- ΤΕΧΝΟΛΟΓΙΚΟ ΙΔΡΥΜΑ ΠΑΤΡΩΝ

- ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ
  ΤΜΗΜΑ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

• Η ΧΡΗΣΗ ΜΕΘΟΔΩΝ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ ΣΤΗ ΜΕΛΕΤΗ ΚΑΙ ΠΡΟΒΛΕΨΗ ΤΗΣ ΑΠΟΔΟΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ.

- ΟΝΟΜΑΤΕΠΩΝΥΜΟ ΣΠΟΥΔΑΣΤΡΙΑΣ

  ΒΙΤΩΡΑΤΟΥ ΚΩΝΣΤΑΝΤΙΝΑ

- ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ

  ΑΝΤΩΝΟΠΟΥΛΟΣ ΓΡΗΓΟΡΙΟΣ

ΠΑΤΡΑ - ΙΑΝΟΥΑΡΙΟΣ 2012

ΠΡΟΛΟΓΟΣ

Ευχαριστώ θερμά τον επιβλέποντα καθηγητή της πτυχιακής εργασίας κύριο Αντωνόπουλο Γρηγόριο, για την πολύτιμη και ουσιαστική βοήθεια που μου προσέφερε σε όλη τη διάρκεια εκπόνησης αυτής, καθώς και την καθοδήγησή του, τη φιλική στάση και τα ουσιαστικά σχόλιά του.

ΠΕΡΙΛΗΨΗ

Ένας από τους βασικούς σκοπούς της Οικονομετρικής ανάλυσης ενός οικονομικού φαινομένου, είναι η δημιουργία των προβλέψεων, η χρησιμοποίηση δηλαδή του εκτιμημένου οικονομικού υποδείγματος, για τη πρόβλεψη των μελλοντικών τιμών των οικονομικών μεγεθών. Η ανάγκη για έγκυρες προβλέψεις, οδήγησε στην ανάπτυξη και εκτίμηση όχι μόνο οικονομικών υποδειγμάτων μίας ή πολλών ταυτοχρόνων εξισώσεων, αλλά και στην ανάπτυξη και άλλων τεχνικών ή μεθόδων, όπως οι τεχνικές ανάλυσης χρονολογικών σειρών.

Η εργασία αυτή, ασχολείται με την ανάλυση των χρονολογικών σειρών και τη χρήση των μεθόδων τους, ως αντικείμενο για τη μελέτη και πρόβλεψη της απόδοσης των επιχειρήσεων. Στα πρώτα κεφάλαια, γίνεται μία εισαγωγή στις χρονολογικές σειρές, ποια η χρησιμότητά τους γενικά ως οικονομετρικό φαινόμενο, τί μπορούν αυτές να περιγράψουν και να προβλέψουν, καθώς αναλύονται ορισμένες βασικές έννοιες αυτών, όπως διάφορα μέτρα και στασιμότητα, και αναλύονται χαρακτηριστικά όπως η τάση, η περιοδικότητα κ.ά. Έπειτα εξετάζονται συγκεκριμένες σειρές χρονολογικών σειρών, όπως είναι οι στάσιμες και οι μη στάσιμες, και γίνεται εκτίμηση των παραμέτρων των παραπάνω σειρών με διάφορους μεθόδους. Τέλος, παρουσιάζονται διάφοροι μέθοδοι πρόβλεψης και γίνεται και εφαρμογή αυτών των μεθόδων.

ΕΙΣΑΓΩΓΗ

Θεωρία και πράξη είναι δύο αντιπαραβαλλόμενες έννοιες στη ζωή μας. Η χρήση των υπολογιστών ανέδειξε έντονα αυτό το πρότυπο στην επιστήμη. Στις οικονομικές επιστήμες και όχι μόνο, αντικειμενικός στόχος είναι αυτές οι δύο έννοιες να έρθουν όσο πιο κοντά είναι δυνατόν. Εργαλεία για την εργασία αυτή είναι κυρίως η προσομοίωση και η πρόβλεψη. Συνδετικός κρίκος της θεωρίας και της πράξης, είναι η δυνατότητα των οικονομετρικών και μαθηματικών εξισώσεων της θεωρίας να συμπέσουν με τις παρατηρήσεις προσομοίωσης της πράξης. Με άλλα λόγια, μέσω των κατάλληλων μοντέλων χρονολογικών σειρών, η δυνατότητα κατανόησης του παρελθόντος, να μας οδηγήσει στη μελέτη και πρόβλεψη της απόδοσης μίας επιχείρησης.

Η μεγάλη σημασία που έχει για την Οικονομική Επιστήμη η ανάλυση των χρονολογικών σειρών και ο λόγος που διεξάγεται η εργασία αυτή, προέρχεται από το γεγονός ότι μέσω αυτής είναι δυνατό, μέσα σε ορισμένα όρια και με ορισμένες προφυλάξεις, να διατυπώνονται προβλέψεις για μελλοντικές εξελίξεις φαινομένων. Η σημασία αυτών των προβλέψεων φαίνεται από το γεγονός ότι καμιά επιχείρηση, δημόσια ή ιδιωτική, δεν μπορεί να αποφύγει τη διατύπωση προγραμμάτων για το μέλλον.

Επιπλέον, βασικός στόχος κάθε ανθρώπινης δραστηριότητας είναι η διενέργεια προβλέψεων, δηλαδή η χρησιμοποίηση ενός εκτιμώμενου μοντέλου για την πρόβλεψη των μελλοντικών τιμών των μεγεθών. Η πρόβλεψη είναι απαραίτητη, γιατί αποτελεί το βασικό εργαλείο για κάθε μελλοντική εξέλιξη και απόφαση που θα αυξήσει την απόδοση της επιχείρησης. Παράλληλα, οι προβλέψεις που γίνονται θα πρέπει να είναι έγκυρες και όσο πιο κοντά γίνεται στο μέλλον που θα προκύψει. Η ποιότητα της πρόβλεψης εξαρτάται κυρίως από τον τρόπο που συλλέχθηκαν και αναλύθηκαν οι πληροφορίες. Έτσι, η ανάγκη για έγκυρες, άρα και πρακτικά χρήσιμες προβλέψεις οδήγησε στην ανάπτυξη πολλών τεχνικών και μεθόδων συλλογής και επεξεργασίας διαφόρων πρωτογενών πληροφοριών. Στην παρούσα εργασία θα μελετηθούν αποκλειστικά μέθοδοι επεξεργασίας των πρωτογενών αυτών πληροφοριών.

Τέτοιοι μέθοδοι, λοιπόν, προβλέψεων μπορεί να είναι ποιοτικές ή ποσοτικές. Με τις ποιοτικές μεθόδους ασχολούνται επιστήμες όπως η φιλοσοφία, η θεολογία, η οικονομία, οι πολιτικές επιστήμες κ.α., οι οποίες για την ανάλυση των δεδομένων χρησιμοποιούν την 'ανθρώπινη' λογική και τον τρόπο με τον οποίο οι διάφορες σχολές ανάλυσης αντιλαμβάνονται τον φυσικό κόσμο. Με τον όρο 'ανθρώπινη' λογική εννοείται κάθε απόφαση που λαμβάνει κανείς μη χρησιμοποιώντας μαθηματικές εξισώσεις. Για παράδειγμα, πρόβλεψη, ή καλύτερα εκτίμηση για την εξέλιξη μιας γλώσσας, της πολιτικής κατάστασης μιας χώρας, την πορεία της Ευρωπαϊκής Ένωσης και άλλα, είναι αποτέλεσμα λογικών διεργασιών σε διάφορα φόρουμ σκέψης. Σε όλες τις παραπάνω προβλέψεις δε χρειάζεται καμία μαθηματική ή οικονομετρική επεξεργασία. Αντίθετα, στις ποσοτικές μεθόδους η ανάλυση των δεδομένων γίνεται με την βοήθεια μαθηματικών εκφράσεων υποστηριζόμενες από υπολογιστές. Μία από αυτές τις ποσοτικές μεθόδους, είναι η ανάλυση χρονολογικών σειρών.

Με τον όρο χρονολογικές σειρές, εννοούμε μία σειρά από παρατηρήσεις που λαμβάνονται σε ορισμένες χρονικές στιγμές ή περιόδους που ισαπέχουν ή μη μεταξύ τους. Έτσι, η τρέχουσα τιμή μιας μεταβλητής Y εκφράζεται ως συνάρτηση των προηγούμενων τιμών της, δηλαδή των τιμών της με χρονική υστέρηση.

Μία ευρεία οικογένεια στοχαστικών υποδειγμάτων γνωστών στην διεθνή βιβλιογραφία ως υποδείγματα ARIMA, τα οποία θα αναλυθούν και θα χρησιμοποιηθούν στην εργασία αυτή για την διεξαγωγή και κατάληξη συμπερασμάτων για τη μελέτη και πρόβλεψη της απόδοσης των επιχειρήσεων, έχουν εισαχθεί και μελετηθεί από τους σύγχρονους μελετητές BOX-Jenkins. Η δύναμη των υποδειγμάτων αυτών έγκειται στο γεγονός ότι αυτά μπορούν να περιγράψουν μια οποιαδήποτε χρονολογική σειρά, η οποία ήθελε παρουσιαστεί στη πράξη. Εξάλλου, η αναζήτηση και κατασκευή υποδειγμάτων είναι ενδιαφέρουσα, διότι τα υποδείγματα μπορούν να αποκαλύψουν την υπάρχουσα νομοτέλεια η οποία υφίσταται μεταξύ διαδοχικών παρατηρήσεων μιας μεταβλητής, να δώσουν άριστες προβλέψεις για τη μελλοντική συμπεριφορά της και συνεπώς να χρησιμεύσουν σημαντικά στην διαδικασία της λήψης αποφάσεων μιας επιχείρησης.

Η ανάπτυξη και η χρησιμοποίηση τέτοιων μοντέλων υπήρξε ραγδαία τις τελευταίες τρείς δεκαετίες, ιδίως μετά την δημοσίευση τη εργασίας των Box&Jenkins. Οι προβλέψεις με χρονολογικές σειρές αποδείχτηκαν πολλές φορές ανώτερες άλλων παλαιότερων μεθόδων. Επίσης, προβλέψεις με χρονολογικές σειρές μπορούν να γίνουν σε ένα μεγάλο πεδίο επιστημών, από την οικονομετρία, την μηχανική, την φυσική, την ιατρική και πολλά άλλα. Βέβαια, στο σημείο αυτό είναι αναγκαίο να τονιστεί ότι οι χρονολογικές σειρές δεν χρησιμοποιούνται αποκλειστικά και μόνο για την εκτίμηση μελλοντικών τιμών, αλλά είναι ένα πολύ χρήσιμο εργαλείο για την ανάλυση και καλύτερη κατανόηση αυτού κάθε αυτού του φαινομένου. Οι χρονολογικές σειρές καταγράφουν την ιστορία των φαινομένων και όπως είναι γνωστό η ιστορία είναι αυτή που διδάσκει και βοηθά για την καλύτερη κατανόηση των γεγονότων. Η ιστορία είναι αυτή που με την καλύτερη ανάλυση μπορεί να βοηθήσει για την καλύτερη εκτίμηση του μέλλοντος.

Στόχος λοιπόν των χρονολογικών σειρών είναι, ενώνοντας όλα τα γεγονότα που έχουν συλλεχθεί στο χρόνο, να δημιουργηθεί ένα χρονοδιάγραμμα, η μελέτη του οποίου θα δώσει μία γενική εικόνα της διαχρονικής εξέλιξης των φαινομένων ή χαρακτηριστικών της εκάστοτε επιχείρησης. Με άλλα λόγια, η ανάλυση χρονολογικών σειρών χρησιμοποιείται για να καθοριστούν μοντέλα που μετατρέπουν πληροφορίες από κανονικά χρονικά διαστήματα σε στατιστικά μέτρα. Οι κυριότερες μέθοδοι ανάλυσης χρονοσειρών είναι η μέθοδος της αυτοσυσχέτισης (auto correlation) όπου η χρονοσειρά αναπαρίσταται με ένα δυναμικό μοντέλο (μοντέλο ARIMA, διαδικασία Box& Jenkins), όπου οι παρατηρήσεις θεωρούνται ως συναρτήσεις του παρελθόντος τους(και πιθανόν του παρελθόντος και άλλων μετρούμενων ή παρατηρούμενων μεταβλητών) και η μέθοδος φασματικής ανάλυσης (spectral analysis) όπου αναπαριστά την χρονοσειρά με ένα κινητικό μοντέλο που οι παρατηρήσεις θεωρούνται συναρτήσεις του χρόνου $x= f(t)$. Στην παρούσα εργασία θα παρουσιαστούν μέθοδοι αυτοσυσχέτισης για την πρόβλεψη της απόδοσης μιας επιχείρησης.

Συμπερασματικά, γίνεται κατανοητό από τα παραπάνω, ότι σκοποί της ανάλυσης χρονολογικών σειρών είναι η περιγραφή μιας χρονολογικής σειράς, η εξήγησή της, η πρόγνωση, ο σχεδιασμός, ο έλεγχος και τέλος η κατανόηση της διαδικασίας η οποία θα μας οδηγήσει στα κατάλληλα συμπεράσματα σχετικά με τη μελλοντική πορεία της επιχείρησης.

Η περιγραφή μιας χρονολογικής σειράς μπορεί να γίνει με μια απλή παρατήρηση της γραφικής παράστασης 'με το μάτι', όπου ανακαλύπτονται περιοδικότητες, τάσεις και άλλα. Επίσης, υπάρχουν περισσότερο σύνθετες και προχωρημένες περιγραφές που γίνονται με τη βοήθειας στοχαστικών μοντέλων. Η εξήγηση με τη σειρά της, είναι η συσχέτιση δύο ή περισσοτέρων χρονολογικών σειρών μεταξύ τους και η δυνατότητα η μία να εξηγεί την άλλη. Παραδείγματος χάριν, η συσχέτιση του όγκου πωλήσεων και του όγκου των διαφημίσεων μίας εταιρείας σε ευρώ για ένα συγκεκριμένο χρονικό διάστημα, δίνουν απάντηση και εξήγηση για την καλύτερη προώθηση των προϊόντων της. Επίσης, σημαντικός σκοπός στην ανάλυση των χρονολογικών σειρών είναι ο έλεγχος. Παραδείγματα στην σπουδαιότητα του ελέγχου μπορούν να παρουσιαστούν στις διαδικασίες παραγωγής και ποιότητας ενός προϊόντος. Τέλος, ένας ακόμη πολύ βασικός σκοπός της ανάλυσης των χρονολογικών σειρών είναι η κατανόηση της διαδικασίας. Την κατανόηση παρέχει η στατιστική δίνοντας φορμαλιστικές περιγραφές των χρονοσειρών καθαυτών. Έτσι, μπορεί να παραχθεί μία χρονολογική σειρά από ένα άγνωστο σύστημα και συνεπώς να κατανοηθεί καλύτερα το ίδιο το σύστημα που έχει παράξει τη χρονολογική σειρά αυτή.

Κύριο ενδιαφέρον όμως για την Οικονομική Επιστήμη και ειδικότερα για τη Διοίκηση Επιχειρήσεων, παρουσιάζει η πρόβλεψη. Παραδείγματα χρονολογικών σειρών που βοηθούν στη πρόβλεψη της μελλοντικής εξέλιξης και συμπεριφοράς μιας επιχείρησης είναι οι πωλήσεις ανά μήνα. Επίσης, βασικός σκοπός στη Διοίκηση Επιχειρήσεων και στο marketing εκτός της πρόβλεψης είναι ο σχεδιασμός, η σύγκριση και η ερμηνεία, με απώτερο σκοπό τη βελτιστοποίηση των διαδικασιών αλλά και την αποφυγή λαθών στο άμεσο και έμμεσο μέλλον της επιχείρησης.

Σκοπός λοιπόν της εργασίας αυτής, είναι να αναδείξει τη σπουδαιότητα και τη χρησιμότητα των χρονολογικών σειρών στη μελέτη αλλά και στην πρόβλεψη της μελλοντικής απόδοσης των επιχειρήσεων.

ΚΕΦΑΛΑΙΟ 2^{o}

Θα ασχοληθούμε αρχικά, με τις **στάσιμες στοχαστικές διαδικασίες**. Κάθε στάσιμη στοχαστική διαδικασία μπορεί να εκφραστεί ως γραμμικός συνδυασμός μιας ακολουθίας ασυσχέτιστων τυχαίων μεταβλητών. Ένας τέτοιος γραμμικός συνδυασμός είναι επίσης γνωστός ως *γραμμικό φίλτρο* (linear filter).

Έστω $[Y_t]$, μια όχι αναγκαστικά αυστηρώς στάσιμη στοχαστική διαδικασία με μέσο μ. Το γραμμικό φίλτρο θα μπορούσε να διατυπωθεί ως εξής:

$$Y_t-\mu=\varepsilon_t+\Psi_1\varepsilon_{t-1}+\Psi_2\varepsilon_{t-2}+\ldots.. \qquad (2.1)$$

Αν θέσουμε $\Psi_0 = 1$, τότε η παραπάνω σχέση μπορεί να γραφεί και ως :

$$Y_t - \mu=\sum_{i-0}^{\infty}\Psi_i e_{\ t-i} \qquad (2.2)$$

<u>Ορισμός:</u> Υποθέτουμε ότι η ακολουθία $\{\varepsilon_t\}$ *για* $t = 0\pm1\pm2,\ldots..$ είναι μία ακολουθία τυχαίων μεταβλητών για την οποία ισχύουν οι παρακάτω προϋποθέσεις για κάθε t:

1. $E(\varepsilon_t) = 0$

2. $V(\varepsilon_t) = \sigma^2$

3. $Cov(\varepsilon_t, \varepsilon_{t-s}) = 0$, για κάθε $s\neq 0$

Μία τέτοια ακολουθία για την οποία ισχύουν οι τρεις αυτές προϋποθέσεις, ονομάζεται διαδικασία λευκού θορύβου (white noise process) ή απλώς λευκός θόρυβος.

Παρατήρηση : Οι συντελεστές Ψ_i είναι γνωστοί και ως συντελεστές στάθμισης και το πλήθος τους μπορεί να είναι άπειρο ή πεπερασμένο. Αν είναι άπειρο, υποθέτουμε ότι το άθροισμά τους συγκλίνει απολύτως, δηλαδή θα ισχύει η σχέση :

$$\sum_{i-0}^{\infty}\left|\Psi_i\right| <\infty \qquad (2.3)$$

Στη συνέχεια, θα δούμε πώς υπολογίζονται η συνδιακύμανση (ή αυτοδιακύμανση), γ_0 και γ_s των παρατηρήσεων Y_t και Y_{t+s} καθώς επίσης και ο συντελεστής αυτοσυσχέτισης :

$$\rho_s = \frac{Cov(Y_t, Y_{t+s})}{\sqrt{V(Y_t)V(Y_{t+s})}} = \frac{g_s}{g_0} \qquad (2.4)$$

Από τη σχέση (2.1) έχουμε ότι :

$\gamma_0 = V(Y_t) = E\,(Y_t - \mu)^2 = E\,(\varepsilon_t + \Psi_1\varepsilon_{t-1} + \Psi_2\varepsilon_{t-2} + \ldots.) =$

$$= E(e_t^2) + \Psi_1^2 E(e_{t-1}^2) + \Psi_2^2 E(e_{t-2}^2) + \ldots\ldots \qquad (2.5)$$

Σύμφωνα με τις υποθέσεις που ισχύουν για το λευκό θόρυβο, $E(\varepsilon_t\varepsilon_{t-s}) = 0$, για κάθε $s \neq 0$.

Άρα,

$\gamma_0 = \sigma^2 + \Psi_1^2\sigma^2 + \Psi_2^2\sigma^2 + \ldots =$

$$= \sigma^2 \quad \begin{aligned} & Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + \ldots + a_p Y_{t-p} e_t \quad ^2 \\ & Y_t = a_0 + a_1 X_t \\ & u_t = Y_t - a_0 - a_1 X_t \\ & \hat{u}_t = Y_t - \hat{a}_0 - \hat{a}_1 X_t \\ & \Delta\hat{u}_t = r^* \hat{u}_{t-1} + e_t \\ & H_0 : r = 1 \end{aligned}$$

$\gamma_s = E(\, Y_t - \mu\,)\,(\, Y_{t-s} - \mu\,) =$

$=E\ (\varepsilon_t + \Psi_1\varepsilon_{t-1} + \Psi_2\varepsilon_{t-2} + \ldots + \Psi_s\varepsilon_{t-s} + \Psi_{s=1}\varepsilon_{t-s-1} + \ldots)\ x$

$(\varepsilon_{t-s} + \Psi_1\varepsilon_{t-s-1} + \Psi_2\varepsilon_{t-s-2} + \ldots + \Psi_s\varepsilon_{t-s-s} + \Psi_{s+1}\varepsilon_{t-s-s-1} + \ldots) =$

$=E\ (\Psi_s e^2_{t-s}) + E\ (\Psi_1\ \Psi_2 e^2_{t-s-1}) + \ldots =$

$= \sigma^2 \sum_{i-0}^{\infty} \Psi_i y_{\ i+s}$

Όλοι οι άλλοι όροι είναι μηδέν, σύμφωνα με την υπόθεση ότι $Cov\ (\varepsilon_{t,}\ \varepsilon_{t-s}) = 0$.

$$\rho_S = \frac{g_s}{g_0} = \frac{\sum_{i-1}^{\infty} \Psi_i \Psi_{i+s}}{\sum_{i-0}^{\infty} \Psi_i^2}$$

Από τις παραπάνω σχέσεις, είναι φανερό ότι το άθροισμα των συντελεστών σταθμίσεως δεν συγκλίνει, τότε η διακύμανση θα τείνει στο άπειρο και ο συντελεστής αυτοσυσχέτισης θα τείνει στο μηδέν. Αν η σειρά είναι στάσιμη, τότε η διακύμανση θα είναι πεπερασμένη.

Η εξίσωση γραμμικού φίλτρου : $Y_t - \mu = \varepsilon_t + \Psi_1\varepsilon_{t-1} + \Psi_2\varepsilon_{t-2} + \ldots$, αποτελεί μία γενική μορφή από την οποία με διάφορες υποθέσεις σχετικά με τους συντελεστές στάθμισης προκύπτουν διάφορα στοχαστικά υποδείγματα χρονολογικών σειρών.

Γενικά, υπάρχουν τρείς βασικές κατηγορίες στοχαστικών υποδειγμάτων χρονολογικών σειρών. Αυτά είναι :

1. Τα Αυτοπαλίνδρομα Υποδείγματα ή Υποδείγματα AR (Autogressive Models)
2. Τα Υποδείγματα Κινητών Μέσων ή Υποδείγματα MA (Moving Average Models)

3. Τα Μεικτά Υποδείγματα ή Υποδείγματα ARMA (Autogressive Moving Average Models) που είναι συνδυασμός των δύο προηγούμενων.

Στη συνέχεια, θα εξετάσουμε αναλυτικά κάθε μία από αυτές τις τρει΄ς κατηγορίες υποδειγμάτων χρονολογικών σειρών.

2.1 Αυτοπαλίνδρομα Υποδείγματα AR(p)

Η γενική μορφή ενός αυτοπαλίνδρομου υποδείγματος *p* τάξεως, ή αλλιώς ενός AR(*p*) είναι:

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + ... + a_p Y_{t-p} e_t$$

Η τάξη *p* αναφέρεται στο μήκος της χρονικής υστέρησης. Η μορφή ενός αυτοπαλίνδρομου υποδείγματος μας παραπέμπει ουσιαστικά σε ένα υπόδειγμα παλινδρόμησης, όπου οι ερμηνευτικές μεταβλητές είναι οι τιμές της εξαρτημένης μεταβλητής Y_t με χρονική υστέρηση. Εξαιτίας αυτού του γεγονότος επικράτησε ο όρος *αυτοπαλίνδρομο*. Η μεταβλητή e_t θεωρείται ότι είναι *λευκός θόρυβος*.

- Στη συνέχεια, θα εξετάσουμε πιο αναλυτικά τη γενική μορφή και τις ιδιότητες ενός αυτοπαλίνδρομου υποδείγματος τάξεως 1, δηλαδή το **AR(1).** Η γενική μορφή ενός AR(1) θα είναι :

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \varepsilon_t \qquad \textbf{(1)}$$

Αν υποθέσουμε ότι ο μέσος είναι μηδέν είτε οι μεταβλητές εκφράζονται ως αποκλίσεις από τους μέσους, τότε η σχέση (1) γράφεται :

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t \qquad \textbf{(2)}$$

όπου $y_t = Y_t - \bar{Y}_t$ και $y_{t-1} = Y_{t-1} - \bar{Y}_{t-1}$

Για το υπόδειγμα AR(1) ισχύουν επίσης οι παρακάτω σχέσεις:

a. $\mu = \dfrac{a_0}{1-a_1}$

b. $\gamma_0 = V(y_t) = \sigma^2 \dfrac{1}{1-a_1^2}$

c. $\gamma_s = Cov(y_t, y_{t-s}) = a_1^s \gamma_0$

d. $\rho_s = \dfrac{g_s}{g_0} = a_1^s$

Η **<u>απόδειξη</u>** των παραπάνω σχέσεων είναι απλή :

a. $\alpha_0 = (1-\alpha_1)\mu$

$\mu = \dfrac{a_0}{1-a_1}$

b. Υψώνουμε στο τετράγωνο και τα 2 μέλη της σχέσης (2) και στη συνέχεια παίρνουμε τις μέσες τιμές ως εξής :

$E(y_t^2) = E(\alpha_1 y_{t-1} + \varepsilon_t)^2$

$E(y_t^2) = a_1^2\ E y_{t-1}^2 + 2\alpha_1 E y_{t-1}\varepsilon_t + E e_t^2$

$V(y_t) = a_1^2 V(y_t) + \sigma^2$

$V(y_t) - a_1^2 V(y_t) = \sigma^2$

$(1-a_1^2)\ V(y_t) = \sigma^2$

$V(y_t) = \dfrac{s^2}{(1-a_1^2)} = \gamma_0$

διότι, $Ey_{t-1}\varepsilon_t = 0$ αφού το y_{t-1} εξαρτάται μόνο από το ε_{t-1}, το οποίο είναι λευκός θόρυβος και $E\,y_t^2 = E\,y_{t-1}^2 = V(y_t)$

c. Ξεκινάμε πάλι από τη σχέση (2), πολλαπλασιάζουμε και τα δύο της μέλη με το y_{t-s} και παίρνουμε τις μέσες τιμές:

$$y_t\, y_{t-s} = \alpha_1 y_{t-1}\, y_{t-s} + \varepsilon_t\, y_{t-s}$$

$$E(y_t\, y_{t-s}) = \alpha_1 E(y_t\, y_{t-s}) + E(y_t\, y_{t-s})$$

$$\gamma_s = \alpha_1 \gamma_0, \qquad \text{για } s > 0$$

διότι, $E(y_t\, y_{t-s}) = 0$ και $E(y_t\, y_{t-s}) = \gamma_{s-1}$.

Από την παραπάνω σχέση προκύπτει ότι:

$\gamma_1 = \alpha_1 \gamma_0$

$\gamma_2 = \alpha_1 \gamma_1 = a_1^2 \gamma_0$

$\gamma_3 = \alpha_1 \gamma_3 = a_1^3 \gamma_0$

και γενικά:

$\gamma_s = a_1^s \gamma_0$

d. $\rho_s = \dfrac{g_s}{g_0} = \dfrac{a_1^s g_0}{g_0} = a_1^s$

Επειδή αρχικά θα ασχοληθούμε με αυστηρά στάσιμες διαδικασίες, θα πρέπει να εξετάσουμε κάτω από ποιες προϋποθέσεις η χρονολογική σειρά που παριστάνεται με τη γενική μορφή ενός AR(1) θα είναι στάσιμη. Για να είναι η σειρά στάσιμη θα πρέπει $|\alpha_1| < 1$. Αυτό συμβαίνει γιατί, για να είναι μία σειρά στάσιμη θα πρέπει η διακύμανση γ_0 να είναι ένας σταθερός θετικός αριθμός. Για $\alpha_1 > 0$, η συνάρτηση αυτοσυσχέτισης φθίνει γεωμετρικά και τείνει προς το μηδέν καθώς και το s αυξάνει.

Ανάλογα, όταν $\alpha_1 < 0$ η συνάρτηση αυτοσυσχέτισης πάλι θα τείνει προς το μηδέν αλλά με εναλλασσόμενο πρόσημο αυτή τη φορά.

v Στη συνέχεια θα ασχοληθούμε με το αυτοπαλίνδρομο υπόδειγμα δεύτερης τάξεως, δηλαδή το **AR(2).** Στη γενική του μορφή μπορεί να γραφεί ως εξής :

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \varepsilon_t \qquad \textbf{(3)}$$

ή ως $y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \varepsilon_t$ **(4)**

Για το AR(2) ισχύουν τα εξής:

1. $\mu = \dfrac{a_0}{1 - a_1 - a_2}$

2. $\gamma_0 = V(Y_t) = \alpha_1\gamma_1 + \alpha_2\gamma_2 + \sigma^2$ **(5)**

$$\gamma_0 = \frac{(1-a_2)}{(1-a_2)(1-a_1-a_2)(1+a_1-a_2)} s^2 \qquad \textbf{(6)}$$

3. $\gamma_s = Cov(Y_t, Y_{t-s}) = \alpha_1\gamma_{s-1} + \alpha_2\gamma_{s-2}$ για $s > 0$

4. $\rho_s = a_1\rho_{s-1} + \alpha_2\rho_{s-2}$ για $s > 0$

Οι **αποδείξεις** αυτών των σχέσεων δίνονται παρακάτω.

1. Στη σχέση **(3)** παίρνουμε τις μέσες τιμές οπότε έχουμε:

$E(Y_t) = \alpha_0 + \alpha_1 E(Y_{t-1}) + \alpha_2 E(Y_{t-2}) + E(\varepsilon_t)$

$\mu = \alpha_0 + \alpha_1\mu + \alpha_2\mu$

μ = $g_s = a_1 g_{s-1} + a_2 g_{s-2} + \ldots + a_p r_p$ αφού Ε (ε_t) = 0 , ως λευκός θόρυβος.

2. Πολλαπλασιάζουμε και τα δύο μέλη της σχέσης **(4)** με το y_t , και στη συνέχεια παίρνουμε τις μέσες τιμές οπότε έχουμε:

$y_t^2 = \alpha_1 y_t y_{t-1} + \alpha_2 y_t y_{t-2} + y_t \varepsilon_t$

$E\, y_t^2 = V(y_t) = \alpha_1 E y_t y_{t-1} + \alpha_2\, E\, y_t y_{t-2} + E y_t \varepsilon_t$

$\gamma_0 = \alpha_1 \gamma_1 + \alpha_2 \gamma_2 + E y_t \varepsilon_t$

$\gamma_0 = \alpha_1 \gamma_1 + \alpha_2 \gamma_2 + \sigma^2$, διότι $E y_t \varepsilon_t = E\,(\alpha_1\, y_{t-1} + \alpha_2\, y_{t-2} + \varepsilon_t\,)\, \varepsilon_t = \sigma^2$

3. Πολλαπλασιάζουμε και τα δύο μέλη της σχέσης **(4)** με το y_{t-s} , και στη συνέχεια παίρνουμε τις μέσες τιμές:

$y_t\, y_{t-s} = \alpha_1\, y_{t-1}\, y_{t-s} + \alpha_2\, y_{t-2}\, y_{t-s} + \varepsilon_t\, y_{t-s}$

$E\, y_t\, y_{t-s} = Cov\,(Y_t\, ,\, Y_{t-s}\,) = \alpha_1 \gamma_{s-1} + \alpha_2 \gamma_{s-2}$

Αφού $E\varepsilon_t\, y_{t-s} = 0$ και $E\, y_{t-1}\, y_{t-s} = \gamma_{s-1}$, $E\, y_{t-2}\, y_{t-s} = \gamma_{s-2}$ εξ ορισμού.

4. Διαιρούμε την παραπάνω σχέση με το γ_0 οπότε :

$$\frac{g_s}{g_0} = \alpha_1 \frac{g_{s-1}}{g_0} + \alpha_2 \frac{g_{s-2}}{g_0}$$

$\rho_s = \alpha_1 \rho_{s-1} + \alpha_2 \rho_{s-2}$

Για να είναι *στάσιμη* αυτή η χρονολογική σειρά, θα πρέπει να ικανοποιούνται οι παρακάτω προϋποθέσεις :

$\alpha_1 + \alpha_2 < 1$

$-\alpha_1 + \alpha_2 < 1$ **(7)**

$-1 < \alpha_2 < 1$

Οι σχέσεις **(7)** πρέπει να ικανοποιούνται για τον εξής λόγο : Για να είναι η σειρά στάσιμη, θα πρέπει η διακύμανση γ_0 να είναι ένας σταθερός αριθμός. Για να συμβαίνει αυτό θα πρέπει κάθε όρος στις παρενθέσεις της σχέσης **(6)** να είναι θετικός αριθμός. Δηλαδή θα πρέπει να ισχύουν οι ανισότητες :

$1 - \alpha_2 > 0$

$1 + \alpha_2 > 0$

$1 - \alpha_1 - \alpha_2 > 0$

$1 + \alpha_1 - \alpha_2 > 0$

Συναληθεύοντάς τες προκύπτουν οι σχέσεις **(7)** οπότε και αποτελούν τις προυποθέσεις για να είναι η σειρά στάσιμη.

Από τη σχέση των αυτοσυσχετίσεων $\rho_s = \alpha_1\rho_{s-1} + \alpha_2\rho_{s-2}$ προκύπτουν οι σχέσεις:

$$\rho_1 = \frac{a_1}{1-a_2}$$

$$\rho_2 = \alpha_2 + \frac{a_1^2}{1-a_2}$$

Αντίθετα, αν είναι γνωστές οι τιμές των αυτοσυσχετίσεων p_1 και p_2, μπορούμε να βρούμε τις τιμές των συντελεστών.
Η συνάρτηση αυτοσυσχέτισης μιας AR(2) διαδικασίας θα τείνει προς το μηδέν, καθώς αυξάνεται η χρονική υστέρηση s.
Οι αυτοσυσχετίσεις ρ_s για $s > 2$ μπορούν να υπολογιστούν από τη σχέση $\rho_s = \alpha_1\rho_{s-1} + \alpha_2\rho_{s-2}$, αφού προηγουμένως έχουν υπολογιστεί οι δύο πρώτες αυτοσυσχετίσεις ρ_1 και ρ_2 από το σύστημα εξισώσεων Yule – Walker.

- Τέλος, θα ασχοληθούμε και με το αυτοπαλίνδρομο υπόδειγμα τάξεως p, δηλαδή το AR(p). Η γενική του μορφή δίνεται από τον τύπο:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \ldots + \alpha_p Y_{t-p} + \varepsilon_t$$

Για μία διαδικασία AR(p) ισχύουν οι παρακάτω σχέσεις:

$$\gamma_0 = \sigma^2 + \alpha_1\gamma_1 + \alpha_2\gamma_2 + \ldots + \alpha_p\gamma_p = \frac{s^2}{1-a_1 r_1 - a_2 r_2 - \ldots - a_p r_p}$$

$$g_s = a_1 g_{s-1} + a_2 g_{s-2} + ... + a_p r_p \qquad \text{για } s > 0$$

$$\rho_s = \alpha_1 \rho_{s-1} + \alpha_{s-2} + \ldots + \alpha_p \rho_{s-p} \qquad \text{για } s > 0$$

Από τη τελευταία σχέση για $s = 1, 2, \ldots, p$ προκύπτουν p εξισώσεις Yule – Walker οι οποίες είναι οι εξής :

$\rho_1 = \alpha_1 + \rho_1\alpha_2 + \rho_2\alpha_3 + \ldots + \rho_{p-1}\alpha_p$

$\rho_2 = \rho_1\alpha_1 + \alpha_2 + \rho_1\alpha_3 + \ldots + \rho_{p-2}\alpha_p$

$\rho_3 = \rho_2\alpha_1 + \rho_1\alpha_2 + \alpha_3 + \ldots + \rho_{p-3}\alpha_p$

. . . . .

. . . . .

. . . . .

$\rho_p = \rho_{p-1}\alpha_1 + \rho_{p-2}\alpha_2 + \rho_{p-3}\alpha_3 + \ldots + \alpha_p$

Οι παραπάνω εξισώσεις αποτελούν ένα σύστημα p εξισώσεων και λύνοντάς το προκύπτουν οι τιμές για τις αυτοσυσχετίσεις ρ_p, αν είναι γνωστές οι τιμές των συντελεστών $\alpha_1, \alpha_2, \ldots, \alpha_p$ που ονομάζονται και *συντελεστές αυτοπαλινδρόμησης (autoregressive coefficients)* .

Αν οι αυτοσυσχετίσεις είναι γνωστές, τότε οι συντελεστές αυτοσυσχέτισης δίνονται από τη σχέση:

$$A = \Pi^{-1} R$$

Όλες οι αυτοπαλίνδρομες διαδικασίες έχουν συναρτήσεις αυτοσυσχέτισης οι οποίες βαίνουν φθίνουσες καθώς αυξάνει το μήκος υστέρησης s, με αποτέλεσμα να είναι πολλές φορές δύσκολο να καθοριστεί η τάξη του υποδείγματος που περιγράφει τη σειρά με βάση τη συνάρτηση αυτοσυσχέτισης. Ένα πρόσθετο κριτήριο για το σκοπό αυτό είναι η συνάρτηση μερικής αυτοσυσχέτισης (partial autocorrelation function). Η μερική αυτοσυσχέτιση ανάμεσα στην Y_t και στην Y_{t-s} αναφέρεται στη συσχέτιση ανάμεσα στην Y_t και στην Y_{t-s} όταν έχουν αφαιρεθεί οι γραμμικές επιδράσεις των ενδιάμεσων μεταβλητών Y_{t-1}, $Y_{t-2}, \ldots, Y_{t(s-1)}$. Αν παραστήσουμε με ρ_{ss} το συντελεστή μερικής αυτοσυσχέτισης s τάξεως, δηλαδή το συντελεστή αυτοσυσχέτισης ανάμεσα στην Y_t και την Y_{t-s} για s =1, 2, …, τότε το ρ_{ss} θα είναι ο συντελεστής μερικής παλινδρόμησης της μεταβλητής y_{t-s} στο υπόδειγμα:

$$y_t = \rho_{1s} y_{t-1} + \rho_{2s} y_{t-2} + \rho_{3s} y_{t-3} + \ldots + \rho_{ss} y_{t-s} + \varepsilon_t$$

Οι συντελεστές μερικής αυτοσυσχέτισης $\rho_{11}, \rho_{22}, \ldots, \rho_{ss}$ προκύπτουν από τη λύση του συστήματος που προκύπτει από τη παραπάνω σχέση.

Είναι προφανές ότι, για μία αυτοπαλίνδρομη διαδικασία p τάξεως, η συνάρτηση μερικής αυτοσυσχέτισης είναι μηδέν για $s > p$, σύμφωνα με τον ορισμό της μερικής αυτοσυσχέτισης.Πιο συγκεκριμένα ισχύουν τα εξής:

Για την **AR(1)** : $\rho_{11} = \rho_1 = \alpha_1$

$\rho_{ss} = 0$ *για* $s > 1$

Για την **AR(2)** : $\rho_{11} = \dfrac{r_2 - r_1^2}{1 - r_1^2}$

$\rho_{ss} = 0$ *για* $s > 2$

Για την **AR(p)** : $\rho_{11} = \rho_1$

$\rho_{22} \neq 0$

.

.

$\rho_{pp} \neq 0$ *για* $s > p$

Στην πράξη, επειδή τόσο οι αληθινές μερικές αυτοσυσχετίσεις όσο και οι αληθινές αυτοσυσχετίσεις ρ_s δεν είναι γνωστές, χρησιμοποιούνται οι αντίστοιχες εκτιμήσεις τους από το δείγμα. Με βάση τις εκτιμήσεις αυτές μπορεί να γίνει έλεγχος σημαντικότητας των παραμέτρων στον πληθυσμό. Για μεγάλα δείγματα, οι εκτιμήσεις $\hat{r}_s$ των αυτοσυσχετίσεων ρ_s κατανέμονται κανονικά με μέση τιμή το μηδέν και διακύμανση $\frac{1}{T}$, όπου T είναι το μέγεθος του δείγματος. Το ίδιο ισχύει και για τις εκτιμήσεις των μερικών αυτοσυσχετίσεων ρ_{ss} για υστερήσεις s μεγαλύτερες από την τάξη p της AR διαδικασίας. Συμβολικά έχουμε:

$$\hat{r}_{ss} \sim N\left(0, \frac{1}{\mathrm{T}}\right)$$

$$\hat{r}_s \quad N \quad \hat{r}_s = \frac{\sum_{i-1}^{T} (Y_t - \bar{Y})(Y_{t+s} - \bar{Y})}{\sum_{i-1}^{T} (Y_t - \bar{Y})^2}$$

$$|r|$$

$$\hat{r}_{ss} \sim N\left(0, \frac{1}{\mathrm{T}}\right) \quad \text{για } s > p$$

Στη συνέχεια θα αναπτύξουμε τον έλεγχο της σημαντικότητας του συντελεστή ρ_s, δηλαδή θα ελέγξουμε την υπόθεση:

$$H_0 : \rho_s = 0$$

$$H_1 : \rho_s \neq 0$$

Ο έλεγχος αυτός θα γίνει με τη βοήθεια της στατιστικής: $t_s = \frac{\hat{r}_s}{\sqrt{\frac{1}{T}}} = \hat{r}_s \sqrt{T}$

Για δεδομένο επίπεδο σημαντικότητας α = 5% η μηδενική υπόθεση απορρίπτεται αν $|t_s| > 2$, διαφορετικά την αποδεχόμαστε.

Το 95% διάστημα εμπιστοσύνης για αυτόν τον έλεγχο είναι:

$$\hat{r}_s - \frac{2}{\sqrt{T}} \leq \hat{r}_s$$

Ακριβώς τα ίδια ισχύουν και για τον έλεγχο σημαντικότητας του συντελεστή μερικής αυτοσυσχέτισης ρ_{ss}. Δηλαδή, ο συντελεστής ρ_{ss} είναι σημαντικός αν $\left|\hat{r}_{ss}\sqrt{T}\right| > 2$.

Με τη βοήθεια του παραπάνω ελέγχου σημαντικότητας των συντελεστών μερικής αυτοσυσχέτισης μπορεί να καθοριστεί η τάξη μιας AR διαδικασίας. Θα επιλεγεί ως τάξη της σειράς αυτή που αντιστοιχεί στην τελευταία σημαντική τιμή του t_s.Έστω, για παράδειγμα ότι η τελευταία σημαντική τιμή του t είναι για $s = 2$, δηλαδή έστω ότι ο συντελεστής ρ_{22} είναι σημαντικός, ενώ ο συντελεστής ρ_{ss} δεν είναι σημαντικός. Τότε, συμπεραίνουμε ότι η τάξη p του υποδείγματος είναι 2.

Στη συνέχεια, θα εκτιμήσουμε τις παραμέτρους ενός υποδείγματος AR, ενώ γνωρίζουμε την τάξη του p. Υπάρχουν δύο τρόποι για να γίνει αυτό.

Ο πρώτος τρόπος είναι να χρησιμοποιήσουμε τη σχέση $A = \Pi^{-1}R$ αντικαθιστώντας τις αυτοσυσχετίσεις ρ_s με τις εκτιμήσεις $\hat{r}_s$ από το δείγμα, αφού τις έχουμε υπολογίσει από τη σχέση: $\sum_{t-1}^{T}(Y_t - m - q_1 e_{t-1} - q_2 e_{t-2} - \ldots - q_q q_{t-q})^2 =$ $\frac{\sum_{t-1}^{T}(Y_t - \bar{Y})(Y_{t+s} - \bar{Y})}{\sum_{t-1}^{T}(Y_t - \bar{Y})^2}$. Έτσι, θα μπορέσουμε να υπολογίσουμε τους εκτιμητές των παραμέτρων α_p του υποδείγματος.

Ένας **δεύτερος τρόπος** για τον υπολογισμό των εκτιμητών των παραμέτρων α_p είναι να χρησιμοποιηθεί η μέθοδος των ελαχίστων τετραγώνων. Οι εκτιμητές που προκύπτουν από τη μέθοδο αυτή έχουν τις ιδιότητες μεγάλων δειγμάτων, είναι δηλαδή, συνεπείς εκτιμητές και ακολουθούν προσεγγιστικά την κανονική κατανομή. Αυτή η μέθοδος μπορεί να χρησιμοποιηθεί από τη στιγμή που το υπόδειγμα AR(p) στη γενική του μορφή μπορεί να θεωρηθεί ως ένα γραμμικό υπόδειγμα με p ανεξάρτητες στοχαστικές μεταβλητές

Για ένα δείγμα Τα παρατηρήσεων έχουμε το ακόλουθο σύστημα T-p εξισώσεων:

$$Y_{p+1} = \alpha_0 + \alpha_1 Y_p + \alpha_2 Y_{p-1} + \ldots + \alpha_p Y_1 + \varepsilon_{p+1}$$

$$Y_{p+2} = \alpha_0 + \alpha_1 Y_{p+1} + \alpha_2 Y_p + \ldots + \alpha_p Y_2 + \varepsilon_{p+2}$$

$$Y_{p+3} = \alpha_0 + \alpha_1 Y_{p+2} + \alpha_2 Y_{p+1} + \ldots + \alpha_p Y_3 + \varepsilon_{p+3}$$

.

.

$$Y_T = \alpha_0 + \alpha_1 Y_{T-1} + \alpha_2 Y_{T-2} + \ldots + \alpha_p Y_{T-p} + \varepsilon_r$$

Οι εκτιμητές των παραμέτρων $\boldsymbol{\alpha_p}$ όπως θα προκύψουν από τη μέθοδο των ελαχίστων τετραγώνων θα δίνονται από τη σχέση:

$\hat{b} = (X'X)^{-1}X'Y$, όπου $\hat{b}$ θα είναι το διάνυσμα των εκτιμητών $\hat{a}_p$ των παραμέτρων α_p.

Ο πίνακας διακυμάνσεων-συνδιακυμάνσεων των εκτιμητών θα είναι:

$$S_{\hat{b}\hat{b}} = s^2(X'X)^{-1}, \text{ όπου } s^2 = \frac{(Y - X\hat{b})'(Y - X\hat{b})}{T - p - (p+1)}.$$

Ένας εκτιμητής της μέσης τιμής της AR(p) διαδικασίας δίνεται από τη σχέση:

$$\hat{m} = \frac{\hat{a}_0}{1 - \hat{a}_1 - \hat{a}_2 - \ldots - \hat{a}_p}.$$

2.2 Υποδείγματα κινητού μέσου MA(q)

Η γενική μορφή ενός υποδείγματος κινητού μέσου τάξης q ή MA(q) είναι:

$$Y_t = \mu + \varepsilon_1 + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots + \theta_q\varepsilon_{t-q}$$

Η τάξη *q* αναφέρεται στο μήκος της υστέρησης της μεταβλητής, για την οποία υποθέτουμε ότι είναι λευκός θόρυβος. Ο όρος κινητός μέσος αναφέρεται στο γεγονός ότι κ Y_t εμφανίζεται ως ένα σταθμισμένο άθροισμα των τιμών της ε_1.

♦ Στη συνέχεια, θα εξετάσουμε πιο αναλυτικά τη γενική μορφή και τις ιδιότητες μιας MA(1) διαδικασίας, δηλαδή μιας διαδικασίας κινητού μέσου πρώτης τάξης. Η γενική μορφή μιας MA(1) είναι :

$$Y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1}$$

Για το υπόδειγμα MA(1) ισχύουν επίσης τα εξής:

1. $E(Y_t) = \mu$

2. $\gamma_0 = V(Y_t) = (1 + q_1^2)\sigma^2$

3. $\gamma_s = Cov(Y_t, Y_{t-s}) = \theta_1\sigma^2$ για $s = 1$

$=0$ για $s > 1$

4. $\rho_s = \dfrac{q_1}{1+q_1^2}$ για s = 1

$=0$ για s > 1

Μία MA(1) διαδικασία είναι *αντιστρέψιμη, ή χαρακτηρίζεται από αντιστρεψιμότητα* αν μπορεί να διατυπωθεί ως μία αυτοπαλίνδρομη διαδικασία με άπειρους όρους. Οπότε, τότε η MA(1) μετατρέπεται σε AR(∞).

v Ανάλογα, η γενική μορφή ενός υποδείγματος κινητού μέσου τάξεως 2, δηλαδή μιας **MA(2)** διαδικασίας, θα είναι:

$$Y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}$$

Οι σχέσεις που ισχύουν για το MA(2) είναι:

1. $E(Y_t) = \mu$

2. $\gamma_0 = V(Y_t) = (1 + q_1^2 + q_2^2)\sigma^2$

3. $\gamma_1 = (\theta_1 + \theta_2\theta_1)\sigma^2$

4. $\gamma_2 = \theta_2\sigma^2$

5. $\gamma_s = 0$ για $s > 2$

6. $\rho_1 = \dfrac{q_1 + q_2 q_1}{1 + q_1^2 + q_2^2}$

7. $\rho_2 = \dfrac{q_2}{1 + q_1^2 + q_2^2}$

8. $\rho_s = 0$ για $s > 2$

v Τέλος, θα ασχοληθούμε με το υπόδειγμα κινητού μέσου τάξης *q*, δηλαδή το MA*(q)*. Η γενική του μορφή και οι σχέσεις που ισχύουν ακολουθούν στη συνέχεια.

$$Y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots + \theta_q\varepsilon_{t-q}$$

$$E\,(Y_t) = \mu$$

$$\gamma_0 = V(Y_t) = (1 + q_1^2 + q_2^2 + \ldots + q_q^2)\sigma^2 = \sigma^2 \sum_{s-0}^{q} q_s^2 \;, \quad \text{όπου } \theta_0 = 1$$

$$\gamma_s = Cov\,(Y_t, Y_{t-s}) = (\theta_s + \theta_{s+1}\theta_1 + \theta_{s+2}\theta_2 + \ldots + \theta_q\theta_{q-s})\sigma^2 \quad \text{για } s = 1, 2, \ldots, q$$

$$= 0 \qquad \text{για } s > q$$

$$\rho_s = \frac{q_s + q_{s+1}q_1 + q_{s+2}q_2 + \ldots + q_q q_{q-s}}{1 + q_1^2 + q_2^2 + \ldots + q_q^2} \qquad \text{για } s = 1, 2, \ldots, q$$

$$= 0 \qquad \text{για } s > q$$

Γενικά, οι αυτοδιακυμάνσεις και συνεπώς η συνάρτηση αυτοσυσχέτισης είναι μηδέν μετά από q υστερήσεις, δηλαδή όταν $s > q$.

Θα αναφέρουμε στη συνέχεια κάποιες διαφορές ανάμεσα σε μία AR(p) διαδικασία και σε μία MA(q). Η συνάρτηση αυτοσυσχέτισης μιας AR(p) διαδικασίας μπορεί να εκτείνεται στο άπειρο, ενώ η συνάρτηση αυσυσχέτισης μιας MA(q) διαδικασίας μηδενίζεται μετά από q υστερήσεις. Αντίθετα, η συνάρτηση μερικής αυτοσυσχέτισης μιας AR(p) διαδικασίας τερματίζεται μετά από p υστερήσεις, ενώ η συνάρτηση μερικής αυτοσυσχέτισης μιας MA(q) διαδικασίας εκτείνεται στο άπειρο.

Η τάξη q του υποδείγματος μπορεί να προσδιοριστεί από τη συμπεριφορά της δειγματικής συνάρτησης αυτοσυσχέτισης, όπως και στις αυτοπαλίνδρομες διαδικασίες. Είδαμε ότι η συνάρτηση αυτοσυσχέτισης μιας MA(q) διαδικασίας μηδενίζεται μετά από q υστερήσεις, οπότε αυτό σημαίνει ότι οι αυτοσυσχετίσεις θα είναι σημαντικές.

Όταν θα έχει καθοριστεί η τάξη της διαδικασίας, οι παράμετροι του υποδείγματος μπορούν να εκτιμηθούν με τον ίδιο τρόπο που εκτιμήθηκαν και οι παράμετροι μιας

AR(p) διαδικασίας, δηλαδή από τη σχέση Α = Π^{-1}R, αντικαθιστώντας τις αυτοσυσχετίσεις ρ_s με τις εκτιμήσεις $\hat{r}_s$ από το δείγμα, αφού τις έχουμε υπολογίσει από τη σχέση:

$$\hat{r}_s = \frac{\sum_{i-1}^{T}(Y_t - \bar{Y})(Y_{t+s} - \bar{Y})}{\sum_{i-1}^{T}(Y_t - \bar{Y})^2}$$

Η μέθοδος των ελαχίστων τετραγώνων δεν μπορεί να χρησιμοποιηθεί για την εκτίμηση των παραμέτρων ενός υποδείγματος κινητού μέσου, όπως στην περίπτωση των αυτοπαλίνδρομων υποδειγμάτων, γιατί η προς ελαχιστοποίηση συνάρτηση:

$\sum_{t-1}^{T} e_t^2 = \sum_{t-1}^{T}(Y_t - m - q_1 e_{t-1} - q_2 e_{t-2} - ... - q_q q_{t-q})^2$ δεν είναι γραμμική ως προς τις παραμέτρους. Για την εκτίμηση των παραμέτρων απαιτείται η χρήση διαφόρων μη γραμμικών μεθόδων, τις οποίες δε θα αναλύσουμε αυτή τη στιγμή.

2.3 Μεικτά αυτοπαλίνδρομα-κινητού μέσου υποδείγματα ARMA(p,q)

Η γενική μορφή ενός υποδείγματος ARMA(p, q) είναι το υπόδειγμα:

$Y_t = \alpha_0 + \alpha_2 Y_{t-2} + \ldots + \alpha_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \varepsilon_{t-q}$

Το υπόδειγμα ARMA(p, q) είναι συνδυασμός p αυτοπαλίνδρομων όρων και q όρων κινητού μέσου. Είναι προφανές ότι ένα καθαρά αυτοπαλίνδρομο υπόδειγμα ή ένα καθαρό υπόδειγμα κινητού μέσου μπορούν να θεωρηθούν ως ειδικές περιπτώσεις μιας ARMA διαδικασίας. Δηλαδή, θα ισχύουν τα εξής:

AR(p) = ARMA(p, 0) και

MA(q) = ARMA (0,q)

Η απλούστερη μορφή μιας ARMA (*p,q*) διαδικασίας είναι το υπόδειγμα **ARMA (1,1).** Η μορφή αυτή του υποδείγματος προφανώς θα είναι :

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \varepsilon_1 + \theta_1 \varepsilon_{t-1} \quad \textbf{(1)} \quad \text{ή αλλιώς:}$$

$$Y_t = \alpha_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1} \quad \textbf{(2)}$$

Αυτό το υπόδειγμα μπορεί να γραφεί ως μία καθαρά MA διαδικασία, αλλά και ως μια αυτοπαλίνδρομη διαδικασία με άπειρους όρους AR(∞).

♦ Ας δούμε πρώτα πώς μπορεί να μεταμορφωθεί σε μία διαδικασία κινητού μέσου.

Υστερούμε διαδοχικά τη μορφή της (1) και αντικαθιστώντας για Y_{t-1}, Y_{t-2}, Y_{t-3},… καταλήγουμε στη μορφή:

$$Y_t = \alpha_0 \sum_{i-0}^{\infty} a_1^i + \varepsilon_t + \sum_{i-0}^{\infty} a_1^i (a_1 + q_1) e_{t-1-i}$$

Για να είναι στάσιμη αυτή η σειρά, θα πρέπει το άθροισμα $\sum_{i-0}^{\infty} a_1^i (a_1 + q_1)$ να συγκλίνει, οπότε αυτό σημαίνει ότι πρέπει το $|\alpha_1| < 1$. Όταν ισχύει αυτό, και επομένως η σειρά είναι στάσιμη, τότε η μορφή του υποδείγματος θα είναι η :

$$Y_t = \frac{a_0}{1-a_1} + \varepsilon_t + \sum_{i-0}^{\infty} a_1^i (a_1 + q_1) e_{t-1-i}$$

Αυτή η μορφή δημιουργήθηκε αντικαθιστώντας τη σειρά $\sum_{i-0}^{\infty} a_1^i = \frac{1}{1-a_1}$ από τη στιγμή που είναι άθροισμα όρων φθίνουσας γεωμετρικής προόδου. Η μορφή αυτή είναι μία MA διαδικασία με άπειρους όρους, που θα μπορούσε να προσεγγιστεί με ένα περιορισμένο αριθμό όρων, δεδομένου ότι η σημασία των συντελεστών όλο και μικραίνει. Αυτό σημαίνει ότι από κάποιο σημείο θα μπορούσαν να παραλειφθούν οι επόμενοι όροι. Μπορούμε εύκολα να καταλάβουμε ότι απαιτείται υψηλής τάξεως MA διαδικασία προκειμένου να προσεγγιστεί η αντίστοιχη ARMA (1,1) διαδικασία. Επομένως, είναι προφανής η οικονομία που επιτυγχάνεται με τη χρήση των μεικτών υποδειγμάτων, αφού το ARMA(1,1) υπόδειγμα έχει μόνο δύο συντελεστές.

♦ Στη συνέχεια, θα δούμε πώς το υπόδειγμα **(1)** μπορεί να διατυπωθεί ως AR(∞). Με διαδοχικές αντικαταστάσεις για ε_{t-1}, ε_{t-2}, ε_{t-3}… στη σχέση **(1),** προκύπτει η ακόλουθη σχέση:

$$Y_t = \frac{a_0}{1-q_1} + \varepsilon_t + \sum_{i-0}^{\infty} q_1^i (a_1 + q_1) Y_{t-1-i}$$

Για να είναι η σειρά αναστρέψιμη θα πρέπει $|\theta_1| < 1$. Και σε αυτή την περίπτωση επιτυγχάνεται οικονομία στους συντελεστές με τη χρήση της ARMA(1,1) διαδικασίας.

Για το μεικτό υπόδειγμα **ARMA(1,1)** αποδεικνύονται εύκολα ότι ισχύουν οι σχέσεις:

I. $E(Y_t) = \mu = \dfrac{a_0}{1-a_1}$

II. $\gamma_0 = \dfrac{1+q_1^2+2a_1q_1}{1-a_1^2}\sigma^2$

III. $\gamma_1 = \alpha_1\gamma_0 + \theta_1\sigma^2$

IV. $\gamma_s = \alpha_1\gamma_{s-1}$ για $s > 1$

V. $r_1 = a_1 + \dfrac{q_1}{g_0}s^2 = \dfrac{(1+a_1q_1)(a_1+q_1)}{1+q_1^2+2a_1q_1}$

VI. $\rho_s = \alpha_1\rho_{s-1}$ για $s > 1$.

Οι **αποδείξεις** αυτών των σχέσεων δίδονται στη συνέχεια:

I. Στη σχέση (1) παίρνουμε τις μέσες τιμές και στα δύο μέλη, οπότε είναι :

$E(Y_t) = \alpha_0 + \alpha_1\mu$

$(1-\alpha_1)\mu = \alpha_0$

$\mu = \dfrac{a_0}{1-a_1}$

II. Υψώνουμε τη σχέση (2) στο τετράγωνο και παίρνουμε τις μέσες τιμές:

$$E(y_t^2) = E(a_1 y_{t-1} + e_t + q_1 e_{t-1})^2$$
$$g_0 = a_1^2 g_0 + s^2 + q_1^2 s^2 + 2a_1 q_1 s^2$$
$$g_0 - a_1^2 g_0 = (1 + q_1^2 + 2a_1 q) s^2$$
$$(1 - a_1^2) g_0 = (1 + q_1^2 + 2a_1 q) s^2$$
$$g_0 = \frac{(1 + q_1^2 + 2a_1 q)}{(1 - a_1^2)} s^2$$

III. Από τη σχέση (2), έχουμε ότι:

$$g_1 = Cov(y_1, y_0) = E(y_1 - m)(y_0 - m) =$$
$$= \mathrm{E}(a_1 y_0 + e_1 + q_1 e_0) y_0 =$$
$$= E(a_1 y_0^2) + E(e_1 y_0) + E(q_1 e_0 y_0) =$$
$$= a_1 g_0 + q_1 s^2$$

IV. Με τη χρήση της σχέσης (2) πάλι, έχουμε ότι:

$$g_s = Cov(y_s, y_{s-1}) = E(y_s - m)(y_{s-1} - m) =$$
$$= E(a_1 y_{s-1} + e_s + q_1 e_{s-1})(a_1 y_{s-2} + e_{s-1} + q_1 e_{s-2}) =$$
$$= a_1^2 g_{s-2} + a_1 q_1 s^2 = a_1 (a_1 g_{s-2} + q_1 s^2) =$$
$$= a_1 g_{s-1},$$

$$r_1 = \frac{g_1}{g_0} = \frac{a_1 g_0 + q_1 s^2}{g_0} = a_1 + \frac{q_1}{g_0} s^2 \text{ ή}$$

$$r_1 = a_1 + \frac{q_1}{g_0} s^2 = a_1 + \frac{q_1 s^2}{\frac{1 + q_1^2 + 2a_1 q_1}{1 - a_1^2} s^2} =$$

$$= a_1 + \frac{q_1 (1 - a_1^2)}{1 + q_1^2 + 2a_1 q_1} = \frac{a_1 + a_1 q_1^2 + 2a_1^2 q_1 + q_1 - a_1^2}{1 + q_1^2 + 2a_1 q_1}$$

$$= \frac{a_1 + a_1 q_1^2 + q_1 a_1^2 + q_1}{1 + q_1^2 + 2a_1 q_1} = \frac{a_1 (1 + a_1 q_1) + q_1 (1 + a_1 q_1)}{1 + q_1^2 + 2a_1 q_1}$$

$$= \frac{(a_1 + q_1)(1 + a_1 q_1)}{1 + q_1^2 + 2a_1 q_1}$$

VII. $r_s = \frac{g_s}{g_0} = \frac{a_1 g_{s-1}}{g_0} = a_1 \frac{g_{s-1}}{g_0} = a_1 r_{s-1}, gia\ s > 1.$

Παρατηρούμε ότι στη συνάρτηση αυτοσυσχέτισης υπεισέρχεται ο συντελεστής από την MA(1) διαδικασία, αλλά μόνο για την αυτοσυσχέτιση πρώτης τάξης ρ_1. Οι υπόλοιπες αυτοσυσχετίσεις εξαρτώνται μόνο από το αυτοπαλίνδρομο μέρος. Η συνάρτηση αυτοσυσχέτισης για την ARMA(1,1) διαδικασία φθίνει γεωμετρικά με την αύξηση του s. Η μείωση όμως, σε αντίθεση με την AR(1), αρχίζει από το ρ_1 και όχι από το $\rho_0 = 1$. Η συνάρτηση μερικής αυτοσυσχέτισης για την ARMA(1,1) διαδικασία φθίνει γεωμετρικά όπως στην περίπτωση της MA(1) διαδικασίας.

Στη συνέχεια, θα ασχοληθούμε με το γενικό μεικτό υπόδειγμα ARMA(p,q). Η γενική του μορφή, όπως είδαμε προηγουμένως είναι:

$Y_t = \delta + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \ldots + \alpha_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$

Οι πρώτες q αυτοσυσχετίσεις για $s \leq q$, εξαρτώνται τόσο από τους συντελεστές α_i του αυτοπαλίνδρομου υποδείγματος, όσο και από τους συντελεστές θ_i του υποδείγματος του κινητού μέσου. Για τις τιμές του $s > q$, οι αυοτδιακυμάνσεις και οι αυτοσυσχετίσεις είναι ακριβώς ίδιες με αυτές μιας AR(p) διαδικασίας και επομένως θα δίνονται και αυτές από τους τύπους:

$\gamma_s = \alpha_1 \gamma_{s-1} + \alpha_2 \gamma_{s-2} + \ldots + \alpha_p \gamma_{s-p}$ για $s > q$

$\rho_s = \alpha_1 \rho_{s-1} + \alpha_2 \rho_{s-2} + \ldots + \alpha_p \rho_{s-p}$ για $s > q$

Γενικά, η συνάρτηση αυτοσυσχέτισης μιας ARMA(p,q) διαδικασίας συμπεριφέρεται όπως αυτή μιας AR(p) διαδικασίας, ενώ η συνάρτηση μερικής αυτοσυσχέτισης συμπεριφέρεται όπως αυτή μιας MA(q) διαδικασίας, για $s > q - p$.

Για την εκτίμηση των παραμέτρων ενός ARMA(p,q) υποδείγματος μπορούν να εφαρμοστούν οι ίδιες τεχνικές που χρησιμοποιούνται και για την εκτίμηση των παραμέτρων ενός MA(q) υποδείγματος. Δηλαδή μπορούν να χρησιμοποιηθούν οι σχέσεις που συνδέουν τις αυτοσυσχετίσεις με τις παραμέτρους του υποδείγματος, αλλά επίσης μπορούν να εφαρμοστούν κάποιες με γραμμικές μέθοδοι εκτίμησης, από τη στιγμή που το υπόδειγμα είναι μη γραμμικό ως προς τις παραμέτρους.

<u>ΚΕΦΑΛΑΙΟ 3</u>

<u>Μη στάσιμες χρονολογικές σειρές</u>.

Τα τρία στοχαστικά υποδείγματα που εξετάσαμε μέχρι τώρα αναφέρονται όλα σε στάσιμες διαδικασίες που σημαίνει ότι ο μέσος, η διακύμανση και οι αυτοδιακυμάνσεις δεν εξαρτώνται από το χρόνο *t*, δηλαδή ο μέσος και η διακύμανση παραμένουν σταθεροί, ενώ οι αυτοδιακυμάνσεις εξαρτώνται μόνο από τη χρονική υστέρηση *s*. Στη συνέχεια, θα εξετάσουμε μη στάσιμες διαδικασίες.

Στην Οικονομετρία και στην Οικονομία, μας ενδιαφέρει οι σειρές να είναι στάσιμες, γιατί έτσι αποφεύγονται πολλά προβλήματα, όπως αυτό της φαινομενικής παλινδρόμησης.

Δυστυχώς, οι περισσότερες αν όχι όλες οι οικονομικές χρονολογικές σειρές, όπως η κατανάλωση, η ανεργία, ο δείκτης τιμών, τα κέρδη δεν έχουν χαρακτηριστικά στάσιμων διαδικασιών. Όμως, υπάρχει τρόπος μετατροπής μιας μη στάσιμης σειράς σε στάσιμη και αυτό γίνεται με τη χρήση πρώτων, δεύτερων κ.τ.λ. διαφορών.

<u>Ορισμός:</u> Μια μη στάσιμη σειρά λέγεται ολοκληρωμένη ***d*** τάξεως (integrated d^{th} order) και παριστάνεται ως I(d), όταν μετατρέπεται σε στάσιμη με τη χρήση ***d*** αριθμού διαφορών. Ο όρος ολοκληρωμένη σειρά προέρχεται από τον τρόπο με τον οποίο μία μη στάσιμη διαδικασία προκύπτει από μία στάσιμη. Αυτό γίνεται ολοκληρώνοντας ***d*** φορές τη μη στάσιμη σειρά για να προκύψει η στάσιμη. Μία στάσιμη σειρά, όπως ο λευκός θόρυβος, θεωρείται ολοκληρωμένη σειρά μηδενικής τάξεως, I(0).

3.1 *Υποδείγματα ARIMA*

Γενικά, ένα υπόδειγμα ARMA(*p,q*) που εφαρμόζεται σε μία ολοκληρωμένη σειρά *d* τάξεως (*p, d, q*) και συμβολίζεται ως ARIMA(*p, d, q*). Συγκεκριμένα, οι τρεις μορφές των παραμέτρων αυτού του υποδείγματος είναι: οι *p* παράμετροι του αυτοπαλίνδρομου υποδείγματος, ο αριθμός *d* των διαφορών που απαιτούνται για να γίνει η σειρά στάσιμη, και τέλος οι *q* παράμετροι του υποδείγματος κινητού μέσου. Για παράδειγμα, ένα υπόδειγμα που περιγράφεται ως ARIMA(0,1,2) σημαίνει ότι περιέχει μηδέν αυτοπαλίνδρομες παραμέτρους και δύο παραμέτρους κινητού μέσου που έχουν υπολογιστεί για την προκύπτουσα σειρά των πρώτων διαφορών.

Όπως είδαμε προηγουμένως μια ARMA(*p,q*) διαδικασία στη γενική της μορφή γράφεται ως εξής:

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + ... + a_p Y_{t-p} + e_t + q_X e_{t-1} + q_X e_{t-2} + ... + q_q q_{t-q}$$

Μία ARIMA(*p, d, q*) διαδικασία μπορεί να διατυπωθεί με τρεις διαφορετικούς τρόπους και να πάρει τρεις διαφορετικές μορφές:

1. Ως συνάρτηση των παρελθουσών τιμών της και των τιμών του διαταρακτικού όρου, τρέχουσας και παρελθουσών. Η μορφή αυτή είναι γνωστή και ως **εξίσωση διαφοράς** (difference equation form).
2. Ως συνάρτηση των παρελθουσών τιμών της και της τρέχουσας τιμής του διαταρακτικού όρου. Η μορφή αυτή είναι γνωστή ως η **αντίστροφη μορφή** (inverted form).
3. Ως συνάρτηση μόνο των τιμών του διαταρακτικού, τρέχουσας και παρελθουσών. Η μορφή αυτή είναι γνωστή ως **τυχαία διαταραχή** (random shock form).

Η ανάπτυξη και η κατασκευή υποδειγμάτων ARIMA ως εργαλεία πρόβλεψης των τιμών οικονομικών μεταβλητών είναι γνωστή ως ***μεθοδολογία Box- Jenkins,*** την οποία θα αναλύσουμε στη συνέχεια.

Μεθοδολογία Box- Jenkins

Η προσέγγιση των Box- Jenkins στην ανάλυση χρονοσειρών είναι μία μέθοδος εύρεσης ενός στατιστικού υποδείγματος ARIMA(p, d, q) που να παριστάνει ικανοποιητικά τη στοχαστική διαδικασία από την οποία προήλθαν τα δεδομένα, δηλαδή το δείγμα μας. Η μέθοδος αυτή περιλαμβάνει τρία στάδια, την ταυτοποίηση (identification), την εκτίμηση (estimation), και τον διαγνωστικό έλεγχο (diagnostic checking) και τα οποία θα αναλύσουμε στη συνέχεια.

1ο Στάδιο: Ταυτοποίηση.

Σε αυτό το στάδιο γίνεται η εξειδίκευση ενός ARIMA υποδείγματος με βάση τις πληροφορίες που παίρνουμε από το δείγμα. Αυτό σημαίνει ότι καθορίζονται οι τιμές των d, p και q. Δηλαδή, καθορίζεται ο αριθμός d των διαφορών που απαιτούνται για να μετατραπεί η σειρά σε στάσιμη, από τη στιγμή βέβαια που δεν είναι, και στη συνέχεια καθορίζεται η τάξη p της αυτοπαλίνδρομης διαδικασίας και η τάξη q της διαδικασίας κινητού μέσου. Για να διατυπωθεί αν σειρά είναι στάσιμη ή όχι, θα εξεταστεί η συμπεριφορά της δειγματικής συνάρτησης αυτοσυσχέτισης. Αν οι αυτοσυσχετίσεις συγκλίνουν ταχύτατα προς το μηδέν, σημαίνει ότι η σειρά μάλλον είναι στάσιμη. Αντίθετα, αν οι αυτοσυσχετίσεις φθίνουν με αργό ρυθμό, είναι σοβαρή ένδειξη ότι η σειρά είναι μη στάσιμη, οπότε πρέπει να γίνει στάσιμη. Σε αυτήν την περίπτωση θα χρησιμοποιήσουμε τις πρώτες ή τις δεύτερες ή κ.τ.λ. διαφορές για να μετατραπεί η σειρά σε στάσιμη. Αφού η σειρά έχει γίνει στάσιμη, προσδιορίζεται στη συνέχεια η τάξη του υποδείγματος ARIMA, δηλαδή προσδιορίζονται οι τιμές το p και του q. Ο προσδιορισμός τους βασίζεται στις δειγματικές απλές και μερικές αυτοσυσχετίσεις.

2ο Στάδιο: Εκτίμηση.

Μετά την εξειδίκευση του υποδείγματος και την εύρεση της τάξης του ακολουθεί η εκτίμηση των p παραμέτρων $\alpha_1, \alpha_2, \ldots \alpha_p$ της αυτοπαλίνδρομης διαδικασίας και των q παραμέτρων $\theta_1, \theta_2, \ldots \theta_q$ της διαδικασίας κινητού μέσου. Αν η σειρά που εξετάζουμε είναι μόνο αυτοπαλίνδρομη, οι παράμετροί της, όπως είδαμε προηγουμένως, μπορούν να εκτιμηθούν με τη μέθοδο των ελαχίστων τετραγώνων. Αν όμως, η σειρά περιέχει και όρους κινητού μέσου τότε για την εκτίμηση των παραμέτρων του κινητού μέσου, θα χρησιμοποιηθούν μη γραμμικές μέθοδοι εκτίμησης.

3ο Στάδιο: Διαγνωστικός Έλεγχος.

Στο στάδιο αυτό γίνεται έλεγχος καλής προσαρμογής του υποδείγματος. Αυτό σημαίνει ότι ελέγχεται το πόσο καλά ταιριάζει το εκτιμώμενο υπόδειγμα με τα δεδομένα, γιατί είναι πιθανό κάποιο άλλο υπόδειγμα ARIMA να προσαρμόζεται καλύτερα. Θα εφαρμόσουμε κάποιους στατιστικούς ελέγχους για τη σημαντικότητα των παραμέτρων, τη συμπεριφορά των καταλοίπων και την τάξη του υποδείγματος.

♦ Θα ασχοληθούμε πρώτα με τον έλεγχο των καταλοίπων. Αν το εκτιμώμενο υπόδειγμα είναι το πιο κατάλληλο για τα δεδομένα μας, αν δηλαδή εκφράζει ικανοποιητικά τη διαδικασία από την οποία προέρχονται τα δεδομένα, τότε τα κατάλοιπα θα πρέπει να συμπεριφέρονται ως μία διαδικασία λευκού θορύβου. Αυτό σημαίνει ότι τα κατάλοιπα δεν πρέπει να αυτοσυσχετίζονται. Αυτός ο έλεγχος για τα κατάλοιπα γίνεται με τη στατιστική Q των Box- Pierce, με την οποία ελέγχεται από κοινού η σημαντικότητα ενός αριθμού συντελεστών αυτοσυσχέτισης, έστω m. Η μηδενική υπόθεση τότε, θα είναι:

$H_0 : \rho_1 = \rho_2 = \ldots = \rho_m = 0,$

Όπου $\rho_i, I = 1, 2, \ldots, m$ είναι οι συντελεστές συσχέτισης των καταλοίπων. Η στατιστική Q των Box- Pierce ορίζεται ως:

$Q_{BP} = T\sum_{s-1}^{m} \hat{r}_s^{\,2}$, όπου $\hat{r}_s$ είναι οι δειγματικές αυτοσυσχετίσεις των καταλοίπων και T ο αριθμός των παρατηρήσεων. Συνήθως, ο αριθμός των αυτοσυσχετίσεων των καταλοίπων ισούται με την τετραγωνική ρίζα του αριθμού των παρατηρήσεων, δηλαδή ισχύει $m = \sqrt{T}$. Η στατιστική Q_{BP} ακολουθεί την X^2 κατανομή με $m\text{-}p\text{-}q$ βαθμούς ελευθερίας. Η μηδενική υπόθεση απορρίπτεται αν η τιμή της Q_{BP} είναι μεγαλύτερη από την κρίσιμη τιμή της κατανομής X_a^2, για δεδομένο επίπεδο σημαντικότητας α. Με λίγα λόγια ισχύουν τα εξής:

- Απόρριψη H_0, αν $Q_{BP} > X_a^2$

- Αποδοχή H_0, αν $Q_{BP} \leq X_a^2$

Μια τροποποιημένη μορφή της στατιστικής των Box- Pierce είναι η στατιστική των Ljung και Box, η οποία ορίζεται ως:

$$Q_{LB} = T(T+2)\sum_{s-1}^{m}\frac{\hat{r}_s^2}{T-s}$$

Κι αυτή η στατιστική ακολουθεί προσεγγιστικά την X^2 κατανομή με m-p-q βαθμούς ελευθερίας. Και εδώ απορρίπτεται η μηδενική υπόθεση όταν $Q_{LB} > X_a^2$ για δεδομένο επίπεδο σημαντικότητας α. Η στατιστική Q_{LB} θεωρείται περισσότερο κατάλληλη για μικρά δείγματα. Για μεγάλα δείγματα δεν υπάρχει διαφορά ανάμεσα στις δύο στατιστικές.

Στη συνέχεια θα ασχοληθούμε με τον **έλεγχο της τάξης του υποδείγματος**. Αυτός ο έλεγχος γίνεται με μια διαδικασία που ονομάζεται υπερπροσαρμογή (overfittimg). Σύμφωνα με αυτή τη διαδικασία, ο έλεγχος της καταλληλότητας του εκτιμημένου υποδείγματος γίνεται συγκρίνοντάς το με ένα άλλο υπόδειγμα μεγαλύτερης τάξης. Με λίγα λόγια το εκτιμημένο υπόδειγμα ARMA*(p, q*) θα συγκριθεί με τα υποδείγματα ARMA (*p + 1, q*) και ARMA(*p, q + 1*) της αμέσως επόμενης τάξης. Αν το εκτιμημένο υπόδειγμα είναι τελικά το καταλληλότερο για τα δεδομένα μας, δηλαδή αν περιγράφει τη διαδικασία από την οποία παράχθηκαν τα δεδομένα, θα πρέπει οι επιπλέον συντελεστές στα μεγαλύτερα υποδείγματα, να μην είναι στατιστικά μεγαλύτεροι από το μηδέν. Αν αυτοί οι συντελεστές δεν είναι μηδέν, τότε θα υπάρχει κάποιο άλλο υπόδειγμα που να είναι πιο κατάλληλο για τα δεδομένα μας, απ’ ότι το εκτιμημένο.

Τέλος, θα αναφέρουμε κάποια κριτήρια που μας βοηθούν να **επιλέξουμε το κατάλληλο υπόδειγμα.** Είναι προφανές ότι αν αυξήσουμε την τάξη του υποδείγματος προσθέτοντας υστερήσεις είτε για το αυτοπαλίνδρομο τμήμα, είτε για το τμήμα κινητού μέσου, θα μειώνεται το άθροισμα των τετραγώνων των καταλοίπων, αλλά ταυτόχρονα θα μειώνονται και οι βαθμοί ελευθερίας αφού εκτιμώνται περισσότερες παράμετροι. Δύο κριτήρια που χρησιμοποιούνται ευρέως στην ανάλυση των χρονολογικών σειρών, είναι το κριτήριο πληροφοριών Akaike

(Akaike Information Criterion) ή αλλιώς AIC και το Μπαϊεσιανό κριτήριο Schwartz (Schwartz Bayesian Criterion) ή αλλιώς SBC. Τα κριτήρια αυτά ορίζονται ως εξής:

$$AIC = 1n\frac{\sum \hat{e}_t^2}{T} + \frac{2k}{T}$$

$$SBC = 1n\frac{\sum \hat{e}_t^2}{T} + \frac{2k}{T}1nT$$

Όπου $\sum \hat{e}_t^2$: το άθροισμα των τετραγώνων των καταλοίπων

T : ο αριθμός των παρατηρήσεων και,

K : ο αριθμός των παραμέτρων που εκτιμούνται (*p + q + 1*)

Και τα δύο κριτήρια επιβάλλουν κάποια ποινή για τη μείωση του αθροίσματος των τετραγώνων των καταλοίπων $\sum \hat{e}_t^2$ που επιτυγχάνεται με την προσθήκη υστερήσεων. Η ποινή αυξάνει με τον αριθμό των παλινδρομητών k. Το κριτήριο SBC επιβάλλει μεγαλύτερη ποινή από το κριτήριο AIC για $T \geq 7$, από τη στιγμή που πάντα $1nT > 2$ και επομένως

$\frac{2k}{T} < \hat{Y}_{T+1}$ Τελικά, σαν συμπέρασμα, με βάση τα δύο κριτήρια, βγαίνει ότι θα επιλέγεται το υπόδειγμα με τη μικρότερη τιμή.

Ο κύριος σκοπός της εξειδίκευσης και εκτίμησης ενός υποδείγματος ARIMA που αναφέραμε προηγουμένως πώς γίνεται, είναι να κάνουμε βραχυχρόνιες προβλέψεις. Δηλαδή, με βάση το εκτιμημένο υπόδειγμα και τις υπάρχουσες πληροφορίες μέχρι τη χρονική περίοδο *T*, να γίνει η πρόβλεψη της τιμής της *Y* για κάποια μελλοντική τιμή της, για παράδειγμα τη χρονική στιγμή *T+1*, *T+2*, κ.τ.λ. και γενικότερα να γίνει πρόβλεψη στην περίοδο *T+ h*. Στη συνέχεια, θα δούμε πώς μπορούμε να κάνουμε προβλέψεις και στα τέσσερα είδη υποδειγμάτων που αναφέραμε, δηλαδή τα αυτοπαλίνδρομα, τα υποδείγματα κινητού μέσου, τα μεικτά υποδείγματα, καθώς και στα ARIMA.

♦ Έστω το υπόδειγμα **AR**(1):

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \varepsilon_1$$

Για $t = T+1$, το υπόδειγμα γίνεται:

$$Y_{T+1} = \alpha_0 + \alpha_1 Y_{t-1} + \varepsilon_t$$

Αν οι παράμετροι α_0 και α_1 είναι γνωστές, τότε με βάση τις πληροφορίες που έχουμε μέχρι την περίοδο *T*, μία πρόβλεψη για την περίοδο *T+1* θα είναι η υπό συνθήκη προσδοκώμενη τιμή της Y_{T+1}. Αν θέσουμε ως $\hat{Y}_{T+1}$ την πρόβλεψη, τότε:

$\hat{Y}_{T+1} = E_T(Y_{T+1})$ και για να γενικεύσουμε, θα ισχύει: $\hat{Y}_{T+h} = E_T(Y_{T+h})$ που θα αποτελεί μία πρόβλεψη *h* περιόδους μπροστά με βάση τις πληροφορίες που έχουμε μέχρι την περίοδο *T*. Αυτή η πρόβλεψη που ελαχιστοποιεί την ποσότητα $E_T(\hat{Y}_{T+h} - Y_{T+h})^2$, δηλαδή ελαχιστοποιεί το μέσο του τετραγώνου του σφάλματος (Mean Square Error), θεωρείται άριστη πρόβλεψη (optimal forecast).

Γενικά, για τις προβλέψεις που γίνονται h περιόδους μπροστά, ισχύουν τα εξής:

$$MSE = \sum_{t-1}^{M}\left[(Y_t^f - \overline{Y}^f) - (Y_t^a - \overline{Y}^a) + (\overline{Y}^f - \overline{Y}^a)\right]^2$$

$$MSE = (\overline{Y}^f - \overline{Y}^a)^2 + (s^f - s^a)^2 + 2(1-r)s^f s^a = \alpha_0 + \alpha_1 \hat{Y}_{T+h-1}$$

$$Y^a$$

$$V(\hat{e}_{T+h}) = s^2(1 + a_1^2 + a_1^4 + ... + a_1^{2(h-1)})$$

Από την τελευταία σχέση είναι φανερό ότι η διακύμανση του σφάλματος πρόβλεψης αυξάνει μη γραμμικά, καθώς αυξάνει η περίοδος πρόβλεψης.

♦ Έστω το υπόδειγμα **MA**(1):

$$Y_t = m + e_t + q_1 e_{t-1}$$

Η πρόβλεψη για την περίοδο Τ+1, όπως και προηγουμένως θα είναι:

$$\hat{Y}_{T+1} = E_T(Y_{T+1})$$

Γενικά, για προβλέψεις h > 1 περιόδους μπροστά ισχύει:

$$\hat{Y}_{T+h} = m$$
$$V(\hat{e}_{T+h}) = s^2(1 + q_1^2)$$

Από την τελευταία σχέση είναι φανερό ότι ένα υπόδειγμα κινητού μέσου πρώτης τάξης είναι κατάλληλο για προβλέψεις μόνο μία περίοδο μπροστά, αφού για h > 1 η πρόβλεψη θα είναι πάντοτε ο μέσος.

- Έστω το υπόδειγμα **ARMA** (1,1):

$$Y_t = a_0 + a_1 Y_{t-1} + e_t + q_1 e_{t-1}$$

Για *t = T+1*, το υπόδειγμα γίνεται: $Y_{T+1} = a_0 + a_1 Y_T + e_{T+1} + q_1 e_T$

Η άριστη πρόβλεψη για μια χρονική περίοδο μπροστά θα είναι:

$$\hat{Y}_{T+1} = E(a_0 + a_1 Y_T + e_{T+1} + q_1 e_T) \text{ ή } \hat{Y}_{T+1} = a_0 + a_1 Y_T + q_1 e_T$$

Το σφάλμα πρόβλεψης και η διακύμανσή του θα δίνονται από τις ακόλουθες σχέσεις:

$$\hat{e}_{T+1} = Y_{T+1} - \hat{Y}_{T+1} = e_{T+1}$$
$$V(\hat{e}_{T+1}) = V(e_{T+1}) = s^2$$

Μπορούμε να κάνουμε προβλέψεις για h περιόδους μπροστά.

- Έστω το υπόδειγμα **ARIMA**(1,1,1):

$$w_t = a_0 + a_1 w_{t-1} + e_1 \text{, όπου } w_t = \Delta Y_t = Y_t - Y_{t-1}$$

Παίρνουμε τις πρώτες διαφορές οι οποίες είναι AR(1) για να γίνει στάσιμη η μη στάσιμη σειρά Y_t. Οπότε, θα γίνει πρόβλεψη για τη διαφορά w και στη συνέχεια θα γίνει πρόβλεψη της Y. Η άριστη πρόβλεψη για την w_{T+1} είναι:

$$\hat{w}_{T+1} = a_0 + a_1 w_T$$

Η πρόβλεψη τώρα για την αρχική σειρά Y, δηλαδή η πρόβλεψη για την Y_{T+1} θα είναι:

$$\hat{Y}_{T+1} = Y_T + \hat{w}_{T+1} = Y_T + a_0 + a_1 w_T = Y_T + a_0 + a_1(Y_T - Y_{T-1}) = a_0 + (1 + a_1)Y_T - a_1 Y_{T-1}$$

Γενικά, η πρόβλεψη για h περιόδους μπροστά θα είναι:

$$\hat{Y}_{T+h} = Y_T + \hat{w}_{T-1} + \hat{w}_{T+2} + \ldots + \hat{w}_{T+h}$$

Η διακύμανση είναι άγνωστη, θα βρούμε μία εκτίμησή της από το άθροισμα των τετραγώνων των καταλοίπων, η οποία θα είναι:

$$\hat{s}^2 = \frac{\sum_{i-1}^{T} \hat{e}_i^2}{T - p - q}$$

Επομένως, αν αντικαταστήσουμε αυτήν την εκτίμηση στη θέση της διακύμανσης σ^2, θα έχουμε μία εκτίμηση της s_h^2.

Αυτό το διάστημα εμπιστοσύνης που υπολογίστηκε, είναι ένα καλό κριτήριο σχετικά με την ακρίβεια της πρόβλεψης. Δηλαδή, το πόσο κοντά βρίσκονται οι προβλεπόμενες από το εκτιμημένο υπόδειγμα τιμές με τις πραγματικές τιμές. Μπορούμε όμως, να συγκρίνουμε τις προβλεπόμενες με τις πραγματικές τιμές για την περίοδο για την οποία έχουμε στοιχεία. Για αυτή την αξιολόγηση της προβλεπτικής ικανότητας του υποδείγματος, υπάρχουν διάφορα κριτήρια. Τα κυριότερα από αυτά είναι:

- **Ρίζα του μέσου τετραγώνου του σφάλματος** (Root Mean Square Error):

Έστω η ποσότητα: RMSE = $U^2 = \dfrac{\frac{1}{M}\sum_{t-1}^{M}(Y_t^f - Y_t^a)^2}{\frac{1}{M}\sum_{t-1}^{M}(Y_t^a)^2}$

Όπου Y_t^f είναι οι προβλεπόμενες τιμές

Y_t^a είναι οι παρατηρούμενες τιμές

M είναι ο αριθμός των χρονικών περιόδων

- **Μέσο απόλυτο σφάλμα** (Mean Absolute Error):

$$MAE = \frac{1}{M}\sum_{t-1}^{M}\left|Y_t^f - Y_t^a\right|$$

- **Μέσο απόλυτο ποσοστιαίο σφάλμα** (Mean Absolute Percentage Error):

$$MAPE = \frac{1}{M}\sum_{t-1}^{M}\left|\frac{Y_t^f - Y_t^a}{Y_t^a}\right|$$

- **Συντελεστής ανισότητας του Theil** (Theil's Inequality Coefficient):

Αυτός ο συντελεστής ορίζεται ως εξής: $U = \dfrac{\sqrt{\frac{1}{M}\sum_{t-1}^{M}(Y_t^f - Y_t^a)^2}}{\sqrt{\frac{1}{M}\sum_{t-1}^{M}(Y_t^a)^2}}$ **(1)**

Ο συντελεστής ανισότητας U είναι ανεξάρτητος από τις μονάδες μέτρησης, σε αντίθεση με τα προηγούμενα κριτήρια που εξαρτώνται από τις μονάδες μέτρησης, και επομένως θα είναι ο πιο κατάλληλος για τη συγκριση της προβλεπτικής ικανότητας του υποδείγματος. Η τιμή του *U* είναι μηδέν, αν οι προβλεπόμενες τιμές συμπίπτουν απόλυτα με τις πραγματικές. Στη περίπτωση που το *U* είναι μεγαλύτερο της μονάδος, οι προβλέψεις τότε είναι πολύ κακές. Τέλος, αν ο συντελεστής *U* ισούται με τη μονάδα, όλες οι προβλέψεις θα είναι μηδέν. Η περίπτωση αυτή έχει περισσότερο νόημα όταν για τον υπολογισμό του *U* δε χρησιμοποιούνται οι αρχικές τιμές των Y_t^f και Y_t^a, αλλά οι μεταβολές από την προηγούμενη περίοδο. Σε αυτήν την περίπτωση αν *U = 1*, τότε αυτό θα σημαίνει ότι οι προβλεπόμενες μεταβολές είναι μηδέν, δηλαδή θα συνεχιστεί η υπάρχουσα κατάσταση.Τα παραπάνω γράφονται σύντομα ως εξής:

= 0, οι προβλεπόμενες τιμές συμπίπτουν απόλυτα

U: > 1, πολύ κακές προβλέψεις

= 1, οι προβλέψεις είναι μηδέν

Μία παρατήρηση που μπορούμε να κάνουμε στη συνέχεια, είναι ότι ο συντελεστής ανισότητας U, μπορεί να διασπαστεί σε τρείς συνιστώσες, καθεμιά από τις οποίες να εκφράζει μία πηγή ή αιτία της ανακρίβειας των προβλέψεων.Αυτό γίνεται υψώνοντας στο τετράγωνο και τα δύο μέλη της σχέσης (1), οπότε θα προκύψει η

σχέση: $$U^2 = \frac{\frac{1}{M}\sum_{t-1}^{M}(Y_t^f - Y_t^a)^2}{\frac{1}{M}\sum_{t-1}^{M}(Y_t^a)^2} \qquad \textbf{(2).}$$

Ο αριθμητής της σχέσης αυτής, είναι ο μέσος του τετραγώνου του σφάλματος και προσθαφαιρώντας τους μέσους των Y^f και Y^a θα προκύψει:

$$MSE = \frac{1}{M}\sum_{t-1}^{M}\left[(Y_t^f - \overline{Y}^f) - (Y_t^a - \overline{Y}^a) + (\overline{Y}^f - \overline{Y}^a)\right]^2$$ ή αλλιώς

$$MSE = (\overline{Y}^f - \overline{Y}^a)^2 + (s^f - s^a)^2 + 2(1-r)s^f s^a$$

όπου: $\overline{Y}^f$ είναι ο μέσος των προβλεπόμενων τιμών

$\overline{Y}^a$ είναι ο μέσος των πραγματικών τιμών

s^f είναι η τυπική απόκλιση των Y^f

s^a είναι η τυπική απόκλιση των Y^a

ρ είναι ο συντελεστής συσχέτισης των $\hat{u}_t = Y_t - \hat{b}_0 - \hat{b}_1 t$ και Y^a

Ο πρώτος όρος στη σχέση αυτή, δηλαδή το τετράγωνο της διαφοράς των μέσων, αποτελεί ένα μέτρο μεροληψίας. Ο δεύτερος όρος, δηλαδή το τετράγωνο της διαφοράς των διακυμάνσεων, αποτελεί ένα μέτρο της άνισης μεταβλητικότητας των Y^f και Y^a. Τέλος, ο τρίτος όρος περιέχει το συντελεστή αυτοσυσχέτισης και μπορεί να θεωρηθεί ως ένα μέτρο της ατελούς συμμεταβλητικότητας των Y^f και Y^a. Η σχέση αυτή, διαιρώντας και τα δύο μέλη της με το MSE, μπορεί να γραφεί και ως εξής:

$$1 = \frac{(\overline{Y}^f - \overline{Y}^a)^2}{MSE} + \frac{(s^f - s^a)^2}{MSE} + \frac{2(1-r)s^f s^a}{MSE}$$

Ενώ η σχέση **(2),** γίνεται:

$$U^2 = \frac{(\overline{Y}^f - \overline{Y}^a)^2}{A} + \frac{(s^f - s^a)^2}{A} + \frac{2(1-r)s^f s^a}{A}, \text{ όπου } A = \frac{1}{M}\sum_{t-1}^{M}(Y_t^a)^2$$

Ο τρίτος όρος στις παραπάνω σχέσεις, παριστάνει τον τυχαίο παράγοντα που δεν μπορεί να αποφευχθεί, ενώ οι δύο πρώτοι όροι παριστάνουν συστηματικά σφάλματα που πρέπει να αποφεύγονται. Το ιδανικό θα ήταν οι δυο πρώτοι όροι να ήταν μηδέν, και η μόνη αιτία σφάλματος να ήταν το μη συστηματικό (τυχαίο) μέρος.

Μέχρι τώρα εξετάσαμε τις βασικές τεχνικές ανάλυσης χρονολογικών σειρών, αλλά και μη στάσιμων που μετατρέπονται σε στάσιμες με τη χρήση των διαφορών. Στη συνέχεια θα ασχοληθούμε με την επίλυση του προβλήματος που προκύπτει στα υποδείγματα παλινδρόμησης όταν χρησιμοποιούνται σειρές που **δεν είναι στάσιμες.**

Αν οι μεταβλητές της χρονολογικής σειράς δεν είναι στάσιμες, οι εκτιμητές που προκύπτουν από τη μέθοδο των ελαχίστων τετραγώνων δε θα είναι συνεπείς, με αποτέλεσμα ο στατιστικός έλεγχος να μην είναι έγκυρος. Γνωρίζουμε ότι τα αποτελέσματα από την εκτίμηση μιας παλινδρόμησης ανάμεσα σε δύο ή περισσότερες μεταβλητές αξιολογούνται με βάση τα συνηθισμένα στατιστικά κριτήρια, όπως τα R^2, t, ή F. η αξιολόγηση αυτή όμως θα είναι έγκυρη μόνο αν οι μεταβλητές μας είναι στάσιμες. Επομένως, στην περίπτωση που οι μεταβλητές δεν είναι στάσιμες, τα στατιστικά αποτελέσματα που θα πάρουμε από αυτά τα κριτήρια μπορεί να είναι πολύ ικανοποιητικά, δηλαδή να έχουμε υψηλή τιμή του συντελεστή προσδιορισμού R^2 και σημαντικές τιμές του t, αλλά να μην έχουν καμία ουσιαστική σημασία. Αυτό σημαίνει ότι η παρατηρούμενη στατιστικά σημαντική σχέση ανάμεσα στις μεταβλητές μπορεί να οφείλεται στην ασυνέπεια των εκτιμητών και να μη συνεπάγεται αναγκαστικά και την ύπαρξη αιτιώδους σχέσης ανάμεσα στις μεταβλητές. Αυτό το αποτέλεσμα περιγράφεται με τον όρο φαινομενική παλινδρόμηση (spurious regression), σύμφωνα με τους Granger και Newbold.

Οι οικονομικές χρονολογικές σειρές χαρακτηρίζονται συνήθως από τάση (trend) η οποία τις καθιστά μη στάσιμες. (ολοκληρωμένες). Με τον όρο τάση, εννοούμε τη συνεχή, διαχρονική αύξηση ή μείωση των τιμών μιας χρονολογικής σειράς. Οπότε, αν μια μεταβλητή χαρακτηρίζεται από τάση, ο μέσος και ίσως και η διακύμανσή της θα μεταβάλλονται με το χρόνο, πράγμα που σημαίνει ότι η σειρά δεν είναι στάσιμη. Αυτό συνεπάγεται ότι η παρατηρούμενη σχέση ανάμεσα σε δύο μεταβλητές μπορεί να μην είναι πραγματική αλλά φαινομενική. Μια συνηθισμένη πρακτική για να αντιμετωπιστεί αυτό το πρόβλημα, είναι να συμπεριλάβουμε στο υπόδειγμα το χρόνο ως μια ερμηνευτική μεταβλητή, αφού με την εισαγωγή του χρόνου ως ανεξάρτητη μεταβλητή απαλείφουμε την τάση από τα δεδομένα. Αυτή η μέθοδος είναι γνωστή ως detrending και θεωρείται έγκυρη μόνο στην περίπτωση που η χρονολογική σειρά είναι στάσιμη ως προς την τάση(trend stationary ή TSP) και όχι στάσιμη ως προς τις διαφορές (difference stationary ή DSP). Στάσιμη ως προς την τάση θεωρείται μία σειρά, όταν καθίσταται στάσιμη αφαιρώντας την τάση, ενώ στάσιμη ως προς τις διαφορές θεωρείται μια σειρά, όταν καθίσταται στάσιμη με τη χρήση των διαφορών. Θα δούμε στη συνέχεια, δύο παραδείγματα υποδειγμάτων, που το ένα είναι στάσιμο ως προς την τάση, ενώ το άλλο είναι στάσιμο ως προς τις διαφορές.

a. Έστω το υπόδειγμα:

$Y_t = b_0 + b_1 t + u_t$, όπου u_t είναι λευκός θόρυβος

Η εκτιμημένη σειρά των καταλοίπων η οποία θα προκύψει από τη μέθοδο των ελαχίστων τετραγώνων θα είναι η: $H_1 : a \neq 1$ και είναι λευκός θόρυβος, που όπως γνωρίζουμε είναι στάσιμη σειρά. Τα κατάλοιπα $\hat{u}_t$ εκφράζουν τη μεταβλητή Y_t από την οποία έχει αφαιρεθεί η τάση. Επομένως, η σειρά Y_t του υποδείγματος είναι στάσιμη ως προς την τάση.

b. Έστω τώρα το υπόδειγμα:

$Y_t = a + Y_{t-1} + u_t$

Η σειρά των πρώτων διαφορών θα είναι: $\Delta Y_t = Y_t + Y_{t-1} = a + u_t$

Όπως γνωρίζουμε, η σειρά αυτή είναι στάσιμη. Οπότε, η σειρά Y_t αυτού του υποδείγματος είναι μια σειρά στάσιμη ως προς τις διαφορές. Το πρόβλημα που παρουσιάζεται είναι πως στην πράξη δεν είναι εύκολος ο διαχωρισμός τους, γιατί επιδεικνύουν παρόμοια συμπεριφορά.

Όταν η τάση που χαρακτηρίζει μια χρονολογική σειρά είναι της μορφής του πρώτου υποδείγματος ονομάζεται *προσδιοριστική* (deterministic), με την έννοια ότι είναι απόλυτα προβλέψιμη και δεν μεταβάλλεται. Αντίθετα, η τάση που χαρακτηρίζει το δεύτερο υπόδειγμα ονομάζεται *στοχαστική* (stochastic), με την έννοια ότι μεταβάλλεται. Επομένως, όταν η τάση είναι προσδιοριστική και γίνεται απαλοιφή της τάσεως με την εισαγωγή του χρόνου ως ερμηνευτικής μεταβλητής, δεν εμφανίζεται το πρόβλημα της φαινομενικής παλινδρόμησης. Η παλινδρόμηση θα είναι φαινομενική όταν οι σειρές χαρακτηρίζονται από στοχαστική τάση. Σε αυτή την περίπτωση θα πρέπει να χρησιμοποιούνται οι πρώτες διαφορές.

Είδαμε προηγουμένως πώς γίνεται έλεγχος στασιμότητας χρονολογικών σειρών με τη βοήθεια της δειγματικής συνάρτησης αυτοσυσχέτισης. Ένας άλλος τρόπος που χρησιμοποιείται ευρύτατα στην ανάλυση χρονολογικών σειρών είναι οι *έλεγχοι μοναδιαίας ρίζας* (unit root tests). Υπάρχουν δύο έλεγχοι μοναδιαίας ρίζας, και αυτοί είναι οι έλεγχοι **Dickey-Fuller** (Dickey-Fuller tests), από τα ονόματα των ερευνητών που πρότειναν τους πρώτους ελέγχους αυτού του είδους, και οι έλεγχοι **Phillips-Perron**. Στη συνέχεια, θα εξετάσουμε αναλυτικά τους παραπάνω ελέγχους.

- Έστω ότι μια δεδομένη οικονομική χρονολογική σειρά ,μπορεί να περιγραφεί με το **AR(1)** υπόδειγμα:

$$Y_t = aY_{t-1} + u_t$$

Η σειρά θα είναι στάσιμη αν -1 < α <1. Αν α = 1, η σειρά δε θα είναι στάσιμη. Με τον έλεγχο Dickey-Fuller γίνεται έλεγχος της μηδενικής υπόθεσης $H_0 : b = 0$ έναντι της εναλλακτικής $H_1 : a \neq 1$. Με λίγα λόγια, γίνεται έλεγχος για την ύπαρξη μοναδιαίας ρίζας, που συνεπάγεται μη στασιμότητα. Ο πιο απλός τρόπος για να ελεγχθεί αυτή η υπόθεση είναι να εκτιμηθεί το αυτοπαλίνδρομο υπόδειγμα με τη μέθοδο των ελαχίστων τετραγώνων και στη συνέχεια, να γίνει ο συνήθης έλεγχος με την κατανομή t. Όμως, το πρόβλημα εδώ είναι ότι αν ισχύει η μηδενική υπόθεση ο έλεγχος αυτός δεν είναι έγκυρος, γιατί σε αυτήν την περίπτωση οι κατανομές του *t* ή του *F* δε συμπίπτουν με τις γνωστές κατανομές *t* ή *F*, και επομένως οι κρίσιμες τιμές τους δε θα είναι κατάλληλες για τον έλεγχο της υπόθεσής μας. Ένα ακόμα πρόβλημα είναι ότι, ο εκτιμητής του a, ο $\hat{a}$, που θα προκύψει από τη μέθοδο των ελαχίστων τετραγώνων δε θα έχει τις επιθυμητές ιδιότητες, και συγκεκριμένα θα είναι μεροληπτικός και ασυνεπής.

Λύση στα παραπάνω προβλήματα έδωσαν οι Dickey και Fuller κάνοντας επαναπαραμετροποίηση (reparameterization) του υποδείγματος ως εξής:

$Y_t = Y_{t-1} + u_t$

Αφαιρούμε την ποσότητα Y_{t-1} και από τα δύο μέλη, οπότε έχουμε:

$Y_t - Y_{t-1} = aY_{t-1} - Y_{t-1} + u_t$, ή αλλιώς

$\Delta Y_t = bY_{t-1} + u_t$ όπου $\beta = \alpha$ *-1.*

Αυτή η διαδικασία έγινε γιατί τώρα ο έλεγχος της μηδενικής υπόθεσης $H_0 : a = 1$ μετατράπηκε σε έλεγχο της μηδενικής υπόθεσης $H_0 : b = 0$. Ο έλεγχος αυτής της μηδενικής υπόθεσης μπορεί να γίνει με τη βοήθεια των πινάκων της κατανομής που κατασκεύασαν οι Dickey και Fuller. Για διάκριση, οι κρίσιμες τιμές αυτής της κατανομής συμβολίζονται με το ελληνικό γράμμα τ. Ο έλεγχος είναι ως εξής: Αν τ > $|t|$, τότε αποδεχόμαστε τη μηδενική υπόθεση ότι β = 0, οπότε καταλήγουμε στο συμπέρασμα ότι η σειρά δεν είναι στάσιμη. Διαφορετικά, αν τ ≤ $|t|$, τότε απορρίπτουμε τη μηδενική υπόθεση και συμπεραίνουμε ότι η σειρά είναι στάσιμη. Αν μετά από αυτόν τον έλεγχο καταλήξουμε στο συμπέρασμα ότι η σειρά δεν είναι στάσιμη, αυτό δεν σημαίνει ότι η σειρά των πρώτων διαφορών ΔY_t θα είναι οπωσδήποτε στάσιμη. Θα πρέπει να επαναλάβουμε τον έλεγχο στη σειρά ΔY_t, και αν ούτε αυτή είναι στάσιμη, θα συνεχίσουμε τον έλεγχο στις δεύτερες, τρίτες κ.ο.κ. διαφορές, μέχρις ότου να απορριφθεί η υπόθεση της μη στασιμότητας.

Έστω τώρα ότι στο αρχικό αυτοπαλίνδρομο υπόδειγμα πρώτης τάξης πρόσθέσουμε σταθερό όρο, έστω δ.

Τότε, το υπόδειγμα θα γίνει:

$Y_t = d + aY_{t-1} + u_t$

Και το μετασχηματισμένο υπόδειγμα θα είναι:

$\Delta Y_t = d + bY_{t-1} + u_t$, όπου β = α-1.

Σε αυτό το υπόδειγμα, ο έλεγχος για την ύπαρξη μοναδιαίας ρίζας, γίνεται ακριβώς με τον ίδιο τρόπο που γίνεται και ο έλεγχος στο υπόδειγμα που δεν έχει σταθερό όρο, με τη μόνη διαφορά ότι σε αυτήν τη περίπτωση οι κρίσιμες τιμές της τα είναι διαφορετικές.

Οι Dickey- Fuller παρέχουν επίσης, πίνακα με τροποποιημένες τιμές της κατανομής F για τον έλεγχο της από κοινού υπόθεσης $H_0 : d = b = 0$. Οι τιμές αυτές παριστάνονται με Φ_1.

Αυτός ο έλεγχος της μοναδιαίας ρίζας των Dickey- Fuller που εφαρμόσαμε σε ένα αυτοπαλίνδρομο υπόδειγμα πρώτης τάξης, μπορεί επίσης να εφαρμοστεί και στη γενική περίπτωση μιας AR(p) διαδικασίας.

Ως γνωστόν, ένα αυτοπαλίνδρομο υπόδειγμα p τάξης διατυπώνεται ως εξής:

$$Y_t = d + aY_{t-1} + a_2 Y_{t-2} + ... + a_p Y_{t-p} u_t$$

Και η τροποποιημένη του μορφή θα είναι:

$$\Delta Y_t = d + bY_{t-1} + a_1^* \Delta Y_{t-1} + a_2^* \Delta Y_{t-2} ... + a_{p-1}^* \Delta Y_{t-p+1} u_t$$

όπου $\Delta Y_{t-1} = Y_{t-1} - Y_{t-2}$

$\Delta Y_{t-2} = Y_{t-2} - Y_{t-3}$ κ.τ.λ

και $b = (a_1 + a_2 + ... + a_p - 1)$

επίσης, a_j^*, για j = 1, 2,…p, είναι συναρτήσεις των αρχικών συντελεστών a_i για i= 1, 2,…p. Ο έλεγχος για την ύπαρξη μοναδιαίας ρίζας, για το αν δηλαδή η σειρά είναι στάσιμη ή όχι, θα γίνει με τον έλεγχο της μηδενικής υπόθεσης $H_0 : b = 0$ έναντι της εναλλακτικής $H_1 : b \neq 0$. Για να γίνει αυτός ο έλεγχος θα εκτιμηθεί το τροποποιημένο υπόδειγμα με τη μέθοδο των ελαχίστων τετραγώνων. Η απόρριψη της μηδενικής υπόθεσης συνεπάγεται ότι η σειρά είναι στάσιμη. Διαφορετικά, αν την αποδεχτούμε, η σειρά θα είναι μη στάσιμη. Ο έλεγχος αυτός για την ύπαρξη μοναδιαίας ρίζας σε αυτοπαλίνδρομα υποδείγματα τάξεως $p > 1$ είναι γνωστός ως επαυξημένος έλεγχος Dickey- Fuller (Augmented Dickey- Fuller test).

Ο έλεγχος Dickey- Fuller μπορεί να εφαρμοστεί και στην περίπτωση που στο υπόδειγμα περιλαμβάνεται και ο χρόνος ως ερμηνευτική μεταβλητή. Δηλαδή, στο υπόδειγμα της μορφής:

$$Y_t = d + bY_{t-1} + gt + u_t$$

Θα αφαιρέσουμε από τα δύο μέλη το Y_{t-1}, όπως και προηγουμένως, οπότε θα προκύψει η σειρά:

$$\Delta Y_t = d + bY_{t-1} + gt + u_t \text{, όπου } b = b_1 - 1$$

Αν β = 0 και γ = 0, τότε η σειρά που σχηματίστηκε χαρακτηρίζεται από στοχαστική και όχι προσδιοριστική τάση, επομένως είναι στάσιμη ως προς τις διαφορές. Στη συνέχεια μπορεί να γίνει ο έλεγχος για την ύπαρξη μοναδιαίας ρίζας με τον τρόπο που περιγράφηκε προηγουμένως. Επίσης, η μέθοδος Dickey- Fuller μπορεί να εφαρμοστεί και για τον έλεγχο της από κοινού υπόθεσης $H_0 : b = g = 0$, χρησιμοποιώντας τις τροποποιημένες τιμές της F κατανομής. Τέλος, με αυτή τη μέθοδο γίνεται επίσης ο έλεγχος της υπόθεσης ότι όλοι οι συντελεστές του υποδείγματος είναι μηδέν, δηλαδή ο έλεγχος $H_0 : d = b = g = 0$.

Οι τιμές της στατιστικής τα για τον έλεγχο της μοναδιαίας ρίζας σε μία χρονολογική σειρά εξαρτώνται από τη μορφή της εξίσωσης Dickey- Fuller, δηλαδή αν στην παλινδρόμηση περιλαμβάνεται σταθερός όρος ή ο χρόνος συμπεριλαμβάνεται ως παλινδρομητής. Αυτά τα δύο είναι πιθανό να συμβαίνουν ταυτόχρονα.

Οι κρίσιμες τιμές της τα αυξάνουν στην περίπτωση που προστίθεται σταθερός όρος ενώ, γίνονται ακόμα μεγαλύτερες όταν προστίθεται και χρονική τάση με την είσοδο του χρόνου. Τα αποτελέσματα του ελέγχου για την ύπαρξη μοναδιαίας ρίζας εξαρτώνται άμεσα από την εξειδίκευση της εξίσωσης Dickey- Fuller επομένως, αν η εξειδίκευση δεν είναι σωστή δε θα εκφράζει τη στοχαστική διαδικασία που παρήγαγε τα δεδομένα, και τελικά θα καταλήξουμε σε λανθασμένα συμπεράσματα. Στη συνέχεια, θα αναφέρουμε τη διαδικασία που πρέπει να ακολουθηθεί έτσι ώστε να γνωρίζουμε πότε η εξίσωση παλινδρόμησης θα πρέπει να περιλαμβάνει σταθερό όρο ή/ και χρονική τάση, στην περίπτωση φυσικά που η μορφή της στοχαστικής διαδικασίας που παρήγαγε τα δεδομένα είναι άγνωστη. Η διαδικασία περιλαμβάνει πέντε βήματα τα οποία θα αναπτύξουμε αναλυτικά παρακάτω:

Βήμα 1: Έστω η εξίσωση Dickey- Fuller στη γενική της μορφή:

$$\Delta Y_t = d + bY_{t-1} + gt + \Delta Y_{t-i}u_t$$

Αφού εκτιμηθεί η παραπάνω εξίσωση, θα γίνει έλεγχος της μηδενικής υπόθεσης $H_0 : b = 0$. Αν η μηδενική υπόθεση απορριφθεί, η διαδικασία τερματίζεται και συμπεραίνουμε ότι δεν υπάρχει μοναδιαία ρίζα. Αντίθετα, αν η μηδενική υπόθεση αποδεχτεί, συνεχίζουμε με το βήμα 2.

Βήμα 2: Στη συνέχεια, γίνεται έλεγχος για την ύπαρξη τάσης δεδομένου ότι β = 0.

Ελέγχουμε με λίγα λόγια την υπόθεση ότι γ = 0, δεδομένου ότι β = 0. Ο έλεγχος αυτός θα πρέπει να επιβεβαιώνεται ελέγχοντας την υπόθεση β = γ = 0. Αν απορριφθεί η μηδενική υπόθεση, γίνεται επανέλεγχος, χρησιμοποιώντας αυτή τη φορά την τυποποιημένη κανονική κατανομή. Αν η υπόθεση β = 0 δε γίνει αποδεκτή, η διαδικασία τερματίζεται και καταλήγουμε στο συμπέρασμα ότι δεν υπάρχει μοναδιαία ρίζα. Διαφορετικά, αν αποδεχτούμε ότι β = 0, συμπεραίνεται ότι υπάρχει μοναδιαία ρίζα. Αν επίσης, η μηδενική υπόθεση γ = 0, δεδομένου ότι β = 0, γίνει αποδεκτή, προχωρούμε στο επόμενο βήμα.

Βήμα 3: Σε αυτό το βήμα εκτιμάται η παλινδρόμηση, χωρίς να λάβουμε υπ' όψιν μας το χρόνο ως ερμηνευτική μεταβλητή. Με λίγα λόγια, εκτιμάται η παλινδρόμηση:

$$\Delta Y_t = d + bY_{t-1} + \sum a_1^* Y_{t-1} + u_t$$

Στη συνέχεια, γίνεται έλεγχος της μηδενικής υπόθεσης $H_0 : b = 0$. Αν αυτή απορριφθεί, συμπεραίνουμε ότι δεν υπάρχει μοναδιαία ρίζα και η διαδικασία τερματίζεται. Διαφορετικά, αν αποδεχτούμε τη μηδενική υπόθεση, η διαδικασία συνεχίζεται με το επόμενο βήμα.

Βήμα 4: σε αυτό το βήμα γίνεται έλεγχος για το σταθερό όρο. Ελέγχεται δηλαδή, η υπόθεση ότι ο σταθερός όρος είναι μηδέν. Η υπόθεση θα είναι ότι δ = 0, δεδομένου ότι β = 0. Ο έλεγχος αυτής της υπόθεσης θα πρέπει να επιβεβαιώνεται ελέγχοντας την υπόθεση $\delta = \beta = 0$.

Αν η μηδενική υπόθεση απορριφθεί, δηλαδή αν ισχύει $\delta \neq 0$, ελέγχεται ξανά η υπόθεση $H_0 : b = 0$ χρησιμοποιώντας την τυποποιημένη κανονική κατανομή. Αν αποδεχτούμε αυτήν την υπόθεση, τότε συμπεραίνουμε ότι υπάρχει μοναδιαία ρίζα. Αν τελικά αποδεχτούμε τη μηδενική υπόθεση δ = 0, δεδομένου ότι β = 0, επομένως αν δεχτούμε ότι ο σταθερός όρος είναι μηδέν, τότε συνεχίζουμε τη διαδικασία με το βήμα 5 που είναι και το τελευταίο.

Βήμα 5: σε αυτό το βήμα θα εκτιμηθεί η εκ νέου παλινδρόμηση:

$\Delta Y_t = d + bY_{t-1} + \sum a_1^* Y_{t-1} + u_t$ χωρίς σταθερό όρο και χωρίς χρονική τάση. Στη συνέχεια, θα γίνει έλεγχος της μηδενικής υπόθεσης $H_0 : b = 0$. Αν η μηδενική αυτή υπόθεση γίνει δεκτή, θα καταλήξουμε στο συμπέρασμα ότι υπάρχει μοναδιαία ρίζα. Διαφορετικά, αν δηλαδή β ≠ 0, συμπεραίνουμε ότι δεν υπάρχει μοναδιαία ρίζα. Σε αυτό το σημείο τελειώνει και η διαδικασία.

Θα πρέπει να σημειωθεί ότι αυτή η διαδικασία που περιγράφηκε παραπάνω είναι πραγματικά μια πολύ χρήσιμη διαδικασία για την εξακρίβωση ύπαρξης των προσδιοριστικών όρων, δηλαδή την ύπαρξη σταθερού όρου και χρονικής τάσης που μπορεί να υπεισέρχονται στην εξίσωση παλινδρόμησης για τον έλεγχο μοναδιαίας ρίζας. Όμως, δεν πρέπει να εφαρμόζεται μηχανική σε κάθε περίπτωση γιατί, μπορεί να προκύψει ποιοι θα πρέπει να είναι οι κατάλληλοι παλινδρομητές είτε από την οικονομική θεωρία είτε από τη γραφική απεικόνιση των δεδομένων.

- Αφού εξετάσαμε αναλυτικά πώς γίνονται οι έλεγχοι Dickey- Fuller, θα ασχοληθούμε τώρα με τους ελέγχους Phillips –Perron. Η διαφορά τους έγκειται στο ότι ενώ οι έλεγχοι Dickey- Fuller υποθέτουν ότι οι διαφορετικοί όροι δεν αυτοσυσχετίζονται και ότι έχουν σταθερή διακύμανση, οι έλεγχοι Phillips –Perron γενικεύουν τη προσέγγιση αυτή χωρίς τις αυστηρές προϋποθέσεις για την κατανομή των διαταρακτικών όρων. Οι Phillips –Perron υποθέτουν ότι ο διαταρακτικός όρος e_t έχει μέσο το μηδέν και ότι τα δεδομένα έχουν παραχθεί από τη σχέση $Y_t = Y_{t-1} + e_t$. Με βάση τις υποθέσεις αυτές αναπτύσσουν στατιστικές ελέγχου οι οποίες είναι τροποποιημένες στατιστικές t ή F, οι κρίσιμες τιμές των οποίων είναι ακριβώς οι ίδιες με αυτές των Dickey- Fuller. Οι πιο συνηθισμένοι αλλά και χρήσιμοι έλεγχοι Phillips –Perron αφορούν τον έλεγχο υποθέσεων για τους συντελεστές στα εξής δύο υποδείγματα:

$$Y_t = a_0 + a_1 Y_{t-1} + u_t \quad \text{και}$$

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 t + u_t$$

Στη συνέχεια, θα μιλήσουμε για τη συνολοκλήρωση δύο χρονολογικών σειρών και για το λόγο για τον οποίο γίνεται. Είδαμε προηγουμένως ότι, όταν τα δεδομένα που χρησιμοποιούνται σε ένα υπόδειγμα παλινδρόμησης δεν προέρχονται από στάσιμες χρονολογικές σειρές, τότε προκύπτει το πρόβλημα της φαινομενικής παλινδρόμησης και σε αυτήν την περίπτωση χρησιμοποιούνται οι πρώτες, οι δεύτερες, ή κ.τ.λ. διαφορές για την αντιμετώπιση του προβλήματος και όχι τα επίπεδα των μεταβλητών. Τις περισσότερες φορές όμως, αυτό που ενδιαφέρει τους οικονομολόγους είναι ότι οι μακροχρόνιες σχέσεις μεταξύ των μεταβλητών και αυτό σημαίνει τα επίπεδά τους και όχι οι διαφορές.

Έστω για παράδειγμα το υπόδειγμα:

$$Y_t = b_0 + b_1 X_t + b_2 X_{t-1} + b_3 X_{t-2} + g_1 Y_{t-1} + g_2 Y_{t-2} + u_t$$
$$a_1(1-g_1)$$
$$\Delta Y_t = b_1 \Delta X_t - (1-g_1)(Y_{t-1} - a_0 - a_1 X_{t-1}) + u_t$$
$$Y_{t-1}$$
$$0 < g_1 < 1$$
$$\Delta Y_t = (g_1 - 1)\Delta Y_{t-1} + b_1 \Delta X_t + (b_1 + b_2)\Delta X_{t-1} - (1 - g_1 - g_2)(Y_{t-2} - a_0 - a_1 X_{t-2}) + u_t$$
$$Y_t = a_0 + a_1 X_t$$

Όπως είναι γνωστό από την οικονομική θεωρία, οι τιμές των μεταβλητών δε μεταβάλλονται, δηλαδή παραμένουν σταθερές, όταν βρίσκονται σε κατάσταση μακροχρόνιας σταθεράς ισορροπίας (steady state equilibrium). Έστω ότι Y^e και X^e είναι τα επίπεδα ισορροπίας των Y και X αντίστοιχα. Τότε, θα ισχύει ότι:

$$Y_t = Y_{t-1} = Y^e, \quad X_t = X_{t-1} = X^e \quad \text{και} \quad e_t = 0$$

Επομένως, η μορφή του υποδείγματος σε κατάσταση ισορροπίας θα είναι:

$$Y^e = b_0 + b_1 X^e + b_2 X^e + g_1 Y^e \quad \text{ή}$$

$$Y^e - g_1 Y^e = b_0 + b_1 X^e + b_2 X^e \quad \text{ή}$$

$(1-g_1)Y^e = b_0 + b_1 X^e + b_2 X^e$ και διαιρώντας και τα δύο μέλη με το $1-g_1$ γίνεται:

$$Y^e = \frac{b_0}{1-g_1} + \frac{b_1 + b_2}{1-g_1} X^e \qquad \text{ή αλλιώς}$$

$$Y^e = a_0 + a_1 X^e, \quad \text{όπου} \qquad a_0 = \frac{b_0}{1-g_1} \quad \text{και} \quad a_1 = \frac{b_1 + b_2}{1-g_1}$$

Αν τώρα χρησιμοποιήσουμε τις πρώτες διαφορές, το αρχικό υπόδειγμα θα πάρει τη μορφή:

$$\Delta Y_t = b_1 \Delta X_t + b_2 \Delta X_{t-1} + g_1 \Delta Y_{t-1} + u_t$$

Είναι προφανές ότι σε κατάσταση ισορροπίας όλες οι διαφορές είναι μηδέν. Αυτό σημαίνει ότι δεν υπάρχει λύση σε όρους Y^e και X^e, δηλαδή δεν μπορούμε να βγάλουμε κανένα συμπέρασμα για τη μακροχρόνια σχέση ανάμεσα στις μεταβλητές *Y* και *X*. Αυτό το πρόβλημα που ανακύπτει όταν οι χρονολογικές σειρές είναι ολοκληρωμένες και προκύπτει αδυναμία χρησιμοποίησης των πρώτων διαφορών, αλλά και η μη καταλληλότητα των επιπέδων, μπορεί να παρακαμφθεί αν οι σειρές είναι *συνολοκληρωμένες* (cointegrated).

<u>Ορισμός</u>: Δύο χρονολογικές σειρές ονομάζονται **συνολοκληρωμένες τάξεως** *d, b* αν

1. Είναι και οι δύο ολοκληρωμένες τάξεως *d*
2. Υπάρχει ένας γραμμικός συνδυασμός των δύο που αποτελεί ολοκληρωμένη σειρά τάξεως (*d* – *b*), όπου *b* > 0.

Εμείς θα περιοριστούμε στην περίπτωση που *d* = *b*= *1*, οπότε ο γραμμικός συνδυασμός των δύο σειρών θα είναι Ι(0), δηλαδή θα είναι σταθερή σειρά. Από τη στιγμή βέβαια, που και οι δύο σειρές είναι μη στάσιμες λογικό είναι και ο γραμμικός τους συνδυασμός να είναι μη στάσιμη σειρά. Όμως, για να θεωρηθούν οι σειρές συνολοκληρωμένες θα πρέπει να υπάρχει ένας γραμμικός συνδυασμός που να είναι στάσιμη σειρά. Η έννοια της συνολοκλήρωσης αποτελεί το συνδετικό κρίκο ανάμεσα στις σχέσεις που υπάρχουν ανάμεσα σε ολοκληρωμένες μεταβλητές και στην έννοια της μακροχρόνιας ισορροπίας. Η συνολοκλήρωση, με λίγα λόγια, είναι η συνέπεια ύπαρξης μακροχρόνιας σχέσης ανάμεσα σε οικονομικές μεταβλητές.

Έστω, για παράδειγμα μια μακροχρόνια σχέση ανάμεσα στις μεταβλητές Y και X:

$$Y_t = a_0 + a_1 X_t$$

Αν υποθέσουμε ότι η μεταβλητή Y χαρακτηρίζεται από τάση και ότι είναι I(1), ενώ η μεταβλητή X αντιπροσωπεύει μια στάσιμη χρονολογική σειρά, τότε η παραπάνω σχέση μακροχρόνιας ισορροπίας δεν μπορεί να ισχύει ούτε προσεγγιστικά. Αυτό σημαίνει, ότι στην περίπτωση των μεταβλητών Y και X, η σχέση ισορροπίας ανάμεσά τους προϋποθέτει και οι δύο σειρές να είναι ολοκληρωμένες της ίδιας τάξεως, και αντίστροφα. Αν τελικά, οι μεταβλητές Y και X βρίσκονται συνεχώς σε κατάσταση ισορροπίας, τότε προκύπτει η σχέση:

$$Y_t - a_0 - a_1 X_t = 0$$

Προφανώς, η σχέση αυτή δεν μπορεί να ισχύει συνεχώς. Η διαφορά $u_t = Y_t - a_0 - a_1 X_t$ παριστάνει την έκταση της ανισορροπίας ανάμεσα στις μεταβλητές Y και X. Το u_t ονομάζεται σφάλμα ανισορροπίας (disequilibrium error) ή σφάλμα ισορροπίας. Θα πρέπει τα σφάλματα ανισορροπίας που παρατηρούνται διαχρονικά, να κυμαίνονται γύρω από το μηδέν και να αποτελούν στάσιμη σειρά, διαφορετικά δε θα έχει νόημα η σχέση ισορροπίας. Στην περίπτωση που τα u_t δεν είναι στάσιμη σειρά, αλλά για παράδειγμα εμφανίζουν αυξητική τάση, τότε οι μεταβλητές Y και X διαχρονικά θα απομακρύνονται η μία από την άλλη συνεχώς. Είναι φανερό ότι μια τέτοια συμπεριφορά δε συμβαδίζει με την ύπαρξη μακροχρόνιας ισορροπίας ανάμεσα στις δύο μεταβλητές. Αν όμως το σφάλμα ισορροπίας u_t είναι I(*0*), τότε ο γραμμικός συνδυασμός των μεταβλητών *Y* και *X*, που παριστάνεται από τη σχέση $u_t = Y_t - a_0 - a_1 X_t$, θα πρέπει να είναι επίσης στάσιμη σειρά. Με λίγα λόγια, η ύπαρξη σχέσης ισορροπίας ανάμεσα σε δύο μεταβλητές προϋποθέτει ένα γραμμικό συνδυασμό των δύο που να αποτελεί στάσιμη σειρά, και αντίστροφα.

Στη συνέχεια, θα ασχοληθούμε με το πώς γίνεται ο έλεγχος της συνολοκλήρωσης ανάμεσα σε δύο μεταβλητές.

Έστω η προηγούμενη μακροχρόνια σχέση ανάμεσα στις μεταβλητές *Y* και *X*:

$$Y_t = a_0 + a_1 X_t$$

Αν οι μεταβλητές *Y* και *X* είναι *I(1),* θα είναι συνολοκληρωμένες μόνο αν ο γραμμικός συνδυασμός τους $u_t = Y_t - a_0 - a_1 X_t$ είναι *I(0),* μόνο δηλαδή αν είναι στάσιμη σειρά. Η υποθετική σχέση $Y_t = a_0 + a_1 X_t$ ανάμεσα στις δύο μεταβλητές ονομάζεται παλινδρόμηση συνολοκλήρωσης (cointegrated regression), ή στατική παλινδρόμηση (static regression). Η παράμετρος a_1 αναφέρεται ως παράμετρος συνολοκλήρωσης (cointegrated parameter).Εμείς θα περιοριστούμε στην περίπτωση του διμεταβλητού υποδείγματος. Στην περίπτωση που υπάρχουν περισσότερες από μία ερμηνευτικές μεταβλητές θα υπάρχουν και περισσότερες από μια παράμετροι. Σε αυτήν τη περίπτωση θα υπάρχει το διάνυσμα ή τα διανύσματα συνολοκλήρωσης. Ο έλεγχος για τη συνολοκλήρωση των μεταβλητών *Y* και *X* βασίζεται στη συμπεριφορά των καταλοίπων που προκύπτουν από τη μέθοδο των ελαχίστων τετραγώνων. Ελέγχουμε δηλαδή, αν τα εκτιμώμενα κατάλοιπα $\hat{u}_t = Y_t - \hat{a}_0 - \hat{a}_1 X_t$ αποτελούν στάσιμη σειρά, με λίγα λόγια ελέγχουμε αν τα κατάλοιπα είναι *I(0).*

Οι πιο συνηθισμένοι τρόποι ελέγχου είναι δύο και θα τους δούμε αναλυτικά στη συνέχεια. Ο πρώτος είναι ο τροποποιημένος έλεγχος Dickey- Fuller ή αλλιώς έλεγχος Engle- Granger και ο δεύτερος είναι ο έλεγχος της παλινδρόμησης συνολοκλήρψσης κατά Durbin- Watson ο οποίος συμβολίζεται και ως CRDW (cointegrating regression Durbin- Watson).

Έλεγχος Engle- Granger

Έστω ότι τα κατάλοιπα αποτελούν ένα αυτοπαλίνδρομο υπόδειγμα πρώτης τάξεως, δηλαδή είναι AR(*1*).Τότε:

$$\hat{u}_t = r\,\hat{u}_{t-1} + e_t$$

Παίρνουμε δύο περιπτώσεις:

♦ αν το $\rho = 1$, τότε τα κατάλοιπα $\hat{u}_t$ δεν αποτελούν στάσιμη χρονολογική σειρά.

♦ Αν το $|r| < 1$, τότε τα $\hat{u}_t$ θα είναι στάσιμη σειρά.

Επομένως, ο έλεγχος θα γίνει για να διαπιστωθεί αν $\rho = 1$ ή όχι.

Στη συνέχεια, αφαιρούμε το $\hat{u}_{t-1}$ και από τα δύο μέλη της σχέσης $\hat{u}_t = r\hat{u}_{t-1} + e_t$,

r^*

$|t| < t_a$

$\hat{u}_t = r\hat{u}_{t-1} + e_t$

οπότε θα έχουμε τη σχέση: $\Delta\hat{u}_t = r^*\hat{u}_{t-1} + e_t$, όπου $r \to 1$ $= \rho - 1$.

$d \to 0$

$a = 5\%$

$d \leq 0.386$

Τελικά, ο έλεγχος με τη μηδενική υπόθεση την $H_0: r = 1$ μετατρέπεται τώρα στον έλεγχο με μηδενική υπόθεση την $H_0: r^* = 0$. Για τον έλεγχο της παραπάνω υπόθεσης συγκρίνεται η στατιστική t με τις κρίσιμες τιμές τα των Dickey- Fuller όπως έχουν τροποποιηθεί από τους Engle- Granger. Οι κρίσιμες αυτές τιμές εξαρτώνται από το μέγεθος του δείγματος, από το επίπεδο σημαντικότητας του ελέγχου, αλλά και από τον αριθμό των μεταβλητών που περιλαμβάνει η παλινδρόμηση συνολοκλήρωσης. Η μηδενική υπόθεση είναι ότι τα κατάλοιπα δεν είναι στάσιμη σειρά, και επομένως ακολουθεί το συμπέρασμα ότι οι μεταβλητές δεν είναι συνολοκληρωμένες. Ο έλεγχος θα είναι ο εξής:

$H_0: r^* = 0$ έναντι της εναλλακτικής $H_1: r^* \neq 0$

Η μηδενική υπόθεση απορρίπτεται αν $|t| \geq t_a$, όπου a το επίπεδο σημαντικότητας του ελέγχου. Διαφορετικά, αν δηλαδή $|t| < t_a$, τότε αποδεχόμαστε τη μηδενική υπόθεση.

Αν η αυτοπαλίνδρομη διαδικασία δεν είναι πρώτης τάξεως, αλλά είναι μεγαλύτερης τάξης, στη σχέση $\Delta\hat{u}_t = r^*\hat{u}_{t-1} + e_t$ θα προστεθούν όροι με χρονικές υστερήσεις ως μεταβλητές. Επίσης, σε αυτήν την περίπτωση θα εφαρμοστεί ο επαυξημένος έλεγχος των Dickey- Fuller και επομένως το r^* θα εκτιμηθεί τώρα από τη σχέση:

$$\Delta \hat{u}_t = r^* \hat{u}_{t-1} + \sum_{i-1}^{p-1} r_i^* \Delta \hat{u}_{t-1} + e_t$$

Έλεγχος CRDW

Έστω πάλι ότι τα κατάλοιπα αποτελούν ένα αυτοπαλίνδρομο υπόδειγμα πρώτης τάξεως. Θα έχουμε τη σχέση $\hat{u}_t = r\hat{u}_{t-1} + e_t$

Ένας απλούστερος τρόπος για να ελεγχθεί η μηδενική υπόθεση $H_0 : r = 1$ είναι να χρησιμοποιηθεί το γνωστό κριτήριο των Durbin- Watson για έλεγχο αυτοσυσχέτισης πρώτης τάξεως στα κατάλοιπα. Όπως γνωρίζουμε, όταν $r \to 1$, τότε το $d \to 0$. Η μηδενική υπόθεση θα απορριφθεί για υψηλές τιμές του d. Συνήθως, για επίπεδο σημαντικότητας $a = 5\%$ και αριθμό παρατηρήσεων 100, η μηδενική υπόθεση απορρίπτεται αν $d \leq 0.386$ η μηδενική υπόθεση γίνεται αποδεκτή και οι μεταβλητές δεν είναι συνολοκληρωμένες.

Το συμπέρασμα που βγαίνει από τα παραπάνω είναι ότι οι μεταβλητές Y και X δεν βρίσκονται πάντοτε σε ισορροπία, οπότε αυτό που παρατηρούμε στην πραγματικότητα είναι μία σχέση ανισορροπίας μεταξύ τους. Μια τέτοια σχέση παριστάνεται με την ακόλουθη ισότητα, με την οποία ασχοληθήκαμε και προηγουμένως:

$$Y_t = b_0 + b_1 X_t + b_2 X_{t-1} + g_1 Y_{t-1} + u_t \quad \textbf{(1)}$$

Η σχέση αυτή, όπως είδαμε, σε κατάσταση ισορροπίας θα πάρει τη μορφή:

$$Y^e = a_0 + a_1 X^e, \quad \text{όπου} \quad a_1 = \frac{b_1 + b_2 + b_3}{1 - g_1 - g_2} \quad \text{και} \quad a_1 = \frac{b_1 + b_2}{1 - g_1}$$

Αν αφαιρέσουμε και από τα δύο μέλη της σχέσης (1) την Y_{t-1}, η σχέση θα πάρει τη μορφή:

$$\Delta Y_t = b_0 + b_1 X_t + b_2 X_{t-1} + (1 - g_1) Y_{t-1} + u_t$$

Προσθαφαιρούμε στο δεξί μέλος αυτής της σχέσης το $b_1 X_{t-1}$, οπότε καταλήγουμε στη σχέση:

$$\Delta Y_t = b_0 + b_1 \Delta X_t + (b_1 + b_2) X_{t-1} + (1 - g_1) Y_{t-1} + u_t$$

Αντικαθιστούμε το b_0 με το $a_0(1-g_1)$ και στη συνέχεια αντικαθιστούμε το $(b_1 + b_2)$ με το $a_1(1-g_1)$ από τις προηγούμενες σχέσεις, οπότε τελικά η σχέση **(1)** γίνεται:

$$\Delta Y_t = b_1 \Delta X_t - (1 - g_1)(Y_{t-1} - a_0 - a_1 X_{t-1}) + u_t$$

Αυτή η σχέση προήλθε από την επαναπαραμετροποίηση της σχέσεως **(1),** οπότε αντιπροσωπεύει τη σχέση ανισορροπίας των δύο μεταβλητών X και Y. Σε αυτή τη μορφή μπορούμε να διακρίνουμε καθαρά ότι οι μεταβολές της Y εξαρτώνται από τις μεταβολές της X και από το λάθος ανισορροπίας της προηγούμενης περιόδου που παριστάνει ο όρος $Y_{t-1} - a_0 - a_1 X_{t-1}$. Επομένως, η τιμή της Y διορθώνεται για το λάθος ανισορροπίας της προηγούμενης περιόδου, για αυτό το λόγο υποδείγματα αυτής της μορφής ονομάζονται Υποδείγματα Διόρθωσης Λαθών (Error Correction Models ή ECM). Η διόρθωση είναι μερική, από τη στιγμή που το μέγεθός της εξαρτάται από το συντελεστή g_1, για τον οποίο υποθέτουμε ότι παίρνει τιμές μεγαλύτερες του μηδενός αλλά μικρότερες της μονάδας, δηλαδή $0 < g_1 < 1$. Στο υπόδειγμα που κατασκευάσαμε, το σφάλμα ανισορροπίας αναφέρεται στην προηγούμενη χρονική περίοδο, την $t-1$. Θα μπορούσε όμως να αναφέρεται και στην περίοδο $t-2$, αλλά και γενικά στην περίοδο $t-s$. Έστω για παράδειγμα, ότι η αρχική σχέση ανάμεσα στις μεταβλητές X και Y ήταν η ακόλουθη:

$$Y_t = b_0 + b_1 X_t + b_2 X_{t-1} + b_3 X_{t-2} + g_1 Y_{t-1} + g_2 Y_{t-2} + u_t$$

Αυτή η σχέση μπορεί να επαναπαραμετροποιηθεί ακριβώς όπως και προηγουμένως και να πάρει την ακόλουθη μορφή υποδείγματος διόρθωσης λαθών:

$$\Delta Y_t = b_0 + b_1 \Delta X_t - (1-g_1) Y_{t-1} + (1-g_1) a_1 X_{t-1} + u_t$$
$$\Delta Y_t = b_1 \Delta X_t - (1-g_1)\hat{u}_t + e_t$$
$$b_0 = a_0(1-g_1)$$
$$Y_t = (1-B)^d (1-B^s)^D X_t$$
$$q(z) = 1 - q_1 z - \ldots - q_p z^q$$
$$f(B)\Phi(B^s) Y_t = q(B)\Theta(B^s) Z_t$$
$$j^*(B) Y_t = q^*(B) Z_t$$
$$Z_t \sim N(0, s^2)$$
$$|z| \leq 1$$
$$\Phi(z) \neq 0$$
$$q + sQ$$
$$q < s$$
$$f_1, \ldots, f_p, \Phi_1, \ldots, \Phi_p, q_1, \ldots, \Theta_1, \ldots, \Theta_q$$

$$\hat{a}_1 = \frac{(1-\hat{g}_1)\hat{a}_1}{1-\hat{g}_1}$$
$$\{Y_t\}$$
$$q^*_{is+j} = q^*_{is} q^*_j$$
$$i = 1, 2, \ldots$$
όπου $j = 1, 2, \ldots, s-1$
$$X_{j+12t} = \Phi_1 X_{j+12(t-1)} + \ldots + \Phi_P X_{j+12(t-P)} + U_{j+12t} + \Theta_1 U_{j+12(t-1)} + \ldots + \Theta_Q U_{j+12(t-Q)}$$
$$j^{-th}$$
$$\{U_{j+12t}, t = \ldots, -1, 0, 1 \ldots\} \sim WN(0, s_u^2)$$
$$\Phi(B^{12}) X_t = \Theta(B^{12}) U_t$$
$$\Theta(z) = 1 - \Theta_1 z - \ldots - \Phi_Q z^Q$$

και $a_1 = \dfrac{b_1 + b_2 + b_3}{1 - g_1 - g_2}$

Το λάθος ισορροπίας στην παραπάνω σχέση που είναι η ποσότητα $Y_{t-2} - a_0 - a_1 X_{t-2}$, αναφέρεται στην περίοδο $t-2$. Εντελώς ανάλογα, μπορεί να υπολογιστεί και το σφάλμα ισορροπίας σε κάθε χρονική περίοδο $t-s$.

Στη συνέχεια, θα διατυπωθεί το θεώρημα αντιπροσώπευσης του Granger (Granger representation theorem). Το θεώρημα λέει ότι *αν δύο μεταβλητές είναι συνολοκληρωμένες, τότε η μεταξύ τους βραχυχρόνια σχέση ανισορροπίας μπορεί πάντοτε να διατυπωθεί ως ένα υπόδειγμα διόρθωσης λαθών*. Αυτό το θεώρημα αποτελεί το αποτέλεσμα της διαπίστωσης ότι η διατύπωση υποδειγμάτων διόρθωσης λαθών βασίζεται ή προϋποθέτει την ύπαρξη κάποιας σχέσης ισορροπίας ανάμεσα στις μεταβλητές. Αυτό σημαίναι ότι οι μεταβλητές θα είναι σίγουρα συνολοκληρωμένες. Αν όμως, δεν υπάρχει μακροχρόνια σχέση ισορροπίας ανάμεσα στις μεταβλητές, τότε δε θα πρέπει η βραχυχρόνια συμπεριφορά των μεταβλητών να παριστάνεται με υποδείγματα διόρθωσης λαθών.

Αφού έχει γίνει ο έλεγχος για το αν οι μεταβλητές είναι συνολοκληρωμένες και έχει βρεθεί το υπόδειγμα διόρθωσης λαθών, στη συνέχεια θα γίνει η εκτίμηση αυτού του υποδείγματος. Η εκτίμηση ενός τέτοιου υποδείγματος μπορεί να επιτευχθεί με δύο τρόπους:

♦ Ο **πρώτος τρόπος** έχει προταθεί από τους Engle- Granger και ακολουθείται μία διαδικασία που περιλαμβάνει δύο στάδια.

Στο πρώτο στάδιο, εκτιμάται η παλινδρόμηση συνολοκλήρωσης που δίνεται από τη

$$\hat{u}_t = Y_t - \hat{a}_0 + \hat{a}_1 X_t$$

$$w_t = a_0 + a_1 w_r + e_t + q_1 e_{t-1}$$

$$w_t = Y_t - Y_{t-1}$$

$$w_i = 0.1513 + 0.5687 w_{t-2} + e_t$$

$$Q_{LB} = T(T+2)\sum_{s=1}^{16} \frac{\hat{r}_s^{2}}{T-s}$$

$$r_i, i = 1, 2, ..., 16$$

$$H_0 = r_1 = r_2 = ... = r_{16} = 0$$

$$Q_{LB} = 15.47 \le X^2 = 26.3$$

$$|t| = 3.85 < t_{0.05} = 1.96$$

$$w_t = 0.1474 + 0.7946 w_{t-1} + 0.1117 e_{t-1}$$

$$\hat{w}_{T+1} = a_0 + a_1 w_r$$

$$\hat{w}_{1995=} 0.1513 + 0.5687^{*} 0.12005 = 0.219$$

$$\hat{Y}_{1994=} Y_{1994} + \hat{w}_{1995} = 16.7899 + 0.2196 = 17.0095$$

σχέση

$$GDP_{1995} = \mathrm{Antilog\acute{a}riqmoV}(\hat{Y}_{1995}) = \mathrm{Antilog\acute{a}riqmoV}(17.0095) = 24385518$$

και η εκτίμηση θα γίνει με την κλασσική μέθοδο των ελαχίστων τετραγώνων.

Στη συνέχεια, θα υπολογιστούν τα εκτιμώμενα κατάλοιπα που δίνονται από τη σχέση:

$$\hat{u}_t = Y_t - \hat{a}_0 + \hat{a}_1 X_t$$

Στο δεύτερο στάδιο γίνεται η αντικατάσταση των αληθινών λαθών ισορροπίας στη σχέση $\Delta Y_t = b_1 \Delta X_t - (1-g_1)(Y_{t-1} - a_0 - a_1 X_{t-1}) + u_t$, με τα εκτιμημένα κατάλοιπα, οπότε το προς εκτίμηση υπόδειγμα λαθών γίνεται:

$$\Delta Y_t = b_1 \Delta X_t - (1-g_1)\hat{u}_t + e_t$$

Σε αυτό το υπόδειγμα όλες οι μεταβλητές είναι στάσιμες αφού οι Y και X είναι συνολοκληρωμένες. Επομένως, οι βραχυχρόνιες παράμετροι b_1 και g_1 μπορούν εύκολα να εκτιμηθούν με τη μέθοδο των ελαχίστων τετραγώνων.

♦ Ένας δεύτερος τρόπος για να εκτιμηθεί το υπόδειγμα διόρθωσης λαθών είναι ο ακόλουθος:

Εκτιμάται απευθείας το υπόδειγμα $\Delta Y_t = b_1 \Delta X_t - (1-g_1)(Y_{t-1} - a_0 - a_1 X_{t-1}) + u_t$, οπότε εκτιμώνται από κοινού οι μακροχρόνιες και οι βραχυχρόνιες παράμετροι. Αυτό μπορεί να γίνει ως εξής: Το υπόδειγμα διόρθωσης λαθών μπορεί να επαναδιατυπωθεί ως: $\Delta Y_t = b_0 + b_1 \Delta X_t - (1-g_1)Y_{t-1} + (1-g_1)a_1 X_{t-1} + u_t$, όπου $b_0 = a_0(1-g_1)$

Στη συνέχεια, σε αυτή τη μορφή του υποδείγματος, εφαρμόζεται η μέθοδος των ελαχίστων τετραγώνων, οπότε υπολογίζονται οι εκτιμήσεις τω μακροχρόνιων παραμέτρων a_0 και a_1, και οι οποίες δίνονται από τις παρακάτω σχέσεις:

$$\hat{a}_0 = \frac{\hat{b}_0}{1-\hat{g}_1} \qquad \text{και} \qquad \hat{a}_1 = \frac{(1-\hat{g}_1)\hat{a}_1}{1-\hat{g}_1}$$

3.2 *Εποχικά υποδείγματα ARIMA (SARIMA)*

Στη συνέχεια, θα μιλήσουμε για τα εποχικά υποδείγματα ARIMA (seasonal ARIMA models ή SARIMA models), τα οποία είναι μια άλλη κατηγορία των υποδειγμάτων ARIMA, και θα υπογραμμίσουμε τα κύρια χαρακτηριστικά τους. Το εποχικό μέρος ενός υποδείγματος ARIMA έχει την ίδια δομή με αυτή του μη- εποχικού υποδείγματος, δηλαδή μπορεί να έχει έναν παράγοντα AR, έναν παράγοντα MA και/ή μια τάξη διαφορών. Στο εποχικό μέρος του υποδείγματος, όλοι αυτοί οι παράγοντες διεξάγουν πολλαπλασιασμούς της χρονικής υστέρησης *s* (δηλαδή τον αριθμό των περιόδων σε μία εποχή). Ένα εποχικό ARIMA υπόδειγμα ορίζεται σαν ένα υπόδειγμα ARIMA(*p.d.q*)x(*P,D,Q*), όπου *P* είναι ο αριθμός των εποχικών αυτοπαλίνδρομων όρων (SAR), D είναι ο αριθμός των εποχικών διαφορών και *Q* είναι ο αριθμός των εποχικών όρων κινητού μέσου (SMA).

Για να εξακριβωθεί ένα εποχικό υπόδειγμα, το πρώτο βήμα είναι να καθορίσουμε αν είναι απαραίτητη ή όχι μια εποχική διαφορά μαζί με μια μη-εποχική διαφορά ή μπορεί και αντί αυτής. Δεν πρέπει ποτέ να χρησιμοποιηθούν περισσότερες από μια εποχικές διαφορές ούτε περισσότερες από δύο διαφορές συνολικά (εποχικές και μη). Εάν το εποχικό υπόδειγμα είναι δυνατό και σταθερό με το χρόνο, (για παράδειγμα ψηλά το καλοκαίρι και χαμηλά το χειμώνα ή και αντίθετα), τότε πιθανόν να πρέπει να χρησιμοποιηθεί μια εποχική διαφορά ανεξάρτητα με το αν χρησιμοποιηθεί μια μη-εποχική διαφορά, από τη στιγμή που αυτή θα προστατέψει το εποχικό πρότυπο από το να πάψει να ισχύει σε προβλέψεις μεγάλης διάρκειας. Εάν η αυτοσυσχέτιση κατά την εποχική περίοδο είναι θετική, θα προσθέσουμε έναν όρο SAR στο υπόδειγμά μας. Αντίθετα, εάν η αυτοσυσχέτιση κατά την εποχική περίοδο είναι αρνητική, τότε θα προστεθεί ένας όρος SMA στο υπόδειγμα. Θα ήταν καλύτερα να μην ανακατεύουμε SAR και SMA όρους στο ίδιο υπόδειγμα και επίσης να αποφεύγουμε να χρησιμοποιούμε περισσότερους από έναν του κάθε είδους. Συνήθως ένας SAR(*1*) και ένας SMA(*1*) όρος είναι επαρκείς. Σπάνια θα συναντήσουμε μια γνήσια SAR(*2*) ή SMA(*2*) διαδικασία, και ακόμα πιο σπάνια θα έχουμε αρκετά δεδομένα έτσι ώστε να μας επιτρέψουν να εκτιμήσουμε δύο ή περισσότερους εποχικούς συντελεστές χωρίς ο αλγόριθμος εκτίμησης να πέσει σε ‘ανατροφοδότηση’. Αν και ένα εποχικό μοντέλο φαίνεται να έχει ελάχιστες παραμέτρους προς εκτίμηση των πεπλεγμένων παραμέτρων ενός ή δυο εποχών που αξίζουν για να τη μηδενίσουν. Πιθανόν, το εποχικό υπόδειγμα που χρησιμοποιείται συχνότερα είναι το υπόδειγμα (0,1,1)x(0,1,1), δηλαδή ένα MA(*1*)x SMA(*1*) υπόδειγμα με μια εποχική και μια μη-εποχική διαφορά. Αυτό είναι ουσιαστικά ένα ‘εποχικό εκθετικό ομαλό’ υπόδειγμα. Τέλος, όταν ένα εποχικό υπόδειγμα εφαρμόζεται σε στάσιμα δεδομένα, είναι ικανό να παρακολουθήσει έναν πολλαπλό εποχικό πρότυπο.

Ας δώσουμε στη συνέχεια έναν μαθηματικό ορισμό για τα υποδείγματα **SARIMA**:

<u>Ορισμός:</u> Αν d και D είναι δύο μηδενικοί ακέραιοι, τότε η διαδικασία $\{X_t\}$ είναι ένα εποχικό ARIMA(*p.d.q*)x(*P,D,Q*),υπόδειγμα με περίοδο s, αν οι διαφοροποιημένες σειρές $Y_t = (1-B)^d (1-B^s)^D X_t$ αποτελούν μια ARMA διαδικασία που ορίζεται από τη σχέση: $f(B)\Phi(B^s)Y_t = q(B)\Theta(B^s)Z_t$, με $Z_t \sim N(0, s^2)$ όπου

$$f(z) = 1 - f_1 z - \ldots - f_p z^p$$

$$\Phi(z) = 1 - \Phi_1 z - \ldots - \Phi_p z^p$$

$$q(z) = 1 - q_1 z - \ldots - q_p z^q$$

$$\Theta(z) = 1 - \Theta_1 z - \ldots - \Theta_p z^q$$

Μπορούμε να παρατηρήσουμε ότι η διαδικασία $\{Z_t\}$ είναι αιτιολογική (causal) αν και μόνο αν $f(z) \neq 0$ και $\Phi(z) \neq 0$ για $|z| \leq 1$. Στις εφαρμογές, ο D σπάνια θα είναι πάνω από τη μονάδα και ο P και Q είναι τυπικά λιγότεροι από το τρία. Επιπλέον, η εξίσωση $f(B)\Phi(B^s)Y_t = q(B)\Theta(B^s)Z_t$ που ικανοποιείται από τη διαφοροποιημένη διαδικασία $\{Y_t\}$ μπορεί να ξαναγραφεί με έναν ισοδύναμο τύπο, ο οποίος είναι:

$$j^*(B)Y_t = q^*(B)Z_t$$

όπου $j^*(.), q^*(.)$ είναι πολυώνυμα βαθμού $p + sP$ και $q + sQ$ αντίστοιχα, των οποίων οι συντελεστές μπορούν να εκφραστούν σε σχέση με τα $f_1, \ldots, f_p, \Phi_1, \ldots, \Phi_p, q_1, \ldots, \Theta_1, \ldots, \Theta_q$. Αν υποθέσουμε κι ότι $p < s$ και $q < s$ οι περιορισμοί στους συντελεστές των $j^*(.)$και $q^*(.)$ μπορούν όλοι να εκφραστούν σαν σχέσεις:

$$f^*_{is+j} = f^*_{is} f^*_j, \qquad i = 1, 2, \ldots \quad \text{και} \quad j = 1, 2, \ldots, s-1$$

και

$$q^*_{is+j} = q^*_{is} q^*_j, \qquad i = 1, 2, \ldots \quad \text{και} \quad j = 1, 2, \ldots, s-1$$

Στη συνέχεια, θα ακολουθήσει ένα παράδειγμα όπου εφαρμόζεται ένα εποχικό μοντέλο.

<u>Παράδειγμα:</u> Ας υποθέσουμε ότι έχουμε r μηνιαία δεδομένα για 12 χρόνια υα οποία ταξινομούνται ως εξής:

Χρόνος/Μήνας	**1**	**2**	**…**	**12**
1	X_1	X_2	$\cdots$	X_{12}
2	X_{13}	X_{14}	$\cdots$	X_{24}
3	X_{25}	X_{26}	$\cdots$	X_{36}
.	.	.		
.	.	.		
r	$X_{1+12(r-1)}$	$X_{2+12(r-1)}$	$\cdots$	$X_{12+12(r-1)}$

Κάθε στήλη αυτού του πίνακα μπορεί να θεωρηθεί σαν μια πραγματοποίηση μιας χρονολογικής σειράς. Ας υποθέσουμε ότι κάθε μία από τις δώδεκα αυτές χρονολογικές σειρές έχει γενικευτεί από το ίδιο το ARMA(P,Q) μοντέλο, ή ακόμα πιο συγκεκριμένα ότι η σειρά που αντιστοιχεί στον j^{-th} μήνα, δηλαδή η $X_{j+12t,} t = 0, \ldots, r-1$, ικανοποιεί μια διαφορική εξίσωση της μορφής:

$$X_{j+12t} = \Phi_1 X_{j+12(t-1)} + ... + \Phi_P X_{j+12(t-P)} + U_{j+12t} + \Theta_1 U_{j+12(t-1)} + ... + \Theta_Q U_{j+12(t-Q)} \quad \textbf{(1)}$$

όπου

$$\left\{U_{j+12t}, t = ..., -1, 0, 1...\right\} \sim WN(0, s_u^2)$$

Επομένως, από τη στιγμή που το ίδιο ARMA(P,Q) μοντέλο επιβεβαιώνεται ότι εφαρμόζεται σε κάθε μήνα, η σχέση **(1)** εφαρμόζεται για κάθε j = 1,…,12.Μπορούμε επίσης, να γράψουμε τη σχέση **(1)** με τη μορφή ενός πιο συμπυκνωμένου τύπου ως εξής:

$$\Phi(B^{12})X_t = \Theta(B^{12})U_t \quad \textbf{(2)}$$

όπου

$$\Phi(z) = 1 - \Phi_1 z - ... - \Phi_P z^P$$

$$\Theta(z) = 1 - \Theta_1 z - ... - \Phi_Q z^Q$$

και

$$\left\{U_{j+12t}, t = ..., -1, 0, 1...\right\} \sim WN(0, s_u^2), \qquad \text{για κάθε j.}$$

Το μοντέλο που απεικονίζεται από τη σχέση (2) αναφέρεται ως το ‘between-year’ μοντέλο.

<u>α. Εκτίμηση των παραμέτρων.</u>

Περνάμε τώρα στο δεύτερο στάδιο της μεθοδολογίας Box-Jenkins που είναι η εκτίμηση των παραμέτρων της αυτοπαλίνδρομης διαδικασίας, καθώς και της διαδικασίας κινητού μέσου. Η σειρά που εξετάζουμε εμείς είναι μόνο αυτοπαλίνδρομης τάξης 1, οπότε οι παράμετροί της μπορούν να εκτιμηθούν με τη μέθοδο των ελαχίστων τετραγώνων. Έστω ότι το υπόδειγμά μας έχει τη μορφή $w_t = a_0 + a_1 w_{t-1} + e_t$, όπου Y=Log(GDP) και $w_t = Y_t - Y_{t-1}$.

Με τη βοήθεια του στατιστικού πακέτου SPSS παίρνουμε τις εκτιμήσεις των αυτοπαλίνδρομων παραμέτρων a_0 και a_1, οι οποίες είναι για το υπόδειγμα ARMA(1,0) οι παρακάτω:

a_0=0.1513

a_1=0.5687

Επομένως, το υπόδειγμα στο οποίο έχουμε καταλήξει προς το παρόν είναι το πιο κατάλληλο για την όσο το δυνατόν καλύτερη πρόβλεψη, έχει τη μορφή:

$$w_i = 0.1513 + 0.5687 w_{t-1} + e_t \qquad \textbf{(1)}$$

<u>β. Διαγνωστικός έλεγχος.</u>

Το στάδιο αυτό περιλαμβάνει δύο επιμέρους στάδια τα οποία είναι ο έλεγχος των καταλοίπων και ο έλεγχος της τάξης του υποδείγματος.

- **Έλεγχος των καταλοίπων**

Για να είναι το υπόδειγμα της σχέσης (1) αποδεκτό για τα δεδομένα μας, για να θεωρηθεί δηλαδή ότι περιγράφει τη διαδικασία που παρήγαγε τα δεδομένα, θα πρέπει τα κατάλοιπα να είναι λευκός θόρυβος, με λίγα λόγια να μην αυτοσυσχετίζονται. Αυτός ο έλεγχος για τα κατάλοιπα γίνεται με τη στατιστική Q των Box- Pierce ή με την τροποποιημένη της μορφή, που είναι η στατιστική των Ljung- Box. Στη δική μας περίπτωση θα χρησιμοποιήσουμε τη δεύτερη μιας και θεωρείται πιο κατάλληλη για μικρά δείγματα. Η στατιστική των Ljung- Box, όπως έχουμε αναφέρει, ορίζεται ως:

$$Q_{LB} = T(T+2)\sum_{s=1}^{16} \frac{\hat{r}_s^{\,2}}{T-s},$$

όπου $r_i, i = 1, 2, ..., 16$, είναι οι συντελεστές συσχέτισης των καταλοίπων και Τα ο αριθμός των παρατηρήσεων. Η στατιστική αυτή ελέγχει από κοινού τη σημαντικότητα των 16 συντελεστών αυτοσυσχέτισης r_i. Η μηδενική υπόθεση, όπως έχουμε πει, θα είναι η:

$$H_0 = r_1 = r_2 = ... = r_{16} = 0$$

Η στατιστική $Q_{LB} \sim X^2_{m-p-q}$, δηλαδή στη δική μας περίπτωση που m = 16, p = 1, q = 0, $Q_{LB} \sim X^2_{15}$. Η αποδοχή ή η απόρριψη της μηδενικής υπόθεσης γίνεται σύμφωνα με το παρακάτω:

- Απόρριψη H_0, αν $Q_{LB} > X^2_a$
- Αποδοχή H_0, αν $Q_{LB} \leq X^2_a$

Η τιμή της Q_{LB} είναι: Q_{LB} = 23.12. Όμως, η τιμή της X^2 για δεδομένο επίπεδο σημαντικότητας $a = 5\%$ και m = 16 βαθμούς ελευθερίας είναι: X^2 = 26.3. Οπότε, $Q_{LB} \leq X^2_a$ και αποδεχόμαστε τη μηδενική υπόθεση ότι τα κατάλοιπα δεν αυτοσυσχετίζονται.

- **Έλεγχος της τάξης του υποδείγματος.**

Ο έλεγχος αυτός γίνεται με τη διαδικασία της υπερπροσαρμογής (overfitting). Δηλαδή, θα συγκρίνουμε το εκτιμημένο υπόδειγμά μας, που δίνεται από τη σχέση (1) με υποδείγματα μεγαλύτερης τάξεως. Αν το εκτιμημένο υπόδειγμα είναι τελικά το καταλληλότερο για τα δεδομένα μας, θα πρέπει οι επιπλέον συντελεστές στα μεγαλύτερα υποδείγματα να μην είναι στατιστικά διαφορετικοί από το μηδέν. Το υπόδειγμά μας είναι ένα ARMA(1,0), οπότε θα το συγκρίνουμε με τα υποδείγματα ARMA(2,0) και ARMA(1,1)

- Έστω ένα υπόδειγμα ARMA(2,0) της μορφής:

$$w_t = a_0 + a_1 w_{t-1} + a_2 w_{t-2} + e_t$$

Οι εκτιμώμενοι παράμετροι a_0, a_1, a_2 θα είναι ίσοι με:

a_0=0.1490, a_1=0.4608, a_2=0.1899.

Οπότε το εκτιμώμενο υπόδειγμα είναι το : w_t = 0.1490 + 0.4608 w_{t-1} + 0.1899 w_{t-2} + e_t. Η στατιστική Q_{LB} είναι: $Q_{LB} = 15.69 \le X^2 = 26.3$, οπότε τα κατάλοιπα δεν αυτοσυσχετίζονται. Όμως, παρατηρούμε ότι η προσθήκη και δεύτερου αυτοπαλίνδρομου όρου δεν είναι αρκετά σημαντική έτσι ώστε να τον κρατήσουμε αφού $t = 1.35 < t_{0.05} = 196$, για επίπεδο σημαντικότητας $a = 5\%$.

► Έστω τώρα ένα υπόδειγμα ARMA(1,1) της μορφής:

$$w_t = a_0 + a_1 w_{t-1} + e_t + q_1 e_{t-1}$$

Οι εκτιμώμενοι παράμετροι a_0, a_1, q_1 θα είναι ίσοι με:

a_0 = 0.1474, a_1 = 0.7946, q_1 = 0.1117

οπότε το εκτιμώμενο υπόδειγμα είναι το: w_t = 0.1474 + 0.7946 w_{t-1} + e_t +0.1117 e_{t-1}.

Η στατιστική Q_{LB} είναι $Q_{LB} = 15.47 \le X^2 = 26.3$, οπότε τα κατάλοιπα και σε αυτό το υπόδειγμα δεν αυτοσυσχετίζονται. Όμως, η προσθήκη ενός όρου κινητού μέσου είναι σημαντική σε αντίθεση με την προσθήκη ενός ακόμα αυτοπαλίνδρομου, διότι

$$X_f = \hat{b}_0 + \hat{b}_1 \hat{Y}_f$$

$$E(X_f)$$

$$f = X_f - \hat{X}_f$$

$$\{x_1, x_2, ..., x_t\}$$

$$\{X_t\}$$

$$e_t(k) = x_{t+k} - x_t(k)$$

$$X_t(k) = E(X_{t+k}, X_{t,} X_{t-1}, ...)$$

$$X_{t+k}$$

$$E(X_{t+k} - X_t(k)^2) = X_{t+k}$$

$$Var(e_t(k)) = Var(X_{t+k} - X_t(k))$$

Αυτό σημαίνει ότι στη συνέχεια θα πρέπει να αξιολογήσουμε την προβλεπτική ικανότητα και των δύο υποδειγμάτων, δηλαδή του **ARIMA(1,1,0)** και του ARIMA(1,1,1), χρησιμοποιώντας τα κριτήρια που έχουν αναφερθεί παραπάνω.

γ. Προβλέψεις.

Τα εκτιμημένα υποδείγματα των οποίων την προβλεπτική ικανότητα θα αξιολογήσουμε, είναι τα εξής:

ARMA(1,0): $w_i = 0.1513 + 0.5687w_{t-1} + e_t$

ARMA(1,1): $w_t = 0.1474 + 0.7946w_{t-1} + 0.1117e_{t-1}$

ΚΕΦΑΛΑΙΟ 4

Προβλέψεις

Ο βασικός σκοπός της μελέτης των μοντέλων για χρονολογικές σειρές, δηλαδή της εξιδείκευσης και εκτίμησης ενός μοντέλου όπως του αυτοπαλίνδρομου (AR), του κινητού μέσου (MA), του μεικτού(ARMA), του αυτοπαλίνδρομου ολοκληρωμένου κινητού μέσου(ARIMA) και του εποχικού πολλαπλασιαστικού (SARIMA) είναι η διενέργεια προβλέψεων (prediction, forecasting) και η επίδραση αυτών κατ΄επέκταση στη διοίκηση των επιχειρήσεων. Με βάση, δηλαδή, το εκτιμώμενο μοντέλο και τις υπάρχουσες πληροφορίες μέχρι τη χρονική περίοδο t, να γίνει πρόβλεψη της τιμής της χρονολογικής σειράς στην περίοδο t + 1, t + 2, κ.ο.κ.

Όπως και στην ανάλυση του απλού γραμμικού μοντέλου, η καλύτερη πρόβλεψη μιας μελλοντικής τιμής X_f της μεταβλητής Y, είναι:

$$X_f = \hat{b}_0 + \hat{b}_1 \hat{Y}_f \tag{4.1}$$

η οποία στην ουσία είναι εκτίμηση της $E(X_f)$.

$$f = X_f - \hat{X}_f \tag{4.2}$$

Η διαφορά εξάλλου παριστάνει το σφάλμα πρόβλεψης. Έτσι, και στην περίπτωση των μοντέλων των χρονολογικών σειρών προκύπτουν ανάλογα αποτελέσματα.

Όπως γίνεται αντιληπτό, η πρόβλεψη των μελλοντικών τιμών μιας παρατηρούμενης χρονικής σειράς είναι σημαντικό πρόβλημα για πολλές εφαρμογές. Ο δείκτης και ο όγκος χρηματιστηρίου αποτελεί χαρακτηριστικό παράδειγμα. Επίσης, ο αριθμός των ηλιακών κηλίδων είναι σημαντικό πρόβλημα, καθώς επηρεάζει το κλίμα της γης, γι' αυτό και έχει μεγάλη σημασία η πρόβλεψη του αριθμού των ηλιακών κηλίδων για τα επόμενα έτη.

Για να πραγματοποιηθεί η πρόβλεψη χρησιμοποιούνται οι παρατηρήσεις μέχρι την παρούσα χρονική στιγμή. Θεωρώντας την παρατηρούμενη χρονολογική σειρά $\{x_1, x_2, ..., x_t\}$ από μια στοχαστική διαδικασία $\{X_t\}$, το πρόβλημα που μελετάται, είναι η πρόβλεψη της χρονολογικής σειράς για k χρονικά βήματα μπροστά από την χρονική στιγμή *t*, που συμβολίζεται με $x_t(k)$, ενώ η πραγματική αλλά άγνωστη τιμή στη χρονική στιγμή *t* + *k* είναι x_{t+k}. Το σφάλμα της πρόβλεψης (prediction error) είναι:

$$e_t(k) = x_{t+k} - x_t(k) \tag{4.3}$$

Με αναφορά στη στοχαστική διαδικασία $\{X_t\}$, η πρόβλεψη $X_t(k)$ είναι η εκτίμηση του στοιχείου X_{t+k} της $\{X_t\}$ με βάση τα προηγούμενα στοιχεία της $\{X_t\}$. Δηλαδή, η βέλτιστη πρόβλεψη είναι:

$$X_t(k) = E(X_{t+k}, X_{t,} X_{t-1}, ...) \tag{4.4}$$

Επιθυμητές ιδιότητες καλής εκτιμήσεως, δηλαδή καλή πρόβλεψη, είναι η *αμεροληψία* (unbiasedness) και η *αποτελεσματικότητα* (efficiency). Γνωστά από τη στατιστική, η αμεροληψία είναι:

$$E(X_t(k)) = X_{t+k} \tag{4.5}$$

Η αποτελεσματικότητα είναι:

$$Var(e_t(k)) = Var(X_{t+k} - X_t(k)) \tag{4.6}$$

δηλαδή η μικρή διασπορά λάθους πρόβλεψης.

Συνδυάζοντας τις δύο παραπάνω, κ καλύτερη είναι εκείνη που ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα πρόβλεψης:

$$E(X_{t+k} - X_t(k)^2) \tag{4.7}$$

για κάθε βήμα πρόβλεψης *k*.

Προτού γίνει η ανάλυση και η μελέτη διάφορων μοντέλων πρόβλεψης, θα γίνει μια προσπάθεια να διατυπωθούν διάφορες τεχνικές αξιολόγησης των μοντέλων αυτών. Να συγκριθούν δηλαδή, οι προβλεπόμενες με τις πραγματικές τιμές για την περίοδο που υπάρχουν στοιχεία. για να αξιολογηθεί η απόδοση ενός μοντέλου πρόβλεψης σε μια χρονολογική σειρά για *k* χρονικά βήματα μπροστά πρέπει να υπολογιστεί κάποιο μέτρο που να συγκεντρώνει τα σφάλματα πρόβλεψης για ένα ικανοποιητικό αριθμό χρονικών στιγμών. Γίνονται λοιπόν, προβλέψεις για *k* χρονικά βήματα μπροστά σε έναν αριθμό γνωστών παρατηρήσεων για χρόνους *t* + 1, *t* + 2,…*t* + *i* δηλαδή υπολογίζονται τα $x_t(k), x_{t+1}(k), ..., x_{t+i}(k)$.Έχοντας τις αντίστοιχες πραγματικές τιμές $x_t, x_{t+1}, ..., x_{t+i}$, υπολογίζονται τα σφάλματα πρόβλεψης για *k* χρονικά βήματα μπροστά $e_t(k), e_{t+1}(k), ..., e_{t+i-k}(k)$. Υπάρχουν διάφορα στατιστικά μέτρα που συγκεντρώνουν τα σφάλματα πρόβλεψης. Ένα από αυτά είναι η εκτίμηση του μέσου τετραγωνικού σφάλματος (mean square error *mse*).

$$mse(k) = \frac{1}{i-k+1} \sum_{j=t}^{t+i-k} e_j(k)^2 = \frac{1}{i-k+1} \sum_{j=t}^{t+i-k} (x_{j+k}(k))^2 \tag{4.8}$$

Ένα άλλο στατιστικό μέτρο που συχνά χρησιμοποιείται είναι η ρίζα του μέσου τετραγωνικού σφάλματος (root mean square error *rmse*):

$$rmse(k) = \sqrt{\frac{1}{i-k+1} \sum_{j=t}^{t+i-k} e_j(k)^2} = \sqrt{\frac{1}{i-k+1} \sum_{j=t}^{t+i-k} (x_{j+k}(k))^2} \tag{4.9}$$

Ένα επίσης χρήσιμο μέτρο του σφάλματος της πρόβλεψης όταν χρειάζεται να συγκριθούν μοντέλα σε διαφορετικές χρονολογικές σειρές είναι η κανονικοποίηση της ρίζας του μέσου τετραγωνικού σφάλματος (normalized root mean square error *nrmse*) διαιρώντας τη ρίζα του μέσου τετραγωνικού σφάλματος (*nrmse*) με την δειγματική απόκλιση των παρατηρήσεων της χρονολογικής σειράς(ή πιο σωστά των παρατηρήσεων που χρησιμοποιούνται στο σχηματισμό σφαλμάτων):

$$nrmse(k) = \frac{\sqrt{\frac{1}{i-k+1} \sum_{j=t}^{t+i-k} (x_{j+k} - x_j(k))^2}}{\sqrt{\frac{1}{i-k+1} \sum_{j=t}^{t+i-k} (x_{j+k} - \bar{x})^2}} \tag{4.10}$$

όπου $\bar{x}$ είναι η δειγματική μέση τιμή των $x_{t+k}, x_{t+1+k}, ..., x_{t+i}$. Τιμές της κανονικοποιημένης ρίζας του μέσου τετραγωνικού σφάλματος (nrmse) κοντά στο μηδέν δηλώνουν πολύ καλή πρόβλεψη, ενώ τιμές κοντά και γύρω στη μονάδα δηλώνουν ότι η πρόβλεψη είναι τόσο καλή, όσο εάν γινόταν πρόβλεψη με τη μέση τιμή.

Ορισμένα ακόμη κριτήρια- μέτρα αξιολόγησης της προβλεπτικής ικανότητας ενός μοντέλου είναι το μέσο απόλυτο ποσοστιαίο σφάλμα(mean absolute percentage error *mape*)

$$mape = \frac{1}{i-k+1} \sum_{j=t}^{t+i-k} \left| \frac{x_{j+k} - x_j(k)}{x_{j+k}} \right| \tag{4.11}$$

και ο συντελεστής ανισότητας του Theil (Theil Inequality Coefficient) που ορίζεται ως εξής:

$$U = \frac{\sqrt{\frac{1}{i-k+1} \sum_{j=t}^{t+i-k} (x_{j+k} - x_j(k))^2}}{\sqrt{\frac{1}{i-k+1} \sum_{j=t}^{t+i-k} (x_{j+k} - x_{j+k}^2}} \tag{4.12}$$

Ο συντελεστής *U* είναι ανεξάρτητος από τις μονάδες μέτρησης και γι’ αυτό είναι περισσότερο κατάλληλος για σύγκριση της προβλεπτικής ικανότητας δίφορων μοντέλων, σε αντίθεση με τα υπόλοιπα κριτήρια που εξαρτώνται από τις μονάδες μέτρησης. Εάν οι προβλεπόμενες τιμές συμπίπτουν απολύτως με τις πραγματικές, η τιμή του *U* είναι μηδέν. Όταν ο συντελεστής ανισότητας του Theil είναι μεγαλύτερος της μονάδας ($U > 1$), οι προβλέψεις είναι πολύ μακριά από την πραγματικότητα, μη αποδοτικές και άρα πολύ κακές. Εάν ο συντελεστής ανισότητας του Theil ισούται με την μονάδα ($U = 1$), τότε όλες οι προβλέψεις είναι μηδέν. Η περίπτωση αυτή έχει περισσότερο νόημα όταν για τον υπολογισμό του συντελεστή ανισότητας του Theil χρησιμοποιούνται όχι οι αρχικές τιμές αλλά οι μεταβολές, δηλαδή στη θέση των $x_j(k)$ και x_{j+k} χρησιμοποιούνται οι μεταβολές από την προηγούμενη περίοδο. σε αυτή την περίπτωση $U = 1$ σημαίνει ότι οι προβλεπόμενες μεταβολές είναι μηδέν, δηλαδή συνέχιση της υπάρχουσας κατάστασης.

Ο συντελεστής ανισότητας του Theil *U* μπορεί να διασπαστεί σε τρείς συνιστώσες κάθε μια από τις οποίες εκφράζει μια πηγή ή αιτία της ανακρίβειας των προβλέψεων, με τον ίδιο τρόπο που μπορεί να γίνει και στην ανάλυση των γραμμικών-παλίνδρομων μοντέλων. Η διάσπαση αυτή μπορεί να γίνει ως εξής:

Υψώνοντας στο τετράγωνο και τα δύο μέρη της σχέσης (4.12), προκύπτει:

$$U^2 = \frac{\frac{1}{i-k+1}\sum_{j=t}^{t+i-k}(x_{j+k} - x_j(k))^2}{\frac{1}{i-k+1}\sum_{j=t}^{t+i-k} x_{j+k}^2} \qquad (4.13)$$

Ο αριθμητής στην παραπάνω σχέση είναι ο μέσος του τετραγώνου του σφάλματος και μπορεί να γραφεί, προσθέτοντας και αφαιρώντας τους μέσους των $x_j(k)$ και x_{j+k} ως :

$$mse = \frac{1}{i-k+1}\sum_{j=t}^{t+i-k}\left[(x_j(k) - \overline{x_j}(k)) - (x_{j+k} - \overline{x}_{j+k}) + (\overline{x}_{j+k} - \overline{x}_{j+k})\right]^2 \qquad (4.14)$$

ή

$$mse = (\overline{x}(k) - \overline{x}_k)^2 + (s^f - s^a)^2 + 2(1-r)s^f s^a \qquad (4.15)$$

όπου $\overline{x}(k)$ είναι ο μέσος των προβλεπόμενων τιμών, $\overline{x}_k$ είναι ο μέσος των πραγματικών τιμών, s^f είναι η τυπική απόκλιση των προβλεπόμενων τιμών, s^a είναι η τυπική απόκλιση των πραγματικών τιμών και ρ είναι ο συντελεστής συσχετίσεως των πραγματικών και των προβλεπόμενων τιμών.

Ο πρώτος όρος στη σχέση (4.15), δηλαδή το τετράγωνο της διαφοράς των μέσων είναι ένα μέτρο της μεροληψίας. Ο δεύτερος όρος, δηλαδή το τετράγωνο της διαφοράς των διακυμάνσεων, είναι ένα μέτρο της άνισης μεταβλητότητας των πραγματικών και προβλεπόμενων τιμών x_k και $x(k)$. Ο τρίτος όρος περιέχει τον συντελεστή αυτοσυσχέτισης και μπορεί να θεωρηθεί ως ένα μέτρο της ατελούς μεταβλητότητας (covariation) των πραγματικών και προβλεπόμενων τιμών x_k και $x(k)$.

Η σχέση (4.15) μπορεί να γραφεί και ως εξής:

$$1 = \frac{(\overline{x}(k) - \overline{x}_k)^2}{mse} + \frac{(s^f - s^a)^2}{mse} + \frac{2(1-r)s^f s^a}{mse} \qquad (4.16)$$

ενώ η σχέση (4.13) γίνεται:

$$U = \frac{(\overline{x}(k) - \overline{x}_k)^2}{A} + \frac{(s^f - s^a)^2}{A} + \frac{2(1-r)s^f s^a}{A} \tag{4.17}$$

όπου

$$A = \frac{1}{i-k+1} \sum_{j=t}^{t+i-k} x_{j+k}^2 \tag{4.18}$$

Ο τρίτος όρος στις παραπάνω σχέσεις παριστάνει το μη συστηματικό (τυχαίο) παράγοντα που δεν μπορεί να αποφευχθεί, ενώ οι δύο πρώτοι όροι παριστάνουν συστηματικά σφάλματα που πρέπει να αποφεύγονται. Ιδεατά. θα ήταν επιθυμητό η μόνη πηγή σφάλματος να ήταν το μη συστηματικό μέρος, δηλαδή οι δύο πρώτοι όροι να ήταν μηδέν.

Για ευκολία, θα χρησιμοποιηθούν καταχρηστικά, συμβολισμοί που αναφέρονται στο δείγμα, όπως x_t για να συμβολιστούν επίσης οι μεταβλητές, για παράδειγμα X_t, όταν χρειάζεται να δοθούν γενικές σχέσεις για τις προβλέψεις και τα σφάλματα πρόβλεψης.

4.1 Απλές Τεχνικές Πρόβλεψης

4.1.1 Αιτιοκρατική Τάση (Deterministic Trend).

Ξεκινώντας με την πιο απλή περίπτωση, γίνεται η υπόθεση πως η πληροφορία στην χρονολογική σειρά δίνεται μόνο από χρονικές τάσεις (trends), που είτε είναι γνωστές, είτε χρειάζεται να εκτιμηθούν, δηλαδή,

$$x_t = m_t + x_t \tag{4.19}$$

όπου m_t είναι μια αιτιοκρατική τάση συναρτήσει του χρόνου t (τάση) και x_t είναι ο λευκός θόρυβος.

Η πρόβλεψη γίνεται με την επέκταση (extrapolation) του αιτιοκρατικού όρου σε μελλοντικούς χρόνους, δηλαδή η πρόβλεψη x_{t+k} είναι:

$$x_t(k) = E(m_{t+k} + z_{t+k} x_t, x_{t-1}, ..., x_1) = m_{t+k} \qquad (4.20)$$

Το σφάλμα πρόβλεψης είναι:

$$e_t(k) = z_t(k) \qquad (4.21)$$

άρα $e_t(k)$ είναι λευκός θόρυβος με διασπορά s_z^2.

- **Επέκταση καθολικών τάσεων. (Extrapolation of Global Trends)**

Μια εύκολη προσαρμογή της συνάρτησης αιτιοκρατικής τάσης καθολικά (σε όλη τη χρονική σειρά) μπορεί να γίνει με πολυώνυμο κάποιας τάξης *m*, δηλαδή:

$$p_m(t) = c_0 + c_1 t + ... + c_m t^m \qquad (4.22)$$

Πολλών ειδών καμπύλες μπορεί να προσαρμοστούν καλά στα δεδομένα, όπως οι πολυωνυμικές καμπύλες όταν το *m* είναι μεγάλο, αλλά δίνουν πολύ διαφορετικές προβλέψεις όταν επεκτείνονται σε μελλοντικά χρονικά βήματα. Στην πράξη, τέτοιες προβλέψεις είναι γενικά φτωχές. Ειδικά τα πολυώνυμα υψηλής τάξης, ξεφεύγουν γρήγορα προς το συν ή το πλην άπειρο (+∞, ή -∞) όταν επεκτείνονται έξω από το διάστημα παρατήρησης για το οποίο έγινε η εκτίμηση των παραμέτρων τους.

- **Επέκταση Τοπικών Τάσεων (Extrapolation of Local Trends)**

Ένας τρόπος για να βελτιωθεί η απόδοση της επέκτασης της τάσης είναι να γίνει η προσαρμογή του μοντέλου, όπως αυτό της σχέσης (4.19), χρησιμοποιώντας μόνο τις σχετικά πρόσφατες παρατηρήσεις. με αυτόν τον τρόπο αποφεύγεται η επίδραση των παλαιών παρατηρήσεων στην εκτίμηση του μοντέλου πρόβλεψης. Αυτό βελτιώνει τα μοντέλα πρόβλεψης που δίνονται ως συνάρτηση του χρόνου.

Για χρονολογικές σειρές με αιτιοκρατικό εποχικό όρο:

$$x_t = s_t + z_t \qquad (4.23)$$

ή με αιτιοκρατικό εποχικό όρο και αιτιοκρατική τάση:

$$x_t = m_t + s_t + z_t \tag{4.24}$$

η πρόβλεψη γίνεται κατά τον ίδιο τρόπο, δηλαδή επεκτείνοντας σε μελλοντικούς χρόνους τους αιτιοκρατικούς όρους που εκτιμώνται με κάποια συνάρτηση του χρόνου *t*.

4.1.2 Εκθετική Ομαλοποίηση.

Ένας άλλος τρόπος πρόβλεψης είναι να εκτιμηθεί το x_{t+k} από το σταθμισμένο άθροισμα των προηγούμενων παρατηρήσεων της χρονολογικής σειράς

$$x_{t+k} = c_0 x_t + c_1 x_{t-1} + \ldots + c_{t-1} x_1 = \sum_{f=0}^{t-1} c_f x_{t-f} \tag{4.25}$$

όπου οι συντελεστές c_f είναι τα βάρη με

$$\sum_{f=0}^{t-1} c_f = 1 \tag{4.26}$$

Είναι φυσικό να δοθεί περισσότερο βάρος στις πρόσφατες παρατηρήσεις και η επιλογή των βαρών να φθίνει πηγαίνοντας προς τα πίσω στον χρόνο $c_0 \geq c_1 \geq \ldots \geq c_{t-1}$. Μια τέτοια επιλογή των βαρών είναι

$$c_f = a(1-a)^f,\ f = 0,1,\ldots,t, 0 < a < 1. \tag{4.27}$$

όπου δε χρειάζεται να ορισθεί το κάθε c_f ξεχωριστά παρά μόνο το *α*. Τα βάρη αυτά φθίνουν εκθετικά και η επιλογή του *α* ορίζει πόσο γρήγορα φθίνουν. Εάν $a \cong 1$ ουσιαστικά μόνο οι πολύ πρόσφατες παρατηρήσεις χρησιμοποιούνται για την πρόβλεψη. Για να ενημερώνεται η πρόβλεψη *k* χρονικών βημάτων κάθε φορά που μια νέα παρατήρηση είναι διαθέσιμη μπορεί να χρησιμοποιηθεί η ανάδρομη σχέση:

$$x_{t+1}(k) = a x_t(k) + (1-a) x_t(k) \tag{4.28}$$

4.2 Πρόβλεψη Στάσιμων χρονολογικών Σειρών με Γραμμικά Μοντέλα.

Πρώτα, θα θεωρηθεί ότι η χρονολογική σειρά για την οποία πρέπει να γίνουν οι προβλέψεις είναι στάσιμη ή έχει μετασχηματιστεί σε στάσιμη με κάποια από τις μεθόδους που αναλύθηκαν σε προηγούμενα κεφάλαια. Τα γραμμικά μοντέλα στάσιμων χρονολογικών σειρών που μελετήθηκαν είναι τα αυτοπαλίνδρομα AR, του κινητού μέσου MA και τα μεικτά μοντέλα ARMA. Αυτά τα μοντέλα θα χρησιμοποιηθούν για να γίνουν προβλέψεις. Παρακάτω θα θεωρηθεί επίσης, πως η χρονολογική σειρά έχει μέση τιμή μηδέν ώστε να αποφευχθεί η ύπαρξη σταθερού όρου στα μοντέλα πρόβλεψης. Πρακτικά αυτό γίνεται αφαιρώντας τη δειγματική μέση τιμή των παρατηρήσεων της χρονολογικής σειράς από κάθε παρατήρηση. Για να σχηματιστεί η πραγματική πρόβλεψη που αφορά την παρατηρούμενη μεταβλητή, προστίθεται η δειγματική μέση τιμή στην προβλεπόμενη τιμή από το μοντέλο.

4.2.1 *Πρόβλεψη με Αυτοπαλίνδρομα Μοντέλα.*

<u>AR(1) Μοντέλο</u>

Έστω το πιο απλό γραμμικό αυτοπαλίνδρομο μοντέλο, το αυτοπαλίνδρομο μοντέλο πρώτης τάξης AR(1),

$$x_t = j\, x_{t-1} + z_t \tag{4.29}$$

Για την πρόβλεψη της επόμενης χρονικής στιγμής ***t+1***, όταν είναι γνωστή η χρονολογική σειρά ως την χρονική στιγμή t, από την υπόθεση του μοντέλου προκύπτει:

$$x_{t+1} = j\, x_t + z_t \tag{4.30}$$

Έτσι, εάν η παράμετρος φ είναι γνωστή, τότε με βάση τις πληροφορίες μέχρι την περίοδο ***t***, μια πρόβλεψη για την περίοδο ***t + 1*** είναι η υπό συνθήκη προσδοκώμενη τιμή της X_{t+1}. Δηλαδή, εάν X_{t+1} παριστάνει την πρόβλεψη, τότε:

$$x_t(1) = E(X_{t+1})$$

(πληροφορία μέχρι την περίοδο t) (4.31)

ή

$$x_t(1) = E(X_{t+1} \setminus X_{t+1}, X_{t-1}, ...z_t, z_{t-1}) \quad (4.32)$$

Για λόγους απλότητας, το δεξιό μέρος της σχέσης (4.32), γράφεται $E_t(X_{t+1})$, οπότε:

$$x_t(1) = E_t(X_{t+1}) \quad (4.33)$$

Γενικά,

$$x_t(k) = E_t(X_{t+1}) \quad (4.34)$$

είναι μία πρόβλεψη *k* περιόδους μπροστά με βάση τις πληροφορίες μέχρι την χρονική περίοδο *t*.

Επειδή η παραπάνω πρόβλεψη ελαχιστοποιεί τον μέσο του τετραγώνου του σφάλματος (Mean Square Error *mse*), δηλαδή ελαχιστοποιεί την σχέση:

$$E_t(X_t(k) - X_{t+k})^2 \quad (4.35)$$

θεωρείται άριστη (optimal forecast).

Από την σχέση (4.30), προκύπτει ότι:

$$E_t(X_{t+k}) = j\, E_t(X_t) + E_t(z_{t+1}) \quad (4.36)$$

ή

$$E_t(X_{t+k}) = E_t(j\, X_t + z_{t+1}) \quad (4.37)$$

Εάν ληφθεί υπόψη ότι στην περίοδο *t* κάθε όρος με δείκτη μικρότερο του *t* είναι γνωστός, συνεπάγεται ότι για χρονικές περιόδους μικρότερης της *t:*

$$E_t(X_t) = X_t kai E_t(z_t) = z_t \quad (4.38)$$

Για χρονικές περιόδους μεγαλύτερες της *t* όμως ισχύει:

$$E_t(X_t) = E_t(X_t) = 0 \qquad (4.39)$$

Με βάση τα παραπάνω αποτελέσματα, η σχέση (4.37) γίνεται:

$$E_t(X_{t+1}) = j\, X_t \qquad (4.40)$$

και η πρόβλεψη την περίοδο *t + 1* είναι:

$$X_t(1) = j\, X_t \qquad (4.41)$$

Για δύο χρονικά βήματα εμπρός προκύπτει:

$$X_{t+2} = j\, x_{t+1} + z_{t+1} \qquad (4.42)$$

Αντικαθιστώντας το x_{t+1} με την πρόβλεψη $x_t(1)$ και χρησιμοποιώντας την (4.41) προκύπτει:

$$x_t(2) = j\, x_t(1) = j^{2} x_t \qquad (4.43)$$

Επαναλαμβάνοντας αυτή τη διαδικασία προκύπτει ότι η πρόβλεψη για *k* χρονικά βήματα είναι:

$$x_t(k) = j\, x_t(k-1) = j^{2} x_t \qquad (4.44)$$

Για το σφάλμα πρόβλεψης

$$e_t(1) = X_{t+1} - X_t(1) \qquad (4.45)$$

αποδεικνύεται εύκολα ότι:

$$e_t(1) = x_{t+1} \qquad (4.46)$$

δηλαδή το σφάλμα της πρόβλεψης $e_t(1)$ είναι λευκός θόρυβος με μηδενική μέση τιμή και:

$$E(e_t(1)) = E(z_{t+1}) = 0 \qquad (4.47)$$

και διασπορά s_z^2

$$Var(e_t(1)) = Var(z_{t+1}) = s_z^2 \qquad (4.48)$$

Για πρόβλεψη στην περίοδο *t + 2*:

$$x_t(2) = j\, x_t(1) \tag{4.49}$$

$$e_t(2) = z_{t+2} + j\, e_t(1) = z_{t+2} + j\, z_{t+1} \tag{4.50}$$

και

$$Var(e_t(2)) = s^{2}(1 + j^{2}) \tag{4.51}$$

Γενικά, για προβλέψεις k χρονικά βήματα μπροστά, ισχύει ότι:

$$x_t(k) = j\, x_t(k-1) \tag{4.52}$$

και

$$Var(e_t(k)) = s^{2}(1 + j^{2} + j^{4} + \ldots + j^{2(k-1)}) \tag{4.53}$$

ή

$$Var(e_t(k)) = s_z^{2}\frac{1 - j^{2k}}{1 - j^{2}} \tag{4.54}$$

Από τη σχέση (4.54), είναι φανερό ότι η διακύμανση του σφάλματος πρόβλεψης αυξάνει μη γραμμικά, καθώς μεγαλώνει η περίοδος πρόβλεψης. Εάν υποθέσουμε ότι η χρονολογική σειρά έχει μέση τιμή *δ*, στο όριο, καθώς η περίοδος πρόβλεψης μεγαλώνει, η πρόβλεψη συγκλίνει προς τον μέσο.

Δηλαδή:

$$\lim_{h \to \infty} x_t(k) = \frac{d}{1 - j} = m \tag{4.55}$$

<u>AR(p) Μοντέλο</u>

Γίνεται η υπόθεση πως η παρατηρούμενη χρονική σειρά $\{x_1, x_2, \ldots, x_t\}$ είναι η πραγματοποίηση μιας αυτοπαλίνδρομης διαδικασίας τάξης *p*, AR(*p*), ή πιο ρεαλιστικά γίνεται η υπόθεση εργασίας πως το μοντέλο AR(p) εξηγεί ικονποιητικά τη χρονολογική σειρά.

Το μοντέλο AR(p) για το x_{t+1} είναι:

$$x_{t+1} = j\, x_t + \ldots + j_p x_{t-p+1} + z_{t+1} \tag{4.56}$$

Ομοίως με πριν, η βέλτιστη πρόβλεψη του x_{t+1} είναι το αιτιοκρατικό μέρος του αυτοπαλίνδρομου AR μοντέλου.

Γενικά, για k χρονικά βήματα, η πρόβλεψη είναι:

$$x_t(k) = j_1 x_t(k-1) + \ldots + j_p x_t(k-p) \tag{4.59}$$

όπου κάθε τιμή $x_t(j)$ είναι γνωστή είτε από την προηγούμενη πρόβλεψη ή απευθείας από την χρονολογική σειρά. Δηλαδή για $j > 0$, το $x_t(j)$ είναι μια από τις προβλέψεις $x_t(1), x_t(2), \ldots, x_t(k-1)$ που έχουν προηγηθεί, ενώ για $j \le 0$ ισχύει $x_t(j) = x_{t+1}$ που ανήκει στο σύνολο $\{x_1, x_2, \ldots, x_t\}$, και είναι μια γνωστή τιμή της χρονολογικής σειράς. Η πρόβλεψη συνίσταται πάλι στο αιτιοκρατικό μέρος του αυτοπαλίνδρομου μοντέλου AR για το x_{t+k}, όπου οι άγνωστες παρατηρήσεις (οι θεωρητικές μεταβλητές) $x_{t+1}, x_{t+2}, \ldots, x_{t+k-1}$ έχουν αντικατασταθεί από τις αντίστοιχες προβλέψεις.

Το σφάλμα πρόβλεψης για *k* χρονικά βήματα δίνεται ως γραμμικός συνδυασμός των στοιχείων του λευκού θορύβου στις χρονικές στιγμές *t + 1, t + 2,…, t + k*

$$e_t(k) = \sum_{j=0}^{k-1} b_j z_{t+k-j} \tag{4.60}$$

όπου κάθε b_j ορίζεται από τις παραμέτρους του μοντέλου $(b_0 = 1)$. Έτσι, το $e_t(k)$ έχει μηδενική μέση τιμή

$$E(e_t(k)) = 0 \tag{4.61}$$

και διασπορά

$$Var(e_t(k)) = s_z^2 \sum_{j=0}^{k-1} b_j^2 \qquad (4.62)$$

Από την παραπάνω διασπορά πρόβλεψης μπορούν να σχηματιστούν όρια πρόβλεψης (prediction bounds, tolerance intervals) για δεδομένο επίπεδο σημαντικότητας *α*:

$$x_t(k) \pm c_{1-\frac{a}{2}} \sqrt{Var(e_t(k))} \qquad (4.63)$$

όπου $c_{1-\frac{a}{2}}$ είναι η κρίσιμη τιμή. Για παράδειγμα, εάν:

$$z_t \sim N(0, s_z^2) \qquad (4.64)$$

της τυπικής κανονικής. Δηλαδή για το 95% διάστημα πρόβλεψης θα χρησιμοποιηθεί το $c_{0.028} - z_{0.976} - 1,96$.

4.2.2 Πρόβλεψη με μοντέλα μέσου όρου

<u>MA(1) Μοντέλο</u>

Έστω το μοντέλο κινητού μέσου τάξης 1, MA(1),

$$x_t = z_t + q_{t-1} \qquad (4.65)$$

Η πρόβλεψη στην περίοδο t + 1 είναι:

$$x_t(1) = E_t(z_{t+1} + q\ z_t) \qquad (4.66)$$

ή

$$x_t(1) = E_t(q\ z_t) \qquad (4.67)$$

αφού $E_t(z_{t+1}) = 0$

Το σφάλμα πρόβλεψης είναι:

$$e_t(1) = x_{t+1} - x_t(1) = z_{t+1} \qquad (4.68)$$

Συνεπώς, η διακύμανση του σφάλματος της πρόβλεψης είναι:

$$Var(e_t(1)) = Var(z_{t+1}) = s^2 \qquad (4.69)$$

Η πρόβλεψη για δύο χρονικές περιόδους μπροστά, δηλαδή την χρονική στιγμή *t + 2* είναι:

$$x_t(2) = E_t(x_{t+2}) = E_t(z_{t+2} + q\, z_{t+1}) \qquad (4.70)$$

ή

$$x_t(2) = 0 \qquad (4.71)$$

αφού

$$E_t(z_{t+1}) = E_t(z_{t+2}) = 0$$

Η διακύμανση του σφάλματος πρόβλεψης είναι:

$$Var(e_t(2)) = Var(z_{t+2} + q\, z_{t+1}) = s^2(1+q^2) \qquad (4.72)$$

Γενικά, για να γίνει η πρόβλεψη για ένα ή περισσότερα χρονικά βήματα χρησιμοποιείται σαν δεδομένο ότι η τυχαία μεταβλητή z_t είναι ανεξάρτητη του *x*, για χρονικές στιγμές μικρότερες του *t* και έτσι προκύπτει:

$$E(z_{t+j} \mid x_t, x_{t-1}, \ldots) = \begin{cases} 0 & ean\; j > 0 \\ z_{t+j} & ean\; j \le 0 \end{cases} \qquad (4.73)$$

Η πρόβλεψη είναι:

$$x_t(k) = \begin{cases} q\, z_t & gia\; k = 1 \\ 0 & gia\; k > 1 \end{cases} \qquad (4.74)$$

και το σφάλμα:

$$e_t(k) = \begin{cases} z_{t+1} & gia\; k = 1 \\ z_{t+k} = s^2(1+q^2) & gia\; k > 1 \end{cases} \qquad (4.75)$$

Γίνεται φανερό λοιπόν, πως όλες οι προβλέψεις για χρονικές στιγμές μεγαλύτερες της μονάδας είναι μηδέν, και έτσι το μοντέλο κινητού μέσου τάξης 1, MA(1), είναι κατάλληλο μόνο για προβλέψεις μιας περιόδου μπροστά.

Τέλος, παρατηρείται πως η πρόβλεψη μιας χρονικής περιόδου μπροστά εξαρτάται από την τιμή του τυχαίου όρου z στην χρονική περίοδο *t*. Επειδή αυτή η τιμή είναι άγνωστη, στην πράξη αντικαθίσταται από την εκτίμησή της, η οποία προκύπτει από τις προηγούμενες τιμές. Συγκεκριμένα, αρχίζοντας από την χρονική στιγμή *t = 0*, η τιμή της *X* είναι:

$$x_1 = z_1 + q\, z_0 \tag{4.76}$$

Η πρόβλεψη είναι:

$$x_0(1) = \mathrm{E}_0(x_1) = q\, z_0 \tag{4.77}$$

αφού $\mathrm{E}_0(z_1) = 0$.

Η τιμή z_0 όμως, που αναφέρεται στην περίοδο πριν από το δείγμα, δεν είναι συνήθως γνωστή και αντικαθίστανται από την προσδοκώμενη τιμή της που είναι μηδέν.

Επομένως,

$$x_0(1) = 0 \tag{4.78}$$

και

$$e_0(1) = x_1 - x_0(1) = z_1 \tag{4.79}$$

Άρα,

$$x_1 = z_1 \tag{4.80}$$

Η πρόβλεψη της x μιας περιόδου μπροστά από την χρονική περίοδο *t = 1* , $x_1(1)$, είναι:

$$x_1(1) = q\, z_1 = q\, e_0(1) \tag{4.81}$$

και

$$e_1(1) = x_2 - x_1(1) = z_2 \tag{4.82}$$

Συνεχίζοντας με αυτόν τον τρόπο εκτιμώνται όλα τα *z*, δηλαδή τα $e_0(1), e_1(1), e_2(1), ..., e_{t-1}(1)$.Έτσι, για να γίνει η πρόβλεψη στη περίοδο *t + 1*, γίνεται εκτίμηση της z_t .

<u>MA(*q*) Μοντέλο.</u>

Θεωρώντας το μοντέλο κινητού μέσου τάξης *q*, MA(*q*), για την χρονολογική σειρά $\{x_1, x_2, ..., x_t\}$ η επόμενη παρατήρηση δίνεται ως:

$$x_{t+1} = z_{t+1} + q_1 z_t + ... + q_q z_{t-q+1} \tag{4.83}$$

Η βέλτιστη πρόβλεψη ενός βήματος $x_t(1)$, στην περίοδο *t + 1*, είναι:

$$x_t(1) = q_1 z_t + ... + q_q z_{t-q+1} \tag{4.84}$$

αφού $E_t(z_{t+1}) = 0$.

Τα αντίστοιχο σφάλμα πρόβλεψης είναι:

$$e_t(1) = x_{t+1} - x_t(1) = z_{t+1} \tag{4.85}$$

Γενικά, η πρόβλεψη για *k* χρονικές περιόδους μπροστά, δηλαδή την χρονική στιγμή *t + k* είναι:

$$x_t(k) = \begin{cases} q_k z_t + q_{k+1} z_{t-1} + ... + q_q z_{t-q+k} & \text{για } k \le q \\ 0 & \quad\;\; k > q \end{cases} \tag{4.86}$$

Το σφάλμα πρόβλεψης των *k* χρονικών βημάτων δίνεται ως γραμμικός συνδυασμός των στοιχείων του λευκού θορύβου στις χρονικές στιγμές *t + 1, t + 2, …t + k* και είναι όπως του αυτοπαλίνδρομου μοντέλου AR για $k \le q$ εάν αντικατασταθούν τα b_j με τα q_j .

$$e_t(k) = \sum_{j=0}^{k-1} q_j z_{t+k-j} \tag{4.87}$$

Έτσι, το $e_t(k)$ έχει μηδενική μέση τιμή:

$$E(e_t(k)) = 0 \tag{4.88}$$

και διασπορά:

$$Var(e_t(k)) = s_z^2 \sum_{j=0}^{k-1} q_j^2 \tag{4.89}$$

Ομοίως, με τα αυτοπαλίνδρομα μοντέλα, από την παραπάνω διασπορά πρόβλεψης μπορούν να σχηματιστούν όρια πρόβλεψης για δεδομένο επίπεδο σημαντικότητας *α*.

$$x_t(k) \pm c_{1-\frac{a}{2}} \sqrt{Var(e_t(k))} \tag{4.90}$$

όπου $c_{1-\frac{a}{2}}$ είναι η κατάλληλη κρίσιμη στιγμή.

Στα κινητού μέσου μοντέλα, τα σφάλματα z_t, z_{t-1}, z_{t-2}, μπορούν να υπολογιστούν από τις παρατηρήσεις $x_t, x_{t-1}, x_{t-2}, \ldots$ όπου οι αρχικές τιμές $z_1, z_2, \ldots, z_q$ είναι μηδέν. Ειδικότερα, για να βρεθούν τα $z_{q+1}, z_{q+2}, \ldots, z_t$, λύνεται η εξίσωση του MA(q) για t = q ως προς z_{q+1},δηλαδή εδώ είναι $z_{q+1} = x_{q+1}$ και συνεχίζοντας με τον ίδιο τρόπο για χρόνους t – q + 1,…,t – 1.

4.2.2 *Πρόβλεψη με αυτοπαλίνδρομα μοντέλα κινητού μέσου.*

Θεωρώντας ένα μεικτό μοντέλο ARMA(p, q) για την χρονική σειρά $\{x_1, x_2, \ldots, x_t\}$ η επόμενη παρατήρηση δίνεται ως:

$$x_{t+1} = j_1 x_t + \ldots + j_p x_{t-p+1} + \ldots + q_1 z_t + q_q z_{t-q+1} + z_{t+1} \tag{4.91}$$

Η βέλτιστη πρόβλεψη για ένα χρονικό βήμα, όταν δίνονται τα $\{x_1, x_2, \ldots, x_t\}$ είναι:

$$x_t(1) = j_1 x_t + \ldots + j_p x_{t-p+1} + \ldots + q_1 z_t + q_q z_{t-q+1} + z_{t+1} \tag{4.92}$$

και το σφάλμα πρόβλεψης:

$$e_t(1) = z_{t+1} \tag{4.93}$$

Γενικά, για k χρονικά βήματα η βέλτιστη πρόβλεψη είναι:

$$x_t(k) = \begin{cases} j_1 x_t(k-1) + \ldots + j_p x_t(k-p) + q_k z_t + \ldots + q_q z_{t-q+1} + z_{t+1} & \text{για } k \leq q \\ j_1 x_t(k-1) + \ldots + j_p x_t(k-p) & k > q \end{cases} \tag{4.94}$$

Η πρόβλεψη με μεικτά μοντέλα ARMA είναι η σύνθεση των προβλέψεων με το αυτοπαλίνδρομο μέρος και το μέρος του κινητού μέσου.

4.3 ***Πρόβλεψη μη – στάσιμων χρονολογικών σειρών με γραμμικά μοντέλα.***

Οι προβλέψεις στην χρονολογική σειρά που μετατράπηκε σε στάσιμη από μια μη- στάσιμη χρονολογική σειρά, θα πρέπει να μετασχηματιστούν κατάλληλα για να αναφέρονται στην αρχική χρονολογική σειρά. Όταν λοιπόν, η χρονολογική σειρά δεν είναι στάσιμη, για να εφαρμοστεί η πρόβλεψη με τα μοντέλα της προηγούμενης παραγράφου πρέπει να γίνουν τα εξής βήματα:

Το **πρώτο βήμα** είναι να μετασχηματιστεί η χρονολογική σειρά σε στάσιμη, δηλαδή η μη-στάσιμη $\{x_t\}$ σειρά να μετασχηματιστεί στην $\{y_t\}$ που θα είναι στάσιμη.

Δεύτερο βήμα είναι να γίνει η πρόβλεψη του y_{t+k} με κάποιο μοντέλο, για παράδειγμα ενός αυτοπαλίνδρομου μοντέλου AR, έστω $y_t(k)$.

Στη συνέχεια, **τρίτο βήμα** είναι να μετασχηματιστεί η πρόβλεψη $y_t(k)$ για την στάσιμη χρονολογική σειρά, στην πρόβλεψη $x_t(k)$ για την αρχική μη – στάσιμη χρονολογική σειρά.

Στη συνέχεια, θα εφαρμοστούν αναλυτικά τα παραπάνω. Έστω, πως η χρονολογική σειρά $\{x_1, x_2, ..., x_t\}$ δεν είναι στάσιμη. Το κλασικό μοντέλο για το x_t είναι:

$$x_t = m_t + s_t + y_t \tag{4.95}$$

όπου m_t είναι η συνάρτηση τάσης, s_t είναι η περιοδικότητα ή η εποχική συνάρτηση και y_t είναι μια στάσιμη χρονολογική σειρά, απαλλαγμένη από τάσεις και περιοδικότητες. Τυπικά μοντέλα για την χρονολογική σειρά $\{y_1, y_2, ..., y_t\}$ είναι τα γνωστά μοντέλα AR, MA και ARMA.

Αντικειμενικός σκοπός είναι, δοθέντων των $\{x_1, x_2, ..., x_t\}$, να βρεθεί η πρόβλεψη για ένα χρονικό βήμα ή γενικά για k χρονικά βήματα μπροστά, δηλαδή να προβλεφθεί το x_{t+k} όπου

$$x_{t+k} = m_{t+k} + s_{t+k} + y_{t+k} \tag{4.96}$$

Εάν επιλεγεί να εκτιμηθούν τα m_t και s_t ως συναρτήσεις του χρόνου t (για παράδειγμα να προσαρμοστεί στο m_t ένα πολυώνυμο (4.22), τότε επεκτείνονται οι εκτιμήσεις στο χρόνο t + k για να βρεθούν τα m_{t+k} και s_{t+k}. Σε αυτήν τη περίπτωση, αφαιρώντας από τις τιμές $x_1, x_2, ..., x_t$, τις εκτιμήσεις της τάσης και της περιοδικότητας, προκύπτουν οι τιμές των $y_1, y_2, ..., y_t$ (πρώτο βήμα). Στη συνέχεια, γίνεται η πρόβλεψη του y_{t+k} με την χρήση κάποιου μοντέλου τύπου AR, MA ή ARMA(δεύτερο βήμα). Η πρόβλεψη $x_t(k)$ προκύπτει απευθείας από την πρόβλεψη $y_t(k)$ και τις επεκτάσεις της τάσης και της περιοδικότητας m_{t+k} και s_{t+k} (τρίτο βήμα) ως:

$$x_t(k) = m_{t+k} + s_{t+k} + y_t(k) \tag{4.97}$$

Εάν επιλεγεί να απαλειφτούν τα m_t και s_t χρησιμοποιώντας διαφορές, τότε η πρόβλεψη με τα παραπάνω τρία βήματα είναι η πρόβλεψη με μοντέλα ARIMA ή SARIMA. Οι γενικοί τύποι για τις προβλέψεις με αυτά τα μοντέλα είναι πολύπλοκοι, αλλά κάποιος μπορεί να καταλάβει πώς γίνεται η πρόβλεψη με ARIMA μοντέλο θεωρώντας το ARIMA(*p, l, q*). Παίρνοντας τις πρώτες διαφορές με υστέρηση μια μονάδα, από την αρχική χρονολογική σειρά $\{x_1, x_2, ..., x_t\}$ προκύπτει η χρονολογική σειρά $\{y_1, y_2, ..., y_t\}$, όπου:

$$y_t = x_t - x_{t-1} \tag{4.98}$$

σχηματίζοντας έτσι το πρώτο βήμα της διαδικασίας πρόβλεψης μη – στάσιμων χρονολογικών σειρών. Εφαρμόζοντας το μοντέλο ARMA(*p*,*q*) στην $\{y_1, y_2, ..., y_t\}$, βρίσκεται η πρόβλεψη για ένα χρονικό βήμα $y_t(1)$ (δεύτερο βήμα) και η πρόβλεψη για την αρχική χρονολογική σειρά (τρίτο βήμα) είναι:

$$x_t(1) = x_t + y_t(1) \tag{4.99}$$

Το σφάλμα πρόβλεψης του x_{t+1}, το $e_t(1)$, είναι το ίδιο με το σφάλμα πρόβλεψης του y_{t+1}. Γενικά, η πρόβλεψη για k χρονικά βήματα είναι:

$$x_t(k) = x_t(k-1) + y_t(k) \tag{4.100}$$

όπου $y_t(k)$ είναι η πρόβλεψη του y_{t+1} με το μοντέλο ARMA(*p*,*q*) και το $x_t(k-1)$ είναι γνωστό από την πρόβλεψη του x_{t+k-1}.

Για ARIMA (*p*,*d*,*q*) ή SARIMA(*p*,*d*,*q*) (*P*,*D*,*Q*), η διαδικασία της πρόβλεψης του x_{t+1} είναι παρόμοια, δηλαδή βρίσκεται η πρόβλεψη $y_t(1)$ με μοντέλο τύπου ARMA και προστίθεται στην κατάλληλη έκφραση των τελευταίων παρατηρήσεων $x_t, x_{t-1}, ...$, σύμφωνα με τις τιμές των *d* και *D*. Η πρόβλεψη για *k* χρονικά βήματα μπορεί να υπολογιστεί αναδρομικά.

4.4 Εφαρμογές

1.Η ΜΕΘΟΔΟΛΟΓΙΑ ΤΗΣ PUBLIC ISSUE ΜΕ ΤΗ ΧΡΗΣΗ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ ΓΙΑ ΤΗ ΔΙΕΞΑΓΩΓΗ ΕΚΛΟΓΙΚΩΝ ΠΡΟΒΛΕΨΕΩΝ.

1.Οι τρεις βασικές μεθοδολογικές καινοτομίες.

Από τις εκλογές του 2004, η Public Issue εισήγαγε τρεις βασικές καινοτομίες στην εκτίμηση της εκλογικής επιρροής των πολιτικών κομμάτων, που στηρίζεται στις έρευνες πρόθεσης ψήφου:

1. την **τηλεφωνική μέθοδο** (παράλληλα με την εταιρεία RASS), έναντι της έως τότε χρησιμοποιούμενης μεθόδου των ερευνών, με προσωπικές συνεντεύξεις (face-to-face) και «χρήση κάλπης», που αποτελεί μια αμφιλεγόμενη, αμιγώς ελληνική πρωτοτυπία.
2. την κατάργηση της ***στάθμισης*** του δείγματος με την προηγούμενη ψήφο και την αντικατάστασή της, με τη μεθοδολογία της ***ανάλυσης χρονολογικών σειρών*** (time-series analysis) των ερευνών της και, τέλος
3. την αλλαγή του τρόπου παρουσίασης των αποτελεσμάτων των δημοσκοπήσεων, με τη δημοσίευση της εκλογικής επιρροής των κομμάτων σε κάθε έρευνα, αντί της παρουσίασης των αποτελεσμάτων με την λεγόμενη «αδιευκρίνιστη» ψήφο, πρακτική που ακολουθούν οι υπόλοιπες εταιρείες του κλάδου.

Με τον όρο **εκτίμηση εκλογικής επιρροής** εννοείται ο επανυπολογισμός των ποσοστών των κομμάτων (repercentage), ώστε τα αποτελέσματα της έρευνας να είναι συγκρίσιμα με τα εκλογικά αποτελέσματα (δηλαδή να μην συμπεριλαμβάνουν «αδιευκρίνιστη» ψήφο). Η κατανομή της «αδιευκρίνιστης» ψήφου (δηλαδή όσων δεν διευκρίνισαν την πρόθεση ψήφου τους) που προκύπτει, δεν είναι αναλογική (απλή απαλειφή), αλλά πραγματοποιείται με τη χρήση στατιστικών υποδειγμάτων.

Η εκτίμηση της εκλογικής επιρροής γίνεται συνδυάζοντας την ανάλυση των βραχυχρόνιων και των μακροχρόνιων τάσεων της πρόθεσης ψήφου.

2. Η ανάλυση των βραχυχρόνιων τάσεων

Ο όρος **βραχυχρόνιες τάσεις** στη διαχρονική εξέλιξη της εκλογικής επιρροής ενός κόμματος περιγράφει τις τάσεις που διαμορφώνονται, κατά τη διάρκεια ενός εκλογικού κύκλου. Για τη μελέτη των βραχυχρόνιων τάσεων της επιρροής κάθε κόμματος, χρησιμοποιείται η χρονολογική σειρά των διευκρινισμένων απαντήσεων στην ερώτηση πρόθεσης ψήφου, στις μηνιαίες τηλεφωνικές έρευνες του Βαρόμετρου της Public Issue, από τις εκλογές του 2007 και ύστερα.

Η ανάλυση των βραχυχρόνιων τάσεων γίνεται ως εξής: Αρχικά εφαρμόζεται, στη σειρά πρόθεσης ψήφου κάθε κόμματος, η τεχνική εξομάλυνσης Kalman (Kalman smoothing). Η εξομάλυνση Kalman είναι μια τεχνική που μειώνει την επίδραση του δειγματοληπτικού σφάλματος, σε κάθε έρευνα, και, επομένως, επιτρέπει την διάκριση της πραγματικής μεταβολής, από το δειγματοληπτικό σφάλμα, στην διαχρονική εξέλιξη των ποσοστών ενός κόμματος.

Στη συνέχεια, οι σειρές των εξομαλυσμένων, με την τεχνική Kalman, ποσοστών αναλύονται με τη μεθοδολογία Box-Jenkins. Τα υποδείγματα χρονολογικών σειρών που κατασκευάζονται, είναι συνδυασμός ενός ARIMA υποδείγματος (**A**uto**R**egression **I**ntegrated **M**oving **A**verages) και μιας ψευδομεταβλητής (dummy variable), που περιγράφει τις βουλευτικές εκλογές του 2007. Τα υποδείγματα αυτά χρησιμοποιούνται, ώστε να απαλειφθεί η εγγενής μεροληψία των δημοσκοπήσεων.

Επικουρικά, στις μηνιαίες τηλεφωνικές έρευνες του Βαρόμετρου της Public Issue, χρησιμοποιείται και η τοπική παλινδρόμηση εξομάλυνσης γραφήματος (**Lo**cally W**e**ighted **S**catterplot **S**moothing - loess regression). Πρόκειται για μια στατιστική τεχνική που προσπαθεί να «ανιχνεύσει» τάσεις, με την ευρύτερη έννοια, στην διαχρονική εξέλιξη των ποσοστών ενός κόμματος, εξουδετερώνοντας τις ακραίες-έκτροπες τιμές (outliers), είτε αυτές οφείλονται σε αληθινά γεγονότα, είτε σε δειγματοληπτικά σφάλματα.

3. Η ανάλυση των μακροχρόνιων τάσεων

Ο όρος **μακροχρόνιες τάσεις** στη διαχρονική εξέλιξη της <u>εκλογικής επιρροής</u> ενός κόμματος περιγράφει τις τάσεις που διαμορφώνονται σε ένα χρονικό διάστημα, μεγαλύτερο του ενός εκλογικού κύκλου. Για τη μελέτη των μακροχρόνιων τάσεων της επιρροής κάθε κόμματος, χρησιμοποιείται η χρονολογική σειρά των διευκρινισμένων απαντήσεων στην ερώτηση πρόθεσης ψήφου, σε όλες τις έρευνες από τον Ιούνιο του 1995 ως σήμερα, της Public Issue και παλαιότερα της VPRC (Έρευνες VPRC, για την περίοδο 1995-2003, Βαρόμετρο Public Issue, από το 2004 και εξής). Μέχρι τον Δεκέμβριο του 2003, οι έρευνες πραγματοποιήθηκαν με προσωπικές συνεντεύξεις (face to face), ενώ από τον Ιανουάριο του 2004 και ύστερα με τηλεφωνικές συνεντεύξεις.

<u>Οι χρονοσειρές της εκλογικής επιρροής</u> κατασκευάζονται σε τριμηνιαία βάση, παίρνοντας ως τιμή για κάθε τρίμηνο το μέσο όρο των ερευνών του τριμήνου. Οι σειρές πρόθεσης ψήφου αναλύονται, επίσης, με τη μεθοδολογία Box-Jenkins. Τα υποδείγματα χρονολογικών σειρών που κατασκευάζονται, είναι συνδυασμός ενός ARIMA υποδείγματος (**A**uto**R**egression **I**ntegrated **M**oving **A**verages) και ψευδομεταβλητών (dummy variables), που περιγράφουν τα έκτακτα γεγονότα (interventions) που επηρέασαν την εξέλιξη της επιρροής των κομμάτων στην υπό μελέτη περίοδο. Ανάμεσα στα έκτακτα γεγονότα συμπεριλαμβάνονται όλες οι εκλογικές αναμετρήσεις από το 1996 και ύστερα.

4. Συνδυασμός των βραχυχρόνιων και των μακροχρόνιων τάσεων

Τόσο οι βραχυχρόνιες, όσο και οι μακροχρόνιες τάσεις προσεγγίζουν το ίδιο ζητούμενο, δηλαδή την <u>εκλογική επιρροή των κομμάτων</u>, από διαφορετική οπτική γωνία. Επομένως, είναι λογικό να μη δίνουν πάντα το ίδιο αποτέλεσμα και, κατά συνέπεια, οι εκτιμήσεις που προέρχονται από αυτές τις δύο διαφορετικές αφετηρίες, πρέπει να συνεκτιμώνται. Η εκτίμηση που δημοσιοποιείται από το Βαρόμετρο, βασίζεται στο μέσο όρο των βραχυχρόνιων και μακροχρόνιων εκτιμήσεων.

5. Το περιθώριο σφάλματος της εκτίμησης

Το περιθώριο σφάλματος της εκτίμησης προκύπτει από τα υποδείγματα χρονολογικών σειρών, με τα οποία γίνεται η εκτίμηση (ARIMA υποδείγματα και εξομάλυνση Kalman). Το περιθώριο σφάλματος της εκτίμησης δεν είναι το γνωστό δειγματοληπτικό σφάλμα (sampling error). Το περιθώριο σφάλματος της εκτίμησης επηρεάζεται από πολλούς παράγοντες, όπως: το μέγεθος του δείγματος, το ποσοστό του κόμματος στη διευκρινισμένη ψήφο, την καλή προσαρμογή των υποδειγμάτων στη χρονοσειρά των δεδομένων (fitting) και το συγκεκριμένο ARIMA υπόδειγμα που επιλέγεται.

6. Η επιτυχής εφαρμογή της μεθοδολογίας στις εκλογές της 16ης Σεπτεμβρίου 2007

Η Public Issue εφάρμοσε την παραπάνω μεθοδολογία με πρωτοφανή επιτυχία, στις εκλογές της 16ης Σεπτεμβρίου 2007.

Το αποτέλεσμα της εκτίμησης που δημοσιοποιήθηκε, προεκλογικά, από την τηλεόραση και το ραδιόφωνο του ΣΚΑΪ, στις 31/8/2007, τελευταία ημέρα πριν την απαγόρευση δημοσίευσης των δημοσκοπήσεων, παρατίθεται στον πίνακα που ακολουθεί:

Κόμματα	Εκτίμηση 31/8	Περιθώριο σφάλματος +/-	Εκλογές 16/9/07	Απόκλιση εκτίμησης από εκλογικό αποτέλεσμα
ΝΔ	42	1,5	41,8	0,2
ΠΑΣΟΚ	38	1,5	38,1	- 0,1
ΚΚΕ	8,5	1	8,2	0,3
ΣΥΝ	5	1	5	0
ΛΑΟΣ	4	1	3,8	0,2
ΛΟΙΠΑ	2,5	1	3,1	- 0,6
Διαφορά ΝΔ - ΠΑΣΟΚ	4		3,7	0,3

2.ΣΤΑΤΙΣΤΙΚΑ ΔΕΔΟΜΕΝΑ ΤΟΥ ΥΠΟΥΡΓΕΙΟΥ ΑΝΑΠΤΥΞΗΣ ΜΕ ΤΗ ΒΟΗΘΕΙΑ ΤΗΣ ΕΛ.ΣΤΑΤ ΓΙΑ ΤΗΝ ΠΡΟΒΛΕΨΗ ΖΩΙΚΗΣ ΚΑΙ ΦΥΤΙΚΗΣ ΠΑΡΑΓΩΓΗΣ.

Το Υπουργείο Αγροτικής Ανάπτυξης και Τροφίμων χρησιμοποιεί όλες τις χρονολογικές σειρές για τη διεξαγωγή αποτελεσμάτων σχετικά με την εξέλιξη της παραγωγής,για μια σειρά προϊόντων φυτικής και ζωικής παραγωγής.

Σε ό τι αφορά την ποιότητα των στοιχείων αυτών και τη σχετική μεθοδολογία συγκέντρωσής τους, παρέχονται οι εξής διευκρινήσεις:
Η τρέχουσα Γεωργική Στατιστική πληροφόρηση στην Ελλάδα εξυπηρετείται από δύο φορείς, την Ελληνική Στατιστική Αρχή (ΕΛ.ΣΤΑΤ) και το Υπουργείο Αγροτικής Ανάπτυξης & Τροφίμων.
Η ΕΛ.ΣΤΑΤ. διενεργεί το σύνολο των ερευνών της σε επίπεδο γεωργικής εκμετάλλευσης, καθώς και μία έρευνα σε επίπεδο Δήμου ή Κοινότητος, την Ετήσια Γεωργική Στατιστική.
Το Υπουργείο Αγροτικής Ανάπτυξης & Τροφίμων συγκεντρώνει τα στατιστικά στοιχεία και τις πληροφορίες που αφορούν τόσο στις προβλέψεις των εκτάσεων για τις ετήσιες καλλιέργειες και την παραγωγή των φυτικών και ζωικών προϊόντων, όσο και στα τελικά στοιχεία των παραπάνω οικονομικών μεγεθών με την μέθοδο των εκτιμήσεων των εμπειρογνωμόνων των περιφερειακών υπηρεσιών του.
Οι τυχόν διαφορές των στοιχείων αυτών με άλλα παρόμοια, που προέρχονται από άλλες πηγές ή και από την ίδια πηγή, είναι δικαιολογημένες, γιατί οφείλονται είτε στη διαφορετική μεθοδολογία εκτίμησης των μεγεθών (έρευνα σε επίπεδο γεωργικών εκμεταλλεύσεων - area frame sampling, επιδοτούμενες δραστηριότητες, κ.λπ.) είτε στην εκτίμηση των ίδιων οικονομικών μεγεθών από διάφορους άλλους εμπειρογνώμονες.
Παράλληλα με τις παραπάνω πηγές, χρησιμοποιούνται από διάφορους άλλους φορείς (λ.χ. υπηρεσίες της Ευρωπαϊκής Ένωσης) άλλες προσεγγίσεις, όπως είναι οι καταγραφές εκτάσεως διάφορων χρήσεων γης με φωτοερμηνευτικές, κυρίως, μεθόδους κ.λπ.

ΣΤΑΤΙΣΤΙΚΕΣ ΧΡΟΝΟΛΟΓΙΚΕΣ ΣΕΙΡΕΣ

Τα στοιχεία χρονολογικών σειρών φυτικής και ζωικής παραγωγής, που παρουσιάζονται εδώ, είναι, όπως αναφέρεται στην αρχή, εκτιμήσεις εμπειρογνωμόνων των νομαρχιακών υπηρεσιών του Υπουργείου Αγροτικής Ανάπτυξης & Τροφίμων.

Οι τιμές παραγωγού συγκεντρώνονται είτε από τα κέντρα διακίνησης των γεωργικών προϊόντων (έμποροι, συνεταιριστικές οργανώσεις) είτε από τους παραγωγούς μέσω των νομαρχιακών υπηρεσιών του Υπουργείου Αγροτικής Ανάπτυξης & Τροφίμων.
Οι τιμές παραγωγού είναι χονδρικές τιμές, αναφέρονται στην πόρτα της γεωργικής εκμετάλλευσης (δεν περιλαμβάνουν τυχόν δαπάνες μεταφορικών από την εκμετάλλευση στα κέντρα συγκέντρωσης), περιλαμβάνουν τις έμμεσες οικονομικές ενισχύσεις και είναι απαλλαγμένες από επιβαρύνσεις (π.χ. κρατήσεις υπέρ ΕΛ.Γ.Α.).

Η έρευνα γάλακτος και γαλακτοκομικών προϊόντων είναι κοινοτική και διενεργείται σε εκτέλεση της οδηγίας 96/16/ΕΚ του Συμβουλίου. Συγκεντρώνει στοιχεία σχετικά με τις χρήσεις του παραγομένου γάλακτος και των προϊόντων που προέρχονται από την επεξεργασία του τόσο στις αγροτικές εκμεταλλεύσεις, όσο και στις γαλακτοκομικές επιχειρήσεις.

ΓΕΩΡΓΙΚΑ ΙΣΟΖΥΓΙΑ

Η κατάρτιση των ισοζυγίων των γεωργικών προϊόντων είναι κοινοτική εφαρμογή και στηρίζεται στην αρχή της «Συμφωνίας Κυρίων». Ένα ισοζύγιο περιλαμβάνει τα εξής βασικά στοιχεία:
Παραγωγή + Αρχικά Αποθέματα + Εισαγωγές = Εξαγωγές + Εσωτερικές Χρήσεις + Τελικά Αποθέματα.

ΥΠΟΥΡΓΕΙΟ ΑΓΡΟΤΙΚΗΣ ΑΝΑΠΤΥΞΗΣ ΚΑΙ ΤΡΟΦΙΜΩΝ
ΔΙΕΥΘΥΝΣΗ Αγροτικής Πολιτικής & Τεκμηρίωσης
ΤΜΗΜΑ Αγροτικής Στατιστικής

Στοιχεία παραγωγής Αβγών 2001-2009 από τις Νομαρχιακές Δ/νσεις Αγροτικής Ανάπτυξης

Έτος	Όρνιθες χωρικής πτηνοτροφίας	Παραγωγή αβγών χωρικής πτηνοτροφίας	Όρνιθες αναπαραγωγής Συστηματικής πτηνοτροφίας	Αυγά εκκολάψεως Συστηματικής Πτηνοτροφίας	Όρνιθες απλής αβγοπαραγωγής Συστηματικής πτηνοτροφίας	Αβγά καταναλώσεως Συστηματικής Πτηνοτροφίας	Σύνολο γεννητριών Ορνίθων Συστηματικής Πτηνοτροφίας	Σύνολο αβγών Συστηματικής Πτηνοτροφίας	Συνολικός αριθμός Ορνίθων	Συνολική Παραγωγή αβγών (τεμάχια)	Συνολική Παραγωγή αβγών σε τόνους
2001	7.520.913	864.825.950	952.500	160.439.000	5.766.335	1.408.289.650	6.718.835	1.568.728.650	14.239.748	2.433.554.600	121.678
2002	7.284.431	800.261.650	885.200	153.169.000	5.878.246	1.437.334.630	6.763.446	1.590.503.630	14.047.877	2.390.765.280	119.538
2003	7.201.479	789.742.800	992.300	167.365.000	6.231.444	1.531.620.760	7.223.744	1.698.985.760	14.425.223	2.488.728.560	124.436
2004	6.927.059	765.865.900	1.115.300	188.405.000	6.181.544	1.558.066.120	7.296.844	1.746.471.120	14.223.903	2.512.337.020	125.617
2005	6.474.920	727.655.500	1.285.900	215.190.000	5.853.800	1.510.167.500	7.139.700	1.725.357.500	13.614.620	2.453.013.000	122.651
2006	6.062.356	677.531.250	1.178.000	201.765.000	5.538.620	1.375.945.500	6.716.620	1.577.710.500	12.778.976	2.255.241.750	112.762
2007	6.124.726	698.192.710	1.194.288	204.937.600	5.701.850	1.427.735.000	6.896.138	1.632.672.600	13.020.864	2.330.865.310	116.543
2008	5.231.086	590.066.100	1.182.885	203.140.000	6.001.965	1.453.510.750	7.184.850	1.656.650.750	12.415.936	2.246.716.850	112.336
2009	4.937.574	557.062.558	1.155.900	191.082.000	5.890.135	1.427.199.890	7.046.035	1.618.281.890	11.983.609	2.175.344.448	108.767

Βιβλιογραφία

1. C. Chatfield, (1989) The Analysis of Time Series, An Introduction, 4th edition, Chapman & Hall, London.

2. Γεωργίου Κ. Χρήστου, (2007) Εισαγωγή στην Οικονομετρία, Β΄ Τόμος, Γ΄ Έκδοση, Αθηνά, Gutenberg.

3. Hamilton, J. D. (1994) Time Series Analysis, Princeton.

4. Box, G. E. P. and Jenkins, G.M., (1970) Time Series Analysis: Forecasting and Control, San Francisco Holden Day.

5. Mills, T. C., Time Series Techniques for Economists, (1991) Cambridge University Press.

6. Griffiths, W. F., R. C. and Judge, G. C., (1993) Learning and Practicing Econometrics, John Wiley.

7. Nelson, C. R., (1973) Applied Time Series Analysis for Managerial Forecasting, Holden – Day.

8. Ljung, G. M. and Box, G. E. P., (1979) On a Measure of Lack of Fite in Time Series Models Biometrika, Vol. 66, pp. 265 – 270.

9. Pindyck, R. S. and Rubinfeld, D. L., (1998) Econometric Models and Economic Forecast, Irwin – McGraw Hill.

10. Theil, H., (1967) Applied Economic Forecasting, Rand McNally.

11. The Methods of Time Series, Analysis, D. S. G. Pollock, London.

12. Enders, W., Applied Econometric Time Series, 2th edition, John Wiley, New York, 2004.

13. Bartlet, M. S. (1966) Stochastic Processes (2nd edn), Cambridge: Cambridge University Press.

14. Schwarz, R. J. and Friedland, B. (1965) Linear Systems, New York: McGraw-Hill.

15. Box, G. E. P. and Pierce, D.A (1970) Distribution of residual auto-correlation in autoregressive-integrated moving average time-series models. J. Amer. Statist. Ass., 65, 1509-26.

16. Wei, W. W. S. (1990) Time Series Analysis, Redwood City, CA: Addison-

Wesley.
17. Akaike, H. (1967) Some problems iv the application of the cross-spectral methods. In Spectral Analysis of Time Series (ed. B. Harris), New York: Wiley.
18. Akaike, H. (1978) Time series analysis and control through parametric models. In Applied Time Series Analysis (ed. D. B. Findley), New York: Academic Press.
19. Harvey, A. C. (1993) Time Series Models (2nd edn), Hemel Hempstead: Harvester Wheatsheaf.
20. The Numerical Recipes (free downloadable from the net).