μπορούν να δουν μόνο τις προβλέψεις του επιπέδου στο οποίο ανήκουν, καθώς επίσης και τις προβλέψεις των προϊόντων ή των κατηγοριών που ανήκουν στην κατώτερη βαθμίδα της πυραμίδας (ιεραρχία των χρηστών στην επιχείρηση). Αθροίζοντας τις προβλέψεις χαμηλότερων επιπέδων σε αυτές πιο υψηλού επιπέδου, καθορίζονται οι συνολικές προβλέψεις της επιχείρησης)

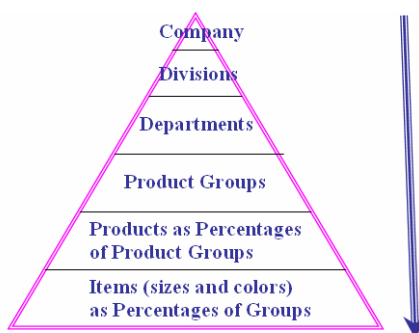
Σχήμα 2.1: Διαδικασία “ Bottom – Up ”



### Διαδικασία από “πάνω προς τα κάτω” (“Top– Down” procedure)

Ο χρήστης έχει υπό την επίβλεψή του μια κατηγορία προϊόντων, η οποία διακρίνεται σε άλλες κατηγορίες ή προϊόντα. Ο χρήστης είναι σε θέση να στείλει τις προβλέψεις του στα κατώτερα επίπεδα. Τα κατώτερα επίπεδα πρέπει να τις λάβουν υπόψη και να ρυθμίσουν τις αρχικές προβλέψεις τους.

Σχήμα 2.2: Διαδικασία “ Top-Down”



- (4) Εάν λειτουργεί το λογισμικό σε περιβάλλον πολλαπλών χρηστών.
- (5) Εάν έχει το λογισμικό δυνατότητα σύγκρισης αποτελεσμάτων διαφορετικών χρηστών.
- (6) Εάν έχει το λογισμικό δυνατότητα ανατροφοδότησης αποτελεσμάτων και επαναξιολόγησής τους;
- (7) Εάν έχει το λογισμικό δυνατότητα ενεργοποίησης και λειτουργίας το λογισμικό σας μέσω Διαδικτύου χωρίς την ανάγκη της εγκατάστασής του.

### ΚΕΦΑΛΑΙΟ 3<sup>ο</sup>: ΑΞΙΟΛΟΓΗΣΗ ΛΟΓΙΣΜΙΚΩΝ ΚΑΙ ΣΥΓΚΡΙΣΗ ΕΠΙΚΡΑΤΕΣΤΕΡΩΝ ΜΕ ΤΟ ΛΟΓΙΣΜΙΚΟ ΡΥΤΗΙΑ

#### 3.1 Εισαγωγή

Στην ενότητα αυτή, κάνοντας χρήση των πινάκων του 2<sup>ο</sup> κεφαλαίου, καταλήγουμε στα 10, κατά την άποψή μας επικρατέστερα λογισμικά, για τα οποία κάνουμε μια σύντομη παρουσίαση. Στη συνέχεια επιχειρούμε επισκόπηση του λογισμικού ΡΥΤΗΙΑ και σύγκρισή του με τα παραπάνω λογισμικά.

#### 3.2 Ποιοτική ανάλυση των συγκριτικών πινάκων και επιλογή των επικρατέστερων λογισμικών.

Λαμβάνοντας υπόψη τα συγκριτικά στοιχεία που παρέχουν οι πίνακες επιλέχθηκαν τελικά 10 λογισμικά με τα παρακάτω κριτήρια:

1. Ανταποκρίνονται στα περισσότερα από τα χαρακτηριστικά της λίστας, που καταρτίστηκε σε συνεργασία με τον επιβλέποντα την εργασία (Παράρτημα “B”) καθώς και με βάση αντίστοιχες παλιότερες δημοσιεύσεις.

2. Εφαρμόζουν πέραν των παραδοσιακών μεθόδων πρόβλεψης (μοντέλα κινητών μέσων όρων, εκθετική εξομάλυνση, απλή και πολλαπλή παλινδρόμηση, αποσύνθεση)- τις οποίες χρησιμοποιεί η πλειονότητα των λογισμικών- και τις προηγμένες τεχνικές προβλέψεων (Box-Jenkins-ARIMA, οικονομετρικά μοντέλα και τεχνητά νευρωνικά δίκτυα).

3. Το περιβάλλον εργασίας τους (user interface) ήταν το πλέον φιλικό και προσιτό προς το μέσο χρήστη παρόμοιων λογισμικών.

4. Οι εταιρείες που τα παράγουν, στις ιστοσελίδες τους, παραθέτουν πλήθος στοιχείων. Αυτό είχε σαν αποτέλεσμα την πιο εύκολη για μας συλλογή -ανά πάσα στιγμή- των απαιτούμενων για τους σκοπούς της εργασίας χαρακτηριστικών. Επιπλέον οι εταιρείες αυτές ανταποκρίθηκαν άμεσα και πρόθυμα στην επιστολή με την οποία τους ζητήσαμε την αποστολή στοιχείων για τα προϊόντα τους.

Τα επιλεγμένα λογισμικά είναι τα παρακάτω:

1. Autobox
2. Eviews
3. ForecastPro
4. Statgraphics
5. Minitab
6. SPSS
7. SAS
8. mySAP SCM
9. MicrosoftForecaster
10. JD Edwards EnterpriseOne Demand Forecasting and Advanced Forecast Modeling (Oracle)

### 3.3 Σύντομη παρουσίαση επικρατέστερων λογισμικών

#### 3.3.1 Παρουσίαση του Autobox

Το Autobox είναι ένα αντικειμενικό, στατιστικό, σύστημα προβλέψεων που χρησιμοποιεί μια προσέγγιση βασισμένη στα μοντέλα ARIMA (μεταβλητού) και στη Λειτουργίας Μεταφοράς (αιτιώδους).

Το σύστημα παρέχει μια αυτοματοποιημένη προσέγγιση στη διαμόρφωση των χρονοσειρών χρησιμοποιώντας τόσο την ιστορία αυτών των χρονοσειρών όσο και την ιστορία των πιθανών προτεινόμενων από το χρήστη αιτιωδών μεταβλητών. Αυτό το μοντέλο καλείται συχνά μια πολυωνυμική διανεμημένη καθυστέρηση ή μια δυναμική παλινδρόμηση ή η τάξη των Box-Jenkins μοντέλων.

Το Autobox αναπτύχθηκε βασιζόμενο σε δύο κεντρικές πεποιθήσεις. Πρώτη, η Box-Jenkins προσέγγιση (και μεταβλητή - ARIMA - και λειτουργία μεταφοράς - αιτιώδη πρότυπα), για να διαμορφώσει τον προσδιορισμό, την εκτίμηση, τα διαγνωστικά πρότυπα και τις προσβλέψεις παρέχει το κατάλληλο πλαίσιο για την διεξαγωγή προβλέψεων. Το παράδειγμα του μοντέλου Box-Jenkins είναι πλούσιο και σύμφωνα με την εταιρεία, εντάσσει τις περισσότερες άλλες κοινές μορφές αριθμητικών τεχνικών πρόβλεψης όπως η παλινδρόμηση και εκθετική εξομάλυνση.

Η δεύτερη πεποίθηση είναι ότι οι διαδικασίες που αποτελούνται από τις μεθόδους που εφαρμόζονται με έναν συνεπή τρόπο υπόκεινται στην αυτοματοποίηση. Σύμφωνα με την εταιρεία, οι τεχνικές πρόβλεψης Box-Jenkins είναι ένα τέτοιο σύνολο μεθόδων. Η μηχανή πρόβλεψης Autobox εφαρμόζει τη φιλοσοφία του μοντέλου των Box και Jenkins στις χρονοσειρές και μπορεί να δημιουργήσει τα πρότυπα αυτόματα.

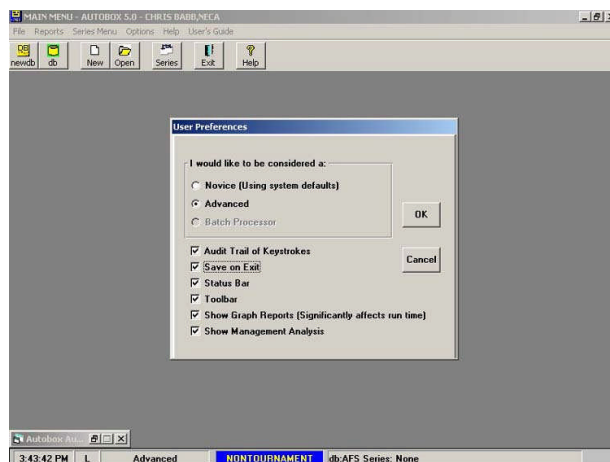
Για να παράξει τις ακριβέστερες προβλέψεις, το AUTOBOX προσαρμόζει αυτόματα το πρότυπο κοντά στο πρόβλημα συμπεριλαμβανομένης της επιλογής των καλύτερων δομών για κάθε σειρά εισαγωγής και των καλύτερων σταθμίσεων. Διορθώνει για τις μεταβλητές που έχουν παραβλεφθεί (π.χ., διακοπές ή μεταβολές των τιμών που έχουν επηρεάσει τα ιστορικά στοιχεία, αλλά για τις οποίες το σύστημα δεν έχει καμία γνώση) με τον προσδιορισμό των σφυγμών, των εποχιακών σφυγμών, των μετατοπίσεων επιπέδων και των τοπικών χρονικών τάσεων και έπειτα με την προσθήκη της αναγκαίας δομής μέσω των υποκαθιστάμενων μεταβλητών. Αντιθέτως, αποβάλλει επίσης την αχρεία δομή (π.χ., μια στατιστικά ασήμαντη αιτιώδη μεταβλητή) για να κρατήσει το πρότυπο εύχρηστο. Εκτελεί όλες αυτές τις λειτουργίες ως τμήμα της κανονικής ρουτίνας του χωρίς ανθρώπινη επέμβαση. Επίσης εκθέτει τα στατιστικά τεστ που χρησιμοποιούνται για να καθοριστούν οι πρότυπες παράμετροι, και αφήνει τους χρήστες να χειρίζονται τους συντελεστές και την δομή του πρότυπου εάν αυτοί θέλουν.

### Χαρακτηριστικά γνωρίσματα:

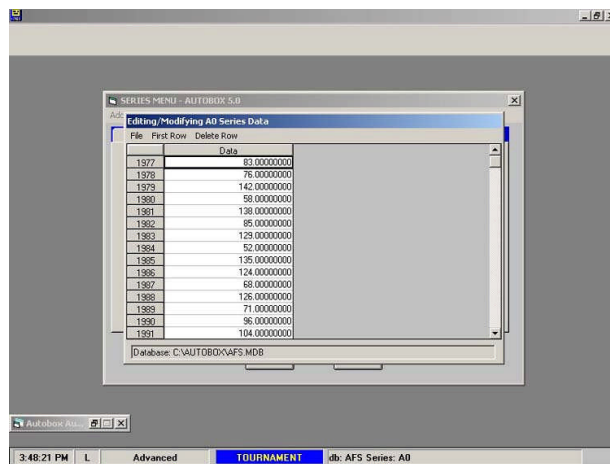
- Αυτόματη διαμόρφωση (Automatic Modeling)
- Διαμόρφωση πρωταθλημάτων (Tournament Modeling)
- Πλήρες σύνολο εργαλείων διαμόρφωσης Box-Jenkins (Complete set of Box-Jenkins modeling Tools)
- Ανίχνευση επέμβασης (Intervention Detection)
- Γραφικά εργαλεία ανάλυσης (Graphical Analysis Tools)
- Προβλέψεις και διαγνωστικά (Forecasting and Diagnostics)
- Βάση δεδομένων και πρότυπη αποθήκευση ( Database and Model Storage)
- Μέτρα για να αξιολογηθεί η πρότυπη απόδοση πρόβλεψης (Measures to Assess Forecast Model Performance)

Το Autobox προσφέρει ένα ολοκληρωμένο εργαλείο διαχείρισης δεδομένων και μια συνολική αυτοματοποίηση των στόχων μοντελοποίησης και πρόβλεψης. Είναι ένα πρόγραμμα βασισμένο στα Windows με μια πλούσια, εύχρηστη δομή επιλογών που χτίζεται γύρω από την ισχυρή μηχανή πρόβλεψών του. Παρατίθενται ενδεικτικά κάποια screen shots του Autobox

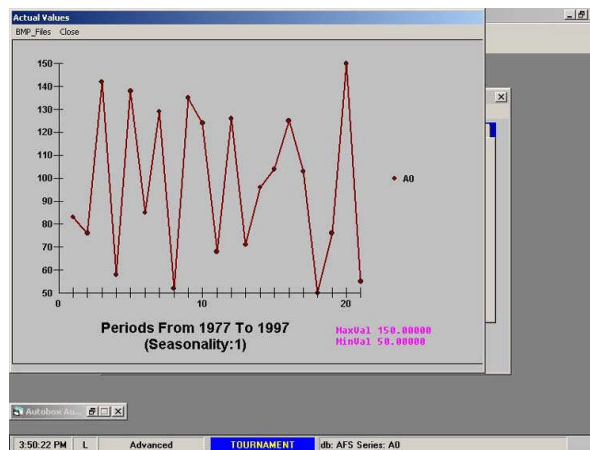
### Setting User Level



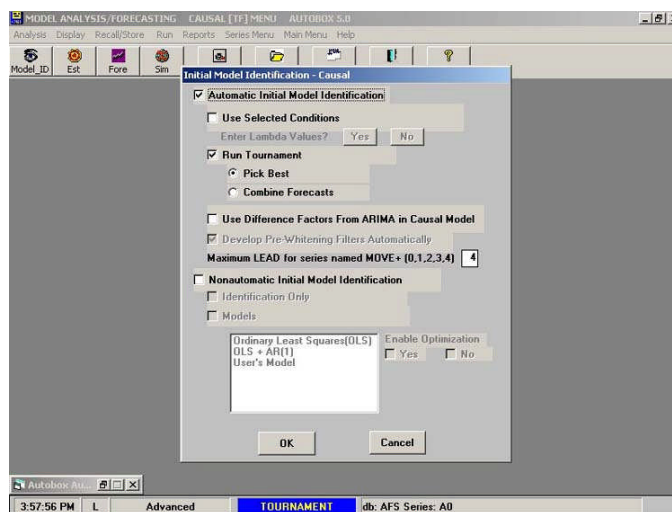
Editing/ Modifying Series



Series Graph



Initial Model Identification Settings



### 3.3.2. Παρουσίαση του EViews

Το EViews είναι ένα πακέτο λογισμικού για την εκτίμηση και την προσομοίωση των οικονομετρικών προτύπων. Το σύστημα προσφέρει ένα φιλικό και καλά σχεδιασμένο GUI (Graphics User Interface) που παρουσιάζει στο σύνολο του μια καινούρια ιδέα στα οικονομετρικά λογισμικά και ενθαρρύνει τη χρήση.

Το EViews υποστηρίζει ένα ευρύ φάσμα γραφημάτων παρουσίασης. Ένα ευρύ φάσμα από στατιστικές τεχνικές και τεχνικές γραφημάτων έπρεπε να είναι διαθέσιμο χωρίς να απαιτείται από τους χρήστες η απομνημόνευση πολύπλοκων εντολών σύνταξης ή επιπέδων πλοήγησης και μενού. Η λύση είναι ένα καινοτόμο αντικειμενοστραφές περιβάλλον χρήστη.

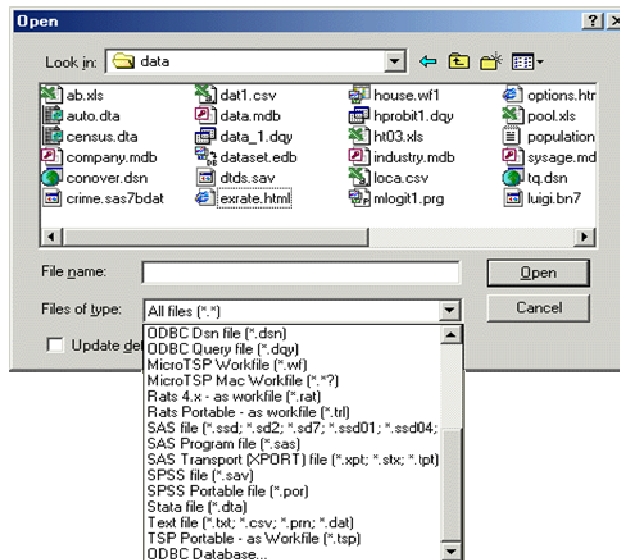
Το EViews 5 είναι δομημένο γύρω από τις έννοιες των αντικειμένων. Σειρές, εξισώσεις και συστήματα είναι μόνο κάποια παραδείγματα αντικειμένων. Κάθε αντικείμενο έχει το δικό του παράθυρο, το δικό του μενού, τις δικές του διαδικασίες και τις δικές του όψεις για τα δεδομένα του. Οι περισσότερες στατιστικές διαδικασίες είναι απλά εναλλακτικές απόψεις-θέσεις του αντικειμένου. Για παράδειγμα, μια απλή επιλογή μενού από ένα παράθυρο σειρών αλλάζει την ένδειξη μεταξύ ενός λογιστικού φύλλου, γραμμών και στηλών γραφημάτων.

#### Περιβάλλον χρήστη (User Interface)

Φυσικά, μπορούμε να αντιγράψουμε και να επικολλήσουμε οποιεσδήποτε από αυτές τις απόψεις στον επεξεργαστή κειμένου με μια απλή επιλογή ενός μενού. Και είναι εξίσου εύκολο να ανταλλάξουμε δεδομένα και αποτελέσματα με τα προγράμματα υπολογισμών με λογιστικά φύλλα (spreadsheet) και με βάσεις δεδομένων.

Το EViews διαβάζει άμεσα και γράφει έναν εκτενή κατάλογο από μορφές δεδομένων (data formats) , συμπεριλαμβανομένων των ASCII/Text, SAS, Stata, SPSS, RATS, Html, Access, Binary, ODBC Databases, ODBC queries ( το ODBC απαιτεί την επιχειρηματική έκδοση), και πολλών άλλων. Για να ανοίξουμε τα περισσότερα αρχεία δεδομένων μπορούμε απλά να τα "σύρουμε" και να τα "ρίξουμε" στο EViews



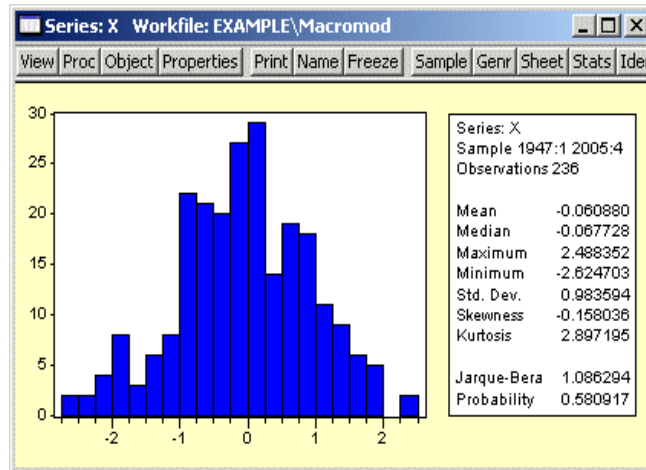


### Οικονομετρικά Εργαλεία (Econometric Tools)

Αντίθετα από κάποια άλλα οικονομετρικά λογισμικά, δεν υπάρχει κανένας λόγος για τους περισσότερους χρήστες να μάθουν μια περίπλοκη γλώσσα εντολών. Οι ενσωματωμένες διαδικασίες του EViews είναι ένα κλικ του ποντικιού μακριά και παρέχουν τα πιο συχνά χρησιμοποιούμενα εργαλεία practical econometric and forecasting work.

- *Βασικά Στοιχεία Στατιστικής (Basic Statistics)*

Οι βασικές περιγραφικές στατιστικές υπολογίζονται εύκολα πέρα από ένα ολόκληρο δείγμα, από μια κατηγοριοποίηση βασισμένη σε μια ή περισσότερες μεταβλητές, ή σε ένα αντιπροσωπευτικό δείγμα ή την περίοδο ενός καταλόγου ή των συγκεντρωμένων στοιχείων. Οι δοκιμές υπόθεσης (τα Hypothesis tests) στο μέσο όρο, τη διάμεσο και τη διακύμανση μπορούν να πραγματοποιηθούν, συμπεριλαμβανομένης της δοκιμής ενάντια στις συγκεκριμένες τιμές, για την ισότητα μεταξύ σειρών, ή κάνοντας τον έλεγχο για την ισότητα μέσα σε μια ενιαία σειρά όταν ταξινομείται από άλλες μεταβλητές.

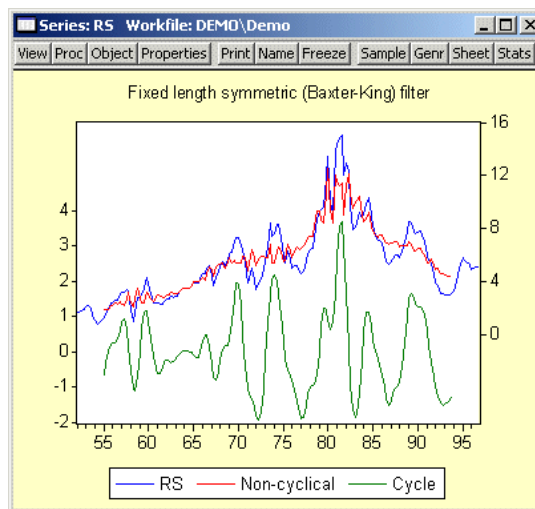


- *Εποχιακή Προσαρμογή (Seasonal Adjustment)*

Το EViews 5 παρέχει εύχρηστη υποστήριξη για τις μεθόδους Census Bureau's X11 και X12-ARIMA. Η εποχιακή προσαρμογή που χρησιμοποιεί τις προσθετικές και πολλαπλασιαστικές μεθόδους διαφοράς υποστηρίζεται επίσης από EViews.

- *Φίλτρα (Filters)*

Το EViews υπολογίζει τις τάσεις από τις χρονοσειρές χρησιμοποιώντας το φίλτρο Hodrick-Prescott

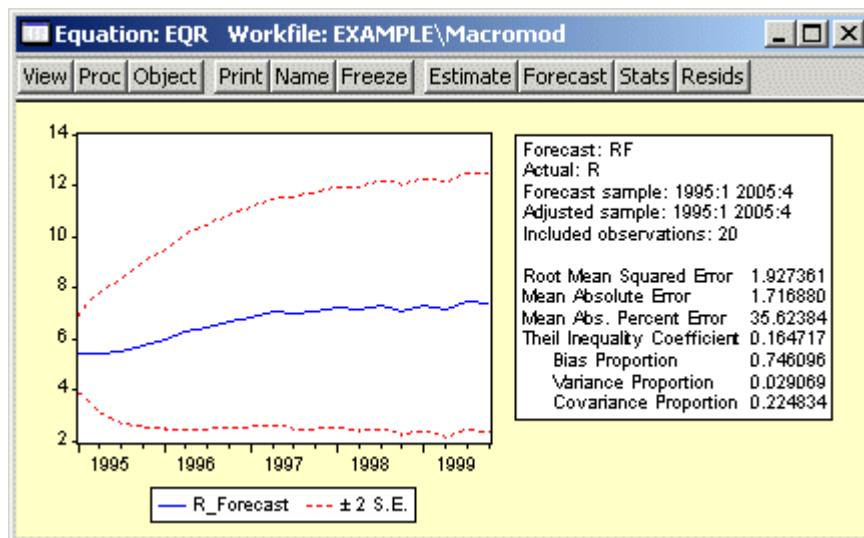


- *Εκτίμηση (Estimation)*

Το EViews περιλαμβάνει ένα ευρύ φάσμα από απλές και πολλαπλές τεχνικές εκτίμησης εξίσωσης τόσο για τις χρονοσειρές όσο και για τα στοιχεία διαγώνιων τμημάτων. Οι βασικοί εκτιμητές περιλαμβάνουν τη μέθοδο ελαχίστων τετραγώνων (πολλαπλή παλινδρόμηση), τη μέθοδο ελαχίστων τετραγώνων δύο σταδίων λ, και τη μέθοδο μη γραμμικών ελαχίστων τετραγώνων. Σταθμισμένη εκτίμηση είναι διαθέσιμη με όλες αυτές τις τεχνικές.

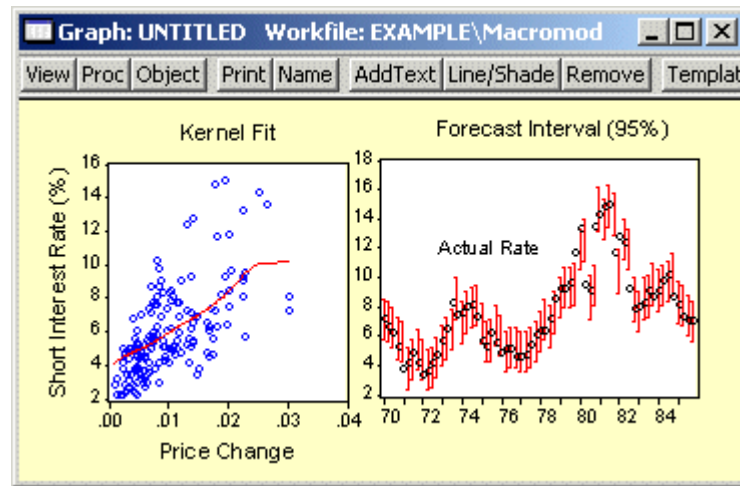
Διαδικασίες Πρόβλεψης και Προσομοίωσης (Forecasting and Simulation)

Στο EViews, δεν χρειάζεται να ανησυχούμε με τις δυσκολίες της παραγωγής των προβλέψεων. Μπορούμε να επικεντρωθούμε στην ουσία του προβλήματος πρόβλεψης. Για τα πρότυπα ενιαίας εξίσωσης, επιλέγουμε ακριβώς ένα αντικείμενο επιλογών και το EViews θα πραγματοποιήσει μια στατική ή δυναμική πρόβλεψη με τα προαιρετικά τυποποιημένα λάθη πρόβλεψης και μια γραφική παράσταση των 95 % διαστημάτων εμπιστοσύνης. Οι επιτυχείς εξισώσεις πρόβλεψης μπορούν να σωθούν ή να αποθηκευτούν σε μια βάση δεδομένων του EViews.



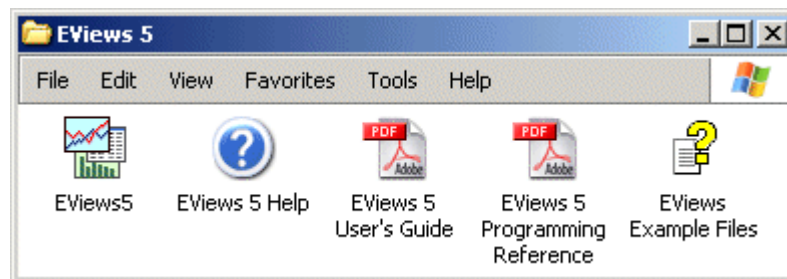
Γραφικά (Graphics)

Το EViews υποστηρίζει ένα ευρύ φάσμα γραφημάτων (line graphs, bar graphs, filled area graphs, pie charts, scatter diagrams, mixed line-bar graphs, high-low graphs, scatter plots and boxplots). Ποικίλες επιλογές μας δίνουν τον έλεγχο των τύπων των γραμμών, του χρώματος, των χαρακτηριστικών συνόρων, των τίτλων κλπ. Τα υπομνήματα δημιουργούνται αυτόματα και μπορείτε να προσθέσετε τις ετικέτες οπουδήποτε στη γραφική παράστασή σας. Οποιοσδήποτε αριθμός γραφημάτων μπορεί να συνδυαστεί σε ένα και μοναδικό γράφημα για την παρουσίαση.



### Windows on-line Help

Το EViews παρέχει ένα πλήρες σύστημα βοήθειας σε στυλ παράθυρου με δείκτες και δυνατότητες αναζήτησης.



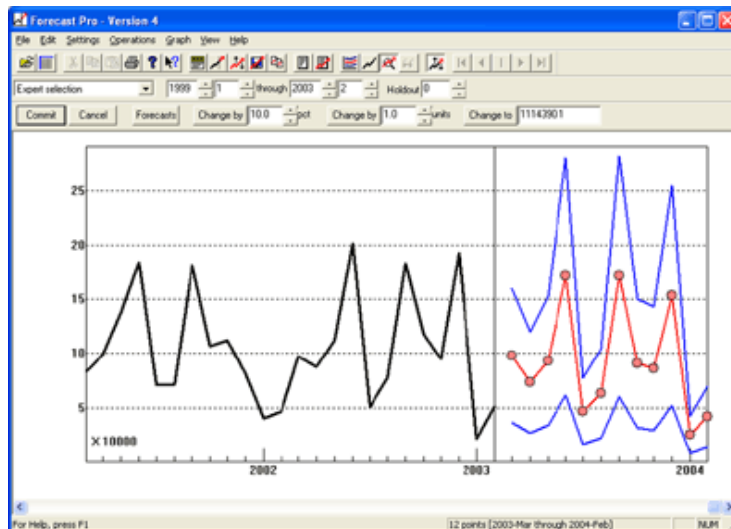
### 3. 3. 3 Παρουσίαση του ForecastPro

Το Forecast Pro είναι ένα γρήγορο, εύκολο, εύχρηστο και ακριβές λογισμικό πρόβλεψης για επαγγελματίες.

Με το Forecast Pro, παρέχουμε μόνο ό,τι προηγούμενα δεδομένα έχουμε για τα στοιχεία που θέλουμε να προβλέψουμε και το λογισμικό κάνει όλες τις υπόλοιπες διαδικασίες. Ο ενσωματωμένος ειδικός τρόπος επιλογής αναλύει τα δεδομένα μας, επιλέγει την κατάλληλη τεχνική πρόβλεψης και υπολογίζει τις προβλέψεις χρησιμοποιώντας τις αποδεδειγμένες στατιστικές μεθόδους. Μπορούμε να συνεργαστούμε με συναδέλφους μας για να διενεργήσουμε προσαρμογές στις στατιστικές προβλέψεις και μπορούμε εύκολα στη συνέχεια να αποθηκεύσουμε τις αλλαγές.

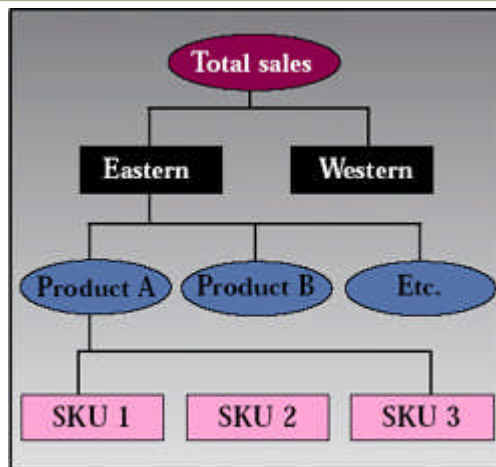
Το Forecast Pro παράγει εκθέσεις και γραφικές παραστάσεις προκειμένου να δημιουργηθούν οι επιθυμητές κάθε φορά παρουσιάσεις. Η αυτοματοποίηση της πρόβλεψης και η ενσωμάτωση των αποτελεσμάτων πρόβλεψης με άλλα συστήματα προγραμματισμού είναι μερικοί από τους απλούς στόχους του Forecast Pro.

Το Forecast Pro ΧΕ χρησιμοποιεί αποδεδειγμένες στατιστικές μεθόδους για να επιτύχει τη μέγιστη ακρίβεια στις προβλέψεις.



Χρησιμοποιούνται οι παρακάτω τεχνικές των μοντέλων πρόβλεψης:

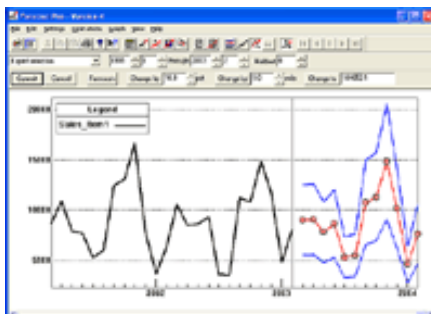
- Simple Methods
- Low Volume Models
- Exponential Smoothing
- Box-Jenkins
- Dynamic regression
- Event models
- Multiple-level models



Example of a multiple level forecasting job.

- Census X-11 seasonal decomposition
- Batch forecasting

Το Forecast Pro επιτρέπει να προσαρμόσουμε τις προβλέψεις μας σε ένα γράφημα ή σε ένα πίνακα όπως φαίνεται παρακάτω:



Name	Var1	Var2	Var3	Var4	Var5
2002-Mar	37896.0	10612.0	19197.0	517.0	4
2002-Apr	87526.0	8589.0	15639.0	566.0	2
2002-May	112219.0	8614.0	14211.0	588.0	2
2002-Jun	201558.0	9301.0	29107.0	582.0	4
2002-Jul	58452.0	7533.0	10231.0	566.0	2
2002-Sep	77349.0	3537.0	13452.0	317.0	2
2002-Sep	102564.0	11227.0	20024.0	929.0	1
2002-Oct	118076.0	10043.0	13736.0	828.0	2
2002-Nov	94666.0	14850.0	15103.0	838.0	2
2002-Dec	192827.0	11716.0	28753.0	1000.0	2
2003-Jan	21124.0	4892.0	7339.0	491.0	1
2003-Feb	51423.0	8954.0	14578.0	745.0	2
2003-Mar					

Καθορισμός του μοντέλου πρόβλεψης

Το Forecast Pro παρέχει ένα καθορισμένο σύνολο από “οθόνες διάγνωσης” (diagnostic screens) που βοηθάνε στη σύγκριση και στην εκτίμηση των μοντέλων. (Adjusted R-Square, MAPE, Sample Size, Mean, Schwartz BIC, Ljung-Box, RMSE, Standard Deviation, MAD, R-Square and Standard Forecast Error).

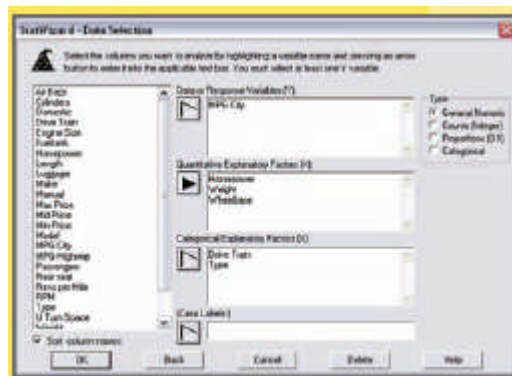
### 3.3.4. Παρουσίαση του Statgraphics Centurion

Το STATGRAPHICS Centurion είναι σχεδιασμένο για κάθε έναν που επιθυμεί να κάνει σοβαρή ανάλυση δεδομένων χωρίς να επενδύσει εβδομάδες μαθαίνοντας πώς να χρησιμοποιήσει ένα στατιστικό πακέτο. Περιέχει πάνω από 150 στατιστικές διαδικασίες που καλύπτουν όλα τα πεδία, από τις συνοπτικές αρχές στατιστικής μέχρι το σχεδιασμό πειραμάτων. Ακόμη, δεν πρέπει να είναι κανείς στατιστικός για να χρησιμοποιήσει το πρόγραμμα. Όλα είναι απολύτως καθοδηγούμενα από το μενού, και υπάρχουν εργαλεία όπως το StatWizard και το StatAdvisor για να βοηθήσουν την αποτελεσματικότερη χρήση του προγράμματος.

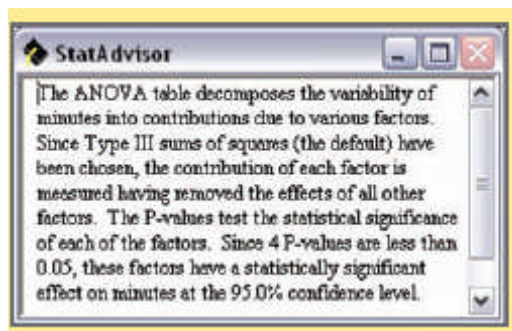
Το STATGRAPHICS Centurion XV είναι η 15η έκδοση STATGRAPHICS για τα PC. Η πρώτη έκδοση κυκλοφόρησε το 1982

#### Μοναδικά Χαρακτηριστικά που κάνουν το χρήστη πιο παραγωγικό

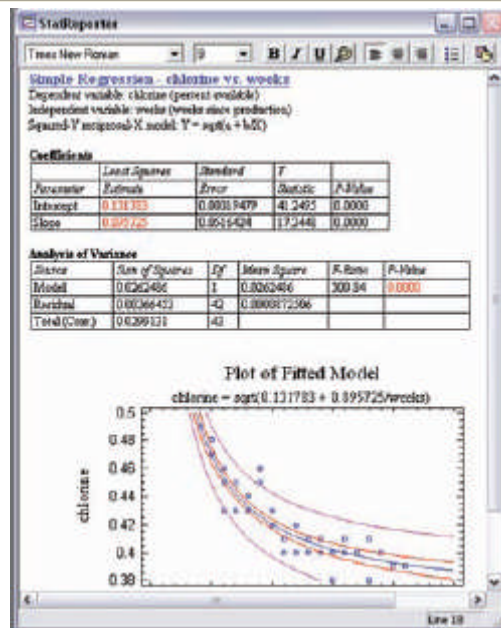
**StatWizard** - Βοηθά να επιλέξει τη κατάλληλη στατιστική ανάλυση για το τύπο των δεδομένων



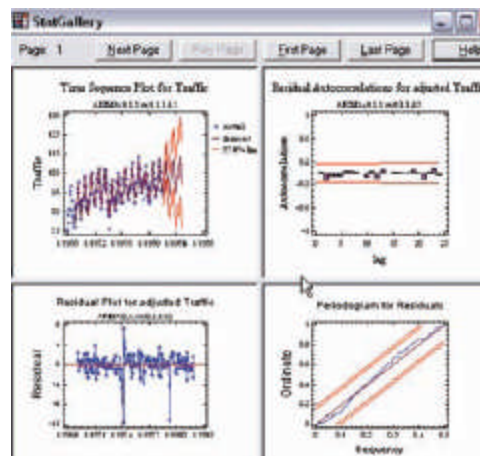
**StatAdvisor** - Προσφέρει στιγμιαίες και ευκολονόητες ερμηνείες των στατιστικών αποτελεσμάτων μας.



**StatReporter** - Συλλέγει τα αποτελέσματα των αναλύσεών μας και δημιουργεί μια παρουσίαση-ποιοτική έκθεση.



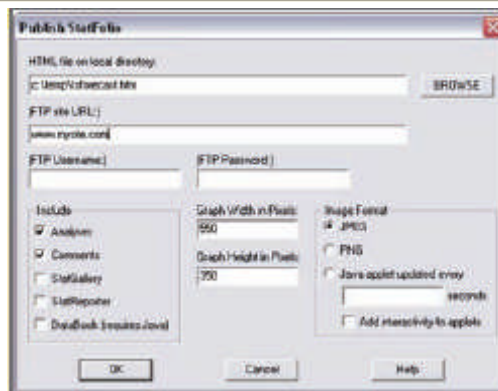
**StatGallery** - Επιτρέπει να συνδυασθούν μέχρι εννέα γραφικές παραστάσεις σε μια ενιαία σελίδα ή να παρατεθούν δύο γραφικές παραστάσεις για να συγκριθούν.



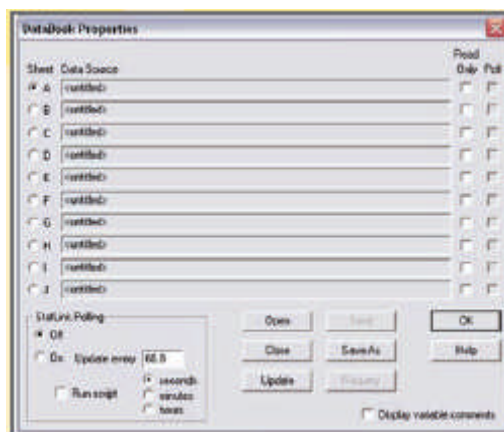
**StatFolios** – Σώζει τις αναλύσεις έτσι ώστε να μπορούν επανακτηθούν σε μια μεταγενέστερη ημερομηνία.

**StatPublish** – Δημοσιεύει τα αποτελέσματά μας στα αρχεία HTML έτσι ώστε να μπορούν να διαβαστούν από τον web browser.

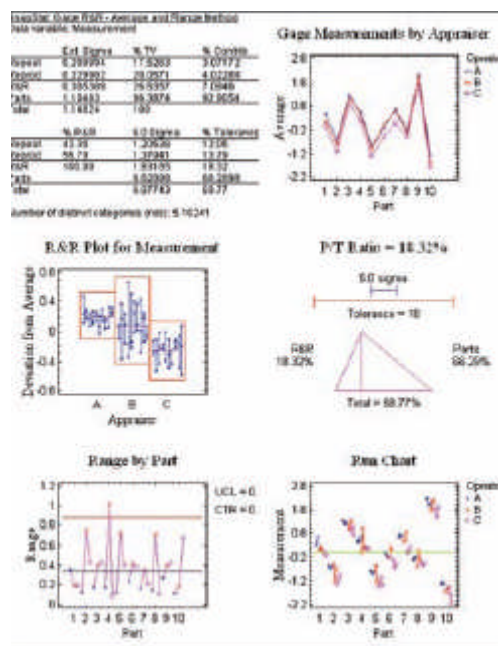




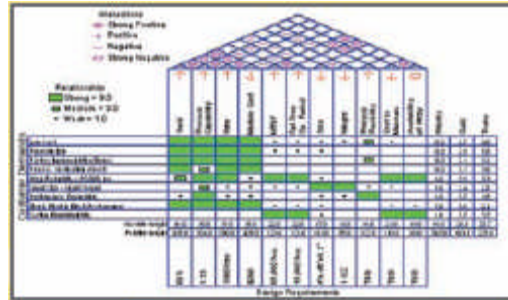
**StatLink** – Συνδέει τις αναλύσεις μας με τις πηγές στοιχείων για στιγμιαία αναπροσαρμογή όποτε τα στοιχεία αλλάζουν.



**SnapStats** – Παρέχει απλές εφαρμογές των συνήθως χρησιμοποιημένων αναλύσεων και το αποτέλεσμα Σχηματοποιείται για να χωράει σε μια ενιαία τυπωμένη σελίδα.



**Six Sigma menu** – Οργανώνει τις στατιστικές διαδικασίες σύμφωνα με το DMAIC (Define, Measure, Analyze, Improve, Control) παράδειγμα.



### 3.3.5. Παρουσίαση του Minitab

Το Minitab είναι ένα εύχρηστο, εύκαμπτο όμως ισχυρό πακέτο για χρήση σε προσωπικούς υπολογιστές που εκτελεί ένα ευρύ φάσμα των στατιστικών αναλύσεων. Είναι ιδιαίτερα κατάλληλο για τους σπουδαστές και τους ερευνητές που δεν είναι εξοικειωμένοι με τη χρήση ενός υπολογιστή. Ο αριθμός φύλλων εργασίας που μπορούμε να ανοίξουμε συγχρόνως περιορίζεται μόνο από τη μνήμη του υπολογιστή μας. Εξ' ορισμού μπορεί να χειριστεί 100.000 κύτταρα ή στοιχεία δεδομένων, μερικά από τα οποία μπορούν να είναι ελλείπουσες τιμές. Τα στοιχεία μπορούν είτε να εισαχθούν μέσω του πληκτρολογίου του PC είτε να ανακτηθούν από ένα αρχείο στο δίσκο, και μπορούν έπειτα να τροποποιηθούν, να μετασχηματιστούν και να σωθούν σε αρχείο μέσα στο Minitab. Τα τυπωμένα αντίγραφα των αποτελεσμάτων που εμφανίζονται στην οθόνη μπορούν να παραχθούν για την εξέταση στον ελεύθερο χρόνο.

### Χαρακτηριστικά και Δυνατότητες (Features and Capabilities)

#### Εύκολο στη χρήση (Ease of Use)

- Εξατομικεύσιμα μενού και γραμμές εργαλείων (Customizable menus and toolbars)
- Project Manager: οργανώνει λογικά την ανάλυση (logically organizes analysis)
- StatGuide™ : εξηγεί την έξοδο (explains output)
- ReportPad™ : δημιουργεί αναφορές (generates reports)
- Online tutorials
- Smart dialog boxes: διατηρεί στη μνήμη πρόσφατες ρυθμίσεις (remember recent settings)
- Σαφές και περιεκτικό HTML Σύστημα βοήθειας (Clear, comprehensive HTML Help system)

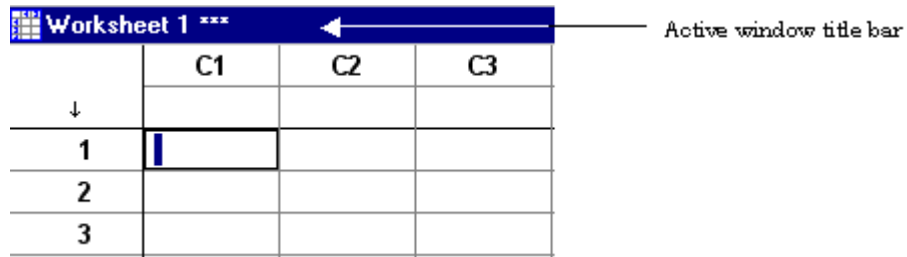
#### Διαχείριση Αρχείων και Δεδομένων (Data and File Management)

## ΜΕΡΟΣ 2ο – ΚΕΦΑΛΑΙΟ 3ο

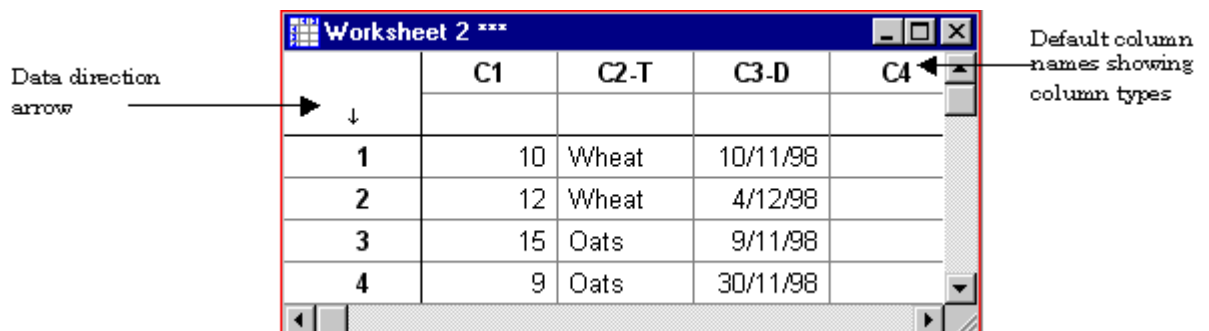
Αποθήκευση αποτελεσμάτων και δεδομένων σε μορφή HTML (Save output and data as HTML)

Απεριόριστο μέγεθος φύλλων εργασίας (Unlimited worksheet size)

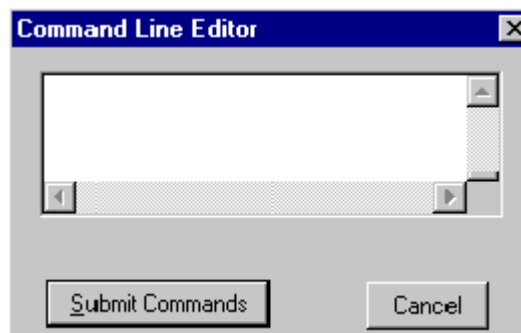
- Χειρισμός δεδομένων: συγχωνεύει, κατατάσει σε υποσύνολα, περικόπτει, μεταθέτει, αλλάζει τον τύπο στοιχείων (Data manipulation: merge, subset, sort, transpose, change data type)
- Νέα και ενημερωμένα σύνολα δειγμάτων δεδομένων (New and updated sample data sets)

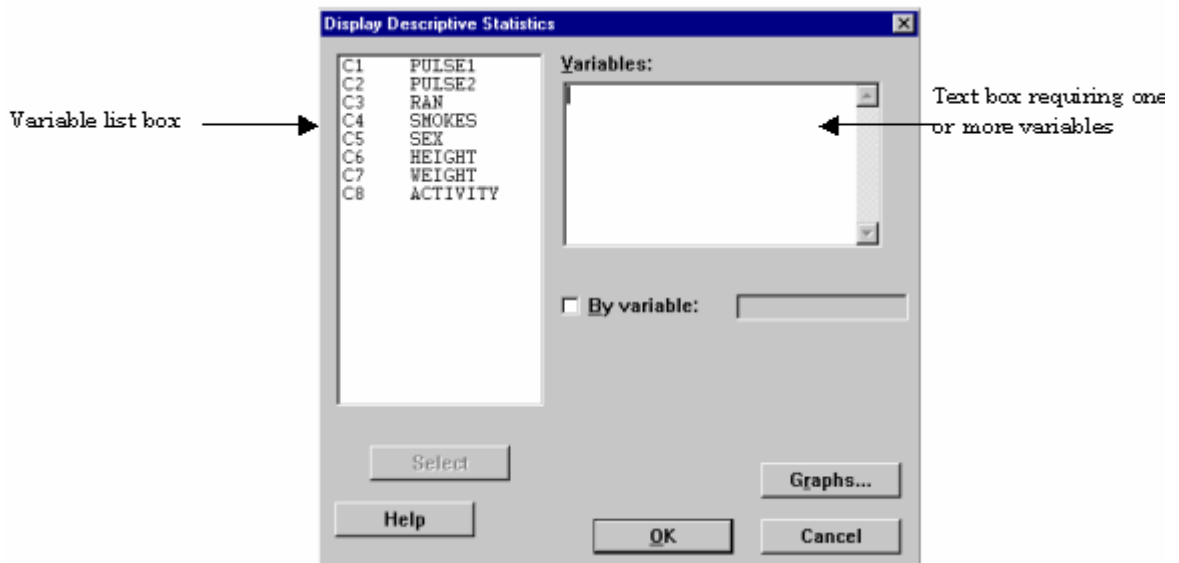
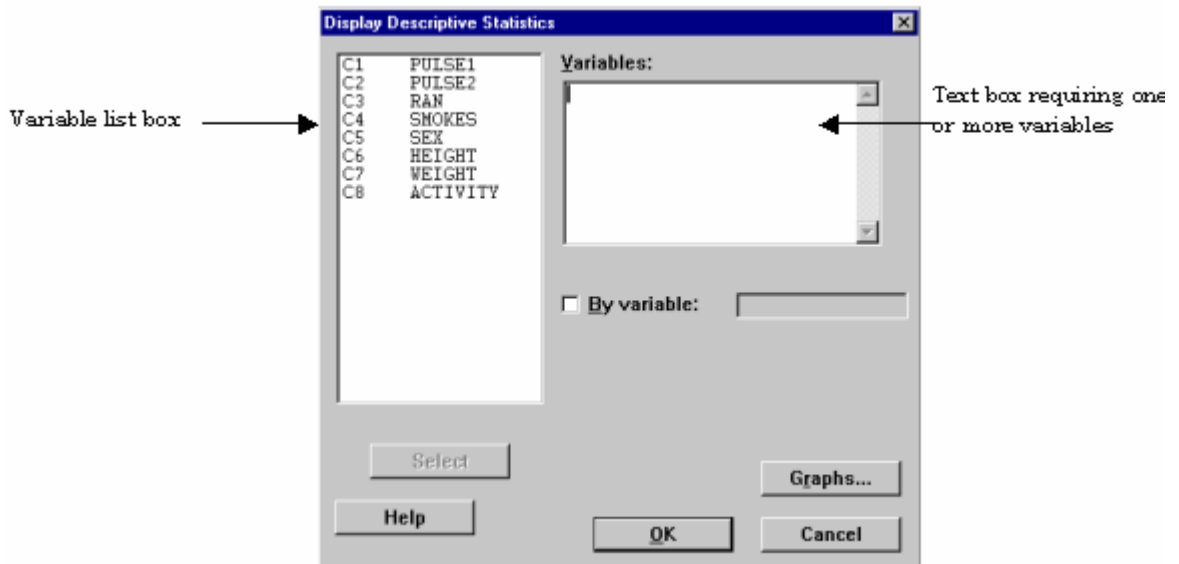
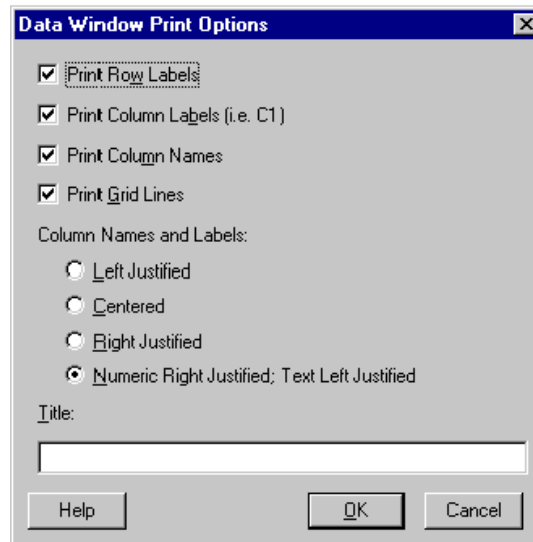


	C1	C2	C3
↓			
1			
2			
3			



	C1	C2-T	C3-D	C4
↓				
1	10	Wheat	10/11/98	
2	12	Wheat	4/12/98	
3	15	Oats	9/11/98	
4	9	Oats	30/11/98	





### Ανάλυση Παλινδρόμησης (Regression Analysis)

- Γραμμική Παλινδρόμηση (Linear regression)
- Πολυωνυμική Παλινδρόμηση (Polynomial regression)
- Λογιστική Παλινδρόμηση (Logistic regression)

### Χρονοσειρές και Διαδικασίες Πρόβλεψης (Time Series and Forecasting)

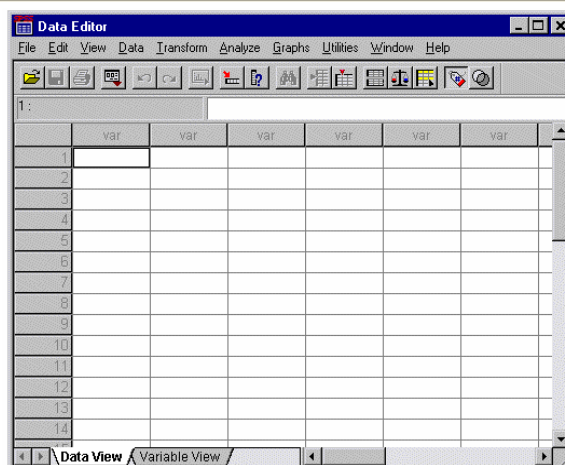
- Διαγράμματα Χρονοσειρών (Time series plots)
- Αυτοσυσχετίσεις, μερικές αυτοσυσχετίσεις (Auto-, partial auto-, and cross correlations)
- Ανάλυση ARIMA (ARIMA analysis)
- Αποσύνθεση (Decomposition)
- Εκθετική Εξομάλυνση (Exponential smoothing)
- Μέθοδος Winter's (Winter's method)
- Κινητοί Μέσοι Όροι (Moving average)

#### 3.3.6. Παρουσίαση του SPSS

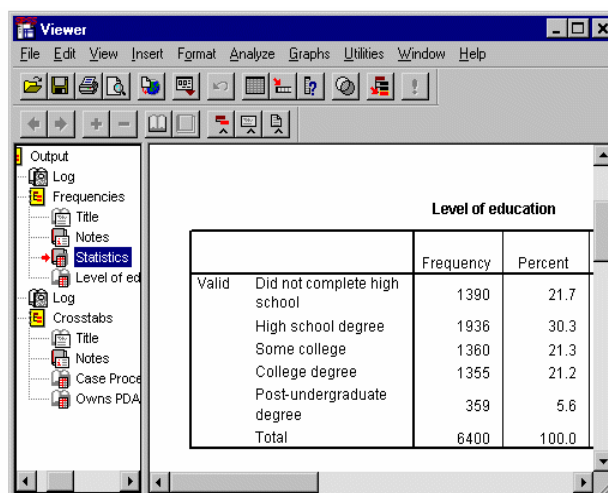
Το SPSS για τα Windows παρέχει ένα ισχυρό στατιστικό σύστημα ανάλυσης και ένα σύστημα διαχείρισης στοιχείων σε ένα περιβάλλον βασισμένο σε γραφικά, χρησιμοποιώντας περιγραφικές επιλογές και απλά πλαίσια διαλόγου για να κάνει το μεγαλύτερο μέρος της δουλειάς για σας. Οι περισσότεροι στόχοι μπορούν να ολοκληρωθούν απλά με την υπόδειξη και το κλικ του ποντικιού.

Το SPSS για τα Windows παρέχει:

**Συντάκτη Δεδομένων (Data Editor).** Ένα ευπροσάρμοστο , με τη μορφή λογιστικού φύλλου (spreadsheet), σύστημα για τον καθορισμό, την εισαγωγή και την παρουσίαση δεδομένων.

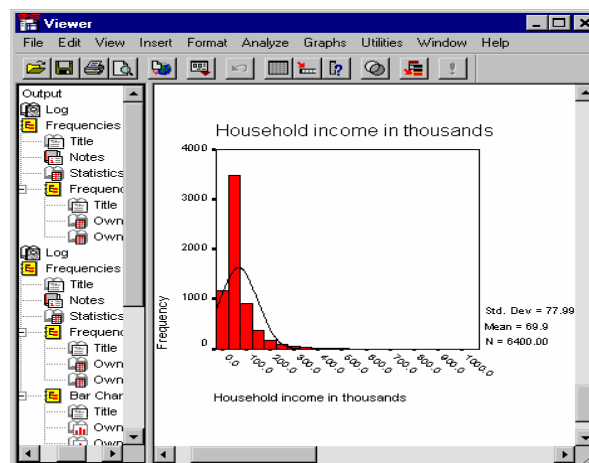
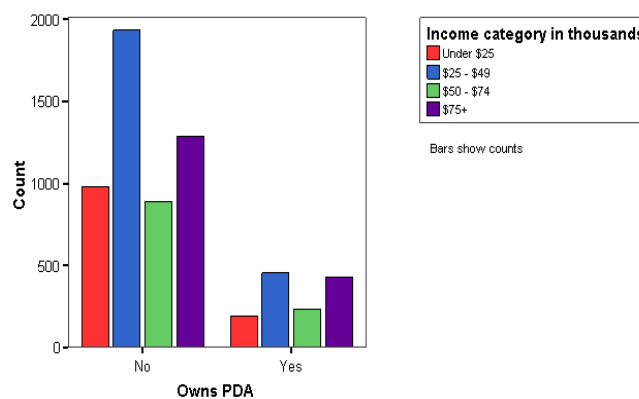
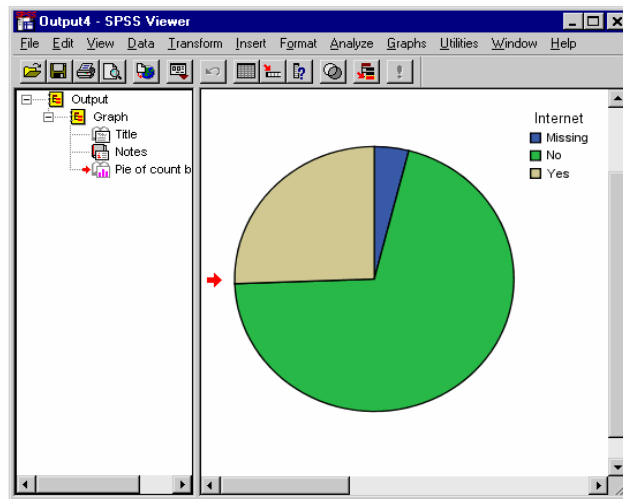


**Θεατή (Viewer)** . Ο θεατής (Viewer) καθιστά εύκολο το γεγονός να ανατρέξουμε στα αποτελέσματά μας, να παρουσιάσουμε και να κρύψουμε επιλεκτικά τα εξαγόμενα, να αλλάξουμε τη διάταξη εμφάνισης των αποτελεσμάτων, και να κινήσουμε πίνακες και διαγράμματα μεταξύ SPSS και άλλων εφαρμογών.



**Πολυδιάστατους πίνακες άξονα (Multidimensional pivot tables)**. Τα αποτελέσματά έρχονται ζωντανά με τους πολυδιάστατους πίνακες άξονα.

**Υψηλής ευκρίνειας γραφικά (High-resolution graphics)**. Τα διαγράμματα υψηλής ευκρίνειας, οι "πίτες" πλήρους - χρώματος, τα bar charts, τα ιστογράμματα, τα τρισδιάστατα γραφικά, και άλλα συμπεριλαμβάνονται ως τυποποιημένα χαρακτηριστικά του SPSS.



**Πρόσβαση βάσεων δεδομένων (Database access).** Μπορούμε να ανακτήσουμε τις πληροφορίες από τις βάσεις δεδομένων με τη χρησιμοποίηση του μάγου βάσεων δεδομένων (Database Wizard) αντί των περίπλοκων SQL ερωτήσεων (SQL queries).

**Μετασχηματισμοί στοιχείων (Data transformations).** Το χαρακτηριστικό γνώρισμα του μετασχηματισμού παίρνει τα στοιχεία σας έτοιμα για την ανάλυση.

**Ηλεκτρονική διανομή (Electronic distribution).** Μπορούμε να στείλουμε τις εκθέσεις μας μέσω ηλεκτρονικού ταχυδρομείου σε άλλους με το πάτημα ενός πλήκτρου, ή να εξάγουμε τους πίνακες και τα διαγράμματα σε μορφή HTML για τη διανομή τους μέσω Διαδικτύου και εσωτερικού δικτύου.

### Σε απευθείας σύνδεση βοήθεια (Online Help).

**Γλώσσα εντολών (Command language).** Αν και οι περισσότεροι στόχοι μπορούν να ολοκληρωθούν με τις απλές point-and-click ενέργειες, το SPSS παρέχει επίσης μια ισχυρή γλώσσα εντολών που μας επιτρέπει να σώσουμε και να αυτοματοποιήσουμε πολλά κοινά θέματα.

### Εκτίμηση Καμπυλών (Curve Estimation)

Η διαδικασία εκτίμησης καμπυλών παράγει τις στατιστικές εκτίμησης μέσω παλινδρόμησης καμπυλών και τα σχετικά διαγράμματα για 11 διαφορετικά πρότυπα εκτίμησης καμπυλών μέσω παλινδρόμησης.

Πρότυπα: linear, logarithmic, inverse, quadratic, cubic, power, compound, S-curve, logistic, growth, and exponential

### Αποθήκευση καμπυλών εκτίμησης

**Αποθήκευση μεταβλητών.** Για κάθε επιλεγμένο πρότυπο, μπορούμε να σώσουμε τις προβλεφθείσες τιμές, τα υπόλοιπα (παρατηρηθείσα αξία της εξαρτώμενης μεταβλητής μείον την πρότυπη προβλεφθείσα αξία), και τα διαστήματα πρόβλεψης (ανώτερα και χαμηλότερα όρια). Τα νέα ονόματα των μεταβλητών και οι περιγραφικές ετικέτες επιδεικνύονται σε έναν πίνακα στο παράθυρο εξόδου (output window).

**Προβλέψη περιπτώσεων.** Εάν επιλέξουμε το χρόνο αντί μιας μεταβλητής στο εν λειτουργία αρχείο στοιχείων, ως ανεξάρτητη μεταβλητή τότε μπορούμε να καθορίσουμε μια περίοδο πρόβλεψης πέρα από το τέλος της χρονοσειράς. Μπορούμε να επιλέξουμε μια από τις ακόλουθες εναλλακτικές λύσεις:

- **Πρόβλεψη από την περίοδο εκτίμησης μέσω της τελευταίας περίπτωσης.** Προβλέπει τις τιμές για όλες τις περιπτώσεις στο αρχείο, βασισμένο στις περιπτώσεις της περιόδου εκτίμησης.
- **Απευθείας πρόβλεψη.** Προβλέπει τις τιμές μέσω της καθορισμένης ημερομηνίας, του χρόνου, ή του αριθμού παρατήρησης, βασισμένο στις περιπτώσεις της περιόδου εκτίμησης. Αυτή μπορεί να χρησιμοποιηθεί για να προβλέψει τις τιμές πέρα από την τελευταία περίπτωση στη χρονοσειρά.

### 3.3.7. Παρουσίαση του SAS

Ο SAS Forecasting Server είναι μια μεγάλης κλίμακας λύση αυτόματης πρόβλεψης, η οποία προσφέρει απεριόριστη εξελιξιμότητα. Επιτρέπει την αυτόματη διάγνωση και τη στατιστική πρόβλεψη εκατομμυρίων δεδομένων, ομαδικά ή μέσω του γραφικού περιβάλλοντος του χρήστη. Για κάθε στοιχείο για το οποίο γίνεται πρόβλεψη, ο SAS Forecasting Server επιλέγει αυτόματα το

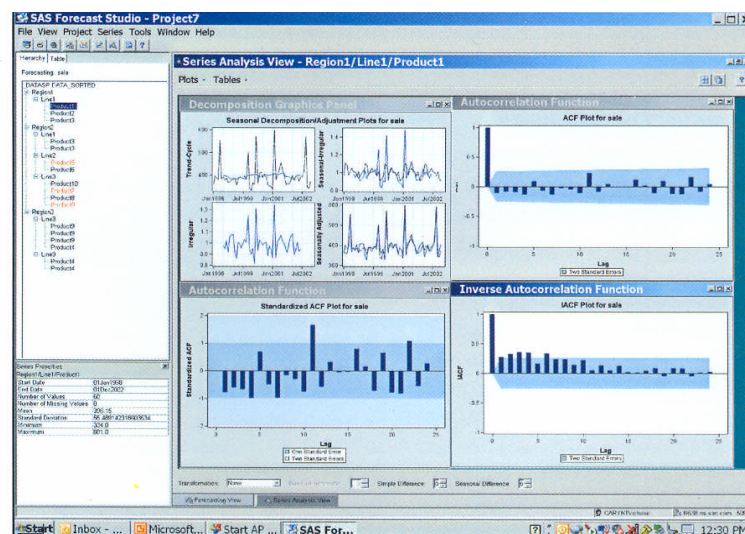
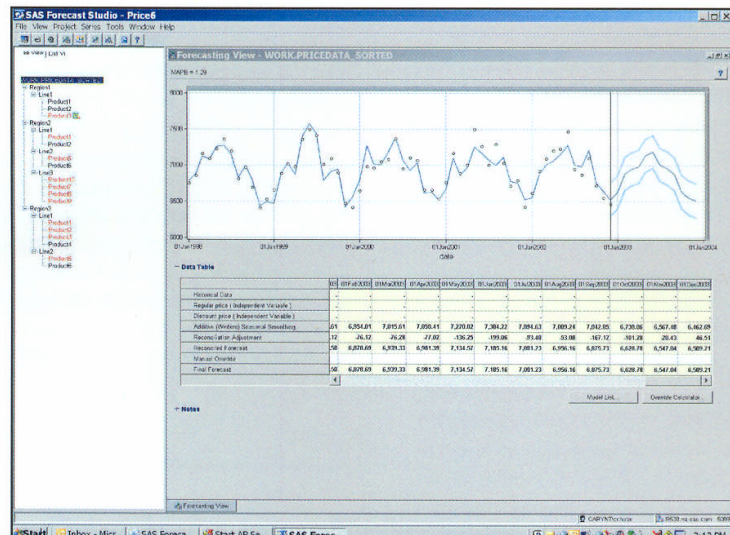


## ΜΕΡΟΣ 2ο – ΚΕΦΑΛΑΙΟ 3ο

καλύτερο μοντέλο πρόβλεψης από μια απεριόριστη λίστα υποψηφίων μοντέλων, βελτιστοποιεί από μαθηματική άποψη όλες τις παραμέτρους των μοντέλων πρόβλεψης και παράγει υψηλής ποιότητας προβλέψεις.

### Αυτόματη πρόβλεψη

Ο SAS Forecasting Server παράγει τις προβλέψεις καθορίζοντας με έξυπνο τρόπο ποιο μοντέλο πρόβλεψης ταιριάζει καλύτερα στα προηγούμενα δεδομένα. Για κάθε δεδομένο παράγεται το καταλληλότερο μοντέλο, με βάση τα κριτήρια που καθορίζονται από το χρήστη. Οι παράμετροι του μοντέλου βελτιστοποιούνται αυτόματα για να παρέχουν το μοντέλο που ταιριάζει περισσότερο, έτσι ώστε να έχουμε τις πιο ακριβείς προβλέψεις. Τα δείγματα που λαμβάνονται μπορούν να διευκρινιστούν έτσι ώστε τα μοντέλα πρόβλεψης να επιλέγονται όχι μόνο με βάση πόσο καλά αποδίδουν τα προηγούμενα δεδομένα, αλλά και από το πόσο καλά είναι πιθανό να προβλέψουν το μέλλον.



Ο SAS Forecasting Server παρέχει ένα γρήγορο και εύκολο τρόπο να δούμε και να προσαρμόσουμε τις προβλέψεις μέσω ενός νέου γραφικού περιβάλλοντος (SAS Forecast Studio).

### 3.3. 8 Παρουσίαση του mySAP SCM

Το mySAP SCM είναι ένα λογισμικό για τη συνεργασία, τον προγραμματισμό, την εκτέλεση, και το συντονισμό μέσα στο προσαρμοστικό δίκτυο αλυσίδων ανεφοδιασμού. Είναι λογισμικό που μπορεί να βοηθήσει την επιχείρησή μας να μετατρέψει μια γραμμική αλυσίδα ανεφοδιασμού σε ένα προσαρμοστικό δίκτυο αλυσίδων ανεφοδιασμού, στο οποίο οι βασισμένες στη ζήτηση και στους πελάτες επιχειρήσεις, μοιράζονται τη γνώση, προσαρμόζονται έξυπνα στις μεταβαλλόμενες συνθήκες της αγοράς και ανταποκρίνονται σε πιο σύντομους και λιγότερο προβλέψιμους κύκλους ζωής.

Πολλές επιχειρήσεις σε όλο τον κόσμο έχουν χρησιμοποιήσει την εφαρμογή mySAP SCM για να βελτιώσουν τις εργασίες τους και τις λειτουργικές τους διαδικασίες. Στην πραγματικότητα, η εφαρμογή mySAP SCM είναι η μόνη εφαρμογή λογισμικού αλυσίδων ανεφοδιασμού που ενσωματώνει τη συνεργασία, τον προγραμματισμό, την εκτέλεση, και το συντονισμό του ολόκληρου δικτύου αλυσίδων ανεφοδιασμού.

Αποτελεί μια εφαρμογή που μας δίνει τη δυνατότητα να :

- Συγχρονίσουμε τον ανεφοδιασμό με τη ζήτηση.
- Κατανοήσουμε και να ανταποκριθούμε σε ένα προσαρμοστικό δίκτυο αλυσίδων ανεφοδιασμού .
- Έχουμε συνεργασία και ανάλυση μέσω διαδικτύου.

Βασισμένο στην πλατφόρμα SAP NetWeaver, το mySAP SCM επιτρέπει στα προσαρμοστικά δίκτυα αλυσίδων ανεφοδιασμού, με την παροχή όχι μόνο δυνατοτήτων προγραμματισμού και εκτέλεσης, να εκτελούν επιχειρηματικές διαδικασίες.

Το αποτέλεσμα που επιτυγχάνεται είναι: μετρήσιμες βελτιώσεις μέσω των μειώσεων δαπανών, αύξηση του επιπέδου παρεχομένων υπηρεσιών, κέρδη παραγωγικότητας – τα οποία τελικά οδηγούν σε μεγαλύτερα περιθώρια κέρδους.

Με το mySAP SCM, μπορούμε να βελτιστοποιήσουμε μια πλήρη σειρά της διαδικασίας προγραμματισμού δραστηριοτήτων, όπως:

- Προγραμματισμός Ζήτησης και Πρόβλεψης
- Προγραμματισμός Αποθεμάτων Ασφάλειας

## ΜΕΡΟΣ 2ο – ΚΕΦΑΛΑΙΟ 3ο

- Προγραμματισμός Δικτύων Ανεφοδιασμού
- Προγραμματισμός Διανομής
- Συνεργασία Δικτύων Ανεφοδιασμού

Demand & Supply Planning	Demand Planning & Forecasting	Safety Stock Planning	Supply Network Planning	Distribution Planning	Supply Network Collaboration
Service Parts Planning	Parts Demand Planning	Parts Inventory Planning	Parts Supply Planning	Parts Distribution Planning	Parts Monitoring
Procurement	Strategic Sourcing	Purchase Order Processing		Invoicing	
Manufacturing	Production Planning & Detailed Scheduling		Manufacturing Operations		
Warehousing	Inbound Processing & Receipt Confirmation	Outbound Processing	Cross-Docking	Warehousing & Storage	Physical Inventory
Order Fulfillment	Sales Order Processing		Billing		
Transportation	Transportation Planning	Transportation Execution		Freight Costing	
SC Design & Analytics	Strategic Supply Chain Design		Supply Chain Analytics		

SAP NetWeaver®

### 3. 3. 9 Παρουσίαση του MicrosoftForecaster

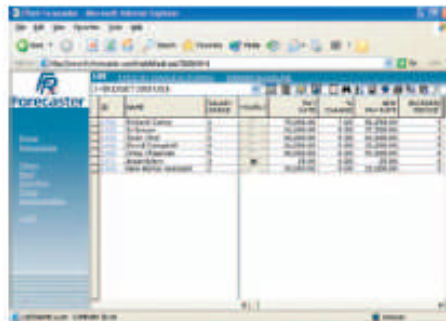
Το Microsoft Forecaster είναι μια εφαρμογή σύνταξης προϋπολογισμού και προγραμματισμού βασισμένη στο διαδίκτυο (web-based), το οποίο μπορεί να βοηθήσει την επιχείρησή μας να ελέγξει τους πόρους χρηματοδότησής της, να περικόψει τις δαπάνες, και να αποτρέψει τις φορολογικές εκπλήξεις. Εξοπλισμένη με το Microsoft Forecaster, η διοικητική ομάδα της επιχείρησης μπορεί να προγραμματίσει και να σχεδιάσει το μέλλον αποτελεσματικότερα και ακριβέστερα.

Ο καθορισμός ρεαλιστικών οικονομικών προσδοκιών απαιτεί έγκυρα στοιχεία και δεδομένα τα οποία να συμβαδίζουν με ένα υγιές επιχειρησιακό σχέδιο. Εντούτοις, οι επιχειρησιακοί προγραμματισμοί και προϋπολογισμοί συχνά αποκλίνουν σημαντικά, κυρίως όταν οι επιχειρήσεις δημιουργούν τους προϋπολογισμούς τους βάση αυτού του προηγούμενου έτους.

Με το Microsoft Forecaster, οι διευθυντές πρώτης γραμμής μπορούν να καταρτίσουν προϋπολογισμούς βασισμένους στους συγκεκριμένους στόχους και στην από πρώτο χέρι γνώση τους συνθηκών της αγοράς. Το αποτέλεσμα: βελτιωμένη ακρίβεια και υψηλότερο επίπεδο δέσμευσης στον προϋπολογισμό από τα άτομα που είναι αρμόδια για την εφαρμογή του. Το Microsoft Forecaster καθιστά πολύ ευκολότερη τη ρύθμιση των προβλέψεων και των προϋπολογισμών στις δεδομένες κάθε φορά αλλαγές του επιχειρησιακού κλίματος, επιτρέποντας στους διευθυντές να ανταποκρίνονται στα καθήκοντά τους αποτελεσματικότερα.

Το σημερινό δυναμικό επιχειρησιακό περιβάλλον απαιτεί την ευέλικτη κατάρτιση προϋπολογισμών καθώς και γρήγορες και εύκολες ικανότητες αναθεώρησης. Το Microsoft Forecaster μας επιτρέπει να διανείμουμε αβίαστα τα στοιχεία σύνταξης προϋπολογισμού και τις οικονομικές πληροφορίες σε όλη την επιχείρησή μας, να συνεργαστούμε με τη διαχείριση, να προγραμματίσουμε αποτελεσματικότερα και να επιτύχουμε τους επιχειρησιακούς στόχους μας.

Παράδειγμα γραφημάτων



Τα λεπτομερή αρχεία για κάθε τμήμα της αποζημίωσης υπαλλήλων αποθηκεύονται για να καταστήσουν εύκολη την προβολή αυτών των σημαντικών πληροφοριών στο επίπεδο απολογισμού.

Οι οθόνες εισαγωγής μπορούν να προσαρμοστούν για να χρησιμοποιήσουν τα χρονικά διαστήματα και τους υπολογισμούς που καθορίζουμε, και μπορούμε να καθορίσουμε τους τύπους για τους υπολογισμούς και να τους ορίσουμε σε οποιοδήποτε γραμμή, στήλη ή κελί.

### 3.3.10 Παρουσίαση του JD Edwards EnterpriseOne Demand Forecasting and Advanced Forecast Modeling (Oracle)

Οι εφαρμογές JD Edwards EnterpriseOne Demand Forecasting Forecasting and Advanced Forecast Modeling της Oracle, μας επιτρέπουν να αναπτύξουμε αποτελεσματικά μια βασισμένη σε στατιστικά στοιχεία πρόβλεψη για να προβλέψουμε και να προγραμματίσουμε μελλοντικές απαιτήσεις μας. Μπορούμε να χρησιμοποιήσουμε τα αρχεία των πωλήσεών μας ή άλλους τύπους δεδομένων σε χρονοσειρές για να παράξουμε τις στατιστικές προβλέψεις. Τα σύνθετα σχέδια απαίτησης, οι τάσεις, η εποχιακότητα, η ζήτηση, και οι μετατοπίσεις ζήτησης προσδιορίζονται με τη χρησιμοποίηση διαφορετικών στατιστικών προτύπων, όπως εκθετική εξομάλυνση, ARIMA, και Croston's intermittent. Η ειδική μηχανή πραγματοποίησης προβλέψεων αναλύει και συνδυάζει τα διάφορα πρότυπα για να παραγάγει μια πρόβλεψη που απεικονίζει καλύτερα την επιχείρησή μας. Και ένα βασισμένο σε γραφικά περιβάλλον χρήστη μας δίνει την ευελιξία να απεικονίσουμε τις πληροφορίες μας με διάφορους τρόπους ώστε να γίνουν κατανοητά καλύτερα τα βασικά ζητήματα.

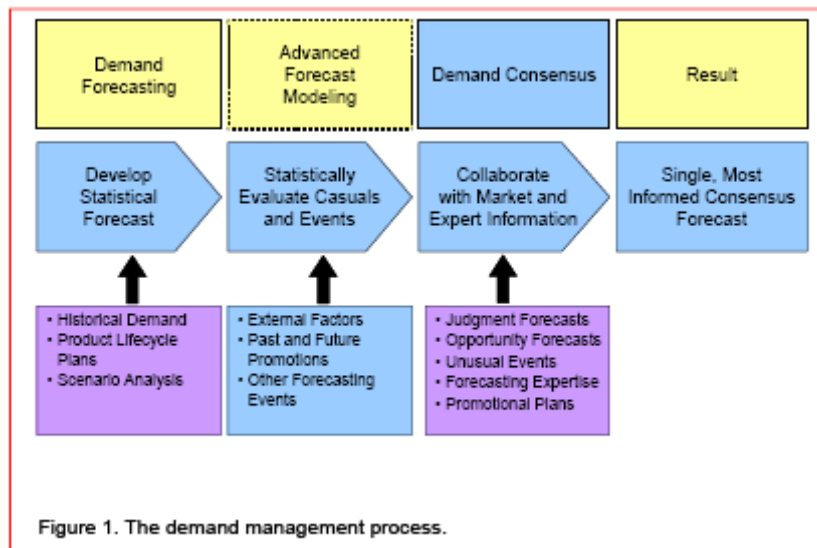
#### Εύκαμπτη Μοντελοποίηση (Flexible Modeling)

Με την εφαρμογή JD Edwards EnterpriseOne Demand Forecasting, έχουμε τη δυνατότητα να αναλύσουμε εύκολα και να διαμορφώσουμε την επιχείρησή μας σε τρέχουσα βάση. Η ιεραρχική δομή δεδομένων μας επιτρέπει να οργανώσουμε και να εξασφαλίσουμε τα στοιχεία μας για να εγκαταστήσουμε την επιχείρησή μας. Παραδείγματος χάριν, καθορίζουμε τέτοιες βασικές σχέσεις όπως τις ιεραρχίες προϊόντων, τις γεωγραφικές οργανώσεις, τους τύπους πελατών, και τα δίκτυα πώλησης.

Καθώς η δομή της επιχείρησής μας αλλάζει, μπορούμε να προσαρμόσουμε το πρότυπο γρήγορα και εύκολα για να προσθέσουμε νέα προϊόντα, να αλλάξουμε εδάφη, να αναδομήσουμε τους

Σχήματισμούς ομάδας προϊόντων, και να προσθέσουμε ή να αλλάξουμε τους συνεργάτες δικτύων. Ο προγραμματισμός κύκλου της ζωής των προϊόντων, μας επιτρέπει να προγραμματίσουμε εύκολα τη ζήτηση για τα νέα προϊόντα ενώ άλλα προϊόντα καταργούνται σταδιακά. Ένας ισχυρός προγραμματισμός μας αφήνει να επιλέξουμε ένα ευρύ φάσμα "κάδων" πρόβλεψης που απαιτούνται για την επιχείρησή μας: εβδομαδιαίος, μηνιαίος, τριμηνιαίος, για τα διάφορα φορολογικά τέταρτα, και ετησίως. Και για να συστήσει τα βέλτιστα επίπεδα αποθεμάτων ασφάλειας, το σύστημα χρησιμοποιεί τη στατιστική διαφορά στην απαίτησή μας από κοινού με τη χρονική ανοχή και το επιθυμητό επίπεδο υπηρεσιών.

Με την ανάλυση σεναρίου, μπορούμε να στρέψουμε το χρόνο και την προσοχή μας στα συγκεκριμένα μέρη της επιχείρησής μας, δημιουργώντας διαφορετικές προβλέψεις για να απεικονίσουμε διαφορετικά σενάρια. Εάν, παραδείγματος χάριν, μπήκαμε μόλις σε μια νέα αγορά, μπορεί να θελήσουμε να δημιουργήσουμε τα αισιόδοξα και απαισιόδοξα σενάρια για τη συγκριτική ανάλυση πριν δεσμεύουμε σε μια ορισμένη πρόβλεψη.



### Πρόβλεψη Ζήτησης με την Advanced Forecast Modeling

Για σύνθετη, πολύπλευρη πρόβλεψη, πρόσθετη λειτουργικότητα είναι διαθέσιμη μέσω μιας συμπληρωματικής προσφοράς προϊόντων. Με την JD Edwards EnterpriseOne Advanced Forecast Modeling, οι προηγούμενες προωθήσεις και άλλα γεγονότα πρόβλεψης είναι στατιστικά προσδιορισμένα και αναλυμένα ώστε να καθορίσουν τον αντίκτυπό τους στις απαιτήσεις μας. Επιπλέον, πολλαπλοί εξωτερικοί παράγοντες, όπως ο δείκτης δημογραφικών και τιμών κατανάλωσης, μπορούν να αξιολογηθούν. Συνεπώς, μπορούμε καλύτερα να καταλάβουμε την αποτελεσματικότητα των προωθήσεών μας και των εκστρατειών μας να προβλέψουμε τη μελλοντική απαίτηση επιτυχέστερα

### Υψηλότερη Αποδοτικότητα με Βελτιωμένη Ακρίβεια Πρόβλεψης (Higher Profitability with Improved Forecast Accuracy)

Η JD Edwards EnterpriseOne Demand Forecasting μας βοηθά να παράξουμε μια πιο οριστική πρόβλεψη. Με μια ακριβέστερη πρόβλεψη, μπορούμε να έχουμε το σωστό προϊόν σε ετοιμότητα.

Επιπλέον, τοποθετούμαστε καλύτερα ώστε να ανταποκριθούμε στην απροσδόκητη ζήτηση. Και είμαστε σε θέση να αυξήσουμε την ανταπόκριση μειώνοντας τις δαπάνες. Με τη βελτίωση της ακρίβειας πρόβλεψης, μπορούμε να βελτιστοποιήσουμε τα σχέδια παραγωγής και διανομής.

### **Προστιθέμενη αξία μέσω της Συνεργάσιμης πρόβλεψης (Added Value Through Collaborative Forecasting)**

Τα μοντέλα Demand Forecasting, Advanced Forecast Modeling, and Demand Consensus του JD Edwards EnterpriseOne καθιστούν ευχερή τη βασισμένη σε περιορισμούς πρόβλεψη. Είμαστε σε θέση να βελτιστοποιήσουμε τον προγραμματισμό των αλυσίδων ανεφοδιασμού μας με το μοντέλο JD Edwards EnterpriseOne Demand Forecasting and Advanced Forecast Modeling για την παραγωγή στατιστικής πρόβλεψης. Η εφαρμογή JD Edwards EnterpriseOne Demand Forecasting βοηθά στο να μας φέρει πιο κοντά στις απαιτήσεις των πελατών μας, αυξάνοντας τη διαφάνεια κατά μήκος του δικτύου ανεφοδιασμού σας .

### **Κυριότερα σημεία των χαρακτηριστικών γνωρισμάτων /λειτουργιών (Feature/Function Highlights) της εφαρμογής: JD Edwards EnterpriseOne Demand Forecasting**

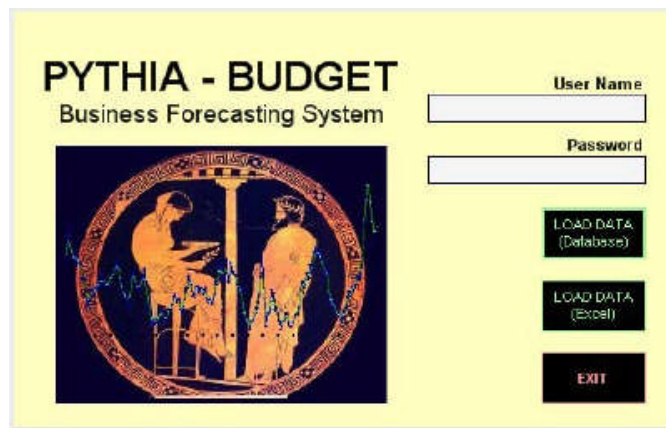
- Πολλαπλά Στατιστικά Μοντέλα (Multiple statistical models).
- Επιλογή εξειδικευμένων μοντέλων (Expert model selection).
- Προγραμματισμός απαίτησης κύκλου της ζωής προϊόντων (Product lifecycle demand planning).
- Βασισμένο σε γραφικά περιβάλλον χρήστη (Graphical user interface).
- Βασισμένο σε στατιστικά στοιχεία απόθεμα ασφάλειας (Statistically based safety stock).
- Ιεραρχική δόμηση δεδομένων (Hierarchical data structuring).
- Εύκαμπτη διαμόρφωση στοιχείων (Flexible data modelling).
- Αντικειμενοστραφής βάση δεδομένων (Object-oriented database).
- Προσομοιώσεις και σενάρια (Simulations and scenarios).
- Προβλέψεις κάτω από περιορισμούς (Constraint-based forecasting).
- Ισχυρός προγραμματισμός εργασιών (Robust calendaring).
- Αναγνώριση των «σκοπέλων» (Outlier identification).

### **Κυριότερα σημεία των χαρακτηριστικών γνωρισμάτων /λειτουργιών (Feature/Function Highlights) της εφαρμογής JD Edwards EnterpriseOne Advanced Forecast Modeling**

- Αιτιώδης ανάλυση εξωτερικών παραγόντων (Causal analysis of external factors)
- Διαχείριση απαιτήσεων προώθησης προϊόντων (Promotions demand management)
- Διαχείριση γεγονότων πρόβλεψης (Forecasting events management)

### 3. 4 Παρουσίαση του ΡΥTHΙΑ

Το λογισμικό ΡΥTHΙΑ ενσωματώνει όλες τις διαθέσιμες γνώσεις και την εμπειρία στον τομέα της πρόβλεψης, χρησιμοποιώντας πλήρως τις νέες ικανότητες των υπολογιστών και των λογισμικών. Το ΡΥTHΙΑ στοχεύει στην άσκηση των διευθυντικών στελεχών(σε επίπεδο οικονομικών διευθυντών, διευθυντών προϊόντων, διευθυντών παραγωγής και αρμόδιων για το σχεδιασμό καταλόγων/αναλυτών) και σχεδιάζεται και αναπτύσσεται έχοντας στο μυαλό ένα συγκεκριμένο σκοπό: να καταστήσει τη διαδικασία της διευθυντικής πρόβλεψης όσο το δυνατόν απλούστερη, φιλικότερη προς το χρήστη και πρακτικότερη, μη συμβιβασόμενη σε θέματα επιστημονικού σθένους και στατιστικής ακρίβειας.



#### Συστατικές Ενότητες

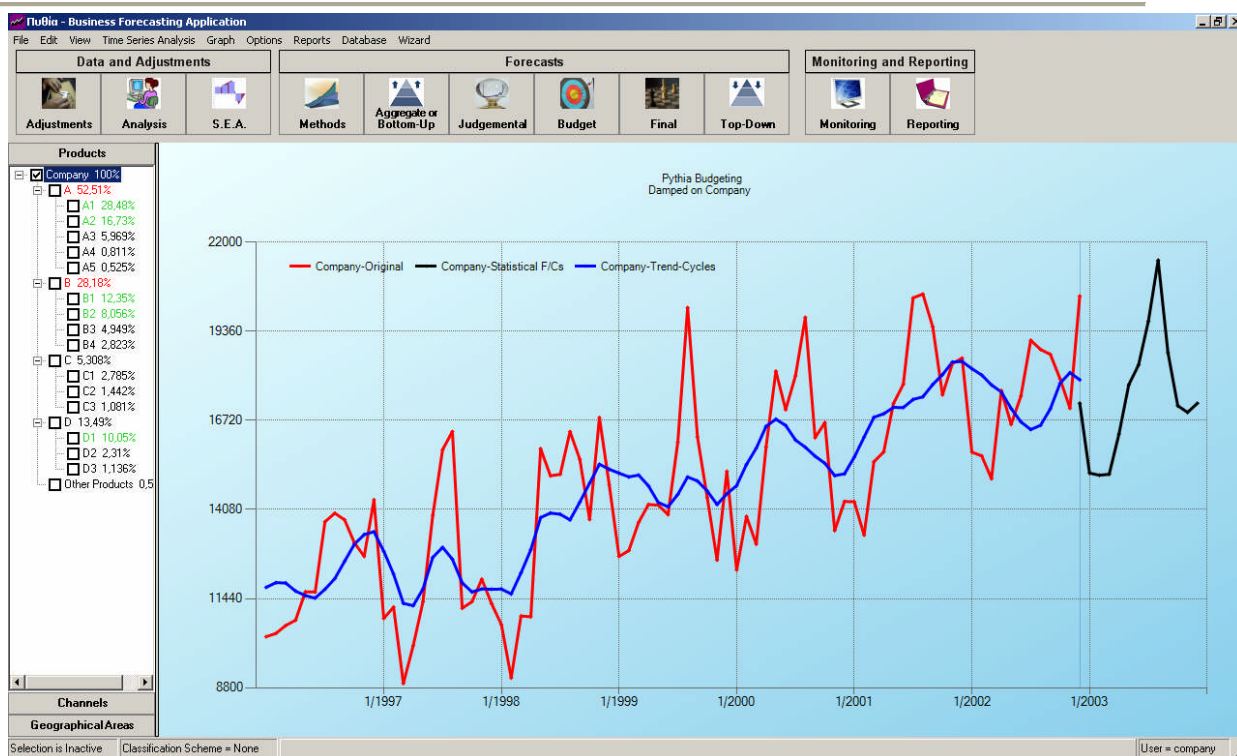
Το ΡΥTHΙΑ αποτελείται από τις ακόλουθες επτά ενότητες, κάθε μια από τις οποίες εστιάζει σε μια ιδιαίτερη πτυχή της ανάλυσης στοιχείων της διευθυντικής πρόβλεψης, του ελέγχου και της υποβολής εκθέσεων:

- I. **Ανάλυση στοιχείων και ρυθμίσεις:** Όλα τα δεδομένα επηρεάζονται από εξωτερικά γεγονότα ή/και ενέργειες (απεργίες, ακραίες καιρικές συνθήκες, μεγάλες μεταβολές των τιμών κ.λπ.) που "αλλοιώνουν" τις προηγούμενες μορφές και τις καθιερωμένες σχέσεις. Για αυτόν τον λόγο πρέπει αυτές να προσαρμοστούν. Η πρώτη ενότητα του ΡΥTHΙΑ συστήνει στο χρήστη τρόπους με τους οποίους προσαρμόζει τα δεδομένα του/της πριν την πρόβλεψη. Επιπλέον, στις συνηθισμένες ρυθμίσεις (εμπορικές συναλλαγές/εργάσιμες ημέρες) το ΡΥTHΙΑ παρέχει επίσης την επιλογή για τα ειδικά γεγονότα/ ενέργειες (Special Events/Actions-SEA) που ασκούν συνήθως μια σημαντική επιρροή στις πωλήσεις/εισοδήματα και έχουν επιπτώσεις στην ακρίβεια και την αξιοπιστία της πρόβλεψης που είναι βασισμένη σε τέτοια δεδομένα. Όταν οι ρυθμίσεις δεδομένων εκτελεστούν, τηρούνται και χρησιμοποιούνται στη συνέχεια από όλες τις ενότητες.
- II. **Προϋπολογισμοί:** Το ΡΥTHΙΑ παρέχει τις προβλέψεις για ολόκληρη την επιχείρηση, για κάθε τμήμα, για χώρα/γεωγραφική περιοχή, για σημαντική κατηγορία προϊόντων, για προϊόντα και στοιχεία. Επιτρέπει επίσης τις προβλέψεις, χρησιμοποιώντας μια από κάτω προς τα επάνω προσέγγιση (bottom-up), ή βασίζοντας τις προβλέψεις του σε οποιοδήποτε



διευκρινισμένο επίπεδο συνάθροισης. Συνεπώς, ο εκάστοτε επικεφαλής μπορεί να τροποποιήσει τις στατιστικές προβλέψεις για να ενσωματώσει κρίσιμες πληροφορίες ή/και στόχους προϋπολογισμών. Μόλις οριστικοποιηθούν οι ρυθμιζόμενες προβλέψεις, μπορούν να διατεθούν αναλογικά σε όλες τις κατώτερες κατηγορίες, ή με κάποιο συγκεκριμένο τρόπο που διευκρινίζεται από το χρήστη.

- III. **Παραγωγή/κατάλογος:** Πολλοί οργανισμοί απαιτούν πολλές χιλιάδες προβλέψεις σε μηνιαία, εβδομαδιαία ή και σε μερικές περιπτώσεις σε καθημερινή βάση. Αυτές οι προβλέψεις πρέπει να παραχθούν αυτόματα, αποτελεσματικά και όσο το δυνατόν ακριβέστερα. Το ΡΥΤΗΙΑ παρέχει τέτοιες προβλέψεις σε μηχανική μορφή, όταν και όπου απαιτούνται. Παρέχει επίσης μια επιλογή για να επιλυθεί η διαφορά μεταξύ των από κάτω προς τα επάνω προβλέψεων και των συνολικών προβλέψεων, συμπεριλαμβανομένων των κρίσιμων ή/και ρυθμιζόμενων στόχων, και διαθέτει τις πιθανές διαφορές αναλογικά ή σύμφωνα με τις προτιμήσεις του χρήστη.
- IV. **Μακροπρόθεσμες Προβλέψεις:** Η πρόβλεψη των μακροπρόθεσμων τάσεων και ο υπολογισμός των ποσοστών αύξησης απαιτούν διαφορετικές μεθόδους και αρκετά περισσότερες υποθέσεις όπως περιβαλλοντικές, ανταγωνιστικές και τεχνολογικές αλλαγές, οι οποίες έχουν επιπτώσεις στις μακροπρόθεσμες τάσεις. Αυτή η ενότητα παρέχει πληροφορίες και συμβουλές για την παραγωγή των μακροπρόθεσμων προβλέψεων παρουσιάζοντας τις πληροφορίες για την ακρίβεια και την αξιοπιστία τους.
- V. **Υπολογισμός των σχέσεων:** Οι ανωτέρω τρεις ενότητες εξετάζουν τη πρόβλεψη χρονοσειρών ενώ η παρούσα ενότητα επιτρέπει την εκτίμηση των σχέσεων βασισμένων στην πολλαπλή παλινδρόμηση.
- VI. **Έλεγχος:** Η ενότητα ελέγχου ελέγχει συνεχώς τη διαφορά μεταξύ των πραγματικών και προβλεφθέντων τιμών και παρέχει ένα σήμα προειδοποίησης όταν τέτοιες διαφορές (δηλ., τα λάθη πρόβλεψης) παύουν να είναι τυχαίες. Σε αυτή την περίπτωση ο χρήστης μπορεί να επαναπροβλέψει ή/και να λάβει κατάλληλα μέτρα, εάν είναι απαραίτητο, για να διορθώσει την κατάσταση.
- VII. **Υποβολή έκθεσης:** Το ΡΥΤΗΙΑ παρέχει λεπτομερείς πληροφορίες εκθέσεως για όλες τις πτυχές της ανάλυσης δεδομένων και της πρόβλεψης. Οι εκθέσεις μπορούν να προσαρμοστούν και να εξαχθούν σε ένα φύλλο Excel για περαιτέρω χρήση.



### Χαρακτηριστικά (και καινοτομίες) του ΡΥΤΗΙΑ

Το ΡΥΤΗΙΑ έχει ως σκοπό και κατασκευάστηκε για να ικανοποιήσει τέσσερις στόχους. Κατ' αρχήν, οποιοσδήποτε διευθυντής, χωρίς τεχνικό/στατιστικό υπόβαθρο, πρέπει να είναι σε θέση να το χρησιμοποιήσει εύκολα και διαισθητικά. Δεύτερον, δεδομένου ότι σχεδόν όλα τα δεδομένα περιέχουν "συστηματικό θόρυβο" από εξωτερικές πηγές, ο χρήστης πρέπει να είναι σε θέση να προσδιορίσει και να ρυθμίσει τα δεδομένα του/της κατά τρόπο απλό. Τρίτον, οι πληροφορίες για τα δεδομένα και τις προβλέψεις πρέπει να είναι εύκολα διαθέσιμες στο χρήστη που πρέπει, επιπλέον, να είναι σε θέση να εισάγει τη γνώση και την εμπειρία του/της σε τέτοιες προβλέψεις με την ενσωμάτωση της κριτικής γνώσης, προκειμένου να βελτιωθεί η ακρίβεια των στατιστικών μεθόδων. Τέλος, οι ειδικοί στις προβλέψεις είναι συχνά υπεραισιόδοξοι ενώ συμπεριλαμβάνουν τις προκαταλήψεις τους στις προβλέψεις τους. Στους ειδικούς στις προβλέψεις, επομένως, δίνονται οι συγκριτικές πληροφορίες για την ακρίβεια των προβλέψεών τους, σε σύγκριση με τις στατιστικές, προκειμένου να προσδιοριστούν οι συστηματικές προκαταλήψεις και αυτές να διορθωθούν στο μέλλον όταν θα προβλέψουν πάλι.

Οι ανωτέρω τέσσερις στόχοι επιτυγχάνονται μέσω των ακόλουθων διακριτικών χαρακτηριστικών που έχουν ενσωματωθεί σε ΡΥΤΗΙΑ:

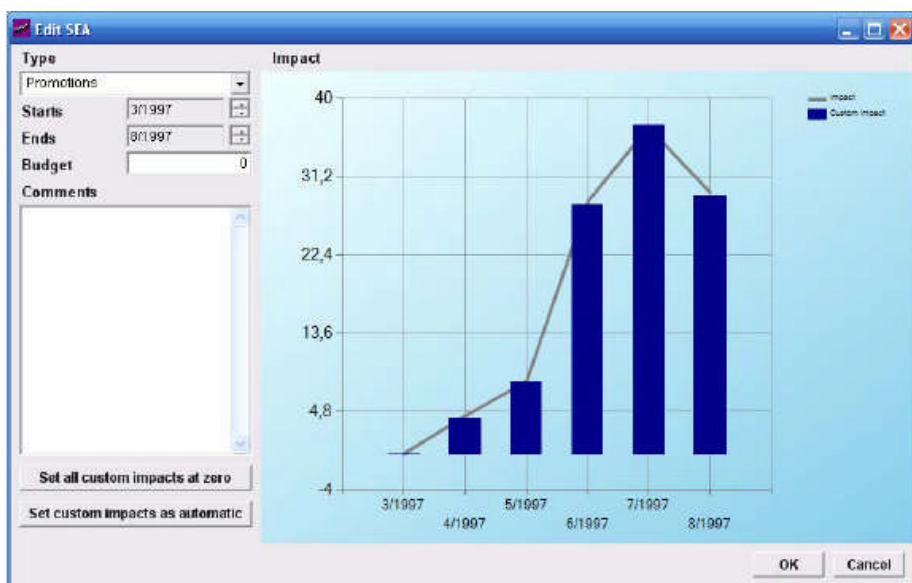
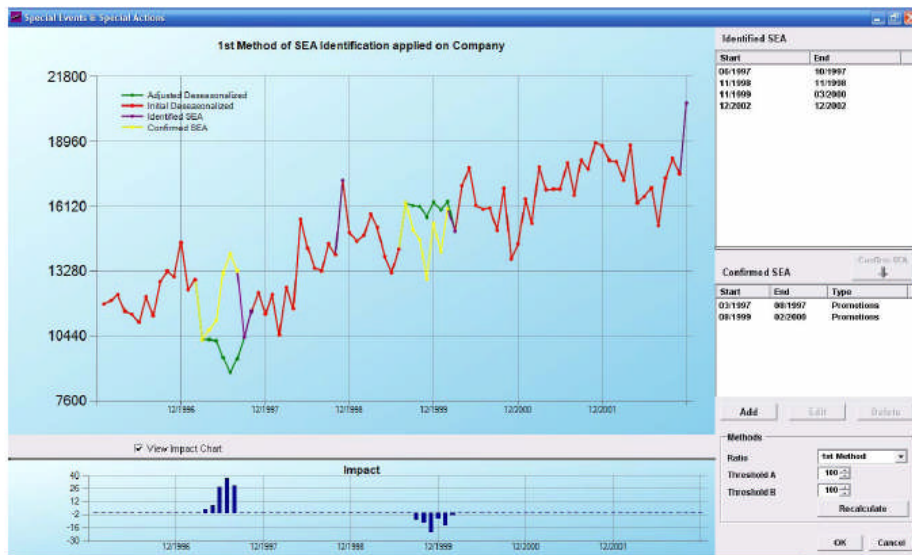
#### 1. Ανάλυση στοιχείων

Σημαντικές πληροφορίες για τη σειρά δεδομένων και τα χαρακτηριστικά τους παρέχονται πάντα από το ΡΥΤΗΙΑ. Τέτοιες πληροφορίες που περιλαμβάνουν, μεταξύ άλλων τα αρχικά και τα ρυθμισμένα στοιχεία, τις αποεποχικοποιημένες τιμές, τη τιμή τάσης-κύκλου, τα

ποσοστά αύξησης, το στατιστικό μοντέλο και τις προβλέψεις, παρουσιάζονται σε γραφικό ή/και σε μια μορφή υπολογισμών με λογιστικό φύλλο (spreadsheet) και μπορούν να εξεταστούν στην οθόνη, ή να αποθηκευτούν για μελλοντικές χρήσεις

## 2. Ρυθμίσεις Δεδομένων

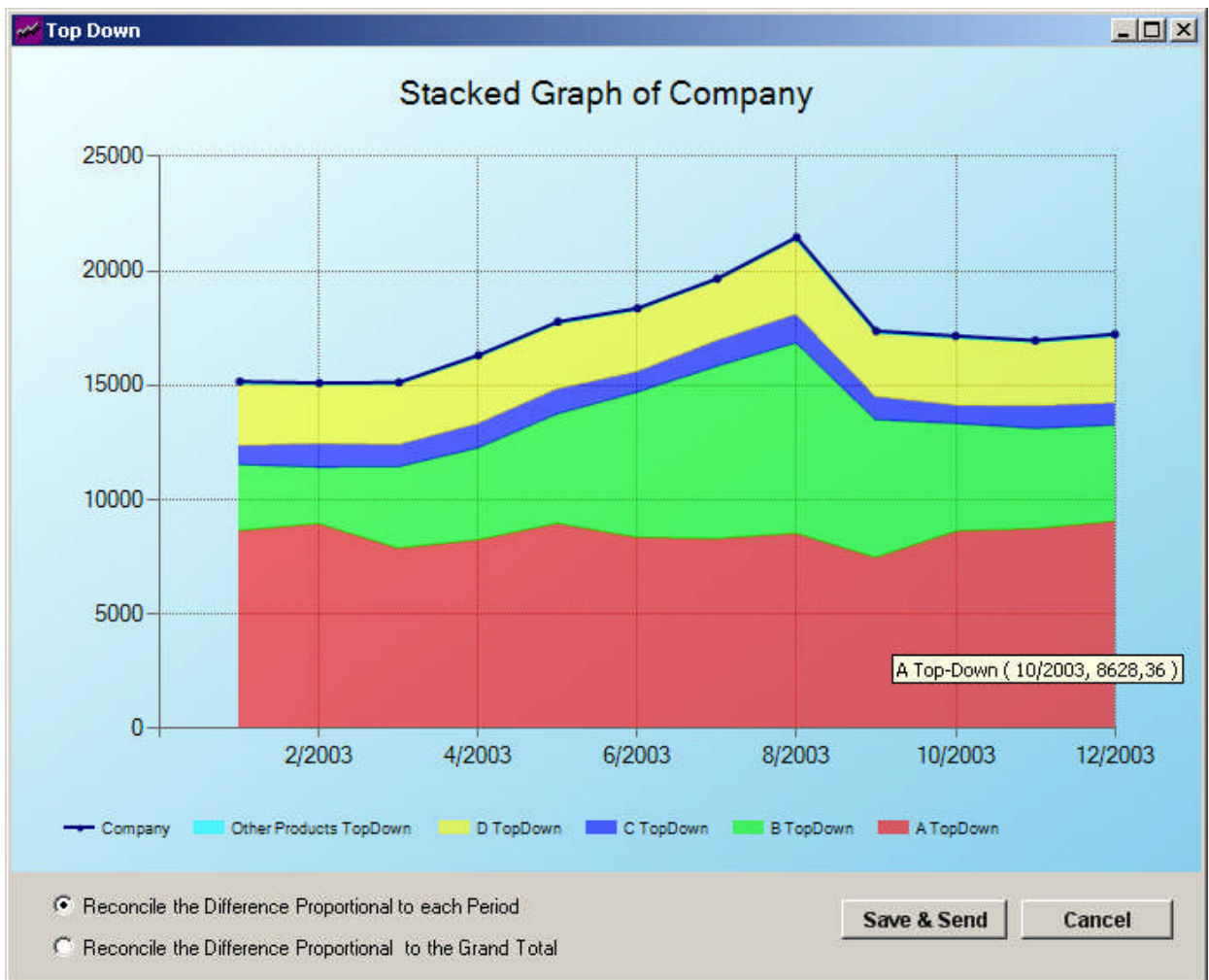
Υπάρχουν διαθέσιμες διάφορες ρυθμίσεις που επιτρέπουν στο χρήστη "να καθαρίσει" τα δεδομένα από τις μη-συστηματικές επιρροές που διαστρεβλώνουν τη μορφή τους. Παραδείγματος χάριν τα ειδικά γεγονότα (SEA- διαφημιστικές εκστρατείες, κακοκαιρίες, απεργίες κ.λπ.) λόγω της κρίσιμης σημασίας τους προσδιορίζονται και τονίζονται από το πρόγραμμα έτσι ώστε ο χρήστης μπορεί να τα καταχωρήσει για μελλοντική χρήση καθώς επίσης και να διορθώσει τα δεδομένα από την επιρροή τους πριν κάνει πρόβλεψη.



**3. Τροποποίηση των στατιστικών προβλέψεων με τις εισαγωγές από την γνώση των προσεχών ειδικών γεγονότων/ ενεργειών (SEA)**

Οι στατιστικές προβλέψεις είναι βασισμένες στην παρέκταση των προηγούμενων σχεδίων, ή των καθιερωμένων σχέσεων. Εντούτοις, τα μελλοντικά αποτελέσματα επηρεάζονται από τα SEA (προωθητικές ή διαφημιστικές εκστρατείες, απεργίες, αυξήσεις τιμών/μειώσεις, ανταγωνιστικές ενέργειες κ.λπ...) όπως και τις διαθέσιμες για διαχείριση εσωτερικές πληροφορίες. Τα αποτελέσματα αυτών των παραγόντων πρέπει να υπολογιστούν κριτικά και να ενσωματωθούν στις στατιστικές προβλέψεις. Το ΡΥΤΗΙΑ επιτρέπει στο χρήστη να υπολογίσει την πιθανή επιρροή των SEA με τη μελέτη των προηγούμενων παρόμοιων γεγονότων/των ενεργειών και να ρυθμίσει στη συνέχεια κατάλληλα τις στατιστικές προβλέψεις.

**4. Συνάθροιση των από κάτω προς τα επάνω (bottom-up) έναντι των από επάνω προς τα κάτω(top-down) προβλέψεων**



Το ΡΥΤΗΙΑ αφήνει το χρήστη να κάνει τις από κάτω προς τα επάνω προβλέψεις με την προσθήκη προβλέψεων σε οποιοδήποτε επιθυμητό επίπεδο συνάθροισης. Επιπλέον υπολογίζει άμεσα τις προβλέψεις στο επιθυμητό επίπεδο συνάθροισης. Εφόσον οι δύο προβλέψεις (από

κάτω προς τα επάνω και από επάνω προς τα κάτω) μπορούν να διαφέρουν, συμβιβάζει τη διαφορά και την διαθέτει κατάλληλα, ή όπως διευκρινίζεται από το χρήστη, στα διάφορα επίπεδα υπό-συνάθροισης. Ο χρήστης μπορεί να έχει πρόσβαση μόνο στα προϊόντα/τις υπηρεσίες στο επίπεδο συνάθροισης του, ή κάτω από αυτό, εφόσον έχει προμηθευτεί τη κατάλληλη άδεια. Μόνο οι διευθυντές, με τις απαραίτητες άδειες, θα μπορούσαν να έχουν πρόσβαση στο πιο υψηλό επίπεδο συνάθροισης (δηλ. προβλέψεις επιχείρησης) και σε όλα τα κατωτέρω επίπεδα.

### **5. Κατανομή των τελικών διοικητικών προβλέψεων σε μια μορφή από επάνω προς τα κάτω.**

Η διοίκηση σε διάφορα επίπεδα, αποφασίζει για το επίπεδο των τελικών προβλέψεων που εφαρμόζονται σε όλη την επιχείρηση. Στην απόφασή της λαμβάνει υπόψη διάφορους παράγοντες συμπεριλαμβανομένου του επιθυμητού ποσοστού αύξησης για ολόκληρη επιχείρηση, τα τμήματα ή τις γεωγραφικές περιοχές. Όταν έχουν συμφωνηθεί οι τελικές προβλέψεις για ολόκληρη την επιχείρηση, ή τα τμήματα κ.λπ., το ΡΥΤΗΙΑ τις διαθέτει σε όλα τα κατωτέρω επίπεδα σε μια ανάλογη ή προκαθορισμένη μορφή. Τέτοια κατανομή μπορεί να εφαρμοστεί σε οποιαδήποτε από τις ακόλουθες ταξινομήσεις:

- α. σημαντικά προϊόντα/υπηρεσίες (και κατώτερες κατηγορίες)
- β. κανάλια/βασικοί πελάτες
- γ. γεωγραφικές περιοχές
- δ. συνδυασμοί των ανωτέρω

3.5 Σύγκριση επικρατέστερων λογισμικών με το ΡΥΘΙΑ

Για τα παραπάνω 10 λογισμικά, παραθέτουμε συγκεντρωτικά τα χαρακτηριστικά τους, προς διευκόλυνση μας στην σύγκρισή τους.

**ΠΙΝΑΚΑΣ 3.1 : ΑΠΑΙΤΗΣΕΙΣ ΣΕ ΥΛΙΚΟ/ΛΟΓΙΣΜΙΚΟ (HARDWARE/SOFTWARE REQUIREMENTS) ΤΩΝ ΕΠΙΚΡΑΤΕΣΤΕΡΩΝ ΛΟΓΙΣΜΙΚΩΝ**

Α/Α	Λογισμικό	Απαιτήσεις Υλικού		Άλλα Λειτουργικά Συστήματα				
		RAM	Χώρος Σκληρού Δίσκου για Εγκατάσταση	DOS	Windows	OS/2	UNIX	MacIntoch
1	Autobox	512 MB	2 MB	Y	Y			
2	Forecast Pro Standard Edition	16 MB	15 MB	Y	Y			
3	Statgraphics Centurion	64 MB	65 MB	Y	Y			
4	Minitab	64 MB	85 MB	Y	Y		Y	Y
5	EViews	64 MB	60.5 MB		Y			
6	SAS Software	32 MB	110 MB	Y	Y	Y	Y	Y
7	SPSS for Windows	16 MB	55 MB		Y		Y	Y
8	SAP SCM	*	*		Y			
9	Microsoft Forecaster	128 MB	100 MB		Y			
10	JD Edwards EnterpriseOne Demand Forecasting	*	*	*	*	*	*	*
11	ΡΥΘΙΑ	128 MB	100 MB		Y			

(Y) Χαρακτηριστικά που πληρεί το λογισμικό

(\*) Στοιχεία που δεν αναφέρονται στις ιστοσελίδες των εταιρειών ή δεν απαντήθηκαν στην επιστολή του Παρτήματος “Α”

**ΠΙΝΑΚΑΣ 3.2: ΚΙΝΗΤΟΙ ΜΕΣΟΙ ΟΡΟΙ-ΕΚΘΕΤΙΚΗ ΕΞΟΜΑΛΥΝΣΗ (MOVING AVERAGES-EXPONENTIAL SMOOTHING) ΤΩΝ ΕΠΙΚΡΑΤΕΣΤΕΡΩΝ ΛΟΓΙΣΜΙΚΩΝ**

Α/Α	Λογισμικό	Κινητοί Μέσοι Όροι	Εκθετική Εξομάλυνση				
			Single	Holt's	Winter's	Damped	Other
1	Autobox	Y	Y	Y	Y	Y	
2	Forecast Pro Standard Edition	Y	Y	Y	Y		
3	Statgraphics Centurion standard	Y	Y	Y	Y		Brown's, Quadratic
4	Minitab	Y	Y		Y		Double
5	Eviews	Y	Y	Y	Y		
6	SAS Software	Y	Y	Y	Y		Brown's, Double, triple, Holt two-parameter Support for up to three levels of seasonality for the Holt-Winters method: time-of year, day-of-week, or time-of day
7	SPSS for Windows	*	*	*	*	*	*
8	SAP SCM	*	*	*	*	*	*
9	Microsoft Forecaster	*	*	*	*	*	*
10	JD Edwards EnterpriseOne Demand Forecasting		Y				Croston's method
11	PYTHIA		Y	Y	Y	Y	

(Y) Χαρακτηριστικά που πληρεί το λογισμικό

(\*) Στοιχεία που δεν αναφέρονται στις ιστοσελίδες των εταιρειών ή δεν απαντήθηκαν στην επιστολή του Παρτήματος "Α"

**ΠΙΝΑΚΑΣ 3.3: ΜΕΘΟΔΟΙ ΑΠΟΣΥΝΘΕΣΗΣ (DECOMPOSITION METHODS) ΤΩΝ  
ΕΠΙΚΡΑΤΕΣΤΕΡΩΝ ΛΟΓΙΣΜΙΚΩΝ**

Α/Α	Λογισμικό	Μέθοδοι Αποσύνθεσης		
		Classical	Census X-11	Other
1	Autobox	Y		
2	Forecast Pro Standard Edition	*	*	*
3	Statgraphics Centurion professional	Y		
4	Minitab	Y		Trend Analysis
5	Eviews		Y	
6	SAS Software	Y	Y	Decomposes raw and transformed series variables and displays the seasonally adjusted series, the trend-cycle component, the seasonal component, or the irregular component
7	SPSS for Windows	*	*	*
8	SAP SCM	*	*	*
9	Microsoft Forecaster	*	*	*
10	JD Edwards EnterpriseOne Demand Forecasting	*	*	*
11	PYTHIA	Y		

(Y) Χαρακτηριστικά που πληρεί το λογισμικό

(\*) Στοιχεία που δεν αναφέρονται στις ιστοσελίδες των εταιρειών ή δεν απαντήθηκαν στην επιστολή του Παρτήματος “Α”



**ΠΙΝΑΚΑΣ 3.4 : ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ-ΟΙΚΟΝΟΜΕΤΡΙΚΑ ΜΟΝΤΕΛΑ (NEURAL NETWORKS-ECONOMETRIC MODELS) ΤΩΝ ΕΠΙΚΡΑΤΗΣΤΕΡΩΝ ΛΟΓΙΣΜΙΚΩΝ**

Α/Α	Λογισμικό	Τεχνητά Νευρωνικά Δίκτυα	Box-Jenkins (ARIMA) ( number of integrations)	Οικονομικά Μοντέλα	
				Single Equation	Multi- Equation
1	Autobox		Y	Y	
2	Forecast Pro Standard Edition			Y	
3	Statgraphics Centurion standard		Y		
4	Minitab		Y		
5	Eviews			Y	Y
6	SAS Software	Y	Y	Y	Y
7	SPSS for Windows	*	*	*	*
8	SAP SCM	*	*	*	*
9	Microsoft Forecaster	*	*	*	*
10	JD Edwards EnterpriseOne Demand Forecasting		Y		
11	PYTHIA	*	*	*	*

(Y) Χαρακτηριστικά που πληρεί το λογισμικό

(\* ) Στοιχεία που δεν αναφέρονται στις ιστοσελίδες των εταιρειών ή δεν απαντήθηκαν στην επιστολή του Παρτήματος “Α”

**ΠΙΝΑΚΑΣ 3.5 : ΜΕΘΟΔΟΙ ΠΑΛΙΝΔΡΟΜΗΣΗΣ (REGRESSION METHODS) ΤΩΝ  
ΕΠΙΚΡΑΤΕΣΤΕΡΩΝ ΛΟΓΙΣΜΙΚΩΝ**

Α/Α	Λογισμικό	Μέθοδοι Παλινδρόμησης		
		Simple	Multiple	Other
1	Autobox	Y	Y	Y
2	Forecast Pro Standard Edition	*	*	*
3	Statgraphics Centurion standard			Ridge Adaptive Poisson Switching Stepwise
4	Minitab	Y	Y	Stepwise ANOVA GLM Nested ANOVA MANOVA Best Subset Partial least square Logistic regression
5	Eviews			Poisson Truncated
6	SAS Software			Ridge Poisson Stepwise Dynamic models Stepwise autoregressive models with constant, linear, or quadratic trend and autoregressive errors to any order
7	SPSS for Windows			Ridge Poisson
8	SAP SCM	*	*	*
9	Microsoft Forecaster	*	*	*
10	JD Edwards EnterpriseOne Demand Forecasting	*	*	*
11	PYTHIA	Y	Y	

(Y) Χαρακτηριστικά που πληρεί το λογισμικό

(\* ) Στοιχεία που δεν αναφέρονται στις ιστοσελίδες των εταιρειών ή δεν απαντήθηκαν στην επιστολή του Παρτήματος “Α”

**ΠΙΝΑΚΑΣ 3.6: ΕΙΔΙΚΕΣ ΛΕΙΤΟΥΡΓΙΕΣ (SPECIAL FEATURES) ΤΩΝ  
ΕΠΙΚΡΑΤΕΣΤΕΡΩΝ ΛΟΓΙΣΜΙΚΩΝ**

Α/ Α	Λογισμικό (1)	Ειδικά Γεγονότα και Συμβόνα (Special Events And Actions) (2)		Λειτουργία με χρήση της διαδικασίας bottom-up και top-down (3)	Περιβάλλον Πολλαπλών Χρηστών (Multi-user interface) (4)	Σύγκριση των αποτελε- σμάτων διαφόρων χρηστών (5)	Ανατροφοδότηση των αποτελεσμάτων και παρακολούθηση (Feedback of the results and monitoring) (6)	Λειτουργία λογισμικού μέσω διαδικτύου (7)
		(α)	(β)					
1	Autobox	Y	Y			Y		Y
2	Forecast Pro Standard Edition	*	*	*	*	*	*	*
3	Statgraphics Centurion standard	*	*	*	*	*	*	*
4	Minitab	*	*	*	*	*	*	*
5	Eviews	*	*	*	*	*	*	*
6	SAS Software	Y	Y	Y	Y		Y	
7	SPSS for Windows	*	*	*	*	*	*	*
8	SAP SCM	*	*	*	*	*	*	*
9	Microsoft Forecaster			Y	Y		Y	Y
10	JD Edwards EnterpriseO ne Demand Forecasting	Y						
11	PYTHIA	Y	Y	Y	Y	Y	Y	Y

(Y) Χαρακτηριστικά που πληρεί το λογισμικό

(\*) Στοιχεία που δεν αναφέρονται στις ιστοσελίδες των εταιρειών ή δεν απαντήθηκαν στην επιστολή του Παρτήματος “Α”

Λαμβάνοντας υπόψη τα στοιχεία που παρέχουν οι παραπάνω πίνακες συγκρίνουμε τα χαρακτηριστικά και τις δυνατότητες του λογισμικού ΡΥΘΙΑ με τα 10 λογισμικά που περιγράφηκαν παραπάνω:

- Χρησιμοποιούμενες μέθοδοι προβλέψεων από τα λογισμικά

1. Μοντέλα Κινητών Μέσων Όρων

Το ΡΥΘΙΑ σε αντίθεση με τα υπόλοιπα λογισμικά δεν εφαρμόζει τις μεθόδους κινητών μέσων όρων.

2. Εκθετική Εξομάλυνση

Πέραν των μεθόδων Single, Winter's και Holt's, τις οποίες εφαρμόζουν τα περισσότερα λογισμικά, μόνο το ΡΥΘΙΑ και το Autobox χρησιμοποιούν τη μέθοδο της φθίνουσας εκθετικής εξομάλυνσης (Damped).

3. Μέθοδοι Αποσύνθεσης

Το ΡΥΘΙΑ ενώ εφαρμόζει τη κλασσική μέθοδο αποσύνθεσης, δεν χρησιμοποιεί τη μέθοδο Census X-11.

4. Μέθοδοι Παλινδρόμησης

Το ΡΥΘΙΑ εφαρμόζει τόσο την απλή όσο και την πολλαπλή μέθοδο παλινδρόμησης, αλλά δεν χρησιμοποιεί άλλες μεθόδους παλινδρόμησης όπως, η Stepwise, η Ridge, η Adaptive, η Poisson, η Switching, η Logistic κ.α.

5. Προηγμένες Μέθοδοι Πρόβλεψης

Η παρούσα έκδοση του ΡΥΘΙΑ δεν εφαρμόζει προηγμένες μεθόδους πρόβλεψης, όπως τα μοντέλα Box-Jenkins ARIMA (θα τοποθετηθούν στην επόμενη έκδοση του λογισμικού), τα νευρωνικά δίκτυα ή τα οικονομετρικά μοντέλα (Single Equation, Multi-Equation).

- Διάφορες Δυνατότητες

Όσον αφορά τα χαρακτηριστικά του πίνακα 3.6 (χειρισμός Ειδικών Γεγονότων και Συμβάντων, λειτουργία του λογισμικού με χρήση της διαδικασίας bottom-up και top-down, ύπαρξη περιβάλλοντος πολλών χρηστών, σύγκριση των αποτελεσμάτων διαφόρων χρηστών, ανατροφοδότηση των αποτελεσμάτων και παρακολούθηση, λειτουργία λογισμικού μέσω διαδικτύου), μόνο τα λογισμικά ΡΥΘΙΑ και SAS τα εφαρμόζουν στο σύνολό τους.

- Απαιτήσεις σε Υλικό/Λογισμικό (Hardware/Software requirements)

Το ΡΥΤΗΙΑ μπορεί να λειτουργήσει μόνο στο περιβάλλον των Windows ενώ τα περισσότερα από τα υπόλοιπα λογισμικά χρησιμοποιούν εναλλακτικά τουλάχιστον ένα ακόμα λειτουργικό σύστημα (π.χ OS/2, UNIX, MacIntosh).

### 3.6 Προοπτική Εξέλιξης-Επέκτασης της Έρευνας

Σαν ευρύτερη προοπτική εξέλιξης της έρευνας που επιχειρείται στο 2<sup>ο</sup> μέρος, προτείνεται η περαιτέρω διερεύνηση των δυνατοτήτων των 10 επικρατέστερων λογισμικών, κάνοντας πια, χρήση των λογισμικών επί συγκεκριμένων χρονοσειρών και μελετώντας τα τεχνικά εγχειρίδιά τους (technical manuals και demos), καταλήγοντας έτσι σε αναλυτική κριτική παρουσίασή τους, σύμφωνα με ανάλογες αναφορές σχετικών παλαιότερων δημοσιεύσεων ([4]).

Τέλος να σημειώσουμε ότι θα μπορούσε, πέρα από ποιοτική ανάλυση των συγκριτικών πινάκων, να γίνει και ποσοτική ανάλυσή τους, καταλήγοντας στην εξαγωγή στατιστικών στοιχείων για τις δυνατότητες των λογισμικών. Κάτι τέτοιο στάθηκε αδύνατο να γίνει στη παρούσα εργασία, διότι η έλλειψη σημαντικού αριθμού χαρακτηριστικών θα οδηγούσε σε εσφαλμένα συμπεράσματα.

## **Βιβλιογραφία 2ου Μέρους**

- [1] Robert S. Rycroft , Microcomputer Software of Interest to Forecasters in Comparative Review: Updated Again, International Journal of Forecasting , 2000
- [2] Spyros Makridakis, Michele Hibon, The M-3 Competition: results, conclusions and implications, International Journal of Forecasting, (16),451-476, 2000
- [3] Robert S. Rycroft , Student Editions of Forecasting Software: A Survey, International Journal of Forecasting ,11, p.p 337-351, 1995
- [4] Leonard J. Tashman, Michael L. Leach, Automatic Forecasting Software: A survey and Evaluation, International Journal of Forecasting, vol.7, p.p 209-230, 1991
- [5] Ramesh Shadra , Joseph F. Rock, Forecasting Software for Microcomputers, International Journal of Forecasting ,vol.13, p.p 197-209, 1986

## **Πηγές από το διαδίκτυο**

- [6] <http://www.google.com>
- [7] <http://www.in.com>
- [8] <http://www.sciencedirect.com>
- [9] <http://www.robhyndman.info/forecasting/>
- [10] <http://www.neural-forecasting.com>
- [11] <http://www.forecastingeducation.com>
- [12] <http://www.informs.org/>
- [13] <http://www.timberlake.pt>

- [14] <http://www.modler.com>
- [15] <http://www.andersonbell.com>
- [16] <http://www.Aptech.com>
- [17] <http://www.autobox.com>
- [18] <http://www.forecastpro.com>
- [19] <http://www.delphus.com>
- [20] <http://www.limdep.com>
- [21] <http://www.estima.com>
- [22] <http://www.mpassociates.gr>
- [23] <http://www.lmt-expo.com>
- [24] <http://www.insightful.com>
- [25] <http://www.mathworks.com/finprod>
- [26] <http://www.ncss.com>
- [27] <http://www.minitab.com>
- [28] <http://www.eviews.com>
- [29] <http://www.sas.com>
- [30] <http://www.scausa.com>
- [31] <http://www.statsoft.com>

- [32] <http://www.spss.com>
- [33] <http://www.smartcorp.com>
- [34] <http://www.mpassociates.gr>
- [35] <http://www.spss.com>
- [36] <http://www.texasoft.com>
- [37] <http://www.tspintl.com>
- [38] <http://www.sap.com>
- [39] <http://www.frxsoftware.com>
- [40] <http://www.oracle.com>
- [41] <http://fsu.ece.ntua.gr>



## ΠΑΡΑΡΤΗΜΑ “Α”

### Επιστολή προς τις Εταιρείες Λογισμικών



**National Technical University of Athens**  
**Department of Electrical and Computer Engineering**

---

Electrical Power Division, Forecasting Systems unit

9, Iroon Polytechniou Str. 15773 Zografou

Athens, Greece

Tel.: +3210 7723637, 7723738 - Fax: +3210 7723740

<http://fsu.ece.ntua.gr>

Athens, 5/22/2006

Dear Mr/Mrs,

We are researchers of National Technical University of Athens in Greece and we are working on an academic project with title “Automatic forecasting software: A survey and evaluation”.

The goals of the research are:

- (a) to clarify for the practitioner the virtues and limitations of automatic forecasting softwares,
- (b) to assess whether the software encourages if not nurtures good forecasting practice in the identification, evaluation, and defense of a forecasting model.
- (c) to evaluate which softwares correspond to the needs and the purposes of our laboratory, in order to buy them.

The research team includes PhD and Master of Science researchers under the supervision of the professor Mr. V. Assimakopoulos.

In order to carry out my research we'll need some information about the features of your product. We are particularly interested in the features mentioned in the tables given at the attached document. You are kindly requested to mark the features that correspond to your product.

Moreover, we would be grateful if you send us a demo and the documentation of your product (if they are free of charge) .

If you want to contact us please use our email, [el01727@mail.ntua.gr](mailto:el01727@mail.ntua.gr), [el01723@mail.ntua.gr](mailto:el01723@mail.ntua.gr)

We are looking forward to hearing from you.

Best Regards

Vasileios Soukos  
Dimitrios Kantzios  
Forecasting Systems Unit, N.T.U.A.

**ΠΑΡΑΡΤΗΜΑ “Β”**  
**Στοιχεία που Ζητήθηκαν από τις Εταιρείες Λογισμικών**

**1. Selected Forecasting Techniques (\*)**

Software	Moving Average Models	Exponential Smoothing				
		Single	Holt's	Winter's	Damped	Other

(\*) please mark with Y for positive answer

Software	Regression Methods		
	Simple	Multiple	Other

(\*) please mark with Y for positive answer

Software	Decomposition Methods		
	Classical	Census X-11	Other

(\*) please mark with Y for positive answer

Software	Artificial Neural Networks	Econometric Models		Box-Jenkins (ARIMA) (**)
		Single Equation	Multi-Equation	

(\*) please mark with Y for positive answer

(\*\*) please note up to how to many integrations your product makes

**2. Hardware/Software Requirements (\*)**

Software	Hardware Requirements		Other PC Operating Systems				
	RAM	Hard disk space for normal installation	DOS	Windows	OS/2	UNIX	MacIntoch

(\*) please mark with Y for positive answer

### 3. Miscellaneous Capabilities (\*)

Software	Price	Site license	Limited or student version available	Price for Limited or student version	On-line Help	On-line tutorials	Technical Support

(\*) please mark with Y for positive answer

Software (1)	Special Events And Actions (2)		Data management using the bottom-up and top-down procedure (3)	Multi-user interface (4)	Comparison of users' results (5)	Feedback of the results and monitoring (6)	Operation via internet (7)
	(a)	(b)					

(\*) please mark with Y for positive answer

(1) Product name

(2) Special Event and Special Action (SEA) are the following incidents:

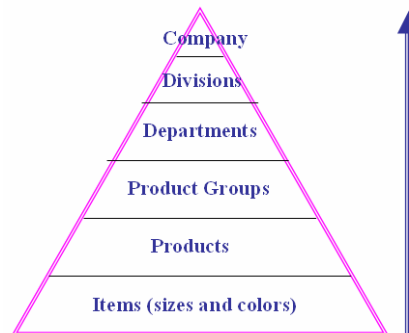
- Promotions (Amount/Type)
- Advertising (Amount/Media)
- Price Increases
- Price Decreases
- Substitute Products: Price Increases
- Substitute Products: Price Decreases
- New Products (Performance/Price Ratio)
- Insufficient Inventories
- Insufficient Inventories of Competitors
- Competitive Promotions (Amount/Type)
- Competitive Advertising (Amount/Media)
- Competitive Price Increases
- Competitive Price Decreases
- Competitive Substitute Products: Price Increases
- Competitive Substitute Products: Price Decreases
- New Products Introduced by Competitors (Performance/Price Ratio)
- Unusual Weather Conditions
- Strikes
- Strikes in Competitors

a. Is your product able to recognize the special events?

b. Is your product able to store the special events and their impact and use them in the smoothing procedure?

(3) "Bottom – Up" procedure:

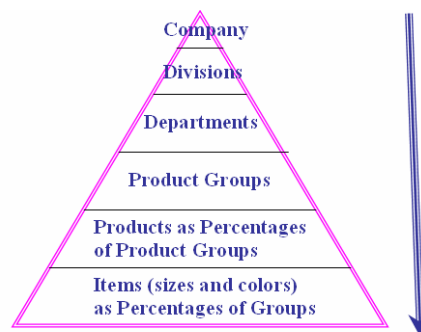
Users can view only the forecasts of the level in which they belong, as well as the forecasts of products or categories that belong in the inferior rung of pyramid (hierarchy of users in the enterprise). Users are able to view (Aggregating lower level forecasts to higher level ones up to determining total company forecasts)



"Top– Down" procedure:

The user has under his supervision a category of products, which is distinguished in other categories or products. The user is able to send his forecasts in the inferior levels. The inferior levels have to take these forecasts into consideration and adjust their initial forecasts.

(Forecasts are done at aggregate levels starting from the top until product groups and then determining the forecasts below as percentages)



(4) Does your product support a multi-user interface?

(5) Is your product able to compare the results of different users?

(6) Is it possible to feedback the results and re-evaluate them?

(7) Is it possible to operate your software via internet without the need of its installation?