

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
Σχολή Χρηματοοικονομικής και Στατιστικής



Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ

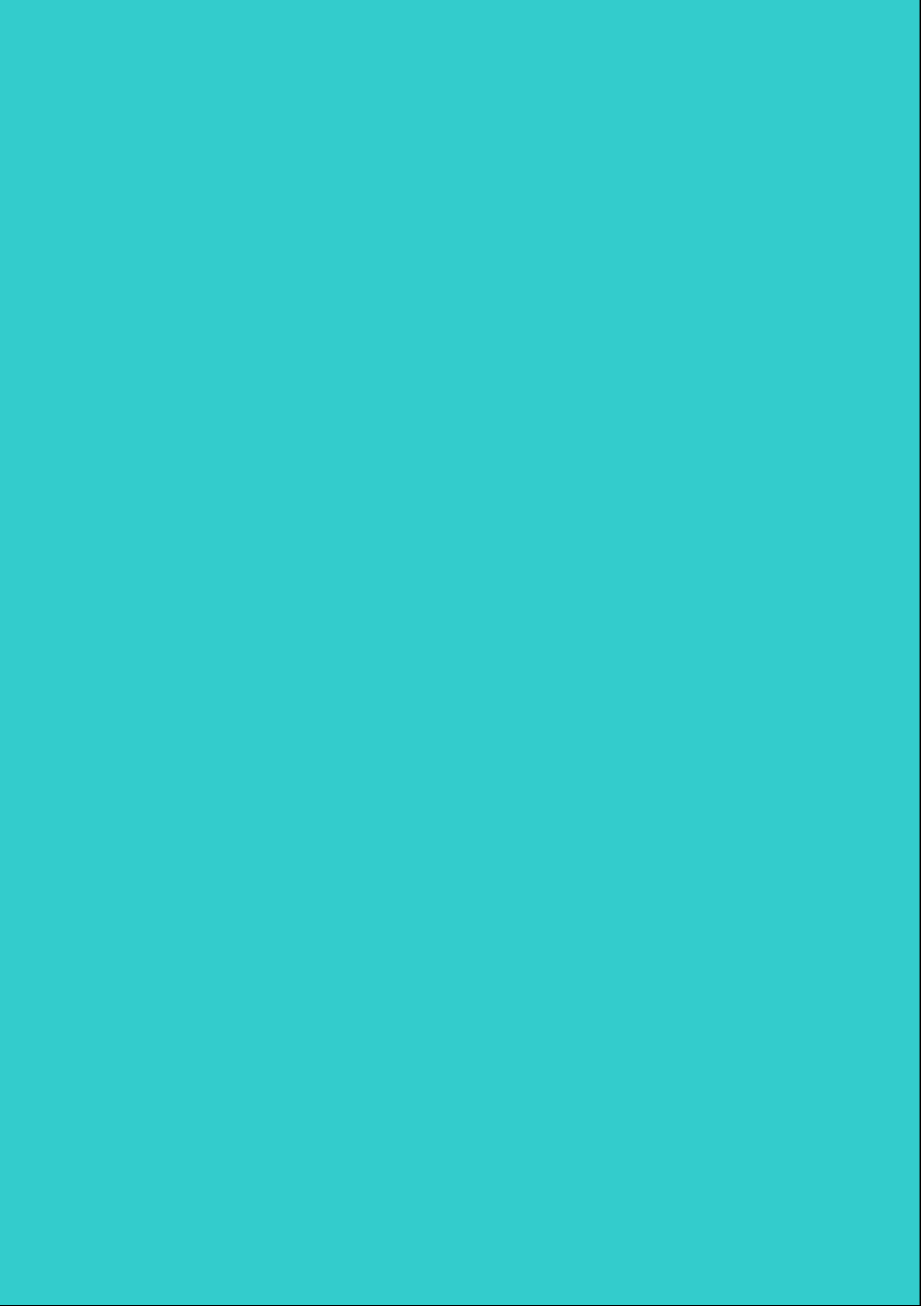
ΤΕΧΝΙΚΕΣ ΑΝΑΛΥΣΗΣ
ΠΟΛΥΔΙΑΣΤΑΤΩΝ ΟΙΚΟΝΟΜΙΚΩΝ
ΧΡΟΝΟΣΕΙΡΩΝ

Διονύσιος Δ. Βαρελάς

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής
Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των
απαιτήσεων για την απόκτηση του Μεταπτυχιακού
Διπλώματος Ειδίκευσης στην *Εφαρμοσμένη Στατιστική*

Πειραιάς
Νοέμβριος 2016



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
Σχολή Χρηματοοικονομικής και Στατιστικής



Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ

ΤΕΧΝΙΚΕΣ ΑΝΑΛΥΣΗΣ
ΠΟΛΥΔΙΑΣΤΑΤΩΝ ΟΙΚΟΝΟΜΙΚΩΝ
ΧΡΟΝΟΣΕΙΡΩΝ

Διονύσιος Δ. Βαρελάς

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής
Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των
απαιτήσεων για την απόκτηση του Μεταπτυχιακού
Διπλώματος Ειδίκευσης στην *Εφαρμοσμένη Στατιστική*

Πειραιάς
Νοέμβριος 2016

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίστηκε από τη ΓΣΕΣ του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμ. συνεδρίασή του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική

Τα μέλη της Επιτροπής ήταν:

- Καθηγητής Κούτρας Μάρκος (Επιβλέπων)
- Καθηγητής Αγιακλόγλου Χρήστος
- Επίκουρος Καθηγητής Στέγγος Δημήτριος

Η έγκριση της Διπλωματικής Εργασίας από το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς δεν υποδηλώνει αποδοχή των γνώμων του συγγραφέα.

UNIVERSITY OF PIRAEUS
School of Finance and Statistics



Department of Statistics and Insurance Science

**POSTGRADUATE PROGRAM IN
APPLIED STATISTICS**

**ANALYSIS OF MULTIVARIATE
FINANCIAL TIME SERIES**

By

Dionysios D. Varelas

MSc Dissertation

submitted to the Department of Statistics and Insurance
Science of the University of Piraeus in partial fulfilment
of the requirements for the degree of Master of Science in
Applied Statistics

Piraeus, Greece
November 2016

Στην οικογένεια μου

Ευχαριστίες

Σε αυτή την ενότητα θα ήθελα να ευχαριστήσω τόσο τα άτομα που με βοήθησαν στην εκπόνηση της παρούσας διπλωματικής εργασίας όσο και αυτούς που με βοήθησαν κατά τις σπουδές μου στο Π.Μ.Σ. «Εφαρμοσμένη Στατιστική». Αρχικά θα ήθελα να ευχαριστήσω τον κύριο Μάρκο Κούτρα, καθηγητή του τμήματος Στατιστικής και Ασφαλιστικής επιστήμης του Πανεπιστημίου Πειραιώς, για την πολύτιμη συνεισφορά του ώστε να λάβει η εργασία την παρούσα μορφή της. Ευχαριστώ επίσης τα μέλη της τριμελούς επιτροπής, τον Καθηγητή κ. Χ. Αγιακλόγλου και τον Επίκουρο Καθηγητή κ. Δ. Στέγγο για το χρόνο που αφιέρωσαν στη διόρθωση της εργασίας. Ευχαριστώ ακόμα τους συμφοιτητές μου Β. Τραπουζανλή και Β. Κοντονή, τον πρώτο για τις εύστοχες παρατηρήσεις και σχόλια που με βοήθησαν στην συγγραφή και τον δεύτερο για όλες τις εποικοδομητικές συζητήσεις που είχαμε κατά τη διάρκεια των σπουδών μας στο Πανεπιστήμιο Πειραιώς. Τέλος, ευχαριστώ θερμα την οικογένεια μου για την υποστήριξη που μου προσέφερε και συνεχίζει να μου προσφέρει, η οποία σε μεγάλο βάθος ευθύνεται για πολλά από αυτά που έχω πετύχει μέχρι σήμερα.

Περίληψη

Τα τελευταία χρόνια πολλοί οργανισμοί και ερευνητικά κέντρα έχουν στη διάθεση τους μεγάλα σύνολα δεδομένων. Τα δεδομένα αυτά μπορούν να χρησιμοποιηθούν προκειμένου να δημιουργηθούν πιο ακριβείς προβλέψεις διαφόρων οικονομικών στοιχείων. Ωστόσο κλασικές μέθοδοι ανάλυσης πολυμεταβλητών χρονοσειρών αδυνατούν να εκμεταλευτούν το μεγάλο όγκο δεδομένων επειδή ο αριθμός των μεταβλητών είναι μεγαλύτερος από τον αριθμό των παρατηρήσεων. Το πρόβλημα αυτό μπορεί να ξεπεράσει το δυναμικό μοντέλο ανάλυσης παραγόντων με τη δημιουργία νέων μεταβλητών, οι οποίες προέρχονται από γραμμικούς συνδυασμούς των αρχικών και συνοψίζουν αρκετά ικανοποιητικά την πληροφορία που περιέχουν οι αρχικές μεταβλητές.

Στόχος της εργασίας αυτής ήταν αρχικά η θεωρητική περιγραφή ενός από τα διαθέσιμα μοντέλα δυναμικής ανάλυσης παραγόντων και έπειτα η εφαρμογή της μεθοδολογίας που θα περιγράψουμε σε ένα σύνολο μετρήσεων, το οποίο θα αποτελείται από πραγματικά δεδομένα του ελληνικού χώρου. Συγκεκριμένα, κατασκευάσαμε ένα μοντέλο που να προβλέπει το ΑΕΠ και το συγκρίναμε με ένα υπόδειγμα ARIMA. Η σύγκριση έγινε χρησιμοποιώντας ως κριτήρια αξιολόγησης τα MSFE, MAE και MAPE. Το υπόδειγμα δυναμικής παραγωγτικής ανάλυσης εκτιμήθηκε σε δύο στάδια. Στο πρώτο εντοπίσαμε κοινούς παράγοντες μεταξύ των διαθέσιμων μεταβλητών, χρησιμοποιώντας την κλασική μέθοδο ανάλυσης παραγόντων. Στο δεύτερο οι παράγοντες αυτοί αποτέλεσαν βασικό μέρος του μοντέλου χώρου-καταστάσεων που χρησιμοποιήσαμε για την πραγματοποίηση προβλέψεων. Η προσαρμογή του μοντέλου ήταν καλύτερη από αυτή του ARIMA γεγονός το οποίο επιβεβαιώνει την χρησιμότητα του δυναμικού μοντέλου ανάλυσης παραγόντων στην πραγματοποίηση οικονομικών προβλέψεων.

Abstract

In recent years many institutions and research centres have at their disposal large datasets. Those data can be exploited in order to obtain more accurate forecasts of various economic data. However, traditional methods of multivariate time series analysis cannot exploit the large volume of data due to the large number of variables compared to the number of observations. That problem can be overcome by the use of a dynamic factor model, which creates new variables as linear functions of the original variables. Those new variables can summarize effectively the information held by the original ones.

The aim of this thesis was firstly to theoretically describe one of the available dynamic factor models and secondly to apply this methodology on a greek dataset. In particular, we have constructed a model which predicts GDP and we have compared that model to an ARIMA model. The comparison was made using the criteria MSFE, MAE and MAPE. The dynamic factor model was estimated in two steps. In the first step we extracted the factors using classic factor analysis. In the second step those factors were an important part of the state-space model used in order to obtain forecasts. The fitting of that model was better compared to the fitting of the ARIMA model, which confirms the fact that dynamic factor models can be useful in forecasting economic data.

Περιεχόμενα

Κατάλογος σχημάτων	iii
Κατάλογος πινάκων	iv
Κατάλογος συντομογραφιών	v
I Θεωρητική περιγραφή	1
1 Η Ανάλυση Παραγόντων	1
1.1 Εισαγωγή	1
1.2 Το βασικό μοντέλο	3
1.3 Μέθοδοι εκτίμησης	5
1.3.1 Ανάλυση Κυρίων Συνιστωσών	5
1.3.2 Η Μέθοδος Μέγιστης Πιθανοφάνειας	6
1.4 Περιστροφή Παραγόντων	7
1.5 Εκτίμηση των τιμών των παραγόντων	8
1.6 Ερμηνεία των Παραγόντων	9
2 Το δυναμικό μοντέλο της ανάλυσης παραγόντων	11
2.1 Εισαγωγή	11
2.2 Μοντέλο χώρου καταστάσεων	12
2.3 Το φίλτρο του Kalman	13
2.4 Εξομάλυνση Kalman	17
2.5 Ο αλγόριθμος EM	22
3 Το μοντέλο των Doz, Giannone, Reichlin	24
3.1 Εισαγωγή	24

3.2	Το μοντέλο και υποθέσεις	24
3.3	Προσεγγίσεις του παραμετρικού χώρου	27
II	Αριθμητική εφαρμογή	31
4	Αριθμητική εφαρμογή σε ελληνικά δεδομένα	31
4.1	Εισαγωγή	31
4.2	Περιγραφή του συνόλου δεδομένων	32
4.3	Προκαταρκτική επεξεργασία των δεδομένων	33
4.4	Παράγοντες	35
4.5	Αποτελέσματα	40
5	Συμπεράσματα και συζήτηση	42
	Παράρτημα	44
	Βασικές έννοιες	44
	Κώδικες και αποτελέσματα	47
	Βιβλιογραφία	62

Κατάλογος σχημάτων

1.1	Προβολή από τον \mathbb{R}^3 στον \mathbb{R}^2	2
1.2	Νέοι άξονες στην διεύθυνση των \vec{e}_1, \vec{e}_2	5
1.3	Ορθογώνια περιστροφή ορθοκανονικού συστήματος συντεταγμένων $O\vec{e}_1\vec{e}_2$	7
4.1	Διάσπαση της μεταβλητής ”προστιθέμενη αξία (γενικός δείκτης)”	34
4.2	Πρώτος παράγοντας	36
4.3	Δεύτερος παράγοντας	38
4.4	Τρίτος παράγοντας	39
4.5	Προσαρμοσμένες τιμές	41
4.6	Πρόβλεψη ΑΕΠ έως το 4ο τρίμηνο του 2016.	42

Κατάλογος πινάκων

1	Συσχετίσεις παραγόντων με τις αρχικές μεταβλητές.	10
2	Εξισώσεις φιλτραρίσματος.	15
3	Εξισώσεις φιλτραρίσματος για το παράδειγμα 2.1.	16
4	Εξισώσεις εξομάλυνσης.	20
5	Πίνακας διαστάσεων.	25
6	Μεταβλητές με τις οποίες παρουσιάζει την υψηλότερη συσχέτιση ο 1ος παράγοντας.	35
7	Μεταβλητές με τις οποίες παρουσιάζει την υψηλότερη συσχέτιση ο 2ος παράγοντας.	37
8	Μεταβλητές με τις οποίες παρουσιάζει την υψηλότερη συσχέτιση ο 3ος παράγοντας.	39
9	Σύγκριση των δύο ανταγωνιστών.	40
10	Μεταβλητές που απαρτίζουν το σύνολο δεδομένων.	46

Κατάλογος συντομογραφιών

ΑΕΠ Ακαθάριστο Εθνικό Προϊόν

ΔΜ Δυναμικό Μοντέλο

ACF Autocorrelation Function

ADF Augmented Dickey-Fuller

ARIMA Autoregressive Integrated Moving Average

EM Expectation-Maximization

MAE Mean Absolute Error

MAPE Mean Absolute Percentage Error

MSFE Mean Squared Forecast Error

PACF Partial Autocorrelation Function

PCA Principal Component Analysis

VAR Vector Autoregressive

VARMA Vector Autoregressive Moving Average

Μέρος I

Θεωρητική περιγραφή

Στο πρώτο μέρος της παρούσας διπλωματικής εργασίας θα δώσουμε μια περιγραφή της κλασικής ανάλυσης παραγόντων μία εκ των βασικότερων μεθόδων της πολυμεταβλητής ανάλυσης. Έπειτα θα περιγράψουμε τα μοντέλα χώρου-καταστάσεων (state-space models) και πως προκύπτει το δυναμικό μοντέλο ανάλυσης παραγόντων ως ειδική περίπτωση αυτών. Στη συνέχεια θα αναφερθούμε στη διαδικασία εκτίμησης των παραμέτρων του μοντέλου χώρου-καταστάσεων με τη βοήθεια του φίλτρου και της εξομάλυνσης Kalman (Kalman filter & smoother) σε συνδυασμό με τον αλγόριθμο EM (Expectation-Maximization). Κλείνοντας, θα δείξουμε πως ο συνδυασμός των παραπάνω οδηγεί φυσικά στο μοντέλο που πρότειναν οι Doz, Giannone, Reichelin[13].

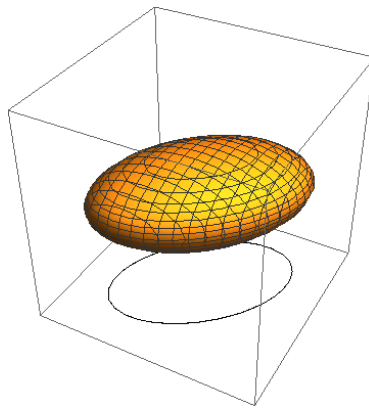
1 Η Ανάλυση Παραγόντων

1.1 Εισαγωγή

Η ανάλυση παραγόντων χρησιμοποιήθηκε για πρώτη φορά από τον Charles Spearman το 1904 για να μελετήσει τις συσχετίσεις ανάμεσα στις βαθμολογίες μαθητών σε διάφορα τεστ, διαπιστώνοντας ότι αυτές μπορούν να περιγραφούν χρησιμοποιώντας ένα γραμμικό συνδυασμό κάποιων μη παρατηρούμενων μεταβλητών. Η παρατήρηση αυτή οδήγησε στην τεχνική της ανάλυσης παραγόντων, η οποία αρχικά χρησιμοποιήθηκε από επιστήμονες με ενδιαφέρον στην ψυχομετρία. Στις μέρες μας η διαρκής εξέλιξη της υπολογιστικής ικανότητας και ταχύτητας των ηλεκτρονικών υπολογιστών αναθέρμανε το ενδιαφέρον των ερευνητών για την ανάλυση παραγόντων, η οποία πλέον έχει ένα ευρύ φάσμα εφαρμογών.

Στόχος της ανάλυσης παραγόντων είναι να περιγράψει τις σχέσεις των υπό εξέταση p το πλήθος μεταβλητών, χρησιμοποιώντας γραμμικούς συνδυασμούς από m το πλήθος μεταβλητές

($m < p$), οι οποίες είναι μη παρατηρούμενες αλλά προκύπτουν ως γραμμικοί συνδιασμοί των αρχικών p μεταβλητών. Οι γραμμικοί συνδιασμοί αυτοί προκύπτουν, όπως θα αναφέρουμε αναλυτικά παρακάτω στην αλγεβρική διατύπωση της μεθόδου, λαμβάνοντας υπόψιν είτε τη δομή του πίνακα συνδιακυμάνσεων είτε αυτού των συσχετίσεων. Για παράδειγμα, οι συσχετίσεις μεταξύ των μεταβλητών "Γαλλικά", "Αγγλικά", "Μαθηματικά" και "Μουσική" που συνέλεξε ο Spearman υποδείκνυαν την ύπαρξη ενός κοινού παράγοντα τον οποίο ονομάτισε "Ευφυΐα".



Σχήμα 1.1: Προβολή από τον \mathbb{R}^3 στον \mathbb{R}^2 .

Στο κεφάλαιο αυτό στόχος μας είναι να δώσουμε τη μορφή του βασικού μοντέλου ανάλυσης παραγόντων. Ιδιαίτερη έμφαση θα δοθεί στην ανάλυση κυρίων συνιστωσών (PCA), η οποία αποτελεί το πρωταρχικό βήμα για την κατασκευή του μοντέλου ανάλυσης παραγόντων. Επιπρόσθετα θα γίνει και μια αναφορά στην μέθοδο μέγιστης πιθανοφάνειας (MLE) ως μέθοδο εκτίμησης των φορτίων των παραγόντων. Στη συνέχεια, θα δείξουμε πως χρησιμοποιώντας τα φορτία των παραγόντων και μέσω ορθογωνίων περιστροφών του συστήματος συντεταγμένων που αυτά ορίζουν μπορούμε να δώσουμε κάποια φυσική ερμηνεία στους παράγοντες που εντοπίσαμε. Κλείνοντας, θα δώσουμε δύο μεθόδους υπολογισμού των τιμών (scores) που λαμβάνουν οι παράγοντες.

1.2 Το βασικό μοντέλο

Δοθέντων p το πλήθος μεταβλητών X_1, X_2, \dots, X_p αυτές μπορούν να αναπαρασταθούν ως εξής,

$$X_i = \mu_i + \sum_{k=1}^m l_{ik} F_k + \varepsilon_i, \quad i = 1, \dots, p \quad (1.1)$$

Στην σχέση (1.1),

1. Το κάθε μ_i , όπου $i = 1, \dots, p$, αποτελεί τη μέση τιμή της i -οστής αντίστοιχης μεταβλητής.
2. Οι τυχαίες μεταβλητές F_1, \dots, F_m λέγονται κοινοί παράγοντες.
3. Τα ε_i λέγονται σφάλματα ή τυχαίοι παράγοντες¹.
4. Το κάθε l_{ik} , όπου $i = 1, \dots, p$, και $k = 1, \dots, m$, καλείται φορτίο της αντίστοιχης i -οστής μεταβλητής στον k -οστό παράγοντα.

Το μοντέλο της σχέσης (1.1) σε μορφή πινάκων λαμβάνει την εξής μορφή,

$$\underset{(p \times 1)}{\mathbf{X}} = \underset{(p \times 1)}{\boldsymbol{\mu}} + \underset{(p \times m)}{\mathbf{L}} \cdot \underset{(m \times 1)}{\mathbf{F}} + \underset{(p \times 1)}{\boldsymbol{\varepsilon}}. \quad (1.2)$$

Χρησιμοποιώντας την σχέση (1.2) μπορούμε να δώσουμε τις ακόλουθες συνθήκες,

1. Τα διανύσματα \mathbf{F} και $\boldsymbol{\varepsilon}$ είναι ανεξάρτητα, δηλαδή $\text{Cov}(\boldsymbol{\varepsilon}, \mathbf{F}) = \mathbf{0}$.
2. $\text{E}(\mathbf{F}) = \mathbf{0}$ και $\text{Cov}(\mathbf{F}) = I_m$ ².
3. $\text{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ και $\text{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}$, όπου $\boldsymbol{\Psi}$ διαγώνιος πίνακας.

¹Όπως αναφέρει ο Maxwell[26] τα ε_i αποτελούν συνδυασμό σφαλμάτων κατά τις μετρήσεις και παραγόντων που σχετίζονται μοναδικά με τις p μεταβλητές.

²Ουσιαστικά με αυτή την συνθήκη εξασφαλίζουμε ότι το σύστημα συντεταγμένων που ορίζεται από τα φορτία των παραγόντων είναι ορθοκανονικό.

Οι συνθήκες της προηγούμενης σελίδας είναι απαραίτητες προκειμένου να εκτιμήσουμε τις παραμέτρους του μοντέλου που δώσαμε στη σχέση (1.2).

Σημαντική είναι επίσης η δομή που λαμβάνει ο πίνακας συνδιακυμάνσεων μέσω της (1.2) και των παραπάνω συνθηκών. Μετά από μερικές αλγεβρικές πράξεις μπορεί να δει κανείς ότι $\Sigma = \text{Cov}(\mathbf{X}) = \mathbf{L}\mathbf{L}' + \mathbf{\Psi}$. Από τη δομή αυτή εξάγουμε δύο πολύ σημαντικά αποτελέσματα για την συνέχεια.

1. Η διακύμανση της i -οστής μεταβλητής είναι ίση με $\psi_i + \sum_{j=1}^m l_{ij}^2$, όπου ψ_i το (i, i) στοιχείο του πίνακα $\mathbf{\Psi}$. Το $\sum_{j=1}^m l_{ij}^2$ καλείται κοινή διακύμανση που οφείλεται στους m παράγοντες, ενώ το υπόλοιπο, το οποίο συμβολίσαμε με ψ_i , καλείται ειδική διακύμανση.
2. Η συνδιακύμανση της i -οστής με την k -οστή μεταβλητή είναι ίση με $\sum_{j=1}^m l_{ij}l_{kj}$.

Λαμβάνοντας υπόψιν τα παραπάνω γίνεται αντιληπτό ότι χρησιμοποιώντας το βασικό (ή ορθογώνιο) μοντέλο της ανάλυσης παραγόντων ενδιαφερόμαστε να επιτύχουμε όσο το δυνατόν καλύτερη αναπαράσταση της διακύμανσης των αρχικών μεταβλητών όταν τις προβάλλουμε σε χώρο μικρότερης διάστασης \mathbb{R}^m .

Κλείνοντας, θα δώσουμε δύο σημαντικούς τύπους για τη συνέχεια μέσω των οποίων είναι δυνατή η ερμηνεία των παραγόντων που θα εντοπίσουμε μέσω της ανάλυσης παραγόντων.

1. $\text{Cov}(X_i, F_j) = l_{ij}$
2. $\rho(X_i, F_j) = \frac{l_{ij}}{\sigma_i}$

Τους τύπους αυτούς θα τους χρησιμοποιήσουμε στην Παράγραφο 1.6 προκειμένου να δώσουμε κάποια φυσική ερμηνεία στους παράγοντες που θα εντοπίσουμε.

1.3 Μέθοδοι εκτίμησης

1.3.1 Ανάλυση Κυρίων Συνιστωσών

Έστω Σ ο πίνακας των συνδιακυμάνσεων των p μεταβλητών, ο οποίος είναι θετικά ημιορισμένος και συμμετρικός. Επίσης δεδομένου ότι ο χώρος, στον οποίο κείνται οι p μεταβλητές είναι ο πραγματικός χώρος, εσωτερικού γινομένου και πεπερασμένης διάστασης \mathbb{R}^p , σύμφωνα με το φασματικό θεώρημα υπάρχουν διανύσματα e_1, e_2, \dots, e_p και αριθμοί $\lambda_1, \lambda_2, \dots, \lambda_p$, έτσι ώστε:

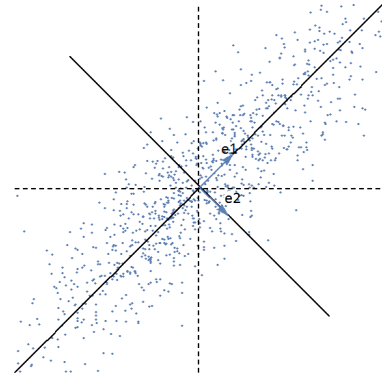
$$\Sigma = \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2' + \dots + \lambda_p \mathbf{e}_p \mathbf{e}_p' \quad (1.3)$$

Συγκεκριμένα τα ζεύγη $(\lambda_i, \mathbf{e}_i)$ εκφράζουν την i -οστή ιδιοτιμή και το αντίστοιχο ιδιοδιάνυσμα του πίνακα συνδιακυμάνσεων. Κάθενα από τα \mathbf{e}_i αποτελεί μια ορθογώνια προβολή υπό την έννοια ότι για κάθε ζεύγος i, j με $i \neq j$ τα ιδιοδιανύσματα $\mathbf{e}_i, \mathbf{e}_j$ είναι μεταξύ τους κάθετα, ιδιότητα που όπως θα δούμε παρακάτω σχετίζεται άμεσα με μία από τις συνθήκες που δώσαμε παραπάνω για τη σχέση (1.2). Γράφοντας τη σχέση (1.3)

υπό μορφή πινάκων βλέπουμε ότι $\Sigma = \mathbf{L}\mathbf{L}' + \mathbf{0}$, όπου $\mathbf{L} = (\sqrt{\lambda_1} \mathbf{e}_1, \sqrt{\lambda_2} \mathbf{e}_2, \dots, \sqrt{\lambda_p} \mathbf{e}_p)'$, μορφή η οποία ταιριάζει στη δομή

του πίνακα συνδιακύμανσης όπως αυτός ορίζεται από το μοντέλο της ανάλυσης παραγόντων για $m = p$. Το Σχήμα 1.1 αναφέρεται στην περίπτωση $p = 2$ και δείχνει το νέο σύστημα αξόνων (μη διακεκομμένες γραμμές) που ορίζεται από τα διανύσματα $\vec{\mathbf{e}}_1, \vec{\mathbf{e}}_2$.

Πηγαίνοτας ένα βήμα παραπέρα αυτό που θα ενδιέφερε είναι πέρα από το να αλλάξουμε το σύστημα αξόνων, να καταφέρουμε να τα προβάλλουμε από τον αρχικό χώρο \mathbb{R}^p σε χώρο μικρότερης διάστασης \mathbb{R}^m . Προκειμένου να το πετύχουμε αυτό αρκεί να χρησιμοποιήσουμε τις m μεγαλύτερες ιδιοτιμές του πίνακα Σ . Μέσω αυτής της διαδικασίας έχουμε πλέον ότι $\Sigma = \mathbf{L}\mathbf{L}' + \Psi$, όπου $\mathbf{L} = (\sqrt{\lambda_1} \mathbf{e}_1, \sqrt{\lambda_2} \mathbf{e}_2, \dots, \sqrt{\lambda_m} \mathbf{e}_m)'$ με $m < p$. Ο πίνακας \mathbf{L} περιέχει τα φορτία όπως αυτά



Σχήμα 1.2: Νέοι άξονες στην διεύθυνση των $\vec{\mathbf{e}}_1, \vec{\mathbf{e}}_2$.

υπολογίστηκαν με τη μέθοδο των κυρίων συνιστωσών.

Να σημειώσουμε ότι στην πράξη δεν θα έχουμε στην διάθεση μας τον πίνακα Σ αλλά μια εκτίμηση αυτού, τον δειγματικό πίνακα διακυμάνσεων-συνδιακυμάνσεων \mathbf{S} . Κατά συνέπεια θα μπορούμε να υπολογίσουμε εκτιμήσεις των φορτίων, δηλαδή τον πίνακα εκτιμημένων φορτίων $\hat{\mathbf{L}} = \left(\sqrt{\hat{\lambda}_1} \hat{\mathbf{e}}_1, \sqrt{\hat{\lambda}_2} \hat{\mathbf{e}}_2, \dots, \sqrt{\hat{\lambda}_m} \hat{\mathbf{e}}_m \right)'$. Περισσότερες λεπτομέρειες δίνονται στο [3].

1.3.2 Η Μέθοδος Μέγιστης Πιθανοφάνειας

Στην περίπτωση που υποθέτουμε ότι οι κοινοί παράγοντες (F_i) και οι τυχαίοι παράγοντες (ε_i) ακολουθούν κανονική κατανομή μπορούμε να πάρουμε εκτιμήσεις των πινάκων \mathbf{L} και Ψ μέσω της μεθόδου μέγιστης πιθανοφάνειας. Η μέθοδος αυτή στηρίζεται, όπως λέει και το όνομα της, στη μεγιστοποίηση της πιθανοφάνειας

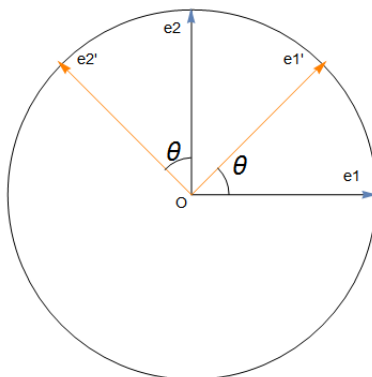
$$L(\mu, \Sigma) = (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} e^{-\frac{1}{2} tr[\Sigma^{-1} (\sum_{j=1}^p (X_j - \bar{X}_j)(X_j - \bar{X}_j)' + n(\bar{X}_j - \mu)(\bar{X}_j - \mu)']}, \quad (1.4)$$

η οποία μπορεί να επιτευχθεί επιβάλλοντας την συνθήκη³ $\mathbf{L}'\Psi\mathbf{L} = \mathbf{\Delta}$, όπου ο $\mathbf{\Delta}$ είναι διαγώνιος πίνακας. Οι εκτιμητές μέγιστης πιθανοφάνειας $\hat{\mathbf{L}}$ και $\hat{\Psi}$ μπορούν να υπολογιστούν χρησιμοποιώντας αριθμητικές μεθόδους, όπως για παράδειγμα τον αλγόριθμο EM. Αν θυμηθούμε τη δομή που λαμβάνει ο πίνακας συνδιακυμάνσεων μέσω της (1.2), δηλαδή $\Sigma = Cov(\mathbf{X}) = \mathbf{L}\mathbf{L}' + \Psi$, μπορούμε να δούμε πως συνδέεται η πιθανοφάνεια της σχέσης (1.4) με τους πίνακες που θέλουμε να εκτιμήσουμε.

³Η συνθήκη αυτή μας επιτρέπει να προσδιορίσουμε μοναδικά το μοντέλο μας, διότι όπως είναι γνωστό ο \mathbf{L} μπορεί να πάρει διαφορετικές τιμές όταν εφαρμόζεται κάποιος ορθογώνιος μετασχηματισμός. Περισσότερες λεπτομέρειες δίνονται στην Παράγραφο 1.4.

1.4 Περιστροφή Παραγόντων

Όπως αναφέραμε στην παράγραφο 2 το σύστημα συντεταγμένων που ορίζεται από τα φορτία των παραγόντων είναι ορθοκανονικό. Έχει αποδειχθεί ότι για κάθε ορθογώνιο πίνακα μετασχηματισμού, έστω \mathbf{T} , το γινόμενο $\hat{\mathbf{L}}\mathbf{T}$ αποτελεί έναν $p \times m$ πίνακα εκτιμημένων φορτίων που έχουν προκύψει από περιστροφή⁴. Γίνεται αντιληπτό πως κάθε τέτοια περιστροφή των παραγόντων αφήνει αναλλοίωτες τόσο τις κοινές διακυμάνσεις που προκύπτουν από τους αρχικούς παράγοντες όσο και τις ειδικές που προκύπτουν από τα σφάλματα. Αφού λοιπόν αλγεβρικά δεν αλλάζει κάτι τίθεται το ερώτημα για ποιό λόγο κανείς θα έκανε περιστροφή των αρχικών παραγόντων.



Σχήμα 1.3: Ορθογώνια περιστροφή ορθοκανονικού συστήματος συντεταγμένων $O\vec{e}_1\vec{e}_2$.

Η περιστροφή των παραγόντων έχει άμεση σχέση με την ερμηνεία αυτών, καθώς πολλές φορές οι αρχικές εκτιμήσεις των φορτίων δεν επιτρέπουν την ερμηνεία των παραγόντων σύμφωνα με τις συσχετίσεις που εμφανίζουν με τις αρχικές μεταβλητές. Μια από πλέον δημοφιλείς τεχνικές περιστροφής προτάθηκε από τον Kaiser και ονομάζεται varimax. Σύμφωνα με αυτή ορίζουμε αρχικά τις ποσότητες $\tilde{l}_{ij}^* = \frac{\hat{l}_{ij}^*}{\hat{h}_i}$, όπου \hat{h}_i η τετραγωνική ρίζα της κοινής διακύμανση που

⁴Η έννοια της περιστροφής προκύπτει από το γεγονός ότι οποιοσδήποτε ορθογώνιος μετασχηματισμός αποτελεί μια περιστροφή του συστήματος συντεταγμένων τέτοια ώστε να διατηρούνται οι αποστάσεις των σημείων (ισομετρία).

οφείλεται στον i -οστό παράγοντα και \hat{l}_{ij}^* το (i, j) φορτίο που προκύπτει αφού πολλαπλασιάσουμε τα αρχικά φορτία με κάποιον πίνακα μετασχηματισμού \mathbf{T} . Στη συνέχεια επιλέγουμε τον πίνακα μετασχηματισμού \mathbf{T} έτσι ώστε να μεγιστοποιήσουμε την πόσότητα

$$V = \frac{1}{p} \sum_{j=1}^m \left[\sum_{i=1}^p \tilde{l}_{ij}^{*4} - \left(\sum_{i=1}^p \tilde{l}_{ij}^{*2} \right)^2 / p \right]$$

Ο παραπάνω ορθογώνιος μετασχηματισμός ελαχιστοποιεί τον αριθμό των μεταβλητών που έχουν μεγάλα φορτία σε κάθε παράγοντα. Με τον τρόπο αυτό η ερμηνεία των παραγόντων γίνεται συνήθως πιο εύκολη, αφού πλέον λιγότερες μεταβλητές ευθύνονται σε μεγάλο μέρος για τη δημιουργία του παράγοντα.

1.5 Εκτίμηση των τιμών των παραγόντων

Σημαντικό ρόλο για την εφαρμογή της μεθόδου της ανάλυσης παραγόντων αποτελεί η εκτίμηση των τιμών που λαμβάνουν οι παράγοντες κάτι το οποίο θα γίνει περισσότερο αντιληπτό στο εμπειρικό κομμάτι της εργασίας αυτής. Συγκεκριμένα, στο μοντέλο χώρου καταστάσεων που θα διαμορφώσουμε οι παράγοντες θα ακολουθούν κάποιο στοχαστικό μοντέλο. Στα πλαίσια αυτού του μέρους θα αναλύσουμε δύο από τις συνηθέστερες τεχνικές εκτίμησης των τιμών των παραγόντων: τη γραμμική παλινδρόμηση και τη μέθοδο του Bartlett[6].

Αν θεωρήσουμε ότι από την εφαρμογή της μεθόδου της ανάλυσης παραγόντων καταλήξαμε στο μοντέλο της μορφής

$$\mathbf{X} = \boldsymbol{\mu} + \hat{\mathbf{L}}\mathbf{F} + \boldsymbol{\varepsilon}$$

μπορούμε να εφαρμόσουμε άμεσα τις δύο μεθόδους.

Στην περίπτωση της γραμμικής παλινδρόμησης αρκεί να θεωρήσουμε ως παράμετρο προς εκτίμηση το διάνυσμα $\mathbf{F} = (F_1, F_2, \dots, F_m)'$. Δεδομένου ότι οι τιμές των \mathbf{X} , $\boldsymbol{\mu}$, $\hat{\mathbf{L}}$ είναι γνωστές

η εκτίμηση των τιμών των παραγόντων μπορεί να προκύψει εύκολα από τον τύπο

$$\hat{\mathbf{F}} = (\hat{\mathbf{L}}'\hat{\mathbf{L}})^{-1}\hat{\mathbf{L}}'(\mathbf{X} - \boldsymbol{\mu})$$

Σύμφωνα με την μέθοδο του Bartlett, η οποία αποτελεί μια σταθμισμένη μέθοδο ελαχίστων τετραγώνων, το γεγονός ότι οι συνιστώσες του διανύσματος $\boldsymbol{\varepsilon}$ δεν είναι απαραίτητα ομοσκεδαστικές οδήγησε τον Bartlett να ελαχιστοποιήσει την ποσότητα

$$\sum_{i=1}^p \frac{\varepsilon_i^2}{\psi_i}$$

καταλήγοντας στον τύπο

$$\hat{\mathbf{F}} = (\hat{\mathbf{L}}'\boldsymbol{\Psi}^{-1}\hat{\mathbf{L}})^{-1}\hat{\mathbf{L}}'\boldsymbol{\Psi}^{-1}(\mathbf{X} - \boldsymbol{\mu})$$

Με την μέθοδο αυτή ο Bartlett ουσιαστικά ελάβε υπόψιν τις ενδεχομένως διαφορετικές διακυμάνσεις των σφαλμάτων.

1.6 Ερμηνεία των Παραγόντων

Αφού έχουμε ολοκληρώσει την εκτίμηση των φορτίων των παραγόντων ιδιαίτερο ενδιαφέρον παρουσιάζει η απόδοση κάποιας λογικής ερμηνείας στις νέες μεταβλητές που κατασκευάσαμε. Παρά το γεγονός ότι τις περισσότερες φορές η μεθοδολογία που θα παρουσιάσουμε χρησιμοποιείται προκειμένου να βελτιώσει τις προβλέψεις μακροοικονομικών μεταβλητών, και για τον λόγο αυτό δεν δίνεται κάποια ερμηνεία στους παράγοντες, υπάρχουν περιπτώσεις που η ερμηνεία των παραγόντων αποτελεί το βασικό στόχο της ανάλυσης. Για παράδειγμα, σε ένα σύνολο δεδομένων με πολλά μακροοικονομικά δεδομένα ο εκάστοτε ερευνητής μπορεί να θέλει να αποφανθεί αν μπορεί να εξάγει κάποιους κοινούς παράγοντες, οι οποίοι να μπορούν να εξηγήσουν τους κοινούς "επιχειρησιακούς κύκλους" (business cycles) των αρχικών μεταβλητών.

Προκειμένου να ερμηνεύσει κανείς τους παράγοντες αρκεί να χρησιμοποιήσει τα φορτία που εκτίμησε. Όπως είδαμε στην §1.2 η συσχέτιση του j παράγοντα με την i μεταβλητή δίνεται από τον τύπο

$$\rho(X_i, F_j) = \frac{l_{ij}}{\sigma_i}.$$

Αν τώρα για ευκολία θεωρήσουμε ότι έχουμε τυποποιήσει τα δεδομένα μας εύκολα φαίνεται ότι τα εκτιμηθέντα φορτία μας δίνουν το συντελεστή συσχέτισης $\rho(X_i, F_j)$. Συνήθως δεν είναι εφικτό να ερμηνεύσουμε τους παράγοντες κατευθείαν μόλις εξάγουμε τα φορτία αυτών και για τον λόγο αυτό μια περιστροφή τους όπως αυτή που δώσαμε στην §1.4 είναι αναγκαία.

Παρακάτω δίνεται ένα παράδειγμα από το βιβλίο των Johnson και Wichern[22]. Το παράδειγμα αυτό αναφέρεται σε μία μελέτη αναφορικά με τις προτιμήσεις των καταναλωτών. Στη μελέτη αυτή ζητήθηκε από τους καταναλωτές που πήραν μέρος να αξιολογήσουν κάποια χαρακτηριστικά ενός νέου προϊόντος. Τα χαρακτηριστικά φαίνονται στον πίνακα 1.

Μεταβλητή	Εκτιμηθέντα φορτία		Εκτιμηθέντα φορτία μετά από περιστροφή	
	F_1	F_2	F_1^*	F_2^*
Γεύση	0.56	0.82	0.02	0.99
Άξιζε τα λεφτά του	0.78	-0.52	0.94	-0.01
Άρωμα	0.65	0.75	0.13	0.98
Κατάλληλο για γεύμα	0.94	-0.10	0.84	0.43
Πλούσιο σε ενέργεια	0.80	-0.54	0.97	-0.02

Πίνακας 1: Συσχετίσεις παραγόντων με τις αρχικές μεταβλητές.

Από τον πίνακα 1 μπορούμε να δούμε ότι μετά τον αρχικό εντοπισμό των δύο παραγόντων ήταν αρκετά δύσκολο να δοθεί μία φυσική ερμηνεία σε αυτούς. Ωστόσο ύστερα από την εφαρμογή της περιστροφής varimax το πρόβλημα επιλύθηκε. Συγκεκριμένα ο πρώτος παράγοντας μπορούμε να πούμε ότι αποτελεί τον "θρεπτικό παράγοντα" και ο δεύτερος τον "παράγοντα γεύσης".

2 Το δυναμικό μοντέλο της ανάλυσης παραγόντων

2.1 Εισαγωγή

Στις προηγούμενες παραγράφους δώσαμε μία περιγραφή του κλασσικού μοντέλου της ανάλυσης παραγόντων. Οι συνθήκες που δώσαμε ωστόσο στην §1.2 είναι αρκετά περιοριστικές όσον αφορά στην περίπτωση δεδομένων που εξελίσσονται στον χρόνο. Αν από εδώ και στο εξής συμβολίσουμε με f_t τους παράγοντες που προκύπτουν από ένα σύνολο δεδομένων που αποτελείται από χρονοσειρές προκειμένου να ληφθεί υπόψιν η χρονική εξάρτηση που υπάρχει μεταξύ των μετρήσεων είναι πιο εύκολο να υποθέσουμε ότι τα f_t παρουσιάζουν σειριακή εξάρτηση. Για παράδειγμα θα μπορούσε κανείς να θεωρήσει ότι τα f_t ακολουθούν ένα υπόδειγμα Vector Autoregressive (VAR). Πιο συγκεκριμένα το δυναμικό μοντέλο της ανάλυσης παραγόντων μπορεί να θεωρηθεί ως ειδική περίπτωση ενός μοντέλου χώρου καταστάσεων (state-space model). Σε αυτό το μοντέλο το διάνυσμα κατάστασης αποτελείται από τους κοινούς παράγοντες, οι οποίοι έχουν μοντελοποιηθεί χρησιμοποιώντας ένα κλασσικό υπόδειγμα ανάλυσης χρονοσειρών (π.χ. VAR, VARMA). Συνήθως απαιτούμε ο αριθμός των παραγόντων που θα εξάγουμε να είναι πολύ μικρότερος του αριθμού των χρονοσειρών που είχαμε αρχικά στη διάθεση μας.

Η ιδέα των δυναμικών μοντέλων της ανάλυσης παραγόντων πρωτοεισήχθηκε το 1977 από τον Geweke[18] και από τους Sargent και Sims[29]. Έκτοτε οι νέες δυνατότητες που έδωσε στην ανάλυση χρονοσειρών η ανάλυση παραγόντων οδήγησε αρκετούς ερευνητές να ασχοληθούν με το συγκεκριμένο ερευνητικό αντικείμενο με κυριότερες να θεωρούνται για παράδειγμα οι συνεισφορές των Engle και Watson[16], Stock και Watson [31] και Doz, Giannone και Reichlin[13]. Τη μεθοδολογία των τελευταίων θα παρουσιάσουμε και θα εφαρμόσουμε στην εργασία αυτή.

Στο κεφάλαιο αυτό θα διατυπώσουμε αρχικά το μοντέλο χώρου-καταστάσεων στη γενική του μορφή. Στη συνέχεια θα δώσουμε τις εξισώσεις του φίλτρου και της εξομάλυνσης του Kalman (Kalman filter & Kalman smoother)[23] προκειμένου να εκτιμήσουμε την μορφή του μοντέλου αυτού. Κλείνοντας το κεφάλαιο αυτό θα δείξουμε πως ο αλγόριθμος EM σε συνδυασμό με το

φίλτρο και την εξομάλυνση του Kalman μπορεί να μας βοηθήσει προκειμένου να λάβουμε καλύτερες εκτιμήσεις των παραμέτρων του μοντέλου χώρου-καταστάσεων.

2.2 Μοντέλο χώρου καταστάσεων

Πρωτού παρουσιάσουμε το μοντέλο που πρότειναν οι Doz, Giannone και Reichelin καλό θα ήταν να εισάγουμε την έννοια των μοντέλων χώρου-καταστάσεων.

Στην βιβλιογραφία μπορεί να συναντήσει κανείς το μοντέλο χώρου-καταστάσεων γραμμένο σε διαφορετικές μορφές[28],[15]. Η παρουσίαση μας είναι αυτής των Durbin και Koopman[15]. Συγκεκριμένα το μοντέλο στην γενική του μορφή γράφεται ως σύστημα δύο εξισώσεων

$$y_t = \mathbf{Z}_t \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \mathbf{H}_t), \quad (2.1)$$

$$\alpha_{t+1} = \mathbf{T}_t \alpha_t + \mathbf{R}_t \eta_t, \quad \eta_t \sim N(0, \mathbf{Q}_t) \quad (2.2)$$

όπου y_t είναι ένα $p \times 1$ διάνυσμα και το α_t είναι ένα $m \times 1$ διάνυσμα, τα οποία καλούνται παρατηρούμενο διάνυσμα (observation vector) και διάνυσμα κατάστασης (state vector) αντίστοιχα. Κατά αντιστοιχία η εξίσωση (2.1) καλείται εξίσωση παρατήρησης και η εξίσωση (2.2) καλείται εξίσωση κατάστασης. Αρχικά θα θεωρήσουμε ότι οι πίνακες $\mathbf{Z}_t, \mathbf{T}_t, \mathbf{R}_t, \mathbf{H}_t$ και \mathbf{Q}_t είναι γνωστοί. Όσον αφορά στην από κοινού στοχαστική διαδικασία των $\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}$ αυτή έχει μέσο το μηδενικό διάνυσμα και είναι σειριακά ασυσχέτιστη διαδικασία λευκού θορύβου με πιθανώς χρονικά μεταβαλλόμενους πίνακες συνδιακυμάνσεων. Τέλος η αρχική κατάσταση $\alpha_1 \sim N(a_1, P_1)$, όπου με P_1 συμβολίσαμε τη διακύμανση της αρχικής κατάστασης, είναι ασυσχέτιστη με τα ε_t, η_t για όλα τα t .

Στην πράξη το πιθανότερο είναι να μην γνωρίζουμε κανέναν από τους πίνακες που δώσαμε παραπάνω. Η εκτίμησή τους σε αυτή την περίπτωση μπορεί να γίνει χρησιμοποιώντας την επαναληπτική διαδικασία που πρότεινε ο Kalman γνωστή ως φίλτρο του Kalman. Τα βασικά χαρα-

κτηριστικά αυτής θα δώσουμε στην επόμενη παράγραφο.

2.3 Το φίλτρο του Kalman

Προκειμένου να εξάγουμε τις απαραίτητες για την εφαρμογή του αλγορίθμου σχέσεις αρχικά θα πρέπει να διατυπώσουμε τη μορφή του μοντέλου χώρου καταστάσεων. Η μορφή του θα είναι αυτή που δίνεται στις σχέσεις (2.1) και (2.2) με την επιπλέον υπόθεση ότι η αρχική κατάσταση είναι γνωστή και συγκεκριμένα,

$$\alpha_1 \sim N(a_1, P_1).$$

Η διατύπωση αυτή πέρα από το γεγονός ότι μας επιτρέπει να εξάγουμε τις σχέσεις που θέλουμε σε μία σχετικά συμπαγή μορφή είναι επίσης αρκετά κοντά στο μοντέλο που θα εφαρμόσουμε στο δεύτερο μέρος της εργασίας.

Στόχος μας λοιπόν είναι να εκτιμήσουμε τις δεσμευμένες κατανομές των α_t και α_{t+1} δεδομένης της πληροφορίας που έχουμε έως τον χρόνο t . Το σύνολο που περιέχει αυτή την πληροφορία θα το συμβολίζουμε με Y_t . Γίνεται σαφές ότι ο πίνακας Y_T θα περιέχει όλα τα y_t έως τον χρόνο T .

Αν θεωρήσουμε ότι τα a_1, P_1 , που αφορούν στην αρχική κατάσταση α_1 μας είναι γνωστά και θέσουμε $a_{t|t} = E(\alpha_t|Y_t)$, $a_{t+1} = E(\alpha_{t+1}|Y_t)$, $P_{t|t} = Var(\alpha_t|Y_t)$ και $P_{t+1} = Var(\alpha_{t+1}|Y_t)$ μπορούμε να δείξουμε προς προκύπτει ο αλγόριθμος του Kalman.

Ξεκινάμε ορίζοντας ως σφάλμα πρόβλεψης για μία περίοδο προς τα μπροστά δεδομένου του Y_{t-1} το $v_t = y_t - E(y_t|Y_{t-1})$ και χρησιμοποιώντας την παραμετρική υπόθεση που δίνεται στην σχέση (2.1) έχουμε ότι

$$v_t = y_t - E(\mathbf{Z}_t\alpha_t + \varepsilon_t|Y_{t-1}) = y_t - \mathbf{Z}_ta_t \quad (2.3)$$

και κατά συνέπεια $E(\alpha_t|Y_t) = E(\alpha_t|Y_{t-1}, v_t)$. Ακολούθως μπορούμε να δούμε ότι $E(v_t|Y_{t-1}) = E(y_t - \mathbf{Z}_t a_t|Y_{t-1}) = E(\varepsilon_t|Y_{t-1}) = 0$, το οποίο μας οδηγεί στο ότι $E(v_t) = 0$ και $Cov(y_j, v_t) = E[y_j E(v_t|Y_{t-1})'] = 0$ για κάθε $j = 1, \dots, t-1$. Γνωρίζουμε ότι η από κοινού κατανομή των α_t και v_t δεδομένου του Y_{t-1} είναι κανονική. Χρησιμοποιώντας το Λήμμα 1 του παραρτήματος Α μπορούμε να δούμε ότι

$$\alpha_{t|t} = E(\alpha_t|Y_{t-1}) + Cov(\alpha_t, v_t)[Var(v_t)]^{-1}v_t. \quad (2.4)$$

Όμως $Cov(\alpha_t, v_t) = E[\alpha_t(\mathbf{Z}_t\alpha_t + \varepsilon_t - \mathbf{Z}_t a_t)'|Y_{t-1}] = E[\alpha_t(\alpha_t - a_t)'\mathbf{Z}_t'|Y_{t-1}] = P_t\mathbf{Z}_t'$ και επιπλέον $F_t = Var(v_t|y_{t-1}) = Var(\mathbf{Z}_t\alpha_t + \varepsilon_t - \mathbf{Z}_t a_t|Y_{t-1}) = \mathbf{Z}_t P_t \mathbf{Z}_t' + \mathbf{H}_t$. Τέλος έχουμε ότι $E(\alpha_t|Y_{t-1}) = a_t$. Λαμβάνοντας υπόψιν τα παραπάνω, η (2.6) αποκτά την παρακάτω μορφή

$$a_{t|t} = a_t + P_t\mathbf{Z}_t'F_t^{-1}v_t. \quad (2.5)$$

Στη συνέχεια χρησιμοποιώντας πάλι το Λήμμα 1 έχουμε ότι

$$P_{t|t} = Var(\alpha_t|Y_{t-1}, v_t) = Var(\alpha_t|Y_{t-1}) - Cov(\alpha_t, v_t)[Var(v_t)]^{-1}Cov(\alpha_t, v_t)'$$

και σύμφωνα με αυτά που είδαμε παραπάνω καταλήγουμε στη σχέση

$$P_{t|t} = P_t - P_t\mathbf{Z}_t'F_t^{-1}\mathbf{Z}_t P_t. \quad (2.6)$$

Κλείνοντας θα δώσουμε τις σχέσεις που μας επιτρέπουν να υπολογίσουμε τα a_{t+1} και P_{t+1} . Χρησιμοποιώντας τη σχέση (2.1) έχουμε ότι

$$a_{t+1} = E(\mathbf{T}_t\alpha_t + \mathbf{R}_t\eta_t|Y_t) = \mathbf{T}_t E(\alpha_t|Y_t)$$

και αντικαθιστώντας την (2.5) σε αυτήν λαμβάνουμε

$$a_{t+1} = \mathbf{T}_t a_{t|t} = \mathbf{T}_t a_t + K_t v_t. \quad (2.7)$$

Επίσης

$$P_{t+1} = \text{Var}(\mathbf{T}_t \alpha_t + \mathbf{R}_t \eta_t | Y_t) = T_t \text{Var}(\alpha_t | Y_t) \mathbf{T}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'$$

και αντικαθιστώντας την (1.6) σε αυτήν λαμβάνουμε

$$P_{t+1} = \mathbf{T}_t P_t (\mathbf{T}_t - K_t \mathbf{Z}_t)' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t' \quad (2.8)$$

όπου $K_t = \mathbf{T}_t P_t \mathbf{Z}_t' F_t^{-1}$, το οποίο στη βιβλιογραφία το συναντάμε ως Kalman gain.

Στον παρακάτω πίνακα δίνονται συνοπτικά όλες οι απαραίτητες σχέσεις που εξάγαμε παραπάνω για την εφαρμογή του φίλτρου του Kalman.

$v_t = y_t - \mathbf{Z}_t a_t$	$F_t = \mathbf{Z}_t P_t \mathbf{Z}_t' + \mathbf{H}_t$
$a_{t t} = a_t + P_t \mathbf{Z}_t' F_t^{-1} v_t$	$P_{t t} = P_t - P_t \mathbf{Z}_t' F_t^{-1} \mathbf{Z}_t P_t$
$a_{t+1} = \mathbf{T}_t a_t + K_t v_t$	$P_{t+1} = \mathbf{T}_t P_t (\mathbf{T}_t - K_t \mathbf{Z}_t)' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'$

Πίνακας 2: Εξισώσεις φιλτραρίσματος.

Χρησιμοποιώντας τις σχέσεις του Πίνακα 2 μπορούμε να ενημερώνουμε τις προγνώσεις μας κάθε φορά που έχουμε στην διάθεσή μας μία καινούργια παρατήρηση.

Για να γίνει πιο κατανοητός ο τρόπος με τον οποίο λειτουργεί το φίλτρο αυτό θα δώσουμε ένα απλό παράδειγμα. Στο παράδειγμά μας θα κάνουμε αρκετές περιοριστικές υποθέσεις προκειμένου αυτό να γίνει εύκολα αντιληπτό και να μη χαθεί ο αναγνώστης σε πολύπλοκους μαθηματικούς υπολογισμούς. Το παράδειγμα που χρησιμοποιήθηκε αποτελεί μέρος του ψηφιακού εκπαιδευτικού υλικού του Cornell University⁵.

⁵<https://www.cs.cornell.edu/courses/cs4758/2012sp/materials/mi63slides.pdf>

Παράδειγμα 2.1. Έστω ότι έχουμε το μοντέλο χώρου-καταστάσεων που δίνεται από τις σχέσεις (2.3) και (2.4). Ας υποθέσουμε για ευκολία ότι το διάνυσμα κατάστασης είναι βαθμωτό και ότι το $T_t = 1$. Επιπλέον, υποθέτουμε ότι το $Z_t = 1$ και το $R_t = 1$. Κάτω από αυτές τις υποθέσεις και μετά από μερικές απλές αλγεβρικές πράξεις οι εξισώσεις του Πίνακα 2 λαμβάνουν την εξής μορφή,

$v_t = y_t - a_t$	$f_t = p_t + h_t$
$a_{t t} = a_t + k_t v_t$	$p_{t t} = (1 - k_t)p_t$
$a_{t+1} = a_t + k_t v_t$	$p_{t+1} = p_{t t} + q_t$

Πίνακας 3: Εξισώσεις φιλτραρίσματος για το παράδειγμα 2.1.

με το Kalman gain να είναι ίσο με $k_t = p_t(p_t + h)^{-1}$. Επίσης, αντικαταστήσαμε τα κεφαλαία με μικρά γράμματα προκειμένου να φανεί ότι δε δουλεύουμε με πίνακες πλέον.

Στο πρώτο βήμα θεωρώντας ότι το $y_1 = 0.9$ με $h_t = h = 0.1$, $q_t = q = 0.0001$ και ότι η αρχική κατάσταση είναι $a_0 = 0$ με διακύμανση, p_0 , ίση με 1000, έχουμε ότι

1. Πρόβλεψη (Predict)

$$a_1 = 0$$

$$p_1 = 1000 + 0.0001$$

2. Αναθεώρηση (Update)

$$K_1 = 1000.0001(1000.0001 + 0.1)^{-1} = 0.9999$$

$$a_{1|1} = 0 + 0.9999(0.9 - 0) = 0.8999$$

$$p_{1|1} = (1 - 0.9999)1000.0001 = 0.1$$

Στο επόμενο βήμα για $y_2 = 0.8$ έχουμε,

1. Πρόβλεψη

$$a_2 = 0.8999$$

$$p_2 = 0.1 + 0.0001 = 0.1001$$

2. Αναθεώρηση

$$K_2 = 0.1001(0.1001 + 0.1)^{-1} = 0.5002$$

$$a_{2|2} = 0 + 0.8999 + 0.5002(0.8 - 0.8999) = 0.8499$$

$$p_{2|2} = (1 - 0.5002)0.1001 = 0.05$$

Συνεχίζοντας κατ' αυτόν τον τρόπο μπορούμε να πάρουμε εκτιμήσεις για το διάνυσμα κατάστασης μέχρι να εξαντλήσουμε όλα τα διαθέσιμα δεδομένα.

2.4 Εξομάλυνση Kalman

Κατά την εξομάλυνση στόχος μας είναι να υπολογίσουμε την υπό συνθήκη πυκνότητα του διανύσματος κατάστασης, α_t , χρησιμοποιώντας όλη την διαθέσιμη πληροφορία Y_T . Χρησιμοποιώντας το Λήμμα 1 και δεδομένου ότι τα v_t, \dots, v_n είναι ανά δύο ανεξάρτητα και ανεξάρτητα του Y_{t-1} με μηδενικό μέσο έχουμε ότι

$$\hat{\alpha}_t = E(\alpha_t | Y_n) = E(\alpha_t | Y_{t-1}, v_t, \dots, v_n) = a_t + \sum_{j=t}^n Cov(\alpha_t, v_j) F_j^{-1} v_j. \quad (2.9)$$

Αν ορίσουμε ως σφάλμα εκτίμησης του διανύσματος κατάστασης το $x_t = \alpha_t - a_t$ με διακύμανση ίση με P_t βλέπουμε ότι

$$v_t = y_t - \mathbf{Z}_t a_t = \mathbf{Z}_t a_t + \varepsilon_t - \mathbf{Z}_t a_t = \mathbf{Z}_t (\alpha_t - a_t) + \varepsilon_t = \mathbf{Z}_t x_t + \varepsilon_t \quad (2.10)$$

και

$$\begin{aligned}
x_{t+1} &= \alpha_{t+1} - a_{t+1} \\
&= \mathbf{T}_t \alpha_t + \mathbf{R}_t \eta_t - \mathbf{T}_t a_t - K_t v_t \\
&= \mathbf{T}_t x_t + \mathbf{R}_t \eta_t - K_t \mathbf{Z}_t x_t - K_t \varepsilon_t \\
&= L_t x_t + \mathbf{R}_t \eta_t - K_t \varepsilon_t
\end{aligned} \tag{2.11}$$

όπου $L_t = \mathbf{T}_t - K_t \mathbf{Z}_t$. Χρησιμοποιώντας τη σχέση (2.12) έχουμε ότι

$$\begin{aligned}
Cov(\alpha_t, v_j) &= E(\alpha_t v_j' | Y_{t-1}) \\
&= E[\alpha_t (\mathbf{Z}_j x_j + \varepsilon_j)' | Y_{t-1}] \\
&= E(a_t x_j' | Y_{t-1}) \mathbf{Z}_j', \quad j = t, \dots, n
\end{aligned} \tag{2.12}$$

και μέσω της (2.13) κάνοντας διαδοχικές αντικαταστάσεις βρίσκουμε ότι

$$\begin{aligned}
E(\alpha_t x_t' | Y_{t-1}) &= P_t \\
E(\alpha_t x_j' | Y_{t-1}) &= P_t \prod_{i=t}^{j-1} L_i', \quad j = t+1, \dots, n.
\end{aligned} \tag{2.13}$$

Χρησιμοποιώντας το παραπάνω αποτέλεσμα στη σχέση (2.11) βλέπουμε ότι

$$\hat{\alpha}_t = a_t + P_t r_{t-1} \tag{2.14}$$

όπου

$$r_n = 0$$

$$r_{n-1} = \mathbf{Z}'_n F_n^{-1} v_n$$

$$r_{n-2} = \mathbf{Z}'_{n-1} F_{n-1}^{-1} v_{n-1} + L'_{n-1} \mathbf{Z}'_n F_n^{-1} v_n \quad (2.15)$$

$$r_t = \mathbf{Z}'_{t+1} F_{t+1}^{-1} v_{t+1} + L'_{t+1} \mathbf{Z}'_{t+2} F_{t+2}^{-1} v_{t+2} + \dots + \prod_{i=t+1}^{n-1} L'_i \mathbf{Z}'_n F_n^{-1} v_n$$

$$r_{t-1} = \mathbf{Z}'_t F_t^{-1} v_t + L'_t \mathbf{Z}'_{t+1} F_{t+1}^{-1} v_{t+1} + \dots + \prod_{i=t}^{n-1} L'_i \mathbf{Z}'_n F_n^{-1} v_n, \quad t = 1, \dots, n-3, n-2.$$

Πηγαίνοντας προς τα πίσω και κάνοντας διαδοχικές αντικαταστάσεις ξεκινώντας από τη σχέση που δίνεται στην (2.15) για το r_t λαμβάνουμε την παρακάτω εξίσωση

$$r_{t-1} = \mathbf{Z}'_t F_t^{-1} v_t + L'_t r_t, \quad t = 1, \dots, n. \quad (2.16)$$

Καταλήγοντας οι εξισώσεις (2.14) και (2.16) μπορούν να χρησιμοποιηθούν για να υπολογιστούν τα $\hat{\alpha}_1, \dots, \hat{\alpha}_n$.

Όσον αφορά στον πίνακα διακύμανσης του διανύσματος κατάστασης, αυτός μπορεί να υπολογιστεί χρησιμοποιώντας πάλι το Λήμμα 1 από το οποίο έχουμε ότι

$$V_t = Var(\alpha_t | Y_{t-1}, v_t, \dots, v_n) = P_t - \sum_{j=t}^n Cov(\alpha_t, v_j) F_j^{-1} Cov(\alpha_t, v_j)'$$

και από τις σχέσεις (2.14),(2.15) λαμβάνουμε άμεσα την ακόλουθη εξίσωση

$$\begin{aligned} V_t &= P_t - P_t \mathbf{Z}'_t F_t^{-1} \mathbf{Z}_t P_t - P_t L'_t \mathbf{Z}'_{t+1} F_{t+1}^{-1} \mathbf{Z}_{t+1} L_t P_t - \dots - P_t L'_t \dots L'_{n-1} \mathbf{Z}'_n F_n^{-1} \mathbf{Z}_n L_{n-1} \dots L_t P_t \\ &= P_t - P_t N_{t-1} P_t \end{aligned} \quad (2.17)$$

όπου

$$N_{t-1} = \mathbf{Z}'_t F_t^{-1} \mathbf{Z}_t + L'_t \mathbf{Z}'_{t+1} F_{t+1}^{-1} \mathbf{Z}_{t+1} L_t + \cdots + L'_t \cdots L'_{n-1} \mathbf{Z}'_n F_n^{-1} \mathbf{Z}_n L_{n-1} \cdots L_{t+1}. \quad (2.18)$$

Έχουμε επίσης

$$\begin{aligned} N_t = & \mathbf{Z}'_{t+1} F_{t+1}^{-1} \mathbf{Z}_{t+1} + L'_{t+1} \mathbf{Z}'_{t+2} F_{t+2}^{-1} \mathbf{Z}_{t+2} L_{t+1} + \cdots \\ & + L'_{t+1} \cdots L'_{n-1} \mathbf{Z}'_n F_n^{-1} \mathbf{Z}_n L_{n-1} \cdots L_{n+1} \end{aligned} \quad (2.19)$$

Πηγαίνοντας προς τα πίσω και κάνοντας διαδοχικές αντικαταστάσεις ξεκινώντας από τη σχέση που δίνεται παραπάνω για το N_t λαμβάνουμε την παρακάτω εξίσωση

$$N_{t-1} = \mathbf{Z}'_t F_t^{-1} \mathbf{Z}_t + L'_t N_t L_t, \quad t = 1, \dots, n \quad (2.20)$$

Για $t = n$ έχουμε ότι $N_n = 0$.

Οι εξισώσεις (2.17) και (2.20) μπορούν να χρησιμοποιηθούν για να υπολογιστεί ο πίνακας διακύμανσης, V_t , του διανύσματος κατάστασης.

Στον παρακάτω πίνακα δίνονται συνοπτικά όλες οι απαραίτητες σχέσεις που εξάγαμε παραπάνω για την εφαρμογή της εξομάλυνσης Κάλμαν.

$r_{t-1} = \mathbf{Z}'_t F_t^{-1} v_t + L'_t r_t$	$N_{t-1} = \mathbf{Z}'_t F_t^{-1} \mathbf{Z}_t + L'_t N_t L_t$
$\hat{a}_t = a_t + P_t r_{t-1}$	$V_t = P_t - P_t N_{t-1} P_t$

Πίνακας 4: Εξισώσεις εξομάλυνσης.

Οι παραπάνω σχέσεις προτάθηκαν από τον de Jong[10],[11] και τους Kohn και Ansley[25].

Σε συνέχεια του Παραδείγματος 2.1 θα πραγματοποιήσουμε δύο βήματα της εξομάλυνσης Kalman.

Παράδειγμα 2.2. Ύστερα από δέκα βήματα του φίλτρου του Kalman, χρησιμοποιώντας τις υποθέσεις του Παραδείγματος 2.1 και τις τιμές που δίνονται στην ηλεκτρονική διεύθυνση που παραθέσαμε προηγουμένως έχουμε ότι $a_{10} = 0.972$ με $p_{10} = 0.011$ και $a_9 = 0.99$ με $p_9 = 0.013$.

1. Από την τελευταία παρατήρηση ξεκινά το πρώτο βήμα του αλγορίθμου με αρχικές τιμές $N_{10} = 0$ και $r_{10} = 0$.

Εφαρμόζοντας τις σχέσεις που δίνονται στον πίνακα 4 έχουμε

$$\begin{aligned} N_9 &= \frac{1}{0.011 + 0.1} = 9.009, \\ r_9 &= \frac{1.15 - 0.972}{0.011 + 0.1} = 1.604, \\ \hat{\alpha}_{10} &= 0.972 + 0.011 \cdot 1.604 = 0.99, \\ V_{10} &= 0.011 - 0.011^2 \cdot 9.009 = 0.01. \end{aligned}$$

2. Στο δεύτερο βήμα προχωράμε ένα βήμα πιο πίσω και εφαρμόζοντας ξανά τις σχέσεις του Πίνακα 4 έχουμε

$$\begin{aligned} N_8 &= \frac{1}{0.013 + 0.1} = 8.85, \\ r_8 &= \frac{1.85 - 0.99}{0.013 + 0.1} = -1.24, \\ \hat{\alpha}_9 &= 0.99 - 0.013 \cdot 1.24 = 0.974, \\ V_9 &= 0.013 - 0.013^2 \cdot 8.85 = 0.01. \end{aligned}$$

Συνεχίζοντας με αυτόν τον τρόπο μέχρι να φτάσουμε στην πρώτη (χρονικά) διαθέσιμη παρατήρηση μπορούμε να λάβουμε εξομαλυσμένες τιμές του διανύσματος κατάστασης. Ουσιαστικά χρησιμοποιώντας την εξομάλυνση του Kalman μπορούμε να πετύχουμε καλύτερη προσαρμογή στα αρχικά δεδομένα. Αυτό μπορούμε να το δούμε και στα δύο βήματα του παραδείγματος 2.2. Στο πρώτο βήμα η τιμή που είχαμε λάβει από το φίλτρο ήταν μικρότερη της πραγματικής και για

τον λόγο αυτό η εξομάλυνση την αύξησε. Αντίθετα στο δεύτερο βήμα το φίλτρο μας έδωσε τιμή μεγαλύτερη της πραγματικής, την οποία ελάττωσε η εξομάλυνση. Ακόμα μια σημαντική ιδιότητα της εξομάλυνσης του Kalman είναι ότι μπορεί να μας δώσει τιμή προγενέστερη της πρώτης (χρονικά) διαθέσιμες παρατήρησης λόγω της αναδρομικής λειτουργίας του.

2.5 Ο αλγόριθμος EM

Έχοντας διαθέσιμες αρχικές εκτιμήσεις για τους πίνακες $\mathbf{Z}_t, \mathbf{T}_t, \mathbf{H}_t, \mathbf{Q}_t$ αυτό που μας ενδιαφέρει είναι να βελτιώσουμε όσο το δυνατόν περισσότερο τις εκτιμήσεις αυτές. Για το σκοπό αυτό θα χρησιμοποιήσουμε τον αλγόριθμο EM. Ο αλγόριθμος EM μεγιστοποιεί την παρακάτω ποσότητα

$$\begin{aligned}
G(\mu, \Sigma, \mathbf{Z}_t, \mathbf{T}_t, \mathbf{Q}_t, \mathbf{H}_t) = & -\frac{1}{2} \log |\Sigma| - \frac{1}{2} \text{tr} \{ \Sigma^{-1} (P_0^n + (x_0^n - \mu)(x_0^n - \mu)') \} \\
& - \frac{n}{2} \log |\mathbf{Q}_t| - \frac{1}{2} \text{tr} \{ \mathbf{Q}_t^{-1} (C - B\mathbf{T}_t' - \mathbf{T}_t B + \mathbf{T}_t A \mathbf{T}_t') \} \\
& - \frac{n}{2} \log |\mathbf{H}_t| \\
& - \frac{1}{2} \text{tr} \{ \mathbf{H}_t^{-1} \sum_{t=1}^T [(y_t - \mathbf{Z}_t x_t^T)(y_t - \mathbf{Z}_t x_t^T)' + \mathbf{Z}_t P_t^T \mathbf{Z}_t'] \},
\end{aligned} \tag{2.21}$$

η οποία προκύπτει αν πάρουμε τη μέση τιμή του λογαρίθμου της πιθανοφάνειας των $\alpha_1, \dots, \alpha_t,$
 y_1, \dots, y_t

$$\begin{aligned}
\log L = & -\frac{1}{2} \log |\Sigma| - \frac{1}{2} (\alpha_t - \mu)' \Sigma^{-1} (\alpha_t - \mu) \\
& - \frac{n}{2} \log |\mathbf{Q}_t| - \frac{1}{2} \sum_{t=1}^T (\alpha_t - \mathbf{T}_t \alpha_{t-1})' \mathbf{Q}_t^{-1} \sum_{t=1}^T (\alpha_t - \mathbf{T}_t \alpha_{t-1}) \\
& - \frac{n}{2} \log |\mathbf{H}_t| - \frac{1}{2} \sum_{t=1}^T (y_t - \mathbf{Z}_t \alpha_t)' \mathbf{H}_t^{-1} \sum_{t=1}^n (y_t - \mathbf{Z}_t \alpha_t).
\end{aligned} \tag{2.22}$$

Συνοπτικά ο αλγόριθμος EM έχει τα επόμενα βήματα:

1 Αρχικοποίηση των πινάκων $\mathbf{T}_t, \mathbf{Q}_t, \mathbf{H}_t$ χρησιμοποιώντας τα αποτελέσματα της κλασσικής

ανάλυσης παραγόντων.

2 Υπολογισμός των α_t, P_t χρησιμοποιώντας τις εξισώσεις που δίνονται στον Πίνακα 2 της Παραγράφου 2.3.

3 Υπολογισμός των $\mathbf{T}_t, \mathbf{Q}_t, \mathbf{H}_t$ μέσω των σχέσεων⁶

$$(\alpha') \mathbf{T}_t = \mathbf{B}\mathbf{A}^{-1},$$

$$(\beta') \mathbf{Q}_t = T^{-1}(\mathbf{C} - \mathbf{B}\mathbf{A}^{-1}\mathbf{B}'),$$

$$(\gamma') \mathbf{H}_t = T^{-1} \sum_{t=1}^T [(y_t - \mathbf{Z}_t x_t^T)(y_t - \mathbf{Z}_t x_t^T)' + \mathbf{Z}_t P_t^T \mathbf{Z}_t'], \text{ όπου}$$

$$\begin{aligned} \mathbf{A} &= \sum_{t=1}^T (P_{t-1} + \alpha_{t-1} \alpha'_{t-1}) \\ \mathbf{B} &= \sum_{t=1}^T (P_{t|t} + \alpha_t \alpha'_{t-1}) \\ \mathbf{C} &= \sum_{t=1}^T (P_t + \alpha_t \alpha' t) \end{aligned} \quad (2.23)$$

4 Επανάλαβε τα βήματα 2 και 3 μέχρι οι εκτιμήσεις και ο λογάριθμος της συνάρτησης πιθανοφάνειας να σταθεροποιηθούν⁷.

⁶Οι σχέσεις αυτές προκύπτουν μέσω της μεγιστοποίησης της ποσότητας (2.22).

⁷Πρακτικά εξετάζουμε αν η διαφορά του λογαρίθμου της πιθανοφάνειας που υπολογίσαμε στο βήμα r με τον λογάριθμο της πιθανοφάνειας στο βήμα $r - 1$ είναι μικρότερη από κάποιο όριο που θέτουμε εμείς.

3 Το μοντέλο των Doz, Giannone, Reichlin

3.1 Εισαγωγή

Το μοντέλο αυτό διατυπώθηκε από τους Doz, Giannone, Reichlin[13] το 2011 και χρησιμοποιείται κατά κύριο λόγο από μεγάλους διεθνείς οργανισμούς και κεντρικές τράπεζες, λόγω του μεγάλου όγκου δεδομένων που έχουν στη διάθεση τους. Συγκεκριμένα η ευρωπαϊκή κεντρική τράπεζα τα τελευταία χρόνια έχει δείξει ιδιαίτερο ενδιαφέρον στην συγκεκριμένη και άλλες παραπλήσιες μεθοδολογίες που βασίζονται σε δυναμικά μοντέλα ανάλυσης παραγόντων (βλέπε [31],[24]).

Η βασική ιδέα του μοντέλου αυτού αφορά στην εκτίμηση του δυναμικού μοντέλου ανάλυσης παραγόντων σε δύο στάδια. Στο πρώτο στάδιο εκτιμούνται οι παράμετροι του μοντέλου εφαρμόζοντας την μέθοδο ελαχίστων τετραγώνων πάνω στους παράγοντες που θα εξάγουμε μέσω της μεθόδου των κυρίων συνιστωσών. Στο δεύτερο στάδιο οι παράγοντες εκτιμούνται μέσω του φίλτρου και της εξομάλυνσης του Kalman προκειμένου να ληφθεί υπόψιν η χρονική μεταβολή των παραγόντων. Το δεύτερο στάδιο επαναλαμβάνεται με την χρήση του αλγορίθμου EM προκειμένου να βελτιωθούν οι εκτιμήσεις των παραμέτρων του μοντέλου.

Στο κεφάλαιο αυτό θα δούμε πως και κάτω από ποιές υπόθεσεις διατύπωσαν το μοντέλο τους οι Doz, Giannone, Reichlin. Επιπλέον, θα αναφέρουμε τέσσερα εναλλακτικά μοντέλα που πρότειναν οι συγγραφείς, τα οποία σχετίζονται με τους περιορισμούς που πρόκειται να επιβάλλει ο εκάστοτε ερευνητής κατά την εφαρμογή της μεθοδολογίας. Οι περιορισμοί αυτοί έχουν να κάνουν με τις ιδιότητες των παραγόντων, δηλαδή αν θα είναι στατικοί ή δυναμικοί, καθώς επίσης και με την μορφή του πίνακα Q_t .

3.2 Το μοντέλο και υποθέσεις

Το μοντέλο έχει την ίδια μορφή με αυτό της σχέσης (1.1) με μόνη διαφορά ότι θα αντικαταστήσουμε τον υποδείκτη i με t ώστε να δείξουμε την ύπαρξη χρονικής εξάρτησης. Συγκεκριμένα

έχουμε το πρότυπο

$$X_t = \mathbf{L}F_t + \varepsilon_t \quad (3.1)$$

στο οποίο οι διαστάσεις των διανυσμάτων που εμφανίζονται φαίνονται στον Πίνακα 5.

-	Διάσταση
X_t	$(n \times 1)$
\mathbf{L}	$(n \times r)$
F_t	$(r \times 1)$
ε_t	$(n \times 1)$

Πίνακας 5: Πίνακας διαστάσεων.

Οι διαδικασίες X_t , F_t και ε_t είναι στάσιμες και οι δύο τελευταίες ανεξάρτητες μεταξύ τους. Ουσιαστικά με την παραπάνω γραφή κάνουμε την υπόθεση ότι οι χρονοσειρές που μας δίνονται μπορούν να διασπαστούν σε δύο μεταξύ τους ανεξάρτητα μέρη. Το πρώτο αποτελείται από τους κοινούς παράγοντες και το δεύτερο από τους ειδικούς.

Επεκτείνοντας τη σχέση που είδαμε στην Παράγραφο 1.3, θα δώσουμε την αντίστοιχη μορφή για τον πίνακα αυτοσυνδιακύμανσης τάξης h :

$$\mathbf{\Sigma}(h) = \mathbf{L}\mathbf{L}' + \mathbf{\Psi}(h) \quad (3.2)$$

όπου $\mathbf{\Sigma}(h) = E[X_t X_{t-h}']$ με $\mathbf{\Sigma}(0) = \mathbf{\Sigma}$ και $\mathbf{\Psi}(h) = E[\varepsilon_t \varepsilon_{t-h}']$ με $\mathbf{\Psi}(0) = \mathbf{\Psi}$.

Το παραπάνω μοντέλο ισχύει κάτω από τις ακόλουθες υπόθεσεις:

- Y1. Για κάθε n , η $\{X_t\}$ είναι στάσιμη με μέση τιμή μηδέν και πεπεράσμενη ροπή δεύτερης τάξης.
- Y2. Τα X_{it} έχουν ομοιόμορφα φραγμένη κύμανση, δηλαδή $\exists M$ τέτοιο ώστε $\forall(i, t)$ να ισχύει $Var(X_{it}) = \sigma_{0,ii} \leq M$.

- Y3. - Οι ακολουθίες $\{F_t\}$ και $\{\varepsilon_t\}$ είναι ανεξάρτητες μεταξύ τους.
- Η $\{F_t\}$ δέχεται αναπαράσταση τύπου Wold, δηλαδή υπάρχει πολυώνυμο $C(L)$ τέτοιο ώστε $F_t = C(L)u_t$ με το $C(L)$ να έχει όλες τις ρίζες του εντός του μοναδιαίου κύκλου και το u_t να είναι στάσιμο έως τάξη τέσσερα.
- Η $\{\varepsilon_t\}$ δέχεται αναπαράσταση τύπου Wold, δηλαδή υπάρχει πολυώνυμο $D(L)$ τέτοιο ώστε $\varepsilon_t = D(L)v_t$ με το $D(L)$ να έχει όλες τις ρίζες του εντός του μοναδιαίου κύκλου και το v_t να είναι λευκός θόρυβος και να ικανοποιεί την συνθήκη $\exists M$ τέτοιο ώστε $\forall(n, i, t)$ να ισχύει $E[v_{it}^4] \leq M$.

Y4. Οι παράγοντες δέχονται αναπαράσταση VAR, δηλαδή υπάρχει πολυώνυμο $A(L)$ με $A(0) = I_r$, τέτοιο ώστε $A(L)F_t = u_t$ με τις ρίζες του $A(L)$ να βρίσκονται εκτός του μοναδιαίου κύκλου.

Y5. Ο πίνακας \mathbf{LL}' έχει διαφορετικές ιδιοτιμές⁸.

Στη συνέχεια, θα δώσουμε τέσσερις ακόμα υποθέσεις προκειμένου να περιορίσουμε την διασυσχέτιση (cross-correlation) μεταξύ των ε_t επιτρέποντας παράλληλα την ύπαρξη μιας μη ισχυρής μορφής αυτοσυσχέτισης μεταξύ των ε_t . Πιο συγκεκριμένα

$$\Sigma 1. \liminf_{n \rightarrow \infty} \frac{1}{n} \lambda_{\min}(\mathbf{LL}') > 0,$$

$$\Sigma 2. \text{ Το } \limsup_{n \rightarrow \infty} \frac{1}{n} \lambda_{\max}(\mathbf{LL}') \text{ είναι πεπερασμένο,}$$

$$\Sigma 3. \text{ Το } \limsup_{n \rightarrow \infty} \sum_{h \in \mathbb{Z}} \|\Psi(h)\| \text{ είναι πεπερασμένο,}$$

$$\Sigma 4. \inf \lambda_{\min}(\Psi) = \underline{\lambda} > 0.$$

Οι παραπάνω υποθέσεις αποτελούν επέκταση που έδωσαν οι Doz, Giannone, Reichlin στις υποθέσεις των Chamberlain και Rothschild[9].

⁸Η υπόθεση αυτή δεν είναι αναγκαία, ωστόσο όπως αναφέρουν οι συγγραφείς γίνεται για να αποφευχθούν ορισμένοι περιττοί υπολογισμοί.

Κλείνοντας, όπως είδαμε και στην Παράγραφο 1.4 οι παράγοντες δεν καθορίζονται μονοσήμαντα. Κάθε ορθογώνια περιστροφή αυτών μας δίνει διαφορετικούς παράγοντες διατηρώντας αναλλοίωτο τον διανυσματικό χώρο που δημιουργούν. Για τον λόγο αυτό οι συγγραφείς προτείνουν μία συγκεκριμένη κανονικοποίηση και την χρησιμοποιούν για να δείξουν την συνέπεια των αποτελεσμάτων τους. Έστω \mathbf{D} ο πίνακας του οποίου τα διαγώνια στοιχεία είναι οι ιδιοτιμές του πίνακα $\mathbf{L}\mathbf{L}'$ σε φθίνουσα σειρά και \mathbf{Q} ο πίνακας που περιέχει τα μοναδιαία ιδιοδιανύσματα που σχετίζονται με τον \mathbf{D} . Ακολουθώντας, ορίζουμε ως $\mathbf{L}^* = \mathbf{L}\mathbf{Q}$ και $\mathbf{P} = \mathbf{L}^*\mathbf{D}^{-1/2}$, έτσι ώστε $\mathbf{L}^*\mathbf{L}'^* = \mathbf{D}$, $\mathbf{L}^*\mathbf{L}'^* = \mathbf{L}\mathbf{L}'$ και $\mathbf{P}'\mathbf{P} = \mathbf{I}_r$. Τελικά έχουμε ότι οι κανονικοποιημένοι παράγοντες δίνονται από την σχέση $G_t = \mathbf{Q}'F_t$. Το μοντέλο μας αποκτά πλέον την εξής μορφή

$$X_t = \mathbf{L}^*G_t + \varepsilon_t \quad (3.3)$$

με πίνακα αυτοσυνδιακύμανσης

$$\mathbf{\Sigma}(h) = \mathbf{L}^*\mathbf{L}'^* + \mathbf{\Psi} \quad (3.4)$$

και να ικανοποιούνται όλες οι υποθέσεις (Y1-Y5) και (Σ 1- Σ 4) (βλ. Doz et al.[13]).

3.3 Προσεγγίσεις του παραμετρικού χώρου

Ιδιαίτερο ενδιαφέρον για την επιλογή του κατάλληλου μοντέλου προς εφαρμογή αποτελεί ο προσδιορισμός του παραμετρικού του χώρου. Στην γενική του μορφή αυτός διατυπώνεται ως εξής

$$\Omega = \{\mathbf{L}^*, A^*(L), D(L)\}. \quad (3.5)$$

Στην περίπτωση που ο παραμετρικός χώρος ήταν γνωστός και τα u_t, v_t ακολουθούσαν κανονική κατανομή, η καλύτερη προσέγγιση του G_t ως γραμμικός συνδιασμός των παρατηρήσεων X_1, \dots, X_T θα ήταν η $Proj_{\Omega}[G_t|X_s, s \leq T] = E_{\Omega}[G_t|X_s, s \leq T]$. Παράλληλα μία τέτοια προβολή ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα. Ωστόσο, η μορφή του παραμετρικού χώρου

δεν θα μας είναι γνωστή εκ των προτέρων και για το λόγο αυτό θα πρέπει να κάνουμε κάποιες υπόθεσεις σχετικά με αυτή. Στην εργασία αυτή θα ασχοληθούμε με τέσσερις διαφορετικές περιπτώσεις, όπως δίνονται στο [13].

Η απλούστερη εκδοχή θα ήταν ο

$$\Omega^{R1} = \{\mathbf{L}^*, I_r, \sqrt{\bar{\psi}}I_n\}$$

, όπου $\bar{\psi} = \frac{1}{n} \sum_{j=1}^n E(\varepsilon_{jt}^2)$. Στην περίπτωση αυτή θεωρούμε ότι οι κοινοί παράγοντες αποτελούν διαδικασία λευκού θορύβου και οι ειδικοί παράγοντες επίσης διαδικασία λευκού θορύβου με κοινή διακύμανση $\bar{\psi}$. Κάτω από αυτές τις υποθέσεις έχουμε ότι

$$Proj_{\Omega^{R1}}[G_t|X_s, s \leq T] = (\mathbf{L}^*\mathbf{L}^* + \bar{\psi}I_r)^{-1}\mathbf{L}^*X_t \quad (3.6)$$

το οποίο είναι ασυμπτωτικά αντίστοιχο του εκτιμητή ελαχίστων τετραγώνων που προκύπτει από την παλινδρόμηση της X_t στα φορτία των παραγόντων \mathbf{L}^* . Ο συγκεκριμένος εκτιμητής παραμένει συνεπής ακόμα και αν $A(L) \neq I_r$ και $D(L) \neq \sqrt{\bar{\psi}}I_n$ όταν το $n \rightarrow \infty$.

Στην περίπτωση που θεωρήσουμε ότι οι ειδικοί παράγοντες είναι ετεροσκεδαστικοί και ορίσουμε ως $\Psi_d = \text{diag}(\psi_{11}, \dots, \psi_{nn})$, ο παραμετρικός χώρος λαμβάνει την εξής μορφή

$$\Omega^{R2} = \{\mathbf{L}^*, I_r, \Psi_d^{1/2}\}. \quad (3.7)$$

Κάτω από αυτή την υπόθεση έχουμε ότι

$$Proj_{\Omega^{R2}}[G_t|X_s, s \leq T] = (\mathbf{L}^*\Psi_d^{-1}\mathbf{L}^* + I_r)^{-1}\mathbf{L}^*\Psi_d^{-1}X_t \quad (3.8)$$

Η παραπάνω εξίσωση μπορεί να δει κανείς ότι έχει την ίδια μορφή με τον τύπο του Bartlett που δώσαμε στην Παράγραφο 1.5.

Γενικεύοντας τις δύο παραπάνω περιπτώσεις θα λάβουμε υπόψιν την συμπεριφορά των παραγόντων στον χρόνο θεωρώντας ότι δέχονται αναπαράσταση VAR. Συγκεκριμένα μπορούμε να εκτιμήσουμε τους συντελεστές της εξίσωσης

$$\hat{G}_t = \hat{A}_1 \hat{G}_{t-1} + \dots + \hat{A}_p \hat{G}_{t-p} + \hat{w}_t$$

μέσω της μεθόδου ελαχίστων τετραγώνων και να τις χρησιμοποιήσουμε στο μοντέλο χώρου-καταστάσεων που θα δημιουργήσουμε. Αναλυτικότερα κάνοντας χρήση των παραμετρικών χώρων

$$\Omega^{R3} = \{\mathbf{L}^*, A(L), \sqrt{\bar{\psi}} I_n\} \quad (3.9)$$

$$\Omega^{R4} = \{\mathbf{L}^*, A(L), \Psi_d^{1/2}\} \quad (3.10)$$

και μετά την εφαρμογή της ανάλυσης κυρίων συνιστωσών ώστε να λάβουμε αρχικές εκτιμήσεις των παραμέτρων των χώρων Ω^{R3} και Ω^{R4} μπορούμε να εφαρμόσουμε την εξομάλυνση του Kalman προκειμένου να λάβουμε νέες εκτιμήσεις των παραγόντων, οι οποίες θα λαμβάνουν υπόψιν την αναπαράσταση VAR των παραγόντων. Αποτελέσματα σχετικά με την συνέπεια των παραπάνω εκτιμητών δίνονται στο [13]. Οι τέσσερις παραπάνω προσεγγίσεις μπορούν να βοηθήσουν τον ερευνητή να επιλέξει το μοντέλο που προσαρμόζεται καλύτερα στα δεδομένα που έχει στην διάθεση του. Οι προσεγγίσεις που χρησιμοποιούνται κυρίως είναι η Ω^{R3} και η Ω^{R4} .

Κλείνοντας την παράγραφο αυτή θα δώσουμε μία συνοπτική περιγραφή της μεθόδου εκτίμησης των παραγόντων που είδαμε παραπάνω. Αρχικά, εκτιμούμε τις παραμέτρους και τους παράγοντες μέσω της μεθόδου των κυρίων συνιστωσών. Στη συνέχεια, εκτιμούμε τους συντελεστές του υποδείγματος VAR μέσω της μεθόδου ελαχίστων τετραγώνων και τελικά αφού γράψουμε το μοντέλο σε μορφή μοντέλου χώρου-καταστάσεων χρησιμοποιούμε την εξομάλυνση και το φίλτρο του Kalman για να εκτιμήσουμε ξανά τις τιμές των παραγόντων. Οι παράμετροι του μοντέλου

υπολογίζονται ξανά με τη χρήση του αλγορίθμου EM μέχρι να ικανοποιηθεί το κριτήριο σύγκλισης του αλγορίθμου που αναφέρεται στην Παράγραφο 2.5. Βασικό πλεονέκτημα της μεθόδου αυτής, πέρα από το γεγονός ότι καταφέρνει να χρησιμοποιήσει πληροφορία από έναν μεγάλο αριθμό χρονοσειρών συνοψίζοντας αυτές σε έναν πολύ μικρότερο αριθμό χρονοσειρών, είναι η δυνατότητα να διαχειριστεί ελλιπείς τιμές λόγω της συνδυαστικής εφαρμογής του φίλτρου και της εξομάλυψης του Kalman.

Μέρος II

Αριθμητική εφαρμογή

Στο δεύτερο μέρος της διπλωματικής εργασίας θα εφαρμόσουμε τη μεθοδολογία που αναπτύξαμε στο πρώτο μέρος σε πραγματικά δεδομένα που αφορούν σε ελληνικά μακροοικονομικά στοιχεία.

4 Αριθμητική εφαρμογή σε ελληνικά δεδομένα

4.1 Εισαγωγή

Το δυναμικό μοντέλο ανάλυσης παραγόντων χρησιμοποιείται κατά κύριο λόγο από μεγάλες τράπεζες, προκειμένου να αξιολογήσουν όσο το δυνατόν καλύτερα τον τεράστιο όγκο δεδομένων που κατέχουν. Συγκεκριμένα, χρησιμοποιώντας το μοντέλο αυτό στοχεύουν τις περισσότερες φορές είτε στη δημιουργία νέων δεικτών για την οικονομία είτε στην πραγματοποίηση προβλέψεων.

Στην περίπτωση της δημιουργίας δεικτών, οι οποίοι να μπορούν να συνοψίσουν όσο το δυνατόν καλύτερα την οικονομική δραστηριότητα, χαρακτηριστικό παράδειγμα αποτελεί ο δείκτης Chicago Fed National Activity (CFNAI). Ο δείκτης αυτός πρωτοεμφανίστηκε το Μάρτιο του 2001 και είχε σχεδιαστεί αρχικά προκειμένου να παρακολουθείται ο πληθωρισμός των ΗΠΑ (βλέπε [30]).

Όσον αφορά στην πραγματοποίηση προβλέψεων έχει φανεί εμπειρικά στο πεδίο της πρόβλεψης μακροοικονομικών χρονοσειρών, χρησιμοποιώντας το δυναμικό μοντέλο ανάλυσης παραγόντων λαμβάνουμε καλύτερα αποτελέσματα από τα κλασικά υποδείγματα ARIMA. Το αποτέλεσμα αυτό θα προσπαθήσουμε να επιβεβαιώσουμε χρησιμοποιώντας δεδομένα της ελληνικής οικονομίας με στόχο να προβλέψουμε την πορεία του ΑΕΠ της Ελλάδας.

Στο κεφάλαιο αυτό θα περιγράψουμε αρχικά το σύνολο δεδομένων που είχαμε στη διάθεσή

μας. Στη συνέχεια, θα αναφερθούμε στους μετασχηματισμούς που ήταν αναγκαίοι προκειμένου να εφαρμόσουμε τον αλγόριθμο. Ακολούθως, θα γίνει μία αναφορά των παραγόντων που εντοπίστηκαν και των κύριων χαρακτηριστικών αυτών. Κλείνοντας, θα δώσουμε τα αποτελέσματα της εφαρμογής του αλγορίθμου. Στην ενότητα αυτή θα συγκρίνουμε το δυναμικό μοντέλο ανάλυσης παραγόντων, το οποίο από εδώ και στο εξής θα το συμβολίζουμε με τα αρχικά ΔΜ, με ένα κλασσικό μοντέλο ανάλυσης χρονοσειρών.

4.2 Περιγραφή του συνόλου δεδομένων

Το σύνολο δεδομένων που χρησιμοποιήθηκε στην εφαρμογή αποτελείται από 33 μεταβλητές, οι οποίες επιλέχθηκαν από ένα αρχικό σύνολο δεδομένων 80 μεταβλητών με κριτήριο τη γραμμική συσχέτιση που παρουσίαζαν με το ΑΕΠ. Τα δεδομένα αντλήθηκαν από την ελληνική στατιστική υπηρεσία, την Eurostat και την ευρωπαϊκή κεντρική τράπεζα (ECB). Τα δεδομένα καλύπτουν τη χρονική περίοδο Ιανουάριος του 2006 μέχρι Αύγουστος του 2016. Μια ιδιαιτερότητα του συνόλου δεδομένων είναι ότι δεν έχουμε διαθέσιμα δεδομένα για όλες τις χρονοσειρές από τον Ιανουάριο του 2006, γεγονός το οποίο δεν αποτελεί πρόβλημα καθώς όπως αναφέραμε εφαρμόζοντας τη μεθοδολογία των Doz, Giannone, Reichlin μπορούμε να προβλέψουμε τις τιμές που λείπουν αναδρομικά χρησιμοποιώντας το φίλτρο και την εξομάλυνση του Kalman.

Όλες οι μεταβλητές που χρησιμοποιήθηκαν ήταν απαλλαγμένες από εποχικότητα και ήταν στάσιμες. Στην περίπτωση που χρειάστηκε να αφαιρεθεί η εποχικότητα που υπήρχε στις αρχικές μεταβλητές διασπάσαμε τη χρονοσειρά στα δομικά της στοιχεία (βλέπε [1]). Οι μη στάσιμες χρονοσειρές έγιναν στάσιμες λαμβάνοντας πρώτες ή δεύτερες διαφορές όπου ήταν αναγκαίο. Περισσότερες πληροφορίες σχετικά με τους μετασχηματισμούς που έγιναν και το χειρισμό των ελλিপών τιμών δίνονται στην επόμενη παράγραφο.

4.3 Προκαταρκτική επεξεργασία των δεδομένων

Σημαντικό μέρος οποιασδήποτε στατιστικής ανάλυσης αποτελεί η προκαταρκτική επεξεργασία των δεδομένων που έχει στην κατοχή του ο ερευνητής. Στην παράγραφο αυτή θα αναφέρουμε τους μετασχηματισμούς, οι οποίοι χρειάστηκε να εφαρμοστούν στις χρονοσειρές που είχαμε στην διάθεσή μας προκειμένου να εφαρμόσουμε αποτελεσματικά τη μεθοδολογία που αναπτύξαμε στο πρώτο μέρος της εργασίας.

Αρχικά, έπρεπε να αφαιρέσουμε από το αρχικό σύνολο των 80 μεταβλητών εκείνες που ενδεχομένως να μην μπορούν να βοηθήσουν στην πρόβλεψη του ΑΕΠ. Τα κύρια προβλήματα που μπορούν να δημιουργηθούν αν κανείς συμπεριλάβει αυτές τις μεταβλητές είναι δύο. Πρώτον, αυξάνει την πολυπλοκότητα των υπολογισμών χωρίς κάποιο ουσιαστικό όφελος. Δεύτερον, επειδή η ανάλυση παραγόντων είναι μία μέθοδος μη εποπτευόμενης μάθησης είναι πολύ πιθανόν οι παράγοντες που θα προκύψουν να σχετίζονται ελάχιστα με το ΑΕΠ, γεγονός το οποίο θα οδηγήσει σε φτώχεια αποτελέσματα όσον αφορά στην πρόβλεψη αυτού⁹. Ένας γρήγορος και αποτελεσματικός τρόπος επιλογής των μεταβλητών που θα συμπεριληφθούν στο σύνολο δεδομένων είναι η χρήση του συντελεστή γραμμικής συσχέτισης του Pearson. Πιο συγκεκριμένα θέτοντας ένα ελάχιστο αποδεκτό κάτω όριο¹⁰ για τη συσχέτιση που πρέπει να έχει μία μεταβλητή με τη μεταβλητή απόκρισης καταλήξαμε στο τελικό σύνολο δεδομένων, το οποίο περιέχει 33 μεταβλητές¹¹.

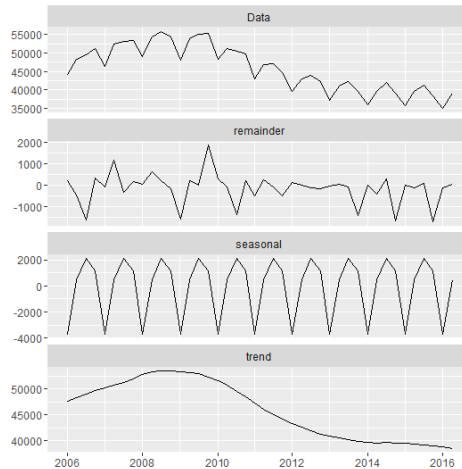
Η εφαρμογή της μεθοδολογίας που αναπτύξαμε στο πρώτο μέρος προϋποθέτει τα δεδομένα μας να είναι στάσιμα. Από τη θεωρία είναι γνωστό ότι οι περισσότερες μακροοικονομικές χρονοσειρές περιέχουν εποχικά στοιχεία. Ο εντοπισμός των χρονοσειρών που παρουσίαζαν εποχικότητα και η αφαίρεση αυτής έγινε χρησιμοποιώντας τη μεθοδολογία που περιγράφεται στο [1]. Με τη μεθοδολογία αυτή πετυχαίνουμε παράλληλα να αφαιρέσουμε τόσο την εποχικότητα όσο και την πιθανή προσδιοριστική τάση που εμφανίζουν οι υπό εξέταση χρονοσειρές. Χαρακτηριστικό παράδειγμα χρονοσειράς με αυτά τα στοιχεία αποτελεί ο γενικός δείκτης προστιθέμενης αξίας.

⁹Βλέπε το [7] για περισσότερες λεπτομέρειες.

¹⁰Στην περίπτωση μας ως κάτω όριο επιλέξαμε το 0.4.

¹¹Βλέπε παράρτημα Α4.

Η μεταβλητή αυτή παρουσιάζει εποχικότητα, η οποία είναι αναγκαίο να αφαιρεθεί προκειμένου να μπορούμε να τη χρησιμοποιήσουμε ως προβλέπουσα στο μοντέλο που θα εκτιμήσουμε αργότερα. Επιπλέον χρειάστηκε να αφαιρέσουμε και την προσδιοριστική τάση που εμφάνισε. Στο Σχήμα 4.1 δίνονται τα δομικά στοιχεία της εν λόγω χρονοσειράς.



Σχήμα 4.1: Διάσπαση της μεταβλητής ”προστιθέμενη αξία (γενικός δείκτης)”

Όσον αφορά στις μη στάσιμες χρονοσειρές, ο εντοπισμός αυτών έγινε χρησιμοποιώντας τον επαυξημένο έλεγχο των Dickey-Fuller (ADF)[12]. Η στασιμότητα επιτεύχθηκε λαμβάνοντας πρώτες ή και δεύτερες διαφορές σύμφωνα με τα αποτελέσματα του ADF¹². Συγκεκριμένα ο έλεγχος ADF μας επιτρέπει να αναγνωρίσουμε ποιές χρονοσειρές παρουσιάζουν μοναδιαία ρίζα γεγονός που επηρεάζει τη μακροχρόνια σύγκλιση αυτών και τις καθιστά μη στάσιμες.

¹²Βλέπε παράρτημα Β2.1.

4.4 Παράγοντες

Στο ΔΜ εκτιμήσαμε τρεις παράγοντες, τους οποίους μοντελοποιήσαμε χρησιμοποιώντας ένα υπόδειγμα VAR τάξης 1. Η επιλογή αυτή βασίστηκε στη βιβλιογραφία (βλέπε [17],[24],[19],[31]) που τονίζει ότι χρησιμοποιώντας δύο ή τρεις παράγοντες για τη μοντελοποίηση με τη βοήθεια ενός μοντέλου VAR(1) μπορούμε να λάβουμε ικανοποιητικές προβλέψεις για το ΑΕΠ. Οι παράγοντες αυτοί καταφέρνουν να ερμηνεύσουν το 95.42%¹³ της αρχικής μεταβλητότητας του συνόλου δεδομένων. Στην προσπάθεια να ερμηνεύσουμε τους παράγοντες αυτούς θα χρησιμοποιήσουμε τον συντελεστή γραμμικής συσχέτισης του Pearson για να εντοπίσουμε με ποιές από τις αρχικές μεταβλητές σχετίζονται περισσότερο οι τρεις παράγοντες. Πρέπει να τονιστεί ωστόσο ότι επειδή στόχος μας σε αυτή την εφαρμογή ήταν η πρόβλεψη, οι παράγοντες στους οποίους καταλήξαμε δεν έχουν κάποια φυσική ερμηνεία. Στην περίπτωση που κάποιος επιθυμεί να καταλήξει σε παράγοντες που να έχουν φυσική ερμηνεία μπορεί να χρησιμοποιήσει κάποιον ορθογώνιο μετασχηματισμό (βλέπε για παράδειγμα την Παράγραφο 1.4).

Ξεκινώντας με τον πρώτο παράγοντα η γραμμική συσχέτιση αυτού με τις αρχικές μεταβλητές που σχετίζεται περισσότερο δίνεται παρακάτω.

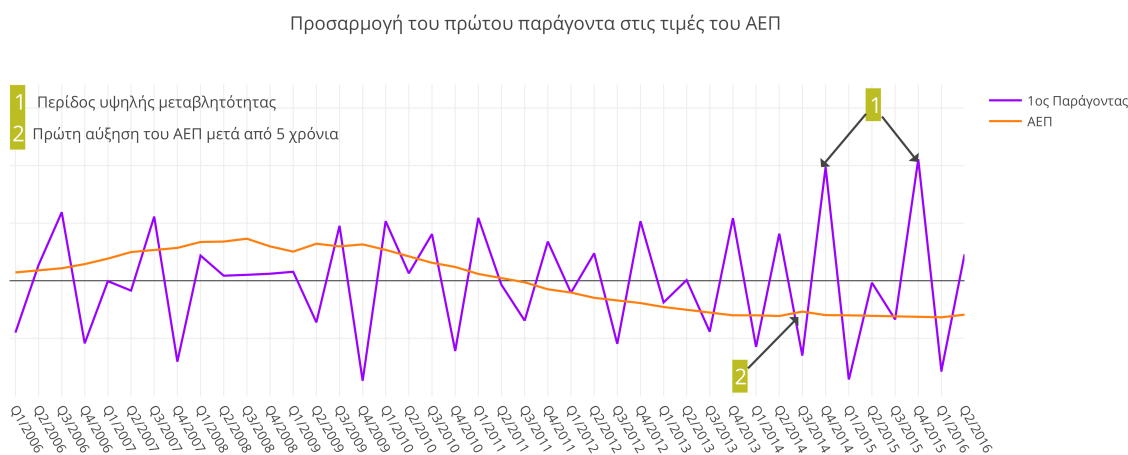
	Μεταβλητές	Παράγοντας 1
1	Industrial Production Index	-0.92
2	Value added services	-0.912
3	Value added all	-0.98
4	Value added industry	0.88
5	Central Government total expenditure	0.79
6	Unit labour wholesale and retail	0.86
7	Unit Labour manufacturing	0.89

Πίνακας 6: Μεταβλητές με τις οποίες παρουσιάζει την υψηλότερη συσχέτιση ο 1ος παράγοντας.

Η σύσταση του παράγοντα αυτού περιλαμβάνει δείκτες προστιθέμενης αξίας στις υπηρεσίες και στην βιομηχανία καθώς και τον γενικό δείκτη προστιθέμενης αξίας. Ακόμα συνδιάζει το

¹³Βλέπε παράρτημα Β2.2.

κόστος εργασίας στους κλάδους του χονδρικού και λιανικού εμπορίου, όπως επίσης και το κόστος εργασίας στον κατασκευαστικό τομέα. Τέλος, ο παράγοντας αυτός σχετίζεται θετικά με τα εξόδα της κεντρικής κυβέρνησης. Το παραπάνω αποτελέσματα είναι δύσκολο να χρησιμοποιηθεί για να δοθεί μια ερμηνεία σε αυτόν τον παράγοντα.



Σχήμα 4.2: Πρώτος παράγοντας

Παρ' όλα αυτά παρατηρώντας το γράφημα στο Σχήμα 4.2 μπορεί να δει κανείς την αυξημένη μεταβλητότητα που εμφανίζεται από το τελευταίο τρίμηνο του 2014 μέχρι σήμερα. Ο παράγοντας δηλαδή φαίνεται να εντοπίζει ένα σημαντικό σημείο της πορείας του ΑΕΠ. Πιο συγκεκριμένα την περίοδο αυτή το ΑΕΠ άρχισε να αυξάνεται για πρώτη φορά μετά από τέσσερα χρόνια. Ωστόσο η ανοδική πορεία αυτή δεν κράτησε παρά μόνο για ένα τρίμηνο. Η αυξημένη μεταβλητότητα κατά τα τελευταία τρίμηνα ενδεχομένως να σχετίζεται με την οικονομική αστάθεια που υπήρξε στην χώρα την περίοδο αυτή κατά την οποία θεωρούνταν πολύ πιθανό η χώρα να βρεθεί εκτός της ζώνης του ευρώ.

Όσον αφορά στο δεύτερο παράγοντα μπορούμε να δούμε στον Πίνακα 7 ότι ο παράγοντας αποτελεί γραμμικό συνδυασμό μεταβλητών, οι οποίες παρουσιάζουν υψηλή γραμμική συσχέτιση

με το ΑΕΠ. Το γεγονός αυτό έχει μια θετική ιδιότητα και μία αρνητική. Ξεκινώντας από τη θετική αναμένουμε ο παράγοντας αυτός να μπορεί να ακολουθήσει ικανοποιητικά την πορεία του ΑΕΠ λόγω της υψηλής συσχέτισης που εμφανίζουν οι μεταβλητές που ευθύνονται κατά κύριο λόγο για τη σύσταση του. Από την άλλη είναι αρκετά δύσκολο να τον ερμηνεύσουμε λόγω των ετερόκλητων μεταβλητών που τον συνθέτουν.

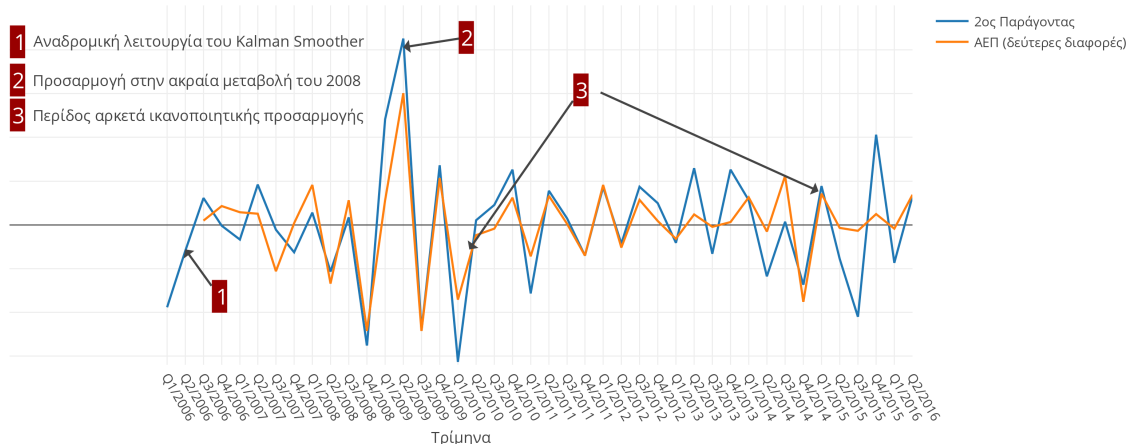
	Μεταβλητές	Παράγοντας 2
1	Manufacturing production index	0.65
2	VIX	-0.88
3	Gross fixed capital formation	0.63
4	Imports of goods and services	0.54
5	Value added construction	0.63
6	Value added all	0.65
7	Total Employment	-0.84
8	Unit labour agriculture	0.80
9	Unit labour industry	-0.80
10	Unit labour information and communication	-0.81
11	Unit labour manufacturing	-0.91

Πίνακας 7: Μεταβλητές με τις οποίες παρουσιάζει την υψηλότερη συσχέτιση ο 2ος παράγοντας.

Στο Σχήμα 4.3 βλέπουμε τη μεταβλητότητα του παράγοντα να αυξάνεται μετά το δεύτερο τρίμηνο του 2008. Ο παράγοντας αυτός καταφέρνει να εντοπίσει την αρχή της οικονομικής κρίσης στην Ελλάδα το 2008. Το σημείο στο οποίο εμφανίζεται το καρφί (spike) στο σχήμα είναι το σημείο εκείνο που το ελληνικό ΑΕΠ άρχισε να μειώνεται (βλ. Σχήμα 4.3). Η περίοδος αυτή ακολουθείται από ένα διάστημα χαμηλής μεταβλητότητας (Q2/2010-Q4/2012), στο οποίο ο παράγοντας φαίνεται γραφικά να έχει αρκετά καλή προσαρμογή.

Το διάστημα αυτό σχετίζεται άμεσα με τη σταθερή πτωτική πορεία του ΑΕΠ. Τέλος, στο σχήμα αυτό μπορούμε να δούμε την αναδρομική λειτουργία του Kalman Smoother μέσω του οποίου λαμβάνουμε εκτιμήσεις των δύο πρώτων τιμών της χρονοσειράς του ΑΕΠ σε δεύτερες

Προσαρμογή του 2ου παράγοντα στις τιμές του ΑΕΠ σε δεύτερες διαφορές



Σχήμα 4.3: Δεύτερος παράγοντας

διαφορές. Οι τιμές αυτές δεν είναι αρχικά διαθέσιμες επειδή η χρονοσειρά που προβλέψαμε προήλθε μετά την λήψη δεύτερων διαφορών στις πραγματικές τιμές του ΑΕΠ.

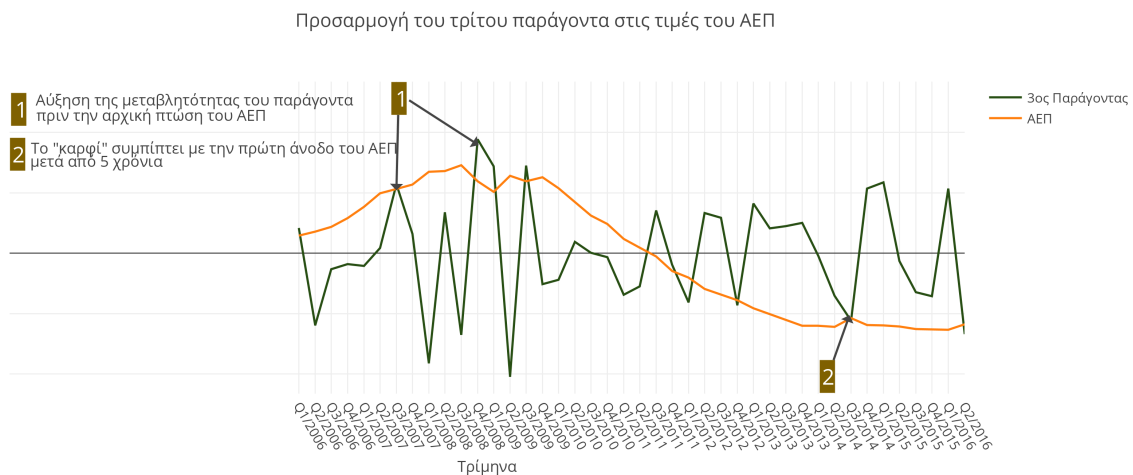
Ο τρίτος παράγοντας παρουσιάζει ιδιαίτερο ενδιαφέρον, καθώς πέρα από κάποιους μακροοικονομικούς δείκτες που αποτελούν κύρια δομικά συστατικά των άλλων δύο παραγόντων είναι ο μόνος από τους τρεις παράγοντες που περιλαμβάνει μεταβλητές σχετικές με την απασχόληση. Οι μεταβλητές με τις οποίες παρουσιάζει τη μεγαλύτερη γραμμική συσχέτιση δίνονται στον Πίνακα 8.

Ωστόσο η φυσική ερμηνεία του παράγοντα αυτού είναι δύσκολη λόγω της ανομοιογένειας των μεταβλητών που είναι υψηλά συσχετισμένες με αυτόν.

Επίσης ένα άλλο στοιχείο που κάνει τον συγκεκριμένο παράγοντα αρκετά ενδιαφέροντα φαίνεται στο Σχήμα 4.4. Η μεταβλητότητα του παράγοντα αυτού είχε ήδη αυξηθεί πριν την αρχική πτώση του ΑΕΠ το 2008. Ακόμα κατά το τρίτο τρίμηνο του 2014 ο παράγοντας φαίνεται να εντοπίζει την αύξηση του ΑΕΠ αρκετά ικανοποιητικά.

	Μεταβλητές	Παράγοντας 3
1	Unemployment rate	0.66
2	Number of employed	-0.75
3	Deposits and repos of non MFIs	-0.62
4	Central government revenue	0.61
5	Value added construction	0.65
6	Value added services	-0.62
7	Unit labour information and communication	-0.70

Πίνακας 8: Μεταβλητές με τις οποίες παρουσιάζει την υψηλότερη συσχέτιση ο 3ος παράγοντας.



Σχήμα 4.4: Τρίτος παράγοντας

Συνοψίζοντας αυτά που είδαμε παραπάνω οι τρεις παράγοντες φαίνεται να εντοπίζουν όλα τα σημαντικά σημεία της πορείας του ΑΕΠ. Επειδή οι παράγοντες αυτοί έχουν προκύψει από το κλασικό μοντέλο ανάλυσης παραγόντων (προτού γίνουν δυναμικοί) φαίνεται και γραφικά ότι είναι ανεξάρτητοι μεταξύ τους αφού ο καθένας καταφέρνει να προσαρμοστεί καλύτερα σε διαφορετικές περιόδους του ΑΕΠ. Ιδανικά θα θέλαμε να δώσουμε και κάποια ερμηνεία σε αυτούς, η οποία δεν είναι εφικτή στην παρούσα εφαρμογή γιατί όπως αναφέραμε και στην αρχή στόχος μας είναι η προβλέψη¹⁴. Στην παρακάτω παράγραφο δίνονται τα αποτελέσματα της εφαρμογής

¹⁴Στην γενική περίπτωση, το ότι κάποιος στοχεύει σε πρόβλεψη δεν αναίρει την δυνατότητα δημιουργίας παραγόντων με κάποια φυσική ερμηνεία.

της μεθόδου, τα οποία φαίνονται να επιβεβαιώνουν τις ενδείξεις που είχαμε γραφικά, ότι δηλαδή οι τρεις παράγοντες αποτελούν καλές εκτιμήτριες του ΑΕΠ.

4.5 Αποτελέσματα

Χρησιμοποιήσαμε το μοντέλο δυναμικής παραγοντικής ανάλυσης με στόχο να πραγματοποιήσουμε πρόβλεψη για την επόμενη περίοδο, δηλαδή για το τρίτο τρίμηνο του 2016. Έχοντας ως στόχο αυτό αξιολογήσαμε το μοντέλο μας ως προς την ικανότητά του να δημιουργεί προβλέψεις μια περίοδο μπροστά. Για τον σκοπό αυτό χρησιμοποιήσαμε ως ανταγωνιστή ένα υπόδειγμα ARIMA, το οποίο ταυτοποιήθηκε χρησιμοποιώντας την μεθοδολογία Box-Jenkins[8]. Συγκεκριμένα για την ταυτοποίηση του μοντέλου ARIMA χρησιμοποιήσαμε τα κορρολογράμματα(ACF & PACF) και ως μέθοδος εκτίμησης χρησιμοποιήθηκε η μέθοδος των ελαχίστων τετραγώνων. Οι διαγνωστικοί έλεγχοι του μοντέλου δίνονται αναλυτικά στο παράρτημα¹⁵. Η σύγκριση των δύο μοντέλων έγινε χρησιμοποιώντας το μέσο τετραγωνικό σφάλμα πρόβλεψης (MSFE), το μέσο απόλυτο σφάλμα (MAE) και το μέσο απόλυτο ποσοστιαίο σφάλμα (MAPE) για μία περίοδο μπροστά.

Ο Πίνακας 9 δίνει το αποτέλεσμα της διαίρεσης της τιμής που λάβαμε σε καθένα από τα κριτήρια για το ΔΜ με αυτή του ARIMA, δηλαδή τους λόγους $\frac{MSFE(\Delta M)}{MSFE(ARIMA)}$, $\frac{MAE(\Delta M)}{MAE(ARIMA)}$, $\frac{MAPE(\Delta M)}{MAPE(ARIMA)}$.

Σε κάθε περίπτωση παρατηρούμε ότι το ΔΜ δημιουργεί πιο ακριβείς προβλέψεις από το ARIMA,

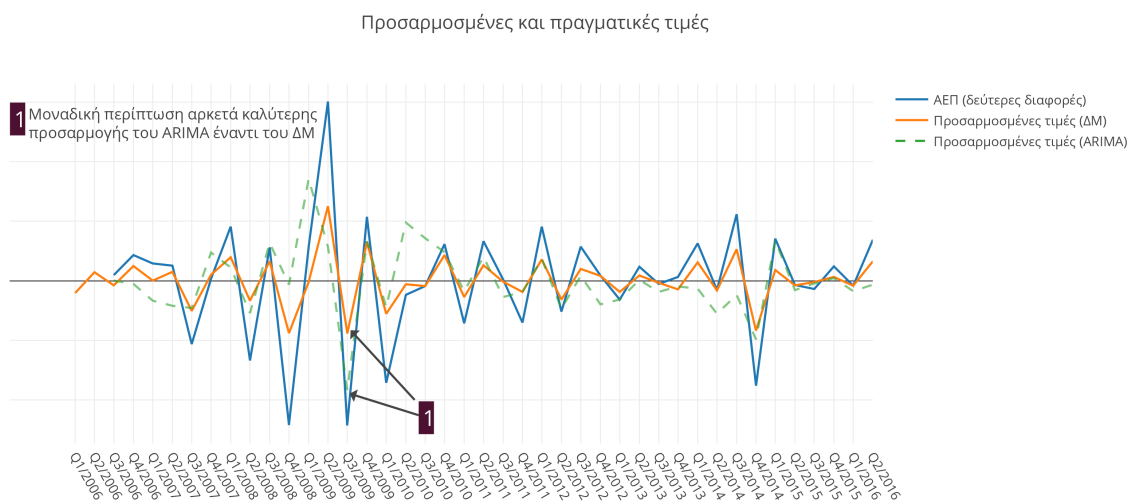
Υπόδειγμα	MSFE	MAE	MAPE
ΔΜ	0.76	0.86	0.36

Πίνακας 9: Σύγκριση των δύο ανταγωνιστών.

¹⁵Παράρτημα Β3.

γεγονός το οποίο επιβεβαιώνει την χρησιμότητα του μοντέλου αυτού στην πραγματοποίηση μακροοικονομικών προβλέψεων.

Στο Σχήμα 4.5 φαίνεται η προσαρμογή των δύο μοντέλων στις τιμές που λαμβάνει το ΑΕΠ σε δεύτερες διαφορές.

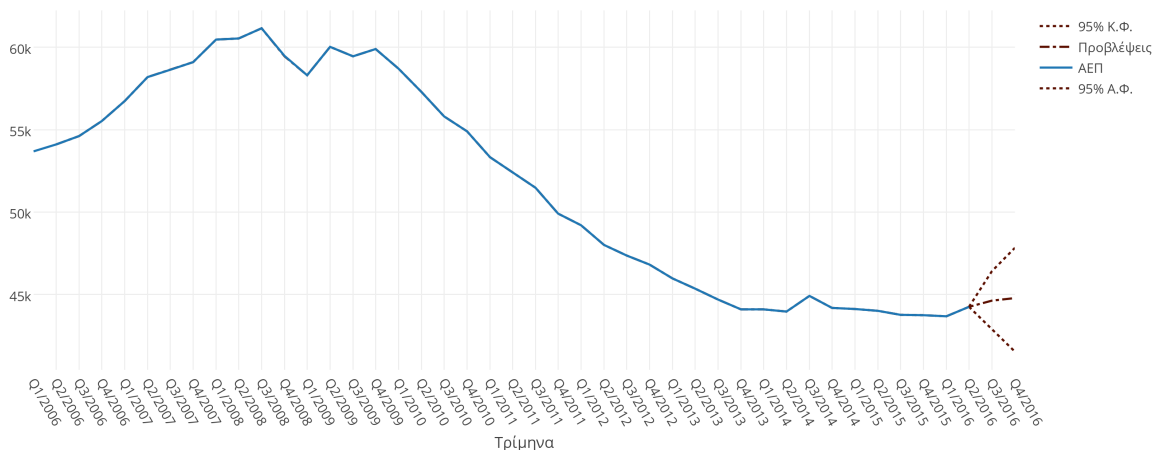


Σχήμα 4.5: Προσαρμοσμένες τιμές

Παρατηρώντας το Σχήμα 4.5 φαίνεται οι τιμές που προήλθαν από το ΔΜ να προσαρμόζονται καλύτερα στις τιμές του ΑΕΠ έναντι αυτών που προήλθαν από το ARIMA. Συγκεκριμένα μόνο σε ένα σημείο το ARIMA έχει αρκετά καλύτερη προσαρμογή από το ΔΜ. Από το σχήμα αυτό πλέον έχουμε και οπτική εικόνα της υπεροχής του ΔΜ, την οποία είδαμε αρχικά στον Πίνακα 9.

Στην περίπτωση που κάποιος θελήσει να λάβει τις προσαρμοσμένες τιμές του ΔΜ στις κανονικές τιμές του ΑΕΠ αρκεί να εφαρμόσει τον μετασχηματισμό που δίνεται στο παράρτημα¹⁶. Εφαρμόζοντας τον ίδιο μετασχηματισμό στις προβλέψεις που λάβαμε από το ΔΜ για τις τιμές του ΑΕΠ σε δεύτερες διαφορές δημιουργήσαμε προβλέψεις για τις αρχικές τιμές του ΑΕΠ. Τις τιμές αυτές τις χρησιμοποιήσαμε για να φτιάξουμε το Σχήμα 4.6.

¹⁶Παράρτημα Β2.6.



Σχήμα 4.6: Πρόβλεψη ΑΕΠ έως το 4ο τρίμηνο του 2016.

Στο Σχήμα 4.6 δίνονται ενδεικτικές προβλέψεις μέχρι το τέταρτο τρίμηνο του 2016 μαζί με το αντίστοιχο 95% διάστημα εμπιστοσύνης. Αναλυτικά ο τρόπος με τον οποίον δημιουργήθηκαν οι προβλέψεις, καθώς και 95% φράγματα εμπιστοσύνης γι' αυτές δίνονται στο παράρτημα¹⁷. Σύμφωνα με το μοντέλο που χρησιμοποιήσαμε το ΑΕΠ αναμένεται να αυξηθεί και να φτάσει την τιμή που είχε λάβει στο τρίτο τρίμηνο του 2014.

5 Συμπεράσματα και συζήτηση

Στόχος της παρούσας εργασίας είναι κατά κύριο λόγο να δώσει μια περιγραφή της θεωρητικής θεμελίωσης του δυναμικού μοντέλου ανάλυσης παραγόντων και κατά δεύτερον να το εφαρμόσει προκειμένου να προβλεφθεί η πορεία του ελληνικού ΑΕΠ. Το μοντέλο των Doz, Giannone, Reichlin παρουσιάζει ιδιαίτερο ενδιαφέρον από θεωρητική άποψη, διότι πετυχαίνει να συνδυάσει δύο πολύ βασικά πεδία της στατιστικής, αυτό της πολυμεταβλητής ανάλυσης και εκείνο της ανάλυσης χρονοσειρών. Πρακτικά, το ΔΜ παρουσίασε καλύτερη προσαρμογή στις τιμές του ΑΕΠ έναντι του κλασσικού υποδείγματος ARIMA, γεγονός το οποίο δείχνει την χρησιμότητα

¹⁷Βλέπε Παράρτημα Β2.5.

των μοντέλων αυτών για τη δημιουργία μακροοικονομικών προβλέψεων. Σημαντικό είναι επίσης το γεγονός ότι ένας μικρός αριθμός κοινών παραγόντων μπορεί να συνοψίσει την κίνηση ενός μεγάλου συνόλου από μακροοικονομικές και χρηματοοικονομικές σειρές.

Η εφαρμογή αυτή στόχο είχε να αναδείξει μία μόνο από τις διαφορετικές χρήσεις της συγκεκριμένης κλάσης μοντέλων. Συγκεκριμένα, τα μοντέλα αυτά μπορούν να χρησιμοποιηθούν επίσης για τη δημιουργία δεικτών, οι οποίοι να είναι σε θέση να συνοψίζουν την πορεία της οικονομίας. Ακόμα, κεντρικές τράπεζες χρησιμοποιούν τη συγκεκριμένη μεθοδολογία προκειμένου να παρακολουθούν την πορεία του ΑΕΠ μηνιαία (Nowcasting), πεδίο στο οποίο έχει πραγματοποιηθεί ένας μεγάλος αριθμός δημοσιεύσεων(ενδεικτικά [14], [17], [19], [24], [31]). Οι δημοσιεύσεις αυτές κατά κύριο λόγο παρουσιάζουν μοντέλα με λιγότερες περιοριστικές υποθέσεις έναντι αυτού που αναπτύξαμε στην παρούσα διπλωματική εργασία αυτή¹⁸.

Κλείνοντας, θα είχε ιδιαίτερο ενδιαφέρον να ερευνηθεί περαιτέρω η δυνατότητα πέρα από την εξίσωση κατάστασης να αποκτήσει δυναμική μορφή και η εξίσωση παρατήρησης περιλαμβάνοντας χρονικές υστερήσεις των παραγόντων.

¹⁸Βλέπε [14]

Παράρτημα

Α΄ Βασικές έννοιες

Α1. Το Φασματικό Θεώρημα

Έστω ένας κανονικός συμμετρικός τελεστής T σε έναν μιγαδικό (πραγματικό) χώρο εσωτερικού γινομένου και πεπερασμένης διάστασης V . Τότε, υπάρχουν γραμμικοί τελεστές E_1, \dots, E_r και αριθμοί $\lambda_1, \dots, \lambda_r$, τέτοιοι ώστε:

$$1. T = \lambda_1 E_1 + \dots + \lambda_r E_r$$

$$2. E_i^2 = E_i, \dots, E_r^2 = E_r$$

$$3. E_1 + E_2 + \dots + E_r = I$$

$$4. E_i E_j = 0 \text{ για } i \neq j$$

Περισσότερες λεπτομέρειες δίνονται στο [27].

Α2. Λήμμα 1

Θεωρώντας ότι τα διανύσματα X, Y ακολουθούν από κοινού διδιάστατη κανονική κατανομή με

$$E \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \quad Var \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix},$$

η από κοινού τους κατανομή μπορεί να γραφεί και ως

$$p(X, Y) = p(Y)p(X|Y)$$

από όπου μπορούμε να υπολογίσουμε τα $E(X|Y)$ και $Var(X|Y)$ χρησιμοποιώντας τους τύπους:

1. $E(X|Y) = \mu_X + \frac{\sigma_{XY}}{\sigma_Y^2}(Y - \mu_Y),$
2. $Var(X|Y) = \sigma_X^2 - \frac{\sigma_{XY}^2}{\sigma_Y^2}.$

A3. Το Θεώρημα Αναπαράστασης του Wold

Κάθε στάσιμη χρονοσειρά X_t , αυτή μπορεί να αναπαρασταθεί ως το άθροισμα δύο ανεξάρτητων μεταξύ τους χρονοσειρών Z_t, Y_t ως $X_t = Z_t + Y_t$, όπου Z_t είναι μία προσδιοριστική χρονοσειρά με την έννοια ότι μπορεί να προβλεφθεί τέλεια από τις προηγούμενες τιμές της και y_t είναι μία χρονοσειρά, η οποία ακολουθεί ένα στοχαστικό υπόδειγμα κινητού μέσου, δηλαδή

$$Y_t = \sum_{i=0}^{+\infty} \Phi_i u_{t-i}$$

όπου $\Phi_0 = I_K$ και την u_t να αποτελεί διαδικασία λευκού θορύβου.

A4. Υπόδειγμα Vector Autoregressive (VAR)

Στην γενική του μορφή το VAR(p) δίνεται από την σχέση,

$$Y_t = v + \mathbf{A}_1 Y_{t-1} + \dots + \mathbf{A}_p Y_{t-p} + u_t, \quad t = 0, 1, 2, \dots$$

όπου $Y_t = (Y_{1t}, \dots, Y_{Kt})'$ είναι ένα $(K \times 1)$ διάνυσμα, τα \mathbf{A}_i είναι $(K \times K)$ πίνακες των συντελεστών του μοντέλου και το $v = (v_1, \dots, v_K)'$ είναι ένα $(K \times 1)$ διάνυσμα σταθερών. Τέλος, το $u_t = (u_{1t}, \dots, u_{Kt})$ αποτελεί διαδικασία λευκού θορύβου. Περισσότερες λεπτομέρειες δίνονται στα [20],[28].

Ουσιαστικά στο υπόδειγμα VAR κάθε μεταβλητή περιγράφεται ως συνάρτηση των προηγούμενων τιμών της αλλά και των τιμών των άλλων μεταβλητών του συστήματος.

Α4. Δεδομένα

	Μεταβλητές	Πηγή
1	GDP	ΕΛΣΤΑΤ
2	Unemployment Rate	ΕΛΣΤΑΤ
3	Number of employed	ΕΛΣΤΑΤ
4	Number of unemployed	ΕΛΣΤΑΤ
5	Manufacturing production index	ΕΛΣΤΑΤ
6	Retail Sales Index (Volume)	ΕΛΣΤΑΤ
7	Current account balance	EUROSTAT
8	Industrial Production Index	ΕΛΣΤΑΤ
9	Turnover index in retail	ΕΛΣΤΑΤ
10	Deposits and repos of non MFIs in MFIs in Greece	EUROSTAT
11	Central government revenue	ΕΛΣΤΑΤ
12	Central government expenditure	ΕΛΣΤΑΤ
13	VIX	ECB
14	Domestic private sector bank deposits	ECB
15	Changes in inventories and acquisition less disposals of valuables	ECB
16	Gross fixed capital formation	ΕΛΣΤΑΤ
17	Government final consumption	ΕΛΣΤΑΤ
18	Private final consumption	ΕΛΣΤΑΤ
19	Imports of goods and services	ΕΛΣΤΑΤ
20	Value added (fishery forestry)	ΕΛΣΤΑΤ
21	Value added (industry constr)	ΕΛΣΤΑΤ
22	Value added (services)	ΕΛΣΤΑΤ
23	Value added (all)	ΕΛΣΤΑΤ
24	Unit labour (agriculture)	ΕΛΣΤΑΤ
25	Unit labour (industry)	ΕΛΣΤΑΤ
26	Unit labour (finance and insurance)	ΕΛΣΤΑΤ
27	Unit labour(information and communication)	ΕΛΣΤΑΤ
28	Unit labour (wholesale and retail)	ΕΛΣΤΑΤ
29	Unit labour (construction)	ΕΛΣΤΑΤ
30	Unit labour(manufacturing)	ΕΛΣΤΑΤ
31	Eonia Rate	ECB
32	3-month interest rates	ECB
33	6-month interest rates	ECB

Πίνακας 10: Μεταβλητές που απαρτίζουν το σύνολο δεδομένων.

B' Κώδικες και αποτελέσματα

B1. Mathematica

Το Mathematica χρησιμοποιήθηκε για τη δημιουργία γραφημάτων που εμφανίζονται στη διπλωματική εργασία. Για τα υπόλοιπα σχήματα που δεν υπάρχει κώδικας στο παράρτημα, χρησιμοποιήθηκε το λογισμικό plotly(<https://plot.ly/ãvarelas>).

1. Σχήμα 1.1

```
x1 = (x*Sqrt[2]/2 + y*Sqrt[2]/2); y1 = (-x*Sqrt[2]/2 + y*Sqrt[2]/2);  
f[x_, y_,  
z_] = ((x*Sqrt[2]/2 + y*Sqrt[2]/2)/  
5)^2 + ((-x*Sqrt[2]/2 + y*Sqrt[2]/2)/3)^2 + (z/2)^2; p1 =  
ContourPlot3D[f[x, y, z] == 1, {x, -5, 5}, {y, -5, 5}, {z, -5, 5},  
Contours -> 1];  
p2 = ContourPlot3D[  
f[x, y, 1] == 1, {x, -5, 5}, {y, -5, 5}, {z, -5, -4.99},  
Contours -> 5];  
  
Show[p1, p2, Axes -> False]
```

2. Σχήμα 1.2

```
\[CapitalSigma] = {{1, 0.9}, {0.9, 1}};  
\[CapitalSigma]2 = MatrixPower[\[CapitalSigma], 1/2];  
n = 1000; M = Table[0, {n}];  
Do[  
R2 = -2*Log[RandomReal[]];  
\[CapitalTheta] = RandomReal[]*2*Pi;  
Z = {R2^0.5*Cos[\[CapitalTheta]], R2^0.5*Sin[\[CapitalTheta]]};  
X = {0, 0} + Z.\[CapitalSigma]2;  
M[[i]] = X;  
, {i, 1, n}]  
p1 = ListPlot[M, PlotRange -> {{-2, 2}, {-2, 2}}, AspectRatio -> 1,  
Ticks -> False, AxesStyle -> Dashed]  
f[x_] = x;  
g[y_] = -y; p2 =
```

```

Plot[f[x], {x, -3*0.7071067811865475, 3*0.7071067811865475},
  Axes -> False, PlotStyle -> Black, GridLines -> {{0}, None},
  GridLinesStyle -> Directive[Black, Dashed]]
p3 = Plot[g[y], {y, -2*0.7071067811865475, 2*0.7071067811865475},
  Axes -> False, PlotStyle -> Black];
Show[p1, p2, p3]
Plot[{0}, {x, 1, 3}, GridLines -> {{2}, None}]

```

3. Σχήμα 1.3

```

Graphics[(*The dashed circle segment*)
  Circle[{0, 0},
    1],(*The solid circle segment*),(*The radial lines and the small \
circle segment*){Black, Line[{{0, 0}, {1, 0}}]}, {Black,
  Line[{{0, 0}, {0, 1}}]}, {Orange,
  Line[{{0, 0}, {Sqrt[2]/2, Sqrt[2]/2}}]}, {Orange,
  Line[{{0, 0}, {-Sqrt[2]/2, Sqrt[2]/2}}]},
  Circle[{0, 0}, .2, {0, 0.8}],
  Circle[{0, 0}, .2, {1.6, 2.4}], {Text[
  Style["\[Theta]", 24], .3 {Cos[1.9], Sin[2.3]}]}, {Text[
  Style["e1", 14], {Cos[0.5], Sin[.1]}]}, {Text[
  Style["e2", 14], {-0.1, 0.9}], {Text[
  Style["e2'",
  14], {(-Sqrt[2] + 0.3)/2, (+Sqrt[2] + 0.05)/2}], {Text[
  Style["e1'",
  14], {(Sqrt[2] - 0.3)/2, (Sqrt[2] + 0.05)/2}], {Text[
  Style["O", 14], {-0.05, -0.05}],(*Various text labels*){Text[
  Style["\[Theta]", 24], .3 {Cos[.5], Sin[.5]}]}]}]

```

B2. Λογισμικό R

1. Προκαταρκτική επεξεργασία δεδομένων

```
library(forecast)
seas = 0
for (i in 1:33) {
  fit <- tbats(na.omit(data[,i]))
  seasonal <- !is.null(fit$seasonal)
  if (seasonal==T) {seas = cbind(seas,i)}
}
seas = seas[,-1]
a = 0
for (i in 1:ncol(data)) {
  if (i %in% seas) {
    a = decompose(data[,i])
    data[,i] = a$random
  }
  else {data[,i]=data[,i]}
}
#####
### unit root test ###
#####
a = 1
for (i in 1:ncol(data)) {
  adf = adf.test(na.omit(data[,i]))
  if (adf$p.value>=0.05) {
    print(i)
    a = append(a,i)
  }
}
a = a[-c(1,5,10,11,15)]
#####
## stationarity ###
#####
n.data = matrix(0, nrow = nrow(data), ncol=ncol(data))
for (i in 1:ncol(data)) {
  if (i %in% a) {
    n.data[,i] = c(NA,diff(data[,i]))
  }
}
```

```

else {
  n.data[,i] = data[,i]
}
}
#####
### unit root test ###
#####
a = 1
for (i in 1:ncol(n.data)) {
  adf = adf.test(na.omit(n.data[,i]))
  if (adf$p.value >= 0.05) {
    print(i)
    a = append(a,i)
  }
}
a = a[-c(5,16,25)]

#####
## stationarity ###
#####
#n.data = matrix(0, nrow = nrow(data), ncol=ncol(data))
for (i in 1:ncol(data)) {
  if (i %in% a) {
    n.data[,i] = c(NA,diff(n.data[,i]))
  }
  else {
    n.data[,i] = n.data[,i]
  }
}
x = apply(n.data, 2, function(z) { (z - mean(z, na.rm=TRUE))/sd(z, na.rm=TRUE) }) ##standardize

```

2. Εκτίμηση του μοντέλου των Doz, Giannone, Reichlin.

```

my_dfm = function (X, r, p, q, max_iter = 100, threshold = 1e-04, rQ
)
{
  if (missing(rQ)) {
    rQ <- ""
  }
  #if (missing(rC)) {

```



```

rC <- ""
#}
if (missing(q)) {
  q <- r
}
if (q > r) {
  stop("r must be larger than q.")
}
T <- dim(X)[1]
N <- dim(X)[2]
x <- apply(X, 2, function(z) {
  (z - mean(z, na.rm = TRUE))/sd(z, na.rm = TRUE)
})
Mx <- apply(X, 2, mean, na.rm = TRUE)
Wx <- apply(X, 2, sd, na.rm = TRUE)
W <- !is.na(x)
Tt <- rbind(matrix(0, nrow = r, ncol = r * p),
  diag(1, nrow = r * (p - 1), ncol = r * p))
Ht <- matrix(0, nrow = p * r, ncol = p * r)
Ht[1:r, 1:r] <- diag(1, r)
eigen.decomp <- eigen(cov(x, use = "complete.obs"))
vec <- eigen.decomp$vectors[, 1:r]
val <- eigen.decomp$values[1:r]
cum = sum(val[1:r]/ncol(na.omit(x)))
chi <- x %*% vec %*% t(vec)
val <- diag(1, r)
F <- x %*% vec
F_pc <- F
F <- na.omit(F)
if (p > 0) {
  if (rQ == "identity") {
    fit <- VAR(F, p)
    Tt[1:r, 1:(r * p)] <- t(fit$Tt)
    Ht[1:r, 1:r] <- diag(1, r)
  }
  else {
    fit <- VAR(F, p)
    Tt[1:r, 1:(r * p)] <- t(fit$Tt)
    Qt <- cov(fit$res)
  }
}

```

```

    }
  }
  Ht <- diag(diag(cov(x - chi, use = "complete.obs")))
  Z <- fit$X
  initx <- Z[1, ]
  initV <- matrix(ginv(kron(Tt, Tt)) %*% as.numeric(Qt),
    ncol = r * p, nrow = r * p)
  Zt <- cbind(vec, matrix(0, nrow = N, ncol = r * (p - 1)))
  previous_loglik <- -.Machine$double.xmax
  loglik <- 0
  num_iter <- 0
  LL <- c()
  converged <- 0
  kf_res <- K_filter(initx, initV, x, Tt, Zt, Ht, Qt)
  ks_res <- K_smoother(Tt, kf_res$xitt, kf_res$xitm, kf_res$Ptt,
    kf_res$Pttm, Zt, Ht, W)
  xsmooth <- ks_res$xifT
  Vsmooth <- ks_res$PtT
  Wsmooth <- ks_res$PtTm
  F_kal <- t(xsmooth[1:r, , drop = FALSE])
  if (rC == "upper" & (r > 1)) {
    dimC <- dim(C[, 1:r])
    rK <- rep(0, (r - 1) * r/2)
    irC <- which(matrix(upper.tri(C[, 1:r]) + 0) == 1)
    rH <- matrix(0, nrow = length(rK), ncol = prod(dim(C)))
    for (i in 1:length(rK)) {
      rH[i, irC[i]] <- 1
    }
  }
  while ((num_iter < max_iter) & !converged) {
    em_res <- Estep(t(x), Tt, Zt, Qt, Ht, initx, initV, W)
    beta <- em_res$beta_t
    gamma <- em_res$gamma_t
    delta <- em_res$delta_t
    gamma1 <- em_res$gamma1_t
    gamma2 <- em_res$gamma2_t
    Plsum <- em_res$V1 + em_res$x1 %*% t(em_res$x1)
    x1sum <- em_res$x1
    loglik <- em_res$loglik_t
  }

```

```

num_iter <- num_iter + 1
if (rC == "upper" & (r > 1)) {
  fp <- matrix(delta[, 1:r] %*% ginv(gamma[1:r, 1:r]))
  kronCR <- kronecker(ginv(gamma[1:r, 1:r]), R)
  sp <- kronCR %*% t(rH) %*% ginv(rH %*% kronCR %*%
                                t(rH)) %*% (rK - rH %*% fp)
  C[, 1:r] <- matrix(fp + sp, nrow = dimC[1], ncol = dimC[2])
}
else {
  Zt[, 1:r] <- delta[, 1:r] %*% ginv(gamma[1:r, 1:r])
}
if (p > 0) {
  Tt_update <- beta[1:r, 1:(r * p), drop = FALSE] %*%
  solve(gamma[1:(r * p), 1:(r * p)])
  Tt[1:r, 1:(r * p)] <- Tt_update
  if (rQ != "identity") {
    H <- (gamma2[1:r, 1:r] - Tt_update %*%
          t(beta[1:r, 1:(r * p), drop = FALSE]))/(T - 1)
    if (r > q) {
      h.decomp <- svd(H)
      P <- h.decomp$v[, 1:q, drop = FALSE]
      M <- h.decomp$d[1:q]
      if (q == 1) {
        P <- P * P[1, ]
        Qt[1:r, 1:r] <- P %*% t(P) * M
      }
      else {
        P <- P %*% diag(sign(P[1, ]))
        Qt[1:r, 1:r] <- P %*% diag(M) %*% t(P)
      }
    }
    else {
      Qt[1:r, 1:r] <- H
    }
  }
}
xx <- as.matrix(na.omit(x))
Ht <- (t(xx) %*% xx - Zt %*% t(delta))/T
RR <- diag(Ht)

```

```

RR[RR < 1e-07] <- 1e-07
Ht <- diag(RR)
Ht <- diag(diag(Ht))
LL <- c(LL, loglik)
initx <- x1sum
initV <- P1sum - initx %*% t(initx)
converged <- em_converged(loglik, previous_loglik,
                          threshold = threshold)
previous_loglik <- loglik
if (num_iter < 25) {
  converged <- FALSE
}
}
if (converged == TRUE) {
  cat("Converged_after", num_iter, "iterations.\n")
}
else {
  cat("Maximum_number_of_iterations_reached.\n")
}
final_object <- list(pca = F_pc, twostep = F_kal,
                   Tt = Tt[1:r, ], Zt = Zt[, 1:r], Qt = Qt[1:q, 1:q], Ht = Ht,
                   p = p, data = x,
                   "Cumulative_Proportion_of_Variance_Explained" = cum)
class(final_object) <- c("dfm", "list")
return(final_object)
}

```

3. Το φίλτρο και η εξομάλυνση του Kalman

```

K_f = function(initx, initV, x, Tt, Zt, Ht, Qt) {

T <- dim(x)[1]
N <- dim(x)[2]
r <- dim(Tt)[1]
W <- !is.na(x)
y <- t(x)

xt.fit <- matrix(0, r, (T+1))
xt <- matrix(0, r, T)

```

```

Pt.fit <- array(0, c(r, r, (T+1)))
Pt <- array(0, c(r, r, T))

xt.fit[,1] <- initx ##### initial value of x
Pt.fit[,1] <- initV ##### initial value of y

logl <- c()
Zit <- Zt
Hit <- Ht

for (j in 1:T) {
  # missing_data <- MissData(y[,j], C, R)
  # C <- missing_data$C
  # R <- missing_data$R
  W <- !is.na(x)
  Zt <- Zit[W[j,],, drop=FALSE]
  Ht <- Hit[W[j,], W[j,], drop=FALSE]
  if (FALSE) #(all(!W[j,])) #(all(is.na(missing_data$y) == TRUE))
  {
    xt[,j] <- Tt %*% xt.fit[,j]
    Pt[,j] <- Zt %*% Pt.fit[,j] %*% t(C) + R
  } else
  {
    # Innovation covariance (inverse)
    Icov <- Zt %*% Pt.fit[,j] %*% t(Zt) + Ht
    L <- solve(Icov)
    # Innovation residual
    Ires <- as.numeric(na.omit(y[,j])) - Zt %*% xt.fit[,j]
    # Optimal Kalman gain
    KFG <- Pt.fit[,j] %*% t(Zt) %*% L
    # Updated state estimate: predicted + (Kalman gain)*fitted
    xt[j] <- xt.fit[,j] + KFG %*% Ires
    # Updated covariance estimate
    Pt[,j] <- Pt.fit[,j] - KFG %*% Zt %*% Pt.fit[,j]
    # State space variable and covariance predictions E[ff_t | t-1]
    xt.fit[,j+1] <- Tt %*% xt[j]
    Pt.fit[,j+1] <- Tt %*% Pt[,j] %*% t(Tt) + Qt
  }
}

```

```

# Compute log-likelihood with Mahalanobis distance
d <- length(Ires)
S <- Zt %*% Pt.fit[,j] %*% t(Zt) + Ht
Sinv <- solve(S)
if (nrow(Ht) == 1)
{
  GG <- t(Zt) %*% solve(Ht) %*% Zt
  detS <- prod(Ht) %*% det(diag(1, r) + Pt.fit[,j] %*% GG)
} else {
  GG <- t(Zt) %*% diag(1/diag(Ht)) %*% Zt
  detS <- prod(diag(Ht)) * det(diag(1, r) + Pt.fit[,j] %*% GG)
}
denom <- (2 * pi)^(d/2) * sqrt(abs(detS))
mahal <- sum(t(Ires) %*% Sinv %*% Ires)
logl[j] <- -0.5 * mahal - log(denom)
}
}
loglik <- sum(logl, na.rm=TRUE)
return(list(xt = xt, xt.fit = xt.fit, Pt = Pt, Pt.fit = Pt.fit, loglik = loglik))
}

#####
#####
#####

K_s <- function(Tt, xt, xt.fit, Pt, Pt.fit, Zt, Ht, W) {
  T <- dim(xt)[2]
  r <- dim(Tt)[1]

  Pt.fit <- Pt.fit[, (1:(dim(Pt.fit)[3] - 1)), drop = FALSE]
  xt.fit <- xt.fit[, (1:(dim(xt.fit)[2] - 1)), drop = FALSE]

  # Whereas J is of constant dimension, L and K dimensions may vary
  # depending on existence of NAs
  J <- array(0, c(r, r, T))
  L <- list()
  K <- list()

```

```

for (i in 1:(T-1)) {
  J[,i] <- Pt[,i] %*% t(Tt) %*% solve(Pt.fit[,,(i+1)], tol = 1e-32)
}

Zit <- Zt
Hit <- Ht
for (i in 1:T) {
  # Only keep entries for non-missing data
  Zt <- Zit[W[i,], drop=FALSE]
  Ht <- Hit[W[i,], W[i,], drop=FALSE]
  L[[i]] <- solve(Zt %*% Pt.fit[,i] %*% t(Zt) + Ht)
  K[[i]] <- Pt.fit[,i] %*% t(Zt) %*% L[[i]]
}

xt.sm <- cbind(matrix(0, r, (T-1)), xt[,T])
Pt.sm <- array(0, c(r, r, T))
PtT.sm <- array(0, c(r, r, T))
Pt.sm[,T] <- Pt[,T]
PtT.sm[,T] <- (diag(1, r) - K[[T]] %*% Zi) %*% Tt %*% Pt[,,(T-1)]

for (j in 1:(T-1)) {
  xt.sm[,,(T-j)] <- xt[,,(T-j)] + J[,,(T-j)] %*% (xt.sm[,,(T+1-j)] - xt.fit[,,(T+1-j)])
  Pt.sm[,,(T-j)] <- Pt[,,(T-j)] + J[,,(T-j)] %*% (Pt.sm[,,(T+1-j)] - Pt.fit[,,(T+1-j)]) %*% t(J[,,(T-j)])
}

for (j in 1:(T-2)) {
  PtT.sm[,,(T-j)] <- Pt[,,(T-j)] %*% t(J[,,(T-j-1)]) + J[,,(T-j)] %*% (PtT.sm[,,(T-j+1)]
    - Tt %*% Pt[,,(T-j)]) %*% t(J[,,(T-j-1)])
}

return(list(xt.sm = xt.sm, Pt.sm = Pt.sm, PtT.sm = PtT.sm))
}

```

4. Ο αλγόριθμος EM

Ο συγκεκριμένος αλγόριθμος αποτελεί μέρος του πακέτου `dynfactoR` της R.

```

#' Computation of the expectation step in the EM-algorithm.

```

```

#' @param y Data matrix

```

```

#' @param A System state matrix
#' @param C Observation matrix
#' @param R Observation equation variance
#' @param Q System state equation variance
#' @param W Logical matrix with  $\dim(W) = \dim(y)$ 
#' indicating missing observations
#' @param initx Initial value for state variable
#' @param initV Initial value for state matrix
#' @return Sufficient statistics used for M-step
Estep <- function(y, A, C, Q, R, initx, initV, W) {

  os <- dim(y)[1]
  T <- dim(y)[2]
  ss <- nrow(A)

  kf <- K_filter(initx, initV, t(y), A, C, R, Q)
  ks <- K_smoother(A, kf$xit, kf$xitm, kf$Ptt, kf$Ptm, C, R, W)

  xsmooth <- ks$xitT
  Vsmooth <- ks$PtT
  Wsmooth <- ks$PtTm

  delta <- matrix(0, os, ss)
  gamma <- matrix(0, ss, ss)
  beta <- matrix(0, ss, ss)

  for (t in 1:T) {
    z <- y[,t]; z[is.na(z)] <- 0
    # There seems to be a problem here
    delta <- delta + z %*% t(xsmooth[,t])
    gamma <- gamma + xsmooth[,t] %*% t(xsmooth[,t]) + Vsmooth[,t]
    if (t > 1) {
      beta <- beta + xsmooth[,t] %*% t(xsmooth[, (t-1)]) + Wsmooth[,t]
    }
  }

  gamma1 <- gamma - xsmooth[, T] %*% t(xsmooth[, T]) - Vsmooth[, , T]
  gamma2 <- gamma - xsmooth[, 1] %*% t(xsmooth[, 1]) - Vsmooth[, , 1]
  x1 <- xsmooth[, 1]

```



```

V1 <- Vsmooth[, , 1]

return(list(beta_t = beta, gamma_t = gamma, delta_t = delta, gamma1_t = gamma1,
            gamma2_t = gamma2, x1 = x1, V1 = V1, loglik_t = kf$loglik, xsmooth = xsmooth))

}

#' Convergence test for EM-algorithm.
#'
#' @param loglik Current value of the log-likelihood function
#' @param previous_loglik Value of the log-likelihood function at the previous
# iteration
#' @param threshold If difference is less than threshold, then algorithm has
#' converged
#' @param check_increased TO DOCUMENT
#' @return A logical statement indicating whether EM algorithm has converged
#' according to slope convergence test
em_converged <- function(loglik, previous_loglik, threshold=1e-4, check_increased=TRUE) {

  converged <- FALSE
  decrease <- 0

  if (check_increased == TRUE) {
    if (loglik - previous_loglik < -0.001) {
      # cat("*** Likelihood decreased from ", previous_loglik, " to ", loglik, "\n")
      decrease <- 1
    }
  }

  delta_loglik <- abs(loglik - previous_loglik)
  avg_loglik <- (abs(loglik) + abs(previous_loglik) + Machine$double.eps)/2

  if ((delta_loglik/avg_loglik) < threshold) {
    converged <- TRUE
  }
  return(converged)
  # return(list('converged'=converged, 'decrease'=decrease))
}

```

5. Δημιουργία 95% διαστήματος εμπιστοσύνης για την πρόβλεψη της επόμενης περιόδου

```
UB = fit$Tt%%fit$twostep[42,]+1.96*sqrt(diag(fit$Qt + fit$Tt%%fit$Q%%t(fit$Tt)))
LB = fit$Tt%%fit$twostep[42,]-1.96*sqrt(diag(fit$Qt + fit$Tt%%fit$Q%%t(fit$Tt)))

UB=(fit$Zt%%UB)[1]*sd(na.omit(n.data[,1]))+mean(na.omit(n.data[,1]))
PE=(fit$Zt%%fit$twostep[42,])[1]*sd(na.omit(n.data[,1]))+mean(na.omit(n.data[,1]))
LB=(fit$Zt%%LB)[1]*sd(na.omit(n.data[,1]))+mean(na.omit(n.data[,1]))
```

6. Μετατροπή δεδομένων που έχουν προκύψει από δεύτερες διαφορές στην κανονική τους μορφή.

```
y1=0
for (i in 2:length(fitted)) {
  y1[i]=diff(data[,1])[i-1]+fitted[i]
}

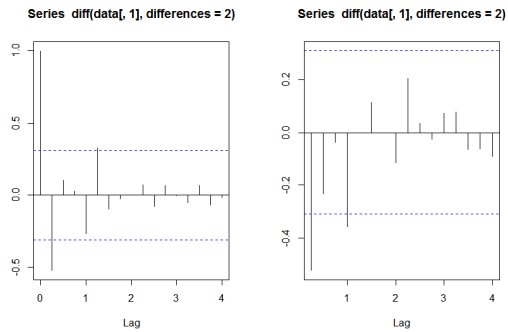
y1[1]=NA

y0=0
for (i in 2:length(fitted)) {
  y0[i]=data[i-1,1]+y1[i]
}

y0[1]=NA
ts.plot(y0)
```

7. Ταυτοποίηση, εκτίμηση και διαγνωστικοί έλεγχοι υποδείγματος ARIMA.

```
par(mfrow=c(1,2))
acf(diff(data[,1],differences=2))
pacf(diff(data[,1],differences=2))
```



```
adf.test(na.omit(x[,1]))
fiit = arima(x[,1], order=c(2,0,4))
ar.msfe = sum(na.omit(fiit$residuals)^2)
Box.test(fiit$res,type="Ljung-Box")
```

B3. Επαυξημένος έλεγχος Dickey-Fuller και έλεγχος Ljung-Box

Augmented Dickey-Fuller Test

data: na.omit(x[, 1]) Dickey-Fuller = -5.9775, Lag order = 3, p-value = 0.01 alternative hypothesis: stationary

Ljung-Box test

data: fiit\$res X-squared = 0.038051, df = 1, p-value = 0.8453

Βιβλιογραφία

Ελληνική

- [1] Αγιακλόγλου Χ.Ν., Οικονόμου Γ.(2004). Μέθοδοι Προβλέψεων και Ανάλυσης Αποφάσεων. *Εκδόσεις Μπένου Κεφ.6.*
- [2] Δημελή Σ.(2004). Σύγχρονες Μέθοδοι Ανάλυσης Χρονολογικών Σειρών. *Εκδόσεις ΟΠΑ.*
- [3] Κούτρας Μ.Β.(2004). Εφαρμοσμένη Πολυμεταβλητή Ανάλυση, Πανεπιστημιακές Σημειώσεις, ΠΜΣ «Εφαρμοσμένη Στατιστική».

Ξένη

- [4] Agiakloglou C., Newbold P. (1992). Empirical evidence on Dickey-Fuller type test. *Journal of Time Series Analysis 13*, 471-483.
- [5] Bai J., Ng S.(2002). Determining the number of factors in approximate factor models. *Econometrica 70*, 191-221.
- [6] Bartlett M.S.(2002). The statistical conception of mental factors. *British Journal of Psychology 28*, 97-104.
- [7] Boiving J., Ng S.(2003). Are more data always better for factor analysis? *Journal of Econometrics 132*, 169-194.
- [8] Box G.E.P., Jenkins G.M.(1970). Time series Analysis: Forecasting and Control. *Holden-Day.*
- [9] Chamberlain G., Rothschild(1983). íArbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica 51*, 1305-1324.

- [10] de Jong P.(1988). Covariances for smoothed estimates in state space models. *Biometrika* 75,. 601-602.
- [11] de Jong P.(1989). Smoothing and interpolation with the state space model. *Journal of American Statistical Association* 84,. 1085-1088.
- [12] Dickey D.A., Fuller W.A.(1981). Distribution of the estimators for autoregressive time series with a unit root. *Econometrica* 49, 1057-1072.
- [13] Doz C., Giannone D., Reichelin L.(2012). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics* 164, 188-205.
- [14] Doz C., Giannone D., Reichelin L.(2011). A Quasi–Maximum Likelihood Approach for Large, Approximate Dynamic Factor Models. *The Review of Economics and Statistics* 94, 1014-1024.
- [15] Durbin J., Koopman S.J.(2012). Time Series Analysis by State Space Methods. *Oxford Statistical Series* 38.
- [16] Engle R.F., M.W. Watson. (1981). A one-factor multivariate time series model of metropolitan wage rates. *The Review of Economics and American Statistical Association* 76, 774-781.
- [17] Forni M.,Hallin M., Lippi M., Reichelin L.(2000). The generalise factor model: identification and estimation. *The Review of Economics and Statistics* 82, 540-554.
- [18] Geweke J.(1977) The dynamic factor analysis of economic time series. *Latent variable in socio-economic models* Ch.19 in Aigner D.J.,Goldberg A.S..
- [19] Giannone D.,Reichelin L., Small D.(2005). Nowcasting GDP and inflation: the real-time information content of macroeconomic data releases. *Journal of Monetary Economics*

- [20] Hamilton J.D.(1994). Time Series Analysis. *Princeton University Press*
- [21] Harvey C.A.(1989). Forecasting structural time series models and the Kalman filter. *Cambridge University Press.*
- [22] Johnson A.R., Wichern W.D.(2014) Applied Multivariate Statistical Analysis(6th Edition). *Pearson*
- [23] Kalman R.E.(1960). A new approach to linear filter and prediction problems. *Journal of Basic Engineering* 82, 35-45.
- [24] Kapetanios G., Marcellino M.(2006). A Parametric Estimation Method For Dynamic Factor Models of Large Dimensions. *Centre for Economic Policy Research* 5620
- [25] Kohn R., C.F. Ansley. (1989). A fast algorithm for signal extraction, influence and cross-validation. *Biometrika* 76, 65-79.
- [26] Lawley D.N., Maxwell A.E. (1962). Factor Analysis as a Statistical Method. *Journal of the Royal Statistical Society. Series D (The Statistician)*, Vol.12, No. 3, 209-229
- [27] Lipschutz S., Lipson M.L.(2013) Schaum's Outline of Linear Algebra (5th Edition). *McGraw-Hill*
- [28] Lutkepohl H.(2005). New Introduction to Multiple Time Series Analysis. *Springer.*
- [29] Sargent T.J., Sims C.A.(1977) Business cycle modelling without pretending to have too much a-priori economic theory. *New methods in business cycle research.* Federal Reserve Bank of Minneapolis.
- [30] Stock J.H., Watson M.W.(1999). Forecasting inflation. *Journal of Monetary Economics* 44,. 293-335.

- [31] Stock J.H., Watson M.W.(2002b). Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association* 97,. 1167-1179.
- [32] Stock J.H., Watson M.W.(2005). An empirical comparison of methods for forecasting using many predictors. *Unpublished manuscript*.
- [33] Stock J.H., Watson M.W.(2010). Dynamic Factor Models. *Oxford Handbook of Economic Forecasting*.
- [34] Tsay S.R.(2010). Analysis of Financial Time Series(Third Edition). *Wiley*.