

ΠΑΝΤΕΙΟΝ ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΟΙΝΩΝΙΚΩΝ ΚΑΙ ΠΟΛΙΤΙΚΩΝ ΣΠΟΥΔΩΝ

PANTEION UNIVERSITY OF SOCIAL AND POLITICAL SCIENCES



ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΗΣ ΚΑΙ ΠΕΡΙΦΕΡΕΙΑΚΗΣ ΑΝΑΠΤΥΞΗΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΚΑΤΕΥΘΥΝΣΗ «ΕΦΗΡΜΟΣΜΕΝΩΝ ΟΙΚΟΝΟΜΙΚΩΝ ΚΑΙ ΔΙΟΙΚΗΣΗΣ»

Now-casting the Greek GDP

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Χρήστος-Σωτήριος Καβάλλος

Αθήνα, 2018

Τριμελής Επιτροπή

Δρ. Ντεγιαννάκης Σταύρος, Επιβλέπων Καθηγητής

Δρ. Λώλος Σαράντης - Ευάγγελος, Καθηγητής

Δρ. Ρίτσαρντσον Στέφανος - Κλάιβ, Καθηγητής



Copyright © Χρήστος-Σωτήριος Καβάλλος

All rights reserved. Με επιφύλαξη παντός δικαιώματος

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

Η έγκριση της διπλωματικής εργασίας από το Πάντειον Πανεπιστήμιο Κοινωνικών και Πολιτικών Επιστημών δεν δηλώνει αποδοχή των γνώμων του συγγραφέα.

Ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε την περίοδο Σεπτέμβριος 2017 – Απρίλιος 2018 στα πλαίσια του Μεταπτυχιακού Προγράμματος Σπουδών «Εφαρμοσμένων Οικονομικών και Περιφερειακής Ανάπτυξης», κατεύθυνσης «Εφαρμοσμένων Οικονομικών και Διοίκησης» του τμήματος Οικονομικής και Περιφερειακής Ανάπτυξης του Παντείου Πανεπιστημίου Κοινωνικών και Πολιτικών Επιστημών Αθηνών. Πριν από την παρουσίαση αυτής της διπλωματικής εργασίας θα ήθελα να ευχαριστήσω όλους όσους συνέβαλαν στην πραγματοποίηση της και ιδιαίτερα:

Τον επιβλέποντα καθηγητή κ.Ντεγιαννάκη Σταύρο για την καθοδήγηση του σε όλη την διάρκεια της έρευνας καθώς και για την αμέριστη συμπαράσταση του στην εκπόνηση της εργασίας.

Ακόμα θα ήθελα να ευχαριστήσω όλους τους συναδέλφους μου, μεταπτυχιακούς φοιτητές, για τις ανταλλαγές απόψεων και για τη σημαντική βοήθεια τους σε όλα τα στάδια της εργασίας και του μεταπτυχιακού προγράμματος γενικότερα.

Τέλος, η εργασία αυτή αφιερώνεται στους γονείς μου, Νικόλαο και Ελένη καθώς και στην Βασιλική Μπαϊρακτάρη για την απεριόριστη στήριξη και εμπιστοσύνη τους σε μένα.

Περιεχόμενα

Ευχαριστίες	3
Περιεχόμενα.....	4
Πίνακες	5
Γραφήματα.....	5
Περίληψη	8
Abstract	9
1.Εισαγωγή	11
1.1. Αντικείμενο εργασίας.....	12
1.2. Δομή εργασίας.....	12
2.Μέθοδοι Μακρο-οικονομικών Προβλέψεων.....	14
3.Πρόβλεψη Ελληνικού ΑΕΠ.....	22
3.1.Περιβάλλον υλοποίησης	23
3.2.Δεδομένα	24
3.3.Μεθοδολογία	33
3.4.Σφάλματα	46
4.Αποτελέσματα - Συμπεράσματα	47
Βιβλιογραφία	57
Παράρτημα.....	60

Πίνακες

Πίνακας 1 : Χρονοσειρές του πειράματος, πηγές και χρονικές υστερήσεις των πραγματικών δεδομένων σε σχέση με τον τρέχοντα μήνα που εκτελείται το πείραμα κάθε φορά.....	32
Πίνακας 2 : Αριθμός μηνών που χρειάζεται να συμπληρωθούν για την πρόβλεψη του ΑΕΠ.....	34
Πίνακας 3 : Εποχικά διορθωμένες χρονοσειρές.....	35
Πίνακας 4 : Μεταβλητές σε μορφή ρυθμού μεταβολής.....	37
Πίνακας 5 : Έλεγχοι στασιμότητας – Stationarity tests. (ADF: Augmented Dickey–Fuller test, DFGLS: Elliott, Rothenberg, and Stock modified DF test, PP: Phillips-Perron test).	40
Πίνακας 6 : Συνιστώσες PCA. Ο αριθμός (πλήθος) των συνιστωσών που συγκεντρώνουν το 70% και 80% αντίστοιχα της πληροφορίας του συνόλου των χρονοσειρών για τους 95 μήνες που δοκιμάστηκαν στο πείραμα.....	44
Πίνακας 7 : Στατιστικά σφάλματα, δείκτες ($e_t = y_t - \hat{y}_t$)	45
Πίνακας 8 : Σφάλματα προβλέψεων των model70 και model80.....	47

Γραφήματα

Γράφημα 1. Πρόβλεψη πραγματικού ΑΕΠ (ΟΟΣΑ - Ελλάδα). Με κόκκινο χρώμα η Ελληνική πρόβλεψη. Με μαύρο χρώμα η πρόβλεψη για τις χώρες του ΟΟΣΑ.	12
Γράφημα 2. Δείκτης Τιμών Παραγωγού στη Βιομηχανία (γενικός δείκτης) – Producer Price Index in Industry (PPI) (overall index)	33
Γράφημα 3. Αφαίρεση εποχικότητας για τον Δείκτη Κύκλου Εργασιών στο Λιανικό Εμπόριο (γενικός δείκτης) - Turnover Index in Retail Trade (overall index) (μπλέ: διορθωμένη χρονοσειρά)	36
Γράφημα 4. Μετασχηματισμός πρώτων διαφορών ρυθμών μεταβολής για την μεταβλητή του Γενικού Δείκτη XAA - ASE General Index.....	37

Γράφημα 5. Box Plot μετασχηματισμένης χρονοσειράς Γενικού Δείκτη ΧΑΑ (αριστερά: με outliers, δεξιά: χωρίς)	38
Γράφημα 6. Outlier cropping Γενικού Δείκτη ΧΑΑ (κόκκινο: τα outliers, μπλε: η χρονοσειρά μετά την επεξεργασία)	39
Γράφημα 7 : Σύγκριση πραγματικού ΑΕΠ (μπλε) και model70_0m (κόκκινο). Το 0m δείχνει ότι οι προβλέψεις έγιναν με δεδομένα που είχανε πραγματικές τιμές μέχρι και τον τρέχοντα μήνα δημοσίευσης του ΑΕΠ).	48
Γράφημα 8 : Σύγκριση πραγματικού ΑΕΠ (μπλε) και model70_1m (κόκκινο). Το 1m δείχνει ότι οι προβλέψεις έγιναν με δεδομένα που είχανε πραγματικές τιμές μέχρι και τον προηγούμενο μήνα από δημοσίευση του ΑΕΠ.	48
Γράφημα 9 : Σύγκριση πραγματικού ΑΕΠ (μπλε) και model70_2m (κόκκινο). Το 2m δείχνει ότι οι προβλέψεις έγιναν με δεδομένα που είχανε πραγματικές τιμές μέχρι και τον προ-προηγούμενο μήνα από δημοσίευση του ΑΕΠ.	49
Γράφημα 10 : Σύγκριση πραγματικού ΑΕΠ (μπλε) και model80_0m (κόκκινο). Το 0m δείχνει ότι οι προβλέψεις έγιναν με δεδομένα που είχανε πραγματικές τιμές μέχρι και τον τρέχοντα μήνα δημοσίευσης του ΑΕΠ.	49
Γράφημα 11 : Σύγκριση πραγματικού ΑΕΠ (μπλε) και model80_1m (κόκκινο). Το 1m δείχνει ότι οι προβλέψεις έγιναν με δεδομένα που είχανε πραγματικές τιμές μέχρι και τον προηγούμενο μήνα από δημοσίευση του ΑΕΠ.	50
Γράφημα 12 : Σύγκριση πραγματικού ΑΕΠ (μπλε) και model80_2m (κόκκινο). Το 2m δείχνει ότι οι προβλέψεις έγιναν με δεδομένα που είχανε πραγματικές τιμές μέχρι και τον προ-προηγούμενο μήνα από δημοσίευση του ΑΕΠ.	50
Γράφημα 13 : Σύγκριση σφαλμάτων 0m των model70 (μπλε) και model80 (κόκκινο). Το 0m δείχνει ότι οι προβλέψεις έγιναν με δεδομένα που είχανε πραγματικές τιμές μέχρι και τον τρέχοντα μήνα δημοσίευσης του ΑΕΠ.	51
Γράφημα 14 : Σύγκριση σφαλμάτων 1m των model70 (μπλε) και model80 (κόκκινο). Το 1m δείχνει ότι οι προβλέψεις έγιναν με δεδομένα που είχανε τιμές μέχρι και τον προηγούμενο μήνα από δημοσίευση του ΑΕΠ.	51

Γράφημα 15 : Σύγκριση σφαλμάτων 2m των model70 (μπλε) και model80 (κόκκινο). Το 2m δείχνει ότι οι προβλέψεις έγιναν με δεδομένα που είχανε τιμές μέχρι και τον προ-προηγούμενο μήνα από δημοσίευση του ΑΕΠ.	52
Γράφημα 16 : Σύγκριση σφαλμάτων 0m των model70 (αριστερά) και model80 (δεξιά) σε ραβδογράμματα. Το 0m δείχνει ότι οι προβλέψεις έγιναν με δεδομένα που είχανε πραγματικές τιμές μέχρι και τον τρέχοντα μήνα δημοσίευσης του ΑΕΠ.	53
Γράφημα 17 : Σύγκριση σφαλμάτων 1m των model70 (αριστερά) και model80 (δεξιά) σε ραβδογράμματα. Το 1m δείχνει ότι οι προβλέψεις έγιναν με δεδομένα που είχανε πραγματικές τιμές μέχρι και τον προηγούμενο μήνα από δημοσίευση του ΑΕΠ.	53
Γράφημα 18 : Σύγκριση σφαλμάτων 2m των model70 (αριστερά) και model80 (δεξιά) σε ραβδογράμματα. Το 2m δείχνει ότι οι προβλέψεις έγιναν με δεδομένα που είχανε πραγματικές τιμές μέχρι και τον προ-προηγούμενο μήνα από δημοσίευση του ΑΕΠ.	55
Γράφημα 19 : Εποχικότητα των προβλέψεων. Από πάνω προς τα κάτω: model70_0m, model70_1m, model70_2m, model80_0m, model80_1m, model80_2m (Ο δείκτης "_Xm " δηλώνει την χρονική υστέρηση των πραγματικών δεδομένων από τον μήνα δημοσίευσης του ΑΕΠ)	55

Περίληψη

Οι ποσοτικές πληροφορίες σχετικά με την τρέχουσα κατάσταση της οικονομίας είναι κρίσιμες για τη χάραξη της οικονομικής πολιτικής και την έγκαιρη κατανόηση της οικονομικής κατάστασης μιας χώρας, αλλά τα στοιχεία για το τριμηνιαίο εθνικό λογαριασμό του ΑΕΠ στη ζώνη του ευρώ δημοσιεύονται με σημαντική καθυστέρηση.

Στην παρούσα διπλωματική εργασία αναζητείται ένα μοντέλο πρόβλεψης του Ελληνικού ΑΕΠ σε χρονικό ορίζοντα τρέχοντος τριμήνου (current-quarter). Για την επίτευξη του στόχου, γίνεται εκτενής αναφορά στις μέχρι σήμερα προσεγγίσεις επί του θέματος από διακεκριμένους οικονομολόγους της Ευρωζώνης και παγκοσμίου βεληνεκούς γενικότερα. Επιδιώκεται η παράθεση και σύγκριση υφιστάμενων μεθόδων πρόβλεψης και τέλος ο συνδιασμός ορισμένων παραγόντων και χαρακτηριστικών τους, που θα αποτελέσουν μέρος του επιλεγμένου μοντέλου, το οποίο επαυξάνεται μέσω της οπτικής του συγγραφέα.

Αξίζει να σημειωθεί πως με την πρόοδο της τεχνολογίας και των υπολογιστικών μέσων, έχει επέλθει πρόοδος στους τομείς της οικονομετρίας και των οικονομικών γενικότερα και αυτό έχει ως συνέπεια την διευκόλυνση της ερευνητικής διαδικασίας και την δυνατότερη παραγωγή θεωρητικής και εμπειρικής γνώσης. Οι επιστήμες της Πληροφορικής και των Οικονομικών συνεργάζονται και συνεργούν, η μεν με την παραγωγή εργαλείων και με την ευκολότερη και αποτελεσματικότερη επεξεργασία μεγάλου όγκου δεδομένων και η δε με την θεωρία.

Τα μοντέλα χρονοσειρών υιοθετούνται συχνά για την πρόβλεψη λόγω της απλότητας και της καλής απόδοσής τους. Ο αριθμός των παραμέτρων σε αυτά τα μοντέλα αυξάνεται ραγδαία με τον αριθμό των μεταβλητών που έχουν μοντελοποιηθεί, έτσι ώστε συνήθως να λαμβάνονται υπόψη μόνο τα μονοδιάστατα μοντέλα ή τα μικρής κλίμακας πολυπαραγοντικά μοντέλα. Ωστόσο, τα δεδομένα είναι τώρα άμεσα διαθέσιμα για έναν πολύ μεγάλο αριθμό μακροοικονομικών μεταβλητών που είναι δυνητικά χρήσιμοι κατά την πρόβλεψη.

Στη συνέχεια της εργασίας παρουσιάζονται υπάρχοντα οικονομικά μοντέλα πρόβλεψης και τέλος μελετάται και προτείνεται μία νέα λύση καθώς και τα αποτελέσματα του πειράματος για την Ελλάδα. Κατασκευάζεται ένα μεγάλο μακροοικονομικό σύνολο δεδομένων με περίπου 175 μεταβλητές, το οποίο μοντελοποιείται χρησιμοποιώντας την ανάλυση κυρίων συνιστωσών σε δύο μοντέλα τα οποία συγκρίνονται σύμφωνα με τα σφάλματά τους. Το πιο αποδοτικό μοντέλο

της τρέχουσας διατριβής παρήγαγε μέση απόλυτη απόκλιση ύψους 760,18 εκατομμυρίων ευρώ ανά τρίμηνο (σφάλμα 1,58%).

Λέξεις Κλειδιά : πρόβλεψη, σύγχρονη πρόβλεψη (τρέχοντος τριμήνου), ανάλυση κύριων συνιστωσών (PCA), μοντέλα χρονοσειρών.

Abstract

Quantitative information on the current state of the economy is crucial for economic policy-making and a timely understanding of a country's economic situation, but data on quarterly national GDP in the euro area are published with a long delay.

In this diploma thesis, a model for forecasting Greek GDP is sought in a next-quarter horizon. In favor of achieving this goal, there is an extensive reference to the approaches that have been taken up to date by distinguished economists in the Eurozone and World-wide range. It is intended to present and compare existing forecasting methods and to combine certain factors and their characteristics, which will form part of the chosen model, which is augmented by the author's view.

It is worth noting that progress in technology and computing has made an impact in econometrics and economics in general, and this has the effect of facilitating the research process as well as the stronger production of theoretical and empirical knowledge. The sciences of Information Technology and Economics collaborate, the former with the production of tools and with the easier and more efficient processing of large volumes of data and the latter with its Theory.

Time series models are often adopted for forecasting because of their simplicity and performance. The number of parameters in these models increases rapidly with the number of variables that have been modeled so that only one-dimensional models or small-scale multifactor models are usually taken into consideration. However, data are now readily available for a very large number of macroeconomic variables that are potentially useful in forecasting.

Following, in the current thesis, existing economic forecasting models are presented and finally a new solution is being researched and proposed and the experimental results regarding Greece are analyzed. A large macro-data set with approximately 175 variables is being formed, from which two principal components models are being produced and are compared considering their errors. The most efficient model of the current thesis produced a mean absolute deviation of 760.18 million Euro per quarter (1.58% error).

Key Words: forecasting, now-casting (next quarter), principal component analysis (PCA), time series models, vector auto-regression, big data analysis

1.Εισαγωγή

Όπως είναι ευρέως γνωστό, ειδικότερα αφότου η οικονομική κρίση έπληξε την Ελληνική οικονομία, το ακαθάριστο εγχώριο προϊόν μιας χώρας είναι ένας ιδιαίτερα σημαντικός, αν όχι ο σημαντικότερος, δείκτης πορείας της οικονομίας αυτής. Ο ΟΟΣΑ ορίζει το ΑΕΠ ως "συνολική μέτρηση παραγωγής ίση με το άθροισμα των ακαθάριστων προστιθέμενων αξιών όλων των κατοίκων και των θεσμικών μονάδων που συμμετέχουν στην παραγωγή (συν τυχόν φόρους και μείον τυχόν επιδοτήσεις για προϊόντα που δεν περιλαμβάνονται στην αξία των αποτελεσμάτων τους) (OECD). Το ΔΝΤ αναφέρει ότι «το ΑΕΠ μετρά τη νομισματική αξία των τελικών αγαθών και υπηρεσιών που αγοράζονται από τον τελικό χρήστη που παράγονται σε μια χώρα σε μια δεδομένη χρονική περίοδο (π.χ. έτος)». (Callen, 2017) Συνεπώς μπορεί εύκολα να αναλογιστεί κανείς την σημαντικότητα της δυνατότητας πρόβλεψης του ΑΕΠ σε όσο το δυνατόν μικρότερο χρονικό διάστημα συνεκτιμώντας πως η οικονομική ανάπτυξη είναι η αύξηση της διορθωμένης από τον πληθωρισμό αγοραίας αξίας των αγαθών και των υπηρεσιών που παράγει μια οικονομία με την πάροδο του χρόνου και συνήθως μετριέται ως το ποσοστό αύξησης του πραγματικού ακαθάριστου εγχώριου προϊόντος ή του πραγματικού ΑΕΠ (IMF, 2012).

Η πρόβλεψη της οικονομίας είναι συχνά ένα επικίνδυνο έργο. Αν και αντιπροσωπεύει ένα από τα βασικά προβλήματα της στατιστικής ανάλυσης, προκύπτουν αρκετές δυσκολίες στην ανάπτυξη ενός μοντέλου που επιτυχώς και σταθερά παρέχει καλές και αξιόπιστες προβλέψεις. Τα δεδομένα μπορεί να είναι κατεστραμμένα και συχνά μολυσμένα με σφάλματα μέτρησης. Επιπλέον, μια ποικιλία κραδασμών - σοκ πλήττει την οικονομία σε κάθε χρονική στιγμή. Τέλος, η οικονομική δομή αλλάζει συχνά, και μαζί της οι σχέσεις συμπεριφοράς που την περιγράφουν.

Ωστόσο, οι πρόσφατες εξελίξεις στην τεχνολογία της πληροφορίας καθιστούν δυνατή την πρόσβαση σε πραγματικό χρόνο, με λογικό κόστος, χιλιάδων οικονομικών χρονοσειρών για μεγάλες ανεπτυγμένες οικονομίες. Αυτό δημιουργεί την προοπτική ενός νέου συνόρου στις μακροοικονομικές προβλέψεις, στην οποία χρησιμοποιείται ένας πολύ μεγάλος αριθμός χρονολογικών σειρών για την πρόβλεψη μερικών βασικών οικονομικών μεγεθών.

Στην διπλωματική εργασία θα προσπαθήσουμε να αναζητήσουμε τρόπους έγκυρους και αποτελεσματικούς για την πρόβλεψη του μακροοικονομικού μεγέθους ΑΕΠ. Τα μοντέλα που θα δημιουργηθούν, θα εξεταστούν εμπειρικά για την περίπτωση της Ελλάδος, συμπεριλαμβάνοντας πληθώρα τρίτων μακροοικονομικών μεγεθών που αντικατοπτρίζουν την εγχώρια οικονομία. Τα

μοντέλα που θα παραχθούν, θα στηριχθούν σε παλαιότερα δοκιμασμένα προβλεπτικά μοντέλα με απαραίτητες διαφοροποιήσεις.

1.1. Αντικείμενο εργασίας

Το κύριο αντικείμενο της εργασίας είναι να μελετηθεί η ανάπτυξη νέων μοντέλων πρόβλεψης του ΑΕΠ, χρησιμοποιώντας ένα διευρημένο σύνολο μακροοικονομικών μεταβλητών με μορφή χρονοσειρών. Συν τοις άλλοις, θα πραγματοποιηθεί πείραμα των μοντέλων αυτών για το Ελληνικό ΑΕΠ και θα αποτυπωθούν τα μέσα σφάλματα των μοντέλων σε βάθος 8-ετίας.

Δευτερεύοντα αλλά σημαντικά αντικείμενα της εργασίας αυτής, είναι να αναλύσει σε βάθος το πλαίσιο το οποίο διαμορφώνεται από τη συνεχή έρευνα πρόβλεψης μακροοικονομικών μεγεθών και ειδικότερα του ΑΕΠ καθώς επίσης και να παραθέσει τρίτες προσεγγίσεις παλαιότερων ετών επί του θέματος. Ένα παράδειγμα προσπάθειας πρόβλεψης του Ελληνικού ΑΕΠ από τον ΟΟΣΑ φαίνεται στο Γράφημα 1.

1.2. Δομή εργασίας

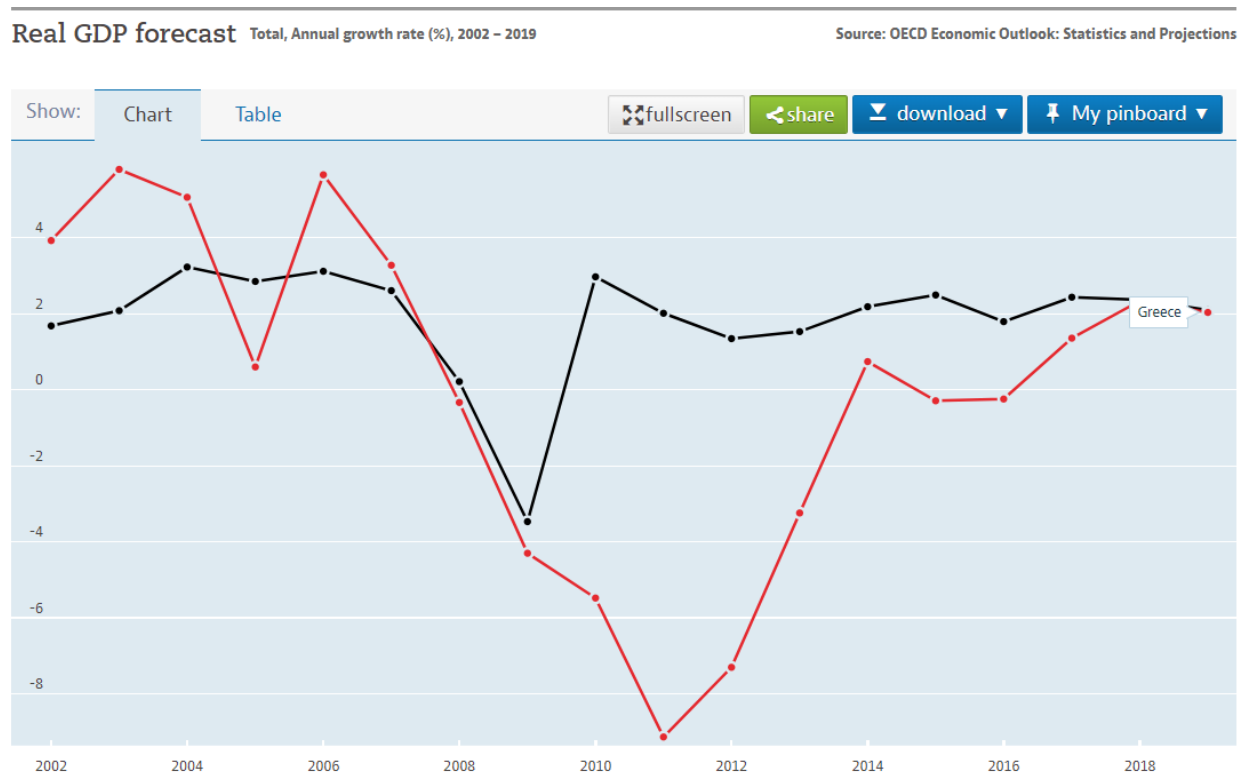
Κεφάλαιο πρώτο: Περιγράφεται σύντομα το αντικείμενο της διπλωματικής εργασίας και γίνεται μία μικρή ανάλυση των επομένων κεφαλαίων.

Κεφάλαιο δεύτερο: Σε αυτό το κεφάλαιο γίνεται γενική αναφορά στις υπάρχουσες μεθόδους πρόβλεψης μακροοικονομικών μεγεθών και ειδικότερα του ΑΕΠ διαφόρων χωρών και στα αποτελέσματα που αποκομίστηκαν από την εμπειρική ανάλυση και χρήση των μοντέλων. Γίνεται ιστορική αναφορά και αναδρομή. Επιπλέον, εντοπίζονται ασυμφωνίες και διαφορετικότητες διακεκριμένων οικονομολόγων.

Κεφάλαιο τρίτο: Γίνεται παρουσίαση της μεθοδολογίας του συγγραφέα για την ανάπτυξη των δύο μοντέλων πρόβλεψης του Ελληνικού ΑΕΠ, του περιβάλλοντος ανάπτυξης και έρευνας, των δεδομένων που χρησιμοποιήθηκαν και διαφόρων μετασχηματισμών που

επιλέχθηκαν, απαραίτητων για την επίτευξη του στόχου. Αναφέρονται και οι περίπου 175 μεταβλητές - χρονοσειρές

Κεφάλαιο τέταρτο: Σε αυτό το κεφάλαιο αναλύονται τα αποτελέσματα των αναπτυγμένων μοντέλων για την Ελλάδα και τα σφάλματά τους. Επιπρόσθετα, πραγματοποιείται ανάλυση συμπερασμάτων και προτάσεις για περαιτέρω έρευνα.



Γράφημα 1: Πρόβλεψη πραγματικού ΑΕΠ (ΟΟΣΑ - Ελλάδα). Με κόκκινο χρώμα η Ελληνική πρόβλεψη. Με μαύρο χρώμα η πρόβλεψη για τις χώρες του ΟΟΣΑ.

Πηγή : Οργανισμός Οικονομικής Συνεργασίας και Ανάπτυξης (ΟΟΣΑ)

2.Μέθοδοι Μακρο-οικονομικών Προβλέψεων

Κατα καιρούς έχουν χρησιμοποιηθεί διάφοροι τρόποι προβλέψεων μακροοικονομικών μεγεθών. Εάν ξεφύγουμε από τη λογική των προβλεπτικών μοντέλων ελαχίστων ερμηνευτικών οικονομικών μεταβλητών, θα βρεθούμε προ εκπλήξεως όταν αντιληφθούμε την πληθώρα διαφορετικών μεθόδων και προσεγγίσεων που έχουν ακολουθηθεί, ιδιαίτερα κατά την τελευταία εικοσαετία, οι οποίες συχνά είναι αντικρουόμενες και εξαρτώνται από τον χρόνο του πειράματος και την χώρα για την οποία εκτελείται. Η ανασκόπησή μας θα επεκταθεί από τις αρχές του 21^{ου} αιώνα μέχρι και τα τελευταία χρόνια. Ωστόσο ακόμη και αυτή η έκταση στον χρόνο αποτελεί μόνο ένα κομμάτι των συγκλονιστικών εξελίξεων στον χώρο των προβλέψεων.

Κατά κύριο λόγο, δύο προσεγγίσεις εμφανίστηκαν πρόσφατα για την εξαγωγή πληροφοριών από μεγάλα μακροοικονομικά σύνολα δεδομένων. Και οι δύο βασίζονται στην ερμηνευση των μεταβλητών με μοντέλα δυναμικών παραγόντων. Στη μέθοδο των Stock και Watson (1998) χρησιμοποιούνται στατικές κύριες συνιστώσες για την εκτίμηση των παραγόντων, ενώ οι Forni et al. (2000) βασίζονται σε δυναμικές κύριες συνιστώσες. Οι Favero et al. (2002), στο πείραμά τους, εφήρμοσαν και τις δύο μεθόδους σε δύο μεγάλα μηνιαία σύνολα δεδομένων, το ένα για τις ΗΠΑ που περιείχε 146 μακροοικονομικές χρονολογικές σειρές και το άλλο για τις τέσσερις μεγαλύτερες χώρες της Ευρωπαϊκής Νομισματικής Ένωσης, για συνολικά 105 μεταβλητές. Και στις δύο περιπτώσεις η ερμηνευτική απόδοση των εκτιμώμενων παραγόντων για τις μακροοικονομικές μεταβλητές ήταν ικανοποιητική, ελαφρώς καλύτερη όταν υιοθετήθηκε η μεθοδολογία των Forni et al. (2000). Η μέθοδος αυτή φαίνεται επίσης να παρέχει μια πιο αποτελεσματική περίληψη των πληροφοριών που περιέχονται στα μεγάλα σύνολα δεδομένων, δεδομένου ότι απαιτούνται λιγότεροι παράγοντες για την επίτευξη της ίδιας πληροφορίας. Όμως, συνολικά, τα αποτελέσματα που προκύπτουν από τις μεθοδολογίες παραπάνω είναι μάλλον παρόμοια, με μέσες συσχετίσεις στην περιοχή 70%-80% της συνολικής πληροφορίας των μεταβλητών. Μία σημαντική διαφορά ωστόσο μεταξύ των δύο είναι πως ενώ η στατική χρησιμοποιεί συνήθως έναν αλγόριθμο μεγιστοποίησης της προσδοκίας για να αντιμετωπίσει τις περιπτώσεις ελλειπόντων δεδομένων, η δυναμική απαιτεί το σύνολο δεδομένων να είναι ισορροπημένο και χωρίς ακραίες παρατηρήσεις.

Αξιοσημείωτο είναι το πείραμα των Artis et al., (2001), οι οποίοι συνέλεξαν 81 μακροοικονομικές χρονοσειρές σε μηνιαία βάση, που παρέχουν μια εξαντλητική περιγραφή της

οικονομίας του Ηνωμένου Βασιλείου, τις απεικόνισαν με ένα μοντέλων παραγόντων και χρησιμοποίησαν τους εκτιμώμενους παράγοντες για την πρόβλεψη διαφόρων πραγματικών, ονομαστικών και χρηματοοικονομικών μεταβλητών, μέθοδος που εφαρμόστηκε για πρώτη φορά στα χρονικά στην οικονομία του Ηνωμένου Βασιλείου. Οι επιλεγμένες χρονοσειρές υπέστησαν επεξεργασία, μετασχηματίστηκαν, για να υπολογίσουν τις στοχαστικές ή τις ντετερμινιστικές τάσεις και χρησιμοποιήθηκαν λογάριθμοι για να μετατραπούν σε ρυθμούς μεταβολής για όσες δεν ήταν ήδη σε αυτή τη μορφή. Επιπλέον, περάστηκαν όλες οι σειρές μέσω μιας διαδικασίας εποχικής προσαρμογής, παρόλο που οι περισσότερες από αυτές αρχικά αναφέρθηκαν ως εποχικά διορθωμένες. Οι μετασχηματισμένες και εποχιακά διορθωμένες σειρές υποβλήθηκαν σε έλεγχο για μεγάλες ακραίες τιμές. Κάθε ακραία παρατήρηση αναδημιουργήθηκε ως κενή και δημιουργήθηκαν παράγοντες από αυτό το μη-ισορροπημένο πλαίσιο χρονοσειρών. Τα αποτελέσματα ήταν ενθαρρυντικά. Τα σημαντικότερα ήταν πρώτον πως με μόλις 4 παράγοντες εξηγήθηκε περίπου το 40% της μεταβλητότητας όλων των 81 μεταβλητών, ποσοστό που αυξήθηκε σε 50% με 6 παράγοντες και σε 68% με 12 και δεύτερον πως η μέθοδος των παραγόντων έφερε σχεδόν 20% καλύτερα αποτελέσματα από τις μέχρι τότε απλές μεθόδους χρονοσειρών.

Ο Iacoviello (2001) από την άλλη πλευρά, για την βραχυπρόθεσμη πρόβλεψη του Ιταλικού ΑΕΠ (ένα – δύο τρίμηνα μελλοντικά), χρησιμοποιεί «μοντέλο-γέφυρα» (Bridge Model - BM) ενώ για την μακροπρόθεσμη πρόβλεψη (ενός – δύο ετών) επιστρατεύει Μπεϋζιανά μοντέλα αυτοπαλινδρόμησης, γεγονός το οποίο συνιστά μια εντελώς διαφορετική προσέγγιση από τους προαναφερθέντες συναδέλφους του. Η προσέγγιση BM, εκμεταλλεύεται πρώιμα δημοσιεύσιμους κυκλικούς οικονομικούς δείκτες, οι οποίοι είναι διαθέσιμοι σε μηνιαία βάση συνήθως, για να παρέχει εκτιμήσεις πραγματικού χρόνου για το ΑΕΠ της Ιταλίας. Επί της ουσίας γεφυρώνει τους μηνιαίους δείκτες με το τριμηνιαίο ΑΕΠ. Εκτός των μηνιαίων δεικτών περιέχει και τριμηνιαίους, οι οποίοι είναι κυρίως Εθνικοί Λογαριασμοί. Όπως αναφέρεται μέσα στην δημοσίευση, τα πιο πρόσφατα μηνιαία δεδομένα κάθε περιόδου είναι απαραίτητα και κρίσιμα για την βραχυπρόθεσμη πρόβλεψη και μειώνουν αισθητά τα σφάλματα πρόβλεψης. Όσον αφορά τα μακροπρόθεσμα VAR μοντέλα, ο συγγραφέας προτίμησε να αυτοπαλινδρομίσει το ΑΕΠ και να το πλαισιώσει με μεταβλητές όπως η Εγχώρια Κατανάλωση και τα Επιτόκια Δανεισμού, κρατώντας τον αριθμό των μεταβλητών σε χαμηλό επίπεδο (6-8 μεταβλητών), καθώς όπως αναφέρει, πρώτον, λιγότερες μεταβλητές ελαχιστοποιούν την

πολυσυγγραμμικότητα του μοντέλου και δεύτερον οι ίδιες χρονικές υστερήσεις του ΑΕΠ, συγκεντρώνουν περισσότερη πληροφορία από ότι οι υπόλοιπες ερμηνευτικές μεταβλητές.

Στην ίδια μεθοδολογία κινούνται και οι Baffigi et al. (2004), αλλά επιχειρούν να προβλέψουν το συνολικό ΑΕΠ της Ευρωζώνης. Σκοπός της δημοσίευσής τους, είναι να εξετάσουν την ικανότητα πρόβλεψης των μοντέλων γέφυρας (BM) για την αύξηση του ΑΕΠ στη ζώνη του ευρώ. Το BM «γεφυρώνει το χάσμα» μεταξύ του περιεχομένου πληροφοριών των έγκαιρα ενημερωμένων με νέες δημοσιεύσεις δεικτών και των καθυστερημένων (αλλά πιο ολοκληρωμένων) Εθνικών Λογαριασμών. Τελικά και αυτοί οδηγούνται στο συμπέρασμα πως τα BM είναι πιο αξιόπιστα από τα μοντέλα Εθνικών Λογαριασμών. Παρ' όλα αυτά, αξίζει να εξετάσουμε την προσέγγισή τους στο ζήτημα καθώς τα προβλεπτικά μοντέλα BM δημιουργούνται με χρήση δεδομένων Προσφοράς αποκλειστικά, ενώ τα υπόλοιπα μοντέλα με δεδομένα Ζήτησης. Επιπρόσθετα, γίνεται διαχωρισμός των εννοιών Now-Casting και Pure-Casting. Το Now-Casting σηματοδοτεί πρόβλεψη τρέχοντος τριμήνου (επόμενης χρονικής στιγμής με χρήση Δεικτών Οικονομικού Κλίματος, ενώ το Pure-Casting χρησιμοποιεί μόνο γνωστά παλαιά σκληρά δεδομένα (hard-data) της οικονομίας. Ένα άλλο χαρακτηριστικό που τονίζεται είναι ο έλεγχος μοναδιαίας ρίζας (Unit Root Testing) των χρονοσειρών, ο οποίος είναι στην ουσία έλεγχος στασιμότητας των χρονοσειρών και υποδηλώνει το πόσο απρόβλεπτη μπορεί να είναι μια χρονοσειρά. Οι Baffigi et al. (2004), επιλέγουν μόνο χρονοσειρές που έχουν περάσει επιτυχώς το Unit Root Testing, έλεγχο στασιμότητας, και ως εκ τούτου τα προβλεπτικά αποτελέσματα τους είναι καλύτερα. Φθάνουν στο συμπέρασμα και αυτοί τελικά ωστόσο πως τα BM δίνουν καλύτερα αποτελέσματα από τα μοντέλα Εθνικών Λογαριασμών (hard data models). Ενισχύοντας και προσμετρώντας την αξία των κυρίων συνιστωσών, και κατ' επέκταση των προβλεπτικών μοντέλων που τις χρησιμοποιούν, οι Breitung και Eickmeier (2005) στο εμπειρικό παράδειγμα τους εξηγούν πως η αξιοποίηση πολλών πληροφοριών μπορεί να οδηγήσει σε ακριβέστερες προβλέψεις και μακροοικονομικές αναλύσεις. Η χρήση πολλών μεταβλητών παραπέμπει περαιτέρω στην πρακτική της κεντρικής τράπεζας να "εξετάζει τα πάντα". Ένα δεύτερο πλεονέκτημα των μοντέλων παραγόντων είναι ότι μπορούν να εξαλειφθούν ιδιοσυγκρασιακές κινήσεις που ενδεχομένως περιλαμβάνουν σφάλμα μέτρησης και τοπικές διαταραχές (shocks). Αυτό δίνει ένα πιο αξιόπιστο μήνυμα στους υπεύθυνους χάραξης πολιτικής και τους εμποδίζει να αντιδράσουν στις ιδιοσυγκρασιακές κινήσεις. Επιπλέον, η εκτίμηση κοινών παραγόντων ή κοινών διαταραχών έχει ουσιαστικό ενδιαφέρον για ορισμένες εφαρμογές.

Ένα τρίτο σημαντικό πλεονέκτημα είναι ότι τα μοντέλα παραγόντων μπορούν να παραμείνουν μη-έχοντα γνώση (agnostic) σχετικά με τη δομή της οικονομίας και δεν χρειάζεται να βασίζονται σε υπερβολικά αυστηρές υποθέσεις, όπως συμβαίνει μερικές φορές στα μοντέλα. Αντιπροσωπεύουν επίσης ένα πλεονέκτημα έναντι των δομικών μοντέλων VAR, όπου ο ερευνητής πρέπει να λάβει θέση ως προς τις μεταβλητές ώστε να συμπεριλάβει εκείνες, που με τη σειρά τους, καθορίζουν το αποτέλεσμα πρόβλεψης και τον αριθμό των κραδασμών (shock). Ένα ακόμη σημαντικό στοιχείο, είναι πως εκτός από το κριτήριο Bai και Ng, ο αριθμός των παραγόντων μπορεί κάλλιστα να διαμορφωθεί λαμβάνοντας υπόψη το ποσοστό της διακύμανσης, και κατά συνέπεια της πληροφορίας, που εμπεριέχουν οι παράγοντες, μετρούμενο από τις ιδιοτιμές τους (eigenvalues).

Παρ'όλα τα παραπάνω πλεονεκτήματα των μοντέλων κοινών παραγόντων, δε μπορούμε να αγνοήσουμε και τα αρνητικά τους στοιχεία καθώς και ορισμένες δυσκολίες στην υλοποίηση και χρήση τους. Όπως αναφέρουν οι Giannone et al. (2005), είναι σημαντικό να εξετάσουμε αν η οριακή συνεισφορά ενός μπλοκ δημοσιευμένων δεδομένων οφείλεται στο κατά πόσο "έγκαιρο" είναι ή στην "ποιότητα" του σαν δεδομένο. Όπως διαπιστώνεται, η σειρά της δημοσίευσης των οικονομικών δεδομένων έχει σημασία και μπορεί να έχουμε μια κατάσταση όπου δεδομένα υψηλής ποιότητας, όπως το ΑΕΠ, δεν έχουν οριακές επιπτώσεις στο ίδιο το ΑΕΠ, δεδομένου ότι απελευθερώνονται με μεγάλη καθυστέρηση (τρεις μήνες). Συνεπώς, πρέπει να προσμετρηθεί και να αξιολογηθεί το αντίκτυπο που έχουν οι νέες δημοσιεύσεις δεδομένων για τις προβλέψεις μας ή το κατά πόσο αυτές αποτελούν θόρυβο για το μοντέλο μας. Σε παρόμοιο συμπέρασμα οδηγείται και ο Reijer (2005) στην προσπάθειά του να προβλέψει την κίνηση του Ολλανδικού ΑΕΠ. Έχοντας συλλέξει περίπου 370 τριμηνιαίες χρονοσειρές της Ολλανδικής οικονομίας, καταλήγει να εμπεριέχει ένα υποσύνολο αυτών, περίπου 150 για το σκοπό του. Η επίδοση των προβλέψεων των μοντέλων παραγόντων παρακωλύεται, επειδή το πλήρες σύνολο δεδομένων είναι υπερδειγματοληπτούμενο ή / και διαθέτει πάρα πολύ θορυβώδη δεδομένα που υπό κανονικές συνθήκες στη θεωρία, τα μοντέλα δυναμικών παραγόντων υποβαθμίζουν τα θορυβώδη δεδομένα και μειώνουν την σειριακή συσχέτιση συγχρονίζοντας τις χρονοσειρές. Επομένως πρέπει να ταιριάζουν ευκολότερα μέσα σε μία δομή παραγόντων.

Μία αντίστοιχη μάλιστα προσέγγιση ακολουθείται και από τους Banerjeea και Marcellino (2006), οι οποίοι προσπαθώντας να μειώσουν τον θόρυβο των δεδομένων τους,

εντοπίζουν, μέσω κατάλληλου λογισμικού, τους ισχυρότερους δείκτες της οικονομίας και απορρίπτουν τους υπόλοιπους. Το καταφέρνουν, μέσω της κοινής εφαρμογής κριτηρίων πληροφόρησης, δοκιμών-τεστ σημαντικότητας των χρονοσειρών-δεδομένων και δοκιμών καταλοίπων για τη σωστή κατασκευή και παραμετροποίηση του προβλεπτικού μοντέλου. Καταλήγουν δε στο συμπέρασμα – αποτέλεσμα πως η ομαδοποίηση μεταβλητών με κοινά χαρακτηριστικά και η χρήση τους σαν ομάδες υπερβαίνει σε προβλεπτική ικανότητα ακόμα και τις κύριες συνιστώσες του συνόλου των μεταβλητών.

Όσον αφορά τους βοηθητικούς και στην πλειοψηφία των εμπειρικών αναλύσεων απαραίτητους μετασχηματισμούς δεδομένων πριν τη δημιουργία των κοινών παραγόντων, οι Dahla et al. (2009) τους συνοψίζουν εξαιρετικά στη μέθοδό τους. Εάν υποθέσουμε πως κάθε χρονοσειρά μπορεί να παρασταθεί ως $x_{i,t} = g_{it} + c_{i,t} + s_{i,t} + e_{i,t}$ (i ο αριθμός της σειράς και t η μονάδα του χρόνου) στην οποία εξίσωση το g_{it} υποδηλώνει μια συνιστώσα της τάσης της χρονοσειράς, το $c_{i,t}$ την συνιστώσα του επιχειρηματικού κύκλου, το $s_{i,t}$ μια εποχική συνιστώσα και το $e_{i,t}$ το σφάλμα-ιδιοσυγκρασιακό παράγοντα, είναι προς το συμφέρον του πειράματος να αφαιρείται το εποχικό στοιχείο συχνά με τη χρήση καταλλήλου φίλτρου, να απομακρύνεται η συνιστώσα της τάσης μέσω εφαρμογής πρώτων-log διαφορών και να γίνεται ενός τύπου διαλογή των τιμών της χρονοσειράς πάνω από ένα κατώφλι (threshold) μέσω της αφαίρεσης των αντίστοιχων παρατηρήσεων.

Μέχρι στιγμής, οι προαναφερθέντες ερευνητές εστίασαν κυρίως στο μεγάλο πλήθος των χρονοσειρών-μεταβλητών και τη συνεισφορά τους στην εξαγωγή ικανοποιητικών αποτελεσμάτων στις προβλέψεις, ανεξαρτήτως μεθόδου. Οι Boivina και Ng (2005), έρχονται να διαταράξουν αυτή τη συνοχή, διαπιστώνοντας ότι λιγότερες μεταβλητές οδηγούν σε πιο αξιόπιστα αποτελέσματα. Εν μέρει, αυτό οφείλεται στο φορτίο υπολογισμού που εμπλέκεται σε μεγάλα μοντέλα και αυτό κυρίως συμβαίνει επειδή όχι όλες οι διαθέσιμες σειρές μπορούν να είναι ενημερωτικές, ούτως ώστε συμπεριλαμβάνοντας άσχετες πληροφορίες στα μοντέλα τα επιβαρύνουμε με κόστος και υπερδειγματοληψία (oversampling). Στην πράξη, τα δεδομένα προέρχονται συνήθως από μικρό αριθμό ευρέων κατηγοριών (όπως η βιομηχανική παραγωγή, οι τιμές, τα επιτόκια). Η ταξινόμηση των σειρών αναλόγως της σημασίας τους για μια κύρια συνιστώσα και η σύνθεση ενός συνόλου δεδομένων που αποτελείται από σειρές με υψηλή κατάταξη από κάθε κατηγορία, πραγματικά έχει επίδραση. Εάν σε διαφορετική περίπτωση,

επεκτείνουμε αυτό το σύνολο δεδομένων προσθέτοντας σειρές χαμηλότερης πληροφορίας ή «θορυβώδη» δύο πράγματα θα συμβούν. Το μέσο μέγεθος των κυρίων συνιστωσών θα μειωθεί καθώς προστίθενται περισσότερες σειρές, και η πιθανότητα συσχετισμένων σφαλμάτων θα αυξηθεί καθώς συμπεριλαμβάνονται περισσότερες σειρές από την ίδια κατηγορία. Συνεπώς, ο αριθμός των παραγόντων μπορεί να εκτιμηθεί με ακρίβεια με το N (πλήθος παραγόντων) τόσο μικρό όσο το 40 όταν τα σφάλματα είναι ανεξάρτητα και κανονικώς κατανομημένα. Αυτό υποδηλώνει ότι το N δεν χρειάζεται να είναι εξαιρετικά μεγάλο για τον κύριο εκτιμητή των συστατικών για να δώσει εύλογα ακριβείς εκτιμήσεις. Στις προσομοιώσεις και στα εμπειρικά παραδείγματα που εξέτασαν, διαπίστωσαν πως οι παράγοντες που εξάγονται από 40 σειρές φαίνεται να μην επιδεινώνονται και σε πολλές περιπτώσεις είναι καλύτεροι από αυτούς που εξάγονται από 147 χρονοσειρές. Ωστόσο, δεν υπάρχει προς το παρόν κανένας οδηγός για τα δεδομένα που πρέπει να χρησιμοποιούνται στην ανάλυση παραγόντων. Την προ-επιλογή χρονοσειρών-μεταβλητών υποστηρίζει και ο Schumacher (2010) για την ανάλυσή του για το ΑΕΠ της Γερμανίας, όπου κάνει χρήση μιας υπο-ομάδας του συνόλου των μεταβλητών του με σκοπό να μειώσει κοινή πληροφορία (λόγω πολυσυγγραμμικότητας) που ενδεχομένως παρέχεται από δύο ή περισσότερες χρονοσειρές. Επιπρόσθετα, υποστηρίζει πως ενώ συχνά επιλέγονται οικονομικές μεταβλητές που αφορούν μία χώρα αποκλειστικά, μπορεί να εξαχθεί σημαντική πληροφορία και από διεθνείς οικονομικές μεταβλητές (για παράδειγμα η τιμή του πετρελαίου) ή άλλων χωρών (για παράδειγμα το ΑΕΠ των Ηνωμένων Πολιτειών της Αμερικής). Αποδεικνύει δε εμπειρικά τους παραπάνω ισχυρισμούς στο πείραμά του.

Εκτός από την προεπιλογή δεδομένων, έχει δοκιμαστεί περαιτέρω η χρήση βαρών στους κοινούς παράγοντες της PCA μεθόδου. Ο Stakenas (2012), αντιβαίνοντας στην επιλογή μεγάλου πλήθους μεταβλητών-χρονοσειρών, τονίζει πως είναι προφανές ότι ορισμένες μεταβλητές είναι πιο σημαντικές για την πρόβλεψη του ΑΕΠ από άλλες: π.χ. η τριμηνιαία μεταβλητή λιανικών πωλήσεων έχει συντελεστή συσχέτισης με το ΑΕΠ της Λιθουανίας ίσο με 0,72, ενώ ο συσχετισμός του δείκτη τιμών παραγωγού με το ΑΕΠ είναι μόνο 0,08. Έτσι, οι υψηλά συσχετιζόμενες μεταβλητές είναι πιθανό να προσδίδουν μεγαλύτερη πληροφορία για την πορεία του ΑΕΠ. Επίσης, πρέπει να σημειωθεί ότι η χρησιμότητα μιας συγκεκριμένης μεταβλητής αλλάζει σε διαφορετικούς ορίζοντες πρόβλεψης. Συνεπώς, προσδίδοντας ειδικό βάρος σε ορισμένες μεταβλητές, αλλάζει άρδην την ποιότητα της πρόβλεψης. Η επίδραση της τροποποίησης του μοντέλου παραγόντων με τη στάθμιση των μεταβλητών, ανάλογα με την

απόλυτη συσχέτισή τους με την αύξηση του ΑΕΠ, έφερε αποτελέσματα τα οποία δείχνουν ότι το μοντέλο weighted-PC θα μπορούσε να δώσει καλύτερες προβλέψεις για το τρέχον τρίμηνο. Από την άλλη πλευρά, το WPC έχει αρκετά μειονεκτήματα: χρησιμοποιεί μια υποκειμενική συνάρτηση βαρών, παράγει λιγότερους ερμηνευτικούς παράγοντες και κάνει τα αποτελέσματα πιο ευαίσθητα στις αλλαγές σε ένα σετ μεταβλητών.

Τα τελευταία χρόνια, η ανάπτυξη οικονομετρικών μοντέλων που βασίζονται σε δεδομένα μικτής συχνότητας έχει προσελκύσει σημαντική προσοχή. Συγκεκριμένα, η προσέγγιση δειγματοληψίας ανάμεικτων δεδομένων (MIDAS) που προτάθηκε από τους Ghysels et al. (2004) αποδείχθηκε χρήσιμη για διάφορους διαφορετικούς σκοπούς πρόβλεψης. Το MIDAS μπορεί να θεωρηθεί ως εργαλείο παλινδρόμησης χρονοσειρών-μεταβλητών που επιτρέπει τη δειγματοληψία των εξαρτημένων μεταβλητών σε μια παλινδρόμηση και των ερμηνευτικών μετεβλητών της παλινδρόμησης σε διαφορετικές συχνότητες και όπου χρησιμοποιούνται χρησιμοποιούνται πολυώνυμα διανεμημένης καθυστέρησης για να εξασφαλιστεί η, όσον το δυνατόν, απλότητα του μοντέλου πρόβλεψης. Οι Kuzina et al. (2011) προβήκανε σε μια εμπειρική χρήση του MIDAS για την πρόβλεψη τρέχοντος τριμήνου του ΑΕΠ της Ευρωζώνης, συγκρίνοντας το με κλασικά VAR μοντέλα (benchmark models) και τα αποτελέσματα ήταν διαφορετικά ως προς την ερμηνεία τους. Οι σχετικές επιδόσεις των MIDAS (Mixed-data sampling models) και MF-VAR (mixed frequency vector autoregressive models) διαφέρουν ανάλογα με τις ερμηνευτικές μεταβλητές και τους ορίζοντες πρόβλεψης και δεν φαίνεται να υπάρχει σαφής νικητής όσον αφορά την πρόβλεψη των επιδόσεων. Συνολικά, τα MF-VAR και MIDAS φαίνεται να είναι ισάξια σε μακροοικονομικά σύνολα δεδομένων, με το MIDAS να υπερτερεί σε βραχυπρόθεσμο ορίζοντα (μικρότερα σφάλματα) ενώ χάνει μακροπρόθεσμα. Ένα μειονέκτημα του όμως είναι πως συχνά οδηγεί σε υπερπληροφόρηση και χάνει από πολυσυγγραμικότητα.

Το σύνολο των παραπάνω μεθόδων που παρουσιάστηκαν διατρέχουν κινδύνους και επηρεάζονται από σφάλματα. Στην δημοσίευσή της, η Diron (2006) συνοψίζει κατατοπιστικά τους τύπους σφαλμάτων που μπορεί να προκύψουν στις προβλέψεις και τους ομαδοποιεί σχετικά με την επίδρασή τους στην πρόβλεψη του μακροοικονομικού μεγέθους ΑΕΠ. Πρώτον, υπάρχουν σφάλματα λόγω διαφορών μεταξύ της αύξησης του πραγματικού ΑΕΠ και των εκτιμώμενων σχέσεων με μηνιαίες μεταβλητές. Δεύτερον, σφάλματα λόγω εσφαλμένων υποθέσεων σχετικά

με τους μηνιαίους δείκτες για τους μήνες-παρατηρήσεις που λείπουν κατά την περίοδο πρόβλεψης. Τρίτον, σφάλματα λόγω αναθεωρήσεων των μηνιαίων δεικτών με την πάροδο του χρόνου και τέταρτον λάθη λόγω της αναθεώρησης του ΑΕΠ μετά το flash-estimate. Η διάκριση μεταξύ αυτών των πηγών σφαλμάτων παρέχει καθοδήγηση στους επαγγελματίες σχετικά με το πού πρέπει να επικεντρωθεί η περαιτέρω προσπάθεια βελτίωσης της συνολικής αξιοπιστίας των βραχυπρόθεσμων προβλέψεων. Καταλήγει δε η συγγραφέας να αξιολογεί και τον βαθμό επίδρασης των σφαλμάτων κατά τις προβλέψεις αναλύοντας τις συνεισφορές τους στα γενικά σφάλματα πρόβλεψης. Η κύρια πηγή σφάλματος είναι τα σφάλματα προδιαγραφών, δηλαδή για παράδειγμα αναντιστοιχία μεταξύ της αύξησης του πραγματικού ΑΕΠ και της αύξησης των χρησιμοποιούμενων δεικτών της οικονομίας. Γενικά, τα σφάλματα προδιαγραφής διαπιστώνονται ότι είναι σημαντικά, αν και κάπως μικρότερα για τα μοντέλα που χρησιμοποιούν «σκληρά δεδομένα» παρά για τα μοντέλα που χρησιμοποιούν δεδομένα ερευνών ή τρίτες οικονομικές μεταβλητές. Στην συνέχεια έρχονται τα σφάλματα που προκύπτουν από τις εκτιμήσεις (των παρατηρήσεων που λείπουν) των δεικτών, τα οποία είναι επίσης σημαντικά όταν οι δείκτες πρέπει να υποβληθούν σε εκτιμήσεις πέντε ή έξι μήνες μπροστά, αλλά μειώνονται ταχύτατα καθώς διατίθενται περισσότερα δεδομένα και αντικαθιστούν τις εκτιμώμενες τιμές τους. Εν τω μεταξύ, οι αναθεωρήσεις των μηνιαίων μεταβλητών και της αύξησης του ΑΕΠ αντιπροσωπεύουν μόνο ένα μικρό ποσοστό των συνολικών σφαλμάτων, αποτέλεσμα το οποίο ποικίλλει ιδίως όταν εξετάζονται μεμονωμένες/ιδιαιτερες περιπτώσεις. Εν τέλει, φαίνεται να υπάρχει αντιστάθμιση μεταξύ καλύτερων προδιαγραφών που βασίζονται σε σκληρά δεδομένα και μικρότερα σφάλματα λόγω αναθεωρήσεων και ελλειπουσών τιμών που λαμβάνονται όταν οι προβλέψεις προέρχονται από έρευνες και οικονομικές μεταβλητές.

3.Πρόβλεψη Ελληνικού ΑΕΠ

Όπως αναλύθηκε στο Κεφάλαιο 2, έχουν χρησιμοποιηθεί αρκετές διαφορετικές στατιστικές μέθοδοι που υποβοηθούν την πρόβλεψη μακρο-οικονομικών μεγεθών και ειδικότερα του ΑΕΠ. Οι πιο διαδεδομένες είναι η ανάλυση κυρίων συνιστωσών PCA και οι παραλλαγές της (π.χ. weighted-PCA), η μοντελοποίηση δυναμικών παραγόντων DFM, η χρήση μοντέλων δειγματοληψίας ανάμεικτων δεδομένων MIDAS, η πολυπαραγοντική αυτοπαλινδρόμηση διανυσμάτων multivariate VAR και τέλος τα κλασικά μοντέλα αυτοπαλινδρόμησης. Το σύνολο των επιλογών επαυξάνεται και πολλαπλασιάζεται αν αναλογιστεί κανείς ότι πρέπει να βρεθεί το κατάλληλο υπο-σύνολο των μακροοικονομικών μεταβλητών που θα τεθούν υπό έλεγχο σύμφωνα με την φύση τους και τις ανάγκες των εκάστοτε ερευνητών, να ορισθεί το πλήθος των χρονοσειρών – ερμηνευτικών μεταβλητών, να γίνει ενδεχομένως προεπιλογή τους βάσει χαρακτηριστικών, να υποστούν πρότερη επεξεργασία, να αντιμετωπισθεί το πρόβλημα των ελλειπών τιμών – καθυστερήσεων, και να γίνουν οι απαραίτητες παραδοχές για την συνέχεια του πειράματος. Παρ' όλα αυτά, δυστυχώς δεν έχει βρεθεί καταλληλότερο μοντέλο ακόμη, και η χρήση ενός εκ των προαναφερθέντων γίνεται κατά περίπτωση χώρας, χρονικής περιόδου και γενικότερων οικονομικών συνθηκών. Επιπλέον, συχνά είναι και αντικρουόμενες μεταξύ τους, σύμφωνα με εμπειρικά αποτελέσματα όσον αφορά την καταλληλότητά τους βραχυπρόθεσμα ή μακροπρόθεσμα. Για την Ελλάδα αποκλειστικά έχουν δοκιμασθεί ελάχιστα μοντέλα με πιο σημαντική ίσως παρέμβαση την έρευνα των Monokroussos και Thomakos (2012) οι οποίοι κάνουν χρήση του MIDAS για την πρόβλεψη του ΑΕΠ τρέχοντος τριμήνου.

Στην παρούσα διπλωματική εργασία γίνεται χρήση της ανάλυσης κυρίων συνιστωσών, PCA, για την πρόβλεψη του ΑΕΠ τρέχοντος τριμήνου. Συλλέχθηκε πληθώρα μηνιαίων δεδομένων - χρονοσειρών, κατά δύναμη και διαθεσιμότητα, από έγκυρους φορείς και θεσμούς. Ακολουθήθηκε κατάλληλος μετασχηματισμός των δεδομένων και υπολογίστηκαν ελλείψεις παρατηρήσεων με χρήση αυτοπαλινδρόμησης. Στη συνέχεια εκτελέστηκαν έλεγχοι στασιμότητας των μετασχηματισμένων πλέον χρονοσειρών και επιλέχθηκαν δύο προσεγγίσεις για την εξαγωγή μοντέλων και την current-quarter πρόβλεψη του ΑΕΠ. Η διαδικασία εκτελέστηκε αναδρομικά για περίοδο οκταετίας σε μηνιαία βάση και έγινε εξαγωγή

αποτελεσμάτων και μέσω των ασφαλών αναλόγων της περιοδικότητας των μηνών της πρόβλεψης.

3.1.Περιβάλλον υλοποίησης

Η παρούσα διπλωματική εργασία εκπονήθηκε σε περιβάλλον λειτουργικού συστήματος Microsoft Windows 7 Ultimate, με updates έως και Service Pack 1. Η γλώσσα προγραμματισμού που χρησιμοποιήθηκε είναι η Python, η οποία είναι δωρεάν στη χρήση (open software). Τέλος το πλήθος των δεδομένων που βρέθηκαν τα κατατάσσουν στην κατηγορία Big Data οπότε συνίσταται ανάλογη διαχείριση.

Η Python είναι μια ερμηνευμένη γλώσσα προγραμματισμού υψηλού επιπέδου για γενικό προγραμματισμό. Δημιουργήθηκε από τον Guido van Rossum και κυκλοφόρησε για πρώτη φορά το 1991. Η Python έχει μια φιλοσοφία σχεδίασης που δίνει έμφαση στην αναγνωσιμότητα του κώδικα, χρησιμοποιώντας κενά διαστήματα. Παρέχει δομές που καθιστούν εφικτό τον σαφή προγραμματισμό σε μικρές και μεγάλες κλίμακες. Η Python διαθέτει σύστημα δυναμικού τύπου και αυτόματη διαχείριση μνήμης. Υποστηρίζει πολλαπλά παραδείγματα προγραμματισμού, συμπεριλαμβανομένου του αντικειμενοστρεφούς, επιτακτικού, λειτουργικού και διαδικαστικού, και έχει μια μεγάλη και ολοκληρωμένη τυποποιημένη βιβλιοθήκη (<https://www.python.org/>). Η έκδοση που χρησιμοποιήθηκε είναι η 3.6.3.

Επιπλέον ο προγραμματισμός έγινε στο ολοκληρωμένο περιβάλλον ανάπτυξης ανοικτού κώδικα Anaconda Navigator (έκδοση 1.6.9.) και στο επιστημονικό περιβάλλον ανάπτυξης Spyder (έκδοση 3.2.4). Το Anaconda είναι μια ελεύθερη και ανοιχτή διανομή των γλωσσών προγραμματισμού Python και R για εφαρμογές σχετικές με την επιστήμη των δεδομένων και την μηχανική μάθηση (επεξεργασία δεδομένων μεγάλης κλίμακας, προγνωστική ανάλυση, επιστημονική πληροφορική) με στόχο την απλοποίηση της διαχείρισης και της ανάπτυξης των πακέτων κώδικα (<https://anaconda.org/>). Το Spyder είναι ένα ισχυρό περιβάλλον διαδραστικής ανάπτυξης για τη γλώσσα Python με προηγμένα χαρακτηριστικά επεξεργασίας, διαδραστικού ελέγχου, εντοπισμού ασφαλών και ενδοσκόπησης και ένα αριθμητικό περιβάλλον

πληροφορικής χάρη στην υποστήριξη του IPython (διαδραστικός διερμηνέας της Python) και τις δημοφιλείς βιβλιοθήκες της Python (<https://pythonhosted.org/spyder/>).

Οι βιβλιοθήκες της Python που ήταν απαραίτητες για το πείραμα ήταν οι εξής :

- Scikit-learn (<http://scikit-learn.org/stable/>): Μια βιβλιοθήκη εκμάθησης μηχανών ελεύθερου λογισμικού για τη γλώσσα προγραμματισμού Python. Έγινε χρήση της για τα σφάλματα των μοντέλων, για τις γραμμικές παλινδρομήσεις και για την εφαρμογή της PCA.
- Arch (<http://arch.readthedocs.io/en/latest/unitroot/tests.html>): Χρησιμοποιήθηκε για έλεγχο μοναδιαίας ρίζας (Unit Root Testing) – έλεγχο στασιμότητας.
- Numpy (<http://www.numpy.org>): Ένα θεμελιώδες πακέτο για την επιστημονική πληροφορική. Έγινε χρήση του για διαχείριση των χρονοσειρών.
- Pandas (<https://pandas.pydata.org/>): Μία βιβλιοθήκη ανάλυσης δεδομένων της Python. Έγινε χρήση του για διαχείριση των χρονοσειρών.
- Statsmodels (<https://www.statsmodels.org>): Είναι ένα Python module που παρέχει κλάσεις και λειτουργίες για την εκτίμηση πολλών διαφορετικών στατιστικών μοντέλων, καθώς και για διεξαγωγή στατιστικών δοκιμών και διερεύνηση στατιστικών δεδομένων. Χρησιμοποιήθηκε για εποχική αποσύνθεση των χρονοσειρών και ARMA μοντέλο.
- Matplotlib (<https://matplotlib.org/>): Το Matplotlib είναι μια βιβλιοθήκη σχεδίασης Python 2D, η οποία παράγει γραφήματα σε επίπεδο ποιότητας δημοσίευσης σε διάφορες μορφές έντυπου και διαδραστικά περιβάλλοντα σε διάφορες πλατφόρμες. Έγινε χρήση του για οπτικοποίηση του πειράματος.

3.2.Δεδομένα

Οι χρονοσειρές - μεταβλητές που χρησιμοποιήθηκαν για την πρόβλεψη του Ελληνικού ΑΕΠ είναι στο σύνολό τους εκατόν εβδομήντα πέντε (175). Επιπλέον είναι σε μηνιαία περιοδικότητα χρόνου και οι παρατηρήσεις ξεκινούν από τον Ιανουάριο του έτους 2004 έως και τον Σεπτέμβριο του έτους 2017 [01:2004-09:2017], προσδίδοντας έτσι κατά μέγιστο εκατόν εξήντα έξι (166) παρατηρήσεις η κάθε μία. Έχουν συλλεχθεί από έγκυρες ιστοσελίδες φορέων και θεσμών όπως η Τράπεζα της Ελλάδος (ΤτΕ), η μακροοικονομική βάση δεδομένων AMECO της Ευρωπαϊκής Επιτροπής (ΕΚ) της Ευρωπαϊκής Ένωσης (ΕΕ), το Ελληνικό Ίδρυμα Οικονομικών και Βιομηχανικών Ερευνών (ΙΟΒΕ), η αποθήκη στατιστικών δεδομένων της Ευρωπαϊκής Κεντρικής Τράπεζας (ΕΚΤ), η Ελληνική Στατιστική Αρχή (ΕΛ.ΣΤΑΤ) και ο Οργανισμός Οικονομικής Συνεργασίας και Ανάπτυξης (<http://www.oecd.org>). (Πίνακας 01)

Η Τράπεζα της Ελλάδος είναι η Κεντρική Τράπεζα της χώρας. Ιδρύθηκε το 1927 βάσει ενός Παραρτήματος του Πρωτοκόλλου της Γενεύης και άρχισε να λειτουργεί τον Μάιο του 1928. Έχει συσταθεί με τη μορφή ανωνύμου εταιρίας. Ως έδρα της ορίζεται από το Καταστατικό της η Αθήνα, ενώ διατηρεί 17 Υποκαταστήματα, 32 Πρακτορεία και 5 Θυρίδες σε όλη την Ελλάδα. Από τον Ιανουάριο 2001 η Τράπεζα της Ελλάδος αποτελεί αναπόσπαστο μέλος του Ευρωσυστήματος, που απαρτίζεται από την Ευρωπαϊκή Κεντρική Τράπεζα (ΕΚΤ) και τις Εθνικές Κεντρικές Τράπεζες των κρατών-μελών της Ευρωπαϊκής Ένωσης (ΕΕ) που ανήκουν στη ζώνη του ευρώ (<https://www.bankofgreece.gr>).

Η AMECO είναι η ετήσια μακροοικονομική βάση δεδομένων της Γενικής Διεύθυνσης Οικονομικών και Δημοσιονομικών Υποθέσεων της Ευρωπαϊκής Επιτροπής. Η βάση δεδομένων αναφέρεται συχνά στις δημοσιεύσεις της Επιτροπής και είναι απαραίτητη για τις αναλύσεις και τις εκθέσεις της Γενικής Διεύθυνσης. Η βάση δεδομένων περιέχει στοιχεία για την ΕΕ-28, τη ζώνη του ευρώ, τα κράτη μέλη της ΕΕ, τις υποψήφιες χώρες και άλλες χώρες του ΟΟΣΑ (Ηνωμένες Πολιτείες, Ιαπωνία, Καναδά, Ελβετία, Νορβηγία, Ισλανδία, Μεξικό, Κορέα, Αυστραλία και Νέα Ζηλανδία). Η Ευρωπαϊκή Επιτροπή (ΕΚ) είναι θεσμός της Ευρωπαϊκής Ένωσης, υπεύθυνος για την πρόταση νομοθεσίας, την εφαρμογή αποφάσεων, την τήρηση των συνθηκών της ΕΕ και τη διαχείριση των καθημερινών δραστηριοτήτων της ΕΕ (<https://ec.europa.eu/>).

Το IOBE είναι ιδιωτικός, μη κερδοσκοπικός, κοινωφελής, ερευνητικός οργανισμός. Ιδρύθηκε το 1975 με δύο σκοπούς: αφενός να προωθεί την επιστημονική έρευνα για τα τρέχοντα και αναδυόμενα προβλήματα της ελληνικής οικονομίας και αφετέρου να παρέχει αντικειμενική πληροφόρηση και να διατυπώνει προτάσεις, οι οποίες είναι χρήσιμες στη διαμόρφωση πολιτικής. Θεωρείται ότι έχει επικοινωνία με παγκόσμιους κύκλους εμπορικής και οικονομικής σημασίας (<http://iobe.gr/>).

Η ΕΚΤ είναι ένα από τα επίσημα θεσμικά όργανα της Ευρωπαϊκής Ένωσης και βρίσκεται στον πυρήνα του Ευρωσυστήματος και του Ενιαίου Εποπτικού Μηχανισμού. Από την 1η Ιανουαρίου 1999 η Ευρωπαϊκή Κεντρική Τράπεζα (ΕΚΤ) είναι υπεύθυνη για την άσκηση της νομισματικής πολιτικής στη ζώνη του ευρώ – τη μεγαλύτερη οικονομία στον κόσμο μετά τις Ηνωμένες Πολιτείες (<https://www.ecb.europa.eu/home/html/index.en.html>).

Η ΕΛΣΤΑΤ έχει σκοπό τη συστηματική παραγωγή επίσημων στατιστικών, καθώς και τη διενέργεια επιστημονικών ερευνών και την κατάρτιση μελετών, οι οποίες αφορούν όλους τους τομείς της δραστηριότητας του δημοσίου και του ιδιωτικού τομέα, υποστηρίζουν τη διαδικασία λήψης αποφάσεων, χάραξης και αξιολόγησης πολιτικών της Κυβέρνησης και των φορέων του Δημοσίου (δείκτες αξιολόγησης), υποβάλλονται σε διεθνείς φορείς σύμφωνα με τις υποχρεώσεις της χώρας και, αφορούν στο γενικό κοινό ή κατηγορίες χρηστών στατιστικών στοιχείων στο εσωτερικό ή στο εξωτερικό (<http://www.statistics.gr/>).

Αποστολή του Οργανισμού Οικονομικής Συνεργασίας και Ανάπτυξης (ΟΟΣΑ) είναι να προωθήσει πολιτικές που θα βελτιώσουν την οικονομική και κοινωνική ευημερία των ανθρώπων σε όλο τον κόσμο. Ο ΟΟΣΑ παρέχει ένα φόρουμ στο οποίο οι κυβερνήσεις μπορούν να συνεργαστούν για να μοιραστούν εμπειρίες και να αναζητήσουν λύσεις σε κοινά προβλήματα (<http://www.oecd.org>).

Η συλλογή των δεδομένων ολοκληρώθηκε στις 15 Οκτωβρίου 2017. Πρέπει να σημειωθεί πως η αξιοποίησή τους στο πείραμα γίνεται ψευδο-πραγματικά (pseudo-real) καθώς κανείς δεν μπορεί να εγγυηθεί πως στο σύνολό τους θα έχουν σταθερές ημερομηνίες δημοσίευσης. Για αυτό τον σκοπό, χρησιμοποιήθηκε στο πείραμα ένας πίνακας χρονικών υστερήσεων για να υποβοηθήσει την αναδρομική εκτέλεση του προγράμματος. Το σύνολο των μεταβλητών εσώκλειαν εποχικότητα εκτός από το ποσοστό ανεργίας και τους δείκτες βιομηχανικής παραγωγής. Επιπρόσθετα, θεώρησα δόκιμο να συλλέξω όσο το δυνατόν περισσότερες μακροοικονομικές μεταβλητές ακολουθώντας το παράδειγμα των Favero et al. (2002), όπως και των Artis et al., (2001) για την συνέχεια της μελέτης. Τέλος, οι παλαιότερες τιμές του ΑΕΠ σε τριμηνιαία βάση βρέθηκαν από το site της ΕΛΣΤΑΤ. Οι μεταβλητές καθώς και οι χρονικές τους υστερήσεις φαίνονται αναλυτικά στον Πίνακα 1.

Όνομα	Πηγή	Χρονική υστέρηση σε μήνες
Turnover Index in Retail Trade (overall index)	ΕΛΣΤΑΤ	3
Turnover Index in Retail Trade (overall index except automotive fuel)	ΕΛΣΤΑΤ	3
Turnover Index in Retail Trade (food sector (supermarkets, food, beverages, tobacco))	ΕΛΣΤΑΤ	3
Turnover Index in Retail Trade (overall index except food sector and automotive fuel)	ΕΛΣΤΑΤ	3
Turnover Index in Retail Trade (super markets)	ΕΛΣΤΑΤ	3
Turnover Index in Retail Trade (department stores)	ΕΛΣΤΑΤ	3
Turnover Index in Retail Trade (automotive fuel)	ΕΛΣΤΑΤ	3
Turnover Index in Retail Trade (food, beverages, tobacco)	ΕΛΣΤΑΤ	3
Turnover Index in Retail Trade (pharmaceutical products, cosmetics)	ΕΛΣΤΑΤ	3
Turnover Index in Retail Trade (clothing and footwear)	ΕΛΣΤΑΤ	3
Turnover Index in Retail Trade (furniture, electrical, equipment, household equipment)	ΕΛΣΤΑΤ	3
Turnover Index in Retail Trade (books stationery, other goods)	ΕΛΣΤΑΤ	3
Turnover Index in Retail Trade (retail sale not in stores)	ΕΛΣΤΑΤ	3
Volume Index in Retail Trade (overall index)	ΕΛΣΤΑΤ	3
Volume Index in Retail Trade (overall index except automotive fuel)	ΕΛΣΤΑΤ	3
Volume Index in Retail Trade (food sector (supermarkets, food, beverages, tobacco))	ΕΛΣΤΑΤ	3
Volume Index in Retail Trade (overall index except food sector and automotive fuel)	ΕΛΣΤΑΤ	3
Volume Index in Retail Trade (super markets)	ΕΛΣΤΑΤ	3
Volume Index in Retail Trade (department stores)	ΕΛΣΤΑΤ	3
Volume Index in Retail Trade (automotive fuel)	ΕΛΣΤΑΤ	3
Volume Index in Retail Trade (food, beverages, tobacco)	ΕΛΣΤΑΤ	3
Volume Index in Retail Trade (pharmaceutical products, cosmetics)	ΕΛΣΤΑΤ	3
Volume Index in Retail Trade (clothing and footwear)	ΕΛΣΤΑΤ	3
Volume Index in Retail Trade (furniture, electrical, equipment, household equipment)	ΕΛΣΤΑΤ	3
Volume Index in Retail Trade (books stationery, other goods)	ΕΛΣΤΑΤ	3
Producer Price Indices (PPI) (OECD)	ΟΟΣΑ	2
Producer Price Index in Industry (PPI) (overall index)	ΟΟΣΑ	2
Producer Price Index in Industry (PPI) (Domestic Market Index)	ΟΟΣΑ	2

Producer Price Index in Industry (PPI) (Foreign Market Index)	ΟΟΣΑ	2
Producer Price Index in Industry (PPI) for Manufacturing (Manufacturing Total Market)	ΟΟΣΑ	2
Producer Price Index in Industry (PPI) for Manufacturing (Domestic Market Manufacturing)	ΟΟΣΑ	2
Producer Price Index in Industry (PPI) for Manufacturing (Foreign Market Manufacturing)	ΟΟΣΑ	2
Import Price Index in Industry (MPI) (Total Import Index)	ΕΛΣΤΑΤ	2
Import Price Index in Industry (MPI) for Manufacturing (Total Manufacturing Imports)	ΕΛΣΤΑΤ	2
Output Price Index in Agriculture and Livestock (API) (Agricultural goods output)	ΕΛΣΤΑΤ	2
Output Price Index in Agriculture and Livestock (API) (Crop output)	ΕΛΣΤΑΤ	2
Output Price Index in Agriculture and Livestock (API) (Animal output)	ΕΛΣΤΑΤ	2
Material Cost Index for the Construction of New Residential Buildings (Overall Index)	ΕΛΣΤΑΤ	2
Material Cost Index for the Construction of New Residential Buildings (Cement, mortars and ready mixed concrete)	ΕΛΣΤΑΤ	2
Material Cost Index for the Construction of New Residential Buildings (Natural stone)	ΕΛΣΤΑΤ	2
Material Cost Index for the Construction of New Residential Buildings (Marble products, granites)	ΕΛΣΤΑΤ	2
Material Cost Index for the Construction of New Residential Buildings (Artificial stone)	ΕΛΣΤΑΤ	2
Material Cost Index for the Construction of New Residential Buildings (Timber and builders' carpentry)	ΕΛΣΤΑΤ	2
Material Cost Index for the Construction of New Residential Buildings (Basic metals)	ΕΛΣΤΑΤ	2
Material Cost Index for the Construction of New Residential Buildings (Plumbing, heating and drainage equipment and supplies)	ΕΛΣΤΑΤ	2
Material Cost Index for the Construction of New Residential Buildings (Door and window fittings)	ΕΛΣΤΑΤ	2
Material Cost Index for the Construction of New Residential Buildings (Electrical equipment)	ΕΛΣΤΑΤ	2
Material Cost Index for the Construction of New Residential Buildings (Glass products)	ΕΛΣΤΑΤ	2
Material Cost Index for the Construction of New Residential Buildings (Paints and varnishes)	ΕΛΣΤΑΤ	2
Material Cost Index for the Construction of New Residential Buildings (Floor and wall tiles and sanitary ware)	ΕΛΣΤΑΤ	2
Material Cost Index for the Construction of New Residential Buildings (Insulating materials)	ΕΛΣΤΑΤ	2

Material Cost Index for the Construction of New Residential Buildings (Elevators)	ΕΛΣΤΑΤ	2
Material Cost Index for the Construction of New Residential Buildings (Fuel for machinery (diesel), electricity, water)	ΕΛΣΤΑΤ	2
Consumer Price Index (CPI) - National Index (Overall Index)	ΕΛΣΤΑΤ	1
Consumer Price Index (CPI) - National Index (Food and non-alcoholic beverages)	ΕΛΣΤΑΤ	1
Consumer Price Index (CPI) - National Index (Alcoholic beverages and tobacco)	ΕΛΣΤΑΤ	1
Consumer Price Index (CPI) - National Index (Clothing and footwear)	ΕΛΣΤΑΤ	1
Consumer Price Index (CPI) - National Index (Housing)	ΕΛΣΤΑΤ	1
Consumer Price Index (CPI) - National Index (Durable goods, household appliances and services)	ΕΛΣΤΑΤ	1
Consumer Price Index (CPI) - National Index (Health)	ΕΛΣΤΑΤ	1
Consumer Price Index (CPI) - National Index (Transport)	ΕΛΣΤΑΤ	1
Consumer Price Index (CPI) - National Index (Communication)	ΕΛΣΤΑΤ	1
Consumer Price Index (CPI) - National Index (Recreation and culture)	ΕΛΣΤΑΤ	1
Consumer Price Index (CPI) - National Index (Education)	ΕΛΣΤΑΤ	1
Consumer Price Index (CPI) - National Index (Hotels, cafés, restaurants)	ΕΛΣΤΑΤ	1
Consumer Price Index (CPI) - National Index (Miscellaneous goods and services)	ΕΛΣΤΑΤ	1
Harmonized Index of Consumer Prices (HICP) (Overall Index)	EKT	1
Industrial Production Index (IPI) (Mining of coal and lignite)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Extraction of crude petroleum and natural gas)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Mining of metal ores)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Other mining and quarrying products)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Food)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Beverages)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Tobacco products)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Textiles)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Wearing apparel)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Leather and related products)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Wood and cork)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Paper and paper products)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Printing and recording services)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Coke and refined petroleum products)	ΕΛΣΤΑΤ	2

Industrial Production Index (IPI) (Chemicals and chemical products)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Basic pharmaceutical products and pharmaceutical preparations)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Rubber and plastic products)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Other non-metallic mineral products)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Basic metals)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Fabricated metal products)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Computers, electronic and optical products)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Electrical equipment)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Machinery and equipment n.e.c.)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Motor vehicles, trailers and semi-trailers)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Other transport equipment)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Furniture)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Other manufactured goods)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Repair and installation services of machinery and equipment)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Production and distribution of electricity)	ΕΛΣΤΑΤ	2
Industrial Production Index (IPI) (Treatment and supply services of natural water)	ΕΛΣΤΑΤ	2
Trade Balance (Total)	TTE	2
Trade Balance (Imports-Arrivals)	TTE	2
Trade Balance (Exports-dispatches)	TTE	2
Composite leading indicator (CLI)	ΟΟΣΑ	2
Business confidence index (BCI)	ΟΟΣΑ	1
Consumer confidence index (CCI)	ΟΟΣΑ	1
Crude Oil (petroleum), Dated Brent Price (US Dollars per Barrel)	EKT	4
Crude Oil (petroleum) Price (US Dollars per Barrel)	EKT	4
Crude Oil (petroleum), Dated Brent Price (Euro per Barrel)	EKT	4
Crude Oil (petroleum) Price (Euro per Barrel)	EKT	4
Exchange Rate yen_eur	EKT	1
Exchange Rate usd_eur	EKT	1
Exchange Rate cny_eur	EKT	1
Exchange Rate ukp_eur	EKT	1

ASE General Index	TTE	2
ASE Banks	TTE	2
ASE Financial services	TTE	2
ASE Non-financial corporations	TTE	2
Reserve assets (Total)	TTE	1
Reserve assets (Monetary gold)	TTE	1
Reserve assets (Monetary gold (millions of the troy ounces))	TTE	1
Reserve assets (Special drawing rights)	TTE	1
Reserve assets (Reserve position in the IMF)	TTE	1
Reserve assets (Foreign exchange)	TTE	1
Memo Items (Euro claims on non-euro area residents)	TTE	1
Memo Items (Non-euro claims on euro area residents)	TTE	1
Balance of Payments - Current account (Total)	TTE	3
Balance of Payments - Current account (Exports of goods)	TTE	3
Balance of Payments - Current account (Imports of goods)	TTE	3
Balance of Payments - Current account (Exports of services)	TTE	3
Balance of Payments - Current account (Imports of services)	TTE	3
Balance of Payments - Current account (Primary income - receipts)	TTE	3
Balance of Payments - Current account (Primary income - payments)	TTE	3
Balance of Payments - Current account (Secondary income - receipts)	TTE	3
Balance of Payments - Current account (Secondary income - payments)	TTE	3
Balance of Payments - Financial account (Total)	TTE	3
Balance of Payments - Financial account (Direct investment)	TTE	3
Balance of Payments - Financial account (Portfolio investment)	TTE	3
Balance of Payments - Financial account (Other investment)	TTE	3
Balance of Payments - Financial account (Change in international reserves)	TTE	3
Balance of Payments - Errors and omissions (Total)	TTE	3
Balance of Payments - Capital account (Total)	TTE	3
Balance of Payments - Capital account (Capital account - receipts)	TTE	3
Balance of Payments - Capital account (Capital account - payments)	TTE	3
M3 (outstanding)	EKT	2

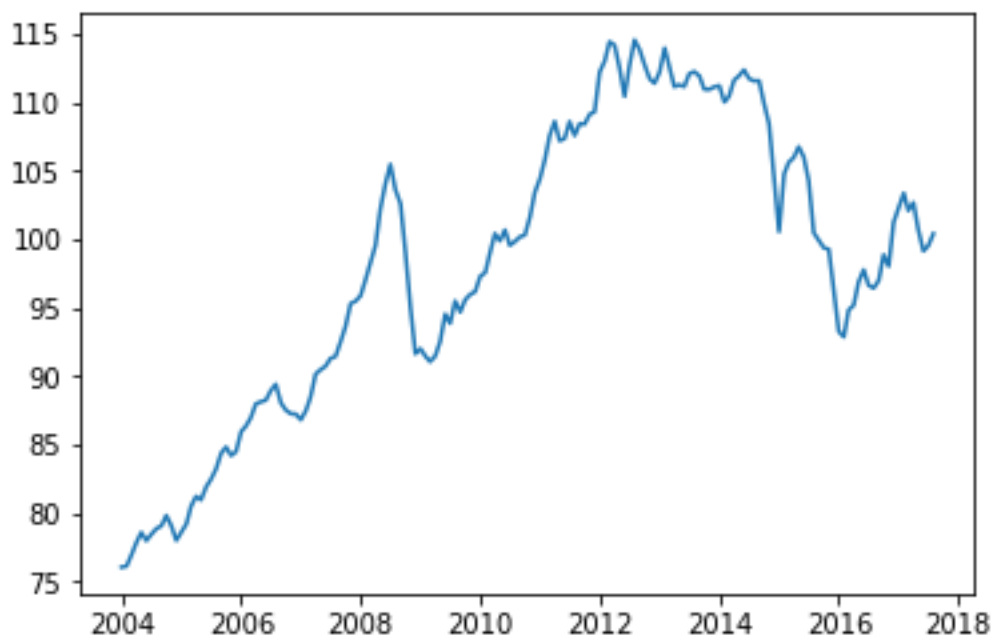
M2 (outstanding)	EKT	2
M1 (outstanding)	EKT	2
Currency in circulation (outstanding)	EKT	2
M3 excluding currency in circulation (outstanding)	EKT	2
Deposits and repos - Domestic Residents - General Government (Total)	TTE	2
Deposits and repos - Domestic Residents - General Government (Central Government)	TTE	2
Deposits and repos - Domestic Residents - General Government (Local Government)	TTE	2
Deposits and repos - Domestic Residents - General Government (Social Security Funds)	TTE	2
Deposits and repos - Domestic Residents - Corporations and Households	TTE	2
Deposits and repos - Domestic Residents - Other financial institutions	TTE	2
Deposits and repos - Domestic Residents - Non-financial corporations	TTE	2
Deposits and repos - Domestic Residents - Households and Non-profit institutions	TTE	2
Deposits and repos - Non-euro area residents	TTE	2
Deposits and repos - Total General	TTE	4
Securities other than shares and derivatives of Monetary Financial Institutions (MFIs) excluding the Bank of Greece (BoG) (Bonds -Total)	TTE	2
Securities other than shares and derivatives of Monetary Financial Institutions (MFIs) excluding the Bank of Greece (BoG) (Bonds - Credit institutions)	TTE	2
Securities other than shares and derivatives of Monetary Financial Institutions (MFIs) excluding the Bank of Greece (BoG) (Bonds - General Government)	TTE	2
Securities Portfolio of Monetary Financial Institutions (MFIs) excluding the Bank of Greece (BoG) (Mutual fund units - Total)	TTE	2
Securities Portfolio of Monetary Financial Institutions (MFIs) excluding the Bank of Greece (BoG) (Shares and other equity excluding mutual funds - Total)	TTE	2
Securities Portfolio of Monetary Financial Institutions (MFIs) excluding the Bank of Greece (BoG) (Shares and other equity excluding mutual funds - Central Bank)	TTE	2
Securities Portfolio of Monetary Financial Institutions (MFIs) excluding the Bank of Greece (BoG) (Shares and other equity excluding mutual funds - Other credit institutions)	TTE	2
Securities Portfolio of Monetary Financial Institutions (MFIs) excluding the Bank of Greece (BoG) (Shares and other equity excluding mutual funds - Insurance corporations)	TTE	2
Securities Portfolio of Monetary Financial Institutions (MFIs) excluding the Bank of Greece (BoG) (Shares and other equity excluding mutual funds - Other financial institutions)	TTE	2

Securities Portfolio of Monetary Financial Institutions (MFIs) excluding the Bank of Greece (BoG) (Shares and other equity excluding mutual funds - Non-financial corporations)	TTE	2
Economic Sentiment Indicator	IOBE	1
Industrial Confidence Indicator	IOBE	1
Construction Confidence Indicator	IOBE	1
Retail Trade Confidence Indicator	IOBE	1
Service Sector Confidence Indicator	IOBE	1
Unemployment rate	ΕΛΣΤΑΤ	3
Long-term interest rates	ΟΟΣΑ	3
Short-term interest rates	ΟΟΣΑ	1
Inflation Rates	ΟΟΣΑ	2

Πίνακας 1 : Χρονοσειρές του πειράματος, πηγές και χρονικές υστερήσεις των πραγματικών δεδομένων σε σχέση με τον τρέχοντα μήνα που εκτελείται το πείραμα κάθε φορά.

3.3.Μεθοδολογία

Το πρώτο στάδιο στο πείραμα πρόβλεψης του ΑΕΠ της Ελλάδος, αφορούσε συλλέχθηκαν οι μεταβλητές-χρονοσειρές ήταν να πραγματοποιηθεί έλεγχος ελλειπουσών τιμών για κάθε μία από αυτές. Όλες οι χρονοσειρές είναι συνεχείς (ένα παράδειγμα φαίνεται στο Γράφημα 2), εκτός των μεταβλητών του Χρηματιστηρίου Αξιών Αθηνών (ΧΑΑ), "ASE Banks", "ASE Financial services" και "ASE Non-financial corporations" οι οποίες έχουν ασυνέχεια λόγω της απαγόρευσης διαπραγματεύσεων κατά τον Ιούλιο 2015 λόγω των capital controls στις Ελληνικές συστημικές τράπεζες (<https://www.bankofgreece.gr>). Κατά σύμβαση διατηρούμε την ίδια τιμή με τον Ιουνίο 2015 και στις τρεις χρονοσειρές για τον Ιούλιο του 2015, πρώτον για διευκόλυνση επεξεργασίας και δεύτερον επειδή ουσιαστικά δεν υπήρξε αυξομείωση.



Γράφημα 2 : Δείκτης Τιμών Παραγωγού στη Βιομηχανία (γενικός δείκτης) – Producer Price Index in Industry (PPI) (overall index)

Μετά την τροποποίηση των μεταβλητών του ΧΑΑ, οι μόνες τιμές που έλλειπαν ήταν στο τέλος των χρονοσειρών λόγω της καθυστέρησης δημοσίευσής τους από τους αρμόδιους φορείς. Δημιουργήθηκε πίνακας για εκχώρηση και αποθήκευση αυτών των τιμών για να χρησιμοποιηθούν στις προγενέστερες χρονικές περιόδους – μήνες πίσω για χάρη της αναδρομικής διαδικασίας. Εκτός από τις ελλειπείς παρατηρήσεις λόγω καθυστέρησης δημοσίευσης, έπρεπε να υπολογιστούν και οι παρατηρήσεις μέχρι το τέλος του τριμήνου για το οποίο πραγματοποιήθηκε πρόβλεψη του ΑΕΠ, με σκοπό την ολοκλήρωση των τιμών ούτως ώστε όλες οι χρονοσειρές να έχουν τον ίδιο αριθμό παρατηρήσεων. Για παράδειγμα, όταν βρισκόμαστε στον Σεπτέμβριο του έτους, δε χρειαζόμαστε να συμπληρώσουμε παρατηρήσεις καθώς τότε λήγει το τρίτο τρίμηνο της οικονομίας, ενώ όταν βρισκόμαστε στον Αύγουστο του έτους, πρέπει να συμπληρωθεί μία τιμή σε κάθε χρονοσειρά (αυτή του Σεπτεμβρίου). Στον παρακάτω Πίνακα 2 παρατίθενται οι παρατηρήσεις που πρέπει να συμπληρωθούν για την πρόβλεψη του τριμηνιαίου ΑΕΠ.

Μήνας	Μήνας τριμήνου ΑΕΠ	Αριθμός μηνών που υπολείπονται
Ιανουάριος	Μάρτιος	2
Φεβρουάριος	Μάρτιος	1
Μάρτιος	Μάρτιος	0
Απρίλιος	Ιούνιος	2
Μάιος	Ιούνιος	1
Ιούνιος	Ιούνιος	0
Ιούλιος	Σεπτέμβριος	2
Αύγουστος	Σεπτέμβριος	1
Σεπτέμβριος	Σεπτέμβριος	0
Οκτώβριος	Δεκέμβριος	2
Νοέμβριος	Δεκέμβριος	1
Δεκέμβριος	Δεκέμβριος	0

Πίνακας 2 : Αριθμός μηνών που χρειάζεται να συμπληρωθούν για την πρόβλεψη του ΑΕΠ

Στην συνέχεια, για κάθε μία εκ των χρονοσειρών υπολογίστηκαν οι ελλειπείς τιμές με την κλασική μέθοδο της αυτοπαλινδρόμησης (auto-regression - AR). Για παράδειγμα για την μεταβλητή "Producer Price Index in Industry (PPI) (overall index)" σε μήνα πρόβλεψης Αύγουστο, έπρεπε να υλοποιηθεί αυτοπαλινδρόμηση για σύνολο τριών (3) μηνών, δύο (2) λόγω καθυστέρησης δημοσίευσης και ενός (01) λόγω του τριμήνου ΑΕΠ τον Σεπτέμβριο. Για την πρόβλεψη των τιμών έγινε χρήση της βιβλιοθήκης statsmodels.tsa.arima_model.ARMA με $p=1$ (AR term) και $q=0$ (MA term) για τον αριθμό των μηνών με επανατοποθέτηση των παρατηρήσεων με φόρμουλα $x_t = a + bx_{t-1} + e_t$.

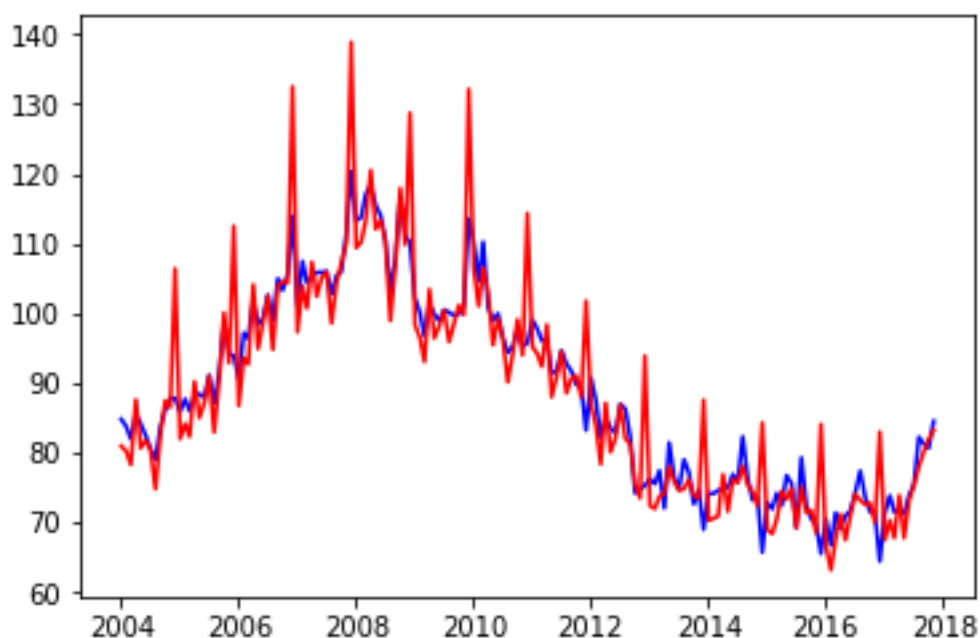
Το σύνολο των χρονοσειρών πριν τις αυτοπαλινδρομήσεις, υπέστη διόρθωση εποχικότητας, ακόμη και όσες ήταν ήδη εποχικά διορθωμένες από τις πηγές δημοσίευσης τους. Στον Πίνακα 3 παρακάτω μπορείτε να δείτε τις χρονοσειρές που ήταν ήδη διορθωμένες.

Εποχικά διορθωμένες χρονοσειρές

Industrial Production Index (IPI) (Mining of coal and lignite)
Industrial Production Index (IPI) (Extraction of crude petroleum and natural gas)
Industrial Production Index (IPI) (Mining of metal ores)
Industrial Production Index (IPI) (Other mining and quarrying products)
Industrial Production Index (IPI) (Food)
Industrial Production Index (IPI) (Beverages)
Industrial Production Index (IPI) (Tobacco products)
Industrial Production Index (IPI) (Textiles)
Industrial Production Index (IPI) (Wearing apparel)
Industrial Production Index (IPI) (Leather and related products)
Industrial Production Index (IPI) (Wood and cork)
Industrial Production Index (IPI) (Paper and paper products)
Industrial Production Index (IPI) (Printing and recording services)
Industrial Production Index (IPI) (Coke and refined petroleum products)
Industrial Production Index (IPI) (Chemicals and chemical products)
Industrial Production Index (IPI) (Basic pharmaceutical products and pharmaceutical preparations)
Industrial Production Index (IPI) (Rubber and plastic products)
Industrial Production Index (IPI) (Other non-metallic mineral products)
Industrial Production Index (IPI) (Basic metals)
Industrial Production Index (IPI) (Fabricated metal products)
Industrial Production Index (IPI) (Computers, electronic and optical products)
Industrial Production Index (IPI) (Electrical equipment)
Industrial Production Index (IPI) (Machinery and equipment n.e.c.)
Industrial Production Index (IPI) (Motor vehicles, trailers and semi-trailers)
Industrial Production Index (IPI) (Other transport equipment)
Industrial Production Index (IPI) (Furniture)
Industrial Production Index (IPI) (Other manufactured goods)
Industrial Production Index (IPI) (Repair and installation services of machinery and equipment)
Industrial Production Index (IPI) (Production and distribution of electricity)
Industrial Production Index (IPI) (Treatment and supply services of natural water)
Unemployment rate

Πίνακας 3: Εποχικά διορθωμένες χρονοσειρές

Η παραπάνω επεξεργασία έγινε με τη χρήση της βιβλιοθήκης της Python statsmodels.tsa.seasonal.seasonal_decompose του πακέτου statsmodels η οποία είναι μια standard αποσύνθεση/ανάλυση των μεταβλητών σε παράγοντες των Holt-Winters. Κάθε μεταβλητή αναλύθηκε σε δύο (2) παράγοντες (παρόμοια αντιμετώπιση με τους Dahla et al. (2006)), έναν για την εποχικότητα και έναν για την τάση ενώ η συχνότητα που δόθηκε στην συνάρτηση ήταν σε δωδεκάμηνη βάση λόγω των μηνιαίων παρατηρήσεων των χρονοσειρών. Τελικά, αφαιρέθηκε ο παράγοντας της εποχικότητας ενώ διατηρήθηκε ο παράγοντας της τάσης (παράδειγμα Γράφημα 3).



Γράφημα 3 : Αφαίρεση εποχικότητας για τον Δείκτη Κύκλου Εργασιών στο Λιανικό Εμπόριο (γενικός δείκτης) - Turnover Index in Retail Trade (overall index) (μπλέ: διορθωμένη χρονοσειρά)

Επόμενο βήμα της επεξεργασίας των χρονοσειρών ήταν η μετατροπή τους σε στάσιμες. Σε όσες από αυτές υπήρχαν αρνητικές τιμές μετά την διόρθωση της εποχικότητας, προστέθηκε σε όλες τις παρατηρήσεις της κάθε μίας η τιμή " $1 - \min(x_{i,t})$ " για να μου δοθεί η δυνατότητα να συμπεριλάβω το σύνολό τους στην μετατροπή στασιμότητας και κατά συνέχεια στο υπόλοιπο πείραμα. Για να μετατρέψω τις μεταβλητές σε στάσιμες επέλεξα τις πρώτες διαφορές των

ρυθμών μεταβολής των χρονοσειρών. Όσες ήταν σε μορφή ρυθμού μεταβολής - ποσοστό χρησιμοποιήθηκαν αυτούσιες (Πίνακας 4), ενώ στις υπόλοιπες εφαρμόστηκαν λογάριθμοι (log). Έτσι οι χρονοσειρές που δεν είχαν την μορφή ποσοστού έλαβαν την μορφή $dlog(x_{i,t}) = log(x_{i,t}) - log(x_{i,t-1})$ ή διαφορετικά $log(x_{i,t})/log(x_{i,t-1})$, ενώ οι υπόλοιπες την μορφή $x_{i,t} - x_{i,t-1}$ (παράδειγμα Γράφημα 4). Αναγκαστικά, χάθηκε η πρώτη παρατήρηση σε όλες αλλά το γεγονός αυτό αντιμετωπίστηκε όπως θα αναλυθεί στη συνέχεια της εργασίας.

Όνομα μεταβλητής

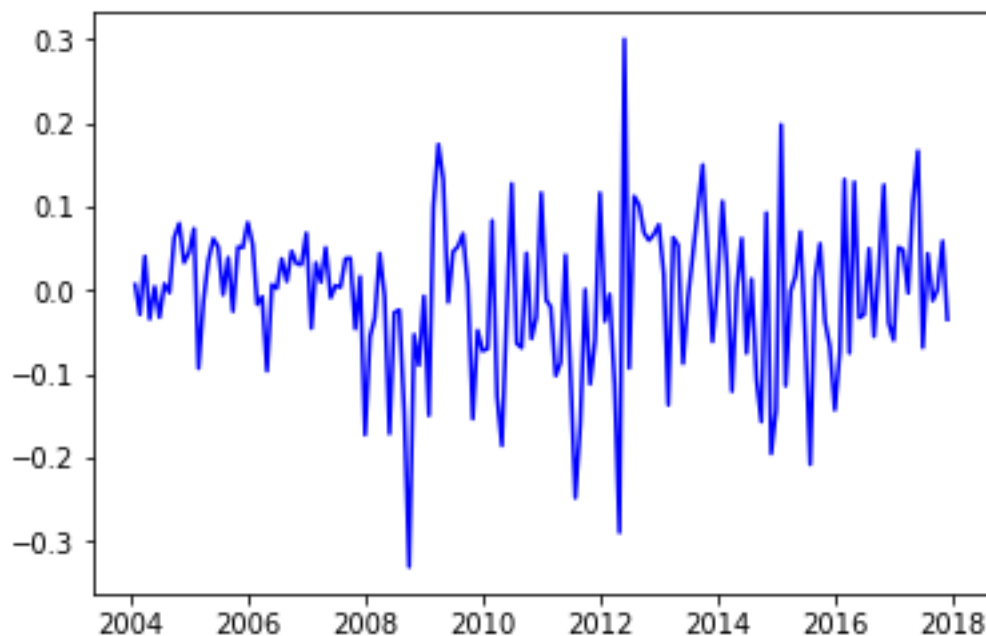
Unemployment rate

Long-term interest rates

Short-term interest rates

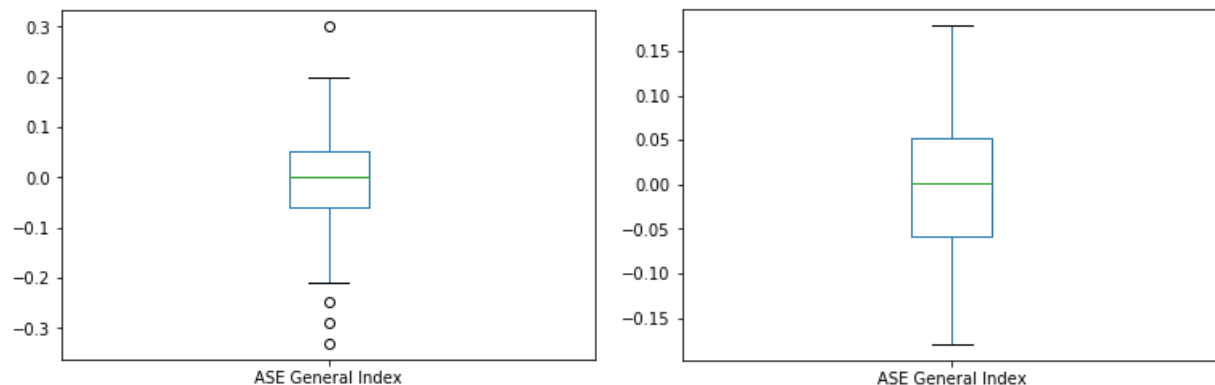
Inflation Rates

Πίνακας 4: Μεταβλητές σε μορφή ρυθμού μεταβολής



Γράφημα 4 : Μετασχηματισμός πρώτων διαφορών ρυθμών μεταβολής για την μεταβλητή του Γενικού Δείκτη ΧΑΑ
- ASE General Index

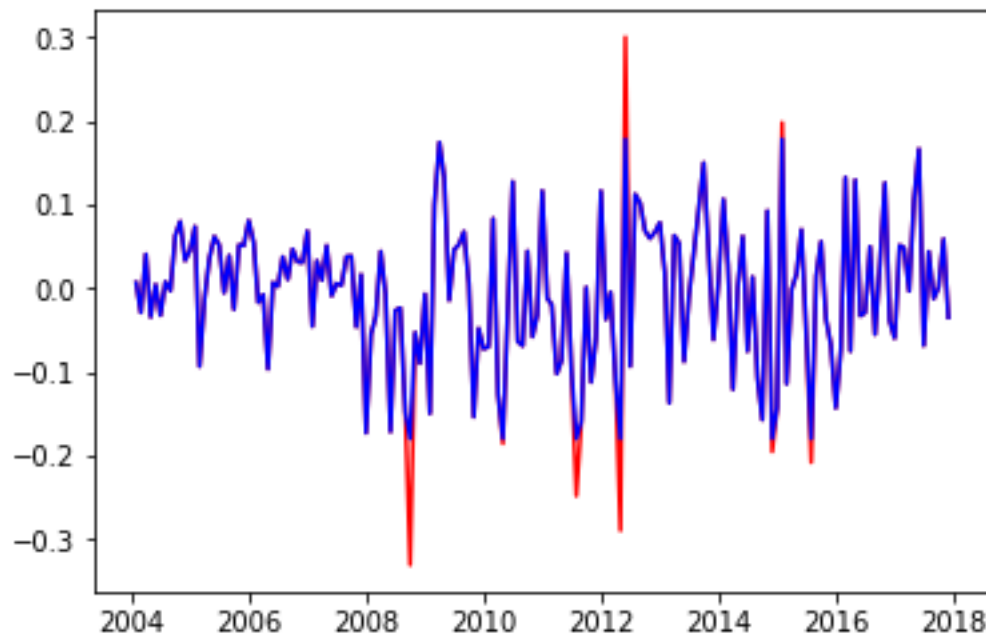
Σειρά στην επεξεργασία είχε η αντιμετώπιση των ακραίων παρατηρήσεων (outliers) κάθε χρονοσειράς. Οι παρατηρήσεις αυτές σύμφωνα με τους Artis et al., (2001) είναι shock μίας οικονομίας. Εκείνοι όσα outliers βρήκανε στην έρευνά τους τα παρέλειψαν και όρισαν τις παρατηρήσεις αυτές ως ελλειπούσες τιμές (missing values) για να προχωρήσουν. Στην παρούσα εργασία τα outliers θα περικοπούν στο μέγιστο και ελάχιστο κατώφλι κάθε χρονοσειράς (min – max threshold) γιατί θεωρώ πως εσωκλείουν πληροφορία και πρέπει να συμπεριληφθούν, αφού μετριάσει η πίεση του εκάστοτε οικονομικού shock. Πρώτα αναλύθηκε κάθε χρονοσειρά σε κατάλληλο γράφημα Box Plot για να εντοπιστούν οι ακραίες τιμές της (Γράφημα 5).



Γράφημα 5 : Box Plot χρονοσειράς Γενικού Δείκτη ΧΑΑ (αριστερά: με outliers, δεξιά: χωρίς)

Κατόπιν, ορίστηκε ως μέγιστο κατώφλι (max threshold) η τιμή της παρατήρησης της χρονοσειράς που βρίσκεται στην θέση 70% του συνόλου των παρατηρήσεων σύν 1,5 επί της διαφοράς των τιμών των παρατηρήσεων της θέσης 70% και 30%. Ταυτόχρονα, ορίστηκε ελάχιστο κατώφλι (min threshold) η τιμή της παρατήρησης που βρίσκεται στην θέση 30% του συνόλου των τιμών των παρατηρήσεων της χρονοσειράς μείον 1,5 επί της διαφοράς των παρατηρήσεων της θέσης 70% και 30%. Η επιλογή διαφέρει από την κλασική αντιμετώπιση μέσω των τεταρτημορίων 25%-75% (Q1 – Q3 percentiles) και του διατεταρτημοριακού εύρους (Interquartile range, IQR = Q3 - Q1) καθώς παρατηρήθηκε πως με τα συγκεκριμένα όρια δεν

εξαλείφονταν πλήρως τα shock. Παράδειγμα της προσαρμογής των outliers φαίνεται καθαρά στο Γράφημα 6 παρακάτω.



Γράφημα 6 : Outlier cropping Γενικού Δείκτη ΧΑΑ (κόκκινο: τα outliers, μπλε: η χρονοσειρά μετά την επεξεργασία)

Πριν ολοκληρωθεί στο σύνολό της η επεξεργασία των εκατόν εβδομήντα πέντε (175) χρονοσειρών, συμπληρώθηκε η χαμένη παρατήρηση του Ιανουαρίου 2004 (αρχή παρατηρήσεων) λόγω της μετατροπής πρώτων διαφορών. Έγινε χρήση της μεθόδου διπλής εκθετικής εξομάλυνσης (Double Exponential Smoothing) των Holt-Winters (Kalekar, 2004). Επιλέχθηκε η συγκεκριμένη έναντι των Single (SES) και Triple Exponential Smoothing (TES) διότι η μεν SES δεν εντοπίζει τάση και εποχικότητα, η δε TES χρησιμοποιείται σε χρονοσειρές που πλήττονται από εποχικότητα, την οποία στο παρόν πείραμα έχει ήδη εξαληφθεί προγενέστερα.

Εν συνεχεία, πραγματοποιήθηκε τυπικός έλεγχος στασιμότητας για την επικείμενη ανάλυση κύριων συνιστωσών. Έγινε χρήση της βιβλιοθήκης arch.unitroot της Python και των μεθόδων ADF, DFGLS και PhillipsPerron που αντιστοιχούν στους ελέγχους στασιμότητας

Augmented Dickey Fuller test, Elliott, Rothenberg and Stock's GLS test και Phillips–Perron test αντίστοιχα. Κάθε μία χρονοσειρά εξετάστηκε (Πίνακας 5) σε κάθε ένα από τα τρία τέστ για επίπεδα εμπιστοσύνης 90%, 95% και 99% ($p = 0.1, 0.05, 0.01$) και οι μόνες που εμφανίστηκαν ως μη-στάσιμες είναι οι "Composite leading indicator (CLI)" και "Consumer confidence index (CCI)". Παρ' όλα αυτά, επειδή διαγραμματικά δείχνουν στάσιμες και επειδή οι παρατηρήσεις τους αποτελούν μόλις 2/175 ~ 1.14% του δείγματος συμπεριλήφθηκαν στην PCA.

Test	Confidence level 90%	Confidence level 95%	Confidence level 99%
ADF	Composite leading indicator (CLI)	Composite leading indicator (CLI)	Composite leading indicator (CLI)
DFGLS	-	-	Composite leading indicator (CLI) Consumer confidence index (CCI)
PP	Composite leading indicator (CLI)	Composite leading indicator (CLI)	Composite leading indicator (CLI)

Πίνακας 5 : Έλεγχοι στασιμότητας – Stationarity tests. (ADF: Augmented Dickey–Fuller test, DFGLS: Elliott, Rothenberg, and Stock modified DF test, PP: Phillips-Perron test).

Ακολούθησε η εφαρμογή της PCA αναδρομικά για 95 μήνες πίσω συμπεριλαμβανομένου και του Σεπτεμβρίου 2017, δηλαδή για σχεδόν 8 χρόνια πίσω. Επέλεξα να εξάγω τόσες συνιστώσες, όσες χρειαζότανε για να εσωκλείουν το 70% και 80% του συνόλου της πληροφορίας των αρχικών μετασχηματισμένων χρονοσειρών με σκοπό την δημιουργία δύο μοντέλων πρόβλεψης αντίστοιχα. Κατέφυγα σε αυτή την τακτική επειδή θεώρησα πως μεγάλος αριθμός συνιστωσών θα οδηγήσει σε μεγάλα σφάλματα πρόβλεψης λόγω θορύβου και πολυσυγγραμμικότητας των μεταβλητών. Η PCA μειώνει δραστικά την πολυσυγγραμμικότητα των δεδομένων μειώνοντας την διάσταση/πλήθος τους συγκρατώντας μεγάλο ποσοστό πληροφορίας. Χαρακτηριστικά, όπως φαίνεται και στον Πίνακα 6, μόλις έξι (6) με οχτώ (8) συνιστώσες μπορούν να περιέχουν το 70% της ολικής πληροφορίας (διακύμανσης) του δείγματος και αντίστοιχα εννέα (9) με δώδεκα (12) συνιστώσες το 80% της πληροφορίας αντίστοιχα. Ο αριθμός εξαρτάται από το πλήθος των παρατηρήσεων καθώς όσο η αναδρομή φθάνει στο τέλος προς τα πίσω χρονικά, ο αριθμός φθίνει. Αξίζει να σημειωθεί πως για να διατηρηθεί το σύνολο (100%) της διακύμανσης πρέπει να διατηρηθούν 72-168 συνιστώσες αναλόγως χρονικής στιγμής, αριθμός εξαιρετικά μεγάλος.

Αναδρομική PCA
Πλήθος συνιστωσών

Αριθμός μηνών πίσω	70% πληροφορία	80% πληροφορία	100% πληροφορία
0	8	12	168
1	8	12	165
2	8	12	165
3	8	12	165
4	8	12	162
5	8	12	162
6	8	12	162
7	8	12	159
8	8	12	159
9	8	12	159
10	8	12	156
11	8	12	156
12	8	12	156
13	8	12	153
14	8	12	153
15	8	12	153
16	8	12	150
17	8	12	150
18	8	12	150
19	8	12	147
20	8	11	147
21	8	11	147
22	8	12	144
23	8	12	144
24	8	12	144
25	8	12	141
26	8	11	141

27	8	11	141
28	8	12	138
29	8	12	138
30	8	12	138
31	8	12	135
32	8	12	135
33	8	12	135
34	8	12	132
35	8	12	132
36	8	11	132
37	8	11	129
38	8	11	129
39	8	11	129
40	7	11	126
41	7	11	126
42	7	11	126
43	7	11	123
44	7	11	123
45	7	11	123
46	8	11	120
47	8	11	120
48	8	11	120
49	8	11	117
50	7	11	117
51	7	11	117
52	8	11	114
53	8	11	114
54	7	11	114
55	7	11	111

56	7	11	111
57	8	11	111
58	8	11	108
59	8	11	108
60	8	11	108
61	8	11	105
62	8	11	105
63	8	11	105
64	8	12	102
65	8	11	102
66	8	11	102
67	7	11	99
68	7	10	99
69	7	10	99
70	7	11	96
71	7	11	96
72	7	11	96
73	7	11	93
74	7	11	93
75	7	11	93
76	7	10	90
77	7	11	90
78	7	10	90
79	6	9	87
80	6	9	87
81	6	9	87
82	6	10	84
83	6	10	84
84	6	9	84

85	6	10	81
86	6	9	81
87	6	10	81
88	6	10	78
89	6	10	78
90	6	10	78
91	6	10	75
92	6	9	75
93	6	9	75
94	6	9	72
95	6	10	72

Πίνακας 6 : Συνιστώσες PCA. Ο αριθμός (πλήθος) των συνιστωσών που συγκεντρώνουν το 70% και 80% αντίστοιχα της πληροφορίας του συνόλου των χρονοσειρών για τους 95 μήνες που δοκιμάστηκαν στο πείραμα

Το τελευταίο στάδιο του πειράματος ήταν η πρόβλεψη του ΑΕΠ με την standard μέθοδο OLS με φόρμουλα $y_t = A + CX_t + E_t$ με $y_t = \log(GDP_{i,t})/\log(GDP_{i,t-1})$ (για κώδικα βλ.Παράρτημα II). Οι συνιστώσες που χρησιμοποιήθηκαν από την PCA ήταν σε μηνιαία μορφή λόγω της μηνιαίας περιοδικότητας των χρονοσειρών – μεταβλητών. Τροποποιήθηκαν σε τριμηνιαίες (Μ.Ο. τριμήνου) και παλινδρομήθηκαν γραμμικά στις διαφορές λογαρίθμων του τριμηνιαίου ΑΕΠ. Τα μοντέλα αντίστοιχα ήταν ένα που περιέχει τις συνιστώσες του 70% πληροφορίας της PCA (στο εξής model70), και ένα που περιέχει το 80% (στο εξής model80). Η παλινδρόμηση έγινε με χρήση της sklearn.linear_model βιβλιοθήκης της Python. Εφόσον η πρόβλεψη αφορούσε τις διαφορές λογαρίθμων του ΑΕΠ στο τέλος έπρεπε να γίνει κατάλληλος μετασχηματισμός για να βρεθούν οι προβλέψεις μετρούμενες σε εκατομμύρια Ευρώ, όπως άλλωστε ήταν και η μορφή του ΑΕΠ από τα δεδομένα της ΕΛΣΤΑΤ. Μετά τον μετασχηματισμό αυτό υπολογίστηκαν τα σφάλματα πρόβλεψης τα οποία θα παρατεθούν στο Κεφάλαιο 4.

3.4.Σφάλματα

Τα σφάλματα – δείκτες σύγκρισης των δύο μοντέλων είναι τα κάτωθι :

- Μέσο απόλυτο σφάλμα (mean absolute error) MAE : Το MAE είναι ένα μέτρο της διαφοράς μεταξύ δύο συνεχών μεταβλητών, στην περίπτωση μας των πραγματικών τιμών του ΑΕΠ και των προβλεπόμενων τιμών.
- Μέσο απόλυτο ποσοστιαίο σφάλμα (mean absolute percentage error) MAPE : Το μέσο σφάλμα απόλυτου ποσοστού, γνωστό και ως μέση απόλυτη ποσοστιαία απόκλιση (MAPD), είναι ένα μέτρο προβλεπτικής ακρίβειας. Είναι το MAE εκφρασμένο σε ποσοστό.
- Μέσο τετραγωνικό σφάλμα (mean squared error) MSE : Το μέσο τετραγωνικό σφάλμα (MSE) ή η μέση τετραγωνική απόκλιση (MSD) ενός εκτιμητή (μιας διαδικασίας για την εκτίμηση μιας μη παρατηρημένης ποσότητας) μετρά τον μέσο όρο των τετραγώνων των σφαλμάτων ή των αποκλίσεων - δηλαδή τη διαφορά μεταξύ του εκτιμητή και του εκτιμώμενου .
- Ρίζα μέσου τετραγωνικού σφάλματος (root mean squared error) RMSE : Η τετραγωνική ρίζα του MSE/MSD

Για την υποβοήθηση του αναγνώστη παρατίθεται ο παρακάτω Πίνακας 7.

Mean squared error	MSE	=	$\frac{1}{n} \sum_{t=1}^n e_t^2$
Root mean squared error	RMSE	=	$\sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2}$
Mean absolute error	MAE	=	$\frac{1}{n} \sum_{t=1}^n e_t $
Mean absolute percentage error	MAPE	=	$\frac{100\%}{n} \sum_{t=1}^n \left \frac{e_t}{y_t} \right $

Πίνακας 7: Στατιστικά σφάλματα, δείκτες ($e_t = y_t - \hat{y}_t$)

Πηγή: <https://tex.stackexchange.com>

4.Αποτελέσματα - Συμπεράσματα

Όπως προαναφέρθηκε οι ποσοτικές πληροφορίες σχετικά με την οικονομική κατάσταση και το γενικότερο οικονομικό κλίμα είναι υψίστης σημασίας για τη χάραξη της οικονομικής πολιτικής μιας χώρας, αλλά τα στοιχεία για το εθνικό ΑΕΠ γίνονται διαθέσιμα με σημαντική καθυστέρηση.

Σκοπός και κύριο αντικείμενο της παρούσης εργασίας ήταν να μελετηθεί εμπειρικά η δυνατότητα πρόβλεψης του Ελληνικού ΑΕΠ δια μέσω ενός υπερμεγέθους συνόλου μακροοικονομικών μεταβλητών - χρονοσειρών. Συλλέχθηκαν 175 χρονοσειρές εγχώριες αλλά και παγκόσμιες (π.χ. ισοτιμία Ευρώ - Δολαρίου) από ποικίλους πλην έγγυρους και έγκαιρους φορείς και θεσμούς από τον Ιανουάριο του 2004 έως και τον Οκτώβριο του 2017. Οι χρονοσειρές αφού πέρασαν όλους τους απαραίτητους και δικαιολογημένους ελέγχους καθώς και μία σειρά μετατροπών, χρησιμοποιήθηκαν στο συνολικά και οι 175 για να προβλεφθεί το ΑΕΠ της Ελλάδος. Θεώρησα απαραίτητο να χρησιμοποιήσω ένα τόσο μεγάλο σύνολο, διότι πιστεύω πως η οικονομία επηρεάζει και επηρεάζεται καθολικά. Κάθε μία μεταβλητή έχει να προσδώσει έστω και ψήγματα ουσιώδους πληροφορίας για την κίνηση της οικονομίας. Σαφέστατα, αρνητικά στοιχεία είναι απρόβλεπτοι παράγοντες (shocks) που μπορεί να προσθέσουν θόρυβο στην διαθέσιμη πληροφορία. Επιπλέον, υπάρχουν πάντα σφάλματα μέτρησης, σφάλματα ελλειπουσών τιμών καθώς και σφάλματα διορθώσεων κατά τις αναθεωρήσεις των τιμών (revisions). Παρ'όλα αυτά, όλοι οι παράγοντες συμπεριλαμβανομένου του ΑΕΠ, κινούνται σαν ένα σώμα, μαζικά έστω και αν έχουν διαφορετικούς ρυθμούς ανάπτυξης/ύφεσης κατά τόπους και τομείς της οικονομίας (για παράδειγμα το ΑΕΠ μπορεί να αυξάνεται αλλά ο αγροτικός τομέας να μειώνεται) . Με αυτή τη θεώρηση προχώρησα στο πείραμα και πιστεύω πως τα αποτελέσματα είναι ιδιαίτερα ενθαρρυντικά.

Τα model70 και model80 είναι μοντέλα ανάλυσης κυρίων συνιστωσών από τις 175 χρονοσειρές, τα οποία εμπεριέχουν 70% και 80% αντίστοιχα της ολικής πληροφορίας τους. Τα ποσοστά επιλέχθηκαν για να μετριάσκει ο αριθμός των κυρίων συνιστωσών και κατά συνέπεια ο θόρυβος της πληθώρας των χρονοσειρών. Τα αποτελέσματα-σφάλματα του πειράματος φαίνονται στον Πίνακα 8. Για κάθε μοντέλο, έπρεπε να υπάρχουν τρεις μετρήσεις σφαλμάτων, αναλόγως του μήνα κατά τον οποίο εφαρμόζεται η διαδικασία (αναδρομικά) και του πόσο

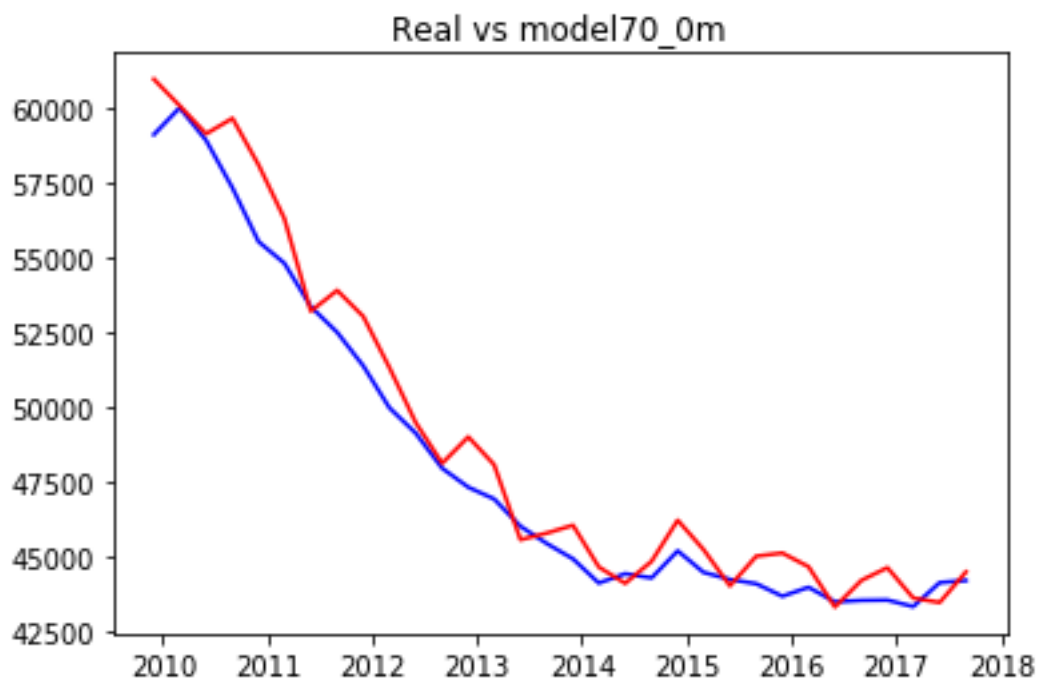
απέχει από τον μήνα δημοσίευσης του τριμηνιαίου ΑΕΠ. Αυτό συμβαίνει λόγω της διαθεσιμότητας των πραγματικών δεδομένων κατά τον εκάστοτε μήνα, επειδή χρειάζεται να συμπληρωθούν τιμές των χρονοσειρών κάθε φορά μέχρις ότου φθάσουμε στον μήνα δημοσίευσης. Οι μήνες Μάρτιος, Ιούνιος, Σεπτέμβριος και Δεκέμβριος είναι current μήνες, οι μήνες Φεβρουάριος, Μάιος, Αύγουστος και Νοέμβριος υπολείπονται ενός μήνα σε πραγματικά δεδομένα και οι μήνες Ιανουάριος, Απρίλιος, Ιούλιος, Οκτώβριος υπολείπονται δύο μηνών σε πραγματικά δεδομένα. Μεταξύ τους αξιολογήθηκαν και οι τρεις κατηγορίες αντίστοιχα.

Models	Months until GDP release	MAE (EUR mil.)	MAPE (%)	MSE	RMSE
Model70	0	875.18	1.78 %	1194957.12	1093.14
Model80	0	935.52	1.91 %	1182635.84	1087.49
Model70	1	828	1.69 %	1012561.09	1006.26
Model80	1	838.76	1.73 %	1005132.75	1002.56
Model70	2	760.18	1.58 %	873942.32	934.85
Model80	2	843.51	1.74 %	1033702.79	1016.71

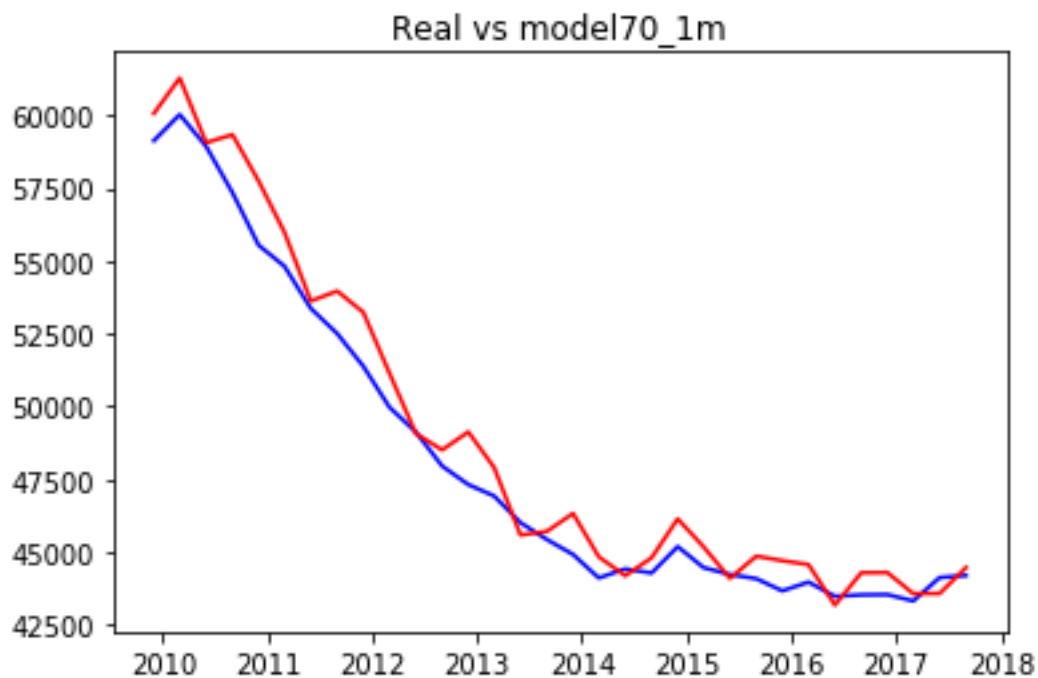
Πίνακας 8 : Σφάλματα προβλέψεων των model70 και model80

Σε γενικές γραμμές, η μέθοδος που ακολουθήθηκε απέφερε καρπούς (για αριθμητικά αποτελέσματα βλ.Παράρτημα Ι). Κανένα μοντέλο σε καμία από τις τρεις κατηγορίες μηνών δεν υπερέβη το ποσοστό σφάλματος του δύο τοις εκατό (2%). Συγκριτικά, το model70 είχε χαμηλότερα σφάλματα MAE και MAPE σε όλες τις κατηγορίες αλλά παρόμοια MSE και RMSE με το model80 εκτός από την κατηγορία που υπολείπονται 2 μήνες για την δημοσίευση του ΑΕΠ. Το καλύτερο ποσοστό είναι 1.58%, το οποίο αντιστοιχεί σε μέση απόκλιση 760.18 εκατομμυρίων Ευρώ, αρκετά χαμηλό για προβλεπτικό μοντέλο. Αξιοσημείωτο ωστόσο είναι το γεγονός πως καλύτερα αποτελέσματα πήραμε όταν είμασταν μακριά από την δημοσίευση του ΑΕΠ και έπρεπε να εκτιμήσουμε τις τιμές των ερμηνευτικών μεταβλητών για 2 μήνες μπροστά. Αυτό σημαίνει πως οι εκτιμήσεις που έγιναν με χρήση κλασικής AR ήταν ωφέλιμες και μάλιστα υποδεικνύει πως η ίδια η μεταβλητή εσωκλείει μεγάλη πληροφορία για τον εαυτό της (αυτοσυσχέτιση - autocorrelation), γεγονός που χρήζει περαιτέρω διερεύνησης. Γραφικά

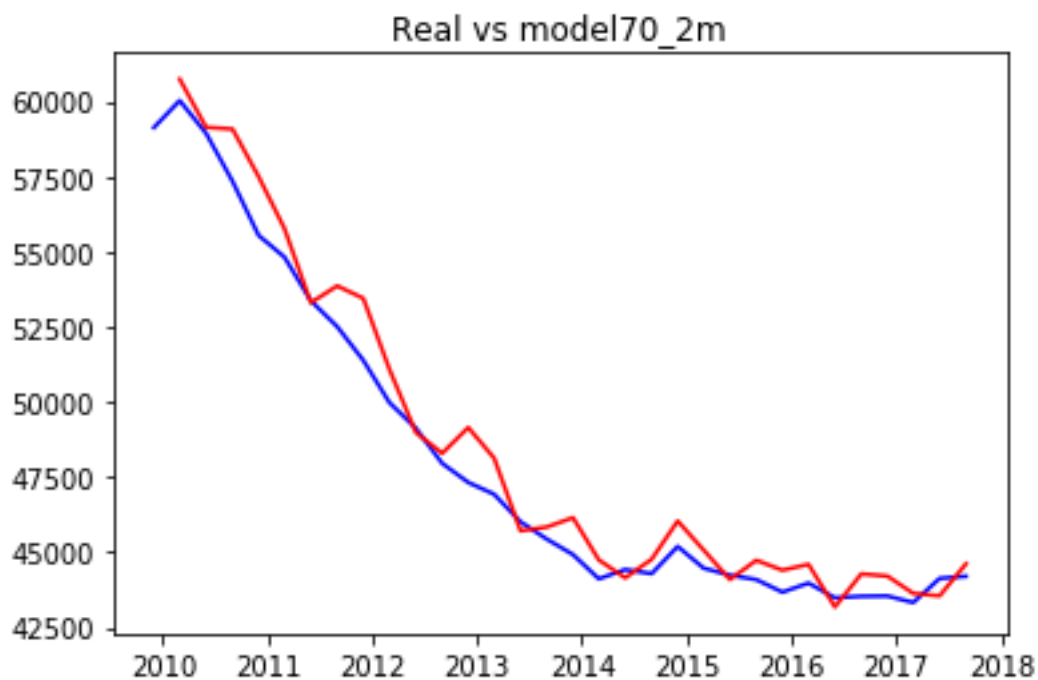
αποτελέσματα μπορείτε να δείτε στα παρακάτω Γραφήματα 7 έως 18. Στα Γραφήματα 7 - 12 συγκρίνονται το πραγματικό ΑΕΠ με τις προβλέψεις των μοντέλων για τις κατηγορίες των μηνών που αναφερθήκανε προ ολίγου, στα Γραφήματα 13 – 15 συγκρίνονται τα σφάλματα μεταξύ ίδιων ομάδων μηνών των μοντέλων μεταξύ τους ενώ στα Γραφήματα 16 - 18 απεικονίζονται οι διαφορές (σφάλματα) των προβλέψεων με το ΑΕΠ ανά μήνα. Τέλος στο γράφημα 19 μπορείτε να δείτε ότι το σύνολο των προβλέψεων εμπεριέχει εποχικότητα.



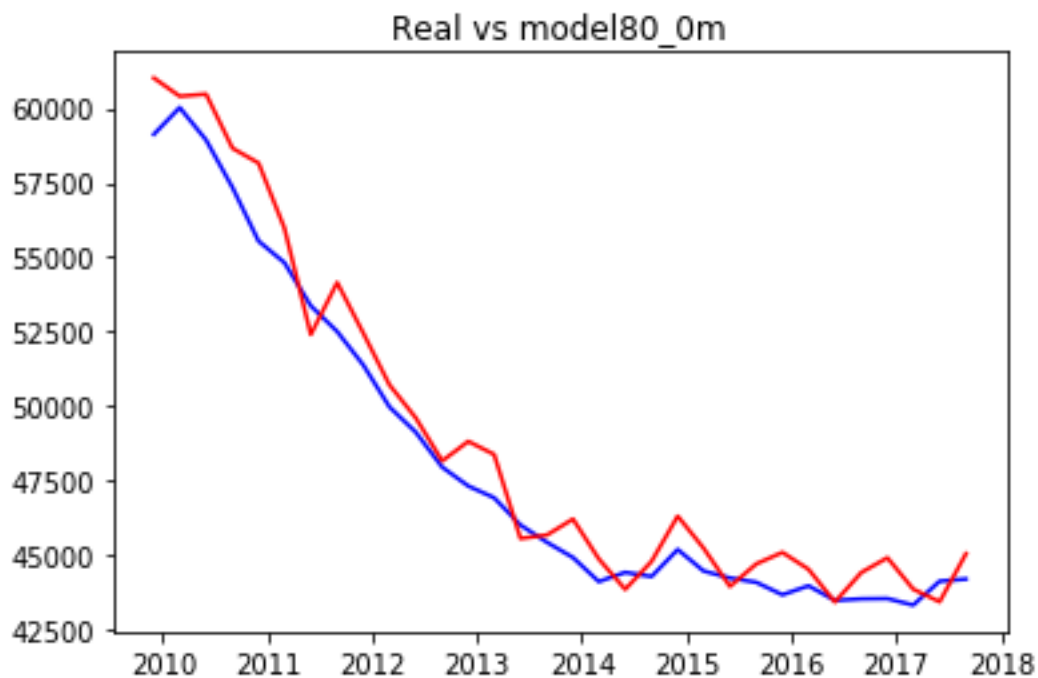
Γράφημα 7 : Σύγκριση πραγματικού ΑΕΠ (μπλε) και model70_0m (κόκκινο). Το 0m δείχνει ότι οι προβλέψεις έγιναν με δεδομένα που είχανε πραγματικές τιμές μέχρι και τον τρέχοντα μήνα δημοσίευσης του ΑΕΠ.



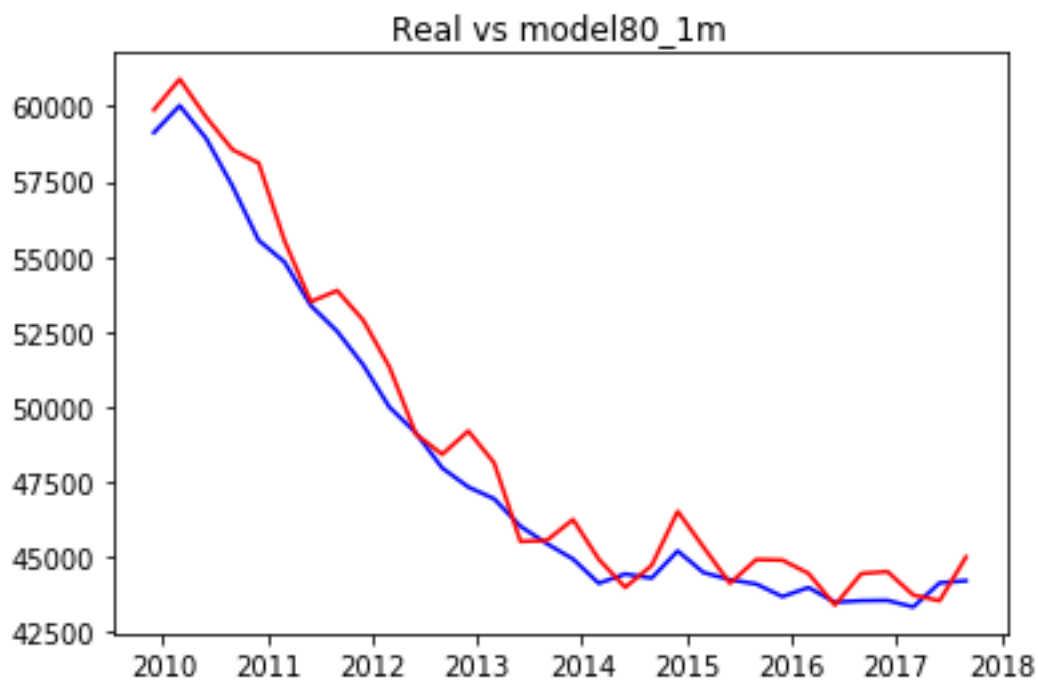
Γράφημα 8 : Σύγκριση πραγματικού ΑΕΠ (μπλε) και model70_1m (κόκκινο). Το 1m δείχνει ότι οι προβλέψεις έγιναν με δεδομένα που είχανε πραγματικές τιμές μέχρι και τον προηγούμενο μήνα από δημοσίευση του ΑΕΠ.



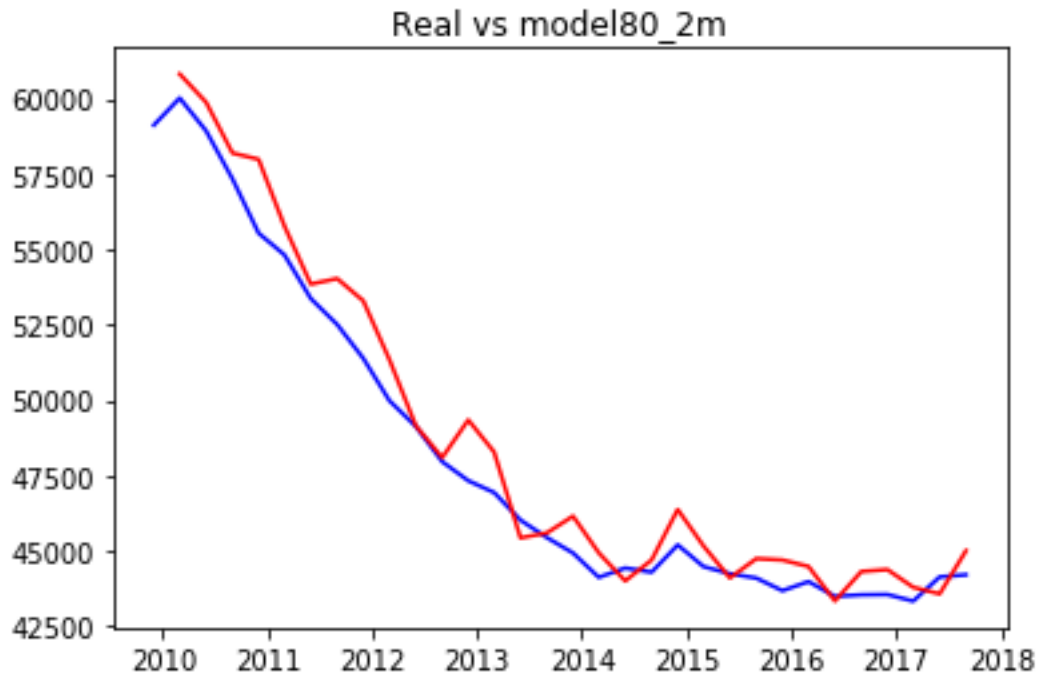
Γράφημα 9 : Σύγκριση πραγματικού ΑΕΠ (μπλε) και model70_2m (κόκκινο). Το 2m δείχνει ότι οι προβλέψεις έγιναν με δεδομένα που είχανε πραγματικές τιμές μέχρι και τον προ-προηγούμενο μήνα από δημοσίευση του ΑΕΠ.



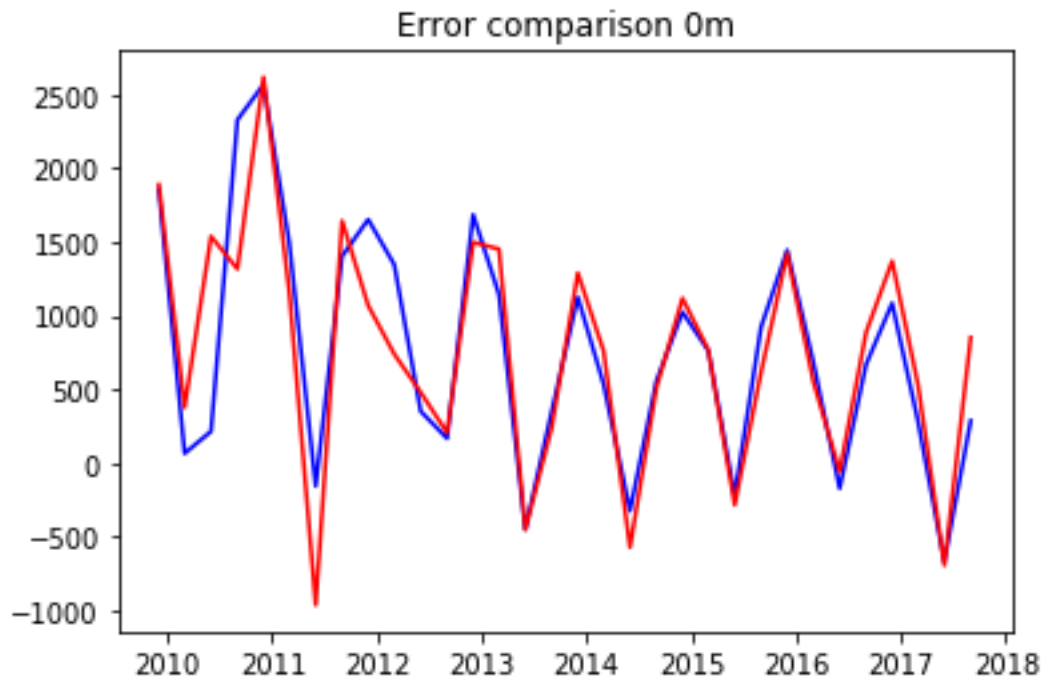
Γράφημα 10 : Σύγκριση πραγματικού ΑΕΠ (μπλε) και model80_0m (κόκκινο). Το 0m δείχνει ότι οι προβλέψεις έγιναν με δεδομένα που είχαν πραγματικές τιμές μέχρι και τον τρέχοντα μήνα δημοσίευσης του ΑΕΠ.



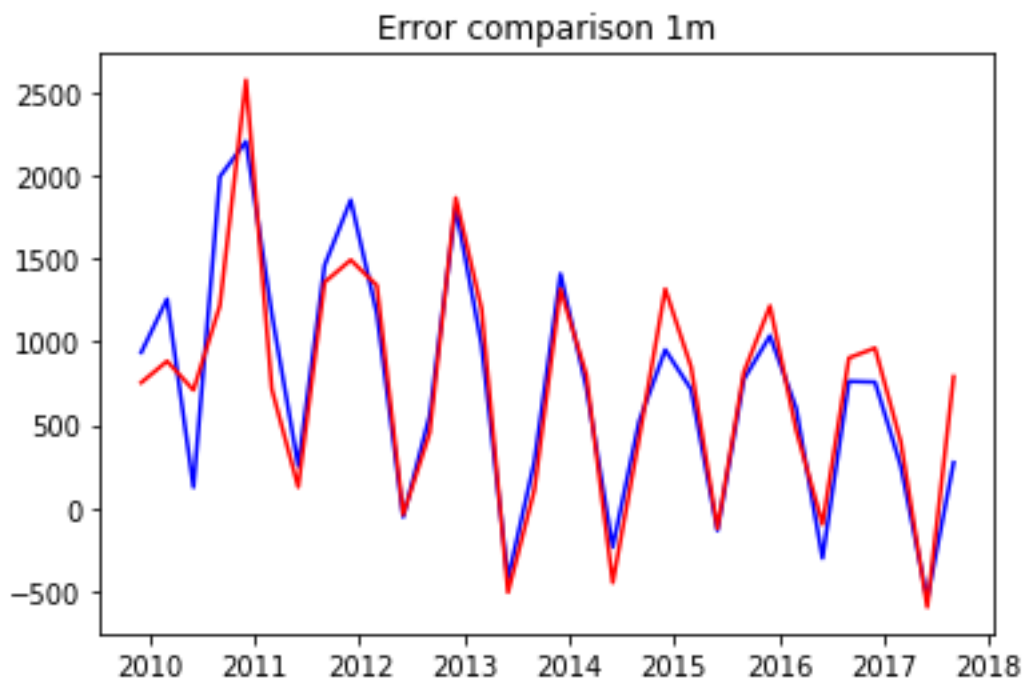
Γράφημα 11 : Σύγκριση πραγματικού ΑΕΠ (μπλε) και model80_1m (κόκκινο). Το 1m δείχνει ότι οι προβλέψεις έγιναν με δεδομένα που είχαν πραγματικές τιμές μέχρι και τον προηγούμενο μήνα από δημοσίευση του ΑΕΠ.



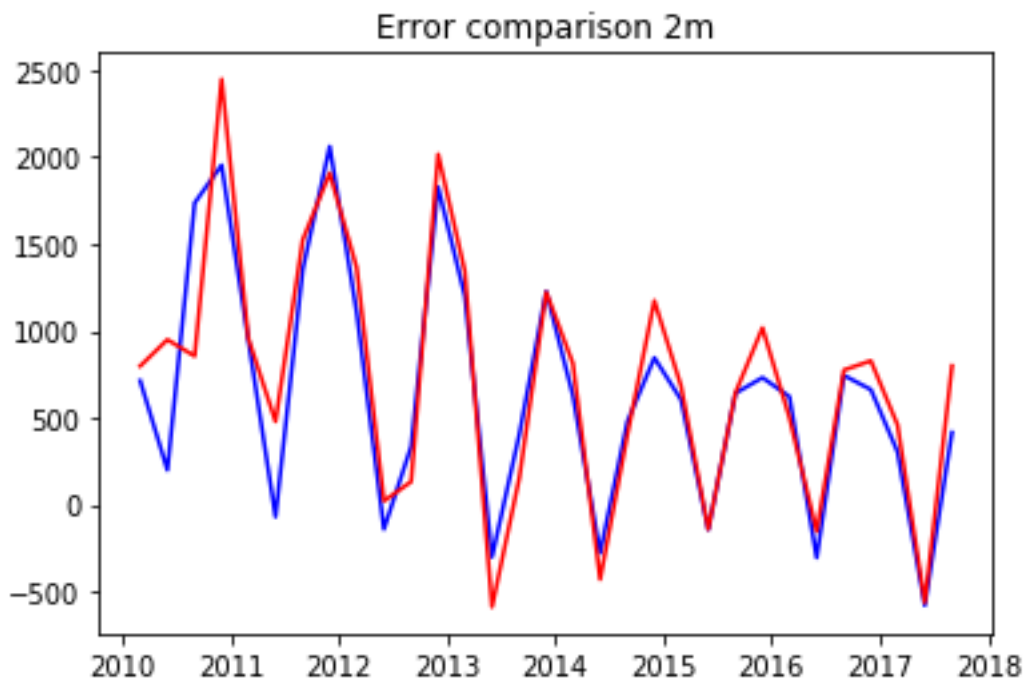
Γράφημα 12 : Σύγκριση πραγματικού ΑΕΠ (μπλε) και model80_2m (κόκκινο). Το 2m δείχνει ότι οι προβλέψεις έγιναν με δεδομένα που είχανε πραγματικές τιμές μέχρι και τον προ-προηγούμενο μήνα από δημοσίευση του ΑΕΠ.



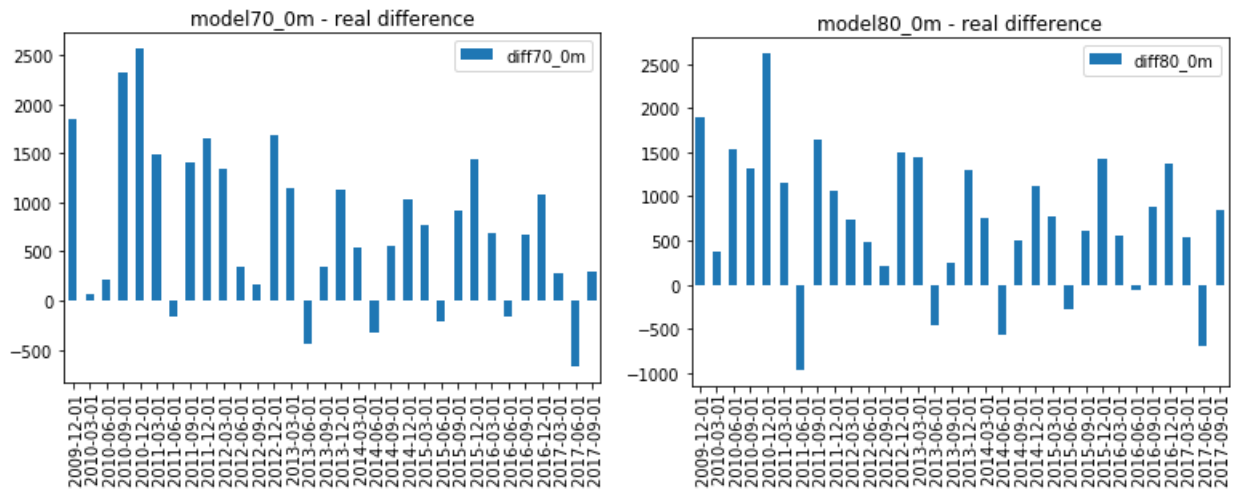
Γράφημα 13 : Σύγκριση σφαλμάτων 0m των model70 (μπλε) και model80 (κόκκινο). Το 0m δείχνει ότι οι προβλέψεις έγιναν με δεδομένα που είχανε πραγματικές τιμές μέχρι και τον τρέχοντα μήνα δημοσίευσης του ΑΕΠ.



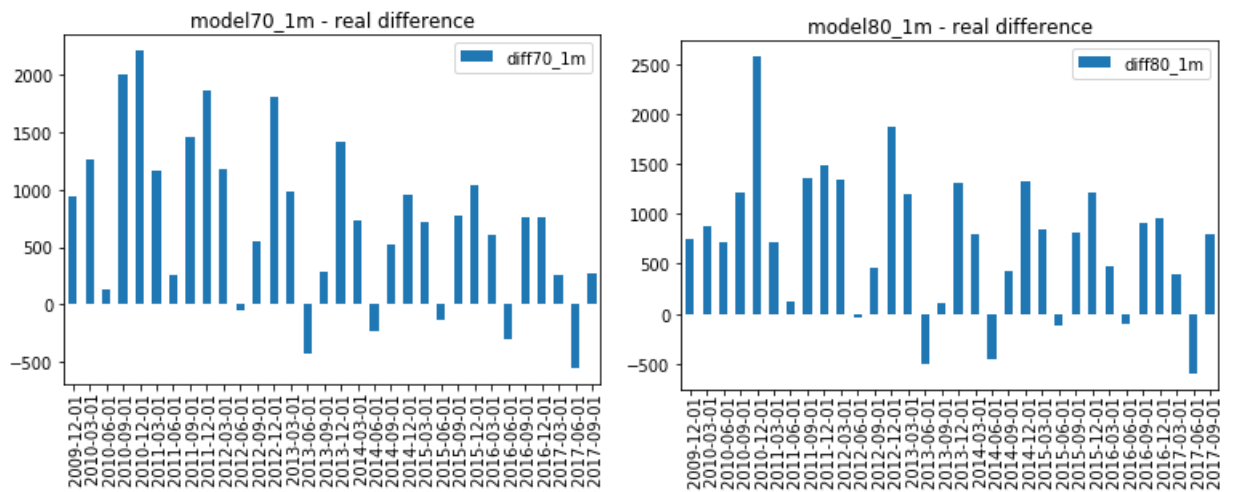
Γράφημα 14 : Σύγκριση σφαλμάτων 1m των model70 (μπλε) και model80 (κόκκινο). Το 1m δείχνει ότι οι προβλέψεις έγιναν με δεδομένα που είχανε τιμές μέχρι και τον προηγούμενο μήνα από δημοσίευση του ΑΕΠ.



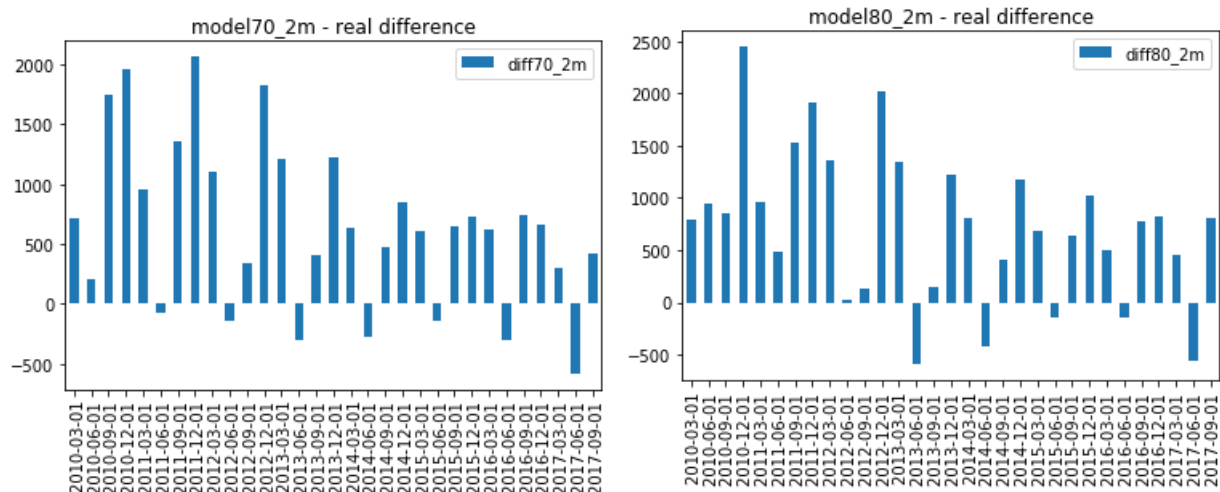
Γράφημα 15 : Σύγκριση σφαλμάτων 2m των model70 (μπλε) και model80 (κόκκινο). Το 2m δείχνει ότι οι προβλέψεις έγιναν με δεδομένα που είχανε τιμές μέχρι και τον προ-προηγούμενο μήνα από δημοσίευση του ΑΕΠ.



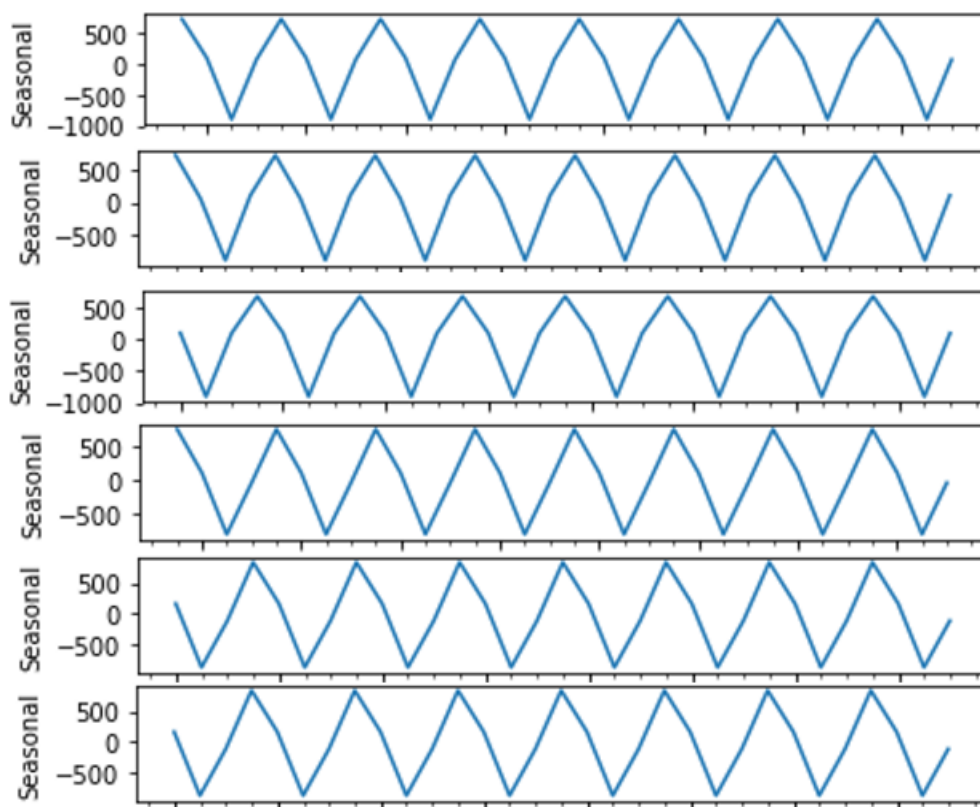
Γράφημα 16 : Σύγκριση σφαλμάτων 0m των model70 (αριστερά) και model80 (δεξιά) σε ραβδογράμματα. Το 0m δείχνει ότι οι προβλέψεις έγιναν με δεδομένα που είχανε πραγματικές τιμές μέχρι και τον τρέχοντα μήνα δημοσίευσης του ΑΕΠ.



Γράφημα 17 : Σύγκριση σφαλμάτων 1m των model70 (αριστερά) και model80 (δεξιά) σε ραβδογράμματα. Το 1m δείχνει ότι οι προβλέψεις έγιναν με δεδομένα που είχανε πραγματικές τιμές μέχρι και τον προηγούμενο μήνα από δημοσίευση του ΑΕΠ.



Γράφημα 18 : Σύγκριση σφαλμάτων 2m των model70 (αριστερά) και model80 (δεξιά) σε ραβδογράμματα. Το 2m δείχνει ότι οι προβλέψεις έγιναν με δεδομένα που είχανε πραγματικές τιμές μέχρι και τον προ-προηγούμενο μήνα από δημοσίευση του ΑΕΠ.



Γράφημα 19 : Εποχικότητα των προβλέψεων. Από πάνω προς τα κάτω: model70_0m, model70_1m, model70_2m, model80_0m, model80_1m, model80_2m (Ο δείκτης "_Xm" δηλώνει την χρονική υστέρηση των πραγματικών δεδομένων από τον μήνα δημοσίευσης του ΑΕΠ)

Όσον αφορά τα επόμενα βήματα για την περίπτωση της Ελλάδας και του ΑΕΠ, υπάρχει πληθώρα επιλογών. Για παράδειγμα, τι θα άλλαζε σε περίπτωση που συνθέταμε ένα μεγαλύτερο πάνελ μεταβλητών, συμπεριλαμβάνοντας τριμηνιαίες, εξαμηνιαίες και ετήσιες (μέθοδος MIDAS), αυξάνοντας το σύνολο περαιτέρω; Θα ήταν ωφέλιμο ή θα εισήγαγε θόρυβο οδηγώντας μας σε μεγαλύτερα σφάλματα; Επιπλέον, θα ήταν βάσιμο να συγκριθεί η Ελλάδα με κάποια τρίτη χώρα με την οποία έχει αλληλεξάρτηση εισαγωγών για παράδειγμα; Ή αναφορικά με το παρόν πείραμα, μήπως θα μας συνέφερε να δημιουργήσουμε κύριες συνιστώσες αφότου δοθούν κατάλληλα βάρη στις χρονοσειρές ή να υπάρχει μια μέθοδος προεπιλογής χρονοσειρών βάσει κριτηρίων; Τέλος, για να εισάγω κάτι από την επιστήμη της Πληροφορικής, θα μπορούσαμε άραγε να εφαρμόσουμε νευρωνικά δίκτυα στις προβλέψεις μακροοικονομικών μεγεθών;

Βιβλιογραφία

Ξενογλώσση

Artis M., Banerjee An., Marcellino M. (2005), "Factor forecasts for the UK", Journal of forecasting, vol. 24, issue 4, p.279-298

Baffigi Al., Golinelli R., Parigi G. (2004), "Bridge models to forecast the euro area GDP", International Journal of Forecasting 20, p.447–460

Banerjeea An, Marcellino M. (2006), "Are there any reliable leading indicators for US inflation and GDP growth?", International Journal of Forecasting 22, p.137– 151

Boivina J., Ng S. (2005), "Are more data always better for factor analysis?", Journal of Econometrics 132, p.169–194

Breitung J., Eickmeier S. (2005), "Dynamic factor models", Deutsche Bundesbank, Discussion Paper, Series 1: Economic Studies, No 38/2005

Callen T. (2017), "Gross Domestic Product: An Economy's All". IMF. Retrieved from <http://www.imf.org/external/pubs/ft/fandd/basics/gdp.htm>. Cimino, JJ (1998)

Dahla C., Hansenb H., Smidtc J. (2009), "The cyclical component factor model", International Journal of Forecasting 25, p.119–127

Diron M. (2006), "Short-term forecasts of euro area real GDP growth. An assessment of real-time performance based on vintage data", ECB Working Paper Series, NO 622 / MAY 2006

Favero C., Marcellino M., Neglia F. (2005), "Principal components at work: The empirical analysis of monetary policy with large datasets", Innocenzo Gasparini Institute for economic research, Journal of Applied Econometrics, vol. 20, Issue5, p.603-620

Forni M., Hallin M., Lippi M., Reichlin L. (2000), "The generalized factor model: identification and estimation", The Review of Economic and Statistics, vol.82, issue 4, p.540-554.

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2004). "The MIDAS touch: mixed data sampling regression models". University of North Carolina. CIRANO working papers, 2004s-20

Giannone D., Reichlin L., Small D. (2005), "Nowcasting GDP and Inflation: The Real-Time Informational Content of Macroeconomic Data Releases", ECB working paper series No 633/2006

IMF (2012), "Statistics on the Growth of the Global Gross Domestic Product (GDP) from 2003 to 2013"

Iacoviello M. (2001), "Short-term fore-casting: Projecting Italian GDP, one quarter to two years ahead", IMF working paper, WP/01/109

Kalekar P. (2004), "Time series Forecasting using Holt-Winters Exponential Smoothing" Retrieved from: <https://labs.omniti.com/>

Kuzina V., Marcellino M., Schumacher C. (2011), "MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area", International Journal of Forecasting 27, p.529–542

Monokroussos P., Thomakos D. (2012), "Eurobank GDP Now-casting model", Eurobank Global Markets Research Nov. 2012

Reijer Ar. (2005), "Forecasting Dutch GDP using large scale factor models", DNB Working Paper No. 028/2005, February 2005

Schumacher C. (2010), "Factor forecasting using international targeted predictors: The case of German GDP", Economics Letters 107, p. 95–98

Stakenas J. (2012), "Generating short-term forecasts of the Lithuanian GDP using factor models", ECB Working Paper Series No 13 / 2012

Stock J., Watson M. (1998), "Diffusion indexes", NBER WP 6702.

Ιστοσελίδες

<http://arch.readthedocs.io/en/latest/unitroot/tests.html> (last visit: 01/05/2018)

<http://iobe.gr/> (last visit: 03/05/2018)

<http://scikit-learn.org/stable/> (last visit: 01/05/2018)

<http://www.numpy.org> (last visit: 01/05/2018)

<http://www.oecd.org>, (last visit: 10/04/2018)

<http://www.statistics.gr/> (last visit: 03/05/2018)

<https://anaconda.org/> (last visit: 01/05/2018)

<https://ec.europa.eu/> (last visit: 03/05/2018)

<https://matplotlib.org/> (last visit: 01/05/2018)

<https://pandas.pydata.org/> (last visit: 01/05/2018)

<https://pythonhosted.org/spyder/> (last visit: 01/05/2018)

<https://www.bankofgreece.gr> (last visit: 03/05/2018)

<https://www.ecb.europa.eu/home/html/index.en.html> (last visit: 03/05/2018)

<https://www.python.org/> (last visit: 01/05/2018)

<https://www.statsmodels.org> (last visit: 01/05/2018)

Παράρτημα

I. Αποτελέσματα προβλέψεων

	model70	model80	real gdp	months back
1/9/2017	44493.42019	45057.72919	44205	1
1/9/2017	44478.70669	44995.1812	44205	2
1/9/2017	44621.01102	45004.27402	44205	3
1/6/2017	43468.01832	43443.44044	44135	4
1/6/2017	43588.12862	43540.16622	44135	5
1/6/2017	43555.59314	43572.79955	44135	6
1/3/2017	43618.89097	43866.98494	43332	7
1/3/2017	43582.6598	43728.69436	43332	8
1/3/2017	43636.14634	43790.98532	43332	9
1/12/2016	44635.28128	44920.647	43550	10
1/12/2016	44307.11244	44511.35618	43550	11
1/12/2016	44212.49436	44377.78623	43550	12
1/9/2016	44208.34712	44426.10699	43540	13
1/9/2016	44301.34582	44442.33289	43540	14
1/9/2016	44283.92241	44314.84988	43540	15
1/6/2016	43319.13201	43434.65759	43490	16
1/6/2016	43189.13164	43392.72238	43490	17
1/6/2016	43185.06938	43336.89114	43490	18
1/3/2016	44666.33597	44534.44346	43979	19
1/3/2016	44580.27366	44449.20317	43979	20
1/3/2016	44600.9821	44480.83797	43979	21
1/12/2015	45122.77036	45100.56736	43678	22
1/12/2015	44712.80273	44892.39738	43678	23
1/12/2015	44409.14926	44695.61955	43678	24
1/9/2015	45020.04954	44714.52383	44096	25
1/9/2015	44870.47998	44912.23732	44096	26
1/9/2015	44737.67174	44741.07433	44096	27

1/6/2015	44024.28005	43957.27596	44241	28
1/6/2015	44106.71854	44119.09002	44241	29
1/6/2015	44098.70071	44099.63647	44241	30
1/3/2015	45240.31593	45245.151	44477	31
1/3/2015	45190.23354	45324.62618	44477	32
1/3/2015	45081.50249	45160.83492	44477	33
1/12/2014	46222.19083	46316.72885	45198	34
1/12/2014	46148.8207	46514.43493	45198	35
1/12/2014	46043.45297	46371.73241	45198	36
1/9/2014	44848.28971	44800.96153	44291	37
1/9/2014	44812.49801	44710.50918	44291	38
1/9/2014	44769.56159	44694.10622	44291	39
1/6/2014	44108.38721	43860.25094	44431	40
1/6/2014	44196.88548	43983.11944	44431	41
1/6/2014	44153.29915	44001.73139	44431	42
1/3/2014	44655.32507	44880.64777	44120	43
1/3/2014	44845.01971	44920.84135	44120	44
1/3/2014	44749.67987	44933.03715	44120	45
1/12/2013	46053.70243	46218.03077	44928	46
1/12/2013	46335.33	46243.05359	44928	47
1/12/2013	46154.36971	46150.84958	44928	48
1/9/2013	45791.2381	45682.13099	45440	49
1/9/2013	45719.28115	45549.02713	45440	50
1/9/2013	45844.21522	45593.43443	45440	51
1/6/2013	45569.87019	45567.44411	46018	52
1/6/2013	45592.9894	45511.92258	46018	53
1/6/2013	45717.56912	45429.77552	46018	54
1/3/2013	48079.38443	48386.26144	46934	55
1/3/2013	47913.66982	48130.35194	46934	56
1/3/2013	48135.79421	48279.37375	46934	57
1/12/2012	49011.15344	48824.12828	47326	58
1/12/2012	49132.90859	49190.16468	47326	59

1/12/2012	49152.63774	49342.55823	47326	60
1/9/2012	48128.28631	48167.43124	47960	61
1/9/2012	48514.85465	48412.58718	47960	62
1/9/2012	48295.41952	48092.81331	47960	63
1/6/2012	49483.53278	49611.65861	49131	64
1/6/2012	49078.30599	49092.18641	49131	65
1/6/2012	48990.51307	49152.18944	49131	66
1/3/2012	51329.70056	50722.94632	49983	67
1/3/2012	51152.56878	51320.89327	49983	68
1/3/2012	51090.71489	51346.75851	49983	69
1/12/2011	53038.97422	52455.33846	51386	70
1/12/2011	53236.84029	52876.40614	51386	71
1/12/2011	53447.34742	53293.88977	51386	72
1/9/2011	53911.83281	54151.74787	52508	73
1/9/2011	53966.80244	53865.02389	52508	74
1/9/2011	53861.37156	54031.97456	52508	75
1/6/2011	53215.16651	52408.60074	53370	76
1/6/2011	53623.94134	53496.44132	53370	77
1/6/2011	53300.51023	53848.88221	53370	78
1/3/2011	56312.8575	55979.40082	54820	79
1/3/2011	55986.81853	55534.65107	54820	80
1/3/2011	55771.24289	55778.14077	54820	81
1/12/2010	58109.54138	58159.25308	55541	82
1/12/2010	57743.87806	58112.92504	55541	83
1/12/2010	57494.91371	57989.60935	55541	84
1/9/2010	59669.4156	58658.42508	57341	85
1/9/2010	59331.99446	58557.67157	57341	86
1/9/2010	59077.17433	58196.02232	57341	87
1/6/2010	59149.84918	60472.31221	58934	88
1/6/2010	59059.29527	59642.96938	58934	89
1/6/2010	59133.51855	59883.51054	58934	90
1/3/2010	60089.70379	60401.60861	60023	91

1/3/2010	61278.63288	60905.42118	60023	92
1/3/2010	60736.56615	60820.48199	60023	93
1/12/2009	60981.39612	61014.22746	59125	94
1/12/2009	60060.54652	59878.96856	59125	95

II. Παράδειγμα κώδικα: PCA

```

"""
@author: Chris Kavallos
"""
import pandas as pd
import numpy as np
from sklearn.decomposition import PCA
from matplotlib import pyplot
m = 0
for m in range(0,96):
    fname = 'C:/Users/ckavall/Desktop/ready_for_pca/mydata_vert_ready_'+str(m)+'m_back.csv'
    df_in = pd.read_csv(fname,na_values=['NaN'],sep=';',index_col=0,parse_dates=True,encoding='utf-8')
    dfs = df_in.T
    val = np.around(dfs.values,3)
    pca = PCA(svd_solver='full')
    pca.fit(val)
    variance = pca.explained_variance_ratio_
    var=np.cumsum(np.round(pca.explained_variance_ratio_, decimals=3)*100)
    varsum2 = 0.7
    pca2 = PCA(n_components = varsum2,svd_solver = 'full')
    pca2.fit(val)
    variance2 = pca2.explained_variance_ratio_
    var2=np.cumsum(np.round(pca2.explained_variance_ratio_, decimals=3)*100)
    projected70 = pd.DataFrame(data=pca2.fit_transform(val), index=df_in.index)
    varsum3 = 0.8
    pca3 = PCA(n_components = varsum3,svd_solver = 'full')
    pca3.fit(val)
    variance3 = pca3.explained_variance_ratio_
    var3=np.cumsum(np.round(pca3.explained_variance_ratio_, decimals=3)*100)
    projected80 = pd.DataFrame(data=pca3.fit_transform(val), index=df_in.index)
    print(str(m)+'m_back ~ 70% var: '+str(len(var2))+' ~ 80% var: '+str(len(var3))+' ~ 100% var: '+str(len(var)))
    projected70.to_csv("C:/Users/ckavall/Desktop/post_pca/var70_"+str(m)+"m_back.csv",sep=';',encoding='utf-8')
    projected80.to_csv("C:/Users/ckavall/Desktop/post_pca/var80_"+str(m)+"m_back.csv",sep=';',encoding='utf-8')

```