



Πολυτεχνείο Κρήτης

Σχολή Μηχανικών Παραγωγής και Διοίκησης

Μεταπτυχιακό Πρόγραμμα: Σχεδίαση και Παραγωγή Προϊόντων

Πρόβλεψη ζήτησης βιβλίων εκδοτικού οίκου

Διατριβή που υπεβλήθη για την μερική ικανοποίηση των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος Ειδίκευσης

υπό

Καδίτη Κωνσταντίνας

Επιβλέπων Καθηγητής: Βασίλειος Κουϊκόγλου

Χανιά, 2025

Μέλη εξεταστικής επιτροπής:

- 1. Βασίλειος Κουϊκόγλου (επιβλέπων), Καθηγητής**
Πολυτεχνείο Κρήτης
Σχολή Μηχανικών Παραγωγής και Διοίκησης
- 2. Δημήτριος Ιψάκης, Αναπληρωτής Καθηγητής**
Πολυτεχνείο Κρήτης
Σχολή Μηχανικών Παραγωγής και Διοίκησης
- 3. Ευστράτιος Ιωαννίδης, Αναπληρωτής Καθηγητής**
Πολυτεχνείο Κρήτης
Σχολή Μηχανικών Παραγωγής και Διοίκησης

Περίληψη

Πολλοί εκδοτικοί οίκοι ανατυπώνουν τα βιβλία σε μικρές ποσότητες για να καλύψουν τη ζήτηση. Επειδή οι απαιτήσεις των πελατών είναι συνεχώς αυξανόμενες και δεδομένων των οικονομικών περιορισμών, είναι σημαντικός ο προγραμματισμός του αριθμού ανατυπώσεων ώστε να αποφεύγονται οι ελλείψεις και συγχρόνως να διατηρούνται χαμηλά αποθέματα. Η παρούσα μεταπτυχιακή εργασία ασχολείται με τον προγραμματισμό έκδοσης βιβλίων για έναν εγχώριο εκδοτικό οίκο. Το μεγαλύτερο μερίδιο των πωλήσεων αφορούν σε βιβλία που απευθύνονται σε φοιτητές μέσω του συστήματος «Εύδοξος» καθώς και μέσω διαδικτυακών βιβλιοπωλείων και φυσικών καταστημάτων. Ο προγραμματισμός ανατυπώσεων γίνεται με βάση την πρόβλεψη ζήτησης από δεδομένα που διαθέτει ο εκδοτικός οίκος και λαμβάνοντας υπ' όψη τα τρέχοντα αποθέματα και τις επιστροφές. Στην εργασία αυτή συγκρίνονται διάφορες μέθοδοι πρόβλεψης με κριτήριο την ελαχιστοποίηση μίας συνάρτησης των σφαλμάτων πρόβλεψης. Από την σύγκριση, θα επιλεγούν τα καλύτερα μοντέλα για κάθε τίτλο βιβλίου και οι προβλέψεις τους θα χρησιμοποιηθούν για την έκδοση εντολών ανατύπωσης με βάση τα τρέχοντα αποθέματα.

Abstract

Many publishers reprint books in small quantities to meet demand. Given the continuously increasing customer requirements and financial constraints, it is important to plan the number of reprints to avoid shortages while also keeping inventory levels low. This master's thesis focuses on planning book releases for a domestic publishing house. A large share of sales consists of books for students, distributed through the "Eudoxus" system, as well as through online bookstores and physical shops. The reprint scheduling is based on demand forecasts from data provided by the publisher, considering current inventory levels and returns. In this study, various forecasting methods are compared, aiming to minimize a function of forecasting errors. From this comparison, the best models for each book title will be selected, and their forecasts will be used to generate reprint orders based on current stock levels.

Ευχαριστίες

Θεωρώ πολύ σημαντικό να ευχαριστήσω από καρδιάς τον επιβλέποντα Καθηγητή μου, κ. Βασίλειο Κουϊκόγλου για την αμέριστη υπομονή του, την καθοδήγησή του και για την άμεση απόκρισή του σε όσες απορίες δημιουργήθηκαν κατά την εκπόνηση της εργασίας. Επίσης, θα ήθελα να ευχαριστήσω τον κ. Μούργκο Ιωάννη για την μεγάλη προθυμία του να παρέχει τα δεδομένα πάνω στα οποία βασίστηκε η παρούσα εργασία καθώς χωρίς αυτά, θα ήταν αδύνατο να υλοποιηθεί. Τέλος, οφείλω να ευχαριστήσω την οικογένειά μου, τους φίλους μου και τον σύντροφό μου, Σταύρο, για την απεριόριστη συμπαράσταση και υποστήριξή τους καθ' όλη τη διάρκεια των σπουδών μου.

Περιεχόμενα

Περίληψη	3
Ευχαριστίες	4
1. Εισαγωγή	7
1.1 Αντικείμενο της διατριβής.....	7
1.2 Βιβλιογραφική ανασκόπηση	7
1.3 Δομή της διατριβής.....	8
2. Πρόβλεψη Ζήτησης Βιβλίων	9
2.1 Περιγραφή προβλήματος	9
2.2 Προεπεξεργασία δεδομένων	9
2.3 Ομαδοποίηση δεδομένων.....	10
3. Μεθοδολογία	12
3.1 Λευκός θόρυβος.....	12
3.2 Μέθοδος ARMA	13
3.2.1 ADF Test.....	16
3.3 Εκθετική εξομάλυνση	16
3.3.1 Απλή εκθετική εξομάλυνση	16
3.3.2 Μέθοδος Holt-Winters.....	17
3.4 Προσέγγιση Syntetos-Boylan.....	18
3.5 Απλός κινητός μέσος.....	19
3.6 Διπλή εκθετική εξομάλυνση για ακανόνιστα (ελλιπή) δεδομένα.....	19
3.7 Σφάλματα	21
3.7.1 Μέσο τετραγωνικό σφάλμα (Root mean square error).....	21
4. Υλοποίηση Αλγορίθμων και Διαγράμματα Ροής.....	22
4.1 Εισαγωγή.....	22
4.2 Συνάρτηση ομαδοποίησης	22
4.3 Αλγόριθμος πρόβλεψης με μεθόδους μέσου όρου, ARIMA και Holt-Winters.....	26
4.3.1 Αρχικοποίηση και αποθήκευση	26
4.3.2 Κύριο μέρος αλγορίθμου	26
4.4 Συνάρτηση Holt-Winters.....	32
4.5 Συνάρτηση Syntetos-Boylan	32
4.5.1 Αρχικοποίηση	32
4.5.2 Κύριο μέρος συνάρτησης.....	33
4.6 Αλγόριθμος προσέγγισης Syntetos-Boylan, απλού κινητού μέσου και εκθετικής εξομάλυνσης ..	35
4.6.1 Αρχικοποίηση και αποθήκευση	35
4.6.2 Κύριο μέρος αλγορίθμου	35
4.7 Αλγόριθμος διπλής εκθετικής εξομάλυνσης για ακανόνιστα δεδομένα.....	38
5. Αποτελέσματα	41
5.1 Συνάρτηση ομαδοποίησης.	41
5.2 Ομαδοποίηση 1	42
5.2.1 Μέσος όρος, ARIMA και Holt-Winters.....	42
5.2.2 Απλός κινητός μέσος, προσέγγιση Syntetos-Boylan και απλή εκθετική εξομάλυνση	48
5.2.3 Διπλή εκθετική εξομάλυνση για ακανόνιστα δεδομένα.....	51
5.3 Ομαδοποίηση 2	53
5.3.1 Μέσος όρος, ARIMA και Holt-Winters	53
5.3.2 Απλός κινητός μέσος, προσέγγιση Syntetos-Boylan και απλή εκθετική εξομάλυνση	59
5.3.3 Διπλή εκθετική εξομάλυνση για ακανόνιστα δεδομένα.....	62

6. Εγχειρίδιο χρήσης αλγορίθμων.....	65
6.1 Μέσος όρος, ARIMA και Holt-Winters	65
6.2 Απλός κινητός μέσος, προσέγγιση Syntetos-Boylan και απλή εκθετική εξομάλυνση.....	66
6.3 Διπλή εκθετική εξομάλυνση για ακανόνιστα δεδομένα	66
7. Σύνοψη.....	68
7.1 Συμπεράσματα	68
7.2 Προτάσεις για μελλοντική έρευνα	69
Βιβλιογραφία	70
Παραρτήματα.....	71
Παράρτημα Α: Συνάρτηση ομαδοποίησης	71
Παράρτημα Β: Αλγόριθμος πρόβλεψης με την χρήση των μεθόδων ARIMA, Holt-Winters και μέσης τιμής	73
Παράρτημα Γ: Αλγόριθμος διπλής εκθετικής εξομάλυνσης για ακανόνιστα δεδομένα	77
Παράρτημα Δ: Συνάρτηση Holt-Winters.....	80
Παράρτημα Ε: Αλγόριθμος απλού κινητού μέσου, προσέγγισης Syntetos-Boylan και απλής εκθετικής εξομάλυνσης.....	81

1. Εισαγωγή

1.1 Αντικείμενο της διατριβής

Το θέμα της παρούσας διατριβής επικεντρώνεται στην πρόβλεψη ζήτησης βιβλίων εκδοτικού οίκου που διαθέτει βιβλία μέσω του συστήματος «ΕΥΔΟΞΟΣ», ηλεκτρονικά μέσω της ιστοσελίδας του, καθώς και σε βιβλιοπωλεία (φυσικά καταστήματα). Το μεγαλύτερο μέρος των βιβλίων απευθύνεται στην ακαδημαϊκή κοινότητα. Ειδικότερα, δεδομένου ότι μεγάλος όγκος βιβλίων διατίθεται μέσω του συστήματος «ΕΥΔΟΞΟΣ», η ζήτηση καθορίζεται από τις επιλογές των φοιτητών και από τον χρόνο στον οποίο αυτές δηλώνονται κάθε εξάμηνο. Ως εκ τούτου, παρατηρείται ότι η ζήτηση διαφέρει κάθε μήνα του έτους και κατά συνέπεια, δημιουργούνται περίοδοι υψηλής και χαμηλής ζήτησης μέσα στο ίδιο έτος με μεγάλες διαφορές.

Στην παραπάνω κατάσταση το πρόβλημα εντοπίζεται κυρίως στο γεγονός ότι, ενώ ο εκδοτικός οίκος θα πρέπει να αποστέλλει τον αριθμό των παραγγελιών που δέχεται, θα δεχτεί πίσω έναν αριθμό βιβλίων από επιστροφές. Αυτό πολλές φορές οφείλεται στο γεγονός ότι τα βιβλία δεν παραλαμβάνονται από όλους τους φοιτητές που τα έχουν δηλώσει.

Ο βασικός στόχος της παρούσας μεταπτυχιακής διατριβής είναι η εφαρμογή διαφόρων μεθόδων για την πρόβλεψη της μηνιαίας ζήτησης τριών διαφορετικών βιβλίων και κατά συνέπεια, η ελαχιστοποίηση των σφαλμάτων πρόβλεψης η εύρεση του καλύτερου αλγορίθμου, βάσει των παραπάνω στοιχείων.

Τα βιβλία αυτά θα αναφέρονται από εδώ και στο εξής ως Α, Β και Γ.

Οι μέθοδοι που χρησιμοποιούνται είναι τα μοντέλα ARMA (αυτοπαλίνδρομα κινητού μέσου μοντέλα, Autoregressive Moving Average models), η μέθοδος της μέσης τιμής και η μέθοδος Holt-Winters. Επιπλέον, εξετάζεται σε άλλο αλγόριθμο η προσέγγιση Syntetos-Boylan και συγκρίνεται με την μέθοδο απλού κινητού μέσου και με την μέθοδο εκθετικής εξομάλυνσης. Τέλος, εφαρμόζεται ο αλγόριθμος διπλής εκθετικής εξομάλυνσης για ακανόνιστα δεδομένα με βάση το άρθρο του Thomas Cipro.

Τέλος, εξετάζονται και συγκρίνονται τα σφάλματα καθώς και οι προβλεπόμενες τιμές για κάθε περίπτωση.

1.2 Βιβλιογραφική ανασκόπηση

Η διαθέσιμη διεθνής βιβλιογραφία είναι αρκετά περιορισμένη ως προς την πρόβλεψη ζήτησης για εκδοτικό οίκο και εν γένει για την ζήτηση βιβλίων. Από την αρχή διαπιστώθηκε ότι οι προβλέψεις εξαρτώνται από τον τρόπο ομαδοποίησης των δεδομένων, τα οποία εμφανίζονταν σε ημερήσια καταγραφή σε αρχείο excel. Στη σχετική βιβλιογραφία η ζήτηση αθροίζεται κατά μήνα. Άλλη

ομαδοποίηση δεν εντοπίστηκε που να σχετίζεται με την πρόβλεψη βιβλίων. Στην εργασία αυτή δοκιμάζονται διάφορες ομαδοποιήσεις κατά μήνες (Petrooulos & Kourentzes, 2014) και συγκρίνονται τα σφάλματα κάθε συνδυασμού. Επειδή, όσον αφορά στις μεθόδους πρόβλεψης που εφαρμόζονται, δεν υπάρχει κάποια που να υπερτερεί γενικώς έναντι των άλλων, εξετάζονται και συγκρίνονται τα μέσα τετραγωνικά σφάλματα κάθε ομαδοποίησης, για επτά διαφορετικές μεθόδους οι οποίες εφαρμόζονται συχνότερα στην πρόβλεψη ζήτησης. Για την εφαρμογή των μεθόδων πρόβλεψης αναπτύχθηκαν τρεις αλγόριθμοι:

Ο πρώτος αλγόριθμος περιλαμβάνει τα μοντέλα ARMA, τριπλή εκθετική εξομάλυνση Holt-Winters και τον απλό μέσον όρο ως τις συνηθέστερα χρησιμοποιούμενες μεθόδους (Terence, 2019). Ο δεύτερος αλγόριθμος υλοποιεί την προσέγγιση των Syntetos-Boylan (Syntetos & Boylan, 2005), τη μέθοδο της απλής εκθετικής εξομάλυνσης και τη μέθοδο του κινητού μέσου όρου. Στον τρίτο αλγόριθμο χρησιμοποιήθηκε η μέθοδος της διπλής εκθετικής εξομάλυνσης για ακανόνιστα δεδομένα σύμφωνα με τον Thomas Cipra (2005).

1.3 Δομή της διατριβής

Το υπόλοιπο μέρος της παρούσας διατριβής οργανώνεται στα εξής κεφάλαια:

Στο 2^ο κεφάλαιο περιγράφεται το πρόβλημα της πρόβλεψης βιβλίων. Αναφέρονται οι τρόποι προεπεξεργασίας των δεδομένων και οι επιμέρους ομαδοποιήσεις που επιχειρήθηκαν.

Στο 3^ο κεφάλαιο αναλύονται οι μέθοδοι πρόβλεψης που εφαρμόστηκαν και το θεωρητικό τους υπόβαθρο. Συγκεκριμένα αναλύονται η απλή εκθετική εξομάλυνση, η τριπλή εκθετική εξομάλυνση Holt-Winters, η μέθοδος ARMA, η προσέγγιση Syntetos-Boylan (Syntetos-Boylan approximation, SBA), η διπλή εκθετική εξομάλυνση για ακανόνιστα δεδομένα, και επεξηγούνται έννοιες όπως ο λευκός θόρυβος που επηρεάζουν τα αποτελέσματα.

Στο 4^ο κεφάλαιο επεξηγούνται οι αλγόριθμοι που δημιουργήθηκαν στην MATLAB και για την ευκολότερη και γρηγορότερη κατανόηση παρουσιάζονται και τα διαγράμματα ροής του κάθε αλγορίθμου.

Στο 5^ο κεφάλαιο παρουσιάζονται τα αποτελέσματα, αναλύονται και συγκρίνονται μεταξύ τους με βάση το μικρότερο RMSE.

Στο 6^ο κεφάλαιο δίνονται αναλυτικές οδηγίες για την εκτέλεση των αλγορίθμων, όπως και για την μορφοποίηση των αρχείων που δίνονται ως είσοδοι σε κάθε αλγόριθμο.

Τέλος, στο 7^ο κεφάλαιο περιλαμβάνονται τα συμπεράσματα και προτάσεις για μελλοντική έρευνα.

Επιπλέον, παρατίθεται παράρτημα με τους αλγόριθμους που αναπτύχθηκαν για τις ανάγκες της παρούσας διατριβής.

2. Πρόβλεψη Ζήτησης Βιβλίων

2.1 Περιγραφή προβλήματος

Ο μεγαλύτερος όγκος βιβλίων που διανέμει ο εκδοτικός οίκος πηγαίνει σε φοιτητές μέσω του προγράμματος διανομής συγγραμμάτων «Εύδοξος». Γίνονται ωστόσο και παραγγελίες από ιδιώτες μέσω της ιστοσελίδας του εκδοτικού οίκου, καθώς και από βιβλιοπωλεία που διαθέτουν τα βιβλία. Όσον αφορά στο σύστημα «Εύδοξος», παρατηρείται κάθε χρόνο μία περιοδική μορφή της ζήτησης όσον αφορά στους μηνιαίους όγκους παραγγελιών, καθώς το σύστημα «Εύδοξος» έχει συγκεκριμένους μήνες για κάθε ακαδημαϊκό εξάμηνο τους οποίους επιτρέπεται στους φοιτητές να δηλώσουν τα συγγράμματά τους. Κατά την διάρκεια, αλλά και μετά την ολοκλήρωση των δηλώσεων πραγματοποιείται η αποστολή των παραγγελιών. Οι δηλώσεις γίνονται από τον Οκτώβριο έως και τον Δεκέμβριο για το χειμερινό εξάμηνο και από τον Φεβρουάριο έως και τον Απρίλιο για το εαρινό εξάμηνο. Αντίστοιχα, οι διανομές συνήθως γίνονται από τον Οκτώβριο έως και τον Ιανουάριο για το χειμερινό εξάμηνο και από τον Φεβρουάριο έως και τον Μάιο για το εαρινό εξάμηνο. Ταυτόχρονα με τις διανομές όμως, υπάρχουν και οι επιστροφές συγγραμμάτων που δεν παραλήφθηκαν ποτέ από τους φοιτητές. Οι επιστροφές των συγγραμμάτων από βιβλιοπωλεία προς τον εκδοτικό οίκο γίνονται καθ' όλη την διάρκεια του έτους και είναι ακανόνιστα κατανομημένες. Στην εργασία αυτή εξετάζονται μόνο οι παραγγελίες βιβλίων και όχι οι επιστροφές.

2.2 Προεπεξεργασία δεδομένων

Προκειμένου τα δεδομένα να γίνουν αξιοποιήσιμα, επιβαλλόταν να πραγματοποιηθεί μία προεπεξεργασία. Η αρχική μορφή των δεδομένων σε αρχείο excel περιλάμβανε αρκετές πληροφορίες οι οποίες ήταν αχρείαστες και πιθανότατα θα δημιουργούσαν προβλήματα κατά την άντλησή τους από την MATLAB. Χρησιμοποιήθηκαν τρία διαφορετικά αρχεία excel, ένα για κάθε βιβλίο. Αρχικά αφαιρέθηκαν οι στήλες του excel που αφορούσαν στον κωδικό του βιβλίου, στο όνομα, στην μονάδα μέτρησης των βιβλίων (τεμάχια), στον τύπο και στον κωδικό του παραστατικού, στον συναλλασσόμενο και στον κωδικό της αποθήκης. Οι παραπάνω είναι πληροφορίες που δεν θα χρυσίμευαν κάπου στην παρούσα εργασία. Επομένως, κρατήθηκαν μόνο η ημερομηνία και η ποσότητα. Σε περίπτωση αποστολής, η ποσότητα εμφανίζεται ως θετικός αριθμός, ενώ σε περίπτωση επιστροφής εμφανίζεται ως αρνητικός αριθμός. Για τις επιστροφές, δηλαδή στους αρνητικούς αριθμούς, ο κάθε αρνητικός αριθμός αντικαταστάθηκε με το 0 στην συνάρτηση ομαδοποίησης (αναλύεται στο κεφάλαιο 4.2), προκειμένου να αποφευχθεί κάποιο στατιστικό/αριθμητικό σφάλμα.

Οι συνολικές εγγραφές για το βιβλίο Α ήταν 617 και η πρώτη συναλλαγή έγινε τον Ιανουάριο του 2018. Αντίστοιχα για το Β ήταν 290 με την πρώτη συναλλαγή να γίνεται τον Απρίλιο του 2019, ενώ για το Γ ήταν 316 και η πρώτη συναλλαγή έγινε τον Ιούνιο του 2019. Επειδή τα βιβλία

προέρχονται από έναν μικρό εκδοτικό οίκο, τα δεδομένα αφορούν σε λίγα έτη, με αποτέλεσμα να επηρεάζονται ανάλογα και τα αποτελέσματα, όπως θα αναλυθεί στην συνέχεια.

Από τα δεδομένα χρησιμοποιήθηκαν, όπως αναφέρθηκε, μόνο η ημερομηνία και η ποσότητα με την ονομασία των αντίστοιχων στηλών ως Date και Sales. Σε όλα τα αρχεία δίνονται οι ίδιες ονομασίες για ευκολότερη και μαζικότερη άντληση. Στόχος είναι να εξετάζονται ταυτόχρονα όλα τα βιβλία για καλύτερη αξιοποίηση του υπολογιστικού χρόνου.

2.3 Ομαδοποίηση δεδομένων

Κατά την βιβλιογραφική ανασκόπηση, δεν βρέθηκε κάποιος συνιστώμενος τρόπος ομαδοποίησης, ο οποίος θα ήταν κατάλληλος για το πρόβλημα που πραγματεύεται η συγκεκριμένη διατριβή. Η μοναδική ομαδοποίηση που αναφερόταν σχεδόν σε όλες τις εργασίες ήταν κατά μήνα. Ωστόσο, αυτό που παρατηρήθηκε είναι ότι οι ομαδοποιήσεις που εντοπίστηκαν στη βιβλιογραφία αφορούσαν σε δεδομένα πολλών ετών, πάνω από 10 έτη συνήθως, με αποτέλεσμα να υπάρχει και μεγαλύτερος όγκος δεδομένων. Επιπλέον, οι παραπάνω ομαδοποιήσεις, αφορούσαν σε προϊόντα που δεν παρουσίαζαν κάποια εποχικότητα και κατ' επέκταση οι ερευνητές δεν προσπάθησαν κάποια επιπλέον ομαδοποίηση των μηνών μέσα στο έτος.

Στην συγκεκριμένη περίπτωση, έχοντας ως δεδομένο ότι υπάρχει εποχικότητα, εξαιτίας του ότι το μεγαλύτερο μέρος των παραγγελιών προέρχεται από το σύστημα «Εύδοξος», θεωρήθηκε χρήσιμο αρχικά να γίνει άθροιση των ποσοτήτων κατά μήνα και στη συνέχεια να γίνει μία προσπάθεια ομαδοποίησης των μηνών. Η ομαδοποίηση των μηνών πραγματοποιείται μέσω ενός αρχείου κειμένου (txt) προκειμένου να διευκολύνεται ο εκάστοτε χρήστης.

Για κάθε βιβλίο δημιουργείται ένα διαφορετικό αρχείο κειμένου. Σε κάθε τέτοιο αρχείο, εισάγονται από τον χρήστη οι διαφορετικές ομαδοποιήσεις που θέλει να εξετάσει. Ορίζεται επομένως ο αριθμός των διαφορετικών ομαδοποιήσεων και ο χρήστης περιγράφει ποιους μήνες έχουν οι ομάδες σε κάθε ομαδοποίηση. Οι μήνες συμβολίζονται με ακέραιους αριθμούς από το 1, για τον Ιανουάριο, έως το 12, για τον Δεκέμβριο. Το αρχείο έχει την μορφή της παρακάτω εικόνας (Εικόνα 2.1).

```
#Number of desired grouping alternatives
3
#Στο τέλος κάθε εναλλακτικής επανατοποθετείται ο πρώτος μήνα για να ολοκληρωθεί ο χρόνος.
# First month of groups
1 6 10 1
# First month of groups
1 3 6 1
# First month of groups
1 4 8 11 1
```

Εικόνα 2.1: Μορφοποίηση αρχείου .txt για την ομαδοποίηση

Επομένως, όπως παρατηρείται στην παραπάνω εικόνα, δίνεται αρχικά ο αριθμός των ομαδοποιήσεων για κάθε βιβλίο και στην συνέχεια η κάθε ομαδοποίηση. Πιο συγκεκριμένα, στο συγκεκριμένο παράδειγμα, η πρώτη ομαδοποίηση περιλαμβάνει τους μήνες από Ιανουάριο έως Μάιο για την πρώτη ομάδα, από Ιούνιο έως Σεπτέμβριο για την δεύτερη ομάδα και από Οκτώβριο έως Δεκέμβριο για την τρίτη ομάδα. Η δεύτερη ομαδοποίηση περιλαμβάνει τους μήνες από Ιανουάριο έως Φεβρουάριο για την πρώτη ομάδα, από Μάρτιο έως Μάιο για την δεύτερη ομάδα και από Ιούνιο έως Δεκέμβριο για την τρίτη ομάδα. Τέλος, η τρίτη ομαδοποίηση περιλαμβάνει τους μήνες από Ιανουάριο έως Μάρτιο για την πρώτη ομάδα, από Απρίλιο έως Ιούλιο για την δεύτερη ομάδα, από Αύγουστο έως Οκτώβριο για την τρίτη ομάδα και από Νοέμβριο έως Δεκέμβριο για την τελευταία ομάδα. Επιπλέον, σύμφωνα με τον τρόπο που έχει δομηθεί ο αλγόριθμος, ο τελευταίος μήνας κάθε ομάδας δεν συμπεριλαμβάνεται σε αυτήν, αλλά συμμετέχει στην επόμενη. Αυτό γίνεται προκειμένου να διευκολυνθεί ο αλγόριθμος.

3. Μεθοδολογία

Στο κεφάλαιο αυτό θα αναλυθούν οι μέθοδοι που χρησιμοποιήθηκαν για την ανάλυση των αποτελεσμάτων: η μέθοδος ARMA, η μέθοδος της τριπλής εκθετικής εξομάλυνσης Holt-Winters, η μέθοδος της Εκθετικής Εξομάλυνσης, η προσέγγιση Syntetos-Boylan και η μέθοδος του Απλού Κινητού Μέσου. Οι παραπάνω μέθοδοι εμφανίζονται σε δύο διαφορετικούς αλγορίθμους. Επίσης, εξετάζεται αν οι χρονοσειρές είναι λευκός θόρυβος, περιλαμβάνουν δηλαδή ανεξάρτητες παρατηρήσεις οι οποίες κυμαίνονται γύρω από κάποια μέση τιμή με εντελώς τυχαίο μοτίβο.

Στο εξής θεωρούμε τυχαίες μεταβλητές X , Y για τις οποίες ορίζονται τα μεγέθη (γνωστά ή άγνωστα):

- πιθανότητες $P(X=x)$, $P(Y=y)$, κοινές $P(X=x, Y=y)$ ορίζονται για κάθε $x, y \in (-\infty, \infty)$
- μέση τιμή $\mu_X = E(X) = \dots xP(X=x) + \dots$
- ιδιότητα^(*): $E(aX + bY + \dots) = aE(X) + bE(Y) + \dots$
- διασπορά της X , $\sigma_X^2 = \text{Var}(X) = E[(X - \mu_X)^2] = (\text{ανάπτυγμα και ιδιότητα}^{(*)}) = E(X^2) - \mu_X^2$
- συσχέτιση (correlation) των X και Y , $E(XY) = \dots xyP(X=x, Y=y) + \dots$ για όλα τα πιθανά x, y
- συνδιακύμανση (ή συμμεταβλητότητα, covariance) των X και Y , $\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = (\text{λόγω}^{(*)}) = E(XY) - \mu_X\mu_Y$
- συντελεστής συσχέτισης των X και Y , $\rho_{XY} = \text{Cov}(X, Y) / (\sigma_X\sigma_Y)$

Ανάλογες έννοιες ορίζονται για χρονοσειρές τυχαίων μεγεθών X_t και Y_t οποία παρατηρούμε σε διαφορετικές χρονικές στιγμές t . Για παράδειγμα η χρονοσειρά $X_1, X_2, \dots, X_t, \dots$ μπορεί να δηλώνει τις συνολικές πωλήσεις κάποιου βιβλίου τους μήνες $t = 1, 2, \dots$ και $E(X_t Y_{t-k})$ είναι η αυτοσυσχέτιση τάξης k της X_t .

3.1 Λευκός θόρυβος

Μία χρονοσειρά X_t λέγεται λευκός θόρυβος όταν (Box et al., 2008) οι τιμές της X_{t1} και X_{t2} σε διαφορετικούς χρόνους είναι *ασυσχέτιστες*, δηλαδή η συσχέτισή τους ισούται με το γινόμενο των μέσων τιμών $E(X_{t1}X_{t2}) = E(X_{t1})E(X_{t2})$ όταν $t_1 \neq t_2$. Αυτό είναι ισοδύναμο με τη συνθήκη ότι η αυτοσυμμεταβλητότητά τους είναι 0, αφού $\text{Cov}(X_{t1}, X_{t2}) = \dots = E(X_{t1}X_{t2}) - E(X_{t1})E(X_{t2}) = 0$. Στο εξής οι όροι αυτοσυσχέτιση και αυτοσυμμεταβλητότητα χρησιμοποιούνται όταν εξετάζεται η ίδια χρονοσειρά X_t .

Προκειμένου να εξεταστεί αν μία χρονοσειρά είναι λευκός θόρυβος υπάρχουν δύο εναλλακτικές. Το Ljung-Box test και το γράφημα της συνάρτησης Αυτοσυσχέτισης (ACF Function).

Για τις χρονοσειρές που είναι λευκός θόρυβος κάθε αυτοσυσχέτιση είναι κοντά στο 0.

Οι τύποι για τον ορισμό του συντελεστή αυτοσυσχέτισης και για τον υπολογισμό του δειγματικού συντελεστή αυτοσυσχέτισης είναι οι εξής:

$$\rho_k = \frac{Cov(X_t, X_{t-k})}{Var(X_t)} \text{ και } \hat{\rho}_k = \frac{\sum_{t=k}^n (X_t - \mu)(X_{t-k} - \mu)}{\sum_{t=1}^n (X_t - \mu)^2} \quad \text{Εξ. 3.1}$$

όπου $Cov(X_t, X_{t-k})$ είναι η αυτοσυμμεταβλητότητα (ή αυτοσυνδιακύμανση) τάξεως k . Είδαμε προηγουμένως ότι όταν η Y_t είναι λευκός θόρυβος, ο αριθμητής είναι 0 για κάθε $k \neq 0$. Αυτό ισχύει όταν ο αριθμός των παρατηρήσεων είναι ∞ . Για $n < \infty$, έχουμε εν γένει μη μηδενικά $\hat{\rho}_k$ γι' αυτό αντί της συνθήκης $Q > 0$ χρησιμοποιούμε την $Q > X_{1-a,h}^2$ ως κριτήριο σημαντικότητας.

Στη συνέχεια υπολογίζεται και σχεδιάζεται και το γράφημα ACF. Σύμφωνα με τους Hyndman & Athanasopoulos (2018), για τις χρονοσειρές που είναι λευκός θόρυβος, το 95% των κορυφών του γραφήματος ACF βρίσκονται μέσα στα όρια που δημιουργούνται από τις ευθείες $\pm 2/\sqrt{T}$, όπου T είναι το πλήθος των τιμών της χρονοσειράς. Αν ένα ποσοστό μεγαλύτερο του 5% βρίσκεται εκτός αυτών των ορίων, τότε πιθανότατα η χρονοσειρά δεν είναι λευκός θόρυβος.

3.2 Μέθοδος ARMA

Η μέθοδος ARMA (Autoregressive Moving Average) χρησιμοποιείται για να αναλύσει και να περιγράψει στάσιμες χρονοσειρές χρησιμοποιώντας την αυτοπαλινδρόμηση και την μέθοδο του κινητού μέσου όρου με ένα πολυώνυμο για την κάθε μέθοδο. Το μοντέλο AR (Autoregressive) χρησιμοποιεί τις προηγούμενες τιμές της χρονοσειράς για την πρόβλεψη των επόμενων, ενώ το μοντέλο MA (Moving Average) χρησιμοποιεί τα προηγούμενα σφάλματα. Επομένως το μοντέλο ARMA συνδυάζει τα παραπάνω και πραγματοποιεί την πρόβλεψη. Χρησιμοποιείται κυρίως για την πρόβλεψη των χρονοσειρών και έχει μεγάλο εύρος εφαρμογών στα χρηματιστηριακά και οικονομικά μοντέλα, στην πρόβλεψη καιρού, αλλά και σε πληθώρα διαφορετικών τομέων.

Βασική προϋπόθεση για την εφαρμογή της μεθόδου είναι η χρονοσειρά να είναι αμετάβλητη (στάσιμη) *κατά την ευρεία έννοια*, δηλαδή να έχει

- μέση τιμή
- αυτοσυσχετίσεις και αυτοσυμμεταβλητότητες όλων των τάξεων ($k = \dots -2, -1, 0, 1, 2, \dots$)

που να είναι ανεξάρτητες του χρόνου για κάθε k . Για παράδειγμα

- $E(X_t) = E(X_0) = \mu$
- $Cov(X_{t+k}, X_t) = Cov(X_k, X_0) = \gamma_k$ (λόγω συμμετρίας) $Cov(X_t, X_{t+k}) = Cov(X_0, X_k) = \gamma_{-k}$

Επί πλέον, θεωρείται ότι $\mu = 0$. Αν αυτές οι προϋποθέσεις δεν ισχύουν τότε η X_t μπορεί να προκύψει από την αρχική χρονοσειρά μετά από α) διαφορίσεις οι οποίες φιλτράρουν τα δεδομένα και αφαιρούν την τάση (αύξηση ή μείωση) και την περιοδικότητα και β) αφαίρεση της μέσης τιμής της φιλτραρισμένης χρονοσειράς που έχει προκύψει μετά το βήμα α).

Ο συμβολισμός $AR(p)$, αναφέρεται στο μοντέλο αυτοπαλινδρόμησης τάξης p με εξίσωση:

$$X_t = \delta + \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t \quad \text{Εξ. 3.2}$$

όπου, το X_t είναι η τιμή της χρονοσειράς την στιγμή t , οι παράμετροι $\varphi_1, \dots, \varphi_p$ και το ε_t είναι λευκός θόρυβος.

Οι παράμετροι $\varphi_1, \dots, \varphi_p$, για το μέρος AR υπολογίζονται από την εξίσωση των Yule-Walker σύμφωνα με την οποία, αρχικά εκτιμώνται από τα δεδομένα της χρονοσειράς οι συνδιακυμάνσεις $\gamma_k = \text{Cov}(X_t, X_{t-k})$, για $k = 0, 1, 2, \dots, p$ ή οι δειγματικοί συντελεστές αυτοσυσχέτισης $\hat{\rho}_1, \dots, \hat{\rho}_p$ (με $\hat{\rho}_0 = 1$) και έπειτα επιλύεται το σύστημα εξισώσεων:

$$\begin{bmatrix} \gamma_0 & \cdots & \gamma_{p-1} \\ \vdots & \ddots & \vdots \\ \gamma_{p-1} & \cdots & \gamma_0 \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \vdots \\ \varphi_p \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_p \end{bmatrix} \quad \text{ή} \quad \begin{bmatrix} 1 & \cdots & \hat{\rho}_{p-1} \\ \vdots & \ddots & \vdots \\ \hat{\rho}_{p-1} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \vdots \\ \varphi_p \end{bmatrix} = \begin{bmatrix} \hat{\rho}_1 \\ \vdots \\ \hat{\rho}_p \end{bmatrix} \quad \text{Εξ. 3.3}$$

Ο τελευταίος συντελεστής φ_p της ανωτέρω εξίσωσης λέγεται συντελεστής μερικής αυτοσυσχέτισης (*partial autocorrelation*). Στην περίπτωση μοντέλου τάξης $p+1$, τότε **προκύπτει διαφορετικό σύστημα εξισώσεων και διαφορετικές λύσεις $\varphi_1, \varphi_2, \dots, \varphi_p$ (καθώς και μία ακόμη για τον συντελεστή φ_{p+1})**. Για να μην υπάρχει κίνδυνος σύγχυσης, ο συντελεστής μερικής αυτοσυσχέτισης συμβολίζεται και ως φ_{pp} ώστε να διευκρινιστεί ότι προέκυψε θεωρώντας μοντέλο τάξης p επίσης.

Για $p = 1$, έχουμε $\varphi_{11} = \gamma_1/\gamma_0 = \rho_1$.

Η συνάρτηση μερικής αυτοσυσχέτισης (PACF) χρησιμοποιείται για τον προσδιορισμό της κατάλληλης τάξης p του μοντέλου $AR(p)$. Συγκεκριμένα, μετά την υστέρηση p , οι τιμές της PACF είναι 0 (ή πολύ μικρές).

Το δεύτερο μέρος, MA, του μοντέλου ARMA αφορά στην μέθοδο του κινητού μέσου όρου. Ο συμβολισμός $MA(q)$, αναφέρεται στο μοντέλο κινητού μέσου όρου τάξης q με εξίσωση:

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad \text{Εξ. 3.4}$$

όπου, X_t είναι η τιμή της χρονοσειράς την στιγμή t , $\theta_1, \dots, \theta_q$ οι παράμετροι, μ η αναμενόμενη τιμή της χρονοσειράς (συχνά θεωρείται ίσο με 0 - για στάσιμες χρονοσειρές), και το ε_t , το οποίο είναι λευκός θόρυβος, ονομάζεται «υπόλοιπο» (residual) και είναι η διαφορά της πραγματικής τιμής του παρελθόντος από την εκτίμησή της (Montgomery et al., 2015).

Οι παράμετροι $\theta_1, \dots, \theta_q$ στο μοντέλο MA υπολογίζονται συχνά με την εκτίμηση της μέγιστης πιθανοφάνειας (Maximum Likelihood Estimation).

Η μέση τιμή της X_t είναι: $E(X_t) = \mu$

Η διακύμανσή της είναι: $Var(X_t) = \gamma_y(0) = \sigma^2(1 + \theta_1^2 + \dots + \theta_q^2)$

Αντίστοιχα και η αυτοσυνδιακύμανση με υστέρηση k είναι:

$$\gamma_y(k) = Cov(y_t, y_{t+k}) = \begin{cases} \sigma^2(-\theta_k + \theta_1\theta_{k+1} + \dots + \theta_{q-k}\theta_q), & k = 1, 2, \dots, q \\ 0, & k > q \end{cases}$$

Η συνάρτηση (ή συντελεστής) αυτοσυσχέτισης της διαδικασίας MA(q) για κάθε χρονική καθυστέρηση (lag) είναι:

$$\rho_k = \frac{\gamma_y(k)}{\gamma_y(0)} \quad \text{Εξ. 3.5}$$

Το παραπάνω χαρακτηριστικό χρησιμοποιείται για την αναγνώριση του κατάλληλου μοντέλου MA(q), καθώς μετά την υστέρηση q, παρουσιάζεται αποκοπή στη συνάρτηση αυτοσυσχέτισης ACF (Montgomery et al., 2015). Δηλαδή οι τιμές $\rho_{q+1}, \rho_{q+2}, \dots$ μηδενίζονται (και οι αντίστοιχοι δειγματικοί συντελεστές είναι κοντά στο μηδέν).

Συνοψίζοντας τις δύο παραπάνω εξισώσεις (Εξ. 3.2, Εξ. 3.4) προκύπτει η παρακάτω εξίσωση για το μοντέλο ARMA:

$$X_t = \varepsilon_t + \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad \text{Εξ. 3.6}$$

Με την χρήση του τελεστή υστέρησης B , η παραπάνω εξίσωση γίνεται ως εξής:

$$(1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p) X_t = \delta + (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_p B^p) \varepsilon_t$$

Ή εναλλακτικά ως:

$$\Phi(B) X_t = \delta + \Theta(B) \varepsilon_t$$

Η χρονοσειρά θεωρείται στάσιμη αν το χαρακτηριστικό πολυώνυμο $\Phi(B)$ έχει ρίζες εκτός του μοναδιαίου κύκλου.

Για την εκτίμηση των παραμέτρων φ_i και θ_i για το μοντέλο ARMA στην MATLAB χρησιμοποιείται η εκτίμηση μέγιστης πιθανοφάνειας (Maximum Likelihood Estimation). Στην εργασία αυτή υποθέτουμε ότι η μηνιαία ζήτηση βιβλίων μετά από κατάλληλες διαφορίσεις και αφαίρεση της μέσης τιμής δίνουν χρονοσειρές X_t που ακολουθούν (κατά προσέγγιση) κανονική κατανομή. Στην περίπτωση αυτή, οι εκτιμήσεις παραμέτρων με τη μέθοδο μέγιστης πιθανοφάνειας

είναι ίδιες με τις εκτιμήσεις που δίνει η μέθοδος ελαχίστων τετραγώνων (ελαχιστοποίηση του αθροίσματος τετραγώνων των σφαλμάτων πρόβλεψης ή των υπολοίπων).

3.2.1 ADF Test

Ο έλεγχος της στασιμότητας μπορεί να επιτευχθεί μέσω του Augmented Dickey-Fuller Test (ADF test).

Το ADF test, θεωρείται επέκταση του αρχικού Dickey-Fuller test και εξετάζει αν μία χρονοσειρά είναι στάσιμη ή όχι.

Η εξίσωση για το ADF test παρουσιάζεται παρακάτω:

$$\Delta X_t = a + \beta t + \gamma X_{t-1} + \delta_1 \Delta X_{t-1} + \delta_2 \Delta X_{t-2} + \dots + \delta_{p-1} \Delta X_{t-p+1} + \varepsilon_t \quad \text{Εξ. 3.7}$$

όπου:

- a είναι σταθερός όρος
- β είναι ο όρος της γραμμικής τάσης
- $\Delta X_t = X_t - X_{t-1}$ είναι οι πρώτες διαφορές της χρονοσειράς για $t = t, t-1, \dots$ (π.χ. $\Delta X_t = X_t - X_{t-1}, \Delta X_{t-1} = X_{t-1} - X_{t-2}, \dots, \Delta X_{t-p+1} = X_{t-p+1} - X_{t-p}$)
- ε_t το σφάλμα

Πιο συγκεκριμένα, ο έλεγχος εξετάζει δύο υποθέσεις. Η μηδενική υπόθεση θεωρεί ότι η χρονοσειρά δεν είναι στάσιμη. Η εναλλακτική υπόθεση θεωρεί ότι η χρονοσειρά είναι στάσιμη (Guo, 2023).

Υπολογίζονται έπειτα τα σφάλματα για το ADF test σύμφωνα με την Εξ. 3.7 και τελικά η στατιστική $ADF = \hat{Y}/SE(\hat{Y})$, όπου $SE(\hat{Y})$ είναι το τυπικό σφάλμα (standard error) της πρόβλεψης ελαχίστων τετραγώνων (τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος). Αυτή η τιμή ακολουθεί την ειδική κατανομή Dickey-Fuller. Από τους ειδικούς πίνακες των Dickey και Fuller, εξάγεται με την χρήση του ADF statistic και του επιπέδου σημαντικότητας $p=0.05$, το συμπέρασμα για την απόρριψη ή όχι της αρχικής υπόθεσης. Αν η τιμή του ADF statistic είναι μεγαλύτερη κατά απόλυτη τιμή, από την κρίσιμη τιμή, τότε η μηδενική υπόθεση απορρίπτεται και γίνεται δεκτή η υπόθεση της στασιμότητας.

3.3 Εκθετική εξομάλυνση

3.3.1 Απλή εκθετική εξομάλυνση

Η απλή εκθετική εξομάλυνση είναι μία μέθοδος κατάλληλη για πρόβλεψη σε γεγονότα, χωρίς τάση και εποχικότητα και ειδικότερα για χρονοσειρές με λίγα δεδομένα. Η μέθοδος εκτελείται

ορίζοντας «βάρη» σε προηγούμενες τιμές της χρονοσειράς. Όσο παλαιότερα βρίσκονται οι τιμές στην χρονοσειρά, τόσο μικρότερο βάρος έχουν. Επομένως, δίνεται μεγαλύτερη βαρύτητα και σημαντικότητα στις πιο πρόσφατες τιμές από αυτές που εξετάζονται κάθε στιγμή. Ο παράγοντας που ορίζεται για το βάρος είναι ο θ και μειώνεται γεωμετρικά για παλαιότερες τιμές (Montgomery et al., 2015). Ο τύπος της απλής εκθετικής εξομάλυνσης είναι ο εξής με $\lambda = 1 - \theta$:

$$\hat{y}_t = \lambda y_t + (1 - \lambda)\hat{y}_{t-1} \quad \text{Εξ. 3.8}$$

όπου y_t είναι η πρόσφατη μέτρηση και \hat{y}_{t-1} είναι η προηγούμενη εκτίμηση της y_{t-1} . Το λ είναι η παράμετρος εξομάλυνσης με απόλυτη τιμή μικρότερη του 1 ($|\lambda| < 1$). Σύμφωνα με την βιβλιογραφία οι τιμές που δίνονται συνηθέστερα στο λ είναι από 0.1 έως 0.4 (Montgomery et al., 2015).

3.3.2 Μέθοδος Holt-Winters

Η μέθοδος Holt-Winters (Hyndman & Athanasopoulos, 2021) χρησιμοποιείται κυρίως για δεδομένα που παρουσιάζουν τάση και εποχικότητα και εξαιτίας αυτών, περιλαμβάνει και την εκτίμηση των αντίστοιχων παραμέτρων. Ο γενικός τύπος είναι:

$$\hat{y}_{t+h}(t) = (l_t + hb_t) + s_{t+h-m(k+1)} \quad \text{Εξ. 3.9}$$

όπου, m αριθμός εποχών σε ένα έτος και k το ακέραιο μέρος της διαίρεσης $(h - 1)/m$ που εξασφαλίζει ότι ο δείκτης που χρησιμοποιείται για την εποχικότητα αφορά στο τελευταίο έτος του δείγματος. Το h αφορά στον ορίζοντα πρόβλεψης.

Η εκτίμηση της παραμέτρου αύξησης της χρονοσειράς (π.χ. χωρίς αύξηση, με γραμμική ή με γεωμετρική αύξηση) l_T δίνεται από τον τύπο:

$$l_t = a(y_t - s_{t-m}) + (1 - a)(l_{t-1} + b_{t-1}) \quad \text{Εξ. 3.10}$$

όπου, $0 \leq a \leq 1$, ο συντελεστής εξομάλυνσης επιπέδου.

Για την εκτίμηση της παραμέτρου για την τάση β_1 , χρησιμοποιείται ο ακόλουθος τύπος:

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \quad \text{Εξ. 3.11}$$

όπου, $0 \leq \beta \leq 1$, ο συντελεστής εξομάλυνσης τάσης.

Τέλος, για την εκτίμηση της παραμέτρου της εποχικότητας S_T :

$$S_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1 - \gamma)S_{t-m}$$

Εξ. 3.12

όπου, $0 \leq \gamma \leq 1$, ο συντελεστής εξομάλυνσης εποχικότητας.

Οι παραπάνω εξισώσεις λειτουργούν αναδρομικά και μέσα από αυτές το μοντέλο προσαρμόζεται στις μεταβολές επιπέδου, τάσης και εποχικότητας.

3.4 Προσέγγιση Syntetos-Boylan

Η προσέγγιση των Syntetos-Boylan (2005) (SBA) χρησιμοποιείται κυρίως για την πρόβλεψη χρονοσειρών με διακοπτόμενη ζήτηση για τον έλεγχο αποθέματος. Δηλαδή, στις χρονοσειρές στις οποίες μπορεί να εφαρμοστεί η προσέγγιση SBA, υπάρχουν περίοδοι χωρίς ζήτηση, γεγονός το οποίο μπορεί να επηρεάσει το πραγματικό μοτίβο της ζήτησης. Η SBA μέθοδος αποτελεί μία παραλλαγή της μεθόδου του Croston (1972).

Ο τύπος για την εφαρμογή της παραπάνω μεθόδου είναι:

$$X'_t = (1 - \frac{b}{2}) \frac{z'_t}{p'_t}$$

Εξ. 3.13

όπου, X'_t η πρόβλεψη ζήτησης, το b είναι η τιμή της σταθεράς εξομάλυνσης που χρησιμοποιείται για την ενημέρωση των διαστημάτων μεταξύ των απαιτήσεων (σύμφωνα με τους συγγραφείς συνίσταται να είναι ανάμεσα στο 0,05 και στο 0,2). Το z'_t αφορά στη ποσότητα του ζητούμενου προϊόντος την περίοδο t , ενώ το p'_t αφορά στον αριθμό περιόδων που έχουν μηδενική ζήτηση ανάμεσα σε δύο περιόδους με υπαρκτή ζήτηση.

Οι τύποι για τον υπολογισμό των z'_t και p'_t είναι οι ακόλουθοι:

$$z'_t = \alpha X_t + (1 - \alpha)z'_{t-1}$$

Εξ. 3.14

όπου το D_t συμβολίζει την τελευταία θετική τιμή ζήτησης της χρονοσειράς.

$$p'_t = bI_t + (1 - b)p'_{t-1}$$

Εξ. 3.15

όπου I_t είναι ο αριθμός των περιόδων από την τελευταία φορά που υπήρξε θετική ζήτηση, ενώ το α είναι η τιμή της σταθεράς εξομάλυνσης των τιμών ζήτησης της χρονοσειράς. (Boylan & Syntetos, 2001).

3.5 Απλός κινητός μέσος

Η μέθοδος του Κινητού Μέσου (Montgomery et al., 2015) είναι η απλούστερη μέθοδος πρόβλεψης. Είναι μία γραμμική μέθοδος εξομάλυνσης δεδομένων. Η πρόβλεψη των επόμενων τιμών μίας χρονοσειράς γίνεται χρησιμοποιώντας τις προηγούμενες N τιμές της χρονοσειράς με βάρος $\frac{1}{N}$ για κάθε παρατήρηση. Ο τύπος είναι:

$$M_T = \frac{1}{N} \sum_{t=T-N+1}^T X_t \quad \text{Εξ. 3.16}$$

όπου N , είναι ο αριθμός περιόδων, σταθερός κάθε φορά, από τις οποίες υπολογίζεται η μέση τιμή της χρονοσειράς για την πρόβλεψη της ζήτησης. Επομένως, κάθε φορά που υπολογίζεται η πρόβλεψη για μία νέα τιμή, αυτή αντικαθιστά την παλαιότερη παρατήρηση στο σύνολο N , για τον υπολογισμό της επόμενης τιμής πρόβλεψης. Όταν το N είναι μικρό, η πρόβλεψη γίνεται πιο ευαίσθητη σε πρόσφατες αλλαγές, ωστόσο είναι και περισσότερο ασταθής. Αντίθετα, όταν το N είναι μεγάλο, η πρόβλεψη είναι πιο ομαλή, αλλά δεν ανταποκρίνεται έγκαιρα σε αλλαγές.

3.6 Διπλή εκθετική εξομάλυνση για ακανόνιστα (ελλιπή) δεδομένα

Ο αλγόριθμος Διπλής εκθετικής εξομάλυνσης για ακανόνιστα δεδομένα δημιουργήθηκε από τον Tomas Cípra. Ακολουθήθηκε πιστά για την εφαρμογή το άρθρο του (Cípra, 2006), σύμφωνα με το οποίο αναπτύσσονται τρεις διαφορετικές τροποποιημένες μέθοδοι εκθετικής εξομάλυνσης και τελικά συγκρίνονται δύο από αυτές. Ο Cípra καταλήγει ότι η μέθοδος που έχει καλύτερη εφαρμογή στην πρόβλεψη δεδομένων είναι η Double exponential smoothing for irregular data. Συγκεκριμένα, προτείνει τροποποιήσεις των συντελεστών εξομάλυνσης που λαμβάνουν υπόψη τους τα «κενά» διαστήματα ανάμεσα στις παρατηρήσεις.

Για την αρχικοποίηση του συντελεστή εκθετικής εξομάλυνσης α χρησιμοποιείται η εξίσωση:

$$\alpha_0 = 1 - (1 - \alpha)^q \quad \text{Εξ. 3.17}$$

Για τον υπολογισμό του, χρησιμοποιείται η σχέση:

$$a_{t_n} = \frac{1}{\sum_{j=1}^n \beta^{t_n - t_j}} \quad \text{Εξ. 3.18}$$

Για τον συντελεστή εκθετικής εξομάλυνσης β χρησιμοποιείται η εξίσωση:

$$\beta = 1 - \alpha$$

Eξ. 3.19

Ο Cίπρα εισάγει επιπλέον δύο βοηθητικές μεταβλητές, τις w και z , οι οποίες χρησιμοποιούνται ως βάρη επιπέδου και τάσης, αντίστοιχα.

Παρακάτω εμφανίζονται οι σχέσεις που τις αρχικοποιούν:

$$w_0 = \frac{(1 - \beta^q)^2}{q\beta^q} \quad \text{Eξ. 3.20}$$

$$z_0 = \frac{(1 - \beta^q)^2}{q\beta^q} \quad \text{Eξ. 3.21}$$

Οι εξισώσεις για την αρχικοποίηση του επιπέδου και της τάσης παρουσιάζονται παρακάτω:

$$S_0 = \widehat{b}_0(0) - \frac{q\beta^q}{1 - \beta^q} \widehat{b}_1(0) \quad \text{Eξ. 3.22}$$

$$S_0^{[2]} = \widehat{b}_0(0) - 2 \frac{q\beta^q}{1 - \beta^q} \widehat{b}_1(0) \quad \text{Eξ. 3.23}$$

Οι μεταβλητές w και z υπολογίζονται από τους τύπους:

$$w_{t_n} = \frac{1}{\sum_{j=1}^n (t_n - t_j) \beta^{t_n - t_j}} \quad \text{Eξ. 3.24}$$

$$z_{t_n} = \frac{1}{\sum_{j=1}^n \frac{a_{t_j}}{w_{t_j}} \beta^{t_n - t_j}} \quad \text{Eξ. 3.25}$$

Μετά τον υπολογισμό των παραπάνω, το επίπεδο και η τάση υπολογίζονται με τους παρακάτω τύπους:

$$S_{t_n} = a_{t_n} y_{t_n} + (1 - a_{t_n}) S_{t_{n-1}} \quad \text{Eξ. 3.26}$$

$$S_{t_n}^{[2]} = a_{t_n} S_{t_n} + (1 - a_{t_n}) S_{t_{n-1}}^{[2]} \quad \text{Eξ. 3.27}$$

Η πρόβλεψη υπολογίζεται από τον τύπο:

$$Y_{t_n+k}(t_n) = S_{t_n} + \left(\frac{z_{t_n}}{w_{t_n}} - \frac{z_{t_n}}{a_{t_n}} t_n \right) (S_{t_n} - S_{t_n}^{[2]}) \quad \text{Eξ. 3.28}$$

Η παραπάνω μέθοδος εφαρμόζεται στην παρούσα εργασία και τα αποτελέσματα συγκρίνονται με αυτά των άλλων δύο αλγορίθμων που δημιουργήθηκαν και αναλύονται στο Κεφάλαιο 5.

3.7 Σφάλματα

Τα είδη σφαλμάτων που θα χρησιμοποιηθούν για την σύγκριση των παραπάνω αλγορίθμων αναλύονται παρακάτω:

3.7.1 Μέσο τετραγωνικό σφάλμα (Root mean square error)

Το μέσο τετραγωνικό σφάλμα είναι κοινώς διαδεδομένο και χρησιμοποιείται για εύρεση του σφάλματος μεταξύ των προβλεπόμενων τιμών και των πραγματικών. Ο τύπος του είναι:

$$RMSE = \sqrt{E((\hat{\theta} - \theta)^2)} \quad \text{Eξ. 3.29}$$

Όπου $\hat{\theta}$, η προβλεπόμενη τιμή και θ , η πραγματική.

4. Υλοποίηση Αλγορίθμων και Διαγράμματα Ροής

4.1 Εισαγωγή

Αναπτύχθηκαν τέσσερις διαφορετικοί αλγόριθμοι στην MATLAB. Τα δεδομένα αρχικά ομαδοποιήθηκαν καταλλήλως και ανάλογα με τις επιλογές που δίνει ο χρήστης μέσα από αρχείο κειμένου .txt. Για την διευκόλυνση της διαδικασίας, δημιουργήθηκε κατάλληλη συνάρτηση ομαδοποίησης. Στη συνέχεια η συνάρτηση χρησιμοποιήθηκε ως είσοδος για δύο αλγορίθμους. Ο πρώτος αλγόριθμος χρησιμοποιώντας την μέθοδο ARMA και την μέθοδο τριπλής εκθετικής εξομάλυνσης (Holt-Winters) και τον μέσο όρο, κάνει πρόβλεψη και εξετάζει τα σφάλματα. Για τον συγκεκριμένο αλγόριθμο, χρειάστηκε να αναπτυχθεί και η αντίστοιχη συνάρτηση για την μέθοδο Holt-Winters. Ο δεύτερος αλγόριθμος χρησιμοποιεί τις μεθόδους SBA (προσέγγιση Syntetos-Boylan), την μέθοδο Απλού κινητού μέσου και την μέθοδο της απλής εκθετικής εξομάλυνσης και κάνει αντίστοιχα πρόβλεψη των δεδομένων της κάθε χρονοσειράς από το κάθε βιβλίο. Ταυτόχρονα, υπολογίζει και το Root mean squared error. Επιπλέον, δημιουργείται ένας τρίτος αλγόριθμος που βασίζεται εξολοκλήρου στο άρθρο του Thomas Cípra (2005) με την μέθοδο Double Exponential smoothing for irregular data, για να συγκριθούν τα αποτελέσματά του με αυτά των προηγούμενων αλγορίθμων:

Μέθοδος ARMA, τριπλή εκθετική εξομάλυνση με την μέθοδο Holt-Winters, απλή Εκθετική , προσέγγιση Syntetos-Boylan και απλός κινητός μέσος.

4.2 Συνάρτηση ομαδοποίησης

Η συνάρτηση ομαδοποίησης χρησιμοποιεί τα αρχεία excel για κάθε βιβλίο μετά την προεπεξεργασία που υπέστησαν, όπως αναφέρεται στο κεφάλαιο 2.2. Επιπλέον, χρησιμοποιεί ως είσοδο και τα αρχεία .txt, ένα για κάθε βιβλίο, όπως αναφέρεται στο κεφάλαιο 2.3.

Η συνάρτηση δέχεται ως είσοδο το όνομα του αρχείου excel “excelFile”, το όνομα του φύλλου μέσα στο αρχείο excel που βρίσκονται τα δεδομένα “sheetName”, το εύρος των γραμμών και των στηλών που βρίσκονται τα δεδομένα “range” και το όνομα του αρχείου txt “txtFile” στο οποίο βρίσκονται οι ομαδοποιήσεις. Ως έξοδο, η συνάρτηση δίνει ένα αρχείο excel στο οποίο φαίνονται οι ομαδοποιήσεις που έγιναν με τον τρόπο που έχει επιλέξει ο χρήστης.

Αρχικά αντικαθίστανται οι αρνητικές τιμές του αρχείου με 0.

Στη συνέχεια, δημιουργείται ένας νέος πίνακας με το όνομα “Sales” με τις νέες τιμές του αρχείου. Έπειτα, αθροίζονται κατά μήνα οι τιμές του πίνακα “Sales” και διαβάζεται το αρχείο txt.

Αρχειοποιούνται οι δείκτες για την ανάγνωση των γραμμών στο αρχείο txt “lineNum”, η μεταβλητή για τον αριθμό των ομάδων που θα δημιουργηθούν “groupCounter”, η μεταβλητή για

την καταμέτρηση των εναλλακτικών που έχει επιλέξει ο χρήστης “alternativesCounter” και, τέλος, δημιουργείται ένα κενό διάνυσμα για την αποθήκευση των εναλλακτικών “alternatives”.

Στο κυρίως μέρος του αλγορίθμου, χρησιμοποιείται μία εντολή επανάληψης, για την ανάγνωση του αρχείου txt, η οποία σταματάει εξετάζοντας αν έχει φτάσει το αρχείο στο τέλος του ή όχι.

Επειδή μέσα στο αρχείο υπάρχουν αρκετές γραμμές που αφορούν σε σχόλια για την διευκόλυνση του χρήστη, υπάρχει εντολή σε αυτό το σημείο προκειμένου να αγνοήσει αυτές τις γραμμές.

Δίνεται η εντολή να αποθηκεύσει στην μεταβλητή num_Alternatives την πρώτη γραμμή χωρίς «#» που θα βρεθεί κατά την ανάγνωση του αρχείου, καθώς αυτή περιέχει τον αριθμό των εναλλακτικών που έδωσε ο χρήστης και μετατρέπεται σε αριθμό κατά την αποθήκευσή του. Έπειτα προχωράει στην επόμενη γραμμή του αρχείου.

Στη συνέχεια εξάγει και αποθηκεύει ως αριθμούς τους μήνες από την επόμενη γραμμή και αποθηκεύει ως εύρος μηνών “numRanges” τον αριθμό των μηνών εκτός του τελευταίου, καθώς είναι ίδιος με τον πρώτο μήνα. Για προγραμματιστικούς σκοπούς που θα αναφερθούν παρακάτω, είναι απαραίτητο ο πρώτος μήνας να οριστεί και ως τελευταίος.

Η συνάρτηση αρχικοποιεί δύο πίνακες, έναν κατάλληλο για ονομαστικές μεταβλητές και έναν για αριθμητικές μεταβλητές. Στην μεταβλητή “groupNames” θα αποθηκευτούν τα ονόματα των ομάδων και στην μεταβλητή “groupSales” θα αποθηκευτούν οι τιμές από το άθροισμα των πωλήσεων για κάθε ομάδα.

Έπειτα, στόχος είναι τα δεδομένα να συγκεντρώνονται με την σωστή χρονολογική σειρά. Εξηγώντας καλύτερα, οι καταγραφές των δεδομένων ξεκινάνε από τυχαίες ημερομηνίες σε κάθε βιβλίο αλλά με τη σωστή χρονολογική σειρά. Σε κάποιο βιβλίο η καταγραφή ξεκινάει Οκτώβριο του 2019, ενώ σε άλλο τον Μάιο του 2018. Επομένως, αν ο χρήστης επιλέξει η ομαδοποίηση να αρχίσει από τον Ιανουάριο, και δεν διαμορφωθεί σωστά ο αλγόριθμος, τότε η ομαδοποίηση θα αφήσει εκτός τους μήνες πριν από την πρώτη στιγμή που εμφανίζεται ο Ιανουάριος στις καταγραφές. Επομένως, θα πρέπει να αποθηκευτούν τα έτη που εμφανίζονται στις καταγραφές. Αυτό επιτυγχάνεται με την εντολή: `uniqueYears = unique(year(SalesByMonth.Date))`; και αποθηκεύονται στην μεταβλητή “uniqueYears”. Στη συνέχεια το μεγαλύτερο και το μικρότερο έτος αποθηκεύονται στις μεταβλητές “maxYear” και “minYear” αντίστοιχα. Έτσι παρακάτω, ξεκινάει μία επανάληψη με βήμα 1 αρχίζοντας από το 1 και τελειώνοντας στην μεταβλητή “numRanges”. Αποθηκεύεται έπειτα ως αρχικός μήνας καταμέτρησης ο τρέχοντας “startMonth” και ως τελευταίος ο αμέσως επόμενος στο από το αρχείο txt “endMonth”.

Στην επόμενη επανάληψη που ακολουθεί για κάθε τιμή που είναι ίση με το πρώτο έτος που έχει αποθηκευτεί στην μεταβλητή “uniqueYears”, αν ο πρώτος μήνας στη σειρά των μηνών που έχει αποθηκευτεί ως “startMonth”, είναι μεγαλύτερος σε τιμή από τον επόμενο μήνα που έχει αποθηκευτεί ως “endMonth”, τότε δημιουργούνται δύο υποομάδες: η “groupData1” και η “groupData2”. Στην πρώτη υποομάδα αποθηκεύονται οι πωλήσεις για τους μήνες που είναι

μεγαλύτεροι ή ίσοι σε νούμερο από τον “startMonth” και ανήκουν στο τρέχον έτος “yr”, ενώ στην δεύτερη υποομάδα αποθηκεύονται οι πωλήσεις για τους μήνες που είναι μικρότεροι κατά νούμερο από τον “endMonth” και ανήκουν στο επόμενο έτος από το τρέχον.

Σε αυτό το σημείο υπάρχουν αρκετές περιπτώσεις που πρέπει να αντιμετωπιστούν για να αποφευχθούν τυχόν σφάλματα κατά την άθροιση των δεδομένων. Όπως, η περίπτωση να μην υπάρχουν δεδομένα για επόμενο έτος τους συγκεκριμένους μήνες που ζητά ο χρήστης, επομένως αν η δεύτερη υποομάδα είναι κενή και αφορά στο επόμενο έτος το οποίο είναι μικρότερο από το μεγαλύτερο έτος που έχει καταγραφεί στα δεδομένα, δηλ. το “maxYear”, ως τρέχουσα ομάδα (η οποία θα υποστεί επεξεργασία) αποθηκεύεται στο “groupData1”, δηλαδή η πρώτη υποομάδα. Σε κάθε άλλη περίπτωση, αποθηκεύονται και οι δύο υποομάδες που δημιουργήθηκαν.

Σε άλλη περίπτωση αν το τρέχον έτος είναι ίσο με το μικρότερο που έχει εντοπιστεί, τότε αποθηκεύονται στην τρέχουσα ομάδα “currentGroup” τα δεδομένα που ανήκουν σε μήνες μικρότερους από τον τελευταίο μήνα και ταυτόχρονα στο μικρότερο έτος. (“groupDataEarlyYear”). Αν, όμως, το τρέχον έτος είναι το μεγαλύτερο που έχει εντοπιστεί και ταυτόχρονα το “groupData2”, το οποίο έχει σχηματιστεί νωρίτερα, είναι κενό, τότε θα πρέπει στο “currentGroup” να αποθηκεύονται τα δεδομένα από το “groupData1”.

Γυρνώντας ξανά στην προηγούμενη επανάληψη, ελέγχεται τώρα αν ο μήνας που εξετάζεται (“startMonth”) είναι μικρότερος από τον επόμενο (“endMonth”). Εφόσον συμβαίνει αυτό, τότε στην τρέχουσα ομάδα (“currentGroup”) αποθηκεύονται οι πωλήσεις για τους μήνες που είναι μεγαλύτεροι ή ίσοι από τον “startMonth” και μικρότεροι από τον “endMonth” και ταυτόχρονα αυτοί οι μήνες ανήκουν στο ίδιο έτος.

Στο τέλος αυτών των επαναλήψεων αποθηκεύεται στο “totalSales” το άθροισμα του “currentGroup.Sales” και στη συνέχεια αποθηκεύονται τα έτη στο “groupNames{end+1}” και οι πωλήσεις (“totalSales”) στο “groupSales(end+1)”, προκειμένου είναι καλύτερα οργανωμένα προς διευκόλυνση του χρήστη.

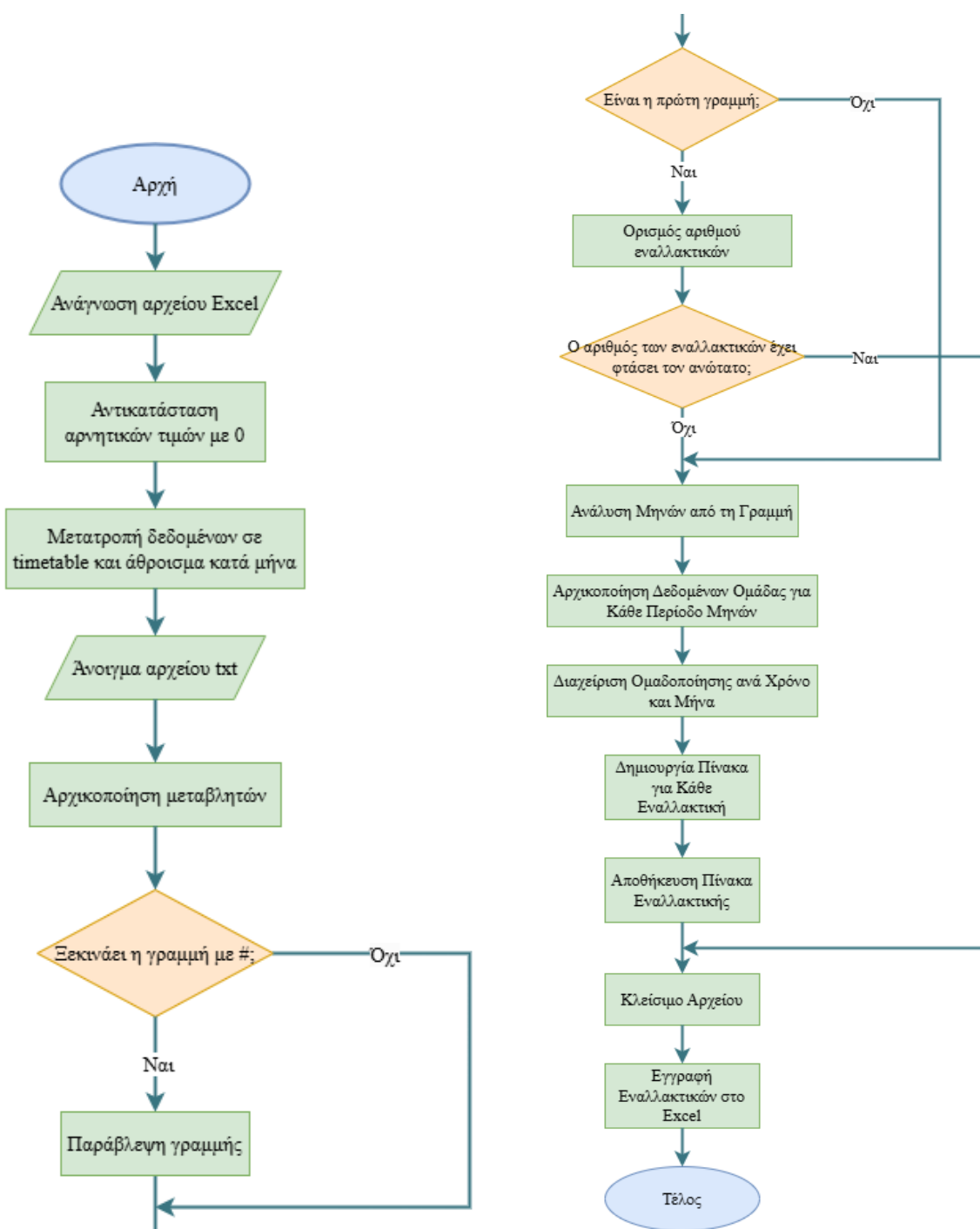
Στη συνέχεια απαραίτητο είναι να δημιουργηθούν πίνακες για την ορθότερη συγκέντρωση των αποτελεσμάτων (“alternativeTable”). Επομένως δημιουργείται πίνακας που αρχικά αποθηκεύει τις μεταβλητές “groupNames” και “groupSales” κάτω από τις στήλες “Group” και “Sales”. Αυξάνεται η τιμή της “alternativeCounter” κατά ένα, για να μεταφερθεί στην επόμενη εναλλακτική κατά την επόμενη επανάληψη.

Έπειτα αποθηκεύεται σε ένα νέο κελί στην μεταβλητή “alternatives{alternativeCounter}” ο πίνακας με τις τιμές που δημιουργήθηκε προηγουμένως (“alternativeTable”).

Τέλος εμφανίζεται μήνυμα για την επιτυχή δημιουργία της εναλλακτικής και προχωράει στην ανάγνωση της επόμενης γραμμής του txt.

Τελειώνει η επανάληψη, το αρχείο κλείνει και αποθηκεύονται όλα τα αποτελέσματα σε ένα αρχείο excel με όνομα Alternatives. Κάθε εναλλακτική αποθηκεύεται σε διαφορετικό φύλλο με όνομα Alternative_(αριθμός εναλλακτικής).

Παρακάτω εμφανίζεται ένα διάγραμμα ροής για την καλύτερη κατανόηση του αλγορίθμου της συνάρτησης ομαδοποίησης.



Εικόνα 4.1: Διάγραμμα ροής Συνάρτησης ομαδοποίησης

4.3 Αλγόριθμος πρόβλεψης με μεθόδους μέσου όρου, ARIMA και Holt-Winters

Ο αλγόριθμος πρόβλεψης χρησιμοποιεί ως είσοδο τα αποτελέσματα της συνάρτησης ομαδοποίησης.

4.3.1 Αρχικοποίηση και αποθήκευση

Αρχικά, αποθηκεύονται τα δεδομένα από τα αρχεία excel για κάθε βιβλίο και τα αρχεία txt στα οποία ορίζει ο χρήστης την ομαδοποίηση που επιθυμεί. Αρχικοποιούνται επιπλέον μεταβλητές για την αποθήκευση των RMSE από κάθε μέθοδο, καθώς και των διαφόρων μεθόδων διαφορίσης ώστε να αφαιρεθούν η εποχικότητα και η διαχρονική τάση. Συγκεκριμένα, ο αλγόριθμος αρχικοποιεί την μεταβλητή RMSE_Results για τα αποτελέσματα από το μέσο τετραγωνικό σφάλμα (root mean squared error) που θα υπολογιστεί αργότερα. Επιπλέον, ορίζει, προκειμένου να αποθηκεύσει μέσα στην ίδια μεταβλητή, την RMSE_Results με επικεφαλίδες για την αποθήκευση των αποτελεσμάτων της RMSE_Results. Αυτές οι επικεφαλίδες είναι: 'File', 'Alternative', 'Differencing_Method', 'WhiteNoise_RMSE', 'Best_ARIMA_RMSE', 'Holt_Winters_RMSE'. Αρχικοποιείται η μεταβλητή rowIndex = 2 και το seasonalLag που αφορά στους μήνες που ο χρήστης θέλει να έχει κάθε εποχή.

4.3.2 Κύριο μέρος αλγορίθμου

Στη συνέχεια ξεκινάει η κύρια επανάληψη μέσα στην οποία πραγματοποιείται το κύριο μέρος του αλγορίθμου. Η επανάληψη ξεκινάει με την μεταβλητή k=1, η οποία παίρνει τιμές μέχρι την εξάντληση των αρχείων excel των δεδομένων, δηλαδή των βιβλίων που θα υποστούν επεξεργασία. Στις μεταβλητές currentExcelFile και currentTxtFile θα αποθηκευτούν τα τρέχοντα αρχεία με τα δεδομένα εισόδου, excel και txt αντίστοιχα.

Έπειτα για λόγους οργάνωσης, δίνεται εντολή να χρησιμοποιηθεί το όνομα των αρχείων excel με τα δεδομένα εισόδου ως ετικέτα.

Στις επόμενες γραμμές καλείται η συνάρτηση ομαδοποίησης και αποθηκεύονται τα αποτελέσματα στη μεταβλητή Alternatives και αποθηκεύεται ο αριθμός των παρατηρήσεων στο numAlternatives.

Στη επόμενη επανάληψη για alt=1 έως τον αριθμό των εναλλακτικών, αποθηκεύονται στο currentTable τα δεδομένα της τρέχουσας εναλλακτικής. Στη συνέχεια, αποθηκεύεται στην μεταβλητή sales_data η στήλη των πωλήσεων από την currentTable και στην μεταβλητή numObservations αποθηκεύεται το μήκος της sales_data. Σε αυτό το σημείο, τα δεδομένα χωρίζονται σε trainData και testData. Σκοπός είναι η εκπαίδευση των μοντέλων για την εύρεση των καλύτερων παραμέτρων κάθε μοντέλου και στην συνέχεια η σύγκρισή τους με ένα μικρό σύνολο δεδομένων, τα οποία τα μοντέλα δεν έχουν δει. Τα testData ορίζονται 5 σε αριθμό - οι 5 τελευταίες παρατηρήσεις- καθώς το σύνολο το δεδομένων είναι γενικά μικρό. Τα υπόλοιπα

δεδομένα αποθηκεύονται στην μεταβλητή `trainData`. Επιπλέον αποθηκεύεται ο αριθμός των `testData` στην μεταβλητή `numTest`.

4.3.2.1 Πρόβλεψη με μέθοδο *Holt-Winters*

Για την πρόβλεψη με την μέθοδο της τριπλής εκθετικής εξομάλυνσης δεν εφαρμόζεται διαφόριση. Τα δεδομένα πρέπει να είναι αυτοτελή, καθώς η μέθοδος εντοπίζει εποχικότητα και τάση από τα ακατέργαστα δεδομένα. Αρχικοποιούνται η `RMSE_HW` για την αποθήκευση του `RMSE` από την μέθοδο *Holt-Winters* και η `bestRMSE_HW` για την αποθήκευση της καλύτερης τιμής `RMSE` μετά από σύγκριση που θα πραγματοποιηθεί μεταγενέστερα. Έπειτα, εκτελείται τριπλή επανάληψη, μία για κάθε παράμετρο `alpha`, `beta` και `gamma`, προκειμένου να βρεθεί η βέλτιστη τιμή σύμφωνα με το `RMSE` για κάθε παράμετρο. Σύμφωνα με την βιβλιογραφία, οι τιμές τους θα πρέπει να είναι ανάμεσα στο 0 και στο 1 (Hyndman & Athanasopoulos, 2018). Μέσα στην τριπλή επανάληψη, καλείται η συνάρτηση `HoltWintersForecast` με είσοδο την μεταβλητή `trainData`, τις μεταβλητές `alpha`, `beta`, `gamma`, την `seasonalLag` και την `numTest`. Η συνάρτηση ως έξοδο δίνει τις προβλεπόμενες τιμές ως `fittedValues`. Έπειτα υπολογίζεται και αποθηκεύεται το `current_rmse` για κάθε τιμή που παίρνουν οι παράμετροι από την αφαίρεση των `fittedValues` από τα `trainData`. Πριν ολοκληρωθεί η τριπλή επανάληψη, τα αποτελέσματα συγκρίνονται με το `bestRMSE_HW` και αν είναι καλύτερα, κρατούνται στην μεταβλητή `bestRMSE_HW`, η οποία στο τέλος θα έχει την μικρότερη τιμή. Έπειτα, ως `RMSE_HW` ορίζεται το `bestRMSE_HW` από την προηγούμενη σύγκριση.

4.3.2.2 Διαφόριση

Στη συνέχεια για κάθε τιμή της αλφαριθμητικής μεταβλητή `differencingMethods` στην οποία αποθηκεύονται οι μέθοδοι διαφόρισης, εκτελείται μία επανάληψη μέσα στην οποία πραγματοποιείται η διαφόριση για τις επόμενες μεθόδους.

Για την μέθοδο `none` δεν ακολουθείται καμία διαφόριση. Για την μέθοδο `first_order` ακολουθείται πρώτης τάξης διαφόριση, δηλαδή από την τρέχουσα παρατήρηση στη χρονοσειρά αφαιρείται η προηγούμενη χρονικά. Για την μέθοδο `seasonal` ακολουθείται εποχική διαφόριση ανάλογα με τον αριθμό που έχει επιλεγεί ως `seasonalLag`, δηλαδή από την τρέχουσα παρατήρηση στη χρονοσειρά αφαιρείται η παρατήρηση που απέχει χρονικά τόσους μήνες όσους ορίζει η μεταβλητή `seasonalLag`. Τέλος, για την μέθοδο `both` αρχικά ακολουθείται διαφόριση πρώτης τάξης και έπειτα εποχική διαφόριση, ανάλογα με τον αριθμό που έχει επιλεγεί ως `seasonalLag`. Σε κάθε μία από τις προηγούμενες μεθόδους, τα αποτελέσματα αποθηκεύονται στην μεταβλητή `trainTimeSeries`.

4.3.2.3 Πρόβλεψη με βάση την μέση τιμή

Υπολογίζεται η μέση τιμή (`meanValue`) της διαφορισμένης χρονοσειράς και αποθηκεύεται η μέση τιμή στην μεταβλητή `predictedDifferenced_WN` ως πρόβλεψη. Έπειτα, αρχικοποιείται η μεταβλητή `predictedOriginal_WN` για την αποθήκευση των αποτελεσμάτων της αντίστροφης διαφόρισης που ακολουθεί.

4.3.2.4 Αντίστροφη διαφόριση

Στην αντίστροφη διαφόριση, εφαρμόζεται αντίστροφη διαφόριση ανάλογα με την μέθοδο διαφόρισης που είχε προηγηθεί στην προηγούμενη επανάληψη.

Στην περίπτωση χωρίς διαφόριση τα αποτελέσματα αποθηκεύονται ως έχουν.

Σε περίπτωση διαφόρισης πρώτης τάξης `first_order` χρησιμοποιείται ως αρχική τιμή η τελευταία παρατήρηση στην μη διαφορισμένη χρονοσειρά, `trainData`, και έπειτα αθροίζεται σε αυτήν η πρώτη τιμή της διαφορισμένης χρονοσειράς, `predictedDifferenced_WN`. Η τελευταία τιμή του `trainData` χρησιμοποιείται σαν πρώτη για την νέα χρονοσειρά που θα προκύψει από την αντίστροφη διαφόριση. Το αποτέλεσμα είναι η πρώτη τιμή των προβλέψεων της αντίστροφης διαφόρισης. Αυτή η νέα τιμή θα χρησιμοποιηθεί, μαζί με τις υπόλοιπες, για να συγκριθεί με τις τιμές του test set και να προκύψει το νέο RMSE. Αυτά αποθηκεύονται στην μεταβλητή `predictedOriginal_WN`. Στην συνέχεια για κάθε επόμενη τιμή μετά την πρώτη και μέχρι το `numTest`, στην τρέχουσα τιμή της `predictedOriginal_WN` αθροίζεται η ακριβώς προηγούμενη τιμή της καθώς και η προηγούμενη τιμή της διαφορισμένης χρονοσειράς `predictedDifferenced_WN`.

Σε περίπτωση εποχικής διαφόρισης `seasonal`, εκτελείται μία επανάληψη με εύρος 1 έως `numTest` και μέσα σε αυτήν, αν ο δείκτης `t_idx` είναι μικρότερος από το `seasonalLag`, τότε η τρέχουσα τιμή της νέας χρονοσειράς είναι το άθροισμα των τιμών της τρέχουσας τιμής της διαφορισμένης χρονοσειράς και η τιμή της χρονοσειράς `trainData` στην θέση που απέχει `seasonalLag + 1` θέσεις από το τέλος. Αν ο δείκτης `t_idx` είναι μεγαλύτερος από το `seasonalLag`, τότε η νέα τιμή της χρονοσειράς είναι το άθροισμα της χρονοσειράς `seasonalLag` θέσεις πριν από τον τρέχοντα δείκτη και της τρέχουσας τιμής της διαφορισμένης χρονοσειράς `predictedDifferenced_WN`. Αυτά αποθηκεύονται στην μεταβλητή `predictedOriginal_WN`.

Σε περίπτωση ταυτόχρονης πρώτης τάξης και εποχικής διαφόρισης, όπως αναφέρεται μέσα στον αλγόριθμο ως `both` πραγματοποιούνται δύο αντίστροφες διαφορίσεις, η μία μετά την άλλη, η `first_order` και η `seasonal`. Για την αντίστροφη εποχική διαφόριση, αρχικοποιείται η μεταβλητή `temp_predicted_first_order` σε 0. Έπειτα, εκτελείται μία επανάληψη με εύρος 1 έως `numTest` και μέσα σε αυτήν αν ο δείκτης `t_idx` είναι μικρότερος από το `seasonalLag`, τότε η τρέχουσα τιμή της νέας χρονοσειράς είναι το άθροισμα των τιμών της τρέχουσας τιμής της διαφορισμένης χρονοσειράς και η τιμή της χρονοσειράς `trainData` στην θέση που απέχει `seasonalLag + 1` θέσεις από το τέλος. Αν ο δείκτης `t_idx` είναι μεγαλύτερος από το `seasonalLag`, τότε η νέα τιμή της χρονοσειράς είναι το άθροισμα της χρονοσειράς `temp_predicted_first_order`, `seasonalLag` θέσεις πριν από τον τρέχοντα δείκτη και της τρέχουσας τιμής της διαφορισμένης χρονοσειράς `predictedDifferenced_WN`. Για την αντιστροφή της πρώτης τάξης διαφόρισης, αρχικοποιείται η `predictedOriginal_WN` ίση με το άθροισμα της τελευταίας τιμής της αρχικής χρονοσειράς `trainData` και της πρώτης τιμής της προσωρινής χρονοσειράς που δημιουργήθηκε προηγουμένως, της `temp_predicted_first_order`. Για κάθε επόμενη τιμή σε εύρος από 2 έως `numTest`, στην τρέχουσα τιμή της χρονοσειράς `predictedOriginal_WN` αθροίζεται η προηγούμενη τιμή της καθώς και η τρέχουσα τιμή της `temp_predicted_first_order`.

Στη συνέχεια υπολογίζεται το μέσο τετραγωνικό σφάλμα για την predictedOriginal, το οποίο είναι η διαφορά της χρονοσειράς που προέκυψε από την αντίστροφη διαφόριση με το σύνολο των δεδομένων testData και αποθηκεύεται στο RMSE0.

4.3.2.5 Πρόβλεψη με μέθοδο ARIMA

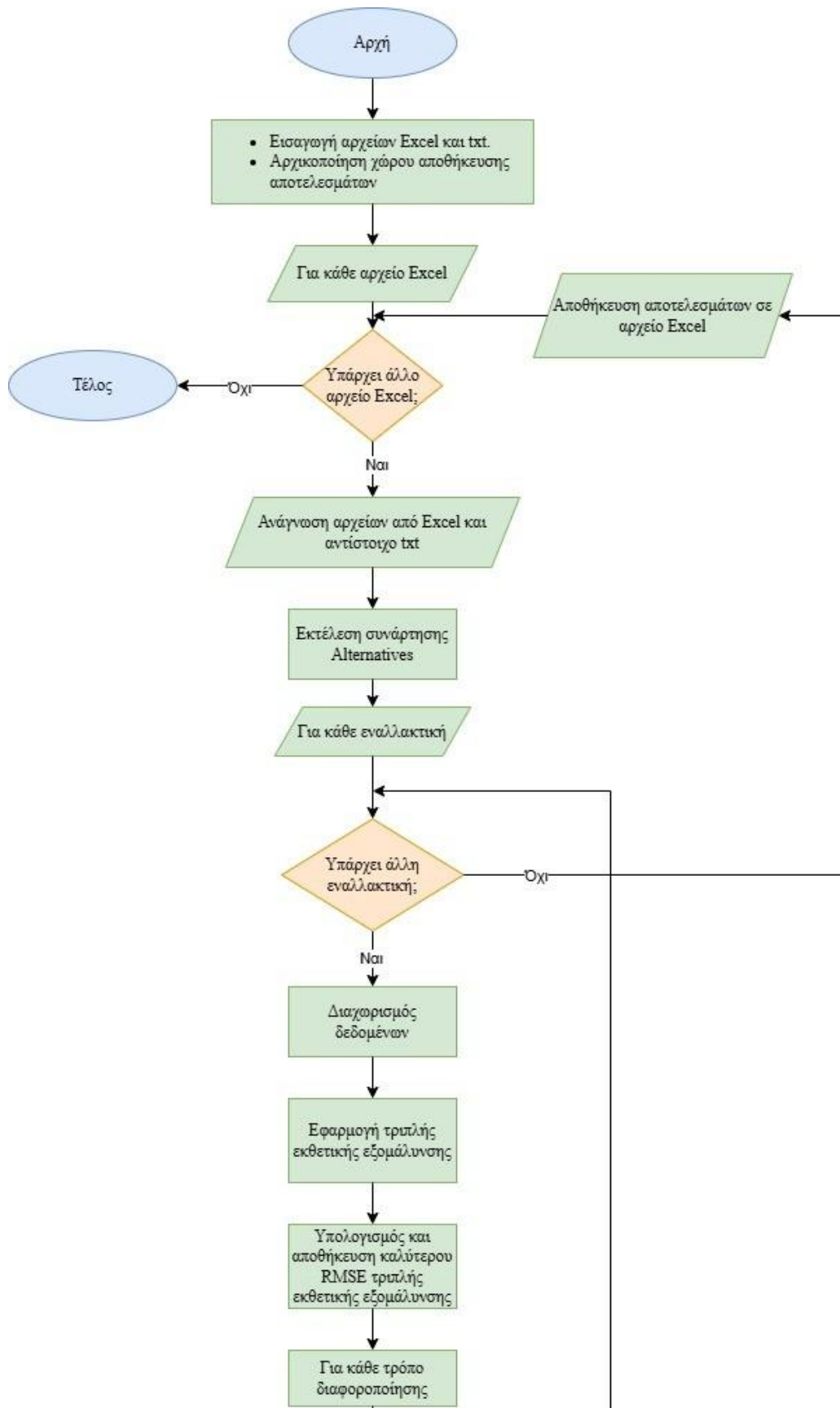
Για την πρόβλεψη με την μέθοδο ARIMA, αρχικοποιούνται οι παράμετροι p , d , q που αφορούν στην τάξη του μέρους που αφορά στην αυτοσυσχέτιση (AR), στην τάξη της διαφόρισης (I) και στην τάξη του μέρους που αφορά στον κινητό μέσο (MA) αντίστοιχα. Η τάξη για τα p και q ορίστηκε από 0 έως 2. Να σημειωθεί σε αυτό το σημείο, ότι η τιμή 0 σε όλα οδηγεί σε αποτελέσματα ίδια με την μέθοδο πρόβλεψης με βάση την μέση τιμή. Ωστόσο, επιλέχτηκε να εισαχθεί και το 0 για να συγκριθούν και να επαληθευτούν τα αποτελέσματα. Για το d ορίστηκε ως 0, εφόσον έχει εφαρμοστεί προηγουμένως διαφόριση.

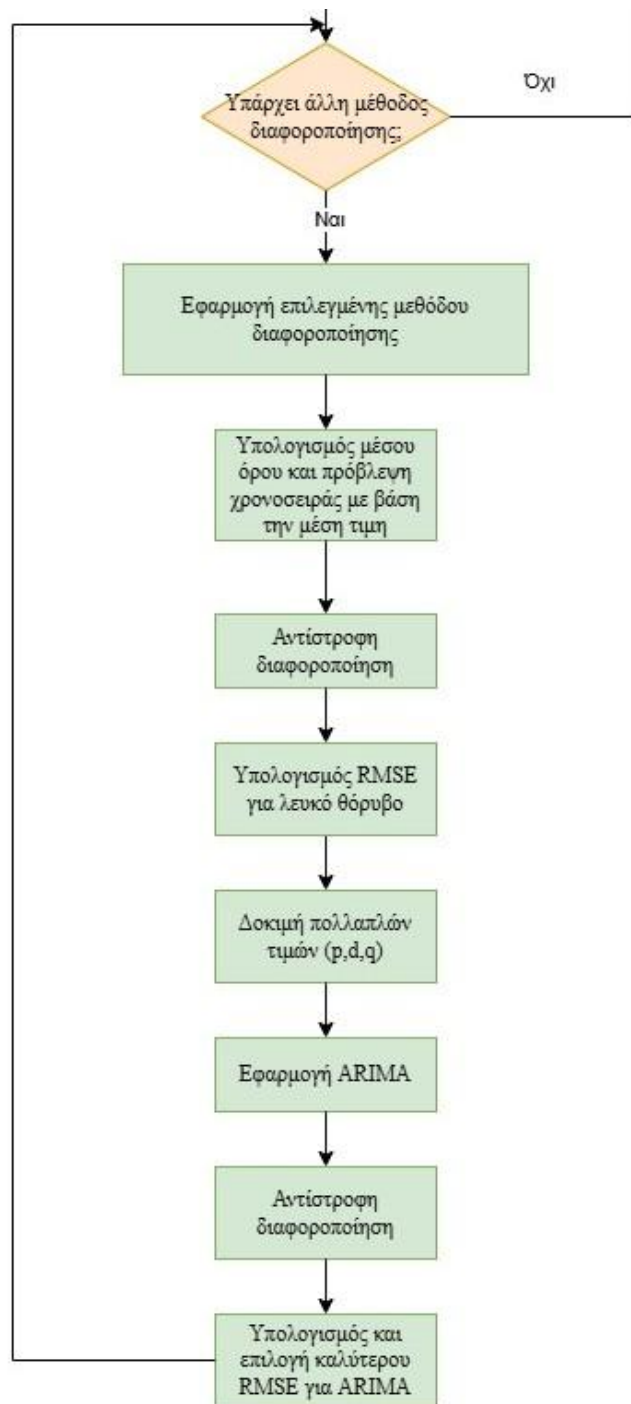
Σε τρεις διαφορετικές επαναλήψεις, η μία μέσα στην άλλη, εφαρμόζεται το μοντέλο ARIMA. Εκτιμώνται οι παράμετροι του μοντέλου και στη συνέχεια γίνεται η πρόβλεψη με βάση τις καλύτερες παραμέτρους που έχουν εκτιμηθεί προηγουμένως. Έπειτα, ανάλογα με την τρέχουσα μέθοδο διαφόρισης που έχει εφαρμοστεί, γίνεται αντίστροφη διαφόριση όπως στην παράγραφο 4.3.2.4. Για τις προβλεπόμενες τιμές υπολογίζεται το τρέχον μέσο τετραγωνικό σφάλμα από την διαφορά των testData με τις προβλεπόμενες τιμές για τον συγκεκριμένο συνδυασμό p, d, q και συγκρίνεται με το καλύτερο bestRMSE_ARIMA που έχει αποθηκευτεί έως αυτή την στιγμή. Εφόσον, είναι μικρότερο από το καλύτερο bestRMSE_ARIMA, αποθηκεύεται στη θέση του. Στη συνέχεια υπολογίζονται οι προβλεπόμενες τιμές για όσες περιόδους μπροστά έχει ορίσει ο χρήστης στην μεταβλητή numTest.

4.3.2.6 Αποθήκευση αποτελεσμάτων

Τέλος, τα αποτελέσματα για κάθε βιβλίο, εναλλακτική και για κάθε διαφόριση που έχει εφαρμοστεί, αποθηκεύονται σε αρχείο Excel με όνομα: RMSE_Results_v10.xlsx. Με αυτόν τον τρόπο ο χρήστης μπορεί να συγκρίνει τα μέσα τετραγωνικά σφάλματα και να επιλέξει για κάθε περίπτωση την καταλληλότερη μέθοδο.

Παρακάτω εμφανίζεται ένα διάγραμμα ροής για την καλύτερη κατανόηση του αλγορίθμου της πρόβλεψης.





Εικόνα 4.2: Διάγραμμα ροής Αλγορίθμου Πρόβλεψης

4.4 Συνάρτηση Holt-Winters

Η συνάρτηση Holt-Winters ακολουθεί ακριβώς την μεθοδολογία που προτείνεται από τον Rob J. Hyndman στο (Hyndman, n.d.). Ως είσοδο δέχεται την χρονοσειρά, τις τιμές για τις παραμέτρους α , β , γ , το μήκος της εποχής m και τον αριθμό των μελλοντικών προβλέψεων. Ως έξοδο δίνει τις προσαρμοσμένες τιμές και τις προβλεπόμενες τιμές.

Αρχικοποιούνται σε πρώτη φάση τα: n (μέγεθος χρονοσειράς), L (επίπεδο), B (τάση), S (εποχικός δείκτης).

Στη συνέχεια για να αρχικοποιηθούν οι μεταβλητές επιπέδου, τάσης και εποχικότητας, χρησιμοποιείται ο κινητός μέσος για να γίνει μία αποσύνθεση των δεδομένων από την τάση και την εποχικότητα. Δηλαδή, εφαρμόζεται ο κινητός μέσος όρος στα τα αρχικά δεδομένα που ανήκουν στις δύο πρώτες εποχές. Αφαιρείται αυτός ο κινητός μέσος όρος από τα αρχικά δεδομένα για να προκύψουν τα δεδομένα χωρίς τάση. Έπειτα, για να αρχικοποιηθεί η συνιστώσα της εποχικότητας, υπολογίζεται ο μέσος όρος της προαναφερθείσας διαφοράς, δηλαδή του κινητού μέσου όρου από τα πραγματικά δεδομένα. Έπειτα αφαιρούνται τα δεδομένα που έχουν προκύψει για την εποχικότητα από τα αρχικά. Στη συνέχεια εφαρμόζεται γραμμική παλινδρόμηση η οποία δίνει το $L(1)$ και το $B(1)$.

Τέλος, σε επανάληψη για κάθε παρατήρηση, αν η παρατήρηση βρίσκεται μετά την πρώτη εποχή, υπολογίζονται οι προσαρμοσμένες τιμές. Διαφορετικά, για την πρώτη παρατήρηση ή για κάθε άλλη παρατήρηση εντός της πρώτης εποχής, υπολογίζονται και ενημερώνονται τα L , B και S σύμφωνα με τις Εξ. 3.10, Εξ. 3.11 και Εξ. 3.12.

Στο τελευταίο στάδιο, υπολογίζονται από την εξίσωση Εξ. 3.9 οι προβλεπόμενες τιμές με βάση τις τελευταίες τιμές των L , B και S .

4.5 Συνάρτηση Syntetos-Boylan

4.5.1 Αρχικοποίηση

Η συνάρτηση που θα αναλυθεί παρακάτω δημιουργήθηκε από τον κώδικα ο οποίος αναφέρεται στην ιστοσελίδα: (MORGAN, χ.χ.) με πολύ μικρές παραλλαγές. Η συνάρτηση των Συντετού και Boylan (SBA) δέχεται ως είσοδο το διάνυσμα της χρονοσειράς που πρόκειται να προβλεφθεί $sales$, την μεταβλητή α η οποία είναι μεταβλητή εξομάλυνσης των πωλήσεων, την μεταβλητή β η οποία είναι μεταβλητή εξομάλυνσης για τα κενά διαστήματα και το $numTest$ ως το μέγεθος της χρονοσειράς που θα χρησιμοποιηθεί για την πρόβλεψη. Για να γίνει καλύτερα κατανοητό, το $numTest$ μπορεί να θεωρηθεί και ως ο αριθμός των περιόδων που θέλει ο χρήστης να προβλέψει. Στην συγκεκριμένη περίπτωση όμως, θα χρησιμοποιηθεί στον κυρίως αλγόριθμο για τον υπολογισμό του μέσου τετραγωνικού σφάλματος ως μέσου αξιολόγησης της μεθόδου. Ως έξοδος δίνεται το διάνυσμα SBA_fitted με τις προσαρμοσμένες τιμές, το $SBA_forecast$ με τις

προβλεπόμενες τιμές καθώς και οι τελευταίες τιμές L και P οι οποίες χρησιμοποιούνται και μέσα στην συνάρτηση για την πρόβλεψη.

Ορίζεται το n για το μήκος της χρονοσειράς sales και αρχικοποιούνται τα: Z για την προβλεπόμενη χρονοσειρά με μήκος όσο η αρχική χρονοσειρά και P για την προβλεπόμενη διακοπτόμενη ζήτηση, δηλαδή για τον προβλεπόμενο χρόνο ανάμεσα σε δύο συνεχόμενες μη μηδενικές τιμές στην χρονοσειρά. Επιπλέον αρχικοποιείται και το διάνυσμα sba_fitted για την αποθήκευση των δεδομένων με την χρήση των διανυσμάτων Z και P, καθώς και ο δείκτης positive_demand_idx για τον εντοπισμό των παρατηρήσεων που παρουσιάζουν ζήτηση. Συγκεκριμένα το διάνυσμα positive_demand_idx αποθηκεύει τις θέσεις που βρίσκονται οι θετικές τιμές στο διάνυσμα των πωλήσεων.

4.5.2 Κύριο μέρος συνάρτησης

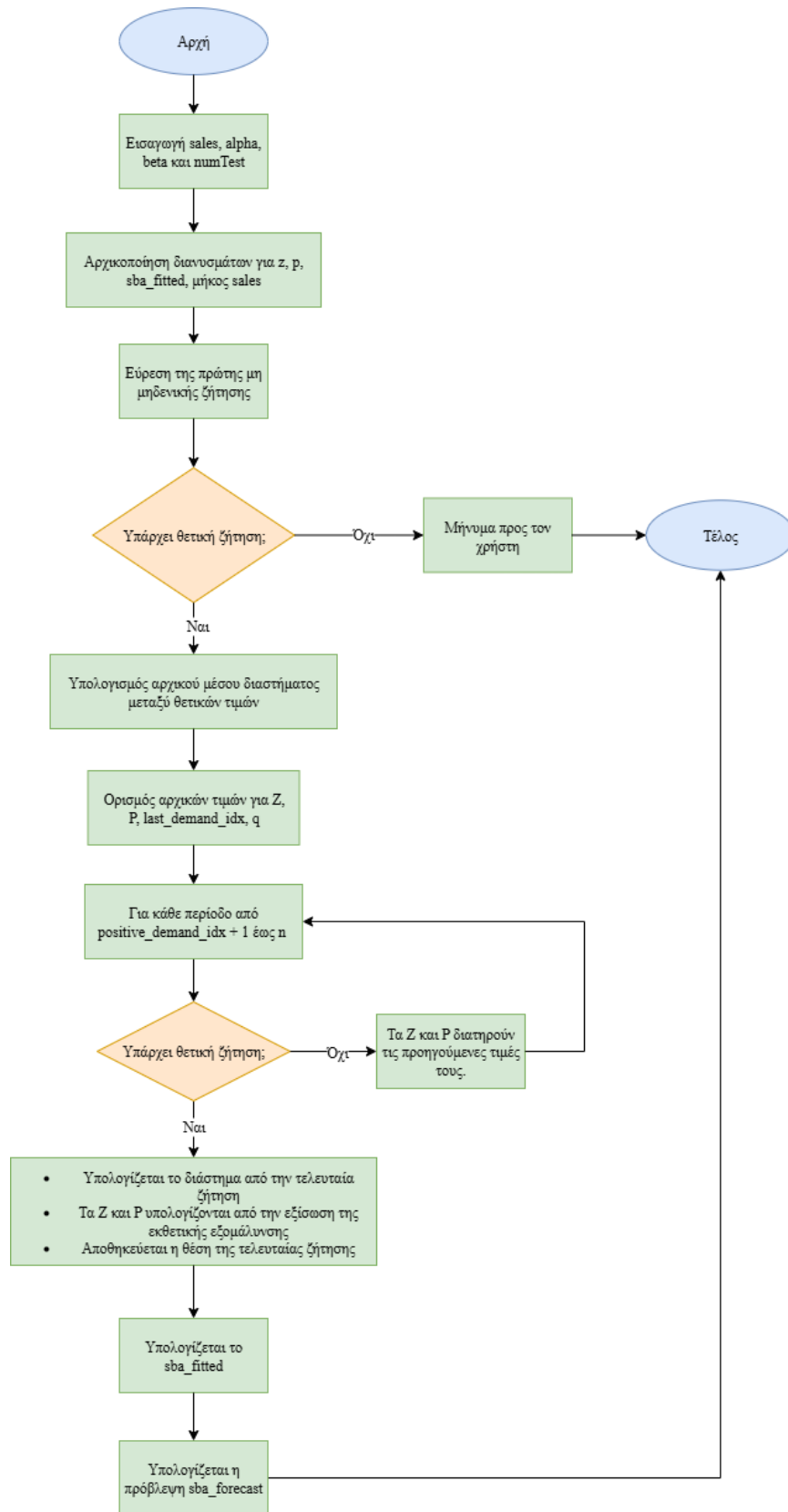
Στο κύριο μέρος ελέγχεται αν τελικά αυτός ο δείκτης είναι κενός ή όχι. Αν είναι κενός, τότε ολόκληρη η χρονοσειρά περιέχει μηδενικές ή αρνητικές τιμές.

Στη συνέχεια, αρχικοποιείται η πρώτη τιμή για το διάνυσμα Z (που αφορά στις πωλήσεις) ως η πρώτη θετική τιμή από την χρονοσειρά sales. Επιπλέον, αρχικοποιείται και το διάνυσμα P (που αφορά στις θέσεις των παρατηρήσεων με θετική ζήτηση) με την θέση της πρώτης παρατήρησης με θετική ζήτηση. Αρχικοποιείται, τέλος η μεταβλητή q που θα χρησιμοποιηθεί ως μετρητής για τα διανύσματα και η μεταβλητή last_demand_idx παίρνει την τιμή της θέσης της πρώτης τιμής με θετική ζήτηση.

Σε επανάληψη για διάστημα από την πρώτη μη μηδενική τιμή που εντοπίστηκε και μέχρι το τέλος της χρονοσειράς, αν η παρατήρηση στο διάνυσμα sales είναι θετική, τότε υπολογίζονται οι μεταβλητές Z και P, η μεταβλητή q παίρνει την τιμή 1 και η last_demand_idx την τιμή του τρέχοντος δείκτη t. Διαφορετικά, οι μεταβλητές Z και P κρατάνε τις τελευταίες τιμές τους και η μεταβλητή q αυξάνεται κατά 1. Έπειτα πραγματοποιείται η πρόβλεψη για το sba_fitted που αφορά στο σύνολο trainData.

Τέλος, χρησιμοποιώντας τα τελευταία Z και P, χρησιμοποιείται ο τύπος πρόβλεψης της προσέγγισης Syntetos-Boylan για την πρόβλεψη των επόμενων περιόδων.

Παρακάτω παρουσιάζεται το διάγραμμα ροής της μεθόδου.



Εικόνα 4.3: Διάγραμμα ροής προσέγγισης Syntetos-Boylan

4.6 Αλγόριθμος προσέγγισης Syntetos-Boylan, απλού κινητού μέσου και εκθετικής εξομάλυνσης

4.6.1 Αρχικοποίηση και αποθήκευση

Στο πρώτο μέρος του αλγορίθμου εισάγονται και αποθηκεύονται τα αρχεία Excel που περιέχουν τα δεδομένα, τα αρχεία κειμένου που περιέχουν τον τρόπο που ο χρήστης επιθυμεί να γίνει ομαδοποίηση και το αρχείο Excel, στο οποίο θα γίνει η αποθήκευση των τελικών αποτελεσμάτων. Κατόπιν ο αλγόριθμος αρχικοποιεί την μεταβλητή RMSE_Results για τα αποτελέσματα από το μέσο τετραγωνικό σφάλμα (root mean squared error) που θα υπολογιστεί αργότερα.

4.6.2 Κύριο μέρος αλγορίθμου

Το κύριο μέρος του αλγορίθμου αποτελείται από μία επανάληψη για κάθε αρχείο Excel ξεχωριστά. Αρχικά καλείται η συνάρτηση Function_Alternatives προκειμένου να γίνει η ομαδοποίηση.

Έπειτα για κάθε εναλλακτική που έχει δημιουργηθεί από την συνάρτηση Function_Alternatives, αρχικοποιούνται οι μεταβλητές για την αποθήκευση των καλύτερων RMSE για κάθε μέθοδο, καθώς και οι μεταβλητές για την αποθήκευση των παραμέτρων που δίνουν το μικρότερο RMSE για κάθε μέθοδο. Δηλαδή για τις παραμέτρους a, b για την προσέγγιση Syntetos-Boylan, alpha για την απλή εκθετική εξομάλυνση και το διάστημα window_size για την μέθοδο κινητού μέσου.

Στη συνέχεια σε διπλή επανάληψη για την εύρεση των καλύτερων μεταβλητών alpha και beta, καλείται η συνάρτηση Syntetos_Boylan_Approximation και υπολογίζεται το RMSE. Η κάθε τιμή του RMSE, συγκρίνεται με αυτήν της μεταβλητής min_rmse_sba και αποθηκεύεται στην ίδια μεταβλητή αν είναι μικρότερη. Στο τέλος των παραπάνω επαναλήψεων, έχει βρεθεί το ελάχιστο RMSE για την προσέγγιση Syntetos-Boylan και τα alpha και beta που οδηγούν σε αυτό το RMSE. Χρησιμοποιώντας τα παραπάνω και το σετ δεδομένων trainData καλείται εκ νέου η συνάρτηση Syntetos_Boylan_Approximation για την πρόβλεψη. Έπειτα υπολογίζεται το τελικό RMSE.

Συνεχίζει ο αλγόριθμος με την διαφορά της χρονοσειράς, προκειμένου να γίνει στάσιμη και να μην είναι λευκός θόρυβος για την εφαρμογή της απλής εκθετικής εξομάλυνσης, όπως στην 4.3.2.2.

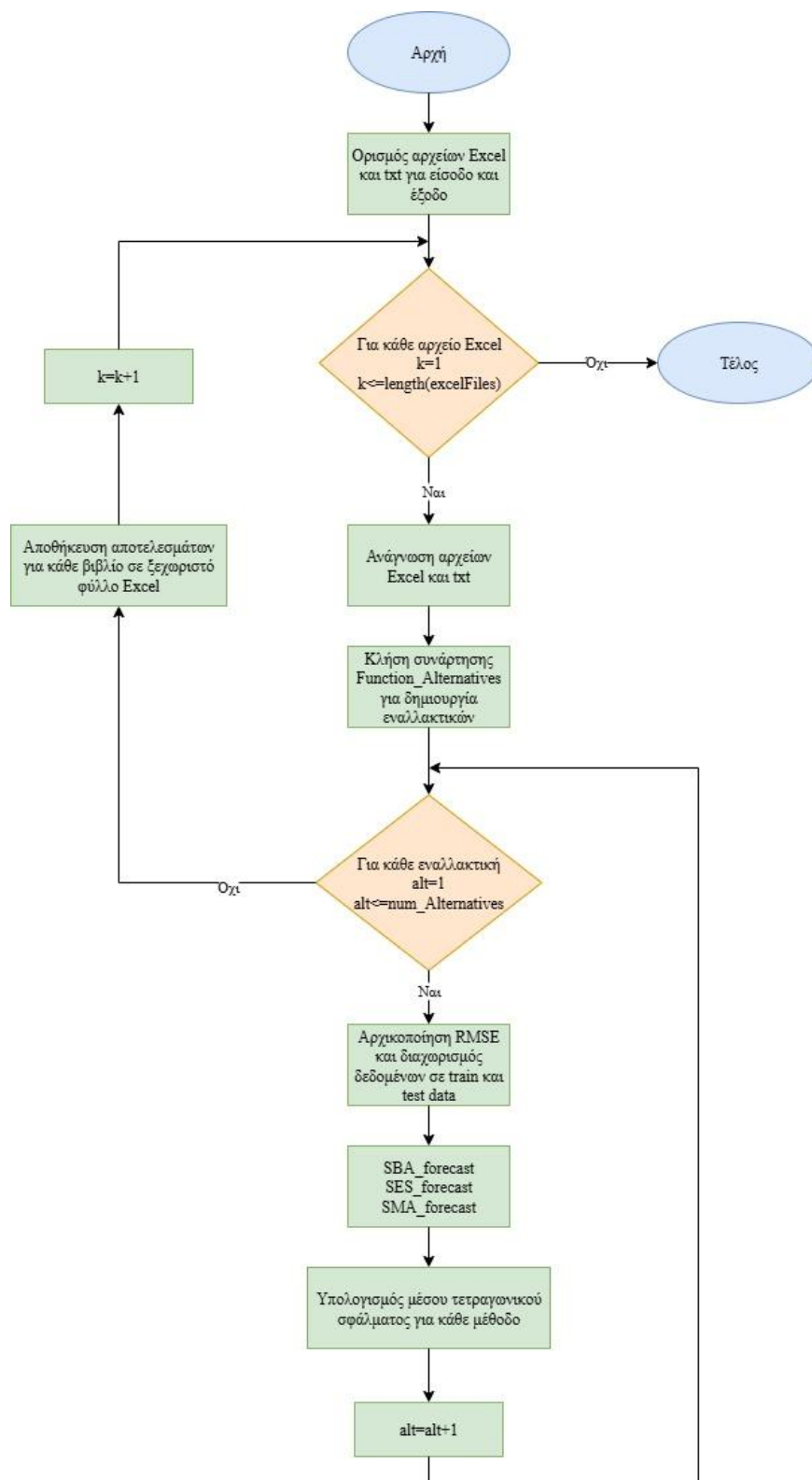
Για την εύρεση της παραμέτρου που δίνει το μικρότερο RMSE για την απλή εκθετική εξομάλυνση, εκτελείται μία επανάληψη μέσα στην οποία εφαρμόζεται η εξίσωση της εξομάλυνσης. Έπειτα, υπολογίζεται το RMSE αφαιρώντας τις εξομαλυσμένες τιμές από τις διαφορισμένες τιμές και αποθηκεύεται το μικρότερο RMSE και στη συνέχεια υπολογίζεται η πρόβλεψη για numTest περιόδους μετά. Τέλος, εφαρμόζεται αντίστροφη διαφορά για τα δεδομένα που εξήχθησαν και υπολογίζεται το τελικό RMSE.

Για την τελευταία μέθοδο, αυτήν του κινητού μέσου, σε επανάληψη για το καλύτερο διάστημα window_size, εφαρμόζεται η εξίσωση του κινητού μέσου και αντίστοιχα υπολογίζεται το μικρότερο RMSE και αποθηκεύεται μαζί με την παράμετρο για το καλύτερο διάστημα.

Χρησιμοποιώντας τις παραπάνω παραμέτρους, εφαρμόζονται οι εξισώσεις των μεθόδων για τα υπάρχοντα δεδομένα της χρονοσειράς καθώς και οι προβλέψεις για κάθε μέθοδο για μία περίοδο μετά. Έπειτα υπολογίζεται το τελικό RMSE.

Το τελευταίο κομμάτι του αλγορίθμου αφορά στην οργάνωση και αποθήκευση των δεδομένων για την διευκόλυνση του χρήστη.

Παρακάτω ακολουθεί το διάγραμμα ροής για την καλύτερη κατανόηση του αλγορίθμου.



Εικόνα 4.4: Διάγραμμα ροής αλγορίθμου πρόβλεψης με την χρήση προσέγγισης Syntetos-Boylan, απλού κινητού μέσου και εκθετικής εξομάλυνσης.

4.7 Αλγόριθμος διπλής εκθετικής εξομάλυνσης για ακανόνιστα δεδομένα.

Για τον αλγόριθμο Double exponential smoothing for irregular data, ακολουθείται πιστά η επιστημονική εργασία του Tomas Cipra (Cipra, 2005).

Στο πρώτο μέρος του αλγορίθμου εισάγονται και αποθηκεύονται τα αρχεία Excel που περιέχουν τα δεδομένα πωλήσεων και τα αρχεία κειμένου που περιέχουν τον τρόπο που ο χρήστης επιθυμεί να γίνει ομαδοποίηση.

Στη συνέχεια αρχικοποιούνται τα αρχεία στα οποία θα αποθηκευτούν τα αποτελέσματα των RMSE, καθώς και οι προβλέψεις.

Το κύριο μέρος του αλγορίθμου αποτελείται από μία επανάληψη για κάθε αρχείο Excel ξεχωριστά. Αρχικά καλείται η συνάρτηση Function_Alternatives προκειμένου να γίνει η ομαδοποίηση.

Έπειτα, σε επανάληψη για κάθε εναλλακτική, αποθηκεύεται η τρέχουσα εναλλακτική και το μέγεθός της, στην μεταβλητή n το μήκος του πίνακα των εναλλακτικών. Στο συγκεκριμένο σημείο διαχωρίζονται τα δεδομένα σε train και test set. Τα 5 τελευταία δεδομένα στη χρονοσειρά αποθηκεύονται στο test set για τον υπολογισμό του τελικού RMSE στο τέλος.

Στην συνέχεια για τον εντοπισμό των διαστημάτων χωρίς πωλήσεις, κάθε στήλη της μεταβλητής currentTable, αποθηκεύεται σε δύο διαφορετικές μεταβλητές t_raw και y_raw . Οι t_raw αφορούν τις παρατηρήσεις για τον μήνα και οι y_raw για τις πωλήσεις. Αντίστοιχα, αποθηκεύονται στο t_raw_test και στο y_raw_test οι στήλες που αφορούν στο test set.

Αρχικοποιούνται οι μεταβλητές για την αποθήκευση του καλύτερου RMSE και για το καλύτερο α . Δημιουργείται, έπειτα, μία επανάληψη για το α_{init} , που λειτουργεί ως συντελεστής εξομάλυνσης, μέσα στην οποία εκτελείται το κύριο μέρος του αλγορίθμου για την εύρεση του καλύτερου α με βάση το καλύτερο RMSE.

Αρχικοποιούνται το διάνυσμα α για την αποθήκευση των τιμών α που επαναυπολογίζονται αργότερα, το q που είναι το μέσο χρονικό διάστημα μεταξύ των δεδομένων και η μεταβλητή $\alpha(1)$ (Εξ. 3.17) ως αρχικοποίηση του α . Επιπλέον, ορίζεται ο συντελεστής απόσβεσης β και οι αρχικές τιμές των βοηθητικών μεταβλητών $w(1)$ και $z(1)$ (Εξ. 3.19 και Εξ. 3.20) ως αρχικοποίηση των w και z . Το w είναι το βάρος που δίνεται ανάλογα με την χρονική απόσταση των τιμών και το z επηρεάζει το πόσο έντονα προσαρμόζεται η τάση της εξομάλυνσης.

Στη συνέχεια χρησιμοποιούνται οι 5 πρώτες παρατηρήσεις για την εκτίμηση της αρχικής τάσης και την κλίση της $b1_0$ με τον τύπο των ελαχίστων τετραγώνων και τη σταθερά της $b0_0$. Τα $b1_0$ και $b0_0$ είναι εκτιμήσεις παλινδρόμησης της τάσης $b_0 + b_1 t$ σε ένα αρχικό σύνολο παρατηρήσεων μήκους n_0 .

Αρχικοποιούνται έπειτα το επίπεδο S_level και η τάση S_trend (Εξ. 3.21 και Εξ. 3.22).

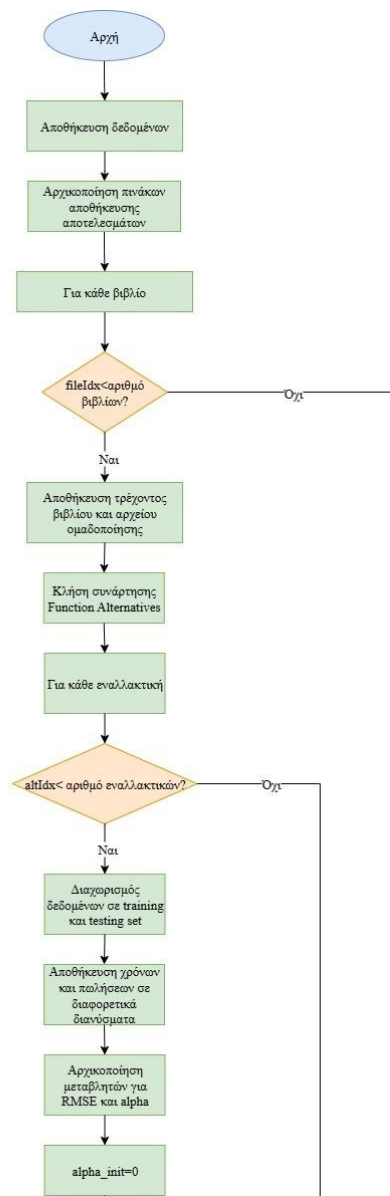
Έπειτα για κάθε τιμή της χρονοσειράς εκτός από την πρώτη, υπολογίζονται εκ νέου οι μεταβλητές: α , w , z , η στάθμη και η τάση S_level και S_trend , σύμφωνα με τις εξισώσεις Εξ. 3.25 και Εξ. 3.26 αντίστοιχα.

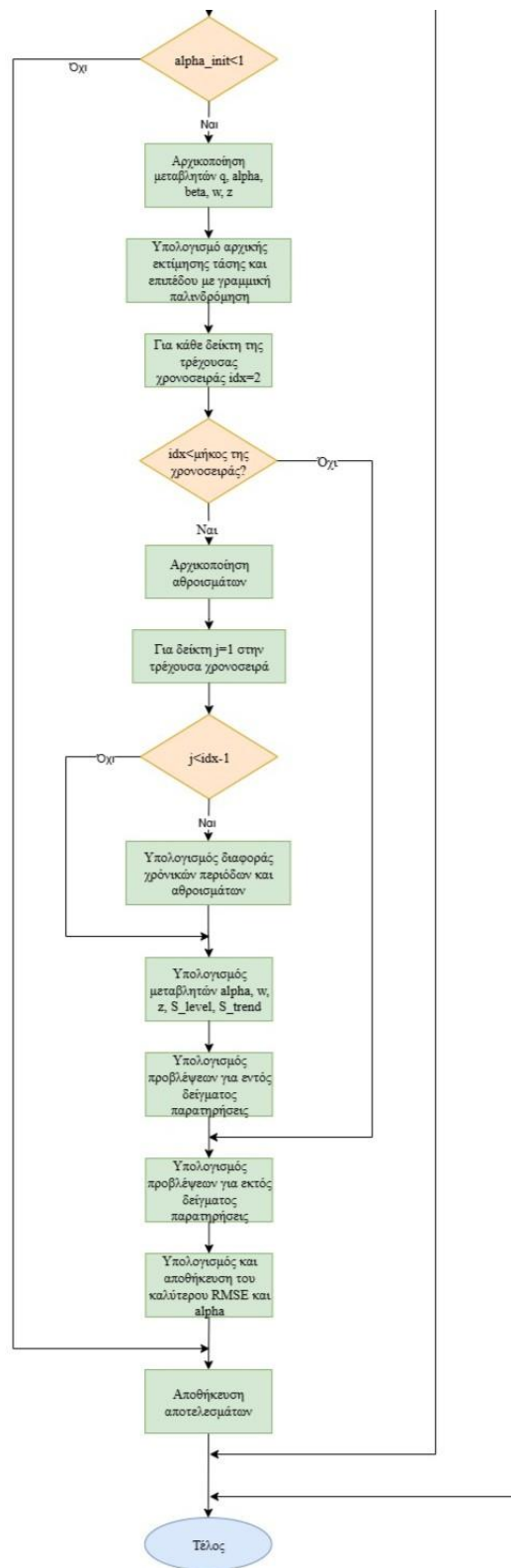
Τέλος, υπολογίζεται η πρόβλεψη για τις υπάρχουσες τιμές της χρονοσειράς, σύμφωνα με την εξίσωση Εξ. 3.27.

Στη συνέχεια, και πριν ολοκληρωθεί η επανάληψη για την τρέχουσα εναλλακτική, υπολογίζεται και αποθηκεύεται το καλύτερο root mean squared error για τις προβλέψεις στο test set.

Τέλος, γίνεται η αποθήκευση των αποτελεσμάτων των RMSE και των προβλέψεων.

Το διάγραμμα ροής του αλγορίθμου παρουσιάζεται παρακάτω.





Εικόνα 4.5: Διάγραμμα ροής αλγορίθμου Διπλής Εκθετικής εξομάλυνσης για ακανόνιστα δεδομένα (Cipra, 2005)

5. Αποτελέσματα

Στο συγκεκριμένο κεφάλαιο παρουσιάζονται τα αποτελέσματα των προηγούμενων αλγορίθμων και συναρτήσεων. Στόχος της παρούσας εργασίας είναι η εξαγωγή προβλέψεων με την χρήση διαφορετικών αλγορίθμων και η σύγκριση των αποτελεσμάτων των αλγορίθμων μεταξύ τους. Συγκεκριμένα, θα παρουσιαστούν αρχικά τα αποτελέσματα του αλγορίθμου πρόβλεψης με την χρήση των μεθόδων μέσου όρου, ARIMA και Holt-Winters και ακολούθως της προσέγγισης Syntetos-Boylan, της απλής εκθετικής εξομάλυνσης και του κινητού μέσου όρου. Ο τελευταίος αλγόριθμος αφορά την μέθοδο της διπλής εκθετικής εξομάλυνσης για ακανόνιστα δεδομένα. Τα αποτελέσματα των παραπάνω αλγορίθμων θα συγκριθούν μεταξύ τους. Όπως αναφέρθηκε και στο Κεφάλαιο 2.2, τα δεδομένα που αναλύθηκαν για το βιβλίο Α ήταν 617, για το Β ήταν 290 και για το Γ ήταν 316. Είναι σημαντικό να σημειωθεί ότι ο μικρός αριθμός των δεδομένων, επηρεάζει την εγκυρότητα των αποτελεσμάτων.

5.1 Συνάρτηση ομαδοποίησης.

Σε αυτή την υποενότητα θα παρουσιαστούν τρεις διαφορετικές επιλογές για την ομαδοποίηση των βιβλίων, όπως αυτή επιλέγεται από τον χρήστη και εξάγεται από την συνάρτηση ομαδοποίησης, η οποία παίζει κρίσιμο ρόλο στην εξαγωγή των αποτελεσμάτων των αλγορίθμων. Όπως αναφέρθηκε και στο Κεφάλαιο 2.3, η ομαδοποίηση ορίζεται από τον χρήστη σε διαφορετικά αρχεία κειμένου, κάθε ένα από τα οποία αφορά σε ένα βιβλίο. Συνολικά εξετάζονται και ελέγχονται τρία βιβλία.

Παρουσιάζονται παρακάτω ενδεικτικά δύο διαφορετικές ομαδοποιήσεις με τρεις εναλλακτικές για κάθε ομαδοποίηση. Ο χρήστης είναι ελεύθερος να επιλέξει όσες εναλλακτικές και όσους μήνες επιθυμεί σε κάθε ομαδοποίηση. Στην παρούσα εργασία επιλέχθηκαν έως 4 διαφορετικές ομάδες μηνών σε κάθε εναλλακτική για εξοικονόμηση πόρων και χρόνου.

Η πρώτη ομαδοποίηση διαμορφώνεται ως εξής:

1 ^η Ομαδοποίηση				
1 ^η εναλλακτική				
Μήνες	1	4	7	10
2 ^η εναλλακτική				
Μήνες	3	6	9	12
3 ^η εναλλακτική				
Μήνες	5	8	11	2

Πίνακας 5.1: Πρώτη ομαδοποίηση

Η δεύτερη ομαδοποίηση παρουσιάζεται παρακάτω:

2η Ομαδοποίηση			
1η εναλλακτική			
Μήνες	1	5	9
2η εναλλακτική			
Μήνες	4	8	12
3η εναλλακτική			
Μήνες	9	1	5

Πίνακας 5.2: Δεύτερη ομαδοποίηση

5.2 Ομαδοποίηση 1

Στους παρακάτω πίνακες εμφανίζονται τα αποτελέσματα της **πρώτης** ομαδοποίησης για κάθε βιβλίο χωριστά, όπως τα εξάγουν οι αλγόριθμοι.

5.2.1 Μέσος όρος, ARIMA και Holt-Winters

Ο συγκεκριμένος αλγόριθμος παράγει ως αποτελέσματα τα Root mean square errors (RMSE) για κάθε βιβλίο, για κάθε εναλλακτική και για κάθε μέθοδο πρόβλεψης και διαφόρισης, καθώς και τον αριθμό των προβλέψεων που έχει ορίσει ο χρήστης. Στην συγκεκριμένη περίπτωση ο αριθμός των προβλέψεων ορίστηκε ως τρεις.

Παρακάτω παρουσιάζονται τα αποτελέσματα των RMSE για τον πρώτο τρόπο ομαδοποίησης, όπως αυτά εξήχθησαν από τον αλγόριθμο πρόβλεψης με την χρήση των μεθόδων μέσου όρου, ARIMA και Holt-Winters. Παρουσιάζονται χωριστά τα αποτελέσματα για κάθε βιβλίο. Σε κάθε πίνακα εμφανίζονται οι διαφορετικές εναλλακτικές για κάθε μέθοδο διαφόρισης.

Βιβλίο: A

Εναλλακτική	Μέθοδος διαφόρισης	WhiteNoise_RMSE	Best_ARIMA_RMSE	Holt_Winters_RMSE
1	none	287.49	285.32	135.75
1	first_order	233.95	175.55	-
1	seasonal	157.9	157.9	-
1	both	244.87	244.87	-
2	none	385.1	68.52	73.71
2	first_order	386.43	176.65	-
2	seasonal	217.71	217.71	-
2	both	1835.11	343.74	-
3	none	334.71	288.5	123.67
3	first_order	88.62	88.62	-
3	seasonal	213.32	213.32	-
3	both	243.46	243.46	-

Πίνακας 5.3: Αποτελέσματα RMSE για το βιβλίο A

Βιβλίο: Β

Εναλλακτική	Μέθοδος διαφορίσης	WhiteNoise_RMSE	Best_ARIMA_RMSE	Holt_Winters_RMSE
1	none	45.35	44.76	21.09
1	first_order	56.1	40.14	-
1	seasonal	101.14	46	-
1	both	423.87	326.43	-
2	none	44.71	44.49	15.61
2	first_order	40.61	38.55	-
2	seasonal	66.94	54.15	-
2	both	371.84	97.78	-
3	none	70.05	58.78	51.10
3	first_order	93.2	78.78	-
3	seasonal	89.03	89.03	-
3	both	132.51	132.51	-

Πίνακας 5.4: Αποτελέσματα RMSE για το βιβλίο Β

Βιβλίο: Γ

Εναλλακτική	Μέθοδος διαφορίσης	WhiteNoise_RMSE	Best_ARIMA_RMSE	Holt_Winters_RMSE
1	none	120.99	120.99	58.47
1	first_order	183.54	123.41	-
1	seasonal	213.82	176.05	-
1	both	752.65	185.02	-
2	none	165.67	139.47	160.48
2	first_order	217.06	161.59	-
2	seasonal	351.03	239.27	-
2	both	2048.37	303.77	-
3	none	192.65	192.65	172.16
3	first_order	191.17	181.27	-
3	seasonal	201.34	196.89	-
3	both	243.71	243.71	-

Πίνακας 5.5: Αποτελέσματα RMSE για το βιβλίο Γ

Σύμφωνα με τους παραπάνω πίνακες, παρατηρείται ότι για το βιβλίο Α, το μικρότερο RMSE εντοπίζεται για την εναλλακτική επιλογή 2, με την μέθοδο ARIMA χωρίς διαφορίση. Το ίδιο ισχύει και για την εναλλακτική 3. Ωστόσο, η εναλλακτική 1, εμφανίζει μικρότερο RMSE με την μέθοδο Holt-Winters.

Για το βιβλίο Β, το καλύτερο RMSE εμφανίζεται στην δεύτερη εναλλακτική με την μέθοδο Holt-Winters. Αντίστοιχα για τις εναλλακτικές 1 και 3 η καλύτερη μέθοδος εμφανίζεται να είναι η Holt-Winters.

Όσον αφορά στο βιβλίο Γ, το καλύτερο RMSE φαίνεται να ανήκει στην πρώτη εναλλακτική, με την μέθοδο Holt-Winters. Αντίστοιχα για την εναλλακτική 3, η καλύτερη μέθοδος εμφανίζεται να είναι, επίσης, η Holt-Winters. Όμως για την εναλλακτική 2, ως η καλύτερη μέθοδος εμφανίζεται η ARIMA.

Σημαντικό είναι να σημειωθεί μία παρατήρηση η οποία αποδεικνύει την σωστή λειτουργία του αλγορίθμου. Παρατηρείται συχνά στα RMSE σε διαφορετικές περιπτώσεις, ότι τα RMSE της μεθόδου ARMA είναι αρκετά κοντά με αυτά της μεθόδου του Μέσου όρου. Αυτό συμβαίνει καθώς το καλύτερο RMSE της μεθόδου ARMA, προέκυψε από το μοντέλο $(p, d, q) = (0, 0, 0)$. Όταν, το παραπάνω μοντέλο εφαρμόζεται, η μέθοδος ARMA ακολουθεί την λογική του μέσου όρου. Επομένως, για τον παραπάνω λόγο εμφανίζονται ίδια αποτελέσματα, δηλαδή η μέθοδος ARMA δεν μπορεί να εντοπίσει κάποιο μοτίβο. Επιπλέον αυτό που παρατηρείται είναι ότι σε αρκετές από τις παραπάνω εναλλακτικές, η μέθοδος που συνδυάζει τις δύο διαφορίσεις (both) εμφανίζει το μεγαλύτερο RMSE, γεγονός που δείχνει ότι πιθανότατα έχει πραγματοποιηθεί υπερβολική διαφορίση, η οποία δεν χρειάζεται και δεν οδηγεί σε καλά αποτελέσματα. Το ίδιο φαινόμενο παρατηρείται και στην διαφορίση κατά εποχή, η οποία παρουσιάζει χειρότερα αποτελέσματα.

5.2.1.1 Προβλέψεις για χρονικό ορίζοντα 5 περιόδων

Τα αποτελέσματα των προβλεπόμενων πωλήσεων (βιβλίων σε τεμάχια) για τις περιόδους 1 έως 5, εμφανίζονται στους παρακάτω πίνακες για κάθε μέθοδο πρόβλεψης χωριστά. Για την μέθοδο ARIMA τα αποτελέσματα φαίνονται στον παρακάτω πίνακα:

Βιβλίο	Εναλλακτική	Μέθοδος διαφορίσης	Πρόβλεψη 1 ^{ης} περιόδου	Πρόβλεψη 2 ^{ης} περιόδου	Πρόβλεψη 3 ^{ης} περιόδου	Πρόβλεψη 4 ^{ης} περιόδου	Πρόβλεψη 5 ^{ης} περιόδου
A	1	none	425.52	355.42	367	365.08	365.4
A	1	first_order	66.33	42.71	25.44	35.55	24.85
A	1	seasonal	120.9	72.9	-13.1	106.8	58.8
A	1	both	119.26	189.53	173.79	275.32	328.84
A	2	none	35.77	92.97	48.2	83.24	55.81
A	2	first_order	186.23	162.21	258.94	185.39	186.41
A	2	seasonal	190	-30	369	159	-61
A	2	both	-418.84	-115.29	362.17	-379.21	1.8
A	3	none	600.98	278.32	-25.53	210.77	562.51
A	3	first_order	89.14	90.27	91.41	92.55	93.68
A	3	seasonal	45.85	384.85	116.85	74.7	413.7
A	3	both	-58	135	60	-249	-219
B	1	none	51.95	56.48	52.76	55.81	53.31
B	1	first_order	75.45	48.9	42.93	57.91	55.25
B	1	seasonal	73.02	-4.47	17.64	74.62	-1.77
B	1	both	-217.05	-191.99	-177.93	-400.91	-371.79
B	2	none	58.53	61.71	62.9	63.34	63.51
B	2	first_order	44.22	46.5	47.39	48.52	49.6
B	2	seasonal	-14.47	79.23	21.37	5.59	75.31
B	2	both	-30.33	95.78	93.6	52.52	168.5
B	3	none	84.33	47.7	36.77	70.91	83.61
B	3	first_order	44.89	56.63	80.96	61.19	69.05
B	3	seasonal	27.38	37.38	111.38	36.75	46.75
B	3	both	101.2	110.4	193.6	174	164.4
Γ	1	none	108.58	108.58	108.58	108.58	108.58
Γ	1	first_order	172.99	197.55	119.29	150.76	190.97

Βιβλίο	Εναλλακτική	Μέθοδος διαφόρισης	Πρόβλεψη 1 ^{ης} περιόδου	Πρόβλεψη 2 ^{ης} περιόδου	Πρόβλεψη 3 ^{ης} περιόδου	Πρόβλεψη 4 ^{ης} περιόδου	Πρόβλεψη 5 ^{ης} περιόδου
Γ	1	seasonal	68.15	244.45	11.47	-3.16	283.14
Γ	1	both	-37.97	117.28	117.53	79.81	235.31
Γ	2	none	19.29	210.26	40.98	225.18	51.24
Γ	2	first_order	100.99	176.86	143.31	175.94	168.54
Γ	2	seasonal	-163.8	2	196	-173.8	-8
Γ	2	both	-159.41	-17.53	318.35	82.81	354.57
Γ	3	none	120.16	120.16	120.16	120.16	120.16
Γ	3	first_order	207	171.42	184.71	210.37	205.49
Γ	3	seasonal	205.73	146.51	213.51	245.24	186.02
Γ	3	both	153.67	149.33	212	80.33	-35.33

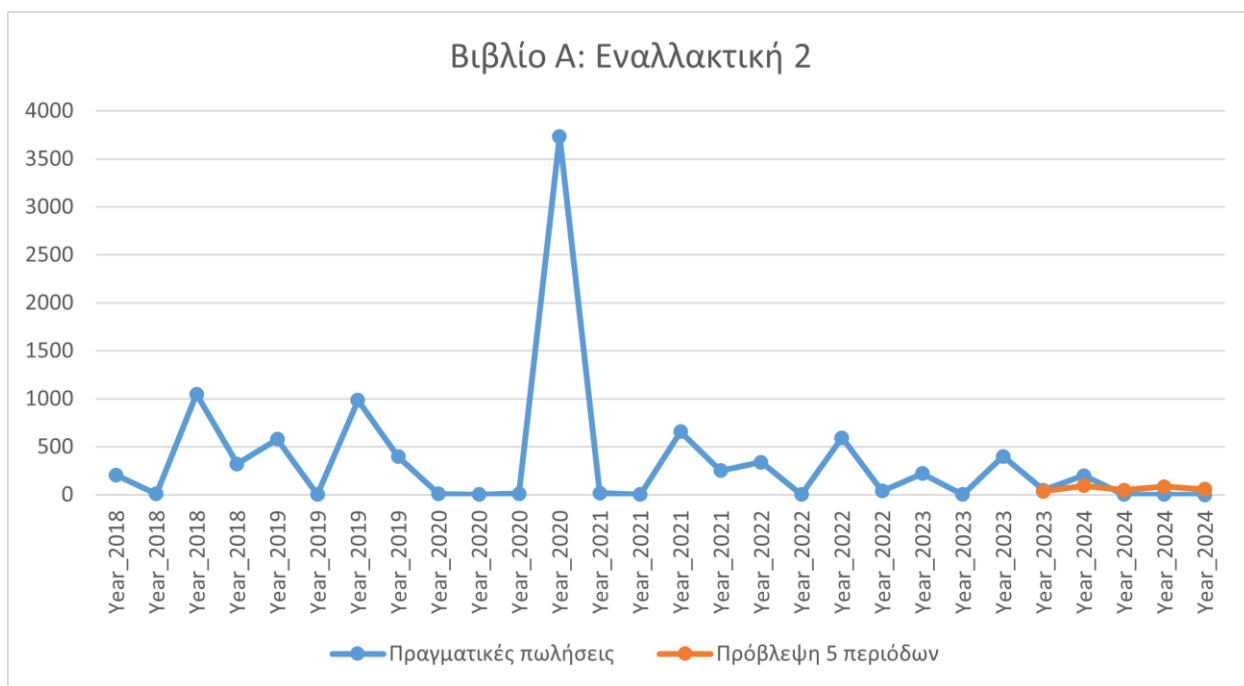
Πίνακας 5.6: Προβλέψεις μεθόδου ARIMA

Για την μέθοδο Holt-Winters τα αποτελέσματα φαίνονται στον Πίνακα 5.7:

Βιβλίο	Εναλλακτική	Μέθοδος διαφόρισης	Πρόβλεψη 1 ^{ης} περιόδου	Πρόβλεψη 2 ^{ης} περιόδου	Πρόβλεψη 3 ^{ης} περιόδου	Πρόβλεψη 4 ^{ης} περιόδου	Πρόβλεψη 5 ^{ης} περιόδου
A	1	none	199.62	237.28	2.34	-124.46	139.79
A	2	none	135.13	93.33	-59.10	-49.43	45.59
A	3	none	341.11	-6.31	-146.17	170.49	245.34
B	1	none	108.61	71.17	46.75	4.57	18.11
B	2	none	90.72	81.17	30.07	-26.28	18.70
B	3	none	90.03	69.56	20.27	23.82	51.42
Γ	1	none	174.02	220.57	168.06	78.70	47.92
Γ	2	none	117.45	61.92	51.98	-16.85	60.54
Γ	3	none	198.19	213.68	99.16	123.50	150.53

Πίνακας 5.7: Προβλέψεις μεθόδου Holt-Winters

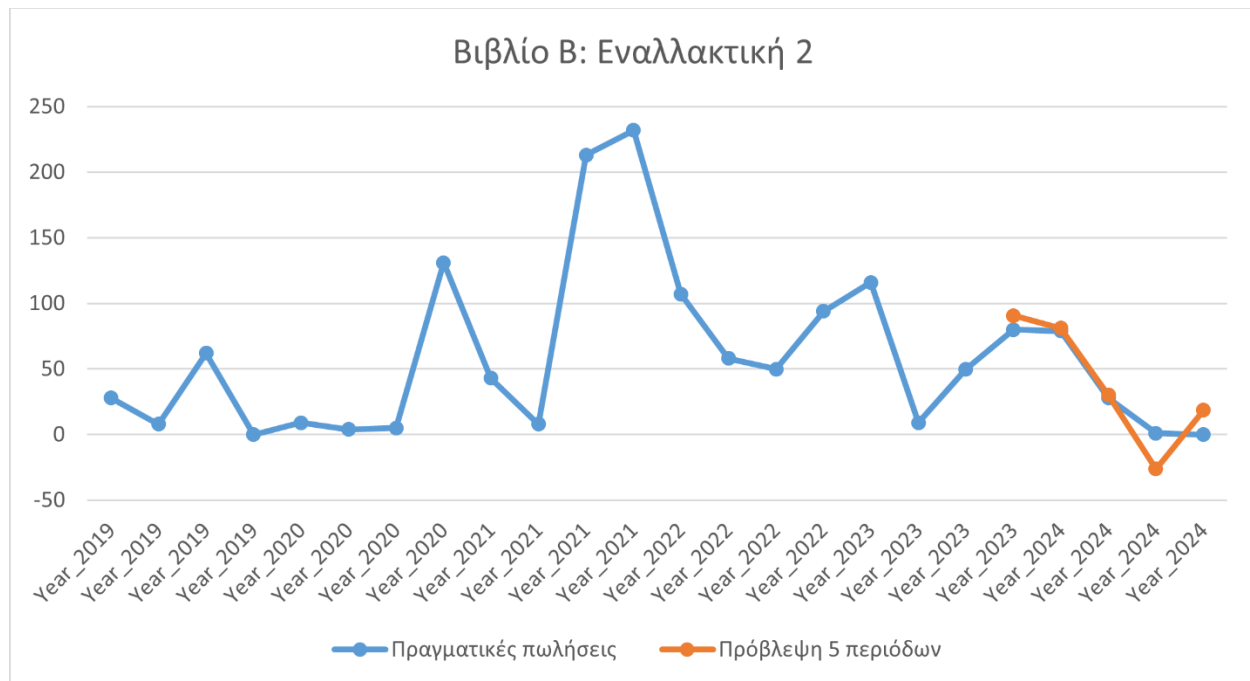
Για την οπτικοποίηση των αποτελεσμάτων και την καλύτερη κατανόηση των αποτελεσμάτων παρακάτω παρουσιάζονται γραφήματα που δείχνουν την χρονοσειρά και την πρόβλεψη για κάθε βιβλίο χωριστά με το καλύτερο RMSE.



Εικόνα 5.1: Πραγματικές πωλήσεις και προβλέψεις 5 περιόδων για το βιβλίο A, $RMSE=68.52$, μέθοδος: ARIMA χωρίς διαφόριση

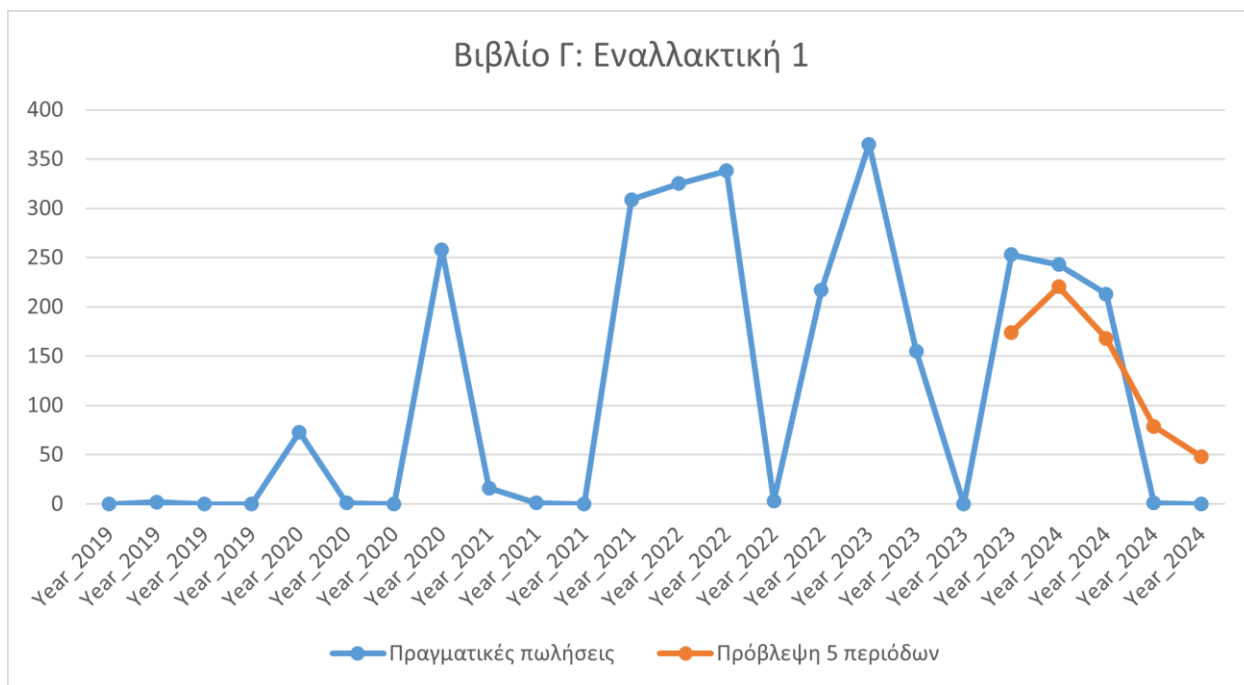
Στο παραπάνω γράφημα, που αφορά στο βιβλίο A και στην εναλλακτική 2, παρατηρείται ότι οι προηγούμενες παρατηρήσεις ακολουθούν ένα μοτίβο με αυξητικές τάσεις κυρίως τους μήνες Ιούλιο έως Δεκέμβριο. Οι προβλέψεις του μοντέλου Holt-Winters για το 2025 προσπαθούν να αντιγράψουν την εποχικότητα αλλά με χαμηλότερη τάση από την πραγματική που παρατηρείται στις προηγούμενες τιμές. Ωστόσο, βλέποντας τις εγγραφές που υπάρχουν για το 2024, αυτές σταματούν τον Σεπτέμβριο του ίδιου έτους. Επομένως, για τους μήνες Οκτώβριο του 2024 έως Ιανουάριο του 2025, δεν υπάρχουν παρατηρήσεις και για τους μήνες Απρίλιο έως Σεπτέμβριο του 2024 οι πωλήσεις είναι 1 για κάθε ομάδα. Για την καλύτερη κατανόηση, το σφάλμα επί των πωλήσεων για το συγκεκριμένο RMSE είναι 18.27%. Το συγκεκριμένο σφάλμα είναι αρκετά καλό, δεδομένου του όγκου των πωλήσεων ανά μήνα για αυτό το βιβλίο.

Παρακάτω ακολουθεί γράφημα για το καλύτερο RMSE που εντοπίστηκε για το βιβλίο B.



Εικόνα 5.2: Πραγματικές πωλήσεις και προβλέψεις 5 περιόδων για το βιβλίο B, $RMSE=15.61$, μέθοδος: Holt-Winters

Στο παραπάνω γράφημα, που αφορά στο βιβλίο B και στην εναλλακτική 2, παρατηρείται ότι οι προηγούμενες παρατηρήσεις ακολουθούν ένα μοτίβο με συχνά αυξητικές τάσεις κυρίως μετά το 2021. Η πρόβλεψη του μοντέλου Holt-Winters για το 2025 προσπαθεί να ακολουθήσει αυτό το μοτίβο. Το σφάλμα επί των πωλήσεων για το συγκεκριμένο RMSE είναι 24.11%, σφάλμα το οποίο είναι αρκετά καλό, δεδομένου του μικρού όγκου παρατηρήσεων.



Εικόνα 5.3: Πραγματικές πωλήσεις και προβλέψεις 5 περιόδων για το βιβλίο Γ, $RMSE=58.47$, μέθοδος: Holt-Winters

Στο παραπάνω γράφημα, που αφορά στο βιβλίο Γ και στην εναλλακτική 1, παρατηρείται ότι οι προηγούμενες παρατηρήσεις ακολουθούν σχετικά αυξητικές τάσεις κυρίως μετά το έτος 2020. Μετά από αυτό το έτος, οι πωλήσεις είναι αυξητικές για όλο το έτος εκτός από Οκτώβριο έως Δεκέμβριο. Οι προβλέψεις του μοντέλου Holt-Winters για το 2025 προσπαθούν να ακολουθήσουν το μοτίβο. Το σφάλμα επί των πωλήσεων είναι 33.48%. Το ποσοστό σφάλματος είναι αρκετά μεγάλο. Πιθανότατα αυτό το μέγεθος σφάλματος να οφείλεται στον περιορισμένο αριθμό δεδομένων που υπάρχουν και σε αρκετές μηδενικές τιμές που εντοπίζονται στις εγγραφές. Οι μηδενικές τιμές στις εγγραφές προέκυψαν από αρνητικές τιμές, οι οποίες κατά την προεπεξεργασία μηδενίστηκαν προκειμένου να μην επηρεαστούν έντονα τα αποτελέσματα.

5.2.2 Απλός κινητός μέσος, προσέγγιση Syntetos-Boylan και απλή εκθετική εξομάλυνση

Ο συγκεκριμένος αλγόριθμος παράγει ως αποτελέσματα τα root mean squared errors για κάθε βιβλίο και για κάθε εναλλακτική και τα συγκρίνει για κάθε μέθοδο πρόβλεψης. Ακολουθείται η λογική του άρθρου των Petropoulos & Kourentzes (2015) για σύγκριση των μεθόδων και εξαγωγή των δεικτών των προβλέψεων όπως αναφέρονται παραπάνω. Σε αυτό που επικεντρώνεται κυρίως ο αλγόριθμος των Syntetos-Boylan Approximation είναι η εκτίμηση και διόρθωση του πραγματικού μέσου όρου, λαμβάνοντας υπόψη τις περιόδους με μηδενική ζήτηση.

Παρακάτω παρουσιάζονται τα αποτελέσματα των δεικτών για τον πρώτο τρόπο ομαδοποίησης, όπως αυτά εξήχθησαν από τον αλγόριθμο πρόβλεψης. Παρουσιάζονται χωριστά τα αποτελέσματα για κάθε βιβλίο. Σε κάθε πίνακα εμφανίζονται οι διαφορετικές εναλλακτικές για κάθε βιβλίο

καθώς και τα γραφήματα με τις τιμές πρόβλεψης της κάθε μεθόδου σε σύγκριση με τις πραγματικές.

Βιβλίο Α

Στο συγκεκριμένο βιβλίο ο αριθμός των παρατηρήσεων για κάθε εναλλακτική κυμαίνεται σε 28. Παρακάτω παρουσιάζονται τα RMSE των μεθόδων: προσέγγιση Syntetos-Boylan (SBA), απλού κινητού μέσου (SMA) και απλής εκθετικής εξομάλυνσης (SES) και τα αντίστοιχα RMSE.

Alternative	RMSE SBA	RMSE SES	RMSE SMA
1	270.86	340.90	184.85
2	246.11	285.08	215.18
3	227.92	258.74	136.08

Πίνακας 5.8: Αποτελέσματα RMSE για το Βιβλίο Α

Για το Βιβλίο Α η καλύτερη εναλλακτική σε αυτόν τον αλγόριθμο είναι η 3. Η μέθοδος του απλού κινητού μέσου εμφανίζει το καλύτερο RMSE, με σφάλμα πωλήσεων 36.29% το οποίο είναι αρκετά μεγάλο, δεδομένων των πωλήσεων ανά ομάδα μηνών, οι οποίες σε κάποιες περιπτώσεις είναι πάνω από 1000. Πιθανότατα η ύπαρξη μόνο μίας μηδενικής τιμής και η μεγάλη διασπορά να οδήγησαν στην αποτυχία των δύο άλλων μεθόδων. Στο παρακάτω γράφημα παρουσιάζονται οι πραγματικές πωλήσεις και στην τελευταία τιμή η πρόβλεψη που δίνει η μέθοδος του κινητού μέσου για μία περίοδο μετά.



Εικόνα 5.4: Βιβλίο Α: Πραγματικές πωλήσεις και πρόβλεψη 1 περιόδου με την μέθοδο του κινητού μέσου

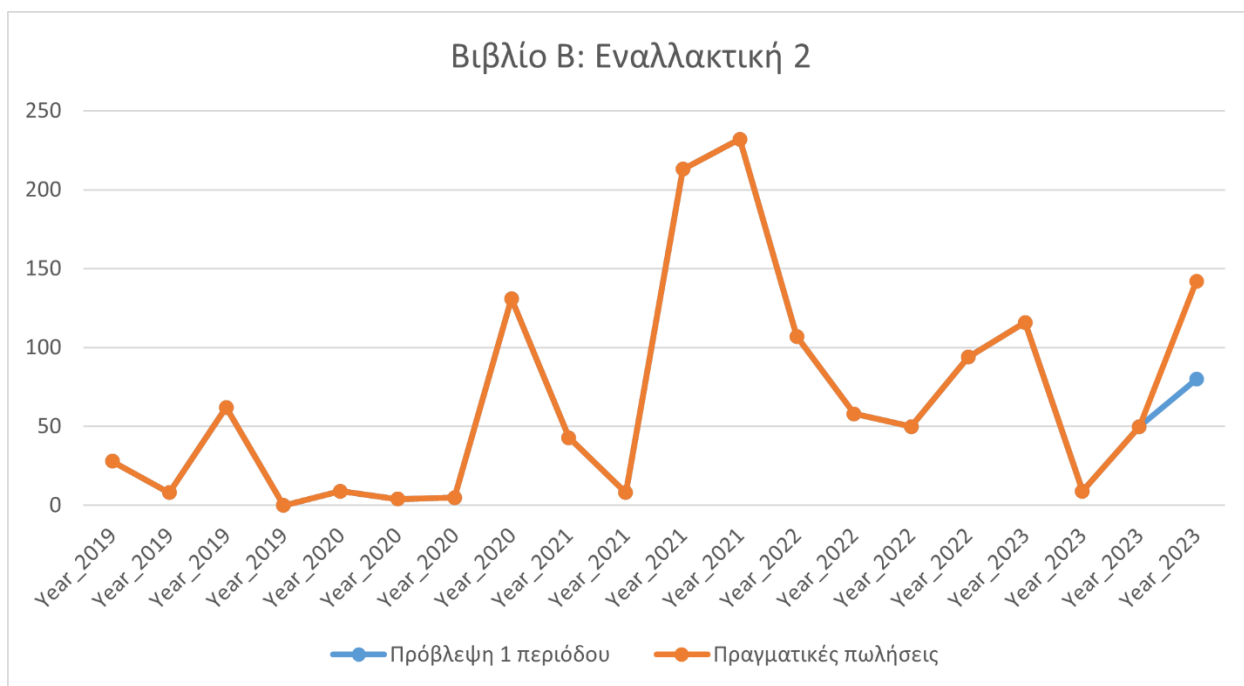
Βιβλίο Β

Για το Βιβλίο Β παρουσιάζονται τα σφάλματα RMSE των μεθόδων: προσέγγιση Syntetos-Boylan (SBA), απλού κινητού μέσου (SMA) και απλής εκθετικής εξομάλυνσης (SES).

Alternative	RMSE SBA	RMSE SES	RMSE SMA
1	44.10	57.45	47.06
2	43.27	49.33	46.37
3	70.59	70.17	71.27

Πίνακας 5.9: Αποτελέσματα RMSE για το Βιβλίο Β

Για το Βιβλίο Β η καλύτερη εναλλακτική σε αυτόν τον αλγόριθμο είναι η 2. Η προσέγγιση των Syntetos-Boylan εμφανίζει το καλύτερο RMSE, με σφάλμα πωλήσεων 66.83% το οποίο είναι αρκετά μεγάλο, δεδομένων των πωλήσεων ανά ομάδα μηνών, οι οποίες σε κάποιες περιπτώσεις είναι πάνω από 100. Στο παρακάτω γράφημα παρουσιάζονται οι πραγματικές πωλήσεις και στην τελευταία τιμή η πρόβλεψη που δίνει η προσέγγιση Syntetos-Boylan για μία περίοδο μετά. Πιθανότατα η έλλειψη πολλών μηδενικών τιμών-εντοπίζονται μόνο 2 στις παρατηρήσεις- να οδήγησε σε αυτό το μεγάλο σφάλμα.



Εικόνα 5.5: Βιβλίο Β: Πραγματικές πωλήσεις και πρόβλεψη 1 περιόδου με την προσέγγιση Syntetos-Boylan

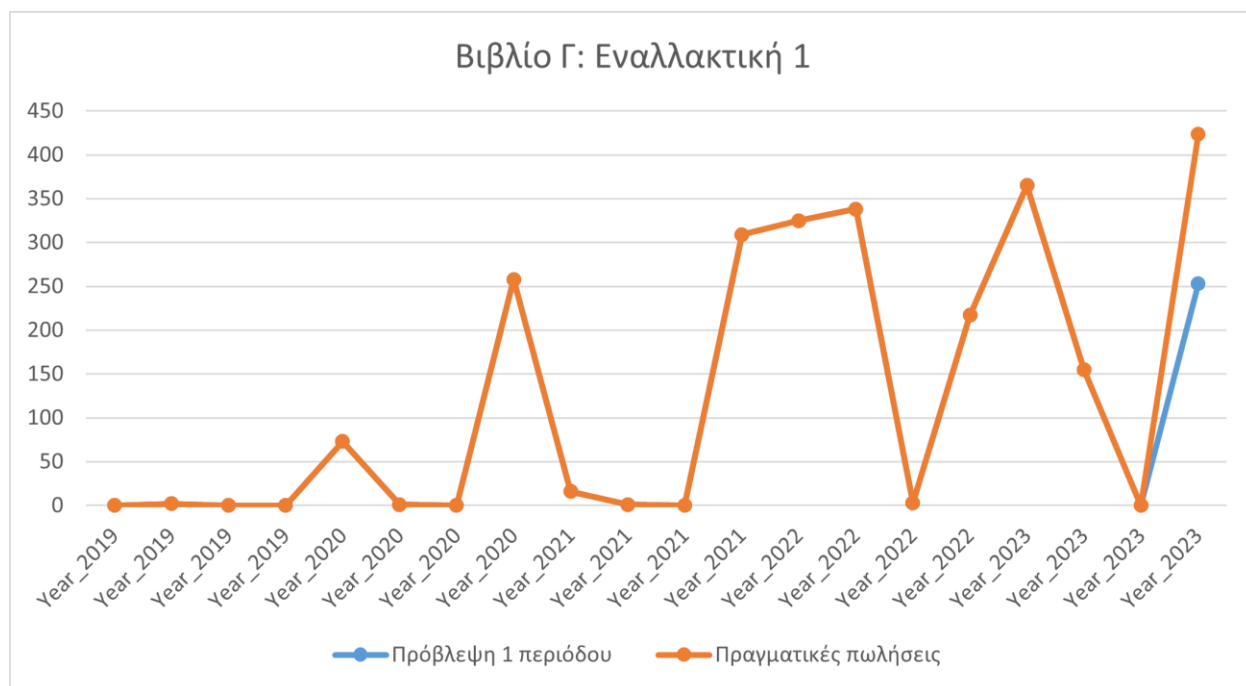
Βιβλίο Γ

Για το Βιβλίο Γ παρουσιάζονται παρακάτω τα σφάλματα RMSE των μεθόδων: προσέγγιση Syntetos-Boylan (SBA), απλού κινητού μέσου (SMA) και απλής εκθετικής εξομάλυνσης (SES).

Alternative	RMSE SBA	RMSE SES	RMSE SMA
1	119.76	127.94	123.72
2	165.95	171.05	180.34
3	187.49	187.32	187.79

Πίνακας 5.10: Αποτελέσματα RMSE για το Βιβλίο Γ

Για το Βιβλίο Γ η καλύτερη εναλλακτική σε αυτόν τον αλγόριθμο είναι η 1. Η προσέγγιση των Syntetos-Boylan εμφανίζει το καλύτερο RMSE, με σφάλμα πωλήσεων 68,58% το οποίο είναι αρκετά μεγάλο, δεδομένων των πωλήσεων ανά ομάδα μηνών, οι οποίες σε αρκετές περιπτώσεις είναι πάνω από 200. Στο παρακάτω γράφημα παρουσιάζονται οι πραγματικές πωλήσεις και στην τελευταία τιμή η πρόβλεψη που δίνει η προσέγγιση Syntetos-Boylan για μία περίοδο μετά.



Εικόνα 5.6: Βιβλίο Γ: Πραγματικές πωλήσεις και πρόβλεψη 1 περιόδου με την προσέγγιση Syntetos-Boylan

Γενικό συμπέρασμα:

Αρχικά λόγω του μικρού όγκου δεδομένων, τα συμπεράσματα είναι επισφαλή. Είναι αναμενόμενο η καλύτερη μέθοδος να παρουσιάζεται ως η μέθοδος του κινητού μέσου στο βιβλίο Α, λόγω του εύρους των δεδομένων που λαμβάνονται υπόψη σε αυτήν. Επιπλέον, πιθανότατα λόγω μερικών μηδενικών τιμών, η Syntetos-Boylan Approximation είναι η πιο έγκυρη μέθοδος για τα βιβλία Β και Γ, η οποία εστιάζει στην αμερόληπτη πρόβλεψη της διακοπτόμενης ζήτησης, ωστόσο τα σφάλματα, παρότι ήταν τα καλύτερα, συνεχίζουν να θεωρούνται μεγάλα. Επίσης, αυτές οι μηδενικές τιμές εντοπίζονταν σποραδικά και όχι πάντα συνεχόμενα, γεγονός το οποίο επηρεάζει την μέθοδο.

5.2.3 Διπλή εκθετική εξομάλυνση για ακανόνιστα δεδομένα

Από τον συγκεκριμένο αλγόριθμο εξήχθησαν δύο ειδών δεδομένα. Αρχικά, τα σφάλματα RMSE και σε δεύτερο χρόνο οι προβλέψεις για 5 περιόδους μετά. Ο λόγος που συνέβη αυτό ήταν για να συγκριθούν τα δεδομένα από αυτόν τον αλγόριθμο με τα δεδομένα που εξήχθησαν από τους προηγούμενους.

Για την συγκεκριμένη ομαδοποίηση, τα αποτελέσματα είναι τα εξής:

Βιβλίο	Εναλλακτική	RMSE	Normalized RMSE
A	Alt_1	143.72	0.38
A	Alt_2	95.88	0.26
A	Alt_3	86.62	0.23
B	Alt_1	39.19	0.61
B	Alt_2	29.42	0.45
B	Alt_3	71.59	1.11
Γ	Alt_1	112.88	0.65
Γ	Alt_2	143.06	0.82
Γ	Alt_3	185.92	1.06

Πίνακας 5.11: Δείκτες πρόβλεψης μεθόδου Διπλής εκθετικής εξομάλυνσης

Συγκρίνοντας το καλύτερο RMSE από τον αλγόριθμο πρόβλεψης με τις μεθόδους ARIMA, μέθοδο κινητού μέσου και Holt-Winters, καταλήγει κάποιος στα εξής συμπεράσματα.

Για το βιβλίο A, με την μέθοδο ARIMA, το καλύτερο Normalized RMSE προκύπτει από την εναλλακτική 2, με τιμή 18.27%. Η αντίστοιχη τιμή για την μέθοδο CIPRA είναι 26%, γεγονός που δείχνει ότι η μέθοδος CIPRA υστερεί έναντι της μεθόδου ARIMA. Όσον αφορά στον δεύτερο αλγόριθμο, η καλύτερη μέθοδος για το βιβλίο A είναι η μέθοδος του κινητού μέσου η οποία δίνει σφάλμα επί των πωλήσεων 36,29%, το οποίο είναι πολύ μεγαλύτερο από τις άλλες δύο μεθόδους.

Οι προβλέψεις του αλγορίθμου ARIMA παρουσιάζονται στον Πίνακας 5.6. Παρακάτω παρουσιάζονται οι προβλέψεις για τον αλγόριθμο CIPRA για το βιβλίο A.

Εναλλακτική	Πρόβλεψη 1 ^{ης} περιόδου	Πρόβλεψη 2 ^{ης} περιόδου	Πρόβλεψη 3 ^{ης} περιόδου	Πρόβλεψη 4 ^{ης} περιόδου	Πρόβλεψη 5 ^{ης} περιόδου
Alt_1	145.32	127.06	108.80	90.54	72.28
Alt_2	160.28	139.74	119.21	98.67	78.13
Alt_3	105.54	93.86	82.18	70.50	58.82

Πίνακας 5.12: Προβλέψεις 5 περιόδων για το βιβλίο A

Για το βιβλίο B, το καλύτερο Normalized RMSE της μεθόδου Holt-Winters, προκύπτει από την εναλλακτική 2, με τιμή 24.11%. Η αντίστοιχη τιμή για την μέθοδο CIPRA είναι 45%, γεγονός που δείχνει ότι η μέθοδος CIPRA υστερεί έναντι της μεθόδου Holt-Winters.

Ο αλγόριθμος πρόβλεψης με την χρήση των μεθόδων απλού κινητού μέσου, Syntetos-Boylan Approximation και Simple exponential smoothing παρουσιάζει καλύτερο σφάλμα επί των πωλήσεων για την εναλλακτική 2, αυτό της προσέγγισης Syntetos-Boylan με τιμή 66.83%. Ο αλγόριθμος Holt-Winters παρουσιάζει καλύτερα αποτελέσματα για αυτό το βιβλίο με την εναλλακτική 2 σε σχέση με όλες τις άλλες μεθόδους.

Παρακάτω παρουσιάζονται οι προβλέψεις για τον αλγόριθμο CIPRA για το βιβλίο B.

Εναλλακτική	Πρόβλεψη 1 ^{ης} περιόδου	Πρόβλεψη 2 ^{ης} περιόδου	Πρόβλεψη 3 ^{ης} περιόδου	Πρόβλεψη 4 ^{ης} περιόδου	Πρόβλεψη 5 ^{ης} περιόδου
Alt_1	49.56	46.15	42.73	39.32	35.91
Alt_2	41.81	36.54	31.27	26.01	20.74
Alt_3	73.92	74.02	74.11	74.21	74.31

Πίνακας 5.13: Προβλέψεις 5 περιόδων για το βιβλίο B

Για το βιβλίο Γ, για τον πρώτο αλγόριθμο πρόβλεψης 5.2.1, το καλύτερο σφάλμα επί των πωλήσεων προκύπτει από την εναλλακτική 1 με την μέθοδο Holt-Winters, με ποσοστό 33.48%. Η αντίστοιχη τιμή για την μέθοδο CIPRA είναι 106%, γεγονός που δείχνει ότι η μέθοδος CIPRA υστερεί σημαντικά έναντι της μεθόδου Holt-Winters.

Ο αλγόριθμος πρόβλεψης με την χρήση των μεθόδων απλού κινητού μέσου, Syntetos-Boylan Approximation και Simple exponential smoothing παρουσιάζει καλύτερο σφάλμα επί των πωλήσεων για την εναλλακτική 1, αυτό της προσέγγισης Syntetos-Boylan με ποσοστό 68.58%. Ο αλγόριθμος CIPRA παρουσιάζει επίσης, χειρότερα αποτελέσματα για αυτό το βιβλίο σε κάθε εναλλακτική.

Εναλλακτική	Πρόβλεψη 1 ^{ης} περιόδου	Πρόβλεψη 2 ^{ης} περιόδου	Πρόβλεψη 3 ^{ης} περιόδου	Πρόβλεψη 4 ^{ης} περιόδου	Πρόβλεψη 5 ^{ης} περιόδου
Alt_1	131.58	127.8	124.01	120.23	116.45
Alt_2	164.53	138.52	112.51	86.49	60.48
Alt_3	172.77	169.82	166.87	163.91	160.96

Πίνακας 5.14: Προβλέψεις 5 περιόδων για το βιβλίο Γ

5.3 Ομαδοποίηση 2

Στο συγκεκριμένο υποκεφάλαιο εμφανίζονται τα αποτελέσματα της **δεύτερης** ομαδοποίησης για κάθε βιβλίο χωριστά, όπως τα εξάγονται από τους αλγόριθμους.

5.3.1 Μέσος όρος, ARIMA και Holt-Winters

Όπως και για το υποκεφάλαιο 5.2.1, ο συγκεκριμένος αλγόριθμος παράγει ως αποτελέσματα τα Root mean square errors για κάθε βιβλίο, για κάθε εναλλακτική και για κάθε μέθοδο πρόβλεψης και διαφόρισης, καθώς και τον αριθμό των προβλέψεων που έχουν οριστεί από τον χρήστη. Στην συγκεκριμένη περίπτωση ο αριθμός των προβλέψεων ορίστηκε ως τρεις.

Παρακάτω παρουσιάζονται τα αποτελέσματα των RMSE για τον δεύτερο τρόπο ομαδοποίησης, όπως αυτά εξήχθησαν από τον αλγόριθμο πρόβλεψης με την χρήση των μεθόδων μέσου όρου, ARIMA και Holt-Winters. Παρουσιάζονται χωριστά τα αποτελέσματα για κάθε βιβλίο. Σε κάθε πίνακα εμφανίζονται οι διαφορετικές εναλλακτικές για κάθε μέθοδο διαφόρισης.

Βιβλίο Α

Εναλλακτική	Μέθοδος διαφόρισης	WhiteNoise_RMSE	Best_ARIMA_RMSE	Holt_Winters_RMSE
1	none	422.94	170.97	162
1	first_order	156.22	156.22	--
1	seasonal	275.29	275.29	--
1	both	946.07	365.36	--
2	none	466.36	466.36	71.12
2	first_order	161.82	97.45	--
2	seasonal	205.44	205.44	--
2	both	2341.8	1506.81	--
3	none	457.91	447.29	42.19
3	first_order	440.98	216.53	--
3	seasonal	187.26	130.61	--
3	both	3409.52	2535.32	--

Πίνακας 5.15: Αποτελέσματα RMSE για το βιβλίο Α

Βιβλίο Β

Εναλλακτική	Μέθοδος διαφόρισης	WhiteNoise_RMSE	Best_ARIMA_RMSE	Holt_Winters_RMSE
1	none	55.81	55.81	35.79
1	first_order	195.26	143.24	--
1	seasonal	150.19	74.02	--
1	both	676.2	237.56	--
2	none	59.13	56.36	21.07
2	first_order	47.77	47.77	--
2	seasonal	67.91	67.91	--
2	both	153.62	153.62	--
3	none	71.37	71.37	46.75
3	first_order	116.88	110.38	--
3	seasonal	125.82	125.82	--
3	both	998.16	397.07	--

Πίνακας 5.16: Αποτελέσματα RMSE για το βιβλίο Β

Βιβλίο Γ

Εναλλακτική	Μέθοδος διαφορίσης	WhiteNoise_RMSE	Best_ARIMA_RMSE	Holt_Winters_RMSE
1	none	147.62	147.62	127.11
1	first_order	413.67	272.98	--
1	seasonal	325.45	325.45	--
1	both	939.52	939.52	--
2	none	135.55	128.06	57.62
2	first_order	144.03	126.06	--
2	seasonal	272.51	203.18	--
2	both	1682.66	1682.66	--
3	none	191.01	152.25	131.77
3	first_order	228.88	197.03	--
3	seasonal	318.54	318.54	--
3	both	2370.82	620.76	--

Πίνακας 5.17: Αποτελέσματα RMSE για το βιβλίο Γ

Παρατηρείται ότι για το βιβλίο Α, το μικρότερο RMSE εντοπίζεται για την εναλλακτική επιλογή 3 με την μέθοδο Holt-Winters με τιμή 42.19. Το ίδιο ισχύει για την εναλλακτική επιλογή 3, δηλαδή το καλύτερο RMSE εμφανίζεται με την μέθοδο Holt-Winters. Για την εναλλακτική επιλογή 1, το καλύτερο RMSE φαίνεται να είναι με την πρώτης τάξης διαφορίση και, μάλιστα, οι μέθοδοι του μέσου όρου και ARIMA εμφανίζουν το ίδιο RMSE. Αυτό, όπως αναφέρθηκε και παραπάνω, προκύπτει από το μοντέλο ARIMA $(p, d, q) = (0, 0, 0)$. Η συγκεκριμένη περίπτωση δείχνει ότι η καλύτερη μέθοδος τελικά, είναι αυτή του μέσου όρου.

Για το βιβλίο Β, το καλύτερο RMSE εμφανίζεται στην δεύτερη εναλλακτική με την μέθοδο Holt-Winters με τιμή 21.07.

Όσον αφορά στο τρίτο βιβλίο Γ, το καλύτερο RMSE ανήκει στην δεύτερη εναλλακτική, με την μέθοδο Holt-Winters με τιμή 57.62.

Πρέπει και εδώ να σημειωθεί ότι παρατηρείται συχνά στα RMSE σε διαφορετικές περιπτώσεις, τα RMSE της μεθόδου ARIMA να είναι αρκετά κοντά, και σε μερικές περιπτώσεις να ταυτίζονται, με αυτά της μεθόδου του Μέσου όρου. Αυτό συμβαίνει καθώς το καλύτερο RMSE της μεθόδου ARIMA, προέκυψε από το μοντέλο $(p, d, q) = (0, 0, 0)$. Όταν, το παραπάνω μοντέλο εφαρμόζεται, η μέθοδος ARIMA ακολουθεί την λογική του μέσου όρου. Επομένως, για τον παραπάνω λόγο εμφανίζονται ίδια αποτελέσματα και κατά πάσα πιθανότητα η χρονοσειρά που δημιουργήθηκε είναι λευκός θόρυβος, δηλαδή η μέθοδος ARIMA δεν μπορεί να εντοπίσει κάποιο μοτίβο. Επιπλέον αυτό που παρατηρείται είναι ότι σε αρκετές από τις παραπάνω εναλλακτικές, αν όχι σε όλες, η μέθοδος που συνδυάζει τις δύο διαφορίσεις (both) εμφανίζει το μεγαλύτερο RMSE, γεγονός που δείχνει ότι πιθανότατα έχει πραγματοποιηθεί υπερβολική διαφορίση, η οποία είναι περιττή και δεν οδηγεί σε επιθυμητά αποτελέσματα.

5.3.1.1 Προβλέψεις για χρονικό ορίζοντα 5 περιόδων

Παρακάτω παρουσιάζονται τα αποτελέσματα των προβλεπόμενων πωλήσεων των βιβλίων σε τεμάχια για 1 έως 5 περιόδους πρόβλεψης για κάθε μέθοδο πρόβλεψης χωριστά.

Για την μέθοδο ARIMA:

Βιβλίο	Εναλλακτική	Μέθοδος διαφορίσης	Πρόβλεψη 1 ^{ης} περιόδου	Πρόβλεψη 2 ^{ης} περιόδου	Πρόβλεψη 3 ^{ης} περιόδου	Πρόβλεψη 4 ^{ης} περιόδου	Πρόβλεψη 5 ^{ης} περιόδου
A	1	none	-148.77	255.44	-67.31	190.4	-15.37
A	1	first_order	188.67	171.33	154	136.67	119.33
A	1	seasonal	130.17	-106.83	460.17	36.17	-39.67
A	1	both	548.14	-20.23	470.32	177.58	283.41
A	2	none	568.81	568.81	568.81	568.81	568.81
A	2	first_order	286.54	95.17	102.44	119.99	38.89
A	2	seasonal	71.17	526.17	104.17	20.17	3.33
A	2	both	1411.57	353.52	832.41	1227.29	2857.75
A	3	none	595.99	514.45	506.9	517.16	516.6
A	3	first_order	290.37	192.62	366.43	294.39	258.26
A	3	seasonal	47.7	105.9	-131.1	235.9	-146.4
A	3	both	432.6	3793.87	927.12	2112.52	3672.53
B	1	none	84.69	84.69	84.69	84.69	84.69
B	1	first_order	176.41	123.8	185.58	212.99	186.8
B	1	seasonal	-5.03	232.11	38.4	126.1	57.87
B	1	both	21.56	332.48	181.07	473.3	180.22
B	2	none	74.82	85.4	87.65	88.12	88.22
B	2	first_order	24.17	23.33	22.5	21.67	20.83
B	2	seasonal	57.67	-25.33	114.67	-52.33	-19.67
B	2	both	103.5	99	234.5	203	225
B	3	none	79.08	79.08	79.08	79.08	79.08
B	3	first_order	124.47	134.37	143.95	153.55	163.14
B	3	seasonal	80	225	120	148	106
B	3	both	-109.35	195.61	395.57	623.53	498.14
Γ	1	none	151.69	151.69	151.69	151.69	151.69
Γ	1	first_order	322.75	288.2	447.83	452.2	397.21
Γ	1	seasonal	481.89	91.89	170.89	382.89	435.78
Γ	1	both	817.63	816.25	893.88	1183.5	1432.75
Γ	2	none	127.1	221.49	222.84	187.43	190.33
Γ	2	first_order	194.57	204.36	189.82	194.79	184.11
Γ	2	seasonal	429.45	-36.17	185.83	-61.17	213.29
Γ	2	both	685.13	1054.25	1645.38	1989.5	2708.75
Γ	3	none	258.2	78.11	53.62	140.67	179.08
Γ	3	first_order	220.03	231.13	247.02	263.44	279.91
Γ	3	seasonal	323	634	244	359	429
Γ	3	both	-506.16	240.93	598.02	1070.12	530.05

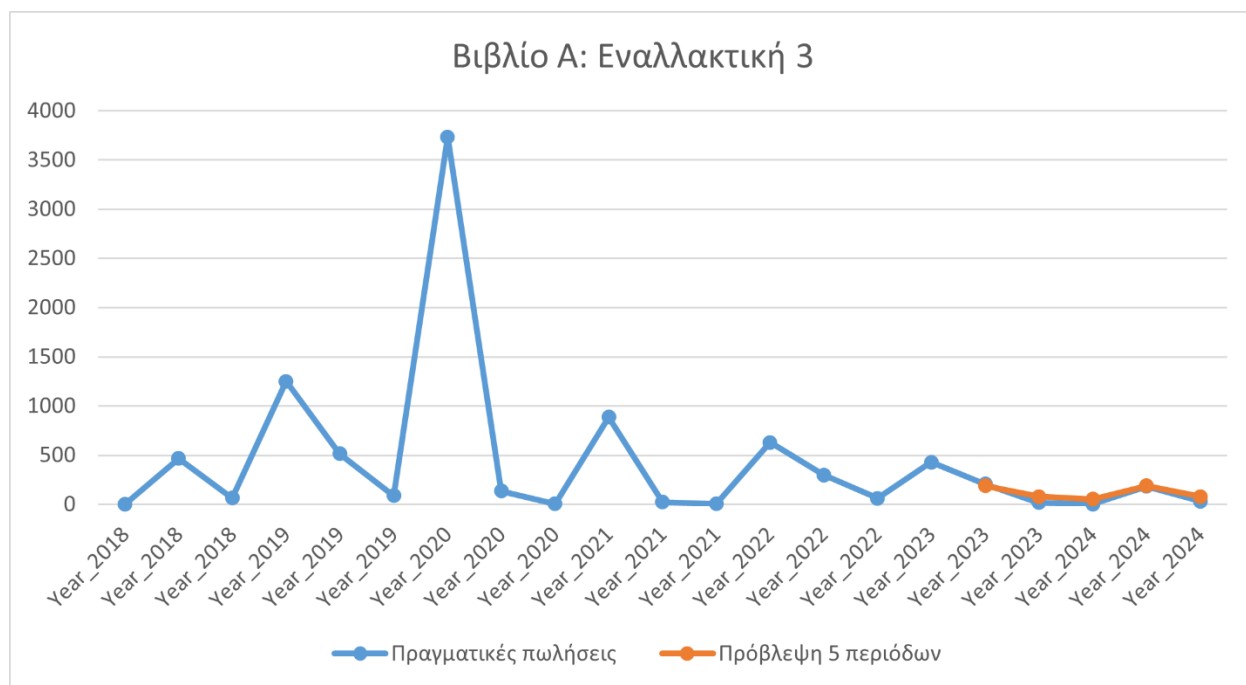
Πίνακας 5.18: Προβλέψεις μεθόδου ARIMA

Για την μέθοδο Holt-Winters:

Βιβλίο	Εναλλακτική	Πρόβλεψη 1 ^{ης} περιόδου	Πρόβλεψη 2 ^{ης} περιόδου	Πρόβλεψη 3 ^{ης} περιόδου	Πρόβλεψη 4 ^{ης} περιόδου	Πρόβλεψη 5 ^{ης} περιόδου
A	1	48.99	496.31	277.62	-239.57	207.75
A	2	287.91	168.69	-26.35	69.14	-50.08
A	3	190.37	80.41	52.07	187.82	77.86
B	1	68.24	83.34	61.64	31.02	46.13
B	2	93.05	99.67	83.54	11.93	18.55
B	3	133.74	86.76	51.95	74.28	27.29
Γ	1	195.17	205.64	167.15	94.36	104.83
Γ	2	209.53	182.1	202.47	61.56	34.12
Γ	3	341.89	239.8	136.63	214.62	112.53

Πίνακας 5.19: Προβλέψεις μεθόδου Holt-Winters

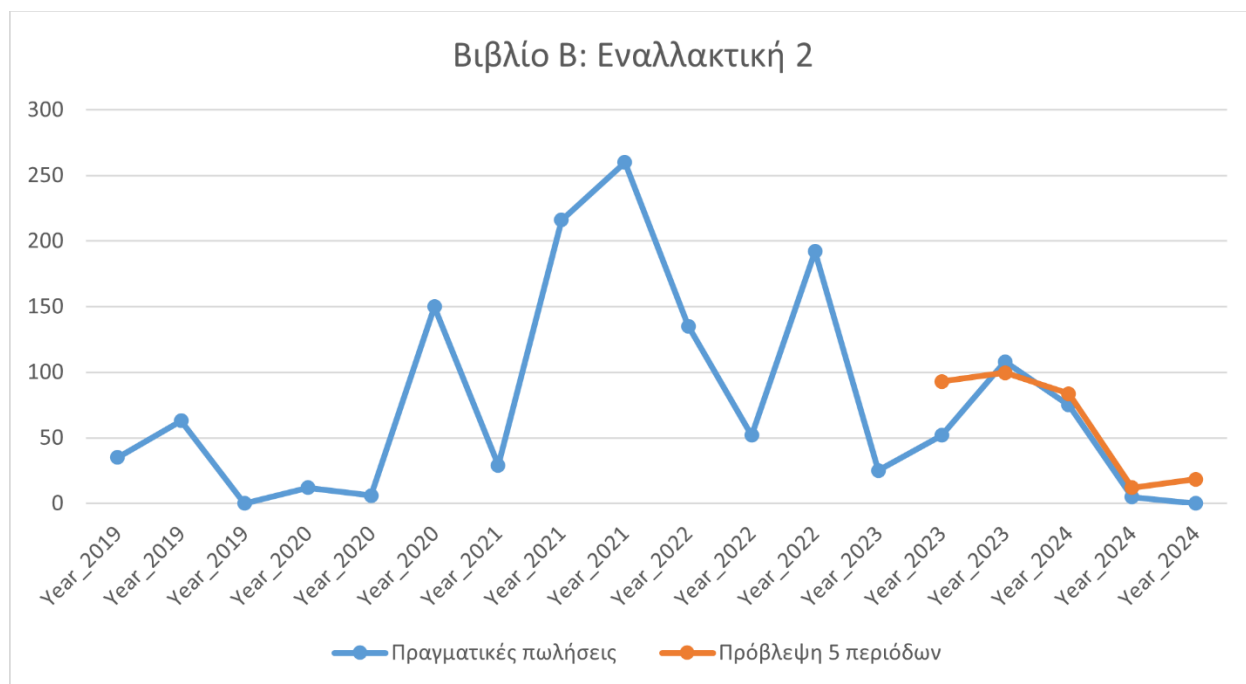
Για την καλύτερη κατανόηση των αποτελεσμάτων παρακάτω οπτικοποιούνται τα γραφήματα που δείχνουν την χρονοσειρά και την πρόβλεψη για κάθε βιβλίο χωριστά με το καλύτερο RMSE.



Εικόνα 5.7: Πραγματικές πωλήσεις και προβλέψεις 5 περιόδων για το βιβλίο A, RMSE=42.19, μέθοδος: Holt-Winters

Στο παραπάνω γράφημα, που αφορά στο βιβλίο A και στην εναλλακτική 3, παρατηρείται ότι οι προηγούμενες παρατηρήσεις ακολουθούν ένα μοτίβο με αυξητικές τάσεις κυρίως τους μήνες Αύγουστο έως και Νοέμβριο και Δεκέμβριο έως και Μάρτιο. Το σφάλμα επί των πωλήσεων για αυτό το βιβλίο είναι 8.44%, το οποίο είναι πολύ ικανοποιητικό σφάλμα, δεδομένου ότι εντοπίζονται και ακραίες τιμές στα δεδομένα. Επιπλέον, από τις προβλέψεις φαίνεται ότι η μέθοδος Holt-Winters προσπαθεί να εντοπίσει ένα μοτίβο.

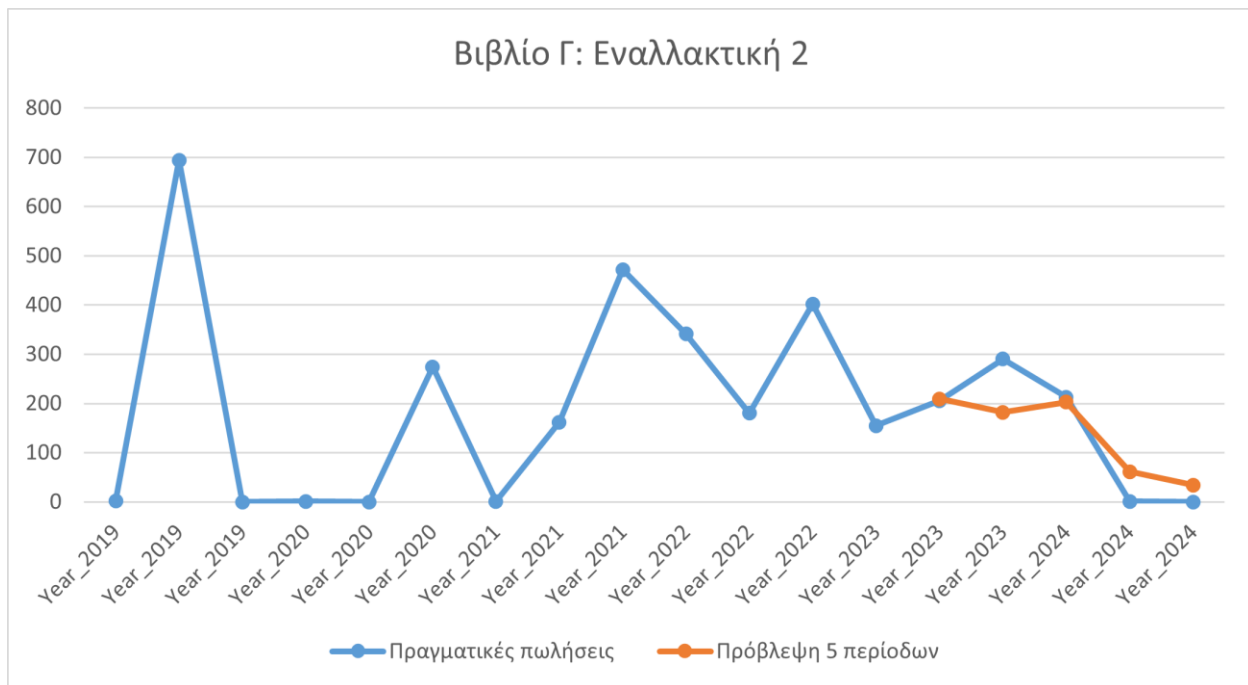
Παρακάτω ακολουθεί γράφημα για το καλύτερο RMSE που εντοπίστηκε για το βιβλίο Β.



Εικόνα 5.8: Πραγματικές πωλήσεις και προβλέψεις 5 περιόδων για το βιβλίο Β, $RMSE=21.07$, μέθοδος: Holt-Winters

Στο παραπάνω γράφημα, που αφορά στο βιβλίο Β και στην εναλλακτική 2, παρατηρείται ότι οι προηγούμενες παρατηρήσεις ακολουθούν ένα μοτίβο με αυξητικές τάσεις κυρίως τους μήνες Αύγουστο έως και Μάρτιο. Οι προβλέψεις του μοντέλου Holt-Winters για το 2025 ακολουθούν μία αυξητική τάση και ένα μοτίβο παρόμοιο με τις προηγούμενες πραγματικές τιμές. Το σφάλμα επί των πωλήσεων για την συγκεκριμένη εναλλακτική είναι 24.41%, το οποίο είναι ικανοποιητικό.

Παρακάτω ακολουθεί γράφημα για το καλύτερο RMSE που εντοπίστηκε για το βιβλίο Γ.



Εικόνα 5.9: Πραγματικές πωλήσεις και προβλέψεις 5 περιόδων για το βιβλίο Γ, $RMSE=57.62$, μέθοδος: Holt-Winters

Στο παραπάνω γράφημα, που αφορά στο βιβλίο Γ και στην εναλλακτική 2, παρατηρείται ότι οι προηγούμενες παρατηρήσεις ακολουθούν ένα μοτίβο με αυξητικές τάσεις και έπειτα πτώσεις. Αυτό πραγματοποιείται συνεχώς σε διάφορες περιόδους. Οι προβλέψεις του μοντέλου Holt-Winters για το 2025 ακολουθούν μία μικρή αυξητική τάση η οποία καταλήγει σε πτώση, το οποίο μοιάζει αναμενόμενο. Το σφάλμα επί των πωλήσεων είναι σε αυτή την περίπτωση 24.75%.

5.3.2 Απλός κινητός μέσος, προσέγγιση Syntetos-Boylan και απλή εκθετική εξομάλυνση

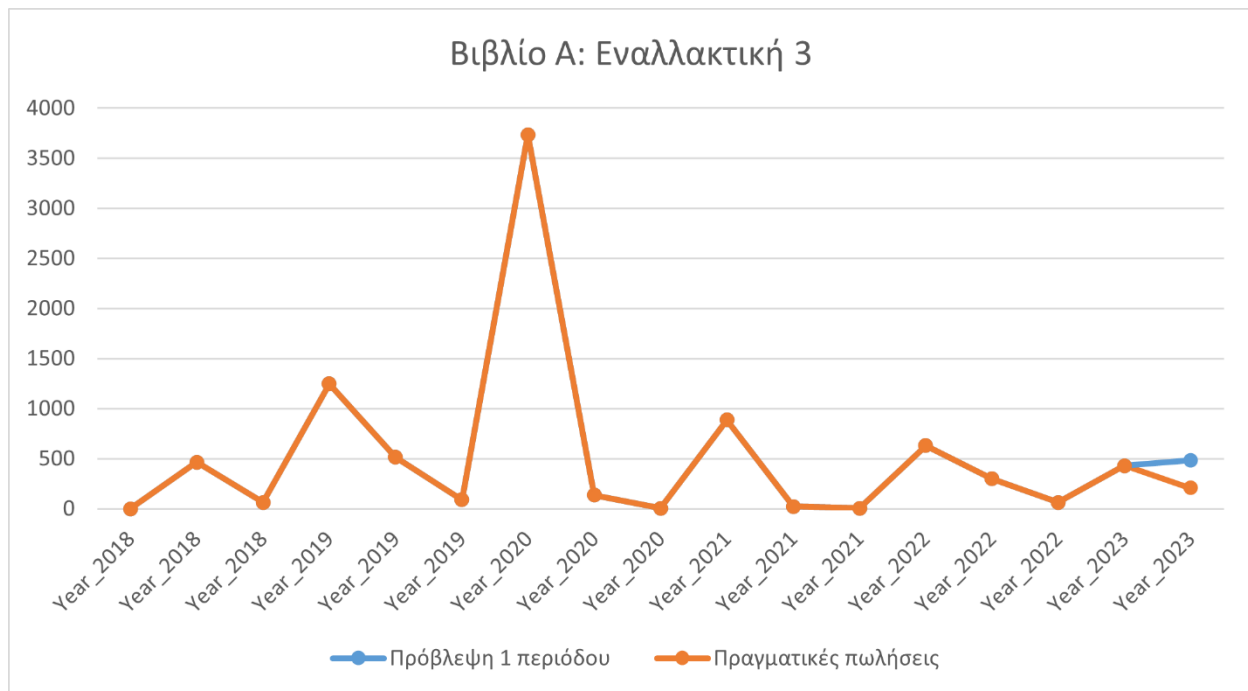
Όπως αναφέρθηκε και στο υποκεφάλαιο 5.2.2, ο συγκεκριμένος αλγόριθμος παράγει ως αποτελέσματα τα root mean squared errors για κάθε βιβλίο και για κάθε εναλλακτική και τα συγκρίνει για κάθε μέθοδο πρόβλεψης.

Βιβλίο Α

Στο συγκεκριμένο βιβλίο ο αριθμός των παρατηρήσεων για κάθε εναλλακτική κυμαίνεται σε 21. Παρακάτω παρουσιάζονται τα RMSE των μεθόδων: προσέγγιση Syntetos-Boylan (SBA), απλού κινητού μέσου (SMA) και απλής εκθετικής εξομάλυνσης (SES) και τα αντίστοιχα RMSE.

Alternative	RMSE SBA	RMSE SES	RMSE SMA
1	370.44	370.44	493.83
2	327.59	406.56	484.19
3	280.14	314.72	207.22

Πίνακας 5.20: RMSE για το βιβλίο Α



Εικόνα 5.10: Πραγματικές πωλήσεις και πρόβλεψη 1 περιόδου για την εναλλακτική 3, Βιβλίο A

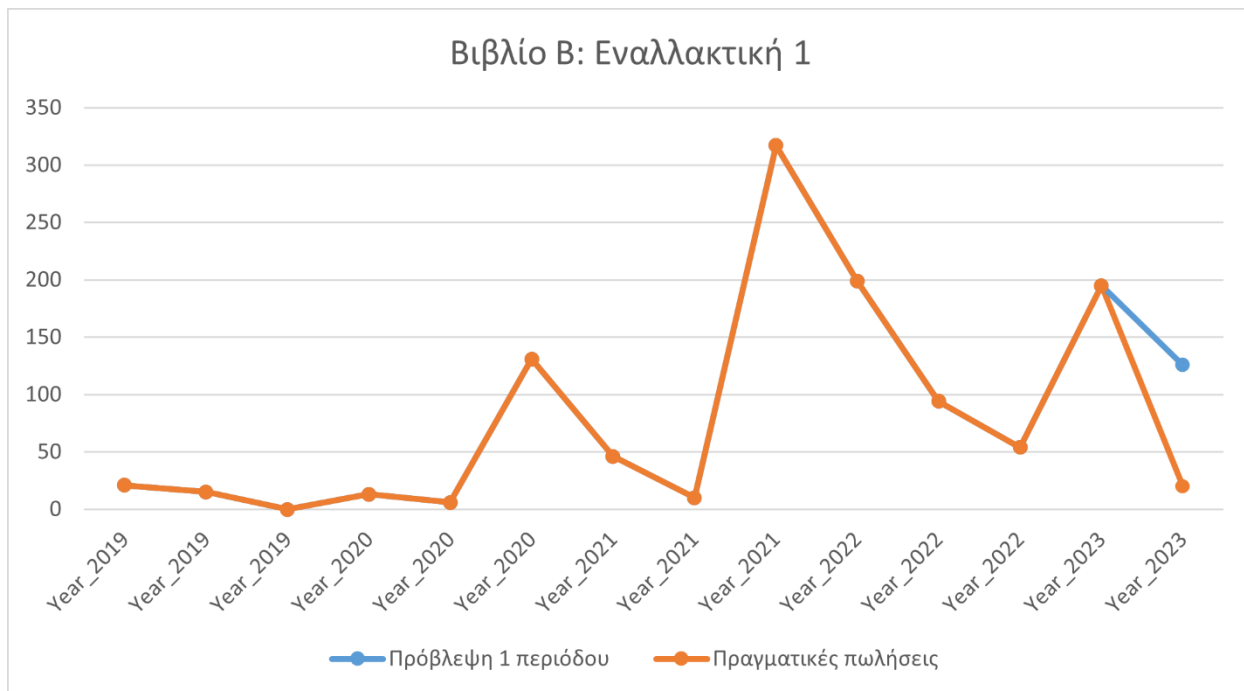
Σε αυτό το βιβλίο και στην εναλλακτική 3, η καλύτερη μέθοδος φαίνεται να είναι η μέθοδος του κινητού μέσου. Φαίνεται και από το γράφημα στην Εικόνα 5.10, ότι ακολουθεί τις πραγματικές τιμές. Το σφάλμα επί των πωλήσεων στην συγκεκριμένη περίπτωση είναι 41.44%. Ένα υψηλό ποσοστό δεδομένων των τιμών των παρατηρήσεων που κάποιες ξεπερνούν τις 1000 πωλήσεις. Η πρόσεγγιση των Syntetos-Boylan πιθανά απέτυχε λόγω της έλλειψης μηδενικών παρατηρήσεων στην συγκεκριμένη χρονοσειρά και λόγω του μικρού όγκου δεδομένων καθώς μετά τον διαχωρισμό σε training και testing sets, τα δεδομένα με βάση τα οποία επιλέχτηκε η καλύτερη μέθοδος ήταν 16.

Βιβλίο B

Στο συγκεκριμένο βιβλίο ο αριθμός των παρατηρήσεων για κάθε εναλλακτική κυμαίνεται σε 18. Παρακάτω παρουσιάζονται τα RMSE των μεθόδων: προσέγγιση Syntetos-Boylan (SBA), απλού κινητού μέσου (SMA) και απλής εκθετικής εξομάλυνσης (SES) και τα αντίστοιχα RMSE.

Alternative	RMSE SBA	RMSE SES	RMSE SMA
1	92.13	70.25	70.95
2	71.23	103.28	81.53
3	72.27	76.44	88.62

Εικόνα 5.11: RMSE για το Βιβλίο B



Εικόνα 5.12: Πραγματικές πωλήσεις και πρόβλεψη 1 περιόδου για το Βιβλίο Β

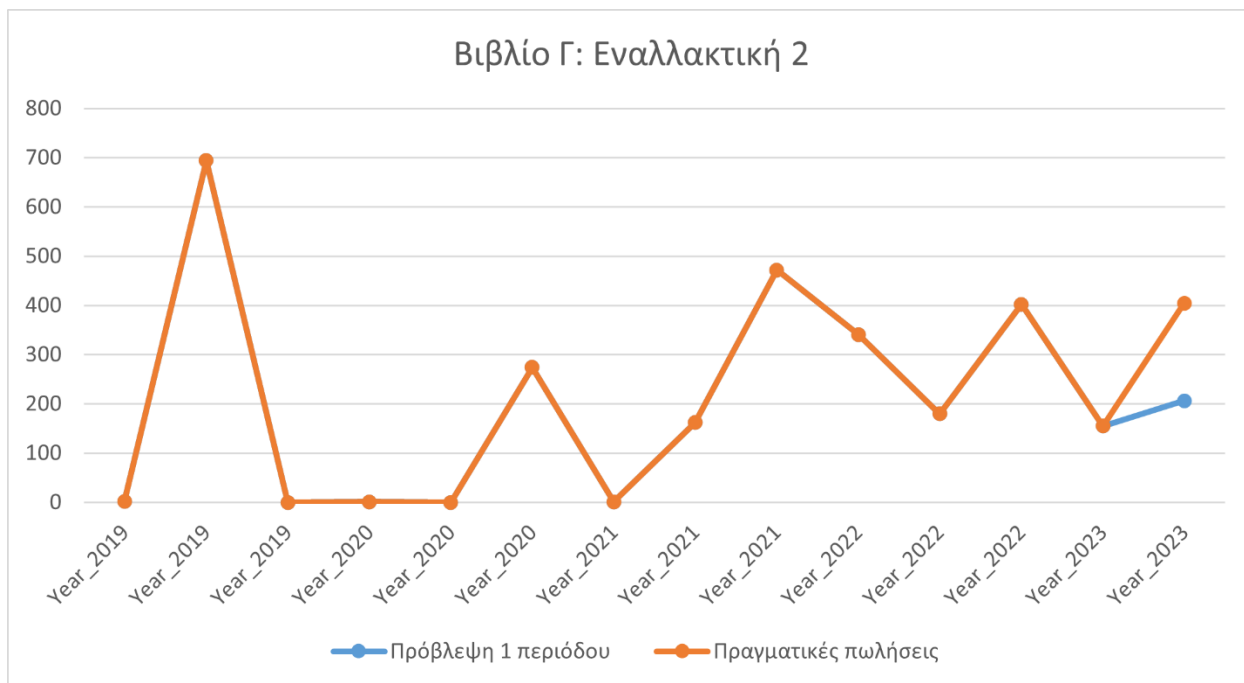
Για την συγκεκριμένη εναλλακτική, η καλύτερη μέθοδος είναι η απλή εκθετική εξομάλυνση. Το σφάλμα επί των πωλήσεων για την συγκεκριμένη περίπτωση είναι 81,38%. Το συγκεκριμένο σφάλμα είναι αρκετά μεγάλο, ωστόσο ο αριθμός των δεδομένων είναι αρκετά μικρός, δεδομένου ότι η εύρεση της καλύτερης μεθόδου στηρίχτηκε σε 13 δεδομένα λόγω του διαχωρισμού των δεδομένων σε training και testing sets.

Βιβλίο Γ

Στο συγκεκριμένο βιβλίο ο αριθμός των παρατηρήσεων για κάθε εναλλακτική κυμαίνεται σε 18. Παρακάτω παρουσιάζονται τα RMSE των μεθόδων: προσέγγιση Syntetos-Boylan (SBA), απλού κινητού μέσου (SMA) και απλής εκθετικής εξομάλυνσης (SES) και τα αντίστοιχα RMSE.

Alternative	RMSE SBA	RMSE SES	RMSE SMA
1	177.56	160.89	151.90
2	147.93	163.56	132.07
3	182.76	182.47	202.63

Πίνακας 5.21: RMSE για το Βιβλίο Γ



Εικόνα 5.13: Πραγματικές πωλήσεις και πρόβλεψη 1 περιόδου για το Βιβλίο Γ

Για την συγκεκριμένη εναλλακτική, η καλύτερη μέθοδος είναι αυτή του κινητού μέσου. Οι δύο άλλες μέθοδοι, δείχνουν σε σύγκριση με την μέθοδο κινητού μέσου ότι δεν παράγουν ικανοποιητικά αποτελέσματα, ωστόσο η διαφορά των RMSE δεν είναι μεγάλη μεταξύ των μεθόδων. Η συνολική εικόνα της εναλλακτικής δείχνει ότι το σφάλμα επί των πωλήσεων είναι 56.72%. Πιθανότατα αυτό οφείλεται στο γεγονός ότι στη χρονοσειρά που εξετάζεται για την εύρεση του καλύτερου μοντέλου χρησιμοποιήθηκαν μόνο 13 τιμές.

5.3.3 Διπλή εκθετική εξομάλυνση για ακανόνιστα δεδομένα

Από τον συγκεκριμένο αλγόριθμο εξήχθησαν δύο ειδών δεδομένα. Αρχικά, τα σφάλματα RMSE και σε δεύτερο χρόνο οι προβλέψεις για 5 περιόδους μετά. Ο λόγος που συνέβη αυτό ήταν για να συγκριθούν τα δεδομένα από αυτόν τον αλγόριθμο με τα δεδομένα που εξήχθησαν από τους προηγούμενους. Για την συγκεκριμένη ομαδοποίηση, τα αποτελέσματα είναι τα εξής:

Βιβλίο	Εναλλακτική	RMSE	Normalized RMSE
A	Alt_1	153.72	0.31
A	Alt_2	106.38	0.21
A	Alt_3	88.94	0.18
B	Alt_1	57.26	0.66
B	Alt_2	34.02	0.39
B	Alt_3	61.83	0.72
Γ	Alt_1	174.23	0.75
Γ	Alt_2	103.32	0.44
Γ	Alt_3	175.5	0.75

Πίνακας 5.22: Δείκτες πρόβλεψης μεθόδου Διπλής εκθετικής εξομάλυνσης

Συγκρίνοντας το καλύτερο RMSE από τον αλγόριθμο πρόβλεψης με τις μεθόδους ARIMA, μέθοδο κινητού μέσου και Holt-Winters, καταλήγει κάποιος στα εξής συμπεράσματα:

Για το βιβλίο A, με την μέθοδο Holt-Winters, το καλύτερο Normalized RMSE προκύπτει από την εναλλακτική 3, με τιμή 8.44%. Η αντίστοιχη τιμή για την μέθοδο CIPRA είναι 18%, γεγονός που δείχνει ότι η μέθοδος CIPRA υστερεί έναντι της μεθόδου Holt-Winters. Όσον αφορά στον δεύτερο αλγόριθμο 5.3.2, η καλύτερη μέθοδος για το βιβλίο A είναι η μέθοδος του κινητού μέσου η οποία δίνει σφάλμα επί των πωλήσεων 41.44%, το οποίο είναι πολύ μεγαλύτερο από τις άλλες δύο μεθόδους.

Οι προβλέψεις του αλγορίθμου Holt-Winters παρουσιάζονται στον Πίνακα 5.15. Παρακάτω παρουσιάζονται οι προβλέψεις για τον αλγόριθμο CIPRA για το βιβλίο A.

Εναλλακτική	Πρόβλεψη 1 ^{ης} περιόδου	Πρόβλεψη 2 ^{ης} περιόδου	Πρόβλεψη 3 ^{ης} περιόδου	Πρόβλεψη 4 ^{ης} περιόδου	Πρόβλεψη 5 ^{ης} περιόδου
Alt_1	228.24	194.72	161.21	127.69	94.18
Alt_2	200.92	159.27	117.62	75.97	34.32
Alt_3	168.58	131.11	93.64	56.17	18.7

Πίνακας 5.23: Προβλέψεις 5 περιόδων για το βιβλίο A

Για το βιβλίο B, με τον αλγόριθμο πρόβλεψης μέσου όρου, ARIMA και Holt-Winters, το καλύτερο Normalized RMSE προκύπτει από την εναλλακτική 2 με την μέθοδο Holt-Winters και με ποσοστό σφάλματος 24.41%. Η αντίστοιχη τιμή για την μέθοδο CIPRA είναι 39%, γεγονός που δείχνει ότι η μέθοδος CIPRA υστερεί με μικρή διαφορά έναντι της μεθόδου Holt-Winters.

Αντίστοιχα, ο αλγόριθμος πρόβλεψης με την χρήση των μεθόδων απλού κινητού μέσου, Syntetos-Boylan Approximation και Simple exponential smoothing παρουσιάζει καλύτερο ποσοστό σφάλματος επί των πωλήσεων με ποσοστό 81.38% για την εναλλακτική 1, με την προσέγγιση Syntetos-Boylan, ενώ η αντίστοιχη τιμή της εναλλακτικής 1 για τον αλγόριθμο του CIPRA είναι 66%. Σε κάθε περίπτωση, η τιμή δεν είναι η ιδανική, ωστόσο η προσέγγιση Syntetos-Boylan υστερεί έναντι της μεθόδου CIPRA.

Οι προβλέψεις του αλγορίθμου Holt-Winters παρουσιάζονται στον Πίνακα 5.13. Παρακάτω παρουσιάζονται οι προβλέψεις για τον αλγόριθμο CIPRA για το βιβλίο B.

Εναλλακτική	Πρόβλεψη 1 ^{ης} περιόδου	Πρόβλεψη 2 ^{ης} περιόδου	Πρόβλεψη 3 ^{ης} περιόδου	Πρόβλεψη 4 ^{ης} περιόδου	Πρόβλεψη 5 ^{ης} περιόδου
Alt_1	130.62	89.45	48.28	7.12	-34.05
Alt_2	61.45	53.08	44.71	36.34	27.97
Alt_3	94.78	80.54	66.29	52.04	37.79

Πίνακας 5.24: Προβλέψεις 5 περιόδων για το βιβλίο B

Για το βιβλίο Γ, με τον αλγόριθμο πρόβλεψης μέσου όρου, ARIMA και Holt-Winters, το καλύτερο Normalized RMSE προκύπτει από την εναλλακτική 2, με ποσοστό 24.75% με την μέθοδο Holt-Winters. Η αντίστοιχη τιμή για την μέθοδο CIPRA είναι 44%, γεγονός που δείχνει ότι η μέθοδος CIPRA υστερεί έναντι της μεθόδου Holt-Winters.

Ο αλγόριθμος πρόβλεψης με την χρήση των μεθόδων απλού κινητού μέσου, Syntetos-Boylan Approximation και Simple exponential smoothing παρουσιάζει καλύτερο Normalized RMSE για την εναλλακτική 2, με την μέθοδο κινητού μέσου με τιμή 56.72%. Στην συγκεκριμένη περίπτωση, ο αλγόριθμος CIPRA υπερτερεί έναντι της μεθόδου του κινητού μέσου όρου.

Παρακάτω παρουσιάζονται οι προβλέψεις για τον αλγόριθμο CIPRA για το βιβλίο Γ.

Εναλλακτική	Πρόβλεψη 1 ^{ης} περιόδου	Πρόβλεψη 2 ^{ης} περιόδου	Πρόβλεψη 3 ^{ης} περιόδου	Πρόβλεψη 4 ^{ης} περιόδου	Πρόβλεψη 5 ^{ης} περιόδου
Alt_1	130.62	89.45	48.28	7.12	-34.05
Alt_2	61.45	53.08	44.71	36.34	27.97
Alt_3	94.78	80.54	66.29	52.04	37.79

Πίνακας 5.25: Προβλέψεις 5 περιόδων για το βιβλίο Γ

6. Εγχειρίδιο χρήσης αλγορίθμων

Πρώτο βήμα σε κάθε εκτέλεση αλγορίθμου είναι η σωστή αποθήκευση των δεδομένων πώλησης και η διαμόρφωση του αρχείου ομαδοποίησης. Τα παραπάνω αρχεία χρησιμοποιούνται με την ίδια μορφή σε όλους τους αλγορίθμους που θα αναλυθούν παρακάτω.

Τα αρχεία Excel πρέπει να έχουν την παρακάτω δομή: Δύο στήλες, η πρώτη με ημερομηνίες και η δεύτερη με τις πωλήσεις. Κάθε άλλη στήλη μετά από αυτές, μπορεί να προστεθεί, αλλά δεν θα χρησιμοποιηθεί στους αλγορίθμους. Οι τίτλοι των στηλών θα πρέπει να είναι Dates και Sales.

Αντίστοιχα, θα πρέπει να δημιουργηθούν αρχεία κειμένου, τα οποία θα περιέχουν την ομαδοποίηση που ο χρήστης επιθυμεί. Κάθε βιβλίο, θα πρέπει να έχει το δικό του αρχείο. Παρατίθεται μία εικόνα για καλύτερη κατανόηση.

```
#Number of desired grouping alternatives
2
#Στο τέλος κάθε εναλλακτικής επανατοποθετείται ο πρώτος μήνας για να ολοκληρωθεί ο χρόνος.
# First month of groups
8 2 8
# First month of groups
9 3 9
```

Εικόνα 6.1: Μορφοποίηση αρχείων ομαδοποίησης

Αρχικά, στην δεύτερη σειρά ορίζεται ο επιθυμητός αριθμός εναλλακτικών. Στη συνέχεια, για κάθε εναλλακτική, τοποθετείται για κάθε ομάδα, ο αριθμός του αρχικού μήνα, θεωρώντας ως μήνα 1 τον Ιανουάριο, 2 τον Φεβρουάριο κ.ο.κ. Στο τέλος κάθε εναλλακτικής τοποθετείται ξανά ο αριθμός του πρώτου μήνα για να ολοκληρωθεί το έτος. Αυτό συμβαίνει για τον ευκολότερο υπολογισμό των ομάδων της εναλλακτικής. Επίσης, σημαντικό είναι να σημειωθεί, ότι στην κάθε ομάδα ανήκουν οι μήνες από τον πρώτο αριθμό μέχρι τον επόμενο, χωρίς αυτός να συμπεριλαμβάνεται. Δηλαδή, για το παράδειγμα της Εικόνα 6.1, και την πρώτη εναλλακτική: Η πρώτη ομάδα μηνών περιλαμβάνει τους μήνες Αύγουστο έως και Ιανουάριο του επόμενου έτους. Η δεύτερη ομάδα στην ίδια εναλλακτική, περιλαμβάνει τους μήνες από Φεβρουάριο έως και Ιούλιο.

Σημαντικό σημείο είναι για κάθε αλγόριθμο, να τοποθετηθούν στον ίδιο φάκελο τα αρχεία και όλες οι συναρτήσεις που καλούνται μέσα στον αλγόριθμο. Αναλυτικές οδηγίες για τις συναρτήσεις που καλούνται σε κάθε αλγόριθμο δίνονται σε κάθε επόμενο υποκεφάλαιο.

6.1 Μέσος όρος, ARIMA και Holt-Winters

Σε αυτόν τον αλγόριθμο, αρχικά ο χρήστης πρέπει να εισάγει τα ακριβή ονόματα των αρχείων Excel και των αρχείων κειμένου κατά αντιστοιχία, όπως φαίνονται στην παρακάτω εικόνα (Εικόνα 6.2). Στο αρχείο Book_A.xlsx αντιστοιχεί το αρχείο ομαδοποίησης month_groups_1.txt και αντιστοίχως για τα υπόλοιπα βιβλία.

```
% Define lists of Excel and TXT files
excelFiles = {'Book_A.xlsx', 'Book_B.xlsx', 'Book_C.xlsx'};
txtFiles = {'month_groups_1.txt', 'month_groups_2.txt', 'month_groups_3.txt'};
```

Εικόνα 6.2: Εισαγωγή αρχείων δεδομένων

Παρακάτω πρέπει να οριστεί ο αριθμός των μηνών που περιέχονται σε μία ομάδα στην μεταβλητή «seasonalLag».

Στην συνέχεια, μπορεί να οριστεί κάποιο άλλο όνομα για το αρχείο εξόδου στην πρότελευταία γραμμή του κώδικα.

Τέλος, οι συναρτήσεις που καλούνται μέσα στο παραπάνω αρχείο και πρέπει να βρίσκονται στον ίδιο φάκελο είναι οι εξής:

- Function_Alternatives_v1
- HoltWintersForecast_2

6.2 Απλός κινητός μέσος, προσέγγιση Syntetos-Boylan και απλή εκθετική εξομάλυνση

Όπως και στον παραπάνω αλγόριθμο, αρχικά εισάγονται τα ονόματα των αρχείων Excel και κειμένου με τον ίδιο τρόπο. Επιπλέον, όπως φαίνεται στην εικόνα, ορίζεται και το όνομα του αρχείου εξόδου.

```
% Define lists of Excel and TXT files
excelFiles = {'Book_A.xlsx', 'Book_B.xlsx', 'Book_C.xlsx'};
txtFiles = {'month_groups_1.txt', 'month_groups_2.txt', 'month_groups_3.txt'};
sheetName = 'Sheet_1';
range = 'A1:B700';
outputFileName = 'Alternatives_SBA_SMA_SES_v5.xlsx';
```

Εικόνα 6.3: Εισαγωγή αρχείων εισόδου και ονομασία αρχείου εξόδου

Παρακάτω πρέπει να οριστεί ο αριθμός των μηνών που περιέχονται σε μία ομάδα στην μεταβλητή «seasonalLag».

Οι συναρτήσεις που πρέπει να βρίσκονται στον ίδιο φάκελο με το αρχείο του συγκεκριμένου κώδικα είναι οι:

- Function_Alternatives
- Syntetos_Boylan_Approximation_V10

6.3 Διπλή εκθετική εξομάλυνση για ακανόνιστα δεδομένα

Οι είσοδοι των δεδομένων σε αυτό το αρχείο κώδικα είναι ακριβώς ίδιες με την είσοδο των δεδομένων στο 6.2 και στην Εικόνα 6.3. Δηλαδή, αρχικά εισάγονται τα ονόματα των αρχείων Excel και κειμένου.

```
% Define Excel and TXT file pairs
excelFiles = {'Book_A.xlsx', 'Book_B.xlsx', 'Book_C.xlsx'};
txtFiles = {'month_groups_1.txt', 'month_groups_2.txt', 'month_groups_3.txt'};
sheetName = 'Sheet_1';
range = 'A1:B700';
forecastHorizon = 3;
```

Εικόνα 6.4: Εισαγωγή αρχείων εισόδου

Επιπλέον, ως «n0» ορίζεται ο αριθμός των αρχικών τιμών που θα χρησιμοποιηθούν για την γραμμική παλινδρόμηση του μοντέλου.

Τέλος, στις τελευταίες γραμμές του αρχείου ορίζονται τα ονόματα των αρχείων εξόδου.

```
% Save Best RMSE metrics to an Excel file
Best_RMSE_Table = cell2table(Best_RMSE_Results(2:end,:), 'VariableNames', Best_RMSE_Results(1,:));
writetable(Best_RMSE_Table, 'Cipra_Best_RMSE_v3_2.xlsx');

% Save Best Forecast values to another Excel file
Best_Forecast_Table = cell2table(Best_Forecast_Results(2:end,:), 'VariableNames', Best_Forecast_Results(1,:));
writetable(Best_Forecast_Table, 'Cipra_Best_Forecast_v3_2.xlsx');

disp('CIPRA Best RMSE metrics saved to Cipra_Best_RMSE_v3_2.xlsx');
disp('CIPRA Best forecast values saved to Cipra_Best_Forecast_v3_2.xlsx');
```

Εικόνα 6.5: Ονομασία αρχείων εξόδου

Η συνάρτηση που πρέπει να βρίσκεται στον ίδιο φάκελο με το αρχείο του συγκεκριμένου κώδικα είναι η:

- Function_Alternatives

7. Σύνοψη

7.1 Συμπεράσματα

Κατά την παρούσα εργασία, χρησιμοποιήθηκαν δεδομένα από εκδοτικό οίκο τα οποία αφορούσαν σε 3 διαφορετικά βιβλία τα οποία εκδίδει και διανέμει ο εκδοτικός οίκος με παραγγελίες μέσω του διαδικτύου ή σε φυσικά σημεία πώλησης, όπως βιβλιοπωλεία. Αρκετές από τις πωλήσεις αφορούν και σε βιβλία που διανέμονται σε φοιτητές τριτοβάθμιας εκπαίδευσης μέσω του προγράμματος του Υπουργείου Παιδείας «Εύδοξος». Τα δεδομένα των βιβλίων, καθώς είναι βιβλία που εκδόθηκαν από το 2019 και έπειτα, είναι περιορισμένα. Τα δεδομένα είχαν την μορφή πωλήσεων και επιστροφών. Οι επιστροφές μετατράπηκαν σε μηδενικές πωλήσεις.

Έπειτα πραγματοποιήθηκαν ομαδοποιήσεις σύμφωνα με τον αλγόριθμο ομαδοποίησης που κατασκευάστηκε και τα στοιχεία που εξήγαγε χρησιμοποιήθηκαν στους αλγορίθμους πρόβλεψης.

Λόγω του μικρού πλήθους των δεδομένων, ειδικά μετά από τις ομαδοποιήσεις και τις διαφορίσεις που υπέστησαν μέσα στους αλγορίθμους, είναι αναμενόμενο να μην δώσουν αρκετά ικανοποιητικά αποτελέσματα. Ωστόσο, είναι δυνατό να εξαχθούν κάποια συμπεράσματα από αυτά και ο χρήστης να μπορεί να κατανοήσει ποια πρόβλεψη θα μπορεί να του δώσει τα πιο κοντινά αποτελέσματα στην πραγματικότητα, μιας και αυτός είναι και τελικά ο στόχος της παρούσας εργασίας.

Όπως αναφέρεται και στο κεφάλαιο 3, οι μέθοδοι που χρησιμοποιήθηκαν για τον πρώτο αλγόριθμο πρόβλεψης, συνοπτικά, ήταν: η μέθοδος πρόβλεψης σε περίπτωση λευκού θορύβου, δηλαδή με την χρήση του μέσου όρου, η μέθοδος ARIMA και η Holt-Winters. Επιπλέον, πριν την εφαρμογή των μεθόδων εφαρμόστηκαν 3 ειδών διαφορίσεις και τα αποτελέσματα αποθηκεύτηκαν σε αρχεία Excel για την καλύτερη και ευκολότερη σύγκρισή τους από τον χρήστη. Για όλες τις εναλλακτικές και για όλα τα βιβλία, τα καλύτερα αποτελέσματα προέκυψαν από χρονοσειρές που δεν υπέστησαν διαφόριση και ταυτόχρονα η μέθοδος πρόβλεψης με τα καλύτερα αποτελέσματα ήταν η Holt-Winters. Ως προς την διαφόριση, είναι αναμενόμενο, όταν ο όγκος των αποτελεσμάτων είναι εξαρχής μικρός, η οποιαδήποτε είδους διαφόριση να μην μπορεί να δώσει ικανοποιητικά αποτελέσματα. Όσον αφορά στην μέθοδο Holt-Winters, ήταν αυτή που μπόρεσε να εντοπίσει καλύτερα από τις άλλες τυχόν εποχικότητα που υπήρχε, η οποία δεν ήταν σταθερή και ευκρινής.

Όσον αφορά στον δεύτερο αλγόριθμο πρόβλεψης, στόχος ήταν να συγκριθούν οι μέθοδοι του Syntetos-Boylan Approximation, του κινητού μέσου όρου καθώς και της μεθόδου απλής εκθετικής εξομάλυνσης. Τα αποτελέσματα που δόθηκαν σε κάποιες περιπτώσεις έδειξαν ότι η μέθοδος του κινητού μέσου αποτύπωσε καλύτερα την εκάστοτε χρονοσειρά, ενώ σε άλλες έδωσε καλύτερα αποτελέσματα η προσέγγιση Syntetos-Boylan. Όσον αφορά στην απλή εκθετική εξομάλυνση, τα σφάλματα δεν ήταν ικανοποιητικά, λόγω του ότι αφενός, δεν είναι σχεδιασμένο

για την οποία διακοπτόμενη ζήτηση υπάρχει και αφετέρου δεν είναι σχεδιασμένο για χρονοσειρές με εποχικότητα.

Σχετικά με τον αλγόριθμο της διπλής εκθετικής εξομάλυνσης για ακανόνιστα δεδομένα, παρατηρήθηκε ότι σε κάποιες περιπτώσεις, όντως παρουσίασε καλύτερα αποτελέσματα από την μέθοδο ARIMA και την μέθοδο του κινητού μέσου. Η διπλή εκθετική εξομάλυνση δεν περιέχει το στοιχείο της εποχικότητας, που στις χρονοσειρές που εξετάστηκαν είναι σημαντικό. Πιθανότατα γι' αυτό στις περιπτώσεις τις οποίες τα αποτελέσματα δεν ήταν τα αναμενόμενα, αυτό να οφείλεται στην εποχικότητα.

7.2 Προτάσεις για μελλοντική έρευνα

Η παρούσα εργασία θα μπορούσε να επεκταθεί σε διάφορους τομείς του εμπορίου και ιδιαίτερα σε αυτούς όπου τα δεδομένα παρουσιάζουν διακοπτόμενη ζήτηση και εποχικότητα. Επιπλέον, θα μπορούσαν να επεκταθούν και οι αλγόριθμοι ώστε να καταλήγουν σε πρόβλεψη των επιστρεφόμενων εμπορευμάτων, γεγονός το οποίο αποτελεί πρόκληση για επιχειρήσεις όπως οι εκδοτικοί οίκοι ή για επιχειρήσεις με περιορισμένο χώρο αποθήκευσης. Επιπλέον, οι παραπάνω αλγόριθμοι λειτουργούν καλύτερα με όσο το δυνατόν περισσότερα δεδομένα. Άρα θα μπορούσε να χρησιμοποιηθεί και μία τελείως διαφορετική αρχική ομαδοποίηση από αυτήν των μηνών, όπως η εβδομαδιαία ομαδοποίηση, είτε μία πιο γενικευμένη ομαδοποίηση των βιβλίων με βάση την κατηγορία στην οποία ανήκουν ή το κοινό για το οποίο προορίζονται. Διαφορετικά δεδομένα λαμβάνονται αν το κοινό αποτελούν φοιτητές τριτοβάθμιας εκπαίδευσης εξολοκλήρου και διαφορετικά όταν το κοινό αποτελείται από ένα ευρύ φάσμα αναγνωστών. Οι αγορές των βιβλίων γίνονται με διαφορετικούς ρυθμούς σε κάθε περίπτωση και ίσως στην δεύτερη περίπτωση να υπάρχει και διαφορετική εποχικότητα ή ίσως και καθόλου.

Επιπρόσθετα, σε περιπτώσεις ύπαρξης μεγαλύτερου πλήθους δεδομένων θα μπορούσαν να εφαρμοστούν και αλγόριθμοι μηχανικής μάθησης, όπως τα νευρωνικά δίκτυα. Αντίστοιχα η χρήση αλγορίθμων όπως ο k-means θα βοηθούσαν στην καλύτερη ομαδοποίηση των παρατηρήσεων, καθώς η ομαδοποίηση παίζει καθοριστικό ρόλο στην επεξεργασία των δεδομένων από τους αλγορίθμους.

Επίσης, θα μπορούσαν να εφαρμοστούν γενετικοί αλγόριθμοι για την εύρεση των βέλτιστων παραμέτρων των παραπάνω αλγορίθμων βάσει των χρονοσειρών που υπόκεινται σε επεξεργασία.

Βιβλιογραφία

- Box, G. E., Jenkins, G. M., Reinsel, C. G., & Ljung, G. M. (2016). *Time Series Analysis: Forecasting and Control (5th Ed.)*. John Wiley & Sons, Inc.
- Cipra, T. (2005). EXPONENTIAL SMOOTHING FOR IRREGULAR DATA. *APPLICATIONS OF MATHEMATICS*, 597-604.
- Croston, J. D. (1972). Forecasting and Stock Control for Intermittent Demands. doi:10.2307/3007885
- Hyndman, R. J. (χ.χ.). Ανάκτηση από robjhyndman.com: <https://robjhyndman.com/hyndsight/hw-initialization/>
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTEXTS.
- Hyndman, R. J., Koehler, A. B., Ord, J., & Snyder, R. D. (2008). *Forecasting with Exponential Smoothing*. Springer.
- Jenkins, G. M., Reinsel, G. C., & Box, G. E. (2008). *Time Series Analysis: Forecast and Control (4th Ed.)*. John Wiley & Sons, Inc.
- Ljung, G., & Box, G. (1978). On a measure of lack of fit in time series model. doi:10.2307/2335207
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer. doi:10.1007/978-3-540-27752-1
- Montgomery, D. C., Jennings, C. L., & Kulahsi, M. (2015). *Introduction to Time Series Analysis and Forecasting (2nd Edition)*. John Wiley & Sons, Inc.
- MORGAN, P. (χ.χ.). www.pmorgan.com.au. Ανάκτηση από <https://www.pmorgan.com.au/tutorials/adaptations-of-croston-method/>
- Petropoulos, F., & Kourentzes, N. (2014, June 11). Forecast combinations for intermittent demand. *Journal of the Operational Research Society*. doi:10.1057/jors.2014.62
- Syntetos, A. A., & Boylan, J. E. (2005). The accuracy of intermittent demand estimates. 303-314. doi:10.1016/j.ijforecast.2004.10.001
- Syntetos, A. A., & Boylan, J. E. (2005). On the stock control performance of intermittent demand estimators. (E. B.V., Επιμ.) doi:10.1016/j.ijpe.2005.04.004
- Terence, M. C. (2019). *Applied time series analysis: A practical guide to modeling and forecasting*. Elsevier Inc.

Παραρτήματα

Παράρτημα Α: Συνάρτηση ομαδοποίησης

```
function [alternatives, num_Alternatives] = Function_Alternatives(excelFile, sheetName, range, txtFile,
outputFileName)
% Read the sales data from the Excel file
Data = readtable(excelFile, 'Sheet', sheetName, 'Range', range);

% Replace negative sales values with zero
Data.Sales = max(0, Data.Sales);

% Create a new table 'Sales'
Sales = Data;
SalesByMonth = table2timetable(Sales);

% Sum sales by month
SalesByMonth = retime(SalesByMonth, 'monthly', 'sum');

fileID = fopen(txtFile, 'r');

lineNum = 1; % Track the line number
groupCounter = 0; % Track the number of groups created
alternativeCounter = 0; % Track the number of alternatives created
alternatives = {}; % Initialize cell array to store alternatives

while ~feof(fileID)
    line = fgetl(fileID);

    if startsWith(line, '#')
        continue;
    end

    if lineNum == 1
        num_Alternatives = str2num(line);
        lineNum = lineNum + 1;
        continue;
    end

    months = str2num(line);
    numRanges = length(months) - 1;

    groupNames = {};
    groupSales = [];

    % Get the unique years in the data
    uniqueYears = unique(year(SalesByMonth.Date));
    maxYear = max(uniqueYears);
    minYear = min(uniqueYears);
```

```

for g = 1:numRanges
    startMonth = months(g);
    endMonth = months(g + 1);

    % Loop through each year
    for yr = uniqueYears'
        if startMonth > endMonth
            groupData1 = SalesByMonth(month(SalesByMonth.Date) >= startMonth & year(SalesByMonth.Date) == yr, :);
            groupData2 = SalesByMonth(month(SalesByMonth.Date) < endMonth & year(SalesByMonth.Date) == yr+1, :);

            if isempty(groupData2) && yr+1 <= maxYear
                currentGroup = groupData1;
            else
                currentGroup = [groupData1; groupData2];
            end

            if yr == minYear
                groupDataEarlyYear = SalesByMonth(month(SalesByMonth.Date) < endMonth & year(SalesByMonth.Date) ==
minYear, :);
                currentGroup = groupDataEarlyYear;
            end

            if yr == maxYear && isempty(groupData2)
                currentGroup = groupData1;
            end

            elseif startMonth < endMonth
                currentGroup = SalesByMonth(month(SalesByMonth.Date) >= startMonth & ...
                    month(SalesByMonth.Date) < endMonth & ...
                    year(SalesByMonth.Date) == yr, :);
            end

            totalSales = sum(currentGroup.Sales);
            groupNames{end+1} = ['Year_' num2str(yr)];
            groupSales(end+1) = totalSales;
        end
    end

    alternativeTable = table(groupNames', groupSales', 'VariableNames', {'Group', 'Sales'});
    alternativeTable = sortrows(alternativeTable, 1);

    alternativeCounter = alternativeCounter + 1;

    alternatives{alternativeCounter} = alternativeTable;

    disp(['Alternative_' num2str(alternativeCounter) ' created:']);
    disp(alternativeTable);

    lineNum = lineNum + 1;
end

```



```
fclose(fileID);
```

```
for a = 1:alternativeCounter
    alternativeTable = alternatives{a};
    sheetName = ['Alternative_' num2str(a)];
    writetable(alternativeTable,outputFileName, 'Sheet', sheetName);
end
end
```

Παράρτημα Β: Αλγόριθμος πρόβλεψης με την χρήση των μεθόδων ARIMA, Holt-Winters και μέσης τιμής

```
clc;
clear all;
```

```
excelFiles = {'Data_LEGO_new.xlsx', 'Data_Per_Filo.xlsx', 'Success_at_Work.xlsx'};
txtFiles = {'month_groups_1.txt', 'month_groups_2.txt', 'month_groups_3.txt'};
```

```
% store RMSE results
RMSE_Results = {'File', 'Alternative', 'Differencing_Method', 'WhiteNoise_RMSE', 'Best_ARIMA_RMSE',
'Holt_Winters_RMSE'};
rowIndex = 2;% Start from row 2
```

```
differencingMethods = {'none', 'first_order', 'seasonal', 'both'};
seasonalLag = 3;
m=12/seasonalLag;
```

```
best_WhiteNoise_Prediction = {};
best_ARIMA_Prediction = {};
best_HoltWinters_Prediction = {};
```

```
% Loop through each Excel file
for k = 1:length(excelFiles)
    currentExcelFile = excelFiles{k};
    currentTxtFile = txtFiles{k};
```

```
sanitizedFileName = regexp(currentExcelFile, '^[a-zA-Z0-9]', '_');
disp(['Processing data from: ', currentExcelFile]);
```

```
Alternatives = Function_Alternatives_v1(currentExcelFile, 'Sheet_1', 'A1:B700', currentTxtFile, 'Alternatives.xlsx');
numAlternatives = length(Alternatives);
```

```
for alt = 1:numAlternatives
    currentTable = Alternatives{alt};
    sales_data = currentTable.Sales;
    numObservations = length(sales_data);
```

```
mean_sales = mean(sales_data);
variance_sales = var(sales_data);
```

```
disp(['Mean value ', num2str(alt), ': ', num2str(mean_sales)]);
```

```

disp(['Variance value ', num2str(alt), ': ', num2str(variance_sales)]);

% Split the data into a training set and a testing set
trainSize = numObservations - 2;
trainData = sales_data(1:trainSize);
testData = sales_data(trainSize+1:end);
numTest = length(testData);

[h_before, pValue_before] = adftest(trainData);

disp(['ADF Test BEFORE Differencing (p-value): ', num2str(pValue_before)]);
if h_before
    disp('The initial timeseries is stationary. ');
else
    disp('The initial timeseries is not stationary. ');
end

% --- Holt Winters Forecasting ---
RMSE_HW = NaN;
bestRMSE_HW = inf;
for alpha = 0:0.1:1
    for beta = 0:0.1:1
        for gamma = 0:0.1:1
            [~,forecasts] = HoltWintersForecast_2(trainData, alpha, beta, gamma, m, numTest);

            currentRMSE = sqrt(mean((testData - forecasts).^2, 'omitnan'));
            if currentRMSE < bestRMSE_HW
                bestRMSE_HW = currentRMSE;
                best_alpha=alpha;
                best_beta=beta;
                best_gamma=gamma;
                best_HoltWinters_Forecast_Current = forecasts;
            end

        end
    end
end
RMSE_HW = min(bestRMSE_HW);
best_HoltWinters_Prediction{end+1} = best_HoltWinters_Forecast_Current;
disp(['Best RMSE: ', num2str(RMSE_HW)]);

% --- Differencing for White Noise and ARIMA ---
for j = 1:length(differencingMethods)
    best_ARIMA_Prediction_Current = [];
    best_WhiteNoise_Prediction_Current = [];

    temp_row = cell(1, 6);
    temp_row{1} = currentExcelFile;
    temp_row{2} = num2str(alt);
    temp_row{3} = differencingMethods{j};

    switch differencingMethods{j}
    case 'none'
        trainTimeSeries = trainData;
        [h_none, pValue_none] = adftest(trainTimeSeries);
        disp(['ADF Test (none) p-value: ', num2str(pValue_none)]);
    end
end

```

```

if ~isnan(RMSE_HW)
temp_row{6} = RMSE_HW;
else
temp_row{6} = NaN;
end
case 'first_order'
trainTimeSeries = diff(trainData, 1);
[h_first, pValue_first] = adftest(trainTimeSeries);
disp(['ADF Test (first_order) p-value: ', num2str(pValue_first)]);
case 'seasonal'
if length(trainData) > seasonalLag
trainTimeSeries = diff(trainData, seasonalLag);
else
continue;
end
case 'both'
if length(trainData) > (1 + seasonalLag)
trainTimeSeries = diff(diff(trainData, 1), seasonalLag);
else
continue;
end
end

% --- White Noise ---
meanValue = mean(trainTimeSeries);
predictedDifferenced_WN = repmat(meanValue, numTest, 1);
predictedOriginal_WN = zeros(numTest, 1);

switch differencingMethods{j}
case 'none'
predictedOriginal_WN = predictedDifferenced_WN;
case 'first_order'
predictedOriginal_WN(1) = trainData(end)+predictedDifferenced_WN(1);
for t_idx = 2:numTest
predictedOriginal_WN(t_idx) = predictedOriginal_WN(t_idx-1) + predictedDifferenced_WN(t_idx-1);
end
case 'seasonal'
for t_idx = 1:numTest
if t_idx <= seasonalLag
predictedOriginal_WN(t_idx) = trainData(trainSize - seasonalLag + t_idx) + predictedDifferenced_WN(t_idx);
else
predictedOriginal_WN(t_idx) = predictedOriginal_WN(t_idx - seasonalLag) + predictedDifferenced_WN(t_idx);
end
end
case 'both'
temp_predicted_first_order = zeros(numTest, 1);
for t_idx = 1:numTest
if t_idx <= seasonalLag
temp_predicted_first_order(t_idx) = predictedDifferenced_WN(t_idx) + trainData(trainSize - seasonalLag + t_idx);
else
temp_predicted_first_order(t_idx) = predictedDifferenced_WN(t_idx) + temp_predicted_first_order(t_idx - seasonalLag);
end
end

predictedOriginal_WN(1) = trainData(trainSize) + temp_predicted_first_order(1);

```

```

for t_idx = 2:numTest
predictedOriginal_WN(t_idx) = predictedOriginal_WN(t_idx - 1) + temp_predicted_first_order(t_idx);
end
end
RMSE0 = sqrt(mean((testData - predictedOriginal_WN).^2));
temp_row{4} = RMSE0;
best_WhiteNoise_Prediction{end+1} = predictedOriginal_WN;

% --- ARIMA Model ---
p_range = 0:2;
d_range = 0;
q_range = 0:2;
bestRMSE_ARIMA = NaN;
best_predictedOriginal_ARIMA = [];
for p = p_range
for d = d_range
for q = q_range
try
model = arima(p, d, q);
fitModel = estimate(model, trainTimeSeries, 'Display', 'off');
[predicted_ARIMA_diff, ~] = forecast(fitModel, numTest, 'Y0', trainTimeSeries);

predictedOriginal_ARIMA = zeros(numTest, 1);
switch differencingMethods{j}
case 'first_order'
predictedOriginal_ARIMA(1) = trainData(end) + predicted_ARIMA_diff(1);
for t_idx = 2:numTest
predictedOriginal_ARIMA(t_idx) = predictedOriginal_ARIMA(t_idx - 1) + predicted_ARIMA_diff(t_idx);
end
case 'seasonal'
for t_idx = 1:numTest
if t_idx <= seasonalLag
predictedOriginal_ARIMA(t_idx) = trainData(trainSize - seasonalLag + t_idx) + predicted_ARIMA_diff(t_idx);
else
predictedOriginal_ARIMA(t_idx) = predictedOriginal_ARIMA(t_idx - seasonalLag) +
predicted_ARIMA_diff(t_idx);
end
end
case 'both'
temp_predicted_first_order = zeros(numTest, 1);
for t_idx = 1:numTest
if t_idx <= seasonalLag
temp_predicted_first_order(t_idx) = predicted_ARIMA_diff(t_idx) + trainData(trainSize - seasonalLag + t_idx);
else
temp_predicted_first_order(t_idx) = predicted_ARIMA_diff(t_idx) + temp_predicted_first_order(t_idx -
seasonalLag);
end
end
predictedOriginal_ARIMA(1) = trainData(trainSize) + temp_predicted_first_order(1);
for t_idx = 2:numTest
predictedOriginal_ARIMA(t_idx) = predictedOriginal_ARIMA(t_idx - 1) + temp_predicted_first_order(t_idx);
end
otherwise
predictedOriginal_ARIMA = predicted_ARIMA_diff;
end
end

```

```

currentRMSE_ARIMA = sqrt(mean((testData - predictedOriginal_ARIMA).^2));

if isnan(bestRMSE_ARIMA) || currentRMSE_ARIMA < bestRMSE_ARIMA
    bestRMSE_ARIMA = currentRMSE_ARIMA;
    best_p = p;
    best_d = d;
    best_q = q;
    best_predictedOriginal_ARIMA = predictedOriginal_ARIMA;
end
catch
continue;
end
end
end
end

temp_row{5} = bestRMSE_ARIMA;
best_ARIMA_Prediction{end+1} = best_predictedOriginal_ARIMA;
RMSE_Results(rowIndex, :) = temp_row;
rowIndex = rowIndex + 1;
end
end
end
% Save to Excel

RMSE_Table = cell2table(RMSE_Results(2:end, :), 'VariableNames', RMSE_Results(1, :));

xlswrite('RMSE_Results_seasonchange_v10_3.xlsx', RMSE_Results, 'Sheet1')

disp('RMSE results saved to RMSE_Results_seasonchange_v10_3.xlsx');

```

Παράρτημα Γ: Αλγόριθμος διπλής εκθετικής εξομάλυνσης για ακανόνιστα δεδομένα

```

clc;
clear all;

excelFiles = {'Data_LEGO_new.xlsx', 'Data_Per_Filo.xlsx', 'Success_at_Work.xlsx'};
txtFiles = {'month_groups_1.txt', 'month_groups_2.txt', 'month_groups_3.txt'};
sheetName = 'Sheet_1';
range = 'A1:B700';
testSetSize = 5;

Best_RMSE_Results = {'File', 'Alternative', 'Best_Alpha_Init', 'Best_Out_of_Sample_RMSE'};
bestRMSE_RowIdx = 2;

Best_Forecast_Results = {'File', 'Alternative'};
bestForecast_RowIdx = 2;

for fileIdx = 1:length(excelFiles)
    currentExcelFile = excelFiles{fileIdx};

```

```

currentTxtFile = txtFiles{fileIdx};
disp(['Processing book: ', currentExcelFile]);

[alternatives, num_Alternatives] = Function_Alternatives(currentExcelFile, sheetName, range, currentTxtFile,
'Alternatives.xlsx');

for altIdx = 1:num_Alternatives
    currentTable = alternatives{altIdx};
    n = height(currentTable); % length of data
    currentTable.Group = (1:n)';

    trainingData = currentTable(1:n-testSetSize, :);
    testingData = currentTable(n-testSetSize+1:end, :);

    n_train = height(trainingData); % length of training data
    n_test = height(testingData); % length of testing data

    t_raw_train = table2array(trainingData(:, 1));
    y_raw_train = table2array(trainingData(:, 2));

    t_raw_test = table2array(testingData(:, 1));
    y_raw_test = table2array(testingData(:, 2));

    t_train = t_raw_train;
    y_train = y_raw_train;

    best_rmse = inf;
    best_alpha = -1;
    best_forecasts = [];

    % --- Cipra Initialization ---
    for alpha_init = 0: 0.01:1
        q = mean(diff(t_train));
        alpha = zeros(n_train, 1);
        alpha(1) = 1 - (1 - alpha_init).^q;
        beta = 1 - alpha_init;
        w = zeros(n_train, 1);
        z = zeros(n_train, 1);
        w(1) = (1 - beta^q)^2 / q * beta^q;
        z(1) = (1 - beta^q)^2 / q * beta^q;

        % Linear regression for initial b0 and b1
        n0=5;
        t_n0 = t_train(1:n0);
        y_n0 = y_train(1:n0);
        t_mean = mean(t_n0);
        y_mean = mean(y_n0);
        b1_0 = sum((t_n0 - t_mean) .* (y_n0 - y_mean)) / sum((t_n0 - t_mean).^2);
        b0_0 = y_mean - b1_0 * t_mean;

        S_level = zeros(n_train, 1);
        S_trend = zeros(n_train, 1);
        S_level(1) = b0_0 - ((q * beta^q) / (1 - beta^q)) * b1_0;
        S_trend(1) = b0_0 - 2 * ((q * beta^q) / (1 - beta^q)) * b1_0;
    end
end

```

```

% Initialize array to store In-Sample forecasts
in_sample_forecasts = zeros(n_train, 1);
in_sample_forecasts(1) = S_level(1);

% CIPRA Calculation
for idx = 2:n_train
    sum_beta = 0;
    sum_w = 0;
    sum_z = 0;
    for j = 1:idx-1
        dt = t_train(idx) - t_train(j);
        sum_beta = sum_beta + beta^dt;
        sum_w = sum_w + dt * beta^dt;
        sum_z = sum_z + (alpha(j)/w(j)) * beta^dt;
    end
    alpha(idx) = 1 / sum_beta;
    w(idx) = 1 / sum_w;
    z(idx) = 1 / sum_z;

    S_level(idx) = S_level(idx-1) * (1 - alpha(idx)) + alpha(idx) * y_train(idx);
    S_trend(idx) = S_trend(idx-1) * (1 - alpha(idx)) + alpha(idx) * S_level(idx);
end

out_of_sample_forecast_values = zeros(n_test, 1);
for i = 1:n_test
    out_of_sample_forecast_values(i) = S_level(n_train) + ((z(n_train)/w(n_train)) +
(z(n_train)/alpha(n_train))*i) * (S_level(n_train) - S_trend(n_train));
end

mse_out = mean((out_of_sample_forecast_values - y_raw_test).^2);
rmse_out = sqrt(mse_out);

if rmse_out < best_rmse
    best_rmse = rmse_out;
    best_alpha = alpha_init;
    best_forecasts = out_of_sample_forecast_values'; % Transpose for horizontal storage
end

end

Best_RMSE_Results{bestRMSE_RowIdx, 1} = currentExcelFile;
Best_RMSE_Results{bestRMSE_RowIdx, 2} = ['Alt_', num2str(altIdx)];
Best_RMSE_Results{bestRMSE_RowIdx, 3} = best_alpha;
Best_RMSE_Results{bestRMSE_RowIdx, 4} = best_rmse;
bestRMSE_RowIdx = bestRMSE_RowIdx + 1;

% Add headers for the forecast values
if bestForecast_RowIdx == 2
    forecast_headers = cell(1, n_test);
    for i = 1:n_test
        forecast_headers{i} = ['Forecast_', num2str(i)];
    end
    Best_Forecast_Results = [Best_Forecast_Results, forecast_headers];
end

```

```

% Store the best forecast values
Best_Forecast_Results{bestForecast_RowIdx, 1} = currentExcelFile;
Best_Forecast_Results{bestForecast_RowIdx, 2} = ['Alt_', num2str(altIdx)];
Best_Forecast_Results(bestForecast_RowIdx, 3:2+n_test) = num2cell(best_forecasts);
bestForecast_RowIdx = bestForecast_RowIdx + 1;

end % End of alternatives loop
end % End of excelFiles loop

% Save Best RMSE
Best_RMSE_Table = cell2table(Best_RMSE_Results(2:end,:), 'VariableNames', Best_RMSE_Results(1,:));
writetable(Best_RMSE_Table, 'Cipra_Best_RMSE_v3_2.xlsx');

% Save Best Forecast values
Best_Forecast_Table = cell2table(Best_Forecast_Results(2:end,:), 'VariableNames', Best_Forecast_Results(1,:));
writetable(Best_Forecast_Table, 'Cipra_Best_Forecast_v3_2.xlsx');

disp('CIPRA Best RMSE metrics saved to Cipra_Best_RMSE_v3_2.xlsx');
disp('CIPRA Best forecast values saved to Cipra_Best_Forecast_v3_2.xlsx');

```

Παράρτημα Δ: Συνάρτηση Holt-Winters

```

function [fitted_values, forecast_values] = HoltWintersForecast_2(data, alpha, beta, gamma, m, forecast_horizon)
n = length(data);

if n < 2 * m
    error('Not enough data. The length of the data must be at least 2*m.');
```

end

```

L = zeros(n + forecast_horizon, 1);
B = zeros(n + forecast_horizon, 1);
S = zeros(m, 1);

fitted_values = zeros(n, 1);
forecast_values = zeros(forecast_horizon, 1);
moving_avg_trend = zeros(n, 1);

for i = m + 1 : 2 * m
    moving_avg_trend(i) = mean(data(i - m : i - 1));
end

detrended_data = data(1:2 * m) - moving_avg_trend(1:2 * m);

for i = 1:m
    S(i) = mean(detrended_data(i:m:2*m));
end

seasonally_adjusted_data = zeros(2 * m, 1);
for t = 1 : 2 * m
    seasonal_idx = mod(t - 1, m) + 1;
    seasonally_adjusted_data(t) = data(t) - S(seasonal_idx);
end

```



```

x = (1 : 2 * m);
coefficients = polyfit(x, seasonally_adjusted_data, 1);
L(1) = coefficients(2); % L(1) ????? ? intercept
B(1) = coefficients(1); % B(1) ????? ? ????? (slope)

for t = 1:n

    if t <= m
        fitted_values(t) = L(1) + (t - 1) * B(1) + S(mod(t - 1, m) + 1);
    else
        fitted_values(t) = L(t - 1) + B(t - 1) + S(mod(t - 1, m) + 1);
    end

    if t < n
        L_prev = L(t);
        B_prev = B(t);
        S_prev_idx = mod(t - 1, m) + 1;

        L(t+1) = alpha * (data(t) - S(S_prev_idx)) + (1 - alpha) * (L_prev + B_prev);
        B(t+1) = beta * (L(t+1) - L_prev) + (1 - beta) * B_prev;
        S(mod(t, m) + 1) = gamma * (data(t) - L(t+1)) + (1 - gamma) * S(mod(t, m) + 1);
    end
end

% --- Forecast ---
for h = 1:forecast_horizon
    season_idx_for_forecast = mod(n + h - 1, m) + 1;
    forecast_values(h) = L(n) + h * B(n) + S(season_idx_for_forecast);
end
end

```

Παράρτημα Ε: Αλγόριθμος απλού κινητού μέσου, προσέγγισης Syntetos-Boylan και απλής εκθετικής εξομάλυνσης

```

clear all;
clc;

excelFiles = {'Data_LEGO_new.xlsx', 'Data_Per_Filo.xlsx', 'Success_at_Work.xlsx'};
txtFiles = {'month_groups_1.txt', 'month_groups_2.txt', 'month_groups_3.txt'};
sheetName = 'Sheet_1';
range = 'A1:B700';
outputFileName = 'Alternatives_SBA_SMA_SES_v9_season4_train.xlsx';

for k = 1:length(excelFiles)
    currentExcelFile = excelFiles{k};
    currentTxtFile = txtFiles{k};

    book_results_table = {'Alternative', 'RMSE_SBA', 'RMSE_SES', 'RMSE_SMA', ...
        'Future_Forecast_SBA', 'Future_Forecast_SES', 'Future_Forecast_SMA', 'Best_Diff_Method_SES'};
    rowIndex = 2;

```

```
[alternatives, num_Alternatives] = Function_Alternatives(currentExcelFile, sheetName, range, currentTxtFile,
outputFileName);
```

```
for alt = 1:num_Alternatives
    currentTable = alternatives{alt};
    sales_data = currentTable.Sales;
    numObservations = length(sales_data);
    seasonalLag = 4;
```

```
    trainSize = numObservations - 5;
    trainData = sales_data(1:trainSize);
    testData = sales_data(trainSize+1:end);
    numTest = length(testData);
```

```
    min_rmse_sba = inf;
    best_alpha_sba = 0;
    best_beta_sba = 0;
    last_z_sba = nan;
    last_p_sba = nan;
```

```
    min_rmse_ses = inf;
    best_alpha_ses = -1;
    best_diffMethod_ses = "";
    last_Y_ses = nan;
```

```
    min_rmse_sma = inf;
    best_windowSize_sma = 0;
```

```
    % --- SBA ---
```

```
    for alpha = 0:0.01:1
        for beta = 0:0.01:1
            [sba_fitted, ~,~,~] = Syntetos_Boylan_Approximation_V10(trainData, alpha, beta, 1);
            rmse_sba = sqrt(mean((trainData - sba_fitted).^2, 'omitnan'));

            if rmse_sba < min_rmse_sba
                min_rmse_sba = rmse_sba;
                best_alpha_sba = alpha;
                best_beta_sba = beta;
            end
        end
    end
```

```
    [~, final_sba_forecast, ~, ~] = Syntetos_Boylan_Approximation_V10(trainData, best_alpha_sba, best_beta_sba,
numTest);
```

```
    RMSE_SBA = sqrt(mean((testData - final_sba_forecast).^2, 'omitnan'));
```

```
    % --- SES ---
```

```
    differencingMethods = {'none', 'first_order', 'seasonal', 'both'};
```

```
    for j = 1:length(differencingMethods)
        currentDifferencingMethod = differencingMethods{j};
```

```
        % Apply differencing
        switch currentDifferencingMethod
            case 'none'
```

```

        timeSeries = trainData;
    case 'first_order'
        timeSeries = diff(trainData, 1);
    case 'seasonal'
        if length(trainData) > seasonalLag
            timeSeries = diff(trainData, seasonalLag);
        else
            continue;
        end
    case 'both'
        if length(trainData) > (1 + seasonalLag)
            tempTimeSeries = diff(trainData, 1);
            timeSeries = diff(tempTimeSeries, seasonalLag);
        else
            continue;
        end
    end
end

for alpha = 0:0.01:1
    n_train = length(timeSeries);
    Y = zeros(n_train, 1);
    fitted_values = zeros(n_train, 1);

    if n_train > 0
        Y(1) = timeSeries(1);
    end

    fitted_values(1) = NaN;

    for t = 2:n_train
        fitted_values(t) = Y(t-1);
        Y(t) = alpha * timeSeries(t-1) + (1 - alpha) * Y(t-1);
    end

    rmse_ses_current = sqrt(mean((timeSeries - fitted_values).^2, 'omitnan'));

    if rmse_ses_current < min_rmse_ses
        min_rmse_ses = rmse_ses_current;
        best_alpha_ses = alpha;
        best_diffMethod_ses = currentDifferencingMethod;
    end
end
end

% --- SES future forecasts ---
switch best_diffMethod_ses
    case 'none'
        timeSeries_best = trainData;
    case 'first_order'
        timeSeries_best = diff(trainData, 1);
    case 'seasonal'
        if length(trainData) > seasonalLag
            timeSeries_best = diff(trainData, seasonalLag);
        end
end

```

```

case 'both'
    if length(trainData) > (1 + seasonalLag)
        timeSeries_best = diff(diff(trainData, 1), seasonalLag);
    end
end

Y_last = timeSeries_best(1);
for i = 2:length(timeSeries_best)
    Y_last = best_alpha_ses * timeSeries_best(i-1) + (1 - best_alpha_ses) * Y_last;
end
fprintf('Best alpha ses: %.2f\n', best_alpha_ses);

forecasts_test = repmat(Y_last, numTest, 1);

% Reverse differencing
predictedOriginal_ES = nan(numTest, 1);
switch best_diffMethod_ses
case 'none'
    predictedOriginal_ES = forecasts_test;
case 'first_order'
    predictedOriginal_ES_temp = cumsum([trainData(end); forecasts_test]);
    predictedOriginal_ES = predictedOriginal_ES_temp(2:end);
case 'seasonal'
    for i = 1:numTest
        if i <= seasonalLag
            predictedOriginal_ES(i) = trainData(end - seasonalLag + i) + forecasts_test(i);
        else
            predictedOriginal_ES(i) = predictedOriginal_ES(i - seasonalLag) + forecasts_test(i);
        end
    end
case 'both'
    tempTimeSeries = diff(trainData, 1);
    temp_first_order_diff = nan(numTest, 1);
    for i = 1:numTest
        if i <= seasonalLag
            temp_first_order_diff(i) = tempTimeSeries(end - seasonalLag + i) + forecasts_test(i);
        else
            temp_first_order_diff(i) = temp_first_order_diff(i - seasonalLag) + forecasts_test(i);
        end
    end
    predictedOriginal_ES_temp = cumsum([trainData(end); temp_first_order_diff]);
    predictedOriginal_ES = predictedOriginal_ES_temp(2:end);
end

RMSE_SES = sqrt(mean((testData - predictedOriginal_ES).^2, 'omitnan'));

% --- SMA ---
for windowSize = 2:min(10, trainSize - 1)
    sma_forecast_current = NaN(trainSize, 1);
    for t = windowSize + 1 : trainSize
        sma_forecast_current(t) = mean(trainData(t - windowSize : t - 1));
    end
    rmse_sma = sqrt(mean((trainData - sma_forecast_current).^2, 'omitnan'));

    if rmse_sma < min_rmse_sma
        min_rmse_sma = rmse_sma;
    end
end

```

```

        best_windowSize_sma = windowSize;
    end
end

% SMA future forecast
if best_windowSize_sma > 0
    sma_forecast_test = repmat(mean(trainData(end - best_windowSize_sma + 1 : end)), numTest, 1);
else
    sma_forecast_test = NaN(numTest, 1);
end

RMSE_SMA = sqrt(mean((testData - sma_forecast_test).^2, 'omitnan'));

book_results_table{rowIndex, 1} = alt;
book_results_table{rowIndex, 2} = RMSE_SBA(1);
book_results_table{rowIndex, 3} = RMSE_SES;
book_results_table{rowIndex, 4} = RMSE_SMA;
book_results_table{rowIndex, 5} = final_sba_forecast(1);
book_results_table{rowIndex, 6} = predictedOriginal_ES(1);
book_results_table{rowIndex, 7} = sma_forecast_test(1);
book_results_table{rowIndex, 8} = best_diffMethod_ses;

rowIndex = rowIndex + 1;
end

% --- Save the final RMSE ---
book_results_table = cell2table(book_results_table(2:end, :), 'VariableNames', book_results_table(1, :));
writetable(book_results_table, outputFileName, 'Sheet', [currentExcelFile(1:end-5), '_Summary']);
end

disp('All results saved successfully to a new Excel file.');
```