

Σχολή Θετικών Επιστημών



ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

«ΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΕΣ»

«Master of Science in Informatics and Telecommunications»

Κατεύθυνση: Ψηφιακές Δεξιότητες

**«Η Επιστήμη των Δεδομένων στις Χρηματοοικονομικές
Υπηρεσίες»**

Αγγελακόπουλος Μάριος του Αθανασίου

Λαμία, Δεκέμβριος 2023

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

Σχολή Θετικών Επιστημών

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΕΣ»

«Master of Science in Informatics and Telecommunications»

Κατεύθυνση: Ψηφιακές Δεξιότητες

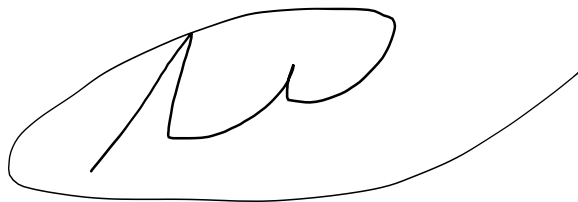
«Επιστήμη των Δεδομένων στις Χρηματοοικονομικές Υπηρεσίες»

Μεταπτυχιακή Διπλωματική Εργασία
που υποβλήθηκε στο Τμήμα Πληροφορικής και Τηλεπικοινωνιών του Πανεπιστημίου
Θεσσαλίας
ως μέρος των απαιτήσεων για την απόκτηση
Διπλώματος Μεταπτυχιακών Σπουδών στην Πληροφορική και Τηλεπικοινωνίες από τον
Αγγελακόπουλο Μάριο του Αθανασίου

Δήλωση Αυθεντικότητας, ζητήματα Copyright

«Ο μεταπτυχιακός Μ.Φ. που εκπόνησε την παρούσα διπλωματική εργασία φέρει ολόκληρη την ευθύνη προσδιορισμού της δίκαιης χρήσης του υλικού, η οποία ορίζεται στη βάση των εξής παραγόντων: του σκοπού και χαρακτήρα της χρήσης (μη-εμπορικός, μη-κερδοσκοπικός, αλλά εκπαιδευτικός-ερευνητικός), της φύσης του υλικού που χρησιμοποιεί (τμήμα του κειμένου, πίνακες, σχήματα, εικόνες κ.λπ.), του ποσοστού και της σημαντικότητας του τμήματος που χρησιμοποιεί σε σχέση με το όλο κείμενο υπό Copyright, και των πιθανών συνεπειών της χρήσης αυτής στην αγορά ή την γενικότερη αξία του υπό Copyright κειμένου».

Ο μεταπτυχιακός φοιτητής



Αγγελακόπουλος Μάριος

Σεπτέμβριος 2023

Σελίδα Τριμελούς Εξεταστικής Επιτροπής

«Η παρούσα Μ.Δ.Ε. εγκρίθηκε ομόφωνα από την τριμελή εξεταστική επιτροπή η οποία ορίστηκε από την Συνέλευση του Τμήματος Πληροφορικής και Τηλεπικοινωνιών του Πανεπιστημίου Θεσσαλίας, σύμφωνα με το νόμο και τον εγκεκριμένο Οδηγό Σπουδών του Π.Μ.Σ. «Πληροφορική και Τηλεπικοινωνίες». Τα μέλη της Επιτροπής ήταν:

- (Επιβλέπων)
- (Μέλος)
- (Μέλος)

Η έγκριση της μεταπτυχιακής διπλωματικής εργασίας από το Τμήμα Πληροφορικής και Τηλεπικοινωνιών του Πανεπιστημίου Θεσσαλίας, δεν υποδηλώνει αποδοχή των απόψεων του συγγραφέα.»

Περίληψη

Ο χρηματοοικονομικός τομέας βρίσκεται αντιμέτωπος με μία σειρά σύνθετες και εξελισσόμενες προκλήσεις, που σχετίζεται με το νομικό και ρυθμιστικό πλαίσιο, τις αυξημένες ανταγωνιστικές πιέσεις, αποτέλεσμα και της διεθνοποίησης των χρηματοοικονομικών αγορών, καθώς και νέα και ιδιαίτερα πολύπλοκα χρηματοοικονομικά προϊόντα και υπηρεσίες, η ικανοποιητική διαχείριση των οποίων είναι ιδιαίτερα δύσκολη. Οι προκλήσεις αυτές καθιστούν την αξιοποίηση του μεγάλου όγκου δεδομένων που είναι πλέον διαθέσιμος καθοριστικής σημασίας, προκειμένου επιχειρήσεις και επενδυτές να μπορούν να διαχειριστούν αποτελεσματικά το νέο ανταγωνιστικό περιβάλλον. Η εφαρμογή των μεθόδων και εφαρμογών της επιστήμης των δεδομένων στον χρηματοοικονομικό τομέα μπορούν να συμβάλλουν στη μείωση του κόστους και την αύξηση της αποτελεσματικότητας, υποστηρίζοντας αποτελεσματικά τη διαδικασία λήψης αποφάσεων και τον αυτοματισμό των διαδικασιών, την καλύτερη διαχείριση του κινδύνου και συμμόρφωση με τις απαιτήσεις του ρυθμιστικού και νομοθετικού πλαισίου. Η παρούσα διπλωματική εργασία ανέδειξε τις διάφορες εφαρμογές της επιστήμης των δεδομένων στις χρηματοοικονομικές υπηρεσίες. Μέσα από την παρουσίαση διαφόρων εφαρμογών της επιστήμης των δεδομένων, εξετάστηκαν τόσο οι μέθοδοι που αξιοποιεί όσο και τα οφέλη που μπορεί να προσφέρει, αξιοποιώντας σύγχρονες υπολογιστικές μεθόδους και τεχνικές για μία σειρά χρηματοοικονομικά προβλήματα. Η επιστήμη των δεδομένων μπορεί να συμβάλει στην κατάλληλη επιλογή ενός χαρτοφυλακίου μετοχών, τη βέλτιστη διαχείριση του κρατικού και εταιρικού χρέους και τη βελτίωση της αποτελεσματικότητας εντοπισμού της χρηματοοικονομικής απάτης. Πρόσθετες εφαρμογές της επιστήμης των δεδομένων αφορούν και την πρόβλεψη της εξέλιξης της τιμής των παραγώγων και των υποκείμενων περιουσιακών στοιχείων και την πρόβλεψη της εξέλιξης των τιμών των ακινήτων στον χρηματοπιστωτικό τομέα. Η επιστήμη των δεδομένων αναμένεται να εξελιχθεί περαιτέρω, συναντώντας όλο και συχνότερα εφαρμογή στον χρηματοοικονομικό τομέα για την επίλυση διάφορων χρηματοοικονομικών προβλημάτων. Η τεχνολογική πρόοδος θα προσδώσει περαιτέρω ώθηση στην εξέλιξη και διάδοση της επιστήμης των δεδομένων, ενός κλάδου με σημαντικές προοπτικές και οφέλη.

Λέξεις κλειδιά: Επιστήμη των δεδομένων, χρηματοοικονομικές υπηρεσίες, επιλογή χαρτοφυλακίου, διαχείριση χρέους, εντοπισμός απάτης, παράγωγα, κτηματομεσιτικός τομέας, διαχείρισης περιουσιακών στοιχείων και υποχρεώσεων.

Abstract

The financial sector is facing a number of complex and evolving challenges, related to the legal and regulatory framework, increasing competitive pressure, as a result of the internationalization of financial markets, as well as new and particularly complex financial products and services, whose effective management is particularly difficult. Due to these challenges utilizing the vast volume of data that are now available becomes critical, in order for firms and investors to be able to manage efficiently the new competitive landscape. Applying the methods and processes of data science in the financial sector can contribute to cost reductions and efficiency improvements, supporting efficiently decision-making and process automation, better risk management and legal and regulatory compliance. The present Master's Thesis highlighted the various applications of data science in the financial sector. Presenting different applications of data science, it investigated the methods data science uses and the benefits it offers, utilizing modern computing methods and techniques for a number of financial problems. Data science can help with the efficient selection of a stock portfolio, optimal state and corporate debt management and improve the efficiency of fraud detection. Additional applications of data science include price forecasting of derivatives and underlying assets as well as property prices in real estate. Data science is expected to further evolve, applied even more frequently in the financial sector to solve various financial problems. Technological progress will promote further the evolution and adoption of data science, a sector with important prospects and benefits.

Keywords: Data science, financial services, portfolio selection, debt management, fraud detection, derivatives, real estate, asset and liabilities management.

Ευχαριστίες

Με τη παρούσα Διπλωματική Εργασία ολοκληρώνεται ένα όμορφο ταξίδι σπουδών στο επιστημονικό πεδίο της επιστήμης των υπολογιστών και των τηλεπικοινωνιών και θα ήθελα να ευχαριστήσω θερμά όλους αυτούς που στάθηκαν αρωγοί σε αυτή μου την προσπάθεια:

1. Πρωτίστως, τον Επιβλέποντα Καθηγητή, κ. Κωνσταντίνο Κολομβάτσο, για την άψογη συνεργασία. Οι ουσιαστικές παρατηρήσεις και επισημάνσεις του υπήρξαν καθοριστικοί παράγοντες για την εκπόνηση αυτής της Διπλωματικής Εργασίας.
2. Δευτερευόντως τους συμφοιτητές μου με τους οποίους υπήρξε μια ευγενής άμιλλα και εξαιρετική συνεργασία όπου το απαίτησε το πρόγραμμα σπουδών
3. Τέλος ευχαριστώ και όλους τους καθηγητές – διδάσκοντες του ΠΜΣ ,οι όποιοι στάθηκαν αρωγοί και πρότυπα μετάδοσης γνώσης σε όλη τη διάρκεια των ακαδημαϊκών εξαμήνων.

Τέλος, αφιερώνω αυτή την εργασία στη μητέρα μου Αμαλία για την ενθάρρυνση και ηθική υποστήριξή της σε όλη τη διάρκεια των σπουδών μου.

Πίνακας περιεχομένων

1. ΕΙΣΑΓΩΓΗ.....	1
2. Η ΕΠΙΣΤΗΜΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ	4
2.1 ΓΕΝΙΚΑ ΓΙΑ ΤΗΝ ΕΠΙΣΤΗΜΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ.....	4
2.2 Η ΕΠΙΣΤΗΜΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΣΤΟΝ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΟ ΤΟΜΕΑ ΚΑΙ ΤΙΣ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΕΣ ΥΠΗΡΕΣΙΕΣ.....	6
2.3 ΠΑΡΑΓΟΝΤΕΣ ΠΟΥ ΣΥΜΒΑΛΛΟΥΝ ΣΤΗΝ ΠΡΟΩΘΗΣΗ ΤΗΣ ΑΞΙΟΠΟΙΗΣΗΣ ΤΗΣ ΕΠΙΣΤΗΜΗΣ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΠΡΟΚΛΗΣΕΙΣ.....	7
3. ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΧΑΡΤΟΦΥΛΑΚΙΟΥ ΜΕΤΟΧΩΝ.....	12
4. ΔΙΑΧΕΙΡΙΣΗ ΚΡΑΤΙΚΟΥ ΚΑΙ ΕΤΑΙΡΙΚΟΥ ΧΡΕΟΥΣ	21
5. ΕΝΤΟΠΙΣΜΟΣ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗΣ ΑΠΑΤΗΣ	33
6. ΕΚΤΙΜΗΣΗ ΑΞΙΑΣ ΠΑΡΑΓΩΓΩΝ ΚΑΙ ΥΠΟΚΕΙΜΕΝΩΝ ΠΕΡΙΟΥΣΙΑΚΩΝ ΣΤΟΙΧΕΙΩΝ	56
7. ΕΦΑΡΜΟΓΕΣ ΣΤΟΝ ΚΤΗΜΑΤΟΜΕΣΙΤΙΚΟ ΤΟΜΕΑ	68
8. ΔΙΑΧΕΙΡΙΣΗ ΠΕΡΙΟΥΣΙΑΚΩΝ ΣΤΟΙΧΕΙΩΝ ΚΑΙ ΥΠΟΧΡΕΩΣΕΩΝ.....	73
9. ΣΥΜΠΕΡΑΣΜΑΤΑ	77
ΒΙΒΛΙΟΓΡΑΦΙΑ	80

Κατάλογος πινάκων

Πίνακας 5.1. Η ακρίβεια των διαφόρων μεθοδολογιών κατάταξης για τον εντοπισμό παραπονημένων χρηματοοικονομικών καταστάσεων με βάση τα δύο υποδείγματα της έρευνας του Gaganis (2009) για το δείγμα επαλήθευσης	35
Πίνακας 5.2. Αποτελέσματα της ιεραρχικής άθροισης του υποδείγματος των Glancy και Yadav (2011).....	39
Πίνακας 5.3. Τα αποτελέσματα του ελέγχου για την ικανότητα ταξινόμησης του υποδείγματος των Glancy και Yadav (2011).....	40
Πίνακας 5.4. Αποτελέσματα σύγκρισης της αποτελεσματικότητας κατάταξης της μεθόδου των Huang et al. (2014) με άλλες μεθόδους	48
Πίνακας 5.5. Τα αποτελέσματα της ταξινόμησης των 3 υποδειγμάτων της έρευνας των Balla et al. (2014) για το δείγμα εκπαίδευσης και το δείγμα επικύρωσης	52
Πίνακας 5.6. Τα αποτελέσματα της ανάλυσης της ευστάθειας των αποτελεσμάτων κατάταξης του υποδείγματος των Balla et al. (2014) στον χρόνο, με τη χρήση διαδοχικών δειγμάτων	53
Πίνακας 5.7. Η ικανότητα κατάταξης του υποδείγματος των Balla et al. (2014) ανάλογα με το σύνολο μεταβλητών που χρησιμοποιείται	54

Κατάλογος διαγραμμάτων

Διάγραμμα 3.1. Οι αποδόσεις χαρτοφυλακίων με 10, 20 και 30 μετοχών που δημιουργήθηκαν με τη χρήση του υποδείγματος SVR και οι αποδόσεις του δείκτη αναφοράς σύμφωνα με την έρευνα του Huang (2012).....	20
Διάγραμμα 4.1. Το μέσο όριο (average frontier) του υποδείγματος των Consiglio και Staino (2012) με τον μέσο όρο των 100 αρχικών εισροών που είχαν ως αποτέλεσμα το χαμηλότερο κόστος, και τα διαστήματα εμπιστοσύνης 95%	23
Διάγραμμα 4.2. Πώς μεταβάλλεται η σύνθεση του χαρτοφυλακίου για διαφορετικές τιμές της υπό συνθήκη αξίας σε κίνδυνο στο υπόδειγμα των Consiglio και Staino (2012)	26
Διάγραμμα 5.1. Σύμφωνα με τη χωρική υπόθεση του υποδείγματος των Huang et al. (2014) τα παραποιημένα δεδομένα βρίσκονται στον εξωτερικό κύκλο	41
Διάγραμμα 5.2. Το δέντρο απάτης και το δέντρο μη απάτης στο υπόδειγμα των Huang et al. (2014)	44
Διάγραμμα 5.3. Ο κανόνας απάτης σχηματικά με βάση το υπόδειγμα των Huang et al. (2014)	45
Διάγραμμα 5.4. Ο κανόνας μη απάτης σχηματικά με βάση το υπόδειγμα των Huang et al. (2014)	46
Διάγραμμα 6.1. Το σφάλμα εκπαίδευσης για την πρόβλεψη της τάσης μεταξύ των διαφόρων δικτύων στην έρευνα των Quek et al. (2008)	64
Διάγραμμα 6.2. Το σφάλμα ελέγχου για την πρόβλεψη της τάσης μεταξύ των διαφόρων δικτύων στην έρευνα των Quek et al. (2008)	64
Διάγραμμα 6.3. Το σφάλμα εκπαίδευσης για την πρόβλεψη της τιμής μεταξύ των διαφόρων δικτύων στην έρευνα των Quek et al. (2008)	65
Διάγραμμα 6.4. Το σφάλμα ελέγχου για την πρόβλεψη της τιμής μεταξύ των διαφόρων δικτύων στην έρευνα των Quek et al. (2008)	65
Διάγραμμα 6.5. Οι τιμές που προβλέπει η μεθοδολογία των Quek et al. (2008) έναντι των πραγματικών τιμών για τα δικαιώματα άσκησης του χρυσού.....	67
Διάγραμμα 6.6. Οι τιμές που προβλέπει η μεθοδολογία των Quek et al. (2008) έναντι των πραγματικών τιμών για τα δικαιώματα του αργού πετρελαίου	67

1. ΕΙΣΑΓΩΓΗ

Ανέκαθεν οι άνθρωποι προσπαθούσαν να κατανοήσουν τα δεδομένα με τη χρήση επιστημονικών μεθόδων και ιδιαίτερα της στατιστικής. Μέσα σε ένα σχετικά σύντομο χρονικό διάστημα, οι ανθρώπινες κοινωνίες πέρασαν από την αναλογική στην ψηφιακή εποχή (Hilbert & López, 2011). Το γεγονός αυτό είχε σημαντικές συνέπειες στις κοινωνικές δραστηριότητες και στον τρόπο που τα άτομα επικοινωνούν και συναλλάσσονται. Δεδομένα για διάφορες δραστηριότητες συλλέγονται συνεχώς και αποθηκεύονται σε σύγχρονα πληροφοριακά συστήματα, με τις διαρκώς αυξανόμενες δυνατότητες των πληροφοριακών αυτών συστημάτων να χαρακτηρίζονται ως «Μεγάλα Δεδομένα».

Ο όρος επιστήμη των δεδομένων αναδείχθηκε σχετικά πρόσφατα ώστε να οριοθετήσει ένα νέο θεματικό πεδίο, που επιδιώκει να συμβάλλει στην κατανόηση και κατάλληλη αξιοποίηση του μεγάλου όγκου δεδομένων που είναι πλέον διαθέσιμος (Press, 2013). Η επιστήμη των δεδομένων μπορεί να θεωρηθεί ως ένας συνδυασμός θεματικών πεδίων όπως η στατιστική και η εξόρυξη δεδομένων και της πληροφορικής όπως είναι οι βάσεις δεδομένων και τα κατανεμημένα συστήματα. Συνδυάζει διάφορες προσεγγίσεις για την παραγωγή αξίας από τα ευρέως διαθέσιμα δεδομένα για οργανισμούς και άτομα, καθώς και για το σύνολο της κοινωνίας (Van Der Aalst, 2016).

Η σημασία των πληροφοριακών συστημάτων τη σύγχρονη εποχή καταδεικνύεται τόσο από το γεγονός της ραγδαίας αύξησης του διαθέσιμου όγκου δεδομένων, όσο και από τον ολοένα και σημαντικότερο ρόλο που διαδραματίζουν στη λειτουργία επιχειρήσεων και οργανισμών στο σύγχρονο οικονομικό περιβάλλον. Οι συναλλαγές λαμβάνουν χώρα κυρίως ψηφιακά, γεγονός που αυξάνει περαιτέρω και τον διαθέσιμο όγκο των δεδομένων (Van Der Aalst, 2016).

Ο χρηματοοικονομικός τομέας βρίσκεται αντιμέτωπος με μία σειρά σύνθετες και εξελισσόμενες προκλήσεις, που σχετίζεται με το νομικό και ρυθμιστικό πλαίσιο, τις αυξημένες ανταγωνιστικές πιέσεις, αποτέλεσμα και της διεθνοποίησης των χρηματοοικονομικών αγορών, καθώς και νέα και ιδιαίτερα πολύπλοκα χρηματοοικονομικά προϊόντα και υπηρεσίες, η ικανοποιητική διαχείριση των οποίων είναι ιδιαίτερα δύσκολη. Οι προκλήσεις αυτές καθιστούν την αξιοποίηση του μεγάλου

όγκου δεδομένων που είναι πλέον διαθέσιμος καθοριστικής σημασίας, προκειμένου επιχειρήσεις και επενδυτές να μπορούν να διαχειριστούν αποτελεσματικά το νέο ανταγωνιστικό περιβάλλον.

Οι μέθοδοι και οι εφαρμογές της επιστήμης των δεδομένων στον χρηματοοικονομικό τομέα μπορούν να συμβάλλουν στη μείωση του κόστους και την αύξηση της παραγωγικότητας, υποστηρίζοντας αποτελεσματικά τη διαδικασία λήψης αποφάσεων και τον αυτοματισμό των διαδικασιών, την καλύτερη διαχείριση του κινδύνου και συμμόρφωση με τις απαιτήσεις του ρυθμιστικού και νομοθετικού πλαισίου. Η εφαρμογή των μεθόδων της επιστήμης των δεδομένων στον τομέα των χρηματοοικονομικών υπηρεσιών μπορεί ακόμη να συμβάλλει στη βελτίωση των διαθέσιμων χρηματοοικονομικών προϊόντων και υπηρεσιών, μέσω της δημιουργίας καινοτόμων προϊόντων και υπηρεσιών και την προσαρμογή τους στις ιδιαίτερες ανάγκες επιχειρήσεων και επενδυτών στο σύγχρονο οικονομικό περιβάλλον (OECD, 2021).

Τα υποδείγματα μηχανικής μάθησης είναι σε θέση να εντοπίσουν σήματα και τις σχέσεις που διέπουν τα δεδομένα και να βελτιστοποιήσουν τη λειτουργική ροή των διαδικασιών και τη διαχείριση του κινδύνου. Η επιστήμη των δεδομένων μπορεί έτσι να προσδώσει σημαντικά ανταγωνιστικά πλεονεκτήματα, απλοποιώντας και αυτοματοποιώντας διαδικασίες, επιτρέποντας την εισαγωγή καινοτομιών και τη χρήση νέων, σύνθετων και αποτελεσματικών μεθόδων (OECD, 2021).

Σκοπός της παρούσας διπλωματικής εργασίας είναι να εξετάσει και να αναδείξει τις διάφορες εφαρμογές της επιστήμης των δεδομένων στις χρηματοοικονομικές υπηρεσίες. Μέσα από την παρουσίαση διαφόρων εφαρμογών της επιστήμης των δεδομένων, αναδεικνύονται τόσο οι μέθοδοι που αξιοποιεί όσο και τα οφέλη που μπορεί να προσφέρει, αξιοποιώντας σύγχρονες υπολογιστικές μεθόδους και τεχνικές για μία σειρά χρηματοοικονομικά προβλήματα.

Η δομή της παρούσας διπλωματικής εργασίας έχει ως ακολούθως. Στο δεύτερο κεφάλαιο παρουσιάζεται η επιστήμη των δεδομένων. Αναφέρονται ο ορισμός και τα χαρακτηριστικά της, παρουσιάζεται η σημασία της και η χρησιμότητα της στον χρηματοοικονομικό τομέα, καθώς και οι παράγοντες που συμβάλλουν στην προώθηση της αξιοποίησης της επιστήμης των δεδομένων, αλλά και οι προκλήσεις που σχετίζονται με την εφαρμογή των μεθόδων της επιστήμης των δεδομένων. Στο τρίτο κεφάλαιο παρουσιάζεται η βελτιστοποίηση χαρτοφυλακίου μετοχών με τη χρήση μεθόδων της επιστήμης των δεδομένων. Στο τέταρτο κεφάλαιο παρουσιάζονται εφαρμογές

της επιστήμης των δεδομένων για τη διαχείρισης εταιρικού και κρατικού χρέους και στο πέμπτο κεφάλαιο για τον εντοπισμό χρηματοοικονομικής απάτης. Στο έκτο κεφάλαιο παρουσιάζεται η χρήση υπολογιστικών μεθόδων για την εκτίμησης της αξίας των παραγώγων και των υποκείμενων περιουσιακών στοιχείων και στο έβδομο κεφάλαιο στον κτηματομεσιτικό τομέα. Τέλος, παρουσιάζονται τα συμπεράσματα της ανάλυσης που προηγήθηκε.

2. Η ΕΠΙΣΤΗΜΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

2.1 ΓΕΝΙΚΑ ΓΙΑ ΤΗΝ ΕΠΙΣΤΗΜΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

Ξεκινώντας από τον ορισμό της επιστήμης των δεδομένων, για τον Van Der Aalst (2016) η επιστήμη των δεδομένων μπορεί να οριστεί ως ένα διεπιστημονικό αντικείμενο που επιδιώκει την παραγωγή αξίας από την αξιοποίηση των δεδομένων. Τα δεδομένα μπορεί να είναι δομημένα ή αδιάρθρωτα, μεγάλου όγκου ή περιορισμένα, στατικά ή χρονικά μεταβαλλόμενα. Η παραγωγή αξίας συνίσταται στην δημιουργία προβλέψεων, την αυτοματοποίηση της διαδικασίας λήψης αποφάσεων, τη δημιουργία υποδειγμάτων από τα δεδομένα και την οπτικοποίηση των δεδομένων για την εξαγωγή συμπερασμάτων και την υποστήριξη αποφάσεων.

Η επιστήμη των δεδομένων περιλαμβάνει διαδικασίες όπως την εξαγωγή και προετοιμασία των δεδομένων, την αποθήκευση και ανάκτηση τους, τις κατάλληλες υπολογιστικές υποδομές, τις διάφορες διαδικασίες εξόρυξης των δεδομένων και εξαγωγής γνώσης από αυτά, την παρουσίαση των προβλέψεων και των συμπερασμάτων από τα δεδομένα, καθώς και την εξέταση και μίας σειράς ηθικών, κοινωνικών, νομικών και επιχειρηματικών ζητημάτων που σχετίζονται με την επεξεργασία των δεδομένων.

Κατά συνέπεια, η επιστήμη των δεδομένων αποτελεί ένα ευρύτερο αντικείμενο από την εφαρμοσμένη στατιστική και την εξόρυξη δεδομένων και αφορά την αξιοποίηση των δεδομένων για την παραγωγή αξίας για τις επιχειρήσεις και περιλαμβάνει αναφορές, προβλέψεις και προτάσεις. Παρά το γεγονός ότι για ορισμένους θεωρείται απλά προέκταση της στατιστικής, μία τέτοια προσέγγιση είναι υπεραπλουστευμένη και ανακριβής. Στην περίπτωση της επιστήμης των δεδομένων, τα υπολογιστικά ζητήματα που σχετίζονται με την επεξεργασία μεγάλου όγκου δεδομένων είναι κρίσιμης σημασίας και δίνεται έμφαση στην ανάπτυξη υποδειγμάτων. Η επιστήμη των δεδομένων αξιοποιεί την πρόοδο στην εξόρυξη δεδομένων, αναδεικνύοντας τη σημασία του εντοπισμού τάσεων και σχέσεων στα δεδομένα μέσα από δέντρα αποφάσεων και κανόνες συσχέτισης (Van Der Aalst, 2016).

Η μεγάλη διαθεσιμότητα και η ολοένα αυξανόμενη σημασία των δεδομένων οδηγεί στην ανάδειξη και την διάδοση της επιστήμης των δεδομένων για την αξιοποίηση των δεδομένων για την παραγωγή αξίας.

Η επιστήμη των δεδομένων αξιοποιεί διαφορετικά επιστημονικά αντικείμενα, τα όρια μεταξύ των οποίων δεν είναι πάντα ξεκάθαρα στα πλαίσια της επιστήμης των δεδομένων. Έτσι, η επιστήμη των δεδομένων αρχικά αξιοποιεί της μεθόδους της στατιστικής, που μπορεί να θεωρηθεί και το αρχικό επιστημονικό πεδίο από το οποίο αναδείχθηκε. Η ανάπτυξη αποτελεσματικών αλγόριθμων είναι καθοριστικής σημασίας για την ανάλυση των δεδομένων, ιδιαίτερα όταν ο όγκος τους αυξάνεται και οι αλγόριθμοι γίνονται περισσότερο σύνθετοι και πολύπλοκοι (Van Der Aalst, 2016). Πρόσθετο πεδίο που αξιοποιεί η επιστήμη των δεδομένων είναι η εξόρυξη δεδομένων. Η εξόρυξη δεδομένων αφορά την ανάλυση συχνά μεγάλου όγκου δεδομένων για την ανακάλυψη σχέσεων που δεν είναι γνωστό εκ των προτέρων ότι υφίστανται μεταξύ των δεδομένων και την παρουσίαση των δεδομένων με πρωτότυπους τρόπους που επιτρέπουν την κατανόηση τους και την κατάλληλη αξιοποίησή τους. Η εξόρυξη δεδομένων αξιοποιεί αλγόριθμους, στατιστικές μεθόδους και τις βάσεις δεδομένων για την αποτελεσματική επεξεργασία και εξέταση των δεδομένων (Hand et al., 2001).

Η επιστήμη των δεδομένων αξιοποιεί ακόμη τη μηχανική μάθηση, την εξόρυξη διαδικασιών, τα analytics, τις βάσεις δεδομένων, τα καταναμημένα συστήματα και την οπτικοποίηση των δεδομένων. Η μηχανική μάθηση αφορά τη δημιουργία υπολογιστικών υποδειγμάτων των οποίων η απόδοση βελτιώνεται με βάση την εμπειρία. Τα υποδείγματα αυτά αναπτύσσονται με βάση τα δεδομένα εισόδου και καθώς το υπόδειγμα εξελίσσεται μπορεί να αξιοποιηθεί για την δημιουργία προβλέψεων και την υποστήριξη αποφάσεων. Η εξόρυξη διαδικασιών, αντίστοιχα, αφορά την ανάπτυξη υποδειγμάτων μέσα από την ανάλυση διαδικασιών και γεγονότων. Τα analytics αφορούν την εξέταση συνόλων δεδομένων για τον εντοπισμό τάσεων και μορφών και την πρόβλεψη μελλοντικών αποτελεσμάτων και τάσεων και σχετίζονται με την επιχειρηματική ευφυΐα (business analytics) (Van Der Aalst, 2016).

Οι βάσεις δεδομένων, αντίστοιχα, χρησιμοποιούνται για την αποθήκευση των δεδομένων. Η διαχείριση των βάσεων δεδομένων έχει ως σκοπό την αποθήκευση των δεδομένων με τέτοιο τρόπο που να είναι εύκολη η διαχείριση τους και η διασφάλιση της απόδοσης κατά την ανάκτηση και αποθήκευση των δεδομένων, καθώς ο όγκος των δεδομένων αυξάνεται. Το πλέον διαδομένο είδος βάσεων δεδομένων είναι οι σχεσιακές βάσεις δεδομένων, αν και η αύξηση του όγκου των δεδομένων έχει οδηγήσει στην υιοθέτηση καταναμημένων βάσεων δεδομένων. Τα καταναμημένα συστήματα αποτελούνται από διάφορα συστήματα τα οποία αλληλεπιδρούν για την επίτευξη ενός

κοινού στόχου. Τα καταμελημένα συστήματα χρησιμοποιούνται για περιπτώσεις όπου η διαχείριση και ανάλυση των δεδομένων μπορεί να είναι τόσο σύνθετη και απαιτητική σε πόρους που δεν μπορεί να πραγματοποιηθεί σε έναν υπολογιστή, όσο ισχυρός και αν είναι. Έτσι, σε ανάλογες περιπτώσεις μία διαδικασία κατανέμεται σε μικρότερες διαδικασίες που εκτελούνται ταυτόχρονα σε διαφορετικές υπολογιστικές μονάδες (Van Der Aalst, 2016). Η οπτικοποίηση των δεδομένων αποτελεί πρόσθετο αντικείμενο της επιστήμης των δεδομένων. Ο συνδυασμός τεχνικών ανάλυσης με την οπτικοποίηση των δεδομένων, που με τις σύγχρονες τεχνικές μπορεί να επιτρέπει και την διάδραση με τους χρήστες, συμβάλλει στην κατανόηση των σχέσεων μεταξύ των δεδομένων και την υποστήριξη αποφάσεων για πολύ μεγάλα και σύνθετα σύνολα δεδομένων (Keim, et al., 2010).

Στην περίπτωση που τα υπό εξέταση δεδομένα αφορούν την ανθρώπινη συμπεριφορά, η κατανόηση τους απαιτεί και την αξιοποίηση των συμπεριφορικών επιστημών για την κατανόηση της ανθρώπινης συμπεριφοράς και της επίδρασης του κοινωνικού πλαισίου στο οποίο λαμβάνει χώρα. Σημαντική είναι, επίσης, και η διασφάλιση της ασφάλειας και της ιδιωτικότητας των δεδομένων, η υιοθέτηση ηθικών πρακτικών, καθώς και η συμμόρφωση με τη νομοθεσία και το κανονιστικό πλαίσιο (Van Der Aalst, 2016).

2.2 Η ΕΠΙΣΤΗΜΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΣΤΟΝ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΟ ΤΟΜΕΑ ΚΑΙ ΤΙΣ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΕΣ ΥΠΗΡΕΣΙΕΣ

Η επιστήμη των δεδομένων παρουσιάζει ιδιαίτερο ενδιαφέρον και χρησιμότητα για τις χρηματοοικονομικές υπηρεσίες και τον χρηματοοικονομικό τομέα. Η πληθώρα νέων και σύνθετων κανονιστικών απαιτήσεων που αντιμετωπίζει ο χρηματοοικονομικός τομέας και το νέο ανταγωνιστικό πεδίο με το οποία βρίσκεται αντιμέτωπος, οδηγούν στην ανάγκη για αξιοποίηση νέων πηγών δεδομένων και σύνθετων μετρήσεων στα χρηματοοικονομικά συστήματα. Τα χρηματοπιστωτικά ιδρύματα επεξεργάζονται συνεχώς μεγάλο όγκο δεδομένων για τη δημιουργία αναφορών και την αξιοποίηση των στοιχείων (Trelewicz, 2017).

Οι κανονισμοί που διέπουν τις χρηματοοικονομικές υπηρεσίες καθιστούν αναγκαίο τον υπολογισμό σύνθετων μετρήσεων, όπως οι προσαρμογές της αξίας των παραγώγων βάσει του πιστωτικού κινδύνου αντισυμβαλλομένου, του κόστους χρηματοδότησης, των περιθωρίων

κέρδους και άλλων χρηματοοικονομικών στοιχείων και χαρακτηριστικών. Ο υπολογισμός των στοιχείων αυτών είναι απαραίτητος για τον καθορισμό των ελάχιστων κεφαλαιουχικών απαιτήσεων ενός χρηματοπιστωτικού ιδρύματος, επιλογή κρίσιμης σημασίας που επηρεάζει την κερδοφορία του οργανισμού. Τα χρονικά δεδομένα αναλύονται για την ανάπτυξη υποδειγμάτων για τη μελέτη της λειτουργίας των αγορών και της συμπεριφοράς των πελατών. Η καταγραφή του όγκου συναλλαγών στην πάροδο του χρόνου μπορεί να συμβάλλει στην πρόβλεψη της πιθανότητας χρεοκοπίας και αθέτησης πληρωμών, προστατεύοντας έναν οργανισμό από τον κίνδυνο απώλειας κεφαλαίων.

Ο χρηματοπιστωτικός τομέας αναγνωρίζοντας τη σημασία της επιστήμης των δεδομένων και την ανάγκη αξιοποίησης του μεγάλου όγκου δεδομένων που είναι διαθέσιμος σήμερα, επενδύει σημαντικά στην επιστήμη των δεδομένων, τόσο σε τεχνολογία όσο και σε ανθρώπινο δυναμικό κατάλληλα καταρτισμένο για την ανάλυση του μεγάλου όγκου δεδομένων που βρίσκεται πλέον στη διάθεση του. Παράλληλα, οι επιχειρήσεις χρηματοοικονομικής τεχνολογίας εκμεταλλευόμενες την πρόοδο της τεχνολογίας και της επιστήμης των δεδομένων, αναπτύσσουν μία σειρά προϊόντων και υπηρεσιών που ανταποκρίνονται στις σύγχρονες ανάγκες διαχείρισης του πλούτου και των περιουσιακών στοιχείων επενδυτών και επιχειρήσεων (Trelewicz, 2017).

Συνεπώς, ο προσανατολισμός στα αποτελέσματα που χαρακτηρίζει τον χρηματοοικονομικό τομέα, η ανάγκη ανάπτυξης σύνθετων λύσεων που ανταποκρίνονται στη λειτουργία και τις ανάγκες του σύγχρονου χρηματοοικονομικού συστήματος, οι σύνθετες απαιτήσεις που θέτει το ρυθμιστικό και κανονιστικό πλαίσιο και η ανάγκη επεξεργασίας και αξιοποίησης του μεγάλου όγκου δεδομένων που αυξάνεται διαρκώς, καθιστούν τον χρηματοοικονομικό τομέα ως έναν από τους πλέον κατάλληλους τομείς για την εφαρμογή και αξιοποίηση της επιστήμης των δεδομένων.

2.3 ΠΑΡΑΓΟΝΤΕΣ ΠΟΥ ΣΥΜΒΑΛΛΟΥΝ ΣΤΗΝ ΠΡΟΩΘΗΣΗ ΤΗΣ ΑΞΙΟΠΟΙΗΣΗΣ ΤΗΣ ΕΠΙΣΤΗΜΗΣ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΠΡΟΚΛΗΣΕΙΣ

Οι Hussain και Prieto (2016) σημειώνουν μία σειρά λόγους για τους οποίους η επιστήμη των δεδομένων και τα μεγάλα δεδομένα χρησιμοποιούνται ολοένα και περισσότερο στον χρηματοοικονομικό τομέα. Οι λόγοι αυτοί αφορούν (Hussain & Prieto, 2016):

- Τον ιδιαίτερα μεγάλο όγκο δεδομένων που είναι πλέον διαθέσιμος, καθώς ο όγκος των χρηματοοικονομικών συναλλαγών αυξάνεται, όπως, για παράδειγμα, μέσω των ηλεκτρονικών συναλλαγών στις αγορές κεφαλαίου. Η χρήση ηλεκτρονικών μέσων και η ψηφιοποίηση των χρηματοοικονομικών προϊόντων και υπηρεσιών έχει οδηγήσει σε αύξηση της συναλλακτικής δραστηριότητας, καθώς μειώνει το κόστος και τη δυσκολία πραγματοποίησης χρηματοοικονομικών συναλλαγών, με αποτέλεσμα τη μεγάλη αύξηση των συναλλαγών και κατά συνέπεια και του διαθέσιμου όγκου χρηματοοικονομικών δεδομένων.
- Τις αυξανόμενες απαιτήσεις του ρυθμιστικού πλαισίου, που προτάσσουν έναν περισσότερο διαφανή και ακριβή τρόπο λειτουργίας του χρηματοοικονομικού τομέα και των χρηματοοικονομικών υπηρεσιών, που απαιτεί την αξιοποίηση του συνόλου των διαθέσιμων δεδομένων και όχι απλά αναφορές που βασίζονται σε ενδεικτικά δεδομένα. Τα χρηματοπιστωτικά ιδρύματα στρέφονται στην επιστήμη των δεδομένων και τα μεγάλα δεδομένα (big data) προκειμένου να είναι σε θέση να ανταποκριθούν στις νέες και σύνθετες ρυθμιστικές απαιτήσεις που αντιμετωπίζουν.
- Την αλλαγή του επιχειρηματικού μοντέλου και της λειτουργίας της αγοράς. Η τεχνολογική πρόοδος, σε συνδυασμό με τις διαρκείς αλλαγές στο ανταγωνιστικό περιβάλλον, θέτει τις επιχειρήσεις αντιμέτωπες με μία ραγδαία μεταβαλλόμενη αγορά που διαφέρει σημαντικά σε σχέση με τις συνθήκες μόλις λίγα χρόνια πριν. Οι αλλαγές αυτές καθιστούν επιτακτική την χρήση και αξιοποίηση της επιστήμης των δεδομένων για τη δημιουργία ενός νέου επιχειρηματικού μοντέλου που επιτρέπει στις επιχειρήσεις να διατηρήσουν την ανταγωνιστικότητα και το μερίδιο αγοράς τους, να καταφέρουν να αντιμετωπίσουν με επιτυχία τις ολοένα αυξανόμενες ανταγωνιστικές πιέσεις και να ανταποκριθούν στις νέες απαιτήσεις της αγοράς.
- Την ανάγκη για κατάλληλη αξιοποίηση των δεδομένων και την κατανόηση των αναγκών και των επιθυμιών των πελατών. Καθώς οι σχέσεις μεταξύ πελατών και επιχειρήσεων, ιδιαίτερα του χρηματοπιστωτικού τομέα, μεταβάλλονται με ταχείς ρυθμούς, εντείνεται και η ανάγκη για την κατάλληλη αξιοποίηση των δεδομένων με τη χρήση της επιστήμης των δεδομένων για τον εντοπισμό των αναγκών και των επιθυμιών των πελατών, καθώς και

των ευρύτερων καταναλωτικών τάσεων. Η επιστήμη των δεδομένων διαδραματίζει καθοριστικό ρόλο προς αυτή την κατεύθυνση.

Η Trelewicz (2017) υπογραμμίζει ότι η επιστήμη των δεδομένων και η ανάλυση μεγάλου όγκου δεδομένων συμβάλλει στην ανάπτυξη των επιχειρήσεων και την αποτελεσματικότητα στη λειτουργία τους, με αποτέλεσμα τη βελτίωση της κερδοφορίας τους.

Από την άλλη πλευρά, μία σειρά δυσκολίες και προκλήσεις δυσχεραίνουν την πλήρη αξιοποίηση της επιστήμης των δεδομένων και των μεγάλων δεδομένων από τις επιχειρήσεις. Οι δυσκολίες και οι προκλήσεις αυτές σχετίζονται με (Hussain & Prieto, 2016):

- Την ξεπερασμένη νοοτροπία και υποδομές σε ορισμένες επιχειρήσεις. Πολλές επιχειρήσεις, συμπεριλαμβανομένων και επιχειρήσεων του χρηματοπιστωτικού τομέα, βασίζονται ακόμα σε ξεπερασμένες υποδομές πληροφορικής και παλιά πληροφορικά συστήματα και θεωρούν τις υποδομές που απαιτεί η επιστήμη των δεδομένων και η αξιοποίηση μεγάλου όγκου δεδομένων ως απλά ένα πρόσθετο στοιχείο της πληροφορικής υποδομής μίας επιχείρησης και όχι σαν μία καινοτομία που αλλάζει δραστικά τον τρόπο λειτουργίας των χρηματοπιστωτικών ιδρυμάτων και των επιχειρήσεων. Ταυτόχρονα, η ξεπερασμένη νοοτροπία αποτελεί εξίσου σημαντικό πρόβλημα, με τα χρηματοπιστωτικά ιδρύματα και τις επιχειρήσεις που δεν έχουν κατανοήσει πλήρως τη λειτουργία της επιστήμης των δεδομένων και τα οφέλη της στη λειτουργία και την αποτελεσματικότητα τους, να μην προχωρούν στις κατάλληλες πρωτοβουλίες για την αξιοποίηση της επιστήμης των δεδομένων.
- Την έλλειψη των κατάλληλων δεξιοτήτων του ανθρώπινου δυναμικού. Χρηματοπιστωτικά ιδρύματα και επιχειρήσεις που δεν διαθέτουν το προσωπικό με τις κατάλληλες δεξιότητες που απαιτούνται για την εφαρμογή των μεθόδων της επιστήμης των δεδομένων, δεν είναι σε θέση να αξιοποιήσουν τις δυνατότητες που προσφέρει, με αποτέλεσμα το κενό δεξιοτήτων να αποτελεί σημαντικό ανασταλτικό παράγοντα για την αξιοποίηση της επιστήμης των δεδομένων.
- Την ικανότητα πλήρους αξιοποίησης των δεδομένων και ανάληψης πρωτοβουλιών με βάση τα δεδομένα. Η επιστήμη των δεδομένων προσφέρει σημαντικά στοιχεία και

πληροφορίες στα χρηματοπιστωτικά ιδρύματα και τις επιχειρήσεις μέσα από την κατάλληλη επεξεργασία και εξέταση των δεδομένων, αλλά εναπόκειται στις ίδιες τις επιχειρήσεις να βρουν τρόπους να αξιοποιήσουν κατάλληλα τα στοιχεία και τις πληροφορίες αυτές με βάση την οργάνωση, τους στόχους και τον τρόπο λειτουργίας τους.

- Τις ανησυχίες αναφορικά με την ασφάλεια και ιδιωτικότητα των δεδομένων. Η ασφάλεια και η ιδιωτικότητα των δεδομένων και η συμμόρφωση με τις απαιτήσεις που θέτει το ρυθμιστικό πλαίσιο επεξεργασίας των δεδομένων αποτελεί ένα σημαντικό ζήτημα. Συνεπώς, απαιτείται να διασφαλιστεί ότι η χρήση των δεδομένων γίνεται με βάση τα όσα προβλέπει το ρυθμιστικό πλαίσιο που διέπει τη λειτουργία των χρηματοπιστωτικών ιδρυμάτων και των επιχειρήσεων.

Όπως σημειώνει η Trelewicz (2017), ορισμένες επιχειρήσεις μπορεί να θεωρούν ότι η επιστήμη των δεδομένων και οι αλγόριθμοι προσφέρουν λύσεις σε τεχνικά προβλήματα, εντούτοις αμφισβητούν ότι προσφέρουν λύσεις σε επιχειρηματικά προβλήματα και ανάγκες και την αποτελεσματική υποστήριξη των λειτουργιών της επιχείρησης. Ταυτόχρονα, ορισμένες επιχειρήσεις μπορεί να μην είναι σε θέση να κατανοήσουν πώς τα δεδομένα θα μπορούσαν να χρησιμοποιηθούν για την παραγωγή υπεραξίας, ενώ μπορεί να θεωρούν, ακόμη, ότι η αξιοποίηση της επιστήμης των δεδομένων βελτιώνει μεν ως ένα βαθμό την τεχνική αποτελεσματικότητα των επιχειρηματικών λειτουργιών, αλλά δεν συμβάλλει ουσιαστικά στην επίτευξη των στόχων των επιχειρήσεων.

Όπως και οι Hussain και Prieto (2016) σημειώνει τη δυσκολία των επιχειρήσεων να βρουν προσωπικό με τις κατάλληλες δεξιότητες, καθώς και τη δυσκολία που συνεπάγονται οι κατεστημένες αντιλήψεις και τρόπος λειτουργίας στην εφαρμογή ενός σύγχρονου σχεδίου τεχνολογικού εκσυγχρονισμού των επιχειρήσεων που θα αξιοποιεί κατάλληλα τα οφέλη που προσφέρει η επιστήμη των δεδομένων και η επεξεργασία μεγάλου όγκου δεδομένων.

Ακόμη, τα υποδείγματα μηχανικής μάθησης μπορεί να οδηγήσουν στην όξυνση των προβλημάτων στις χρηματοοικονομικές αγορές, ιδιαίτερα σε περιόδους αναταραχής, εξαιτίας της δυνατότητας τους να μαθαίνουν και να προσαρμόζονται στις συνθήκες αυτόνομα, με αποτέλεσμα να δημιουργούνται νέες προκλήσεις και κίνδυνοι. Ταυτόχρονα, η χρήση μεγάλου όγκου δεδομένων,

θέτει ζητήματα που αφορούν την ποιότητα των δεδομένων που χρησιμοποιούνται στα υποδείγματα, καθώς και ζητήματα ασφάλειας και εμπιστευτικότητας (OECD, 2021).

Ιδιαίτερα κρίσιμο ζήτημα αναφορικά με τα υποδείγματα μηχανικής μάθησης είναι η δυσκολία κατανόησης και εξήγησης των λόγων και του τρόπου με τον οποίο τα υποδείγματα παράγουν τα αποτελέσματα, γεγονός που ελοχεύει σημαντικούς κινδύνους. Η ευρεία χρήση υποδειγμάτων που δεν λειτουργούν με διαφάνεια και ο τρόπος λειτουργίας τους δεν είναι εύκολα κατανοητός, μπορεί να έχει συνέπειες οι οποίες δεν είναι εύκολο να προβλεφθούν, ιδιαίτερα εάν οι χρήστες των υποδειγμάτων αυτών δεν είναι σε θέση να προβλέψουν τον τρόπο με τον οποίο η χρήση τους μπορεί να επηρεάσει τη λειτουργία των αγορών. Περιορίζοντας τη δυνατότητα των χρηστών τους να κατανοήσουν το πώς τα υποδείγματα αυτά επηρεάζουν τη λειτουργία της αγοράς συμβάλλουν στη διάχυση των σοκ και της αβεβαιότητας και μπορεί να οδηγήσουν στην ένταση των συστημικών κινδύνων επιδεικνύοντας προκυκλική συμπεριφορά. Η αδυναμία των χρηστών των υποδειγμάτων να προσαρμόσουν τη στρατηγική τους σε περιόδους αναταραχής στην αγορά, μπορεί να οδηγήσει σε αύξηση της μεταβλητότητας και περιορισμού της ρευστότητας, εντείνοντας φαινόμενα ξαφνικής κατάρρευσης.

Ταυτόχρονα, τίθενται και ζητήματα ως προς το ρυθμιστικό πλαίσιο, αναφορικά με τη διαφάνεια και τον έλεγχο των υποδειγμάτων μηχανικής μάθησης και αναδεικνύεται το γεγονός ότι η χρήση υποδειγμάτων που δεν λειτουργούν με διαφανή τρόπο, που να είναι απόλυτα κατανοητός, αντίκειται στην ισχύουσα νομοθεσία αναφορικά με τις χρηματοοικονομικές υπηρεσίες, αλλά και τις αρχές της εσωτερικής διοίκησης και διαχείρισης του κινδύνου. Συνεπώς, είναι κρίσιμης σημασίας ο διαρκής έλεγχος και κατάλληλη διαχείριση των υποδειγμάτων, προκειμένου να διασφαλιστεί η αποτελεσματική λειτουργία τους (OECD, 2021).

3. ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΧΑΡΤΟΦΥΛΑΚΙΟΥ ΜΕΤΟΧΩΝ

Η βελτιστοποίηση χαρτοφυλακίου αφορά την κατανομή των διαθέσιμων κεφαλαίων μεταξύ διαφόρων κατηγοριών περιουσιακών στοιχείων, όπως τα ομόλογα, οι μετοχές και οι καταθέσεις. Η επιστήμη δεδομένων με τη χρήση υποδειγμάτων μαθηματικής βελτιστοποίησης επιδιώκει τη βέλτιστη επιλογή περιουσιακών στοιχείων με στόχο τη μείωση του κινδύνου με βάση τη συσχέτιση μεταξύ των διαφόρων περιουσιακών στοιχείων. Σε κάθε κατηγορία περιουσιακών στοιχείων, ο διαχειριστής χαρτοφυλακίου θα πρέπει να επιλέξει θα πρέπει να επιλέξει συγκεκριμένους κλάδους και στη συνέχεια συγκεκριμένες επιχειρήσεις στις οποίες θα επενδύσει. Η κατανομή των κεφαλαίων και η επιλογή των περιουσιακών στοιχείων αποτελεί καθοριστικό παράγοντα της διαχείρισης χαρτοφυλακίου (Edirisinghe & Zhang, 2007). Ο Markowitz (1952) έθεσε τις βάσεις της διαχείρισης χαρτοφυλακίου, με βασική επιδίωξη την επίτευξη μίας ικανοποιητικής ισορροπίας μεταξύ της αναμενόμενης απόδοσης και της διακύμανσης του χαρτοφυλακίου.

Οι Liu et al. (2016) προτείνουν μία νέα μεθοδολογία επιλογής μετοχών με στόχο τη βελτίωση της απόδοσης του χαρτοφυλακίου, συγκριτικά με την παραδοσιακή προσέγγιση του υποδείγματος χαρτοφυλακίου του Markowitz. Όπως σημειώνουν, η επανεξισσορόπηση του χαρτοφυλακίου ενδείκνυται κυρίως για αγορές στις οποίες υπάρχει χαμηλή συσχέτιση των περιουσιακών στοιχείων και σχετικά ανεξάρτητες αποδόσεις, καθώς και υψηλή μεταβλητότητα. Η μεθοδολογία τους αξιοποιεί ένα υπόδειγμα μηχανικής μάθησης για τη δημιουργία ενός χαρτοφυλακίων μετοχών, του οποίου οι μετοχές είναι όσο το δυνατόν γίνεται περισσότερο ανεξάρτητες μεταξύ τους, με στόχο την αντιμετώπιση των προκλήσεων που σχετίζονται με μία στρατηγική επανεξισσορόπησης χαρτοφυλακίου, όπως η δυσκολία να βρεθεί ένα σύνολο ανεξάρτητων μετοχών, καθώς όλες οι μετοχές αντιμετωπίζουν έναν κοινό βαθμό συστημικού κινδύνου και παρουσιάζουν συνδιακύμανση με τη συνολική αγορά, καθώς και τη μεταβαλλόμενη συσχέτιση μεταξύ των μετοχών, με τις μετοχές να παρουσιάζουν μεγαλύτερη συσχέτιση κατά τις αρνητικές περιόδους στην αγορά συγκριτικά με τις θετικές.

Αναπτύσσουν μία γραφική παράσταση ελλειπτικού ζεύγους, η οποία παρέχει μία γραφική αναπαράσταση μίας πολυμεταβλητής ελλειπτικής κατανομής. Για την ερμηνεία της γραφικής παράστασης αναπτύσσουν μία ιεραρχική Γκαουσσιανή αναπαράσταση για κάθε ελλειπτική

κατανομή, η οποία επιτρέπει την ερμηνεία της ελλειπτικής γραφικής παράστασης ως ένα λανθάνων τυχαίο πεδίο Markov και, παράλληλα, παρέχει και έναν αναλυτικό μηχανισμό για την επιλογή ενός δείγματος μετοχών οι οποίες είναι όσο περισσότερο ανεξάρτητες γίνεται.

Ο αλγόριθμος επιλογής μετοχών τον οποίο αναπτύσσουν έχει ως εξής:

1. $\hat{f} \leftarrow$ Επιλογή Μετοχών $(X_1 \dots X_n, \lambda_1 \dots \lambda_m, k)$
2. **Εισαγωγή:** Οι ιστορικές αποδόσεις d μετοχών κατά τη διάρκεια n περιόδων $X_1 \dots X_n \in R^d$, μία αλληλουχία παραμέτρων ρύθμισης (tuning parameters) $\lambda_1 < \lambda_2 < \dots < \lambda_m$, ο επιθυμητός αριθμός μετοχών προς επιλογή k
3. **Αποτέλεσμα:** Ένα σύνολο δείκτη $\hat{f} \subset \{1, \dots, d\}$ με $|\hat{f}| = k$
4. για $r \in 1, \dots, m$ κάνε
5. $G^{\lambda_r} \leftarrow$ Επιλογή-Σταθερότητας (Stability-Selection) $(X_1, \dots, X_n; \lambda_r)$
6. $N_r \leftarrow$ ο αριθμός των κόμβων που δεν συνδέονται σε άλλους κόμβους στο G^{λ_r}
7. εάν $N_r \geq k$ τότε διακοπή
8. \hat{f} οι δείκτες των κόμβων που δεν συνδέονται σε άλλους κόμβους στο G^{λ_r}
9. **επιστροφή** του \hat{f}

Εξετάζουν τη μεθοδολογία τους εμπειρικά με τη χρήση 43 μετοχών από επιχειρήσεις του κλάδου της υγείας, συγκρίνοντας τη με μία στρατηγική επανεξισορρόπησης που βασίζεται σε ένα υπόδειγμα παραγόντων, το οποίο συνδυάζει την ανάλυση κυρίων συνιστωσών και την τεχνική ομαδοποίησης, καθώς και την παραδοσιακή στρατηγική χαρτοφυλακίου αγοράς-και-κατοχής του Markowitz. Με βάση τα εμπειρικά αποτελέσματα της έρευνας, η μέθοδος επιλογής με βάση το γραφικό υπόδειγμα που προτείνουν υπερέχει με βάση τις αποδόσεις, τον λόγο Sharpe και τη μέγιστη μείωση της τιμής ενός χρεογράφου από τη μέγιστη μέχρι τη χαμηλότερη τιμή της (Maximum drawdown). Η παραδοσιακή μεθοδολογία Markowitz παρουσιάζει τη χειρότερη απόδοση, με χαμηλότερη απόδοση των επιλεγμένων μετοχών του χαρτοφυλακίου, μεγαλύτερες

τυπικές αποκλίσεις και τη μεγαλύτερη μείωση της τιμής των μετοχών από τη μέγιστη μέχρι τη χαμηλότερη τιμή τους.

Εξετάζουν, επίσης, τη χρησιμότητα της νέας μεθοδολογίας με ένα σύνολο 1500 μετοχών που καλύπτουν μία περίοδο 35 ετών, από το 1981 μέχρι και το 2014. Με βάση τα ευρήματα τους, η νέα μεθοδολογία υπερέχει σημαντικά της παραδοσιακής στρατηγικής Markowitz για όλη την υπό εξέταση περίοδο, ενώ προσφέρει καλύτερες αποδόσεις και συγκριτικά με την εναλλακτική μεθοδολογία που ακολουθεί μία στρατηγική επανεξισορρόπησης με βάση ένα υπόδειγμα παραγόντων, το οποίο συνδυάζει την ανάλυση κυρίων συνιστωσών και την τεχνική ομαδοποίησης. Σημειώνουν ότι με βάση τα ευρήματα τους, η νέα μεθοδολογία που προτείνουν είναι σε θέση να εντοπίσει σε σημαντικό βαθμό τη διασύνδεση μεταξύ των μετοχών, ενώ το υπόδειγμα που βασίζεται στην ανάλυση συνιστωσών είναι σε θέση να εντοπίσει μόνο μερικώς τη σχέση αυτή.

Αναφέρουν, ακόμη, πως ο αλγόριθμος επιλογής συγκλίνει γρήγορα, ενώ, παράλληλα, μπορεί να εφαρμοστεί αποτελεσματικά σε έναν ιδιαίτερα μεγάλο αριθμό μετοχών. Καταλήγουν πως η νέα μεθοδολογία υπερέχει τόσο τις παραδοσιακής στρατηγικής διαχείρισης χαρτοφυλακίου μετοχών του Markowitz, καθώς και της εναλλακτικής στρατηγικής που εξέτασαν.

Οι Edirisinghe και Zhang (2007) αναπτύσσουν ένα υπόδειγμα ανάλυσης της περιβάλλουσας (Data Envelopment Analysis) για την αξιολόγηση των δημόσιων χρηματοοικονομικών καταστάσεων των εταιρειών, για τη δημιουργία ενός δείκτη σχετικής χρηματοοικονομικής ισχύος (relative financial strength indicator – RSFI) με στόχο την πρόβλεψη της εξέλιξης των τιμών των μετοχών. Ο δείκτης RSFI βασίζεται στη μεγιστοποίηση της συσχέτισης μεταξύ της αξιολόγησης χρηματοοικονομικής ισχύος και της απόδοσης της μετοχής. Ο δείκτης δεν επιδιώκει την αξιολόγηση της χρηματοοικονομικής ισχύος της κάθε εταιρείας ξεχωριστά, αλλά σε συνάρτηση με άλλες επιχειρήσεις του κλάδου στον οποίο δραστηριοποιείται.

Ο δείκτης Σχετικής Χρηματοοικονομικής Ισχύος RFSI για μία επιχείρηση j για την περίοδο t , με βάση τις δημόσιες χρηματοοικονομικές καταστάσεις του κλάδου, ορίζεται ως:

$$RFSI(t, j; \hat{t}) = \frac{1}{\hat{t}} \sum_{\tau=t-\hat{t}}^{t-1} \eta_j(y^*, z^*; x^\tau)$$

Όπου $\frac{1}{\hat{t}} \sum_{\tau=\hat{t}-\hat{t}}^{\hat{t}-1}$ ο απλός κινούμενος μέσος όρος για \hat{t} περιόδους και $\eta_j(y^*, z^*; x^T)$ αποτελεί τη βαθμολογία απόδοσης της επιχείρησης j που αντιστοιχεί στο διανυσματικό ζεύγος $\eta_j(y^*, z^*)$ και μπορεί να υπολογιστεί σύμφωνα με το ακόλουθο πρόγραμμα γραμμικού προγραμματισμού:

$$\begin{aligned} \eta_j(y^*, z^*) &:= \sum_{i=1}^I (z_i x_{ij}) v_{ij} \\ \text{s.t.} \quad &\sum_{i=1}^I (y_i x_{jk}) u_{ij} = 1 - (y_i x_{jk}) u_{ij} + \\ &\sum_{i=1}^I (z_i x_{ik}) v_{ij} \leq 0, k = 1, \dots, K \\ &u_{ij}, v_{ij} \geq 0, i = 1, \dots, I \end{aligned}$$

όπου k το σύνολο των επιχειρήσεων του κλάδου.

Ο δείκτης λαμβάνει τιμές από το 0 έως το 1, με την τιμή της μονάδας να αποτελεί τη μέγιστη τιμή του δείκτη σχετικής χρηματοοικονομικής ισχύος της επιχείρησης j σε σχέση με τον κλάδο στον οποίο δραστηριοποιείται.

Αξιολογούν εμπειρικά τη χρησιμότητα του δείκτη για τη βελτιστοποίηση της σχέσης απόδοσης και κινδύνου ενός χαρτοφυλακίου μετοχών με τη χρήση των τριμηνιαίων δημοσιευμένων χρηματοοικονομικών καταστάσεων για διάφορες εισηγμένες επιχειρήσεις για την περίοδο μεταξύ 1996 και 2002. Το δείγμα περιλαμβάνει επιχειρήσεις από 6 τεχνολογικούς κλάδους, τον κλάδο του λογισμικού, του εξοπλισμού τηλεπικοινωνιών, των ηλεκτρονικών υπολογιστών, των δικτύων, των ημιαγωγών και των υπηρεσιών υπολογιστών. Σύμφωνα με τα ευρήματα της έρευνας, η βελτιστοποίηση χαρτοφυλακίου με βάση τον δείκτη RFSI προσφέρει καλύτερες ημερήσιες αθροιστικές αποδόσεις συγκριτικά με την παραδοσιακή στρατηγική βελτιστοποίησης μέσου – διακύμανσης, αλλά και συγκριτικά με τον δείκτη S&P500. Οι ημερήσιες αθροιστικές αποδόσεις ενός χαρτοφυλακίου μετοχών με βάση τον δείκτη RFSI ξεπέρασαν το 20%, με τις αντίστοιχες αποδόσεις για την παραδοσιακή στρατηγική του Markowitz να κυμαίνονται γύρω στο 10% και τις ημερήσιες αθροιστικές αποδόσεις του δείκτη S&P500 να εμφανίζονται αρνητικές σε ορίζοντα 60 ημερών. Όπως σημειώνουν, ο δείκτης RFSI μπορεί να βοηθήσει τους διαχειριστές κεφαλαίων να

αξιολογήσουν αν θα πρέπει να επενδύσουν σε έναν συγκεκριμένο κλάδο ή μία μετοχή με βάση την χρηματοοικονομική ισχύ του κλάδου ή της επιχείρησης.

Ο Quah (2008) εξετάζει το ζήτημα της εκπαίδευσης των μοντέλων μαλακής υπολογιστικής (soft-computing) με τη χρήση μίας μεθοδολογίας την οποία αποκαλεί καμπύλη σχετικών λειτουργικών χαρακτηριστικών (relative operating characteristics curve - ROC). Εξετάζει 3 υποδείγματα μαλακής υπολογιστικής, το υποδείγμα multi-layer perceptrons (MLP), το υπόδειγμα adaptive neuro-fuzzy integrative systems (ANFIS) και το υπόδειγμα general growing and pruning radial basis function (GGAP - RBF).

Τα μοντέλα μαλακής υπολογιστικής χρησιμεύουν για την επίλυση σύνθετων προβλημάτων πολλαπλών μεταβλητών και παραμέτρων, για τα οποία είναι ιδιαίτερα δύσκολη η εύρεση μίας ακριβούς λύσης. Καθώς το κόστος επίλυσης ανάλογων προβλημάτων είναι ιδιαίτερα μεγάλο, τα συγκεκριμένα υποδείγματα προσεγγίζουν μία λύση με έναν βαθμό αβεβαιότητας, προσφέροντας, όμως, τη δυνατότητα λύσεων που είναι επαρκείς και εφικτό να υπολογιστούν με σχετικά χαμηλό κόστος. Τα βασικά στοιχεία της μαλακής υπολογιστικής είναι η ασαφής λογική (fuzzy logic), τα τεχνητά νευρωνικά δίκτυα, η πιθανοτική λογική, η εξελικτική υπολογιστική, η μηχανική μάθηση και η θεωρία του χάους.

Η θεμελιώδης ανάλυση για την αξιολόγηση των μετοχών βασίζεται στην υπόθεση ότι υπάρχει μία σχέση μεταξύ της τιμής της μετοχής και των θεμελιωδών χαρακτηριστικών μίας επιχείρησης, όπως η κεφαλαιοποίηση, ο λόγος τιμής προς κέρδη ή το μέγεθος της επιχείρησης. Το πρόβλημα ταξινόμησης που εξετάζει αφορά την ταξινόμηση των μετοχών σε 2 κλάσεις, την Κλάση 1, που περιλαμβάνει κάθε μετοχή η τιμή της οποίας αναμένεται να ανατιμηθεί κατά τουλάχιστον 80% και την Κλάση 2 στην οποία ανήκουν οι υπόλοιπες μετοχές. Καθώς τα δεδομένα δεν είναι ισορροπημένα, αφού οι μετοχές που ανήκουν στην Κλάση 1 είναι σαφώς λιγότερες, ο Quah (2008) χρησιμοποιεί την τεχνική της υπερδειγματοληψίας για να αποφύγει την απώλεια πληροφοριών που συνεπάγεται η χρήση της τεχνικής της υποδειγματοληψίας.

Για τον υπολογισμό των μοντέλων χρησιμοποιήθηκαν ο λόγος της τιμής της μετοχής προς τα έσοδα, η λογιστική αξία ανά μετοχή, η απόδοση επί των ιδίων κεφαλαίων, η αναλογία πληρωμής μερισμάτων, το ετήσιο μέρισμα ανά μετοχή επί την τιμή της μετοχής, ο λόγος της τιμής προς τη λογιστική αξία, το κυκλοφορούν ενεργητικό, το συνολικό χρέος, ο σταθμισμένος μέσος αριθμός

των μετοχών, το κυκλοφορούν ενεργητικό επί των βραχυπρόθεσμων υποχρεώσεων και τα έσοδα ανά μετοχή.

Δείγμα της έρευνας αποτέλεσαν 1630 μετοχές για μία περίοδο 10 ετών, από το 1995 μέχρι το 2004. Χρησιμοποιείται το 80% των δεδομένων, οι τιμές των μετοχών για τα πρώτα 8 χρόνια, για την πρόβλεψη του υπόλοιπου 20% των δεδομένων, δηλαδή τις τιμές των μετοχών για τα τελευταία 2 χρόνια του δείγματος. Με βάση τα ευρήματα της έρευνας, ο υπολογιστικός χρόνος που απαιτείται για την εκπαίδευση του μοντέλου GGAP - RBF είναι 360,7 λεπτά, που είναι τάξης μεγέθους μεγαλύτερος των αντίστοιχων χρόνων που απαιτούνται για την εκπαίδευση του μοντέλου MLP, του οποίου ο υπολογιστικός χρόνος που απαιτείται είναι 188,45 δευτερόλεπτα, και του μοντέλου ANFIS, που είναι 396,85 δευτερόλεπτα. Τα μοντέλα MLP και ANFIS παρουσιάζουν, επίσης, σημαντική συσχέτιση μεταξύ των προβλέψεων τους και την πραγματική εξέλιξη των τιμών των μετοχών. Αντίθετα, το μοντέλο GGAP-RBF παρουσιάζει χαμηλή συσχέτιση, συνεπώς δεν είναι κατάλληλο για υποδείγματα πρόβλεψης της εξέλιξης των τιμών των μετοχών.

Καθώς το όριο ανατίμησης των μετοχών αυξάνεται, μειώνεται ο αριθμός των επιλεγόμενων μετοχών, αλλά αυξάνεται το μέσο ποσοστό ανατίμησης των μετοχών. Η καμπύλη σχετικών λειτουργικών χαρακτηριστικών (ROC) περιέχει δύο μεταβλητές από έναν πίνακα σύγχυσης (confusion matix), το ποσοστό αληθώς θετικών και το ποσοστό ψευδώς αρνητικών, ο οποίος χρησιμοποιείται για τον προσδιορισμό της απόδοσης ενός αλγορίθμου ταξινόμησης. Η μεθοδολογία που προτείνει ο Quah (2008) επιδιώκει τη μεγιστοποίηση του ποσοστού αληθώς θετικών επιλεγμένων μετοχών και την ελαχιστοποίηση του ποσοστού ψευδούς αρνητικών απορριφθέντων μετοχών. Το υπόδειγμα ANFIS παρουσιάζει τη μεγαλύτερη ακρίβεια και μέσο ποσοστό ανατίμησης των επιλεχθέντων μετοχών σύμφωνα με τα ευρήματα της έρευνας ενώ, αντίθετα, το υπόδειγμα GGAP-RBF, αποδείχθηκε ιδιαίτερα αναποτελεσματικό για την επιλογή μετοχών. Για το σύνολο ελέγχου η μέση ανατίμηση των μετοχών για το μοντέλο MPL έφτασε το 13%, για το μοντέλο ANFIS το 14,3% και για το μοντέλο GGAPRBF το 11,15%. Η ακρίβεια του μοντέλου MLP φτάνει στο 70%, του μοντέλου ANFIS στο 74,85% και του μοντέλου GGAPRBF στο 61,70%. Συνεπώς, το μοντέλο ANFIS αποδείχθηκε το περισσότερο αποτελεσματικό, τόσο όσον αφορά το μέσο ποσοστό ανατίμησης των μετοχών που επιλέχθηκαν και όσον αφορά την ακρίβεια του μοντέλου, ενώ το μοντέλο GGAPRBF, αντίστοιχα, το λιγότερο αποτελεσματικό. Το

μοντέλο GGAPRBF είναι ιδιαίτερα αναποτελεσματικό και όσον αφορά την υπολογιστικό χρόνο που απαιτείται για την εκπαίδευση του.

Η έρευνα διαπιστώνει ότι η χρήση νευρωνικών δικτύων μπορεί να είναι αποτελεσματική για την επιλογή ενός χαρτοφυλακίου μετοχών. Η πιθανότητα ανατίμησης των μετοχών αυξάνεται όσο αυξάνονται οι προβλεπόμενες τιμές με βάση τα αποτελέσματα των μοντέλων. Το μοντέλο ANFIS αναδεικνύεται ως το πλέον αποτελεσματικό, όσον αφορά την ακρίβεια και το ποσοστό ανατίμησης των μετοχών που επιλέγει.

Το ζήτημα της αποτελεσματικής επιλογής μετοχών για τη δημιουργία ενός χαρτοφυλακίου μετοχών εξετάζει η έρευνα του Huang (2012), ο οποίος προτείνει μία νέα μεθοδολογία με τη χρήση παλινδρόμησης διανύσματος υποστήριξης (support vector regression - SVR) και γενετικών αλγόριθμων. Η μεθοδολογία του συνίσταται στη χρήση της παλινδρόμησης SVR για την αποτελεσματική κατάταξη των μετοχών. Οι γενετικοί αλγόριθμοι χρησιμεύουν για τη βελτιστοποίηση των παραμέτρων του μοντέλου και την επιλογή των κατάλληλων χαρακτηριστικών των μετοχών για τον καθορισμό βέλτιστων υποσυνόλων μεταβλητών εισόδου για το μοντέλο SVR.

Κριτήριο κατάταξης αποτελούν οι αναμενόμενες αποδόσεις των μετοχών στο μέλλον. Τα θεμελιώδη χαρακτηριστικά των μετοχών που χρησιμοποιούνται ως μεταβλητές εισόδου του υποδείγματος, αποτυπώνουν το κατά πόσο η τιμή της μετοχής είναι εύλογη, την ανάπτυξη της επιχείρησης, την κερδοφορία της, τη ρευστότητα και την αποτελεσματικότητα της, καθώς και τη μόχλευση της επιχείρησης. Τα χαρακτηριστικά αυτά προσεγγίζονται με τη χρήση 14 χρηματοοικονομικών λόγων. Οι κορυφαίες μετοχές με βάση το μοντέλο επιλέγονται για τη δημιουργία του χαρτοφυλακίου.

Το μοντέλο SVR προβλέπει την τιμή μίας μετοχής i κατά τον χρόνο t ως $y_{i,t}(F, \theta) \in R$ για $i=1 \dots n$, όπου F και θ αποτελούν τα σύνολα των θεμελιωδών χαρακτηριστικών και των παραμέτρων πυρήνα που χρησιμοποιεί το μοντέλο. Η πρόβλεψη των αποδόσεων των μετοχών δεν απαιτείται να είναι τέλεια, αλλά να προσφέρει μία αποτελεσματική προσέγγιση για την κατάταξη των μετοχών. Η κατάταξη των μετοχών δίνεται από:

$$a_{i,t}(F, \theta) = \rho(y_{i,t}(F, \theta))$$

όπου $\alpha_{i,t} \in N$ είναι η κατάταξη της μετοχής i τον χρόνο t και $\alpha_{i,\tau} \geq \alpha_{j,t}$ εάν $y_{it} \geq y_{jt}$.

Η απόδοση ενός χαρτοφυλακίου δίνεται από τον μέσο όρο των αποδόσεων των μετοχών που περιλαμβάνονται στο χαρτοφυλάκιο.

Ο αλγόριθμος επιλογής μετοχών με βάση το μοντέλο SVR που προτείνει είναι:

Βήμα 1: $i \leftarrow 1$.

Βήμα 2: Χρησιμοποίησε τα δεδομένα εισόδου (F, θ) και τις πραγματικές αποδόσεις των μετοχών το έτος i για την εύρεση των διανυσμάτων υποστήριξης για τα σύνολα της υπό εξέταση περίπτωσης.

Βήμα 3: Χρησιμοποίησε τα δεδομένα εισόδου (F, θ) για το έτος $i + 1$ για τον υπολογισμό των εκτιμώμενων ετήσιων αποδόσεων των μετοχών για την κατάταξη των μετοχών.

Βήμα 4: Επιλογή των κορυφαίων m μετοχών με βάση την κατάταξη στο βήμα 3. Υπολογισμός των μέσων ετήσιων αποδόσεων των επιλεγμένων μετοχών.

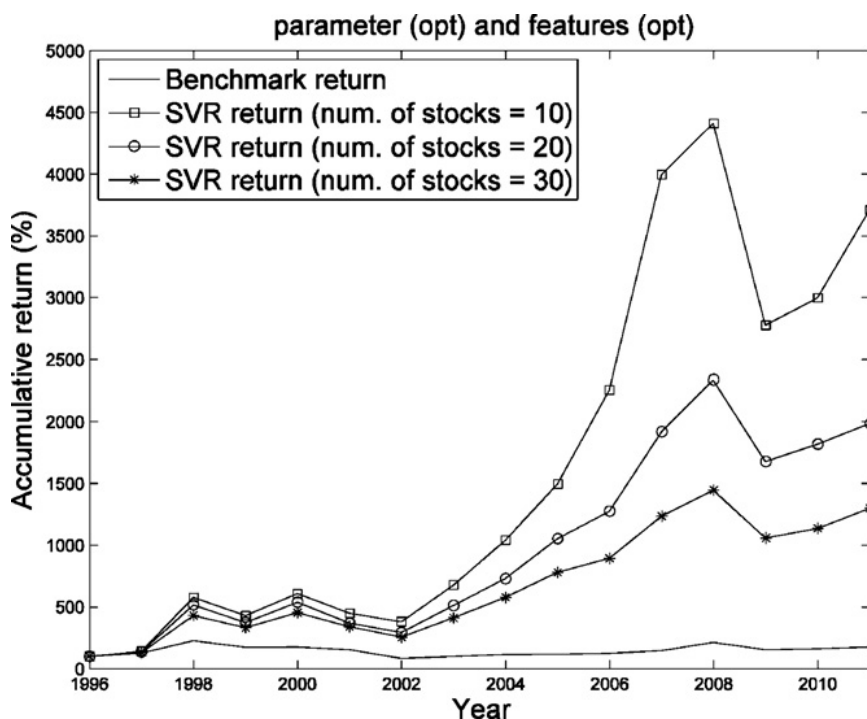
Βήμα 5: $i \leftarrow i + 1$. Επανάληψη των βημάτων 2 έως 4 έως $i = n - 1$.

Βήμα 6: Ανατοκισμός των μέσων ετήσιων αποδόσεων για τον υπολογισμό της αθροιστικής συνολικής απόδοσης του χαρτοφυλακίου σε ορίζονται n χρόνων, ως $R_t = \prod_{t=1}^n R_t$.

Για την αξιολόγηση του υποδείγματος χρησιμοποίησε δεδομένα για τις τιμές 200 μετοχών των εισηγμένων εταιρειών με τη μεγαλύτερη κεφαλαιοποίηση στο Χρηματιστήριο της Ταϊβάν και τις ετήσιες χρηματοοικονομικές καταστάσεις τους. Ως δείκτης αναφοράς χρησιμοποιείται το γινόμενο των μέσων ετήσιων αποδόσεων και των 200 μετοχών. Με βάση τα ευρήματα της έρευνας ένα χαρτοφυλάκιο 10, 20 ή 30 επιλεγμένων μετοχών με βάση το υπόδειγμα SVR προσφέρει καλύτερες αποδόσεις από τον δείκτη αναφοράς, ακόμα και χωρίς βελτιστοποίηση των παραμέτρων και των θεμελιωδών χαρακτηριστικών των μετοχών, αλλά ιδιαίτερα όταν το υπόδειγμα υπολογίζεται με τη βελτιστοποίηση των παραμέτρων και των χαρακτηριστικών των μετοχών.

Στο διάγραμμα που ακολουθεί παρουσιάζονται οι αποδόσεις χαρτοφυλακίων με 10, 20 και 30 μετοχών που δημιουργήθηκαν με τη χρήση του υποδείγματος SVR και οι αποδόσεις του δείκτη αναφοράς.

Διάγραμμα 3.1. Οι αποδόσεις χαρτοφυλακίων με 10, 20 και 30 μετοχών που δημιουργήθηκαν με τη χρήση του υποδείγματος SVR και οι αποδόσεις του δείκτη αναφοράς σύμφωνα με την έρευνα του Huang (2012)



Πηγή: Huang (2012)

Ο Huang (2012) καταλήγει ότι η χρήση του υποδείγματος SVR για τη δημιουργία ενός χαρτοφυλακίου μετοχών μπορεί να προσφέρει σημαντικές αθροιστικές αποδόσεις για τους επενδυτές. Η επιλογή των κατάλληλων θεμελιωδών χαρακτηριστικών των μετοχών αναδεικνύεται ως ιδιαίτερα σημαντική για τη δημιουργία ενός χαρτοφυλακίου μετοχών με τη χρήση του συγκεκριμένου υποδείγματος.

4. ΔΙΑΧΕΙΡΙΣΗ ΚΡΑΤΙΚΟΥ ΚΑΙ ΕΤΑΙΡΙΚΟΥ ΧΡΕΟΥΣ

Η λήψη αποφάσεων υπό καθεστώς αβεβαιότητας αφορά μία σειρά προβλημάτων, όπου η χρονική διάσταση και η αβεβαιότητα και δυσκολία πρόβλεψης της μελλοντικής εξέλιξης μίας μεταβλητής διαδραματίζουν καθοριστικό ρόλο. Ο στοχαστικός προγραμματισμός πολλαπλών σταδίων αποτελεί μία από τις μεθοδολογίες για την επίλυση προβλημάτων που ενέχουν αβεβαιότητα και μεταβάλλονται χρονικά (Bertocci et al., 2006).

Η δομή του κρατικού χρέους μίας χώρας εξαρτάται από το νόμισμα έκδοσης και την ωριμότητα των χρεογράφων που χρησιμοποιούνται για τη χρηματοδότηση των κρατικών δαπανών. Οι Consiglio και Staino (2012) αναπτύσσουν ένα υπόδειγμα γραμμικού προγραμματισμού για τη βελτιστοποίηση της έκδοσης κρατικού χρέους, καθώς η αβεβαιότητα που σχετίζεται με τη μελλοντική εξέλιξη του πληθωρισμού, των επιτοκίων και των υπόλοιπων μακροοικονομικών μεταβλητών καθιστά τη διαχείριση του χρέους και την κατανομή του ανάμεσα στα διάφορα είδη χρεογράφων μία ιδιαίτερα σύνθετη και δύσκολη υπόθεση.

Η δομή του υποδείγματος έχει ως εξής. Έστω ότι $T \equiv \{t_0, t_1, \dots, t_H\}$ τα διάφορα στάδια λήψης αποφάσεων. Οι αποφάσεις αναφορικά με την κατανομή του χρέους ανάμεσα στα διάφορα χρεόγραφα λαμβάνονται αποκλειστικά με βάση τις πληροφορίες που είναι διαθέσιμες μέχρι το κάθε στάδιο. Ο περιορισμός αυτός εισάγεται σε ένα υπόδειγμα με τη χρήση περιορισμών ισότητας ή την αναπαράσταση της στοχαστικής διαδικασίας με τη χρήση ενός δέντρου σεναρίων. Η αβεβαιότητα σχετικά με τα επιτόκια αναπαρίσταται από μία στοχαστική διαδικασία επί του συνόλου T , και το υποσύνολο του πεδίου ορισμού της που περιέχει τα μη μηδενικά στοιχεία (support function) είναι ο χώρος πιθανότητας (Ω, F, P) . Το N είναι το σύνολο των κόμβων και $N_{t_h} \subset N$ είναι το σύνολο των κόμβων στο στάδιο t_h . Ο αρχικός κόμβος $0 \in N_{t_0}$ είναι ο πρόγονος όλων των άλλων κόμβων.

Με βάση ένα δέντρο σεναρίων, κάθε κόμβος $n \in N_{t_h}$ και $t_h \in T/t_0$ έχει έναν μοναδικό γονέα που δίνεται από $a(n) \in N_{t_{h-1}}$ και $t_h \in T/t_H$ και διαθέτει ένα μη κενό σύνολο κόμβων παιδιών $C(n) \subset N_{t_{h+1}}$. Το σύνολο των κόμβων που ανήκουν στη μοναδική διαδρομή από τη ρίζα μέχρι το $n \in N_{t_h}$ δίνεται από $H(n)$. Σε ένα δέντρο σεναρίων υπάρχει μία αντιστοίχιση ένα προς ένα μεταξύ των κόμβων $n \in N_{t_h}$ και των συνόλων κατάτμησης A_{t_h} για κάθε $t_h \in T$.

Στο υπόδειγμα των Consiglio και Staino (2012) η στοχαστική διαδικασία ζ περιγράφει την εξέλιξη των επιτοκίων, η οποία δίδεται από I επιτόκια των οποίων οι τιμές σε κάθε κόμβο n καθορίζονται από το διάνυσμα $R_n = (R_n^1, \dots, R_n^I)$ και βασίζονται στο υπόδειγμα διάχυσης (diffusion model) των Bernaschi et al. (2007), εκμεταλλευόμενοι τη Μαρκοβιανή δομή των δεδομένων, δημιουργώντας το R_n με βάση το προηγούμενο διάστημα $R_{a(n)}$ και την τυχαία μεταβλητή ϵ_n ως:

$R_n = F(R_{a(n)} + \epsilon_n)$ όπου $F(\cdot)$ είναι η συνάρτηση μετάβασης.

Επιδιώκεται η έκδοση ενός χαρτοφυλακίου ομολόγων που θα καλύπτουν το χρέος σε κάθε κατάσταση της οικονομίας $n \in N_t$ και για κάθε στάδιο απόφασης $t_h \in T$. Με βάση τις υποθέσεις του υποδείγματος, ο υπουργός οικονομικών μπορεί να εκδώσει ομόλογα μηδενικού ή σταθερού κουπονιού διαφορετικής ωριμότητας και ο συνολικός αριθμός των διαθέσιμων χρεογράφων είναι J . Το R_0 είναι γνωστό κατά τον χρόνο t_0 , και αποτελεί την καμπύλη των επιτοκίων που παρατηρείται σήμερα, με τη διαδικασία λήψης αποφάσεων να δίνεται από $x_n \equiv \{x_{th}\}_{th} \in T$. Η απόφαση που λαμβάνεται σε κάθε κόμβο αποτελεί ένα διάνυσμα $x_n = (x_n^1, \dots, x_n^J)$. Οι αποφάσεις x καθορίζονται ενδογενώς από το υπόδειγμα βελτιστοποίησης και αναπαριστούν την ονομαστική αξία κάθε ομολόγου για την κάλυψη του χρέους σε κάθε κατάσταση της οικονομίας.

Το χρέος που θα πρέπει να καλυφθεί σε κάθε κόμβο $n \in N$ ορίζεται ως O_n και προκειμένου η επιλογή να είναι βέλτιστη θα πρέπει σε κάθε κατάσταση της οικονομίας το χρέος να καλύπτεται από την έκδοση ομολόγων από το διαθέσιμο σύνολο J . Επιδιώκεται ο καθορισμός της τελικής κατανομής κόστους ώστε να ελαχιστοποιεί την αξία σε κίνδυνο (Value-at-Risk) ή εναλλακτικά την υπό συνθήκη αξία σε κίνδυνο (Conditional VaR), καθώς μία θέση που συνεπάγεται μεγάλη αξία σε κίνδυνο είναι ιδιαίτερα ευάλωτη σε σημαντικές μεταβολές στην αγορά, όπως για παράδειγμα μία απροσδόκητη αύξηση των επιτοκίων.

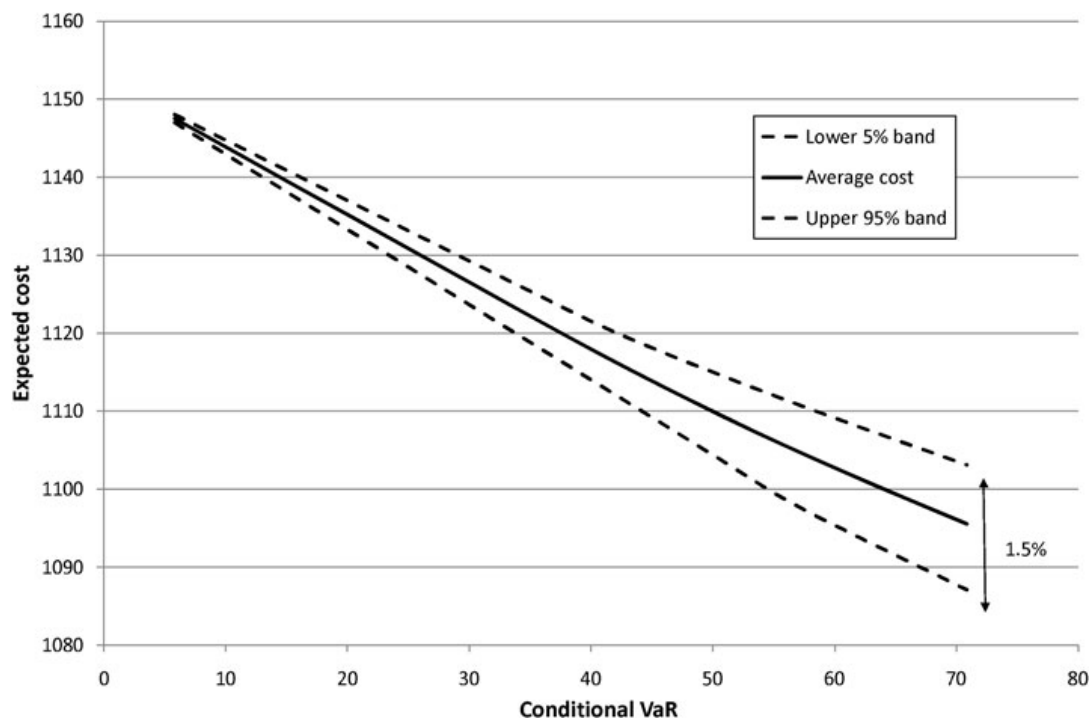
Τα ευρήματα των Consiglio και Staino (2012) καταδεικνύουν ότι τα υποδείγματα στοχαστικού προγραμματισμού μπορούν να αποτελέσουν πολύτιμα εργαλεία για τη βελτιστοποίηση της έκδοσης κρατικού χρέους. Για την πραγματοποίηση των προσομοιώσεων της καμπύλης των επιτοκίων χρησιμοποίησαν το υπόδειγμα των Bernaschi et al. (2007), το οποίο αποτελεί ένα πολυμεταβλητό υπόδειγμα επιστροφής στον μέσο και ως δεδομένα εισόδου την καμπύλη των επιτοκίων, το επίσημο επιτόκιο αναφοράς της Ευρωπαϊκής Κεντρικής Τράπεζας και τον

πληθωρισμό. Η εκτίμηση των παραμέτρων του υποδείγματος έγινε με τη χρονοσειρά των επιτοκίων των swap για τη χρονική περίοδο από τον Ιανουάριο του 1999 μέχρι τον Δεκέμβριο του 2007. Με βάση την εκτίμηση των τιμών των παραμέτρων του υποδείγματος, δημιούργησαν ένα σύνολο σεναρίων διάρκειας 12 μηνών. Τα σχετικά δεδομένα σε κάθε κόμβο του δέντρου είναι η καμπύλη των επιτοκίων, το επιτόκιο αναφοράς και ο πληθωρισμός και τα σενάρια του επόμενου σταδίου προσομοιώνονται με αρχικές τιμές τα δεδομένα για τον αντίστοιχο μήνα, με τη διαδικασία να επαναλαμβάνεται για κάθε σενάριο και κάθε στάδιο $t_h \in T$.

Ο έλεγχος του υποδείγματος έγινε με δέντρα σεναρίων 2 και 3 σταδίων, δηλαδή με ορίζοντα 2 και 3 ετών, αντίστοιχα. Για την εκτίμηση της σταθερότητας της βέλτιστης λύσης δημιούργησαν το βέλτιστο όριο με 100 διαφορετικές αρχικές εισροές της εκτιμήτριας και τελικό αριθμό κόμβων της τάξης των 500, 1000, 1700 και 3000 κόμβων. Το σύνολο των διαθέσιμων εργαλείων αποτελείται από 9 χρεόγραφα, όπως ομόλογα μηδενικού κουπονιού διάρκειας 3, 6, 12 και 24 μηνών και ομόλογα σταθερού κουπονιού διάρκειας 3, 5, 10, 15 και 30 ετών.

Στο παρακάτω διάγραμμα παρουσιάζεται το μέσο όριο (average frontier), το οποίο δημιουργήθηκε με τον μέσο όρο των 100 αρχικών εισροών που είχαν ως αποτέλεσμα το χαμηλότερο κόστος, και τα διαστήματα εμπιστοσύνης 95%. Με 1000 τελικούς κόμβους το μεγαλύτερο εύρος του διαστήματος εμπιστοσύνης φτάνει στο 1,5%, ενώ για δέντρα με πάνω από 1000 τελικούς κόμβους παρατηρείται περαιτέρω μείωση.

Διάγραμμα 4.1. Το μέσο όριο (average frontier) του υποδείγματος των Consiglio και Staino (2012) με τον μέσο όρο των 100 αρχικών εισροών που είχαν ως αποτέλεσμα το χαμηλότερο κόστος, και τα διαστήματα εμπιστοσύνης 95%



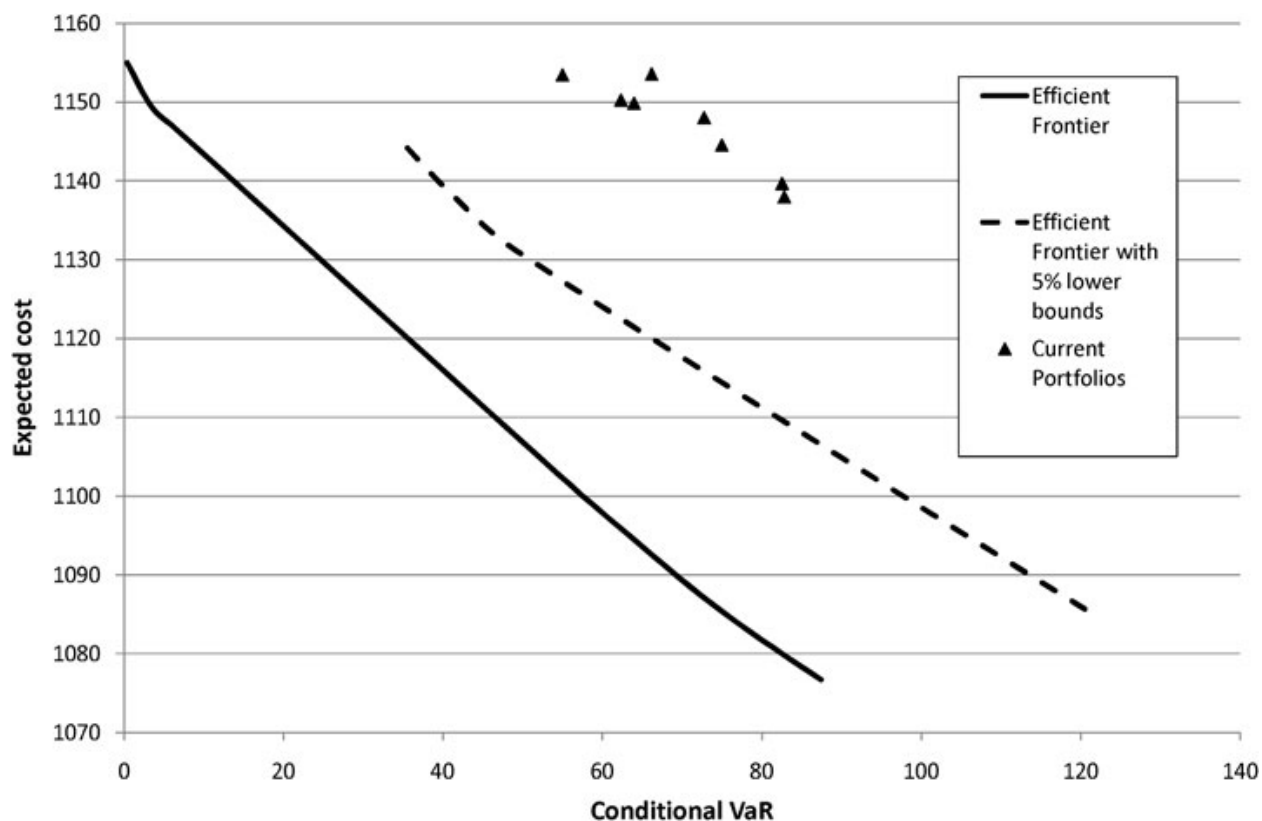
Πηγή: Consiglio και Staino (2012).

Το όριο αποτελεί φθίνουσα συνάρτηση της υπό συνθήκη αξίας σε κίνδυνο (Conditional Value-at-Risk) και είναι κυρτό. Αυτό σημαίνει ότι το αναμενόμενο κόστος ενός χαρτοφυλακίου σταθερού εισοδήματος μειώνεται με την έκδοση μακροχρόνιων ομολόγων, των οποίων η ωρίμανση λαμβάνει χώρα μετά το τέλος του ορίζοντα σχεδιασμού. Κατ' αυτό τον τρόπο, το μόνο κόστος που θα πρέπει να ληφθεί υπόψη είναι το κόστος πληρωμής των κουπονιών, καθώς το κεφάλαιο θα επιστραφεί μετά το τέλος του ορίζοντα σχεδιασμού. Η έκδοση μακροχρόνιων ομολόγων συνεπάγεται αύξηση της μεταβλητότητας του χαρτοφυλακίου, συνεπώς όσο μεγαλύτερη είναι η επιτρεπόμενη αύξηση της μεταβλητότητας τόσο μικρότερο είναι το κόστος που μπορεί να επιτευχθεί.

Στο παρακάτω σχήμα απεικονίζεται το πώς μεταβάλλεται η σύνθεση του χαρτοφυλακίου για διαφορετικές τιμές της υπό συνθήκη αξίας σε κίνδυνο. Καθώς μετακινούμαστε προς μεγαλύτερα επίπεδα της υπό συνθήκη αξίας σε κίνδυνο, η κατανομή μεταξύ των διαφόρων χρεογράφων μεταβάλλεται από χρεόγραφα με μεγαλύτερη μεταβλητότητα, όπως μακροχρόνια ομόλογα, σε χρεόγραφα χαμηλότερου κινδύνου, όπως τα βραχυχρόνια ομόλογα. Η σύνθεση του χρέους μεταξύ των χρεογράφων για το χαρτοφυλάκιο με το χαμηλότερο κόστος αποτελείται κατά κύριο λόγο από

ομόλογα τριετούς διάρκειας. Η επιλογή αυτή είναι η βέλτιστη με βάση το γεγονός ότι ο ορίζοντας σχεδιασμού του υποδείγματος είναι επίσης τριετής και το χαρτοφυλάκιο που ελαχιστοποιεί τον κίνδυνο αποτελείται από χρεόγραφα με ωρίμανση αντίστοιχη του ορίζοντα σχεδιασμού.

Διάγραμμα 4.2. Πώς μεταβάλλεται η σύνθεση του χαρτοφυλακίου για διαφορετικές τιμές της υπό συνθήκη αξία σε κίνδυνο στο υποδείγμα των Consiglio και Staino (2012)



Πηγή: Consiglio και Staino (2012)

Οι Consiglio και Staino (2012) εξετάζουν και πρακτικά τη χρησιμότητα του υποδείγματος συγκρίνοντας το αποτελεσματικό όριο του υποδείγματος τους με πραγματικά χαρτοφυλάκια που επέλεξε το Υπουργείο Οικονομικών της Ιταλίας. Στο παρακάτω διάγραμμα παρουσιάζεται το βέλτιστο όριο και 9 χαρτοφυλάκια αναφοράς του Ιταλικού Υπουργείου Οικονομικών πριν από κάθε έκδοση. Η απόδοση των χαρτοφυλακίων μετράται με βάση την υπό συνθήκη αξία σε κίνδυνο και το αναμενόμενο κόστος με τη χρήση των ίδιων δέντρων σεναρίων που χρησιμοποιήθηκαν για τη δημιουργία του βέλτιστου ορίου. Με βάση τα ευρήματα της έρευνας, τα χαρτοφυλάκια αναφοράς του Ιταλικού Υπουργείου Οικονομικών είναι αναποτελεσματικά συγκριτικά με τα βέλτιστα χαρτοφυλάκια. Καθώς η ελαχιστοποίηση των απωλειών με βάση την υπό συνθήκη αξία σε κίνδυνο αποτελεί έναν αποτελεσματικό στόχο αναφορικά με την διαχείριση του χαρτοφυλακίου, οι Consiglio και Staino (2012) αναφέρουν πως η μεθοδολογία τους μπορεί να αποτελέσει ένα ιδιαίτερα χρήσιμο εργαλείο για τη σύγκριση χαρτοφυλακίων. Η κυριαρχία των

βέλτιστων χαρτοφυλακίων συγκριτικά με τα χαρτοφυλάκια που επέλεξε το Υπουργείο Οικονομικών της Ιταλίας επιβεβαιώνεται και με τη χρήση σεναρίων εκτός δείγματος. Τα αποτελέσματα καταδεικνύουν ότι τα χαρτοφυλάκια που δημιουργούνται με τη συγκεκριμένη μεθοδολογία είναι αποτελεσματικότερα συγκριτικά με τα χαρτοφυλάκια αναφοράς του Υπουργείου Οικονομικών της Ιταλίας.

Οι μεγάλες επιχειρήσεις χρηματοδοτούν τα κεφαλαιακά και λειτουργικά τους έξοδα μέσω της έκδοσης ομολόγων διαφόρων ονομαστικών αξιών, δόσεων και λήξεων. Οι Valladão et al. (2014) προτείνουν ένα υπόδειγμα γραμμικού στοχαστικού προγραμματισμού πολλαπλών σταδίων για τη βελτιστοποίηση της έκδοσης ομολόγων μέσω της ελαχιστοποίησης του μέσου κόστους χρηματοδότησης, διατηρώντας τη μόχλευση υπό έλεγχο και τον κίνδυνο χρεοκοπίας σε αποδεκτά επίπεδα.

Ο στοχαστικός προγραμματισμός πολλαπλών σταδίων προσφέρει ένα κατάλληλο πλαίσιο για την εξέταση προβλημάτων μακροχρόνιου χρηματοοικονομικού σχεδιασμού και ιδιαίτερα της διαχείρισης του εταιρικού χρέους. Ένα αντίστοιχο υπόδειγμα περιγράφει ένα δυναμικό πλαίσιο, όπου σε κάθε στάδιο απόφασης η μελλοντική εξέλιξη είναι άγνωστη. Μετά την εφαρμογή των αποφάσεων του συγκεκριμένου σταδίου, αποκαλύπτονται οι πληροφορίες του επόμενου σταδίου και η διαδικασία επαναλαμβάνεται για το επόμενο στάδιο.

Συνήθως, στα υποδείγματα στοχαστικού προγραμματισμού πολλαπλών σταδίων η αβεβαιότητα αποτυπώνεται με ένα διακριτό δέντρο αποφάσεων, με τους κόμβους του δέντρου να καταδεικνύουν την κατάσταση της διαδικασίας στα σημεία αποφάσεων και την αβεβαιότητα πριν το επόμενο στάδιο. Ένα δέντρο γεγονότων μπορεί να θεωρηθεί ως ένας χώρος πιθανοτήτων που δημιουργεί ένα πιθανοτικό ισοδύναμο του υποδείγματος στοχαστικού προγραμματισμού, με μία πλήρη διαδρομή του δέντρου γεγονότων να αποτελεί ένα σενάριο και μία πολιτική να ορίζεται ως το σύνολο των αποφάσεων για όλα τα στάδια και όλα τα σενάρια. Η δομή αυτή απαιτεί όλες οι αποφάσεις να λαμβάνονται αποκλειστικά με βάση τις διαθέσιμες πληροφορίες των προηγούμενων σταδίων, προϋπόθεση που επιβάλλει ότι οι μεταβλητές αποφάσεων σε ένα δεδομένο στάδιο θα πρέπει να είναι ίσες, εάν τα σενάρια τους μοιράζονται τον ίδιο κόμβο στο δέντρο γεγονότων (Valladão et al., 2014).

Όπως σημειώνουν οι Valladão et al. (2014), το μέγεθος του υποδείγματος αυξάνεται εκθετικά καθώς αυξάνεται ο αριθμός των σταδίων. Για τη μείωση των διαστάσεων προτείνουν μία υβριδική προσέγγιση, που αποτελείται από ένα παραδοσιακό υπόδειγμα πολλαπλών σταδίων για τις πρώτες T^* περιόδους και ένα ανεξάρτητο σενάριο με έναν απλό κανόνα δεδομένης πολιτικής για $t < T^*$. Η αβεβαιότητα αναπαρίσταται από ανεξάρτητα σενάρια που δημιουργούνται από ένα δείγμα του συνολικού δέντρου γεγονότων. Στο υπόδειγμα τους, για $t \leq T^*$ εξετάζεται ένα πλήρες σύνολο χρεογράφων, ενώ για $t > T^*$ περιλαμβάνονται μόνο βραχυχρόνια ομόλογα ώστε να διασφαλιστεί το όριο ελάχιστων ταμειακών διαθεσίμων. Το υπόδειγμα υιοθετεί μία τέτοια δομή, καθώς όπως αναφέρουν οι Valladão et al. (2014) στην πράξη οι περισσότερες επενδύσεις λαμβάνουν χώρα στα αρχικά στάδια του ορίζονται προγραμματισμού.

Η αντικειμενική συνάρτηση του υποδείγματος περιλαμβάνει δύο συνιστώσες. Η πρώτη καταγράφει τη σχέση μεταξύ τελικών ταμειακών διαθεσίμων και του κινδύνου χρεοκοπίας στο τέλος του ορίζοντα, με βάση μία συνάρτηση χρησιμότητας των τελικών ταμειακών διαθεσίμων. Καθώς το σύνολο του κόστους δανεισμού αντικατοπτρίζεται στα ταμειακά διαθέσιμα, η μεγιστοποίηση των τελικών ταμειακών διαθεσίμων ισοδυναμεί με την ελαχιστοποίηση του μελλοντικού κόστους εξυπηρέτησης του χρέους. Η δεύτερη συνιστώσα της αντικειμενικής συνάρτησης λαμβάνει υπόψη της ότι η φερεγγυότητα της επιχείρησης βασίζεται σε δείκτες χρηματοοικονομικής απόδοσης οι οποίοι είναι γνωστοί στην αγορά. Όπως σημειώνουν οι Valladão et al. (2014), ιδανικά το υπόδειγμα θα έπρεπε να περιλαμβάνει την προσαρμογή των επιτοκίων βάσει της πιστοληπτικής ικανότητας της επιχείρησης, αλλά κάτι τέτοιο δεν είναι εφικτό, καθώς θα αυξάνονταν ιδιαίτερα η πολυπλοκότητα του υποδείγματος, αλλά και ανάλογοι υπολογισμοί δεν είναι εφικτή με τα διαθέσιμα δεδομένα. Προτείνουν, εναλλακτικά, μία συνάρτηση ποινής που αποθαρρύνει τη συσσώρευση υπερβολικής μόχλευσης στα ενδιάμεσα στάδια του ορίζοντα προγραμματισμού.

Η συνάρτηση χρησιμότητας $U(C_T)$ αποδίδει μία τιμή σε ένα σενάριο στο τέλος του ορίζοντα προγραμματισμού βάσει των τελικών ταμειακών διαθεσίμων.

$$U(C_T) = C_T^+ - pC_T^-$$

Μία αρνητική τιμή τελικών ταμειακών διαθεσίμων τιμωρείται από την παράμετρο αποστροφής του κινδύνου $p \geq 1$. Στη συνάρτηση χρησιμότητας $U(C_T)$ ο συντελεστής p αποτελεί έναν

συντελεστή αποστροφής κινδύνου, με $p = 1$ να υποδηλώνει ουδετερότητα ως προς τον κίνδυνο. Η αναμενόμενη τιμή της συνάρτησης χρησιμότητας είναι:

$$E[U(C_T)] = E[C_T^+] - pE[C_T^-]$$

συνδυάζοντας τα προσδοκώμενα τελικά ταμειακά διαθέσιμα με την προσδοκώμενη ποινή της χρεοκοπίας.

Το δεύτερο σκέλος της αντικειμενικής συνάρτησης αφορά την επίδραση της άποψης της αγοράς αναφορικά με την ομολογιακή πολιτική μίας επιχείρησης. Το υπόδειγμα των Valladão et al. (2014) περιλαμβάνει στην αντικειμενική συνάρτηση μία ποινή για υπερβολική μόχλευση. Προτείνουν μία συνάρτηση που τιμωρεί όλο και περισσότερο την υπερβολική μόχλευση βάσει μίας αλληλουχίας ορίων για τον λόγο του χρέους προς τα περιουσιακά στοιχεία. Οι τιμές αυτές $\gamma_1 \leq \dots \leq \gamma_K$ ανταποκρίνονται σε κρίσιμα επίπεδα μόχλευσης που ορίζονται από τους διαχειριστές του χρέους.

Στην αντικειμενική συνάρτηση του υποδείγματος επιβάλλεται μία αθροιστική ποινή για την παραβίαση κάθε ενός από αυτούς τους στόχους μόχλευσης σε κάθε σενάριο και σε κάθε χρονική περίοδο. Αρχικά, ορίζεται το ποσό της υπερβολικής μόχλευσης που ξεπερνά τον κάθε στόχο:

$$I_{t,k}(s) = [D_t(s) - \gamma_k A_t(s)]^+ = \max(0, D_t(s) - \gamma_k A_t(s)), \quad \forall t \in H, s \in S, k \in K$$

Η τελευταία έκφραση στο υπόδειγμα γραμμικού προγραμματισμού εισάγεται προσθέτοντας αρχικά ως περιορισμούς τις ακόλουθες ανισότητες:

$$I_{t,k}(s) \geq 0, I_{t,k}(s) \geq D_t(s) - \gamma_k A_t(s), \quad \forall t \in H, s \in S, k \in K$$

Η ισότητα στον ορισμό των μεταβλητών $I_{t,k}(s)$ ισχύει μόνο στην περίπτωση της βέλτιστης λύσης, καθώς το υπόδειγμα περιλαμβάνει ποινές στην αντικειμενική συνάρτηση για υπερβολική μόχλευση.

$$\theta_k I_{t,k}(s) \quad \forall t \in H, s \in S, k \in K$$

Όπου $\theta_1 \leq \dots \leq \theta_K$ είναι ποινές που επίσης ορίζονται από τους διαχειριστές του χρέους.

Η συνολική ποινή υπερβολικής μόχλευσης αθροίζει τις μελλοντικές τιμές των παραβιάσεων κάθε στόχου σε όλα τα σενάρια:

$$\sum_{k \in K} \theta_k \sum_{t \in H} I_{t,k}(s) \prod_{\tau=t+1}^T (1 + \rho_{\tau}(s)) \quad \forall s \in S$$

Στην τελική μορφή της αντικειμενικής συνάρτησης προστίθενται οι δύο συνιστώσες και υπολογίζεται η προσδοκώμενη τιμή:

$$\max S^{-1} \sum_{s \in S} (C_T^+(s) - p C_T^-(s) - \sum_{k \in K} \theta_k \sum_{t \in H} I_{t,k}(s) \prod_{\tau=t+1}^T (1 + \rho_{\tau}(s)))$$

Οι συναρτήσεις ποινών λαμβάνουν αρνητικά πρόσημα στη μεγιστοποίηση της αντικειμενικής συνάρτησης.

Η εμπειρική εφαρμογή του υποδείγματος των Valladão et al. (2014) έγινε με τη χρήση λογισμικού χρηματοοικονομικού σχεδιασμού στην πετρελαϊκή εταιρεία της Βραζιλίας Petrobras. Οι περιορισμοί του υποδείγματος αφορούν σενάρια για τα επιτόκια και τα ασφάλιστρα κινδύνου, τα οποία αναφέρονται ως Παράγοντες Χρηματοοικονομικού Κινδύνου. Μία πρόσθετη κατηγορία κινδύνων, οι Παράγοντες Κινδύνου του Έργου, εισάγονται έμμεσα, μέσω των σεναρίων για τις ταμειακές ροές που δημιουργούνται από το χαρτοφυλάκιο του έργου, συμπεριλαμβανομένων των τιμών για το αργό πετρέλαιο και το φυσικό αέριο. Χρησιμοποιούνται οι μέσες τιμές της κάθε περιόδου ώστε να αντικατοπτρίζουν τη συνεχή ροή εσόδων και εξόδων σε κάθε στάδιο σχεδιασμού.

Το υπόδειγμα πρόβλεψης παρέχει σενάρια για την εξέλιξη των επιτοκίων και του συναλλάγματος, με:

$r_{t,k}(s)$: Η ετήσια αποτελεσματική απόδοση για εταιρικά ομόλογα σε Αμερικανικά δολάρια.

$\rho_t(s)$: Το ασφαλές επιτόκιο, το οποίο αποτελούν τα Αμερικανικά ομόλογα ενός έτους.

$\varphi_{t,k}(s)$: Το ασφάλιστρο κινδύνου για ομόλογα κυμαινόμενου επιτοκίου με πλήρη απόσβεση στη λήξη.

Υπολογίζεται το κουπόνι για το ομόλογο σταθερού επιτοκίου $i \in X$ που εκδόθηκε τον χρόνο $t \in H$ και το σενάριο $s \in S$, $a_t^i(s)$ και το ασφάλιστρο κινδύνου για το ομόλογο κυμαινόμενου επιτοκίου που εκδόθηκε τον χρόνο $t \in H$ και το σενάριο $s \in S$, $\psi_t^i(s)$, με βάση την καμπύλη των επιτοκίων $r_{t,k}(s)$ και τα ασφάλιστρα κινδύνου $\varphi_{t,k}(s)$ για κάθε $t \in H, k \in H, s \in S$.

Πέρα από τις τιμές του αργού πετρελαίου και των υποπροϊόντων του, οι συναλλαγματικές ισοτιμίες θεωρούνται επίσης παράγοντες κινδύνου έργου, καθώς το χαρτοφυλάκιο στο πλαίσιο της εμπειρικής εφαρμογής περιλαμβάνει και επενδύσεις σε διάφορα νομίσματα. Με βάση τους συγκεκριμένους παράγοντες κινδύνου και στοιχεία για τις επενδύσεις και την παραγωγή, ένας προεπεξεργαστής του υποδείγματος βελτιστοποίησης υπολογίζει σενάρια για τις χρηματικές ροές $f_t(s), d_t(s)$ και $I_t(s)$.

Το λογισμικό χρηματοοικονομικού σχεδιασμού για τον υπολογισμό του υποδείγματος χρησιμοποιεί το Matlab για την προετοιμασία των δεδομένων και την παρουσίαση των λύσεων. Το υπόδειγμα γραμμικού προγραμματισμού υλοποιήθηκε με τη γλώσσα μοντελοποίησης MOSEL, με τη χρήση της σουίτας βελτιστοποίησης Xpress για τη λύση. Οι παράμετροι του υποδείγματος πρόβλεψης επιλέχθηκαν ώστε να προσομοιώνουν ένα περιβάλλον αυξανόμενων επιτοκίων.

Στην εμπειρική εφαρμογή, οι Valladão et al. (2014) εξετάζουν ένα σενάριο βάσης με ορίζοντα 48 μηνών, ξεκινώντας το 2010, με βραχυπρόθεσμο ορίζοντα σχεδιασμού τους 6 μήνες, αρχικά ταμειακά διαθέσιμα $c = 5$, ελάχιστα ταμειακά διαθέσιμα $\hat{c}_t = 5$, παράμετρο αποστροφής κινδύνου $p = 10$, σταθμισμένο μέσο κόστος κεφαλαίου $\omega = 8,8$, αριθμό στόχων μόχλευσης $K = 3$ και αριθμό σεναρίων $S = 1024$. Το ντετερμινιστικό γραμμικό υπόδειγμα έχει 82.534 γραμμές, 813.056 στήλες και 4.559.468 μη μηδενικά στοιχεία πίνακα.

Συγκρίνουν τις λύσεις του προβλήματος με βάση δύο υποθέσεις για τις ποινές υπερβολικής μόχλευσης. Στην πρώτη περίπτωση υποθέτουν μηδενικές ποινές, ενώ στη δεύτερη περίπτωση, στόχους μόχλευσης $\gamma = \{35\%, 50\%, 100\%\}$ και αυξανόμενες ποινές μόχλευσης $\theta = \{1, 1, 1\}$. Αρχικά συγκρίνουν τις εκδόσεις ομολόγων για τον ορίζοντα σχεδιασμού. Στην πρώτη περίπτωση, το συνολικό χρέος που εκδίδεται είναι υπερβολικά υψηλότερο από το ποσό που απαιτείται για την επίτευξη των ελάχιστων ταμειακών διαθεσίμων που απαιτούνται από την εταιρεία. Το αποτέλεσμα αυτό δεν υφίσταται στη δεύτερη περίπτωση, λόγω των ποινών υπερβολικής μόχλευσης.

Όπως σημειώνουν οι Valladão et al. (2014), η λύση στην πρώτη περίπτωση αντίκειται στην διαισθητική λογική ότι μία εταιρεία με ένα σταθερό χαρτοφυλάκιο έργου δεν θα πρέπει να βρίσκεται εκτεθειμένη χωρίς λόγο σε κίνδυνο από αβέβαιες χρηματοοικονομικές αποδόσεις. Στη δεύτερη περίπτωση, οι ενδιάμεσες ποινές υπερβολικής μόχλευσης αποτρέπουν επικίνδυνες πολιτικές με μεγάλη μόχλευση όπως συμβαίνει στην περίπτωση ενός μικρού αριθμού σεναρίων. Το τελευταίο εύρος μόχλευσης, $\frac{D_t(s)}{A_t(s)} \geq 100\%$ ορίζει την κατάσταση χρεοκοπίας σε κάθε στάδιο, όταν το χρέος ξεπερνά τα συνολικά περιουσιακά στοιχεία. Τα αποτελέσματα της έρευνας καταδεικνύουν ότι οι διαχειριστές του χρέους θα πρέπει να χρησιμοποιούν το υπόδειγμα καθορίζοντας κατάλληλα τις παραμέτρους αποστροφής του κινδύνου και τις ενδιάμεσες ποινές υπερβολικής μόχλευσης, προκειμένου να αποφύγουν τη δημιουργία χαρτοφυλακίων με υπερβολική μόχλευση.

Οι Valladão et al. (2014) καταλήγουν πως το υπόδειγμα τους αποτελεί ένα υπολογιστικά διαχειρίσιμο, διαισθητικό εργαλείο που βασίζεται στην πρακτική εμπειρία. Σημειώνουν ότι παρά την πρόοδο της πληροφορικής τεχνολογίας και την αύξηση της υπολογιστικής ισχύος, ο υπολογισμός ενός ολοκληρωμένου υποδείγματος βελτιστοποίησης παραμένει από ιδιαίτερα δύσκολος έως αδύνατος και στην πράξη υπολογίζονται απλοποιημένα υποδείγματα που προσεγγίζουν τη βέλτιστη λύση. Το υπόδειγμα που προτείνουν αποτελεί μία υβριδική, ενδιάμεση προσέγγιση, όπου για τα αρχικά στάδια χρησιμοποιείται ένα πλήρες δέντρο γεγονότων, ενώ τα υπόλοιπα στάδια περιγράφονται από επιμέρους δείγματα που προσεγγίζονται από ανεξάρτητες διαδρομές.

Καταλήγουν πως το υπόδειγμα που προτείνουν αποτελεί μία αποτελεσματική προσέγγιση όταν τα περισσότερα έξοδα λαμβάνουν χώρα στα αρχικά στάδια του επενδυτικού σχεδιασμού, όπως συνήθως συμβαίνει στην πράξη. Σημειώνουν πως το υπόδειγμα τους προσφέρει μια πιο αξιόπιστη προσέγγιση συγκριτικά με άλλα υποδείγματα, καθώς εξετάζει ένα πλήρες πρόγραμμα πληρωμών ομολόγων, με τη βέλτιστη λύση να αποτελείται από έναν συνδυασμό ομολόγων με διαφορετικές λήξεις, αποτιμήσεις σε νομίσματα και πρόγραμμα αποπληρωμών.

5. ΕΝΤΟΠΙΣΜΟΣ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗΣ ΑΠΑΤΗΣ

Το ζήτημα των παραποιημένων χρηματοοικονομικών καταστάσεων αποτελεί ένα ιδιαίτερα σημαντικό ζήτημα που επηρεάζει τις οικονομίες πολλών χωρών. Σύμφωνα με εκτιμήσεις, η παραποίηση των χρηματοοικονομικών καταστάσεων μεγάλων εταιρειών, όπως η Enron και η Xerox, οδήγησε σε μεγάλης έκτασης ζημιές για επενδυτές και πιστωτές. Ο εντοπισμός της παραποίησης των χρηματοοικονομικών καταστάσεων με τη χρήση των τυπικών διαδικασιών του οικονομικού ελέγχου είναι μία ιδιαίτερα δύσκολη υπόθεση (Porter & Cameron, 1987). Η χρήση όμως υποδειγμάτων κατάταξης επιτρέπει στους χρηματοοικονομικούς ελεγκτές τον ταυτόχρονο έλεγχο πολλών επιχειρήσεων, δίνοντας τους τη δυνατότητα να εστιάσουν στη συνέχεια να εστιάσουν σε εκείνες τις επιχειρήσεις για τις οποίες υπάρχουν ενδείξεις ότι μπορεί να εντοπιστούν προβλήματα στις χρηματοοικονομικές τους καταστάσεις (Pasiouras et al., 2007).

Όπως σημειώνουν οι Calderon και Cheh (2002) η χρήση της τεχνολογίας είναι απαραίτητη για τον εντοπισμό κρίσιμων ζητημάτων, όπως είναι η απάτη, που θα μπορούσαν να έχουν αρνητικό αντίκτυπο στην ποιότητα και την ακρίβεια του οικονομικού ελέγχου αν δεν γίνουν αντιληπτα και δεν ανιχνευθούν. Τα υποδείγματα κατάταξης με τη χρήση χρηματοοικονομικών μεταβλητών μπορούν να αποτελέσουν χρήσιμα εργαλεία για τους οικονομικούς ελεγκτές κατά τον οικονομικό έλεγχο επιχειρήσεων, για τον καθορισμό του εύρους του ελέγχου, τη διασφάλιση της ποιότητας του ελέγχου και για την αντιμετώπιση των δυσκολιών που συνεπάγεται ο έλεγχος μεγάλου όγκου δεδομένων (Ramamoorti et al., 1999).

Ο Gaganis (2009) συγκρίνει 10 μεθοδολογίες κατάταξης για τον εντοπισμό των παραποιημένων χρηματοοικονομικών καταστάσεων με τη χρήση 398 χρηματοοικονομικών καταστάσεων ελληνικών επιχειρήσεων, από διάφορους τομείς οικονομικής δραστηριότητας επιχειρήσεων εισηγμένων στο Χρηματιστήριο Αξιών Αθηνών και μη. Αρχικά, πραγματοποίησε τη συλλογή 199 χρηματοοικονομικών καταστάσεων επιχειρήσεων που είχαν παραποιημένες χρηματοοικονομικές καταστάσεις κατά την περίοδο μεταξύ 2001 και 2004 και στη συνέχεια προχώρησε στην αντιστοίχιση τους με μία μη παραποιημένη χρηματοοικονομική κατάσταση από μία επιχείρηση του αντίστοιχου οικονομικού τομέα και αντίστοιχου μεγέθους.

Στη συνέχεια προχώρησε στη συλλογή δεδομένων για κάθε μία από τις 398 επιχειρήσεις του δείγματος, που αφορούν 28 χρηματοοικονομικές μεταβλητές που καλύπτουν το σύνολο των διαστάσεων της απόδοσης μίας επιχείρησης, όπως η ρευστότητα, η κερδοφορία, η μόχλευση, η δραστηριότητα της διοίκησης και ετήσιες μεταβολές σε βασικούς της λογαριασμούς, καθώς και στην εξέταση των συνιστωσών των μεταβλητών. Μέσω αυτής της προσέγγισης μπορεί να αποκαλυφθεί η λανθάνουσα δομή ενός συνόλου μεταβλητών μέσω της μείωσης του διαστήματος των ιδιοτήτων, από έναν αριθμό μεταβλητών σε έναν μικρότερο αριθμό παραγόντων.

Η ανάλυση παραγόντων χρησιμοποιήθηκε για τη μείωση των επιπτώσεων της πολυσυγγραμμικότητας και της επικάλυψης των μεταβλητών και εντοπίστηκαν 7 παράγοντες που εξηγούν το 76,53% της διακύμανσης των δεδομένων. Οι λόγοι των εσόδων προς τις πωλήσεις, του κυκλοφορούντος ενεργητικού προς τις βραχυπρόθεσμες υποχρεώσεις, του κυκλοφορούντος ενεργητικού προς το συνολικό ενεργητικό, των μετρητών προς το συνολικό ενεργητικό, των κερδών προ φόρων προς το συνολικό ενεργητικό, των αποθεμάτων προς το συνολικό ενεργητικό και η ετήσια μεταβολή των πωλήσεων χρησιμοποιήθηκαν για την αντιπροσώπευση των 7 παραγόντων και ως εισροές των υποδειγμάτων κατάταξης. Πέρα από τις 7 αυτές χρηματοοικονομικές μεταβλητές, χρησιμοποίησε και 6 πρόσθετες μεταβλητές, όπως η οικονομική δυσπραγία, το αν βρίσκεται η επιχείρηση υπό δικαστική διαδικασία, η εταιρεία που πραγματοποιεί τον οικονομικό έλεγχο, η αλλαγή οικονομικών ελεγκτών, τα κέρδη ή η ζημία τη χρονιά του ελέγχου και εάν μία επιχείρηση είναι εισηγμένη στο χρηματιστήριο ή όχι.

Η εκτίμηση των υποδειγμάτων κατάταξης έγινε με τη χρήση των 7 αρχικών μεταβλητών και στη συνέχεια με τη χρήση και των 6 πρόσθετων μεταβλητών. Η μέση ακρίβεια των μεθοδολογιών με τη χρήση μόνο των 7 αρχικών μεταβλητών κυμαίνεται από 76,83%, για τη μεθοδολογία της παλινδρόμησης με διανύσματα υποστήριξης (support vector regression) και 87,20%, για τη μεθοδολογία UTADIS (Utilités Additives Discriminantes) και τα πιθανοτικά νευρωνικά δίκτυα. Η μέση ακρίβεια των μεθοδολογιών και με τις πρόσθετες 6 μεταβλητές διαμορφώνεται από 84,15% για την περίπτωση της λογιστικής παλινδρόμησης και 90,24% για το UTADIS και τα πιθανοτικά νευρωνικά δίκτυα. Το υπόδειγμα UTADIS σημειώνει τις καλύτερες επιδόσεις και για τα δύο υποδείγματα, δηλαδή και για το υπόδειγμα με τις 7 αρχικές μεταβλητές αλλά και για το υπόδειγμα με τις 6 πρόσθετες μεταβλητές, αλλά στην περίπτωση του δεύτερου υποδείγματος τα πιθανοτικά νευρωνικά δίκτυα επιτυγχάνουν την ίδια ακρίβεια.

Στον πίνακα που ακολουθεί παρουσιάζεται η ακρίβεια των διαφόρων μεθοδολογιών κατάταξης με βάση τα δύο υποδείγματα για το δείγμα επαλήθευσης.

Πίνακας 5.1. Η ακρίβεια των διαφόρων μεθοδολογιών κατάταξης για τον εντοπισμό παραποιημένων χρηματοοικονομικών καταστάσεων με βάση τα δύο υποδείγματα της έρευνας του Gaganis (2009) για το δείγμα επαλήθευσης

Μεθοδολογία	Ακρίβεια					
	Υπόδειγμα 7 χρηματοοικονομικών μεταβλητών			Υπόδειγμα με 6 επιπλέον μεταβλητές		
	Χρηματοοικο νομικές αναφορές χωρίς προβλήματα	Χρηματοοικο νομικές αναφορές με πιθανά προβλήματα	Μέσος όρος	Χρηματοοικο νομικές αναφορές χωρίς προβλήματα	Χρηματοοικο νομικές αναφορές με πιθανά προβλήματα	Μέσος όρος
UTADIS	86,59	87,81	87,20	89,02	91,46	90,24
Πολυμοαδική ιεραρχική διάκριση (MHDIS)	80,49	87,80	84,15	84,15	91,46	90,24
Κ-εγγύτεροι γείτονες	78,05	85,37	81,71	95,12	75,61	85,37
Πιθανοτικά νευρωνικά δίκτυα	78,05	87,81	82,93	91,46	89,02	90,24
Τεχνητά νευρωνικά δίκτυα	78,05	82,93	80,49	84,15	92,68	88,42
Μηχανήματα διανυσμάτων υποστήριξης συνάρτησης ακτινικής βάσης (radial basis function	82,93	79,27	81,10	86,59	86,59	86,59

support vector machines)						
Πολυωνυμικά Μηχανήματα διανυσμάτων υποστήριξης (Polynomial SVM)	75,61	78,05	76,83	89,02	85,37	87,20
Γραμμικά Μηχανήματα διανυσμάτων υποστήριξης (Linear SVM)	84,15	79,27	81,71	93,90	82,93	88,42
Ανάλυση Διακρίνουσας	84,10	79,30	81,70	93,90	81,71	87,80
Λογιστική παλινδρόμηση	78	86,60	82,30	74,39	93,90	84,15

Πηγή: Gaganis (2009).

Παρατηρείται ότι η μεθοδολογία UTADIS παρουσιάζει τη μεγαλύτερη ακρίβεια και για τα δύο υποδείγματα, τόσο όσον αφορά τον εντοπισμό των χρηματοοικονομικών αναφορών για τις οποίες παρατηρούνται ζητήματα πιθανής παραποίησης, όσο και για τις χρηματοοικονομικές αναφορές για τις οποίες δεν προκύπτουν αντίστοιχα ζητήματα και δεν αμφισβητείται η αξιοπιστία τους. Ακολουθεί η μεθοδολογία πολυομαδικής ιεραρχικής διάκρισης (multi-group hierarchical discrimination - MHDIS), η μεθοδολογία k-εγγύτερων γειτόνων (k-nearest neighbours), τα πιθανοτικά νευρωνικά δίκτυα, τα τεχνητά νευρωνικά δίκτυα, τα μηχανήματα διανυσμάτων υποστήριξης συνάρτησης ακτινικής βάσης (radial basis function support vector machines), τα πολυωνυμικά μηχανήματα διανυσμάτων υποστήριξης, τα γραμμικά μηχανήματα διανυσμάτων υποστήριξης, η ανάλυση διακρίνουσας και η λογιστική παλινδρόμηση.

Όπως σημειώνει ο Gaganis (2009) είναι αξιοσημείωτο το γεγονός ότι η μεθοδολογία UTADIS σημειώνει την πιο ισορροπημένη επίδοση αναφορικά και με τους δύο πιθανούς τύπους σφαλμάτων κατάταξης, με τη διαφορά αναφορικά με την ακρίβεια κατάταξης ανάμεσα στους δύο τύπους να

διαμορφώνεται στο 1,22%, ενώ η αντίστοιχη διαφορά για τα πιθανοτικά νευρωνικά δίκτυα είναι σημαντικά μεγαλύτερη και ανέρχεται στο 9,76%.

Οι μεθοδολογίες UTADIS, πολυμοαδικής ιεραρχικής διάκρισης, η ανάλυση διακρίνουσας και η λογιστική παλινδρόμηση παρέχουν πληροφορίες και για το ποιές μεταβλητές είναι στατιστικά σημαντικές. Με βάση τα ευρήματα της έρευνας στατιστικά σημαντικές μεταβλητές είναι ο λόγος του κυκλοφορούντος ενεργητικού προς το συνολικό ενεργητικό και ο λόγος του κυκλοφορούντος ενεργητικού προς τις βραχυπρόθεσμες υποχρεώσεις, που βρίσκεται σε αντιστοιχία με τα ευρήματα ερευνών (Ireland, 2003) που αναφέρουν ότι είναι πιθανότερο να εντοπιστούν προβλήματα κατά τον οικονομικό έλεγχο για επιχειρήσεις που αντιμετωπίζουν προβλήματα ρευστότητας και υψηλό χρηματοοικονομικό κίνδυνο και αναφέρουν στις χρηματοοικονομικές τους καταστάσεις σημαντικές ενδεχόμενες υποχρεώσεις.

Αναφορικά με τη σημασία των 6 πρόσθετων μεταβλητών παρατηρείται ότι η συμπερίληψη τους σε ένα υπόδειγμα για τον εντοπισμό παραποιημένων χρηματοοικονομικών αναφορών βελτιώνει την ακρίβεια των εκτιμήσεων για όλες τις μεθοδολογίες, με τη μεθοδολογία των k-εγγύτερων γειτόνων να σημειώνει τη μεγαλύτερη ακρίβεια, με βάση το δεύτερο υπόδειγμα, για επιχειρήσεις με χρηματοοικονομικές αναφορές χωρίς προβλήματα και τη μεθοδολογία της λογιστικής παλινδρόμησης να παρουσιάζει τη μεγαλύτερη ακρίβεια για τις επιχειρήσεις με χρηματοοικονομικές αναφορές με πιθανά προβλήματα.

Ο Gaganis (2009) σημειώνει πως αν και ορισμένες μεθοδολογίες παρουσιάζουν μεγαλύτερη ακρίβεια κατάταξης, όλες οι μεθοδολογίες προσφέρουν χρήσιμες πληροφορίες, καθώς εξετάζουν το πρόβλημα κατάταξης με τη χρήση διαφορετικών μορφών μοντελοποίησης και αντικειμενικές συναρτήσεις. Με βάση τα αποτελέσματα της έρευνας, οι μεθοδολογίες απόφασης με τη χρήση πολλαπλών κριτηρίων παρουσιάζουν τη μεγαλύτερη ακρίβεια στην περίπτωση του πρώτου υποδείγματος, δηλαδή του υποδείγματος μόνο με τις 7 αρχικές μεταβλητές, και τη μεγαλύτερη και τρίτη μεγαλύτερη ακρίβεια στην περίπτωση του δεύτερου υποδείγματος, δηλαδή με τη χρήση και 6 επιπλέον μεταβλητών.

Οι Glancy και Yadav (2011) αναπτύσσουν ένα υπολογιστικό υπόδειγμα εντοπισμού της απάτης με τη χρήση υπολογιστικών μεθόδων σε δεδομένα κειμένου, που αξιοποιεί όλες τις διαθέσιμες πληροφορίες σε ένα κείμενο. Το υπόδειγμα επιδιώκει τον εντοπισμό τυχόν προσπάθειας αναφοράς

ψευδών ή ανακριβών πληροφοριών σε ετήσιες χρηματοοικονομικές αναφορές προς την Επιτροπή Κεφαλαιαγοράς των Η.Π.Α (US Securities and Exchange Commission) και μπορεί να αποτελέσει εργαλείο αυτοματισμού του εντοπισμού απάτης στις χρηματοοικονομικές αναφορές.

Όπως αναφέρουν, η ανάπτυξη ενός αποτελεσματικού εργαλείου θα μπορούσε να συμβάλλει στον έγκαιρο εντοπισμό των υποθέσεων απάτης και κατά συνέπεια και στην αποτροπή της απάτης. Η ανίχνευση των περιπτώσεων απάτης στο υπόδειγμα των Glancy και Yadav (2011) βασίζεται στην εξέταση των Αναφορών Ελέγχου (Accounting and Auditing Enforcement Release) της Επιτροπής Κεφαλαιαγοράς των Η.Π.Α, που αποτελεί μία διοικητική διαδικασία που αφορά την παραβίαση των κανόνων και σχετίζεται με μία λογιστική ή ελεγκτική παραβίαση που αφορά τα χρεόγραφα που επιβλέπει η Επιτροπή Κεφαλαιαγοράς.

Η ανάπτυξη του υποδείγματος γίνεται με τη χρήση του SAS Enterprise Miner. Το υπόδειγμα περιλαμβάνει τα ακόλουθα βήματα:

- Επιλογή των εταιρειών και του εγγράφου αναφοράς.
- Εξαγωγή του κειμένου αναφοράς από το έγγραφο.
- Προετοιμασία του κειμένου για την εισαγωγή του στο Enterprise Miner.
- Εισαγωγή των κειμένων αναφοράς από τα έγγραφα αναφοράς στο Enterprise Miner και δημιουργία μίας βάσης δεδομένων.
- Μείωση των διαστάσεων και σήμανση των μερών του λόγου.
- Δημιουργία μοναδιαίων διανυσμάτων κατάτμησης της τιμής (singular value decomposition vectors - SVDs) για την περαιτέρω μείωση των διαστάσεων του κειμένου.
- Άθροιση του συνόλου των εγγράφων.
- Αξιολόγηση των αποτελεσμάτων της άθροισης.
- Επεξεργασία των διανυσμάτων κατάτμησης της τιμής και του αλγορίθμου άθροισης, στον βαθμό που απαιτείται.

Δείγμα της έρευνας αποτέλεσαν επιχειρήσεις που κατηγορήθηκαν για απάτη από την Επιτροπή Κεφαλαιαγοράς των Η.Π.Α μεταξύ 2006 και 2008. Κάθε εταιρεία που κατηγορήθηκε για απάτη από την Επιτροπή Κεφαλαιαγοράς των Η.Π.Α αντιστοιχίστηκε με μία εταιρεία αντίστοιχου μεγέθους από τον ίδιο κλάδο που δεν αντιμετώπισε κατηγορίες απάτης. Το έγγραφο αναφοράς για κάθε εταιρεία ήταν η δήλωση K-10 που υπέβαλλε η εταιρεία. Κάθε δήλωση K-10 περιλαμβάνει 2 κύρια μέρη, τη Συζήτηση και Ανάλυση της Διοίκησης (Management's Discussion and Analysis) και τις Σημειώσεις στη Χρηματοοικονομική Δήλωση (Notes to the Financial Statement). Η Συζήτηση και Ανάλυση της Διοίκησης αποτέλεσε το κείμενο αναφοράς της έρευνας, προκειμένου να εξεταστεί η απάτη στις χρηματοοικονομικές αναφορές που προέρχεται από τη διοίκηση της εταιρείας ή η διοίκηση της εταιρείας εμπλέκεται σε αυτή.

Στα μοναδιαία διανύσματα κατάτμησης της τιμής (SVDs) με λογαριθμική στάθμιση (log frequency) και στάθμιση αύξησης της πληροφορίας (information gain term weights) πραγματοποιήθηκε επιβλεπόμενη άθροιση μεγιστοποίησης της προσδοκίας (supervised expectation maximization clustering) και ιεραρχική άθροιση. Η ιεραρχική άθροιση ήταν ευσταθής, αλλά η άθροιση μεγιστοποίησης της προσδοκίας ήταν ασταθής λόγω της ύπαρξης τοπικού ελάχιστου.

Με βάση τα ευρήματα της έρευνας, τα αποτελέσματα της άθροισης είναι σε θέση να εντοπίσουν πιθανή απάτη στις δηλώσεις K-10 ακόμα και σε επίπεδο σημαντικότητας 1%, με τη στατιστική τους δύναμη να ανέρχεται στο 90%. Στον πίνακα που ακολουθεί παρουσιάζονται τα αποτελέσματα της ιεραρχικής άθροισης.

Πίνακας 5.2. Αποτελέσματα της ιεραρχικής άθροισης του υποδείγματος των Glancy και Yadav (2011)

	Σύνολο	Ταξινομήθηκαν ορθά	Ταξινομήθηκαν λανθασμένα	p-τιμή
Έγγραφα	69	66	3	$1,2 * 10^{-14}$

Πηγή: Glancy και Yadav (2011)

Οι Glancy και Yadav (2011) προχώρησαν ακόμη και σε έλεγχο της εγκυρότητας των αποτελεσμάτων του υποδείγματος, ως προς την ικανότητα του υποδείγματος να διακρίνει μεταξύ

των επιχειρήσεων που είχαν διαπράξει απάτη και των επιχειρήσεων που δεν είχαν διαπράξει απάτη. Αρχικά, πραγματοποιήθηκε έλεγχος με ένα νέο δείγμα 10 εταιρειών που πληρούσαν τα κριτήρια επιλογής των αρχικών εταιρειών που έλαβαν προειδοποίηση από την Επιτροπή Κεφαλαιαγοράς. Από τις 10 Συζητήσεις και Αναλύσεις της Διοίκησης, η μέθοδος ταξινόμησε τις 9 ως περιπτώσεις απάτης, με την p -τιμή να είναι 0,0059. Για την ικανότητα εντοπισμού των εταιρειών που δεν έχουν διαπράξει απάτη χρησιμοποιήθηκε ένα δείγμα 20 εταιρειών που δεν είχαν κατηγορηθεί για απάτη, με βάση τα κριτήρια που χρησιμοποιήθηκαν αρχικά για την ανάπτυξη και τον έλεγχο του υποδείγματος. Από τις 20 εταιρείες οι οποίες δεν εμπλέκονταν σε απάτη, οι 4 ταξινομήθηκαν ως περιπτώσεις απάτης σύμφωνα με το υπόδειγμα, με την p -τιμή να διαμορφώνεται σε 0,059. Με βάση τα ευρήματα της έρευνας, το υπόδειγμα είναι σε θέση να διακρίνει μεταξύ περιπτώσεων απάτης και μη απάτης σε στατιστικά σημαντικά επίπεδα.

Στον επόμενο πίνακα παρουσιάζονται τα αποτελέσματα του ελέγχου για την ικανότητα ταξινόμησης του υποδείγματος τόσο σε περιπτώσεις απάτης όσο και σε περιπτώσεις μη απάτης, με βάση τη Συζήτηση και Ανάλυση της Διοίκησης στις χρηματοοικονομικές αναφορές των επιχειρήσεων.

Πίνακας 5.3. Τα αποτελέσματα του ελέγχου για την ικανότητα ταξινόμησης του υποδείγματος των Glancy και Yadav (2011)

Δείγμα	Σύνολο	Ταξινομήθηκαν ορθά	Ταξινομήθηκαν λανθασμένα	p -τιμή
Συζήτηση και Ανάλυση της Διοίκησης που υποκρύπτει απάτη	11	10	1	0,0059
Συζήτηση και Ανάλυση της Διοίκησης που δεν αποκρύπτει απάτη	20	16	4	0,0059

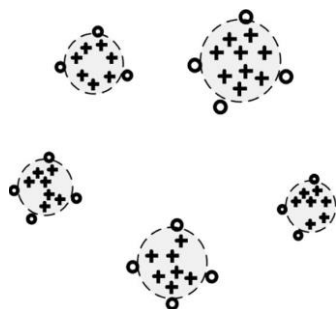
Πηγή: Glancy και Yadav (2011)

Το υπόδειγμα των Glancy και Yadav (2011) μπορεί να συμβάλλει στην αυτοματοποίηση της διαδικασίας εντοπισμού της απάτης, χρησιμοποιώντας το σύνολο των πληροφοριών που

περιέχονται σε ένα κείμενο. Μπορεί, ακόμη, να συμβάλλει στο έργο των ρυθμιστικών αρχών, προσφέροντας μία αρχική προσέγγιση εντοπισμού πιθανών περιπτώσεων απάτης, επιτρέποντας έτσι στις ρυθμιστικές αρχές να κατευθύνουν αποτελεσματικά τους διαθέσιμους πόρους τους και να αυξήσουν την πιθανότητα εντοπισμού των περιπτώσεων απάτης. Ταυτόχρονα, εκτός από την ικανότητα του να διακρίνει τις περιπτώσεις απάτης στις χρηματοοικονομικές αναφορές, μπορεί να χρησιμοποιηθεί και σε άλλους τομείς και πέρα από τις χρηματοοικονομικές αναφορές, όπως η πραγματοποίηση αναλύσεων για τους επενδυτές και η επιχειρηματική ευφυΐα (business intelligence).

Την αντιμετώπιση της απάτης στις χρηματοοικονομικές αναφορές επιδιώκει και το υπόδειγμα των Huang et al. (2014) με την εφαρμογή μη επιβλεπόμενων νευρωνικών δικτύων για την αναγνώριση μορφών μεταξύ παραπονημένων και μη παραπονημένων δεδομένων. Προτείνουν τη χρήση ενός ζεύγους διπλών Αναπτυσσόμενων Ιεραρχικών Αυτοοργανωμένων Χαρτών (Growing Hierarchical Self-Organizing Map) για την εξέταση της χωρικής υπόθεσης των παραπονημένων και μη παραπονημένων δεδομένων. Σύμφωνα με την χωρική υπόθεση των παραπονημένων δεδομένων, τα παραπονημένα δεδομένα βρίσκονται σε ξεχωριστές συστάδες και σε κάθε συστάδα τα μη παραπονημένα δεδομένα βρίσκονται συγκεντρωμένα στον εξωτερικό κύκλο.

Διάγραμμα 5.1. Σύμφωνα με τη χωρική υπόθεση του υποδείγματος των Huang et al. (2014) τα παραπονημένα δεδομένα βρίσκονται στον εξωτερικό κύκλο



Πηγή: Huang et al. (2014)

Το υπόδειγμα των Huang et al. (2014) χρησιμοποιεί παραπονημένα και μη παραπονημένα δεδομένα για την εκπαίδευση ενός ζεύγους Αναπτυσσόμενων Ιεραρχικών Αυτοοργανωμένων Χαρτών, με τις ίδιες παραμέτρους, και στη συνέχεια εξετάζει τις χωρικές μορφές μεταξύ των διαφόρων υποομάδων. Το υπόδειγμα περιλαμβάνει δύο κύριους μηχανισμούς. Τον μηχανισμό

κατάταξης και τον μηχανισμό εντοπισμού χαρακτηριστικών. Για τον μηχανισμό κατάταξης χρησιμοποιούνται αρχικά παραπονημένα και μη παραπονημένα δεδομένα για την εκπαίδευση δύο Αναπτυσσόμενων Ιεραρχικών Αυτοοργανωμένων Χαρτών, με τις ίδιες παραμέτρους, και αναπτύσσεται το δέντρο απάτης και το δέντρο μη-απάτης. Οι αντίστοιχοι κόμβοι κλαδιά του δέντρου απάτης και του δέντρου μη-απάτης χρησιμοποιούνται για την ανίχνευση της χωρικής μορφής.

Στη συνέχεια, συνάγεται ένας κανόνας κατάταξης αναφορικά με την κάθε υποομάδα για την ανίχνευση δειγμάτων απάτης. Αν διαπιστωθεί ότι υφίσταται μία χωρική σχέση εντοπίζονται τα χαρακτηριστικά της κάθε υποομάδας. Για κάθε αντίστοιχο κόμβο κλαδί από το δέντρο απάτης και το δέντρο μη-απάτης, όλα τα δείγματα εκπαίδευσης και οι στατιστικές πληροφορίες από αυτά, χρησιμοποιούνται για την εκτίμηση του κανόνα απάτης και του κανόνα μη-απάτης και στη συνέχεια χρησιμοποιείται ο κανόνας που διαθέτει την υψηλότερη απόδοση κατάταξης.

Για την εξέταση του υποδείγματος, οι Huang et al. (2014) χρησιμοποίησαν 762 χρηματοοικονομικές αναφορές από 144 εισηγμένες εταιρείες, οι μισές από τις οποίες έχουν υποπέσει σε χρηματοοικονομική απάτη και οι άλλες μισές δεν έχουν υποπέσει.

Ο μηχανισμός εντοπισμού των χαρακτηριστικών εντοπίζει τις κοινές κατηγορίες χρηματοοικονομικής απάτης από κάθε κόμβο δέντρο του δέντρου απάτης και τα χαρακτηριστικά του δείγματος απάτης για κάθε υποομάδα. Ο εντοπισμός των χαρακτηριστικών των μεταβλητών εισόδου που συνδέονται με τα δείγματα απάτης για κάθε υποομάδα γίνεται με τη χρήση της ανάλυσης κυρίων συνιστωσών (principal component analysis). Κάθε δείγμα ταξινομείται σε μία συγκεκριμένη υποομάδα μέσω του κυρίαρχου κόμβου φύλου στο δέντρο απάτης και του κυρίαρχου κόμβου φύλου στο δέντρο μη απάτης. Στη συνέχεια, εφαρμόζεται ο κυρίαρχος κανόνας κατάταξης για κάθε υποομάδα, ούτως ώστε να καθοριστεί εάν το δείγμα εντάσσεται στην κατηγορία απάτης ή μη απάτης. Εάν το δείγμα ανήκει στην κατηγορία απάτης, τότε τα χαρακτηριστικά της κάθε υποομάδας εντοπίζονται ώστε να καταδειχθεί η πιθανή παραποίηση των χρηματοοικονομικών αναφορών.

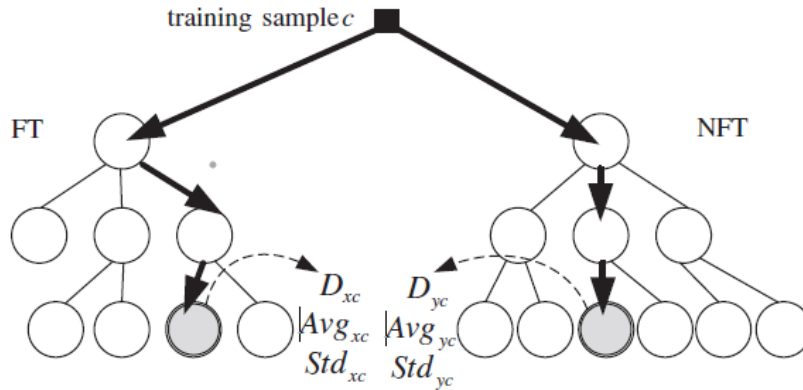
Ο μηχανισμός κατάταξης λειτουργεί ως εξής. Από το πρώτο μέχρι το τρίτο βήμα δημιουργούνται το δέντρο απάτης και το δέντρο μη απάτης από τα δείγματα εκπαίδευσης απάτης και μη απάτης του υποδείγματος. Αρχικά, χρησιμοποιείται το δείγμα με τις εισηγμένες που έχουν υποπέσει σε

απάτη για τον υπολογισμό ενός αποδεκτού δέντρου απάτης. Μετά τον εντοπισμό του αποδεκτού δέντρου απάτης, οι τιμές των παραμέτρων των Αναπτυσσόμενων Ιεραρχικών Αυτοοργανωμένων Χαρτών και το δείγμα με τις επιχειρήσεις που δεν έχουν διαπράξει απάτη χρησιμοποιούνται για τον υπολογισμό του δέντρου μη απάτης.

Στο τέταρτο και πέμπτο βήμα συλλέγονται τα στατιστικά για κάθε κόμβο φύλο από το δέντρο απάτης και το δέντρο μη απάτης. Στο τέταρτο βήμα, το δέντρο απάτης χρησιμοποιείται για την ομαδοποίηση όλων των δειγμάτων απάτης και στη συνέχεια υπολογίζονται και αποθηκεύονται η μέση τιμή των Ευκλείδειων αποστάσεων μεταξύ του διανύσματος στάθμισης και των ομαδοποιημένων δειγμάτων απάτης και η τυπική απόκλιση των Ευκλείδειων αποστάσεων μεταξύ του διανύσματος στάθμισης και των ομαδοποιημένων δειγμάτων απάτης για κάθε κόμβο του δέντρου απάτης. Στο πέμπτο βήμα, αντίστοιχα, χρησιμοποιείται το δέντρο μη απάτης και ακολουθείται η ίδια διαδικασία που ακολουθήθηκε στο τέταρτο βήμα για το δέντρο απάτης, αυτή τη φορά για το δέντρο μη απάτης.

Στο έκτο βήμα χρησιμοποιείται το δέντρο απάτης και το δέντρο μη απάτης για την ταξινόμηση όλων των δειγμάτων εκπαίδευσης του υποδείγματος και υπολογίζονται ο κυρίαρχος κόμβος του δέντρου απάτης και του δέντρου μη απάτης, καθώς και η μέση τιμή (Avg_{xc} και Avg_{yc}) και η τυπική απόκλιση (Std_{xc} και Std_{yc}) των Ευκλείδειων αποστάσεων μεταξύ του δείγματος εκπαίδευσης και του διανύσματος στάθμισης του κυρίαρχου κόμβου κλαδί του δέντρου απάτης και του δέντρου μη απάτης αντίστοιχα (D_{xc} και D_{yc}). Η διαδικασία απεικονίζεται στο παρακάτω σχήμα.

Διάγραμμα 5.2. Το δέντρο απάτης και το δέντρο μη απάτης στο υπόδειγμα των Huang et al. (2014)



Πηγή: Huang et al. (2014).

Στο έβδομο βήμα, αντιστοιχίζονται οι κόμβοι κλαδιά του δέντρου απάτης και του δέντρου μη απάτης με βάση τον κανόνα της πλειοψηφίας. Με βάση τον κανόνα της πλειοψηφίας υπάρχει το ενδεχόμενο η χωρική σχέση να είναι ένα-προς-πολλά και ένας κόμβος του δέντρου απάτης μπορεί να έχει περισσότερους αντίστοιχους κόμβους φύλα στο δέντρο μη απάτης. Στο όγδοο βήμα χρησιμοποιείται ο κανόνας απάτης που δίνεται από:

$$\text{Αν } D_{xc} < Avg_x + \beta_x Std_{yc}$$

Τότε το δείγμα c ταξινομείται ως δείγμα απάτης, διαφορετικά ταξινομείται ως δείγμα μη απάτης.

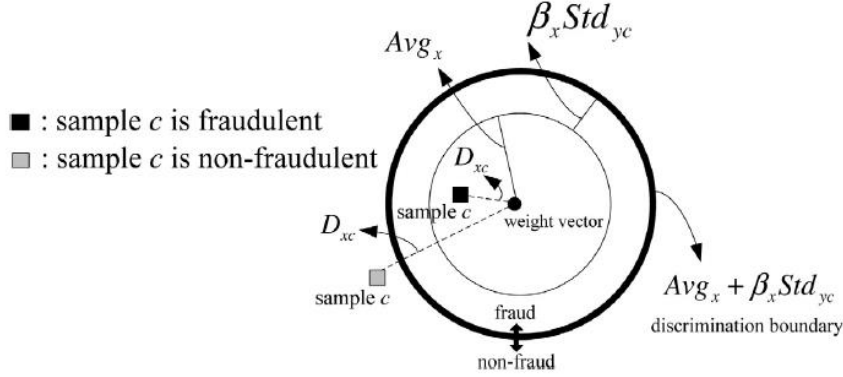
Η παράμετρος β_x δίνεται από τη λύση του προβλήματος βελτιστοποίησης:

$$\min_{\beta_x} w_1 \sum_{c \in F_x} (i^c + 1)^2 + w_2 \sum_{c \in N_x} (i^c - 1)^2$$

$$\text{υ.π. } i^c = 1 \text{ εάν } D_{xc} < Avg_x + \beta_x Std_{yc} \text{ ή } -1 \text{ εάν } D_{xc} \geq Avg_x + \beta_x Std_{yc}$$

Στο διάγραμμα που ακολουθεί παρουσιάζεται σχηματικά ο κανόνας απάτης.

Διάγραμμα 5.3. Ο κανόνας απάτης σχηματικά με βάση το υπόδειγμα των Huang et al. (2014)



Πηγή: Huang et al. (2014)

Οι συντελεστές w_1 και w_2 αποτελούν τις σταθμίσεις των σφαλμάτων ταξινόμησης τύπου I και τύπου II, αντίστοιχα. Τα σφάλματα τύπου I ορίζονται ως το ποσοστό των δειγμάτων μη απάτης που ταξινομούνται λανθασμένα ως δείγματα απάτης και τα σφάλματα τύπου II ορίζονται ως το ποσοστό των δειγμάτων απάτης που ταξινομούνται λανθασμένα ως δείγματα μη απάτης.

Στο ένατο βήμα χρησιμοποιείται ο κανόνας μη απάτης που δίνεται από:

$$\text{Αν } D_{yc} < Avg_y + \beta_y Std_{xc}$$

Τότε το δείγμα c ταξινομείται ως δείγμα μη απάτης, διαφορετικά ταξινομείται ως δείγμα απάτης.

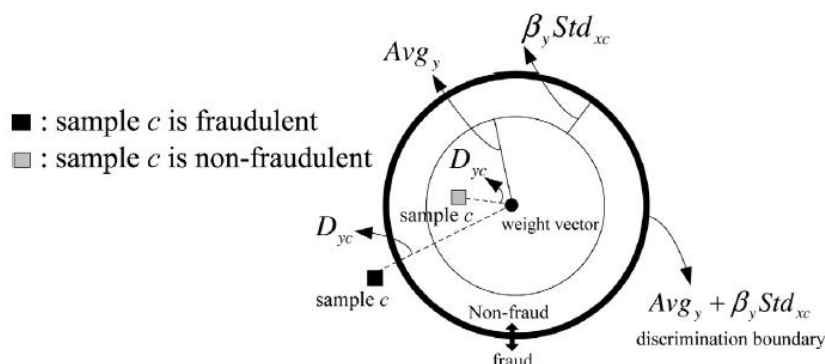
Η παράμετρος β_y δίνεται από τη λύση του προβλήματος βελτιστοποίησης:

$$\min_{\beta_y} w_1 \sum_{c \in N_y} (i^c + 1)^2 + w_2 \sum_{c \in F_y} (i^c - 1)^2$$

$$\text{υ.π. } i^c = 1 \text{ εάν } D_{yc} < Avg_y + \beta_y Std_{xc} \text{ ή } -1 \text{ εάν } D_{yc} \geq Avg_y + \beta_y Std_{xc}$$

Στο διάγραμμα που ακολουθεί παρουσιάζεται σχηματικά ο κανόνας μη απάτης.

Διάγραμμα 5.4. Ο κανόνας μη απάτης σχηματικά με βάση το υπόδειγμα των Huang et al. (2014)



Πηγή: Huang et al. (2014)

Στο δέκατο βήμα επιλέγεται ο κανόνας απάτης ως ο κυρίαρχος κανόνας ταξινόμησης εάν το άθροισμα των σφαλμάτων κατάταξης τύπου I και τύπου II του κανόνα απάτης σε όλα τα δείγματα εκπαίδευσης είναι μικρότερο του κανόνα μη απάτης, διαφορετικά επιλέγεται ο κανόνας μη απάτης ως ο κυρίαρχος κανόνας κατάταξης.

Πέρα από την κατάταξη των δειγμάτων, προχωρούν και στον εντοπισμό των χαρακτηριστικών των δειγμάτων απάτης στο πλαίσιο της υποστήριξης αποφάσεων. Για τον εντοπισμό των χαρακτηριστικών των δειγμάτων απάτης χρησιμοποιούνται τα δείγματα απάτης που ταξινομούνται στον ίδιο κόμβο κλαδί στο δέντρο απάτης.

Ο μηχανισμός εντοπισμού των χαρακτηριστικών αποτελείται από δύο βήματα, τον εντοπισμό των χαρακτηριστικών και τον εντοπισμό των μορφών. Για κάθε κόμβο κλαδί του δέντρου απάτης χρησιμοποιείται η ανάλυση κυρίων συνιστωσών για την εξαγωγή των παραγόντων από τις μεταβλητές εισόδου των ομαδοποιημένων δειγμάτων απάτης. Αποκλείοντας τις λιγότερο σχετικές μεταβλητές εισόδου, η ανάλυση κυρίων συνιστωσών επιλέγει ένα σύνολο μεταβλητών που αντιπροσωπεύει καλύτερα τα σύνθετα χαρακτηριστικά για κάθε υποομάδα. Οι παράγοντες αυτοί αντιπροσωπεύουν τα χαρακτηριστικά των μεταβλητών που αποκαλύπτουν τις ομοιότητες μεταξύ των χρηματοοικονομικών λόγων των δειγμάτων απάτης που ταξινομούνται στον ίδιο κόμβο κλαδί του δέντρου απάτης. Τα χαρακτηριστικά αυτά, όπως αναφέρουν οι Huang et al. (2014), μπορούν

να χρησιμοποιηθούν για την υποστήριξη αποφάσεων κατά τον εντοπισμό της απάτης, σε συνδυασμό με πρόσθετες πληροφορίες που σχετίζονται με περιπτώσεις απάτης.

Για την ανάλυση κυρίων συνιστωσών χρησιμοποιούνται ως ανεξάρτητες μεταβλητές οι μεταβλητές εισόδου που χρησιμοποιήθηκαν για το δέντρο απάτης και το δέντρο μη απάτης. Η μεταβλητή που λαμβάνει την τιμή απάτη/μη απάτη χρησιμοποιείται ως ανεξάρτητη μεταβλητή και το αποτέλεσμα περιλαμβάνει μόνο τους παράγοντες για τους οποίους η ιδιοτιμή είναι μεγαλύτερη της μονάδας. Στη συνέχεια και προκειμένου να είναι ευκολότερη η ερμηνεία των παραγόντων, εφάρμοσαν τη μέθοδο περιστροφής των παραγόντων varimax, για τη μείωση του αριθμού των μεταβλητών που έχουν μεγάλες φορτώσεις (loadings) σε έναν παράγοντα.

Η διαδικασία υποστήριξης αποφάσεων καθορίζει το αν το υπό εξέταση δείγμα αφορά περίπτωση απάτης ή μη απάτης, αλλά παρέχει και το αποτέλεσμα ταξινόμησης και τα χαρακτηριστικά που σχετίζονται με απάτη, ώστε να συμβάλλει στη διαδικασία λήψης αποφάσεων για τα επόμενα δείγματα. Στο πρώτο βήμα, κάθε υπό εξέταση δείγμα εισάγεται τόσο στο δέντρο απάτης όσο και στο δέντρο μη απάτης και καθορίζεται ο κυρίαρχος κόμβος κλαδί. Στο δεύτερο βήμα, εάν ο κυρίαρχος κανόνας ταξινόμησης είναι ο κανόνας απάτης, χρησιμοποιείται ο κανόνας απάτης για την ταξινόμηση του δείγματος, ενώ, αντίθετα, αν είναι κυρίαρχος ο κανόνας μη απάτης χρησιμοποιείται ο κανόνας μη απάτης για την ταξινόμηση του υπό εξέταση δείγματος. Στο τέταρτο βήμα ανακτώνται οι κύριες συνιστώσες και οι διάφορες κατηγορίες απάτης του κυρίαρχου κόμβου κλαδί για το υπό εξέταση δείγμα. Για κάθε δείγμα για το οποίο υπάρχουν υποψίες απάτης, ανεξάρτητα της ταξινόμησης του, ανακτώνται οι κύριες συνιστώσες και οι πιθανές κοινές κατηγορίες χρηματοοικονομικής απάτης και στη συνέχεια μπορούν να χρησιμοποιηθούν ως προειδοποιητικά σήματα για την αποκάλυψη δυνητικών περιπτώσεων απάτης.

Για την αξιολόγηση της ικανότητας κατάταξης του υποδείγματος χρησιμοποιήθηκε ένα δείγμα ελέγχου με 54 εταιρείες που έχουν διαπράξει χρηματοοικονομική απάτη και 128 εταιρείες που δεν έχουν διαπράξει χρηματοοικονομική απάτη, για την περίοδο μεταξύ 2002 και 2008. Με βάση τα αποτελέσματα της αξιολόγησης, τα σφάλματα τύπου I ανέρχονται στο 11,54% των περιπτώσεων και τα σφάλματα τύπου II στο 19,78% των περιπτώσεων. Όπως σημειώνουν οι Huang et al. (2014), καθώς το ποσοστό για κανένα τύπο σφαλμάτων δεν ξεπερνά το 20%, η απόδοση του υποδείγματος θεωρείται ικανοποιητική. Ταυτόχρονα, το σύστημα υποστήριξης αποφάσεων είναι

σε θέση να αναγνωρίσει σωστά το 82,47% των περιπτώσεων για τις οποίες δεν έχει διαπραχθεί χρηματοοικονομική απάτη.

Η απόδοση του υποδείγματος αναλύθηκε και συγκριτικά και με άλλες μεθόδους, όπως η μέθοδος των εγγύτερων γειτόνων (nearest neighbor), νευρωνικά δίκτυα προς τα πίσω διάδοσης (back-propagation neural network - BPNN), τα μηχανήματα διανυσμάτων υποστήριξης (support vector machines - SVM), η γραμμική ανάλυση διακρίνουσας και αυτοοργανωμένων χαρτών (linear discriminant analysis (LDA) and self-organizing maps (SOM)) και Αναπτυσσόμενων Ιεραρχικών Αυτοοργανωμένων Χαρτών με γραμμική ανάλυση διακρίνουσας (GHSOM with LDA). Στον πίνακα που ακολουθεί παρουσιάζονται τα αποτελέσματα της σύγκρισης της αποτελεσματικότητας κατάταξης μεταξύ των διαφορετικών μεθόδων των Huang et al. (2014).

Πίνακας 5.4. Αποτελέσματα σύγκρισης της αποτελεσματικότητας κατάταξης της μεθόδου των Huang et al. (2014) με άλλες μεθόδους

	Δείγμα εκπαίδευσης		Δείγμα ελέγχου	
	Σφάλματα τύπου I	Σφάλματα τύπου II	Σφάλματα τύπου I	Σφάλματα τύπου II
Ζεύγος Αναπτυσσόμενων Ιεραρχικών Αυτοοργανωμένων Χαρτών	13.62%	13.28%	11.54%	19.78%
Μέθοδος εγγύτερων γειτόνων	46.04%	25.66%	28.91%	53.70%
Μέθοδος BPNN	15,20%	32,74%	10,94%	64,81%
Μέθοδος SVM	24,84%	21,24%	45,31%	27,78%
Μέθοδος SOM + LDA	16,49%	30,09%	32,81%	51,85%
Μέθοδος GHSOM + LDA	20,34%	22,12%	22,66%	44,44%

Πηγή: Huang et al. (2014)

Σύμφωνα με τα αποτελέσματα της σύγκρισης, η μέθοδος της εκπαίδευσης ενός ζεύγους Αναπτυσσόμενων Ιεραρχικών Αυτοοργανωμένων Χαρτών που προτείνουν οι Huang et al. (2014),

προσφέρει καλύτερη απόδοση συγκριτικά με τις άλλες μεθόδους, με χαμηλότερα σφάλματα τύπου Ι και τύπου ΙΙ τόσο για το δείγμα εκπαίδευσης όσο και για το δείγμα ελέγχου. Οι Huang et al. (2014) σημειώνουν, ακόμη, ότι το υπόδειγμα τους είναι σε θέση να παρέχει πληροφορίες για πιθανές περιπτώσεις απάτης αντίστοιχης ακρίβειας με εκείνες των ειδικών και συνεπώς μπορεί να αποτελέσει χρήσιμο εργαλείο υποστήριξης αποφάσεων. Μειονέκτημα της μεθόδου αποτελεί το γεγονός ότι οι κανόνες εξαρτώνται από τα δεδομένα, κατά συνέπεια δεν μπορούν να εφαρμοστούν άμεσα σε άλλα πλαίσια στα οποία δεν ισχύουν οι ίδιες χωρικές υποθέσεις.

Μία πρόσθετη εφαρμογή της επιστήμης των δεδομένων για τον εντοπισμό περιστατικών χρηματοοικονομικής απάτης αποτελεί και το υπόδειγμα των Balla et al. (2014), οι οποίοι αναπτύσσουν ένα υπόδειγμα κατάταξης για την πρόβλεψη ομαδικών αγωγών για απάτη εις βάρος των μετόχων και άλλων εμπλεκόμενων μερών (stakeholders) κατά των αμερικανικών τραπεζών. Όπως σημειώνουν, οι ομαδικές αγωγές μπορούν, υπό ορισμένες προϋποθέσεις, να επηρεάσουν αρνητικά τη φήμη και τη φερεγγυότητα των τραπεζών.

Το υπόδειγμα χρησιμοποιεί τις εξής μεταβλητές. Για την κεφαλαιακή ισχύ χρησιμοποιείται ο λόγος των μετοχών προς το συνολικό ενεργητικό, για την ποιότητα των περιουσιακών στοιχείων χρησιμοποιείται ο λόγος των αποθεμάτων δανειακών απωλειών προς τα συνολικά δάνεια, η κερδοφορία αποτιμάται βάσει των αποδόσεων επί των μέσων περιουσιακών στοιχείων, η ρευστότητα από τα ρευστά περιουσιακά στοιχεία προς τις καταθέσεις και τον βραχυπρόθεσμο δανεισμό. Για τη συνολική δανειακή δραστηριότητα των τραπεζών χρησιμοποιείται ο λόγος των καθαρών δανείων προς το συνολικό ενεργητικό και για το μέγεθος μίας τράπεζας ο λογάριθμος του συνολικού ενεργητικού.

Δείγμα της έρευνας αποτέλεσαν 120 ομαδικές αγωγές για το χρονικό διάστημα μεταξύ 2002 και τα μέσα του 2011, οι οποίες αντιστοιχίστηκαν με ένα δείγμα 120 τραπεζών οι οποίες δεν βρέθηκαν ποτέ στο επίκεντρο μίας ομαδικής αγωγής κατά το παρελθόν, με την αντιστοίχιση να γίνεται βάσει του τύπου της τράπεζας.

Το υπόδειγμα βασίζεται στη μέθοδο Utilites Additives DIScriminantes (UTADIS), η οποία χρησιμοποιεί μία συνάρτηση χρησιμότητας της ακόλουθης μορφής:

$$U(x) = \sum_{i=1}^n w_i u'_i(g_i) \in [0,1]$$

Όπου w_i είναι η στάθμιση του κριτηρίου g_i , με τα κριτήρια να αθροίζονται ίσα με 1 και $u'_i(g_i)$ είναι η αντίστοιχη οριακή συνάρτηση χρησιμότητας, η οποία κανονικοποιείται μεταξύ του 0 και του 1.

Η συνάρτηση χρησιμότητας μπορεί να απλοποιηθεί στην ακόλουθη μορφή, η οποία παρέχει μία συνολική βαθμολογία $U(x)$ για κάθε τράπεζα για όλα τα κριτήρια:

$$U(x) = \sum_{i=1}^n u'_i(g_i) \in [0,1]$$

Στην περίπτωση της πρόβλεψης των ομαδικών αγωγών στο συγκεκριμένο υπόδειγμα, η $U(x)$ αποτελεί τη βάση καθορισμού εάν μια τράπεζα θα πρέπει να ταξινομηθεί στην ομάδα τραπεζών για τις οποίες δεν έχει κατατεθεί ομαδική αγωγή (C_1) ή στην ομάδα τραπεζών για τις οποίες έχει κατατεθεί ομαδική αγωγή (C_2), με βάση τον κανόνα ταξινόμησης:

$$U(x) \geq u_i \Rightarrow x \in C_1$$

$$U(x) \geq u_i \Rightarrow x \in C_2$$

Για την εκτίμηση της συνάρτησης και του ορίου αποκοπής (cut-off threshold) (u_i) χρησιμοποιούνται τεχνικές γραμμικού προγραμματισμού. Στόχος της μεθόδου είναι ο κανόνας κατάταξης να είναι σε θέση να αποτυπώσει τις ταξινομήσεις των τραπεζών με όσο το δυνατόν μεγαλύτερη ακρίβεια. Ο γραμμικός προγραμματισμός χρησιμοποιείται για την ελαχιστοποίηση του αθροίσματος όλων των παραβιάσεων του κανόνα κατάταξης για όλες τις τράπεζες στο δείγμα εκπαίδευσης.

Το υπόδειγμα αναπτύχθηκε και ελέγχθηκε στη βάση μίας 10-πλης διασταυρωμένης επικύρωσης (cross-validation), που επιτρέπει τη μέγιστη αξιοποίηση των διαθέσιμων δεδομένων κατά την εκτίμηση του υποδείγματος, διασφαλίζοντας και την κατάλληλη επικύρωση του υποδείγματος εκτός δείγματος. Το συνολικό δείγμα 240 τραπεζών κατατμήθηκε τυχαία σε 10 αμοιβαία αποκλειόμενα υποσύνολα. Στη συνέχεια, αναπτύσσονται διαδοχικά 10 υποδείγματα,

χρησιμοποιώντας 9 υποσύνολα για την εκπαίδευση και ένα για την επικύρωση. Ο μέσος όρος σφαλμάτων και των 10 υποδειγμάτων αποτελεί το ποσοστό σφάλματος της διασταυρωμένης επικύρωσης.

Για τη σύγκριση της απόδοσης του υποδείγματος με άλλα υποδείγματα, οι Balla et al. (2014) ανέπτυξαν και άλλα δύο υποδείγματα με τη χρήση παραδοσιακών μεθόδων, όπως η ανάλυση διακρίνουσας και η λογιστική παλινδρόμηση. Ο έλεγχος των εναλλακτικών υποδειγμάτων έγινε με την ίδια μεθοδολογία που χρησιμοποιήθηκε για την ανάπτυξη του υποδείγματος ταξινόμησης με τη χρήση της μεθόδου UTADIS. Σύμφωνα με τα αποτελέσματα της έρευνας, οι συνολικές ορθές ταξινομήσεις ανέρχονται σε 77,05% για την ανάλυση διακρίνουσας, 78% για το υπόδειγμα της λογιστικής παλινδρόμησης και 79,59% για το υπόδειγμα UTADIS. Το υπόδειγμα UTADIS αναδεικνύεται σχετικά αποτελεσματικότερο όσον αφορά την κατάταξη μεταξύ τραπεζών για τις οποίες έχουν υποβληθεί ομαδικές αγωγές καθώς και των τραπεζών για τις οποίες δεν έχουν υποβληθεί ομαδικές αγωγές, αν και όλα τα υποδείγματα προσφέρουν σχετικά ικανοποιητική απόδοση.

Στη συνέχεια εκτιμήθηκαν τα αποτελέσματα για το δείγμα επικύρωσης, με τα αποτελέσματα αυτά να παρέχουν μία περισσότερο ακριβή εκτίμηση της ικανότητας ταξινόμησης των υποδειγμάτων. Όλα τα υποδείγματα επιτυγχάνουν αντίστοιχα αποτελέσματα με το δείγμα εκπαίδευσης, αλλά το υπόδειγμα UTADIS είναι αυτό που παρουσιάζει τη μικρότερη μείωση της ακρίβειας του, καθώς η ακρίβεια του παρουσιάζει μείωση κατά μόλις 0.10%, ενώ η ακρίβεια του υποδείγματος ανάλυσης διακρίνουσας μειώνεται κατά 1,76% και της λογιστικής παλινδρόμησης κατά 1,44%. Ταυτόχρονα, το υπόδειγμα UTADIS σημειώνει καλύτερες αποδόσεις αναφορικά και με τις 2 κατηγορίες τραπεζών, ταξινομώντας ορθά το 82,89% των τραπεζών για τις οποίες δεν έχει υποβληθεί ομαδική αγωγή και το 76,10% για τις οποίες έχει υποβληθεί ομαδική αγωγή, με τα αντίστοιχα ποσοστά για το υπόδειγμα λογιστικής ανάλυσης να ανέρχονται σε 79,72% και 73,40% και για το υπόδειγμα ανάλυσης της διακρίνουσας σε 79,64% και 70,94%, αντίστοιχα.

Στον πίνακα που ακολουθεί παρουσιάζονται τα αποτελέσματα της ταξινόμησης των 3 υποδειγμάτων της έρευνας των Balla et al. (2014) για το δείγμα εκπαίδευσης και το δείγμα επικύρωσης.

Πίνακας 5.5. Τα αποτελέσματα της ταξινόμησης των 3 υποδειγμάτων της έρευνας των Balla et al. (2014) για το δείγμα εκπαίδευσης και το δείγμα επικύρωσης.

Κατηγορία			
	Δεν έχει υποβληθεί ομαδική αγωγή	Έχει υποβληθεί ομαδική αγωγή	Συνολικά
Δείγμα εκπαίδευσης			
UTADIS	83,71%	75,47%	79,59%
Λογιστική παλινδρόμηση	80,26%	75,74%	78%
Ανάλυση διακρίνουσας	81,50%	72,61%	77,05%
Δείγμα επικύρωσης			
UTADIS	82,89%	76,10%	79,49%
Λογιστική παλινδρόμηση	79,72%	73,40%	76,56%
Ανάλυση διακρίνουσας	79,64%	70,94%	75,29%

Πηγή: Balla et al. (2014)

Όπως σημειώνουν οι Balla et al. (2014), πέρα από τη σχετικά μεγαλύτερη ικανότητα κατάταξης, πρόσθετο πλεονέκτημα της μεθόδου UTADIS αποτελεί και το γεγονός ότι δεν απαιτούνται συγκεκριμένες υποθέσεις, όπως στην περίπτωση των παραδοσιακών μεθόδων στατιστικής και οικονομετρίας, σχετικά με την κανονικότητα των μεταβλητών ή την διασπορά στους πίνακες των κατηγοριών, όπως στην περίπτωση της ανάλυσης διακρίνουσας, ενώ και η μέθοδος UTADIS δεν είναι ευαίσθητη σε ζητήματα πολυσυγγραμικότητας ή ακραίων τιμών, όπως η λογιστική ανάλυση.

Για την εξέταση της ευστάθειας των αποτελεσμάτων του υποδείγματος στον χρόνο οι Balla et al. (2014) χρησιμοποιούν την τεχνική της κίνησης προς τα εμπρός (walk-forward), όπου γίνεται η εξέταση του υποδείγματος σε διαφορετικές διαδοχικές χρονικές περιόδους του δείγματος. Τα ευρήματα της έρευνας καταδεικνύουν ότι η ακρίβεια κατάταξης του υποδείγματος UTADIS κυμαίνεται μεταξύ 66,67% για το έτος 2011 και 84,53% για το έτος 2008, με τη μέση ακρίβεια για το σύνολο της περιόδου να διαμορφώνεται στο 76,03%. Στον πίνακα που ακολουθεί

παρουσιάζονται τα αποτελέσματα της ανάλυσης της ευστάθειας των αποτελεσμάτων κατάταξης του υποδείγματος των Balla et al. (2014) στον χρόνο, με τη χρήση διαδοχικών δειγμάτων.

Πίνακας 5.6. Τα αποτελέσματα της ανάλυσης της ευστάθειας των αποτελεσμάτων κατάταξης του υποδείγματος των Balla et al. (2014) στον χρόνο, με τη χρήση διαδοχικών δειγμάτων

Εκτίμηση						Επικύρωση			
						Δεν έχει υποβληθεί ομαδική αγωγή	Έχει υποβληθεί ομαδική αγωγή	Μέσος όρος	
Περίοδος 1 υποδείγματος	1	2002 - 2007	88,10%	80,95%	84,52%	2008	88,10%	80,95%	84,52%
Περίοδος 2 υποδείγματος	2	2002 - 008	87,84%	79,73%	83,78%	2009	82,61%	78,26%	80,43%
Περίοδος 3 υποδείγματος	3	2002 - 2009	86,60%	75,26%	80,93%	2010	75%	70,00%	72,50%
Περίοδος 4 υποδείγματος	4	2002 - 2010	83,76%	77,78%	80,77%	2011	100,00%	33,33%	66,67%
Μέσος όρος			86,57%	78,43%	82,50%		86,43%	65,64%	76,03%

Πηγή: Balla et al. (2014)

Στη συνέχεια, οι Balla et al. (2014) εξέτασαν και αν η προσθήκη μεταβλητών σχετικά με τα χαρακτηριστικά της εταιρικής διοίκησης βελτιώνει την απόδοση του υποδείγματος. Αρχικά, προχωρούν στην προσθήκη της μεταβλητής Πηλίο Εταιρικής Διακυβέρνησης (Corporate Governance Quotient) που αναπτύσσεται από την εταιρεία RiskMetrics και επιχειρεί την ποσοτικοποίηση της ποιότητας της εταιρικής διακυβέρνησης μίας επιχείρησης, συγκριτικά με άλλες επιχειρήσεις του κλάδου στον οποίο δραστηριοποιείται και υπολογίζεται με τη χρήση 60 μεταβλητών που αφορούν το διοικητικό συμβούλιο μίας επιχείρησης, τις αμοιβές των στελεχών της και την ιδιοκτησία της, τον λογιστικό της έλεγχο και την ικανότητα αντιμετώπισης μίας προσπάθειας επιθετικής εξαγοράς.

Με βάση τα αποτελέσματα της έρευνας, η στάθμιση του δείκτη στο υπόδειγμα UTADIS είναι σχετικά μικρή και ανέρχεται σε 3,22%, με την ακρίβεια κατάταξης στο δείγμα εκπαίδευσης να παρουσιάζει μικρή βελτίωση συγκριτικά με το υπόδειγμα που περιέχει μόνο τις χρηματοοικονομικές μεταβλητές και να ανέρχεται στο 87,54% συγκριτικά με 86,82% στο αρχικό υπόδειγμα. Παρ' όλα αυτά, η απόδοση του υποδείγματος στο δείγμα επικύρωσης παρουσιάζει μείωση και μειώνεται στο 77,29% συγκριτικά με 80,20% για το υπόδειγμα μόνο με τις χρηματοοικονομικές μεταβλητές, μειώνοντας έτσι την ακρίβεια του υποδείγματος εκτός δείγματος.

Στη συνέχεια, προχώρησαν και στην εξέταση του υποδείγματος με μία μεταβλητή που αφορά την ιδιοκτησία της τράπεζας. Τα αποτελέσματα σε αυτή την περίπτωση είναι διαφορετικά, καθώς η προσθήκη της μεταβλητής που αφορά την ιδιοκτησία της τράπεζας βελτιώνει την ικανότητα ορθής ταξινόμησης του υποδείγματος κατά 5,33% για τις τράπεζες για τις οποίες έχει κατατεθεί ομαδική αγωγή και κατά 8,67% για τις τράπεζες για τις οποίες δεν έχει κατατεθεί κάποια ομαδική αγωγή, με τη συνολική ικανότητα ταξινόμησης του υποδείγματος να βελτιώνεται και να ανέρχεται στο 87,20%, συγκριτικά με συνολική ακρίβεια της τάξης του 80,20% για το αρχικό υπόδειγμα που περιλαμβάνει μόνο τις χρηματοοικονομικές μεταβλητές.

Στον πίνακα που ακολουθεί παρουσιάζεται η ικανότητα κατάταξης του υποδείγματος ανάλογα με το σύνολο μεταβλητών που χρησιμοποιείται στην έρευνα των Balla et al. (2014).

Πίνακας 5.7. Η ικανότητα κατάταξης του υποδείγματος των Balla et al. (2014) ανάλογα με το σύνολο μεταβλητών που χρησιμοποιείται

Υπόδειγμα	Δείγμα εκπαίδευσης			Δείγμα επικύρωσης		
	Δεν υποβληθεί ομαδική αγωγή	Έχει υποβληθεί ομαδική αγωγή	Μέσος όρος	Δεν υποβληθεί ομαδική αγωγή	Έχει υποβληθεί ομαδική αγωγή	Μέσος όρος
Μόνο χρηματοοικονομικές μεταβλητές	87,44%	86,20%	86,82%	80,33%	80,07%	80,20%
Χρηματοοικονομικές μεταβλητές και το Πηλίκιο Εταιρικής	86,72%	88,37%	87,54%	76,17%	78,40%	77,29%

Διακυβέρνησης από την
RiskMetrics

Χρηματοοικονομικές μεταβλητές και μεταβλητή για το Διοικητικό Συμβούλιο	88,90%	87,16%	88,03%	78,67%	80,07%	79,37%
Χρηματοοικονομικές μεταβλητές και μεταβλητή για την ιδιοκτησία	91,07%	89,84%	90,45%	89%	85,40%	87,20%
Χρηματοοικονομικές μεταβλητές και μεταβλητή για την ικανότητα άμυνας σε επιθετικές εξαγορές	87,94%	85,73%	86,83%	80,33%	76,74%	78,54%
Χρηματοοικονομικές μεταβλητές και μεταβλητή για λογιστικό έλεγχο	87,44%	86,20%	86,82%	80,33%	80,07%	80,20%

Πηγή: Balla et al. (2014)

Με βάση τα ευρήματα της έρευνας των Balla et al. (2014), το υπόδειγμα με τη χρήση της μεθόδου UTADIS προσφέρει καλύτερη απόδοση συγκριτικά με υποδείγματα που βασίζονται σε παραδοσιακές μεθόδους, όπως η ανάλυση της διακρίνουσας και η λογιστική παλινδρόμηση, ενώ και το υπόδειγμα παρουσιάζεται ευσταθές τόσο στα πλαίσια του δείγματος εκπαίδευσης όσο και στο δείγμα επικύρωσης, αλλά και με τη χρήση διαδοχικών δειγμάτων. Οι μόνες μεταβλητές η προσθήκη των οποίων θα μπορούσε να βελτιώσει την απόδοση του υποδείγματος, συγκριτικά με τη χρήση μόνο χρηματοοικονομικών μεταβλητών, είναι οι αμοιβές των διευθυντικών στελεχών και η εταιρική διακυβέρνηση της τράπεζας.

6. ΕΚΤΙΜΗΣΗ ΑΞΙΑΣ ΠΑΡΑΓΩΓΩΝ ΚΑΙ ΥΠΟΚΕΙΜΕΝΩΝ ΠΕΡΙΟΥΣΙΑΚΩΝ ΣΤΟΙΧΕΙΩΝ

Οι έντονες διακυμάνσεις στις αγορές προβληματίζουν τους επενδυτές, με αποτέλεσμα να στρέφονται στις συναλλαγές με δικαιώματα προαίρεσης και παράγωγα, αντί να εμπορεύονται άμεσα τα υποκείμενα περιουσιακά στοιχεία. Το γεγονός αυτό προκάλεσε το έντονο ενδιαφέρον των ερευνητών και ανέδειξε την ανάγκη για την ανάπτυξη νέων τεχνικών για την αντιμετώπιση των περιορισμών των παραδοσιακών παραμετρικών υποδειγμάτων τιμολόγησης, τα οποία βασίζονται σε υποθέσεις και προσεγγίσεις για τον εντοπισμό των σύνθετων δυναμικών που χαρακτηρίζουν τις διαδικασίες τιμολόγησης. Οι Quek et al. (2008) προτείνουν μία νέα μη παραμετρική μεθοδολογία με τη χρήση ενός ad-hoc επαναλαμβανόμενου νευρωνικού δικτύου (recurrent neural network) για την εξέλιξη της μελλοντικής πορείας των τιμών εμπορευμάτων όπως ο χρυσός και το αργό πετρέλαιο, καθώς και τα νομίσματα, τα οποία συγκεντρώνουν το έντονο ενδιαφέρον των επενδυτών και σημαντικό όγκο επενδυτικής δραστηριότητας στις χρηματαγορές.

Ο χρήστης του υποδείγματος εισάγει τα δεδομένα και το σύστημα εντοπίζει τα κατάλληλα σημεία για την πραγματοποίηση επικερδών συναλλαγών, καθώς και τις τάσεις στην αγορά. Το στάδιο της προεπεξεργασίας περιλαμβάνει 3 βήματα. Το πρώτο βήμα περιλαμβάνει τον καθαρισμό και την κανονικοποίηση των δεδομένων με τη μέθοδο της ισοδιαστημικής (interval scaling) για τη μείωση της επίδρασης των ακραίων τιμών. Όλα τα σημεία κανονικοποιούνται με έναν συγκεκριμένο παράγοντα. Για την κανονικοποίηση των δεδομένων χρησιμοποιείται ο παρακάτω τύπος:

$$y_{if} = \frac{(x_{if} - m_{jf})}{S_{if}}$$

Όπου y_{if} είναι η κανονικοποιημένη μεταβλητή εξόδου, x_{if} το αρχικό σημείο, m_{jf} ο μέσος του συνόλου δεδομένων, n είναι ο συνολικός αριθμός των δεδομένων και S_{if} η μέση απόλυτη απόκλιση που υπολογίζεται με βάση τον τύπο:

$$S_{if} = (|x_1 - m_{if}| + |x_2 - m_{if}| + \dots + |x_n - m_{if}|)/n$$

Στη συνέχεια πραγματοποιείται η επιλογή των χαρακτηριστικών, που αυξάνει την απόδοση του συστήματος τόσο όσον αφορά την ακρίβεια των προβλέψεων και όσον αφορά την υπολογιστική του αποτελεσματικότητα.

Καθώς υπάρχει μία σχέση μεταξύ θορύβου και μη στασιμότητας, το σύστημα δημιουργεί υποδείγματα αποκλειστικά για βραχυχρόνιες περιόδους. Εάν το σύνολο εκπαίδευσης είναι πολύ μικρό, ο θόρυβος δυσκολεύει την εύρεση των πραγματικών μορφών και τάσεων στα δεδομένα, ενώ, αντίθετα, αν το σύνολο εκπαίδευσης είναι πολύ μεγάλο, τότε η μη στασιμότητα των δεδομένων συνεπάγεται χρήση λιγότερο σχετικών δεδομένων για τη δημιουργία της εκτιμήτριας. Με βάση πειράματα που πραγματοποίησαν οι Quek et al. (2008) κατέληξαν στην εκπαίδευση των υποδειγμάτων με τη χρήση δεδομένων 40 ημερών και στη συνέχεια ο πραγματοποιούνταν με τη χρήση δεδομένων των επόμενων 20 ημερών.

Για την επιλογή των χαρακτηριστικών χρησιμοποιήθηκε η Αξιολογική Επιλογή Monte Carlo (Monte Carlo Evaluative Selection), που αποτελεί μία τυχαιοποιημένη μέθοδο δοκιμής και σφάλματος που παρακολουθεί τα αποτελέσματα μίας δράσης, ενεργοποιώντας ή απενεργοποιώντας ένα χαρακτηριστικό, και συγκλίνει σε μία λύση εάν υπάρχουν αρκετά δείγματα εκπαίδευσης. Τα κυρίαρχα χαρακτηριστικά ενισχύονται μέχρι τελικά να είναι τα μόνα χαρακτηριστικά που απομένουν. Το σύνολο των χαρακτηριστικών που εντοπίστηκαν στη συνέχεια παρουσιάζονται στο δίκτυο και υπολογίζεται το αποτέλεσμα.

Οι σταθμίσεις που προέκυψαν από την εκπαίδευση, καταγράφονται μαζί με την ακρίβεια του αποτελέσματος. Μετά από αρκετές ανάλογες επαναλήψεις, τα χαρακτηριστικά που παρέχουν τα πιο ακριβή αποτελέσματα επιλέγονται και αποτελούν τον χώρο χαρακτηριστικών εισόδου. Τα χαρακτηριστικά που επιλέχθηκαν, στη συνέχεια χρησιμοποιούνται για την εκπαίδευση και τον έλεγχο του επαναλαμβανόμενου νευρωνικού δικτύου. Μετά από διάφορους κύκλους του δικτύου, το τελικό σύνολο των σταθμίσεων καθορίζει τα χαρακτηριστικά που επιλέγονται.

Για την πρόβλεψη χρησιμοποιείται μία αρχιτεκτονική επαναλαμβανόμενου δικτύου, παράλληλα με έναν αλγόριθμο μάθησης για την εκπαίδευση των σταθμίσεων. Το δίκτυο χρησιμοποιεί τις τιμές προηγούμενων δεδομένων και τα αποτελέσματα του κρυφού επιπέδου (hidden layer), δημιουργώντας μία μορφή μνήμης, πέρα από τις πληροφορίες που περιέχονται στις τιμές στάθμισης, που αποτελούν το υπόδειγμα της δυναμικής της τιμολόγησης του υποκείμενου

περιουσιακού στοιχείου. Το προτεινόμενο δίκτυο επιχειρεί να συνδυάσει τα χαρακτηριστικά των δικτύων τροφοδότησης προς τα εμπρός (feed-forward network) και ανατροφοδότησης (feedback network) με έναν νέο αλγόριθμο προς τα πίσω διάδοσης σφάλματος (error back propagation) προκειμένου να παρέχει μία νέα δυνατότητα μάθησης.

Οι εισροές στο δίκτυο αποτελούνται από δύο σύνολα δεδομένων. Το πρώτο σύνολο αποτελεί τα δεδομένα που παρέχονται από τον χρήστη και αποτελούνται από n διαδοχικές τιμές, που καθορίζονται από τον μηχανισμό Αξιολογικής Επιλογής Monte Carlo των χαρακτηριστικών που επιλέχθηκαν κατά το πρώτο στάδιο. Το δεύτερο σύνολο αποτελείται από τις εισροές ανατροφοδότησης, δηλαδή τα αποτελέσματα του κρυφού επιπέδου.

Όπως αναφέρουν οι Quek et al. (2008) η μέθοδος εκπαίδευσης που χρησιμοποιείται για τη ρύθμιση του νευρωνικού δικτύου είναι κρίσιμης σημασίας, καθώς καθορίζει την τελική δομή και απόδοση του συστήματος. Καθώς στη μεθοδολογία τους επιθυμούν το δίκτυο να περιγράφει τον μηχανισμό τιμολόγησης ενός περιουσιακού στοιχείου, η εκπαίδευση του γίνεται με επίβλεψη. Η εκπαίδευση θα παρέχει στο δίκτυο πληροφορίες για το πώς η τιμή ανταποκρίνεται σε διαφορετικούς συνδυασμούς των τιμών των παραμέτρων εισόδου. Οι σταθμίσεις στη συνέχεια ανανεώνονται ούτως ώστε να ελαχιστοποιούν τη διαφορά μεταξύ των επιθυμητών και των πραγματικών αποτελεσμάτων.

Σημειώνουν, επίσης, ότι οι παράμετροι του δικτύου θα πρέπει να εξεταστούν και να βελτιστοποιηθούν με προσοχή, καθώς η απόδοση του δικτύου εξαρτάται από τις τιμές αυτών των παραμέτρων. Ταυτόχρονα, η βέλτιστη τιμή αυτών των παραμέτρων διαφέρει ανάλογα με το πεδίο εφαρμογής, για, παράδειγμα, οι βέλτιστες τιμές για την εκπαίδευση του συστήματος για δεδομένα νομισμάτων διαφέρουν από τις τιμές για τα δεδομένα για τον χρυσό.

Οι βασικές παράμετροι που εξετάζονται είναι ο αριθμός n των προηγούμενων εισροών που θα συμπεριληφθούν, όπως καθορίζονται από τον μηχανισμό Αξιολογικής Επιλογής Monte Carlo και επαληθεύονται πειραματικά, ο αριθμός των νευρώνων του κρυφού επιπέδου, που καθορίζονται εμπειρικά καθώς ο καθορισμός τους είναι δύσκολος και η παράμετρος μάθησης α , που καθορίζει τον ρυθμό μεταβολής όλων των σταθμίσεων του δικτύου, που αποτελεί την τιμή που οδηγεί στο μικρότερο σφάλμα, όπως καθορίζεται μέσω μίας διαδικασίας δοκιμής και σφάλματος κατά την

εκπαίδευση του δικτύου. Ακόμη, καθορίζονται οι αρχικές σταθμίσεις στο δίκτυο, που αφορούν και πάλι, όπως και παραπάνω, το σύνολο των σταθμίσεων που παρέχουν το μικρότερο σφάλμα.

Ο αλγόριθμός μάθησης, όπως τονίζουν οι Quek et al. (2008) αποτελεί το σημαντικότερο κομμάτι σε κάθε δίκτυο. Η μεθοδολογία τους προτείνει μία νέα μέθοδο διάδοσης σφάλματος (error propagation) που ταιριάζει στη χρήση χρηματοοικονομικών δεδομένων και ενός επαναλαμβανόμενου νευρωνικού δικτύου. Ο αλγόριθμος μάθησης της μεθοδολογίας τους έχει ως εξής:

Βήμα 1: Αρχικοποίηση

Αριθμός κρυμμένων νευρώνων = h ;

Παράμετρος μάθησης = α ;

data_index = 1;

Τυχαία αρχικοποίηση των σταθμίσεων;

Θέσε $t = 0$; // t = αριθμός επανάληψης

Αριθμός προηγούμενων εισροών δεδομένων = $n - 1$;

Αποτέλεσμα του κρυφού επιπέδου (hidden layer output_ = $y = 0$;

Για $I = 1 : n$

$x[i] = \text{input_data}[\text{data_index}]$;

data_index = data_index + 1;

τέλος

Όπως σημειώνουν οι Quek et al. (2008) κατά τη διάρκεια πειραμάτων διαπίστωσαν ότι το δίκτυο είναι ευαίσθητο στον ρυθμό μάθησης και στις αρχικές σταθμίσεις που ορίζει ο χρήστης. Προκειμένου να αντιμετωπίσουν αυτό το ζήτημα και να θωρακίσουν το δίκτυο απέναντι σε τέτοια ζητήματα, οι τιμές αυτές τυχαιοποιούνται περιοδικά.

Βήμα 2: Ανατροφοδότηση του αποτελέσματος του κρυφού επιπέδου

Για $i = 1 : h$

$$x[n + i] = y[i];$$

Μετά την αρχικοποίηση αυτών των παραμέτρων, τα κανονικοποιημένα δεδομένα εισόδου εισάγονται στο δίκτυο. Η εισροή σε κάθε νευρώνα στο κρυμμένο επίπεδο είναι το άθροισμα των γινομένων των διαφόρων νευρώνων εισροών και των αντίστοιχων σταθμίσεων. Στη συνέχεια, οι συναρτήσεις ενεργοποίησης στους νευρώνες ενεργούν επί αυτών των τιμών. Μία αντίστοιχη διαδικασία ακολουθείται για τον συνδυασμό κρυμμένου επιπέδου και επιπέδου εξόδου.

Βήμα 3: Αποτελέσματα δικτύου για δεδομένο σύνολο δεδομένων εισόδου

$$u = Whi * x;$$

$$y = f(u); // \text{αποτέλεσμα του κρυφού επιπέδου}$$

$$v = Woh * y;$$

$$o = g(v); // \text{αποτέλεσμα του επιπέδου εξόδου}$$

$$\text{Error} = \text{desired} - o;$$

Όπου Whi είναι οι σταθμίσεις μεταξύ του επιπέδου εισόδου και του κρυφού επιπέδου, Who οι σταθμίσεις μεταξύ του κρυμμένου επιπέδου και του επιπέδου εξόδου, $f()$ και $g()$ είναι οι συναρτήσεις ενεργοποίησης στο κρυφό επίπεδο και το επίπεδο εξόδου, αντίστοιχα, $x[]$, $u[]$ και $v[]$ είναι οι εισροές στο δίκτυο, το κρυφό επίπεδο και το επίπεδο εξόδου και $y[]$ και $o[]$ είναι τα αποτελέσματα (outputs) του κρυφού και του επιπέδου εξόδου, αντίστοιχα.

Το τελικό αποτέλεσμα συμβάλλει, επίσης, στην εκτίμηση του σφάλματος στο δίκτυο.

Βήμα 4: Ενημέρωση στάθμισης

$$\text{Έστω } g(v) = \text{desired};$$

$$V' = g^{-1}(\text{desired});$$

Με δεδομένο το τρέχον current y , // εξωτερικό επίπεδο

$$Woh_ideal = v' * y^{-1};$$

$$Woh_diff = Woh_ideal - Woh;$$

Με δεδομένο το τρέχον Who , // κρυμμένο επίπεδο

$$y' = v' * Woh^{-1};$$

$$u' = f^{-1}(y');$$

$$Whi_ideal = u * x^{-1};$$

$$Whi_diff = Whi_ideal - Whi;$$

Η βασική ιδέα του αλγόριθμου είναι η χρησιμοποίηση των τιμών εξόδου για τον υπολογισμό των ιδανικών τιμών του διανύσματος στάθμισης και στη συνέχεια σταδιακά να επιχειρήσει να φέρει τις σταθμίσεις εγγύτερα στον στόχο. Η διαδικασία αυτή επαναλαμβάνεται και για τα δύο σύνολα σταθμίσεων. Η απόκλιση από τις ιδανικές τιμές στάθμισης υπολογίζεται ως ένα θετικό ή αρνητικό σήμα ενίσχυσης για τη συνέχιση ή απόρριψη της ακολουθούμενης τάσης. Μόνο οι σταθμίσεις με τη μεγαλύτερη συνεισφορά στην επιθυμητή τιμή αυξάνονται σταδιακά. Οι νέες σταθμίσεις χρησιμοποιούνται για τις επόμενες επαναλήψεις.

Βήμα 5: Πραγματοποίηση της ενημέρωσης

$$Woh = Woh + \alpha * Woh_diff;$$

$$Whi = Whi + \alpha * Whi_diff;$$

Η διαδικασία αυτή επαναλαμβάνεται μέχρι να επιτευχθεί η σύγκλιση ή να πραγματοποιηθεί ο μέγιστος αριθμός επαναλήψεων που ορίστηκε από τον χρήστη. Η σύγκλιση επιτυγχάνεται όταν το δίκτυο προβλέψει όλες τις τιμές με το απαιτούμενο επίπεδο ακρίβειας.

Βήμα 6: Επανάλαβε μέχρι να επιτευχθεί σύγκλιση

Επανάλαβε για κάθε μοτίβο εκπαίδευσης (training pattern)

$$t = t + 1;$$

για $i = 1 : n-1$

$$x[i] = x[i + 1]; // \text{λάβε τις προηγούμενες } n-1 \text{ συνεχόμενες τιμές}$$

Συνέχισε

$$x[n] = input_data[data_index];$$

$$data_index = data_index + 1;$$

Συνέχισε με τα βήματα 2 – 6

Σημείωσε τις σταθμίσεις με το μικρότερο σφάλμα επανάληψης

τέλος

Οι Quek et al. (2008) αναφέρουν πως υπάρχουν διάφοροι τρόποι με τους οποίους θα μπορούσαν να χρησιμοποιηθούν τα αποτελέσματα του δικτύου, όπως για παράδειγμα για την ανάπτυξη ενός συστήματος συναλλαγών. Ορισμένοι επενδυτές θα μπορούσαν να επωφεληθούν από το σήμα για αγορά ή πώληση, ενώ άλλοι μπορεί να επιθυμούν μία εκτίμηση για την πιθανή διαφορά της τιμής αγοράς και της τιμής πώλησης. Οι επενδυτές θα μπορούσαν να χρησιμοποιήσουν τις πληροφορίες του δικτύου σχετικά με την τιμή για να λάβουν μία θέση με βάση την πρόβλεψη για τη μελλοντική εξέλιξη των τιμών και να αποκομίσουν κέρδη από τις διακυμάνσεις της αγοράς. Άλλοι επενδυτές θα μπορούσαν να υιοθετήσουν μία περισσότερο συντηρητική στρατηγική και να επενδύσουν μόνο σε δικαιώματα προαίρεσης, με μειωμένη απόδοση συγκριτικά με μία άμεση στρατηγική αλλά και μειωμένη έκθεση σε μεταβολές της τιμής και ελαχιστοποίηση των απωλειών εξαιτίας απρόσμενων και απρόβλεπτων γεγονότων.

Το νευρωνικό δίκτυο προβλέπει την τιμή του υποκείμενου περιουσιακού στοιχείου για τις επόμενες 3 εβδομάδες και στη συνέχεια χρησιμοποιούνται τα δεδομένα για την επιλογή των σημείων στα οποία θα πραγματοποιηθούν συναλλαγές. Το σύστημα συναλλαγών πρώτα καθορίζει τα ανώτατα και τα κατώτατα όρια των τιμών των περιουσιακών στοιχείων για την περίοδο και στη συνέχεια υιοθετεί μία επενδυτική στρατηγική με βάση τις συνθήκες, λαμβάνοντας μία θέση ανάλογα με την τάση της τιμής του περιουσιακού στοιχείου. Αν η τιμή του περιουσιακού στοιχείου αναμένεται να μειωθεί το προσεχές χρονικό διάστημα, ο επενδυτής αγοράζει δικαιώματα αγοράς και τα ασκεί όταν η τιμή φτάσει στο επιθυμητό χαμηλότερο επίπεδο. Η χρήση δικαιωμάτων γίνεται για την ελαχιστοποίηση του κινδύνου από απρόβλεπτες μεταβολές της τιμής. Εάν η μεταβολή της τιμής του περιουσιακού στοιχείου δεν είναι προς την επιθυμητή κατεύθυνση, οι απώλειες για τον επενδυτή θα αφορούν τις απώλειες από την επένδυση για την αγορά δικαιωμάτων, σημαντικά μικρότερες από το εάν είχε επενδύσει άμεσα στο περιουσιακό στοιχείο.

Ο έλεγχος της μεθοδολογίας των Quek et al. (2008) έγινε με τη χρήση δεδομένων για συμβόλαια δικαιωμάτων για τον χρυσό και την ισοτιμία Βρετανικής Στερλίνας – Αμερικανικού Δολαρίου για την περίοδο μεταξύ 2000 και 2002. Ακόμη, δεδομένα που χρησιμοποιήθηκαν για τον εμπειρικό έλεγχο της μεθοδολογίας αφορούν τις τιμές κλεισίματος συμβολαίων μελλοντικής εκπλήρωσης

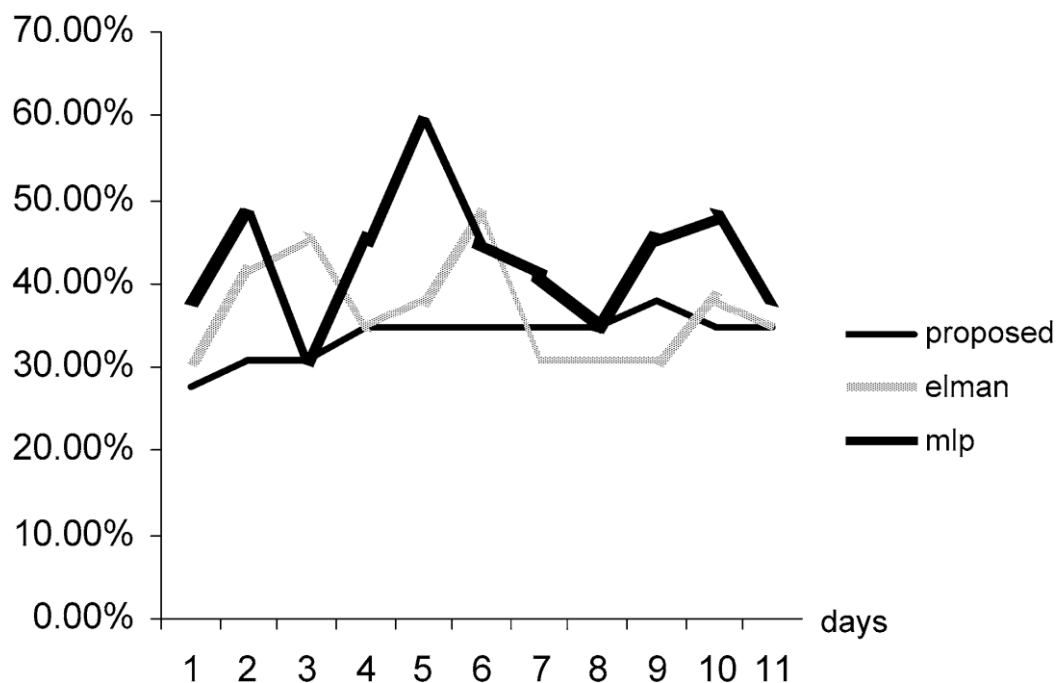
της Βρετανικής Στερλίνας έναντι του Αμερικανικού Δολαρίου, και οι ημερήσιες τιμές αγοράς και πώλησης δικαιωμάτων Αμερικανικού τύπου για την περίοδο μεταξύ Οκτωβρίου του 2002 και Ιουνίου του 2003. Συνολικά, ο έλεγχος έγινε με 792 δείγματα για τα δικαιώματα αγοράς, που περιλαμβάνουν τις ιστορικές τιμές για τα δικαιώματα αγοράς με πέντε διαφορετικές τιμές άσκησης.

Συγκρίνουν το προτεινόμενο δίκτυο και τον αλγόριθμο μάθησης με το δίκτυο Multi-Layer Perceptron (MLP) feed-forward, το επαναλαμβανόμενο δίκτυο (recurrent network) Elman και το RSOP fuzzy neural network. Τα δίκτυα αυτά αποτελούν τυπικά δίκτυα για την αντίστοιχη αρχιτεκτονική και, τις περισσότερες περιπτώσεις, προσφέρουν αξιόπιστα αποτελέσματα. Τα δίκτυα συγκρίνονται βάσει της μελλοντικής τάσης της αγοράς, με το κάθε δίκτυο να προβλέπει εάν τη τιμή θα αυξηθεί ή θα μειωθεί συγκριτικά με την προηγούμενη μέρα, και την πραγματική τιμή του περιουσιακού στοιχείου.

Το δίκτυο MLP υλοποιήθηκε με τη χρήση του MATLAB και η απόδοση του ήταν 48%. Το δίκτυο Elman αποτελεί μία τυπική υλοποίηση ενός επαναλαμβανόμενου δικτύου και σε κάθε επανάληψη παρουσιάζεται στο δίκτυο ολόκληρη η αλληλουχία εισόδου, υπολογίζονται τα αποτελέσματα και συγκρίνονται με την αλληλουχία στόχου για τη δημιουργία ενός προς τα πίσω διαδιδόμενου σφάλματος (back-propagated error) για να βρεθούν οι βαθμίδες σφάλματος που θα χρησιμοποιηθούν για την ενημέρωση των σταθμίσεων. Το δίκτυο Elman είχε καλύτερη απόδοση από το δίκτυο MLP και ήταν σε θέση να προβλέψει τη χρονοσειρά με ακρίβεια της τάξης του 65%. Εξετάστηκε και η απόδοση ενός δικτύου Hopfield με τη χρήση του κανόνα μάθησης Hebbian, αλλά ακόμα και μετά από πολλαπλές επαναλήψεις εκπαίδευσης, η ακρίβεια του ήταν αρκετά χαμηλή και έφτανε στο 30%. Για το δίκτυο RSOP χρησιμοποιήθηκαν 100 επαναλήψεις, με ρυθμό μάθησης 0,2. Το συγκεκριμένο δίκτυο ήταν το ταχύτερο μεταξύ των δικτύων που εξετάστηκαν. Αν και η απόδοση του ήταν ιδιαίτερα ικανοποιητική κατά τη φάση της εκπαίδευσης, η ικανότητα πρόβλεψης του δεν ήταν ιδιαίτερα μεγάλη. Στα διαγράμματα που ακολουθούν, παρουσιάζονται τα αποτελέσματα της σύγκρισης μεταξύ των διαφόρων δικτύων.

Στο επόμενο διάγραμμα παρουσιάζεται το σφάλμα εκπαίδευσης για την πρόβλεψη της τάσης μεταξύ των διαφόρων δικτύων.

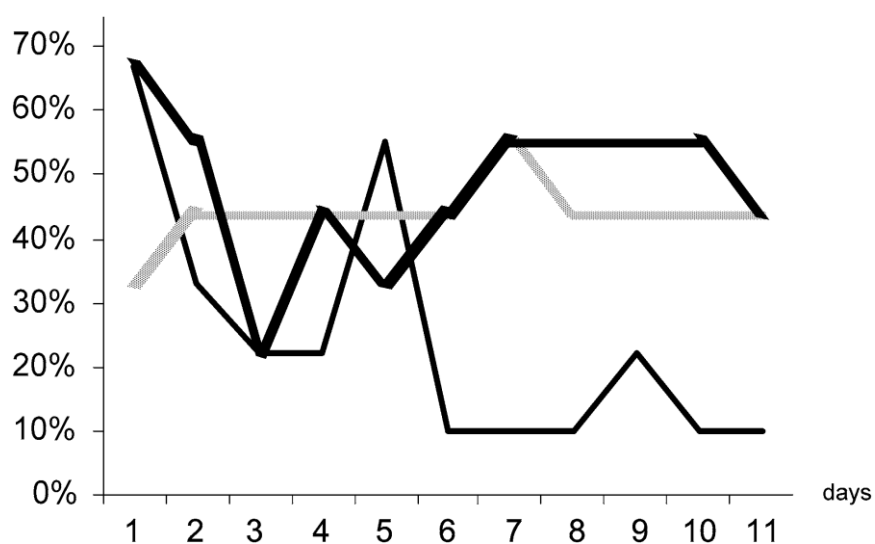
Διάγραμμα 6.1. Το σφάλμα εκπαίδευσης για την πρόβλεψη της τάσης μεταξύ των διαφόρων δικτύων στην έρευνα των Quek et al. (2008)



Πηγή: Quek et al. (2008)

Στο παρακάτω διάγραμμα παρουσιάζεται το σφάλμα ελέγχου για την πρόβλεψη της τάσης μεταξύ των διαφόρων δικτύων.

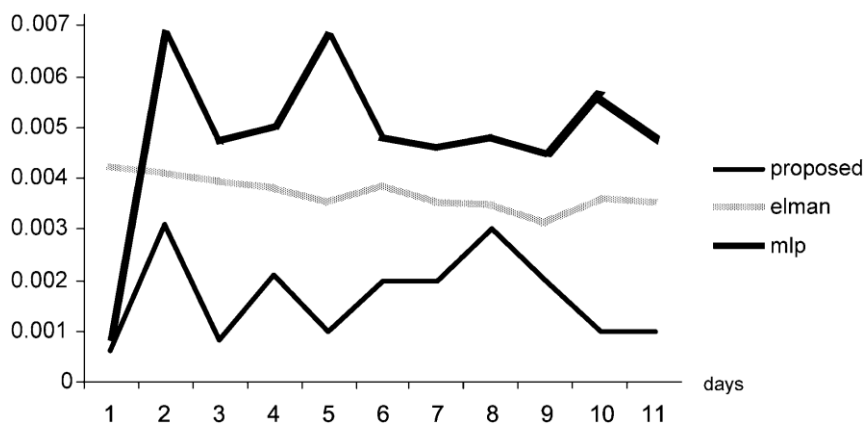
Διάγραμμα 6.2. Το σφάλμα ελέγχου για την πρόβλεψη της τάσης μεταξύ των διαφόρων δικτύων στην έρευνα των Quek et al. (2008)



Πηγή: Quek et al. (2008)

Στο επόμενο διάγραμμα παρουσιάζεται το σφάλμα εκπαίδευσης για την πρόβλεψη της τιμής μεταξύ των διαφόρων δικτύων.

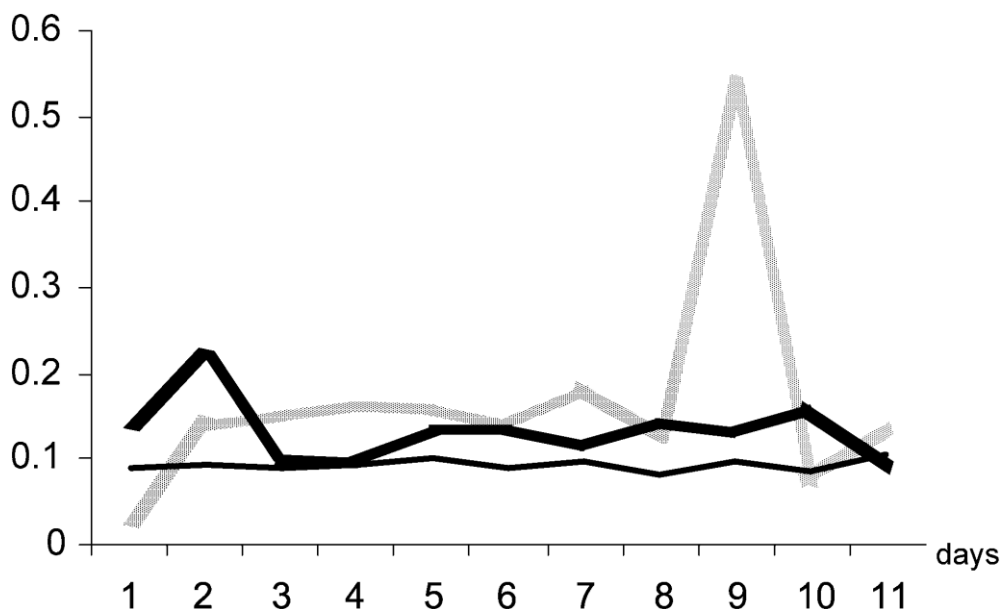
Διάγραμμα 6.3. Το σφάλμα εκπαίδευσης για την πρόβλεψη της τιμής μεταξύ των διαφόρων δικτύων στην έρευνα των Quek et al. (2008)



Πηγή: Quek et al. (2008)

Στο επόμενο διάγραμμα παρουσιάζεται το σφάλμα ελέγχου για την πρόβλεψη της τιμής μεταξύ των διαφόρων δικτύων.

Διάγραμμα 6.4. Το σφάλμα ελέγχου για την πρόβλεψη της τιμής μεταξύ των διαφόρων δικτύων στην έρευνα των Quek et al. (2008)



Πηγή: Quek et al. (2008)

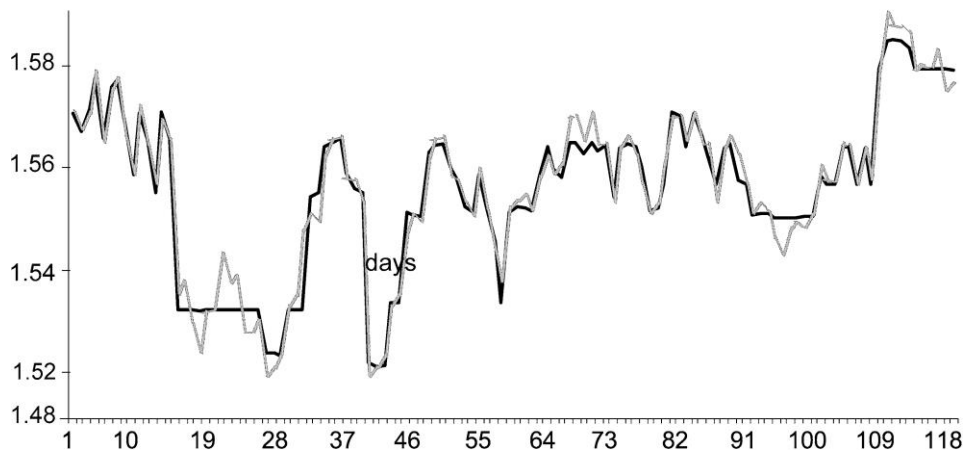
Όπως σημειώνουν οι Quek et al. (2008) το δίκτυο που προτείνουν προσφέρει σημαντικά καλύτερη απόδοση συγκριτικά με τα άλλα δίκτυα.

Στη συνέχεια, εξετάζουν τα αποτελέσματα της μεθοδολογίας τους για διάφορα χρεόγραφα. Το σύστημα εξετάζει την τάση της αγοράς, με το 1 να υποδηλώνει αύξηση της τιμής και το 0 μείωση της τιμής, και την πραγματική τιμή του περιουσιακού στοιχείου. Με βάση τα ευρήματα της έρευνας, το προτεινόμενο δίκτυο ανταποκρίνεται ιδιαίτερα καλά στα πραγματικά δεδομένα της αγοράς, τόσο για τον χρυσό όσο και για το αργό πετρέλαιο. Εξετάζοντας το δίκτυο για διάφορες χρονικές περιόδους, προκειμένου να εντοπίσουν τη βέλτιστη περίοδο εκπαίδευσης και ελέγχου του υποδείγματος, οι Quek et al. (2008) καταλήγουν πως οι βέλτιστες χρονικές περίοδοι είναι οι 40 μέρες για την εκπαίδευση, με τις προβλεπόμενες τιμές να βρίσκονται πολύ κοντά στις πραγματικές, και οι 20 ημέρες για τον έλεγχο του δικτύου.

Αναφορικά με τη χρησιμότητα του δικτύου για επενδυτικούς σκοπούς, τα ευρήματα των Quek et al. (2008) καταδεικνύουν ότι για την περίπτωση του χρυσού το δίκτυο είναι σε θέση να ακολουθήσει την αγοραία τιμή με σχετική ακρίβεια και να αναδείξει την τάση που κυριαρχεί, με τη συσχέτιση να ανέρχεται στο 89%. Για την περίπτωση της Βρετανικής λίρας οι προβλέψεις είναι ακόμα πιο ακριβείς, με τη συσχέτιση να αυξάνεται στο 94%, εξαιτίας του γεγονότος ότι η Βρετανική λίρα παρουσιάζει μικρότερες μεταβολές. Όπως αναφέρουν οι Quek et al. (2008) το δίκτυο είναι σε θέση να βοηθήσει τους επενδυτές να προβλέψουν τη μελλοντική τάση της αγοράς και να λάβουν θέσεις με βάση τη διαφορά μεταξύ τιμής αγοράς και τιμής πώλησης. Η συσχέτιση ανέρχεται στο 92% και η μέση απόδοση της επένδυσης φτάνει στο 4,76% για το δικαίωμα προαίρεσης μεταξύ Βρετανικής λίρας και Αμερικανικού δολαρίου.

Στον πίνακα που ακολουθεί παρουσιάζονται οι τιμές που προβλέπει η μεθοδολογία των Quek et al. (2008) έναντι των πραγματικών τιμών για τα δικαιώματα άσκησης του χρυσού.

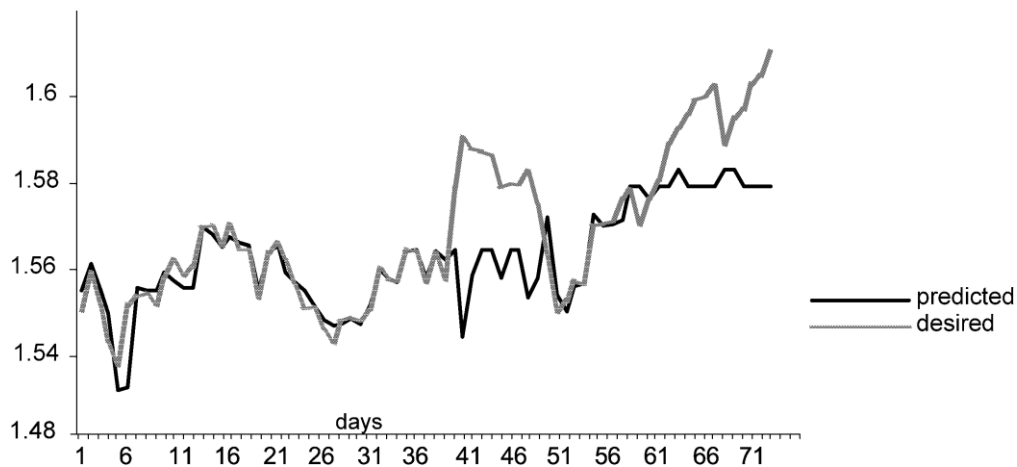
Διάγραμμα 6.5. Οι τιμές που προβλέπει η μεθοδολογία των Quek et al. (2008) έναντι των πραγματικών τιμών για τα δικαιώματα άσκησης του χρυσού



Πηγή: Quek et al. (2008)

Αντίστοιχα, στον πίνακα που ακολουθεί παρουσιάζονται οι τιμές που προβλέπει η μεθοδολογία των Quek et al. (2008) έναντι των πραγματικών τιμών για τα δικαιώματα του αργού πετρελαίου.

Διάγραμμα 6.6. Οι τιμές που προβλέπει η μεθοδολογία των Quek et al. (2008) έναντι των πραγματικών τιμών για τα δικαιώματα του αργού πετρελαίου



Πηγή: Quek et al. (2008)

Οι Quek et al. (2008) καταλήγουν πως η μεθοδολογία που προτείνουν παρουσιάζει ικανοποιητική απόδοση συγκριτικά με τα εναλλακτικά δίκτυα, καθώς είναι σε θέση να προβλέψει τη μεταβολή της τιμής στην αγορά για την επόμενη μέρα με ακρίβεια που φτάνει το 90%. Ένας επενδυτής μπορεί να αξιοποιήσει τις προβλέψεις του δικτύου για να σχεδιάσει ένα χαρτοφυλάκιο λαμβάνοντας τις σωστές θέσεις προκειμένου να εξασφαλίσει κέρδη.

7. ΕΦΑΡΜΟΓΕΣ ΣΤΟΝ ΚΤΗΜΑΤΟΜΕΣΙΤΙΚΟ ΤΟΜΕΑ

Η πρόβλεψη της τιμής των κατοικιών και των επαγγελματικών χώρων είναι καθοριστικής σημασίας για τον κτηματομεσιτικό τομέα, αλλά και τη γενικότερη οικονομική δραστηριότητα. Σε περιόδους οικονομικής αβεβαιότητας, οι κατασκευαστικές εταιρείες αντιμετωπίζουν το δίλημμα για το αν θα πρέπει να αναπτύξουν οικοδομική δραστηριότητα ή όχι. Οι Rafiei και Adeli (2016) αναπτύσσουν ένα υπόδειγμα το οποίο συμβάλλει στην πρόβλεψη της τιμής των νέων κτηρίων κατά τη φάση του σχεδιασμού και τα αρχικά στάδια της κατασκευής του, το οποίο ενσωματώνει έναν μεγάλο αριθμό παραγόντων, μέσω της χρήσης ενός νέου υποδείγματος μηχανικής μάθησης γνωστού ως βαθύς περιορισμένου μηχανήματος Boltzmann (deep restricted Boltzmann machine – DRBM) και ενός γενετικού αλγόριθμου.

Όπως αναφέρουν οι Rafiei και Adeli (2016), το υπόδειγμα μπορεί να χρησιμοποιηθεί από τις κατασκευαστικές εταιρείες ώστε να εκτιμήσουν την τιμή αγοράς πριν την κατασκευή ενός νέου κτηρίου και να αποφασίσουν τελικά αν θα είναι κερδοφόρα η κατασκευή του. Καθώς ο αριθμός των μεταβλητών εισόδου ενός τέτοιου υποδείγματος μπορεί να είναι ιδιαίτερα μεγάλος, παρουσιάζουν και τεχνικές βελτιστοποίησης για τη μείωση του αριθμού των μεταβλητών.

Στο υπόδειγμα υπάρχουν δύο σύνολα μεταβλητών εισόδου και ένα σύνολο μεταβλητών εξόδου. Το πρώτο σύνολο μεταβλητών εισόδου αποτελείται από I στοιχεία που αντιπροσωπεύουν τις δομικές και χρηματοοικονομικές ιδιότητες της αγοράς ακινήτων. Το δεύτερο σύνολο μεταβλητών εισόδου αποτελείται από T υποσύνολα που αντιπροσωπεύουν τις T περιόδους πριν την κατασκευή του κτηρίου. Η περίοδος μπορεί να είναι ένα τρίμηνο, ένας μήνας, ή μία εβδομάδα ανάλογα με τα διαθέσιμα δεδομένα. Κάθε υποσύνολο περιλαμβάνει J μεταβλητές εισόδου που αντιπροσωπεύουν βασικές οικονομικές μεταβλητές που επηρεάζουν την τιμή του κτηρίου. Η μεταβλητή εξόδου είναι η τιμή του κτηρίου κατά τη στιγμή της ολοκλήρωσης του. Ο αριθμός και ο τύπος των μεταβλητών εισόδου μπορεί να διαφέρει ανάλογα με τον τύπο του κτηρίου, αλλά και την περιοχή ανέγερσης του. Μία ενδεικτική λίστα μεταβλητών που αναφέρουν οι Rafiei και Adeli (2016) περιλαμβάνει από την τοποθεσία του ακινήτου και τη συνολική του έκταση, μέχρι και τη διάρκεια της κατασκευής του και την τιμή ανά τετραγωνικό κατά τη στιγμή έναρξης της κατασκευής του.

Όπως σημειώνουν οι Rafiei και Adeli (2016) ο συνολικός αριθμός των μεταβλητών μπορεί να είναι ιδιαίτερα μεγάλος, και να ξεπερνά ακόμα και τις 100. Ένας τόσο μεγάλος αριθμός

μεταβλητών θα απαιτούσε έναν ακόμα μεγαλύτερο αριθμό δειγμάτων εκπαίδευσης, που θα έφτανε τις χιλιάδες, γεγονός που θα καθιστούσε τη διαδικασία συλλογής των δεδομένων ιδιαίτερα δύσκολη έως και αδύνατη και τον υπολογισμό του υποδείγματος ιδιαίτερα απαιτητικό σε πόρους. Το πρόβλημα αυτό είναι γνωστό ως «η κατάρα των διαστάσεων» (dimensionality curse). Προκειμένου να αντιμετωπίσουν την «κατάρα των διαστάσεων» οι Rafiei και Adeli (2016) μειώνουν τον αριθμό των εισροών, ώστε να περιλαμβάνει έναν αρκετά μικρότερο αριθμό μεταβλητών, με βάση τις μεταβλητές με τη μεγαλύτερη επίδραση που προσφέρουν τα αποτελέσματα με τη μεγαλύτερη ακρίβεια.

Ο αριθμός των πιθανών συνδυασμών K μεταβλητών εισόδου από $T \times J$ μεταβλητές εισόδου δίνεται από:

$$W(T \times J, K) = \frac{(T \times J)!}{(T \times J - K)! \times K!}$$

Για τον καθορισμό των μεταβλητών με τη μεγαλύτερη επιρροή που προσφέρουν και τα αποτελέσματα με τη μεγαλύτερη ακρίβεια, οι Rafiei και Adeli (2016) αναπτύσσουν έναν μοναδικό γενετικό αλγόριθμο μη ζευγαρώματος (nonmating genetic algorithm). Ο αλγόριθμος λειτουργεί ως εξής. Αρχικά, όλα τα διαθέσιμα δεδομένα διαχωρίζονται σε δεδομένα εκπαίδευσης και δεδομένα ελέγχου τυχαία, με βάση έναν προκαθορισμένο λόγο δεδομένων ελέγχου προς δεδομένα εκπαίδευσης. Στη συνέχεια η πρώτη γενιά χρωμοσωμάτων του αλγόριθμου δημιουργούνται τυχαία. Κάθε χρωμόσωμα στον πληθυσμό είναι ένας δυαδικός αριθμός, με $T \times J$ bits. Τα χρωμοσώματα με ανάλογη δομή χαρακτηρίζονται ως αποδεκτά χρωμοσώματα. Σε ένα αποδεκτό χρωμόσωμα όταν το i -οστό bit μεταξύ των $T \times J$ bits λαμβάνει την τιμή 1 καταδεικνύει την επιλογή της i -οστής εισροής, είτε αφορά το σύνολο ελέγχου, είτε αφορά το σύνολο εκπαίδευσης. Η διαδικασία αυτή αναφέρεται ως διαδικασία μετάφρασης και χρησιμοποιείται για όλα τα δείγματα εκπαίδευσης και ελέγχου. Η διαδικασία αυτή μειώνει τον αριθμό των μεταβλητών εισόδου για τα δεδομένα εκπαίδευσης και τα δεδομένα ελέγχου σε μία χαμηλότερη διάσταση, περιορίζοντας τον αριθμό τους.

Στη συνέχεια το μηχάνημα Boltzmann εκπαιδεύεται για τόσες φορές όσες είναι και ο αριθμός των χρωμοσωμάτων στον πληθυσμό που επιλέχθηκε, χρησιμοποιώντας κάθε φορά το σύνολο δεδομένων εκπαίδευσης που αντιστοιχεί στο εκάστοτε χρωμόσωμα στον πληθυσμό, με το μηχάνημα να εκπαιδεύεται για κάθε χρωμόσωμα. Στη συνέχεια, τα δεδομένα ελέγχου για κάθε

χρωμόσωμα εφαρμόζονται στο εκπαιδευμένο μηχάνημα Boltzmann και υπολογίζεται το ελάχιστο τετραγωνισμένο σφάλμα των δειγμάτων ελέγχου για κάθε χρωμόσωμα. Η ελάχιστη τιμή των ελάχιστων τετραγωνισμένων σφαλμάτων είναι η τιμή υγείας (fitness) της συνάρτησης υγείας του γενετικού αλγόριθμου. Η συνάρτηση υγείας είναι μία συνάρτηση ένα-προς-ένα που δίνεται από:

$$Y = g(x)|x \in C$$

Όπου x = το αποδεκτό χρωμόσωμα, $g(x)$ η συνάρτηση υγείας, Y το μέσο ελάχιστο τετραγωνισμένο σφάλμα του αποδεκτού χρωμοσώματος και C το σύνολο όλων των αποδεκτών χρωμοσωμάτων.

Στη συνέχεια δημιουργείται μία νέα γενιά χρωμοσωμάτων.

Οι Rafiei και Adeli (2016) εξετάζουν εμπειρικά το υπόδειγμα τους με τη χρήση δεδομένων για 360 συγκροτήματα κατοικιών στην Τεχεράνη για την περίοδο μεταξύ 1993 και 2008. Το πρώτο σύνολο μεταβλητών περιλαμβάνει μία σειρά οικονομικών δεδομένων και δεδομένων σχετικά με την οικοδομική δραστηριότητα. Ο αριθμός των υποσυνόλων στο δεύτερο σύνολο δεδομένων τέθηκε σε 5, δηλαδή 5 τρίμηνα πριν την ημερομηνία έναρξης κατασκευής του έργου. Το δεύτερο σύνολο δεδομένων περιλαμβάνει 95 μεταβλητές εισόδου, με τον συνολικό αριθμό μεταβλητών να ανέρχεται σε 102. Στη συνέχεια, ο αριθμός των μεταβλητών στο δεύτερο σύνολο μειώνεται σε 20, με τον συνολικό αριθμό μεταβλητών να μειώνεται σε 27 μεταβλητές για 360 δείγματα.

Ο αριθμός τρεξιμάτων του γενετικού αλγόριθμου ορίζεται σε 200, με 40.000 υποψήφια χρωμοσώματα και, κατά συνέπεια, σε κάθε λειτουργία του αλγόριθμου εξετάζονται 8.000.000 πιθανά χρωμοσώματα. Το υπολογιστικό υπόδειγμα υλοποιήθηκε στο MATLAB. Το καλύτερο χρωμόσωμα είναι το χρωμόσωμα με αριθμό διανύσματος 87 (**B**) για λόγο δεδομένων ελέγχου προς δεδομένα εκπαίδευσης της τάξης του 10% και μέσο τετραγωνισμένο σφάλμα 0,0029. Με βάση το χρωμόσωμα στο υπόδειγμα περιλαμβάνονται 19 μεταβλητές που αφορούν οικονομικά και οικοδομικά στοιχεία. Με βάση τα αποτελέσματα της εμπειρικής έρευνας, οι Rafiei και Adeli (2016) σημειώνουν πως οι μεταβλητές που αφορούν στοιχεία κοντά στην ημερομηνία έναρξης κατασκευής του έργου επηρεάζουν περισσότερο την αγορά ακινήτων και την απόφαση για την κατασκευή ή μη του έργου συγκριτικά με εκείνες που αφορούν στοιχεία σε προγενέστερη χρονική περίοδο.

Από το σύνολο 360 δειγμάτων για την περίοδο μεταξύ 1993 και 2008 επελέγησαν τυχαία 10 δείγματα για τον έλεγχο του υποδείγματος και τα υπόλοιπα 350 χρησιμοποιήθηκαν για την εκπαίδευση του. Το αρχικό δείγμα περιλαμβάνει 102 χαρακτηριστικά και, με βάση το καλύτερο χρωμόσωμα με τα σημαντικότερα χαρακτηριστικά, ο αριθμός των μεταβλητών μειώθηκε σε 27, με την διαδικασία αυτή να επαναλαμβάνεται 100 φορές. Σύμφωνα με τα αποτελέσματα, το μέσο σφάλμα ανέρχεται σε 5,7%, με βάση τις στρατηγικές που χρησιμοποιήθηκαν για τη μείωση των διαστάσεων και 2 περιόδους εκπαίδευσης, έναντι σφάλματος της τάξης του 17,4% χωρίς τη χρήση των στρατηγικών για τη μείωση των διαστάσεων του υποδείγματος.

Για λόγους σύγκρισης, η τιμή εκτιμήθηκε και με τη χρήση του νευρωνικού δικτύου feedforward backpropagation, με ένα κρυμμένο επίπεδο και 10 κόμβους στο κρυμμένο επίπεδο. Μετά από δύο περιόδους εκπαίδευσης, το σφάλμα για το μηχάνημα Boltzmann ανέρχεται σε 5,7% και για το νευρωνικό δίκτυο feedforward backpropagation σε 56,8%. Συνεπώς, όπως σημειώνουν οι Rafiei και Adeli (2016), το μηχάνημα Boltzmann προσφέρει αρκετά ακριβή αποτελέσματα με μόνο δύο περιόδους εκπαίδευσης, που απαιτούν μόλις 16,1 δευτερόλεπτα, ενώ το νευρωνικό δίκτυο feedforward backpropagation απαιτεί 1820 περιόδους εκπαίδευσης με χρόνο εκπαίδευσης που ανέρχεται σε 4.073,9 δευτερόλεπτα για να επιτύχει το ίδιο σφάλμα ελέγχου, της τάξης του 4,7%, όπως το μηχάνημα Boltzmann. Συνεπώς, ως εκτιμήτρια είναι πιο αργή κατά 253 φορές συγκριτικά με το μηχάνημα Boltzmann για να επιτευχθεί το ίδιο σφάλμα ελέγχου. Κατά συνέπεια, το μηχάνημα Boltzmann αποτελεί έναν σαφώς ταχύτερο και ακριβέστερο αλγόριθμο μηχανικής μάθησης συγκριτικά με το νευρωνικό δίκτυο feedforward backpropagation.

Για την επίλυση του συνολικού υποδείγματος πρόβλεψης της τιμής των κατασκευών απαιτούνται 4.000 εφαρμογές της εκτιμήτριας του μηχανήματος Boltzmann ή του νευρωνικού δικτύου feedforward backpropagation. Το μηχάνημα Boltzmann χρειάζεται περίπου 18 ώρες για τον υπολογισμό του υποδείγματος, ενώ το νευρωνικό δίκτυο feedforward backpropagation απαιτεί πάνω από 4.530 ώρες.

Παράλληλα, οι Rafiei και Adeli (2016) έλεγξαν και την ακρίβεια του γενετικού αλγόριθμου μη ζευγαρώματος (nonmating genetic algorithm) που προτείνουν με 4 δημοφιλείς μεθοδολογίες επιλογής χαρακτηριστικών, την τυπική γενετική έρευνα (standard genetic search), τη γραμμική επιλογή προς τα εμπρός (linear forward selection) και το υποσύνολο χαρακτηριστικών βάσει της συσχέτισης (correlation-based feature subset), με βάση 3 διαφορετικούς ταξινομητές, τον Αφελή

Bayes (Naïve Bayes), τον ταξινομητή bagging και το μηχάνημα διανύσματος υποστήριξης (support vector machine - SVM). Από τις τυπικές μεθοδολογίες, η μεθοδολογία του υποσυνόλου χαρακτηριστικών βάσει της συσχέτισης παρουσιάζει τα μικρότερα σφάλματα, της τάξης του 10,3% με βάση τον ταξινομητή Naïve Bayes, του 18,1% βάσει του bagging και 64,8% βάσει του SVM. Ο γενετικός αλγόριθμος που προτείνουν οι Rafiei και Adeli (2016), αντίστοιχα, παρουσιάζει σφάλμα της τάξης του 3,7%, συνεπώς, παρουσιάζει σαφώς μεγαλύτερη ακρίβεια συγκριτικά με τις τυπικές εναλλακτικές μεθόδους.

Οι Rafiei και Adeli (2016) καταλήγουν πως το υπόδειγμα τους μπορεί να βοηθήσει της κατασκευαστικές εταιρείες να αποφασίσουν για το αν θα πρέπει να προχωρήσουν στην κατασκευή ενός έργου. Με βάση τη μελλοντική τιμή πώλησης που προβλέπει το μοντέλο, μία κατασκευαστική εταιρεία μπορεί να αποφασίσει την αναβολή ενός έργου, ιδιαίτερα σε περιόδους κάμψης στην αγορά, ή την πώληση των κατοικιών μετά την ολοκλήρωση της κατασκευής τους και όχι πριν, προκειμένου να μεγιστοποιήσει το κέρδος της. Συνεπώς, το υπόδειγμα μπορεί να αποτελέσει χρήσιμο οδηγό για μία κατασκευαστική εταιρεία στην απόφαση της για το αν θα πρέπει να προχωρήσει στην κατασκευή ενός έργου, είτε αφορά κατοικίες, είτε άλλου είδους οικοδομικές κατασκευές, υπολογίζοντας με σχετική ακρίβεια τη μελλοντική τιμή του.

8. ΔΙΑΧΕΙΡΙΣΗ ΠΕΡΙΟΥΣΙΑΚΩΝ ΣΤΟΙΧΕΙΩΝ ΚΑΙ ΥΠΟΧΡΕΩΣΕΩΝ

Η διαχείριση περιουσιακών στοιχείων και υποχρεώσεων αποτελεί ένα από τα διαχρονικά θέματα της διαχείρισης χρηματοοικονομικού κινδύνου και αφορά τη διαχείριση των περιουσιακών στοιχείων ώστε να προσφέρουν ικανοποιητικές αποδόσεις, διατηρώντας, παράλληλα, ένα σημαντικό πλεόνασμα περιουσιακών στοιχείων έναντι των τρεχουσών και μελλοντικών υποχρεώσεων. Το συγκεκριμένο πρόβλημα αφορά ιδιαίτερα τα ασφαλιστικά ταμεία και τις ασφαλιστικές εταιρείες. Η εύρεση βέλτιστων στρατηγικών διαχείρισης περιουσιακών στοιχείων και υποχρεώσεων θέτει σημαντικές υπολογιστικές προκλήσεις και πολλές από τις θεωρητικές προσεγγίσεις απαιτούν ιδιαίτερα σημαντικούς υπολογιστικούς πόρους για να εφαρμοστούν στην πράξη (Gülpinar & Pachamanova, 2013).

Αν και υπάρχουν αναλυτικές λύσεις για ορισμένες βέλτιστες δυναμικές επενδυτικές στρατηγικές, συνήθως στην πράξη χρησιμοποιούνται αριθμητικές μέθοδοι, όπως ο δυναμικός προγραμματισμός, προσομοιώσεις και τεχνικές στοχαστικού προγραμματισμού, για την εύρεση βέλτιστων επενδυτικών κανόνων μέσω της εξέτασης διαφόρων σεναρίων για τις μελλοντικές αποδόσεις των περιουσιακών στοιχείων και υποχρεώσεων μίας επιχείρησης. Αν και αυτές οι τεχνικές είναι εφικτό να χρησιμοποιηθούν στην πράξη, η εφαρμογή τους είναι δύσκολη. Αυτό οφείλεται στο γεγονός ότι η διαχείριση περιουσιακών στοιχείων και υποχρεώσεων αποτελεί ένα πρόβλημα πολλαπλών περιόδων, κατά συνέπεια ο αριθμός των πιθανών σεναρίων που απαιτείται για την ρεαλιστική προσέγγιση της πραγματικότητας αυξάνεται εκθετικά με τον αριθμό των χρονικών περιόδων που εξετάζονται. Συνεπώς, οι διαστάσεις του προβλήματος βελτιστοποίησης και η υπολογιστική του δυσκολία αυξάνονται.

Ταυτόχρονα, η δημιουργία σεναρίων απαιτεί τη χρήση σύνθετων στατιστικών τεχνικών, γεγονός που στην πράξη λειτουργεί αποτρεπτικά, καθώς απαιτείται η λήψη αποφάσεων σε σύντομο χρονικό διάστημα, ενώ και τα στοιχεία αναφορικά με τη μελλοντική κατανομή της αβεβαιότητας μπορεί να μην είναι επαρκή για την κατάλληλη εκτίμηση ενός υποδείγματος (Gülpinar & Pachamanova, 2013).

Οι Gülpınar και Pachamanova (2013) εξετάζουν το ζήτημα της διαχείρισης περιουσιακών στοιχείων και υποχρεώσεων με τη χρήση της ισχυρής βελτιστοποίησης (robust optimization), η οποία παρουσιάζει ομοιότητες με τον δυναμικό και τον στοχαστικό προγραμματισμό. Η ισχυρή βελτιστοποίηση μπορεί να χρησιμοποιηθεί για την επίλυση αντίστοιχων προβλημάτων με τον δυναμικό και τον στοχαστικό προγραμματισμό, επιλύοντας το πρόβλημα βελτιστοποίησης υποθέτοντας ότι η αβεβαιότητα που συνδέεται με τα δεδομένα εισόδου ανήκει σε ένα σύνολο αβεβαιότητας και βρίσκει τη βέλτιστη λύση υποθέτοντας ότι η αβεβαιότητα θα λάβει τις χειρότερες τιμές σε αυτό το σύνολο.

Η μέθοδος της ισχυρής βελτιστοποίησης εξετάζεται με τη χρήση δεδομένων επιχειρήσεων για 80 τρίμηνα, από το πρώτο τρίμηνο του 1988 μέχρι και το τέταρτο τρίμηνο του 2007. Αρχικά εκτιμούν τις παραμέτρους μίας διαδικασίας VAR(1). Εξετάζουν την απόδοση τριών στρατηγικών, του ονομαστικού, του παραδοσιακού και του ισχυρού στοχαστικού προγραμματισμού για διαφορετικές τιμές των παραμέτρων λ , που είναι ο πολλαπλασιαστής ποινής για το ποσό των συνεισφορών, ψ , που είναι ο απαιτούμενος λόγος χρηματοδότησης και θ , που είναι οι προϋπολογισμοί που σχετίζονται με τους περιορισμούς που περιέχουν αβέβαιους συντελεστές. Μόνο στη μεθοδολογία της ισχυρής βελτιστοποίησης απαιτείται η παράμετρος θ .

Για την εκτίμηση των παραμέτρων της διαδικασίας VAR(1) και την προσομοίωση 100 σεναρίων για τις αθροιστικές αποδόσεις χρησιμοποίησαν τα εργαλεία οικονομετρίας του MATLAB. Τα υποδείγματα βελτιστοποίησης υλοποιήθηκαν στο MATLAB με τη χρήση της διεπαφής YAMLP και του SeDuMi. Εξετάζουν 4 στάδια αποφάσεων με 3 μήνες χρονικό περιθώριο μεταξύ τους. Για τον έλεγχο της απόδοσης, τα περιουσιακά στοιχεία καταγράφονται την τέταρτη περίοδο, ενώ υπάρχουν 2 περιουσιακά στοιχεία, ένα αγοραίο, που περιέχει κίνδυνο, και ένα ασφαλές περιουσιακό στοιχείο. Οι υπόλοιπες παράμετροι του υποδείγματος αφορούν τα κόστη συναλλαγών για τις αγορές και τις πωλήσεις, και τίθενται στο 1% του συνολικού ποσού της συναλλαγής, τους μισθούς W_t σε κάθε χρονική περίοδο, που τίθενται στα 1000 δολάρια και οι υποχρεώσεις που θα πρέπει να καταβληθούν σε κάθε χρονικό σημείο, που ορίζονται στα 50 δολάρια. Η αρχική αξία διακρατήσεων μετρητών είναι 0 δολάρια και η αξία του επισφαλούς περιουσιακού στοιχείου είναι 100 δολάρια.

Σύμφωνα με τα ευρήματα της έρευνας, μία αύξηση της τιμής της παραμέτρου λ , διατηρώντας τις άλλες παραμέτρους σταθερές, γενικά οδηγεί σε μείωση της αναμενόμενης αξίας των τελικών περιουσιακών στοιχείων για κάθε μία από τις στρατηγικές που εξετάστηκαν, ενώ το ίδιο ισχύει και για το χειρότερο σενάριο και το πέμπτο εκατοστημόριο. Το εύρημα αυτό αποδίδεται στο γεγονός ότι καθώς μειώνεται το εισόδημα, μειώνονται και οι διαθέσιμες επενδυτικές ευκαιρίες και κατά συνέπεια και η πιθανότητα επίτευξης υψηλών αποδόσεων. Οι τιμές της παραμέτρου ψ δεν επηρεάζουν ιδιαίτερα καμία από τις στρατηγικές. Η παράμετρος θ καθορίζει την αβεβαιότητα και τον βαθμό προστασίας που απαιτεί έναντι της αβεβαιότητας αυτός που λαμβάνει τις αποφάσεις. Μεγαλύτερη τιμή του θ συνεπάγεται ότι αυτός που λαμβάνει τις αποφάσεις απαιτεί προστασία έναντι σημαντικών μεταβολών των παραμέτρων που εμπεριέχουν αβεβαιότητα από τις ονομαστικές τιμές τους και μεγαλύτερες τιμές του θ έχουν ως αποτέλεσμα μεγαλύτερα επίπεδα διαφοροποίησης του χαρτοφυλακίου. Όταν το θ αυξάνεται ο αναμενόμενος τελικός πλούτος που συνεπάγεται η εφαρμογή της μεθόδου της ισχυρής βελτιστοποίησης μειώνεται.

Συνεπώς, υπάρχει μία αρνητική σχέση μεταξύ της προστασίας από την αβεβαιότητα που απαιτεί ένας επενδυτής με την εφαρμογή της μεθόδου της ισχυρής βελτιστοποίησης και της αναμενόμενης απόδοσης, αν και η σχέση αυτή δεν είναι μονοτονική. Καθώς το θ αυξάνεται ο αναμενόμενος πλούτος αρχικά μειώνεται και στη συνέχεια αυξάνεται ξανά. Το γεγονός αυτό συνεπάγεται ότι η επιλογή της τιμής της παραμέτρου θ θα πρέπει να γίνει με προσοχή, ούτως ώστε να αντικατοπτρίζει τον βέλτιστο κίνδυνο που επιθυμεί να αναλάβει ο επενδυτής.

Όσο πιο μικρές είναι οι τιμές της παραμέτρου θ τόσο μικρότερη είναι η διαφοροποίηση του χαρτοφυλακίου, ενώ η αύξηση της τιμής του θ συνεπάγεται και αύξηση της διαφοροποίησης του χαρτοφυλακίου. Για μικρές τιμές της παραμέτρου λ , μεταξύ 0,01 και 0,50, η ισχυρή βελτιστοποίηση υπερέχει του στοχαστικού προγραμματισμού, καθώς προσφέρει μεγαλύτερο αναμενόμενο τελικό πλούτο, μεγαλύτερο ελάχιστο, μέγιστο και διάμεσο τελικό πλούτο και χαμηλότερα κόστη συναλλαγών.

Για μεγαλύτερες, όμως, τιμές του λ , ο γραμμικός προγραμματισμός προσφέρει μεγαλύτερο αναμενόμενο τελικό πλούτο, αν και η ισχυρή βελτιστοποίηση εξακολουθεί να υπερέχει ως προς την τυπική απόκλιση του τελικού πλούτου και τα κόστη συναλλαγών. Οι Gülpınar και Pachamanova (2013) αποδίδουν τα ευρήματα αυτά στο γεγονός ότι τα σενάρια που επιλέχθηκαν

για υπολογιστικούς λόγους ευνοούν την προσέγγιση του στοχαστικού προγραμματισμού, καθώς είναι περιορισμένα και περιέχουν ιδιαίτερα αναλυτικά στοιχεία αναφορικά με το μέλλον. Στην πράξη, όμως, τα σενάρια για την εξέλιξη της αβεβαιότητας δεν είναι πάντα γνωστά. Η ισχυρή βελτιστοποίηση είναι ιδιαίτερα αποτελεσματική υπολογιστικά συγκριτικά με τις προσεγγίσεις στοχαστικού προγραμματισμού, καθώς ενώ για την εύρεση της βέλτιστης λύσης ενός υποδείγματος στοχαστικού προγραμματισμού απαιτούνται ώρες, ένα υπόδειγμα ισχυρής βελτιστοποίησης μπορεί να επιλυθεί σε μερικά δευτερόλεπτα για μεγάλο αριθμό περιουσιακών στοιχείων και χρονικών περιόδων. Καθώς ο αριθμός των σεναρίων, των περιουσιακών στοιχείων και των χρονικών περιόδων ενός υποδείγματος αυξάνεται, το μέγεθος του υποδείγματος στοχαστικού προγραμματισμού αυξάνεται εκθετικά, ενώ το υπόδειγμα ισχυρής βελτιστοποίησης αυξάνεται γραμμικά.

Συνεπώς, η ισχυρή βελτιστοποίηση προσφέρει τη δυνατότητα αποτελεσματικής και γρήγορης επίλυσης ενός προβλήματος διαχείρισης περιουσιακών στοιχείων και υποχρεώσεων. Η μέθοδος αυτή όμως αποδίδει καλύτερα όταν δεν είναι γνωστή η διαδικασία που ακολουθούν οι αποδόσεις των περιουσιακών στοιχείων. Όταν ο επενδυτής θέλει να εξετάσει συγκεκριμένα σενάρια, η μέθοδος του στοχαστικού προγραμματισμού μπορεί να αποφέρει καλύτερα αποτελέσματα, ενώ για την αποτελεσματική εκτίμηση ενός υποδείγματος ισχυρής βελτιστοποίησης απαιτείται και η προσεκτική επιλογή της τιμής της παραμέτρου θ , καθώς μία ακατάλληλη τιμή μπορεί να έχει αρνητικές επιπτώσεις στην απόδοση του υποδείγματος.

9. ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην παρούσα διπλωματική εργασία εξετάστηκε η επιστήμη των δεδομένων και οι εφαρμογές της στις χρηματοοικονομικές υπηρεσίες. Όπως κατέδειξε η ανάλυση που προηγήθηκε, η επιστήμη των δεδομένων συναντά πληθώρα εφαρμογών στις χρηματοοικονομικές υπηρεσίες, από τη βέλτιστη επιλογή χαρτοφυλακίου μετοχών και τη διαχείριση κρατικού και εταιρικού χρέους, μέχρι τον εντοπισμό περιπτώσεων απάτης, την εκτίμηση της αξίας παραγώγων και υποκείμενων περιουσιακών στοιχείων και εφαρμογές στον κτηματομεσιτικό τομέα.

Έτσι, η επιστήμη των δεδομένων μπορεί να συμβάλλει στην κατάλληλη επιλογή ενός χαρτοφυλακίου μετοχών που προσφέρει ικανοποιητικές αποδόσεις, μεγαλύτερες συγκριτικά με τις παραδοσιακές μεθόδους επιλογής χαρτοφυλακίου μετοχών, με τη χρήση μεθόδων που είναι αποτελεσματικές και μπορούν να εφαρμοστούν για μεγάλο αριθμό μετοχών και προσφέρουν τη δυνατότητα διαφοροποίησης κινδύνου, καθώς είναι σε θέση να εντοπίσουν τη συσχέτιση μεταξύ των μετοχών. Ακόμη, στο ίδιο πλαίσιο, οι εφαρμογές της επιστήμης των δεδομένων είναι σε θέση να συμβάλλουν στην αξιολόγηση της σχετικής χρηματοοικονομικής ισχύος των εταιρειών ενός κλάδου και να οδηγήσουν τους επενδυτές και τους διαχειριστές κεφαλαίων στην κατάλληλη απόφαση για το εάν θα πρέπει να επενδύσουν σε έναν κλάδο οικονομικής δραστηριότητας, καθώς και σε ποιες μετοχές εταιρειών του κλάδου. Συνεπώς, οι μέθοδοι που αξιοποιεί η επιστήμη των δεδομένων μπορούν να προσφέρουν πολύτιμα εργαλεία στους επενδυτές, συμβάλλοντας στη βέλτιστη επιλογή μετοχών για τη δημιουργία ενός αποτελεσματικού χαρτοφυλακίου.

Επίσης, με τη χρήση των μεθόδων της επιστήμης των δεδομένων μπορεί να γίνει η βέλτιστη διαχείριση του κρατικού χρέους, με τη χρήση υποδειγμάτων που αναδεικνύουν τα κατάλληλα χρηματοοικονομικά εργαλεία για τη μείωση του κόστους εξυπηρέτησης του δημόσιου χρέους, μέσω της έκδοσης κρατικού χρέους τον κατάλληλο χρόνο, με χρεόγραφα που προσφέρουν χαμηλό κόστος δανεισμού και το ενδεδειγμένο και επιδιωκόμενο προφίλ ωρίμανσης του χρέους. Ταυτόχρονα, οι μέθοδοι της επιστήμης των δεδομένων μπορούν να αποτελέσουν πολύτιμα εργαλεία και για τη διαχείριση του εταιρικού χρέους, συμβάλλοντας στη βέλτιστη έκδοση ομολόγων, ούτως ώστε να ελαχιστοποιείται το μέσο κόστος χρηματοδότησης του χρέους, με τη μόχλευση να παραμένει σε ανεκτά επίπεδα και τον κίνδυνο χρεοκοπίας περιορισμένος.

Η επιστήμη των δεδομένων προσφέρει ακόμη τη δυνατότητα αυτοματοποίησης και αύξησης της αποτελεσματικότητας όσον αφορά τον εντοπισμό περιπτώσεων χρηματοοικονομικής απάτης και μπορεί να προσφέρει χρήσιμα εργαλεία για τις εποπτικές και ελεγκτικές αρχές καθώς και για τους επενδυτές, δίνοντας τους τη δυνατότητα να εντοπίσουν περιπτώσεις απάτης με σχετική ακρίβεια. Συμβάλλοντας στον έγκαιρο εντοπισμό της απάτης, συμβάλλει ουσιαστικά και στην αποτροπή της απάτης. Μέσα από την ανάπτυξη υπολογιστικών υποδειγμάτων, καθίσταται εφικτή η αξιοποίηση όλων των δημόσια διαθέσιμων πληροφοριών για τον εντοπισμό περιπτώσεων απάτης εις βάρος του επενδυτικού κοινού και την παραβίαση του ρυθμιστικού και νομικού πλαισίου για την αποκάλυψη όλων των πληροφοριών για τη χρηματοοικονομική κατάσταση μίας επιχείρησης στις χρηματοοικονομικές της καταστάσεις και αναφορές.

Πρόσθετες εφαρμογές της επιστήμης των δεδομένων αφορούν και την πρόβλεψη της εξέλιξης της τιμής των παραγώγων και των υποκείμενων περιουσιακών στοιχείων, με αποτελεσματικότητα και ακρίβεια, για τη δημιουργία των κατάλληλων επενδυτικών στρατηγικών που θα επιτρέψουν στους επενδυτές να επιτύχουν κέρδη, εκμεταλλευόμενοι τις μεταβολές στην αγορά, ελαχιστοποιώντας παράλληλα τον κίνδυνο απωλειών, αλλά και τις πιθανές απώλειες αν τελικά πραγματοποιηθούν.

Ακόμη, οι μέθοδοι της επιστήμης των δεδομένων συναντούν εφαρμογή και στον κτηματομεσιτικό τομέα, όπου μπορούν να χρησιμοποιηθούν για την πρόβλεψη των τιμών των ακινήτων μέσα από τον υπολογισμό υποδειγμάτων που περιλαμβάνουν μεταβλητές που αφορούν μία σειρά οικονομικών μεταβλητών και μεταβλητών που αφορούν τον κτηματομεσιτικό κλάδο και να προσφέρουν πολύτιμες πληροφορίες στις κατασκευαστικές εταιρείες για την πιθανή κερδοφορία ενός έργου πριν την κατασκευή του, αξιοποιώντας όλα τα διαθέσιμα δεδομένα, και κατά συνέπεια να βοηθήσουν τις κατασκευαστικές εταιρείες να αποφασίσουν για το αν θα πρέπει να προχωρήσουν στην κατασκευή του, ώστε να διασφαλίσουν κέρδη ή να αποφύγουν απώλειες.

Καθώς η πληροφορική τεχνολογία εξελίσσεται, επιτρέπει τη συγκέντρωση ολοένα και μεγαλύτερου όγκου δεδομένων και τη χρήση περισσότερο σύνθετων μεθόδων, τεχνικών και υπολογιστικών υποδειγμάτων που δεν ήταν εφικτό να χρησιμοποιηθούν κατά το πρόσφατο παρελθόν. Έτσι, η επιστήμη των δεδομένων αναμένεται να εξελιχθεί περαιτέρω, συναντώντας όλο και συχνότερα εφαρμογή στον χρηματοοικονομικό τομέα, για την επίλυση διάφορων προβλημάτων και τη δημιουργία των κατάλληλων στρατηγικών για τη μεγιστοποίηση των κερδών και την ελαχιστοποίηση του κινδύνου και των πιθανών απωλειών. Η τεχνολογική πρόοδος

αναμένεται να δώσει περαιτέρω ώθηση στην εξέλιξη και διάδοση της επιστήμης των δεδομένων, ενός κλάδου με σημαντικές προοπτικές και οφέλη.

BIBΛΙΟΓΡΑΦΙΑ

1. V. Babalos, N. Philippas, M. Doumpos, and C. Zopounidis, "Mutual funds performance appraisal using stochastic multicriteria acceptability analysis," *Applied Mathematics and Computation*, vol. 218, no. 9, pp. 5693-5703, 2012.
2. V. Balla, C. Gaganis, F. Pasiouras, and C. Zopounidis, "Multicriteria decision aid models for the prediction of securities class actions: evidence from the banking sector," *OR Spectrum*, vol. 36, pp. 57-72, 2014.
3. M. Bernaschi, M. Briani, M. Papi, and D. Vergni, "Scenario-generation methods for an optimal public debt strategy," *Quantitative Finance*, vol. 7, no. 2, pp. 217-229, 2007.
4. T. G. Calderon and J. J. Cheh, "A roadmap for future neural networks research in auditing and risk assessment," *International Journal of Accounting Information Systems*, vol. 3, no. 4, pp. 203-236, 2002.
5. A. Consiglio and A. Staino, "A stochastic programming model for the optimal issuance of government bonds," *Annals of Operations Research*, vol. 193, pp. 159-172, 2012.
6. N. C. Edirisinghe and X. Zhang, "Generalized DEA model of fundamental analysis and its application to portfolio optimization," *Journal of Banking & Finance*, vol. 31, no. 11, pp. 3311-3335, 2007.
7. C. Gaganis, "Classification techniques for the identification of falsified financial statements: a comparative analysis," *Intelligent Systems in Accounting, Finance & Management: International Journal*, vol. 16, no. 3, pp. 207-229, 2009.
8. F. H. Glancy and S. B. Yadav, "A computational model for financial reporting fraud detection," *Decision Support Systems*, vol. 50, no. 3, pp. 595-601, 2011.
9. N. Gülpinar and D. Pachamanova, "A robust optimization approach to asset-liability management under time-varying investment opportunities," *Journal of Banking & Finance*, vol. 37, no. 6, pp. 2031-2041, 2013.

10. D. Hand, H. Mannila, and P. Smyth, "Principles of Data Mining," *MIT Press*, 2001.
11. M. Hilbert and P. López, "The world's technological capacity to store, communicate, and compute information," *Science*, vol. 332, no. 6025, pp. 60-65, 2011.
12. C. F. Huang, "A hybrid stock selection model using genetic algorithms and support vector regression," *Applied Soft Computing*, vol. 12, no. 2, pp. 807-818, 2012.
13. S. Y. Huang, R. H. Tsaih, and F. Yu, "Topological pattern discovery and feature extraction for fraudulent financial reporting," *Expert Systems with Applications*, vol. 41, no. 9, pp. 4360-4372, 2014.
14. K. Hussain and E. Prieto, "Big data in the finance and insurance sectors," in *New horizons for a data-driven economy: A roadmap for usage and exploitation of big data in Europe*, pp. 209-223, 2016.
15. J. C. Ireland, "An empirical investigation of determinants of audit reports in the UK," *Journal of Business Finance & Accounting*, vol. 30, no. 7-8, pp. 975-1016, 2003.
16. D. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann, "Mastering the information age solving problems with visual analytics," *Eurographics Association*, 2010. [Online]. Διαθέσιμο από: [Link](#).
17. H. Liu, J. Mulvey, and T. Zhao, "A semiparametric graphical modelling approach for large-scale equity selection," *Quantitative finance*, vol. 16, no. 7, pp. 1053-1067, 2016.
18. H. Markowitz, "Portfolio selection," *Journal of finance*, vol. 7, pp. 77-91, 1952.
19. OECD, "Artificial Intelligence, Machine Learning and Big Data in Finance: Opportunities, Challenges, and Implications for Policy Makers," 2021. [Online]. Διαθέσιμο από: [Link](#).
20. F. Pasiouras, C. Gaganis, and C. Zopounidis, "Multicriteria decision support methodologies for auditing decisions: The case of qualified audit reports in the UK," *European Journal of Operational Research*, vol. 180, no. 3, pp. 1317-1330, 2007.

21. B. Porter and A. Cameron, "Company fraud—what price the auditor?," *Accountant's Journal*, vol. 12, pp. 44-47, 1987.
22. G. Press, "A Very Short History Of Data Science," *Forbes*, 2013. [Online]. Διαθέσιμο από: [Link](#).
23. T. S. Quah, "DJIA stock selection assisted by neural network," *Expert Systems with Applications*, vol. 35, no. 1-2, pp. 50-58, 2008.
24. C. Quek, M. Pasquier, and N. Kumar, "A novel recurrent neural network-based prediction system for option trading and hedging," *Applied Intelligence*, vol. 29, pp. 138-151, 2008.
25. M. H. Rafiei and H. Adeli, "A novel machine learning model for estimation of sale prices of real estate units," *Journal of Construction Engineering and Management*, vol. 142, no. 2, pp. 04015066, 2016.
26. S. Ramamoorti, A. D. Bailey Jr, and R. O. Traver, "Risk assessment in internal auditing: a neural network approach," *Intelligent Systems in Accounting, Finance & Management*, vol. 8,