

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

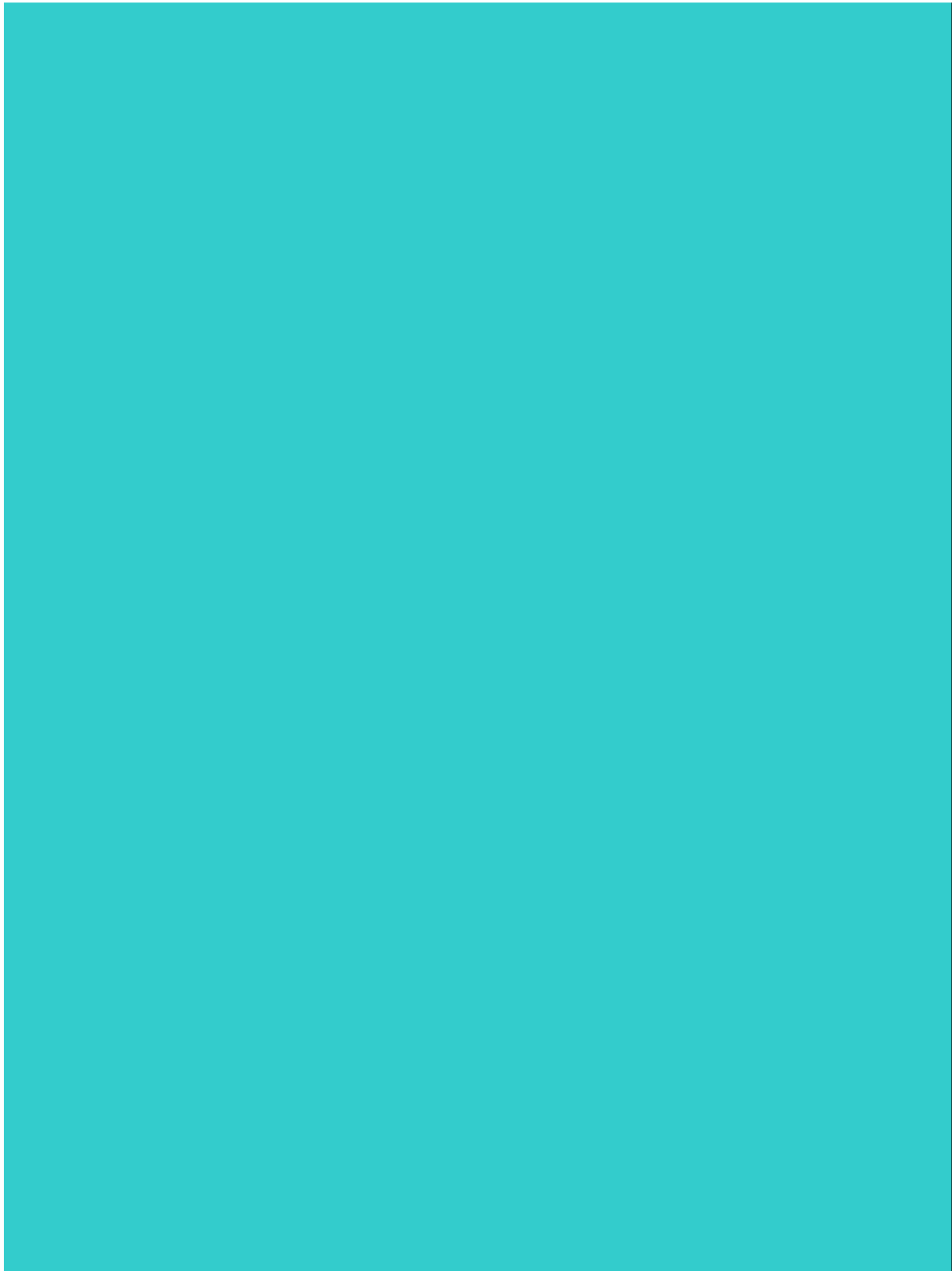
**ΥΠΟΦΑΙΝΟΜΕΝΙΚΕΣ ΣΧΕΣΕΙΣ
ΜΕΤΑΞΥ ΧΡΟΝΟΣΕΙΡΩΝ**

Απόστολος Η. Τσιμπάνος

Διπλωματική Εργασία

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

Πειραιάς
Νοέμβριος 2005



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΥΠΟΦΑΙΝΟΜΕΝΙΚΕΣ ΣΧΕΣΕΙΣ
ΜΕΤΑΞΥ ΧΡΟΝΟΣΕΙΡΩΝ**

Απόστολος Η. Τσιμπάνος

Διπλωματική Εργασία

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

Πειραιάς
Νοέμβριος 2005

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

Τα μέλη της Επιτροπής ήταν:

- Αν. Καθ. Χρήστος Αγιακλόγλου (Επιβλέπων)
- Καθηγητής Μάρκος Κούτρας
- Λέκτορας Γεώργιος Πιτσέλης

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

UNIVERSITY OF PIRAEUS



**DEPARTMENT OF STATISTICS
AND INSURANCE SCIENCE**

**POSTGRADUATE PROGRAM IN
APPLIED STATISTICS**

**SPURIOUS RELATIONS IN
TIME SERIES**

By

Apostolos E. Tsimpanos

MSc Dissertation

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

Piraeus, Greece
November 2005

*Στους γονείς μου
Ηλία και Εριέττα*

Ευχαριστίες

Αισθάνομαι την ανάγκη να ευχαριστήσω τον καθηγητή κύριο Μάρκο Κούτρα και τον λέκτορα κύριο Γεώργιο Πιτσέλη που ήταν μέλη της επιτροπής αξιολόγησης της Διπλωματικής μου Εργασίας. Ιδιαίτερα ευχαριστώ τον επιβλέποντα της Διπλωματικής μου Εργασίας αναπληρωτή καθηγητή κύριο Χρήστο Αγιακλόγλου για τη συνεχή καθοδήγησή του και την άψογη συνεργασία μας. Επίσης, επιθυμώ να ευχαριστήσω την οικογένειά μου για την ηθική συμπαράσταση που μου προσέφερε σ' όλο αυτό το χρονικό διάστημα.

Περίληψη

Στην εργασία αυτή εξετάζεται το φαινόμενο της ύπαρξης υποφαινομενικών σχέσεων μεταξύ χρονοσειρών. Ο όρος υποφαινομενική σχέση φανερώνει ότι δύο ή περισσότερες χρονοσειρές βρίσκονται να συσχετίζονται γραμμικά μεταξύ τους, σύμφωνα με τη στατιστική μεθοδολογία, ενώ στην πραγματικότητα αυτές δεν συσχετίζονται. Το φαινόμενο μπορεί να εμφανιστεί μέσω της εκτίμησης ενός γραμμικού υποδείγματος παλινδρόμησης με τη μέθοδο των ελαχίστων τετραγώνων οπότε ονομάζεται υποφαινομενική παλινδρόμηση αλλά και μέσω τον υπολογισμό του δειγματικού συντελεστή γραμμικής συσχέτισης οπότε ονομάζεται υποφαινομενική συσχέτιση. Τα χαρακτηριστικά της υποφαινομενικής παλινδρόμησης είναι η πολύ υψηλή τιμή του συντελεστή προσδιορισμού, οι υψηλές απόλυτες τιμές των t στατιστικών για τους ελέγχους της στατιστικής σημαντικότητας των συντελεστών και η πολύ χαμηλή τιμή (κοντά στο μηδέν) της στατιστικής των Durbin και Watson για τον έλεγχο της αυτοσυσχέτισης πρώτου βαθμού στα σφάλματα, που προκύπτουν από το εκτιμηθέν υπόδειγμα. Στην περίπτωση της υποφαινομενικής συσχέτισης το φαινόμενο χαρακτηρίζεται από την υψηλή τιμή του δειγματικού συντελεστή γραμμικής συσχέτισης και την υψηλή απόλυτη τιμή της t στατιστικής για τον έλεγχο της στατιστικής του σημαντικότητας. Η εμφάνιση των υψηλών αυτών τιμών για τις t στατιστικές αποτελεί σοβαρό πρόβλημα στην ανάλυση ποσοτικών σχέσεων διότι μπορεί να οδηγήσει στην εσφαλμένη απόρριψη της μηδενικής υπόθεσης των ελέγχων της στατιστικής σημαντικότητας των συντελεστών. Υποφαινομενικές σχέσεις μπορούν να εμφανιστούν με χρονοσειρές που είναι ολοκληρωμένες διαδικασίες, με στάσιμες σε τάση χρονοσειρές αλλά και με απλές αυτοσυσχετιζόμενες διαδικασίες. Η αιτία του προβλήματος εντοπίζεται στο γεγονός ότι οι ιδιότητες αυτοσυσχέτισης των χρησιμοποιούμενων χρονοσειρών επηρεάζουν τη δομή των σφαλμάτων όταν εφαρμόζεται η μέθοδος των ελαχίστων τετραγώνων.

Abstract

This thesis examines the phenomenon of spurious relations in time series. The term spurious relation means that two or more time series are found to be linearly correlated, according to the statistical methodology, while in fact they are not correlated. This phenomenon can occur either through ordinary least squares estimation of a linear regression model and it is called spurious regression or through the sample linear correlation coefficient and it is called spurious correlation. Spurious regression is characterized by high values of the determination coefficient, high absolute values of the t statistic for the test of the statistical significance of the model coefficients and by very low (near to zero) values of the Durbin – Watson statistic of the residuals for a first order autocorrelation problem. In the case of spurious correlation this phenomenon is characterized by high values of the sample linear correlation coefficient and by high absolute values of the associated t statistic for testing zero correlation. The high values of the t statistic is a very serious problem because it can lead to the wrong rejections of the null hypothesis of the coefficients statistical significance. Spurious relations can arise with integrated time series, trend stationary time series and also with simple autoregressive time series. The cause of the problem is the fact that autocorrelated time series affect the structure of the residuals when ordinary least squares method is applied.

Περιεχόμενα

Κατάλογος Πινάκων	xvii
Κατάλογος Διαγραμμάτων	xix
Κατάλογος Συνοτομογραφιών	xxi

ΚΕΦΑΛΑΙΟ 1

Η μεθοδολογία της παλινδρόμησης και η χρησιμοποίηση χρονοσειρών

1.1	Εισαγωγή	1
1.2	Οι χρονοσειρές ως στοχαστικές διαδικασίες	3
1.3	Βασικές στοχαστικές διαδικασίες στην ανάλυση χρονοσειρών	7
1.3.1	Η i.i.d. διαδικασία	7
1.3.2	Η διαδικασία λευκού θορύβου	7
1.3.3	Η διαδικασία Innovations	7
1.3.4	Η αυτοπαλίνδρομη διαδικασία p τάξης $AR(p)$	7
1.3.5	Η διαδικασία κινητού μέσου q τάξης $MA(q)$	11
1.3.6	Η αυτοπαλίνδρομη κινητού μέσου διαδικασία $ARMA(p, q)$	14
1.3.7	Η διαδικασία μερικού αθροίσματος	16
1.3.8	Ο τυχαίος περίπατος	17
1.3.9	Η διαδικασία Wiener	18
1.4	Υποδείγματα παραγωγής μη στάσιμων χρονοσειρών	19
1.5	Η μεθοδολογία των Box και Jenkins	22
1.6	Η αυτοπαλίνδρομη ολοκληρωμένη κινητού μέσου διαδικασία $ARIMA(p, d, q)$	24
1.7	Το πρόβλημα της αυτοσυσχέτισης των σφαλμάτων	25
1.7.1	Ο έλεγχος των Durbin και Watson	26
1.7.2	Ο έλεγχος για λευκό θόρυβο των Ljung και Box	28
1.8	Βασικές έννοιες ασυμπτωτικής θεωρίας	29
1.8.1	Σύγκλιση κατά πιθανότητα	29
1.8.2	Σύγκλιση κατά κατανομή	30
1.8.3	Σύγκλιση τυχαίων διανυσμάτων και πινάκων	30

1.8.4	Τάξεις σύγκλισης μιας ακολουθίας	31
1.8.5	Ασθενής σύγκλιση	32
1.8.6	Το κεντρικό οριακό θεώρημα	32
1.8.7	Το συναρτησιακό κεντρικό θεώρημα	33
1.8.8	Το θεώρημα της συνεχούς αντιστοίχισης	34
1.9	Ανακεφαλαίωση	34

ΚΕΦΑΛΑΙΟ 2

Το πρόβλημα της υποφαινομενικής παλινδρόμησης

2.1	Εισαγωγή	35
2.2	Υποφαινομενικές παλινδρομήσεις με ολοκληρωμένες διαδικασίες	36
2.3	Υποφαινομενικές παλινδρομήσεις με ολοκληρωμένες διαδικασίες με μη μηδενικό μέσο όρο	51
2.4	Υποφαινομενικές παλινδρομήσεις με στάσιμες αυτοσυσχετιζόμενες διαδικασίες	59
2.5	Ανακεφαλαίωση	64

ΚΕΦΑΛΑΙΟ 3

Άλλες μορφές υποφαινομενικής παλινδρόμησης

3.1	Εισαγωγή	65
3.2	Υποφαινομενικές παλινδρομήσεις λόγω λανθασμένης ταυτοποίησης των χρονοσειρών	66
3.3	Υποφαινομενικές παλινδρομήσεις με στάσιμες σε τάση διαδικασίες	73
3.4	Υποφαινομενικές παλινδρομήσεις λόγω κοινής σχέσης με τρίτη μεταβλητή	76
3.5	Μη γραμμικές υποφαινομενικές παλινδρομήσεις	79
3.5.1	Η επίδραση του τετραγωνικού μετασχηματισμού της ανεξάρτητης μεταβλητής	80
3.5.2	Λογαριθμικές υποφαινομενικές παλινδρομήσεις	83
3.5.3	Η συμπεριφορά των ελέγχων μη γραμμικότητας για ολοκληρωμένες διαδικασίες	85
3.6	Ανακεφαλαίωση	86

ΚΕΦΑΛΑΙΟ 4

Υποφαινομενικές συσχετίσεις

4.1	Εισαγωγή	89
4.2	Η συσχέτιση δύο απλών αρμονικών συναρτήσεων του χρόνου που διαφέρουν κατά ένα τέταρτο στη φάση	91
4.3	Τα συμπεράσματα του Yule και οι ολοκληρωμένες διαδικασίες	96
4.4	Η κατανομή των συσχετίσεων μεταξύ δύο χρονοσειρών και η εμφάνιση υποφαινομενικών συσχετίσεων	97
4.5	Ανακεφαλαίωση	111

ΚΕΦΑΛΑΙΟ 5

Η χρήση διαφορών στην εκτίμηση του υποδείγματος

5.1	Εισαγωγή	113
5.2	Οι συνέπειες των διαφορών στο υπόδειγμα της γραμμικής παλινδρόμησης	114
5.3	Η επιλογή ενός υποδείγματος ανάμεσα στα επίπεδα και στις διαφορές	116
5.4	Ο συντελεστής πολλαπλού προσδιορισμού για την σύγκριση με ένα υπόδειγμα σε διαφορές	120
5.5	Ανακεφαλαίωση	122

Παραρτήματα

		123
Π1.	Αποτελέσματα προσομοιώσεων	125
Π2.	Προγράμματα προσομοιώσεων	129

Βιβλιογραφία

145

Κατάλογος Πινάκων

2-1	Παλινδρόμηση δύο ανεξαρτήτων τυχαίων περιπάτων	39
2-2	Πολλαπλές παλινδρομήσεις με μία έως πέντε ανεξάρτητες μεταβλητές	40
2-3	Υποφαινομενικές παλινδρομήσεις για διάφορα μεγέθη δείγματος	42
2-4	Υποφαινομενικές παλινδρομήσεις μεταξύ δύο τυχαίων περιπάτων με μέσο όρο για διάφορα μεγέθη δείγματος, $g_Y = 400$ και $g_X = 200$	52
2-5	Υποφαινομενικές παλινδρομήσεις μεταξύ δύο τυχαίων περιπάτων με μέσο όρο για διάφορα μεγέθη δείγματος, $g_Y = 0$ και $g_X = 200$	54
2-6	Υποφαινομενικές παλινδρομήσεις μεταξύ δύο τυχαίων περιπάτων με μέσο όρο για διάφορα μεγέθη δείγματος, $g_Y = 400$ και $g_X = 0$	55
2-7	Υποφαινομενικές παλινδρομήσεις μεταξύ δύο ανεξαρτήτων AR(1) χρονοσειρών ($f_Y = f_X = f$)	60
2-8	Υποφαινομενικές παλινδρομήσεις μεταξύ δύο ανεξαρτήτων AR(1) χρονοσειρών ($f_Y \neq f_X$)	61
2-9	Υποφαινομενικές παλινδρομήσεις μεταξύ δύο ανεξαρτήτων MA χρονοσειρών	62
3-1	Υποφαινομενικές παλινδρομήσεις με στάσιμες σε τάση διαδικασίες, ποσοστό $ t > 1.96$	75
3-2	Υποφαινομενικές παλινδρομήσεις λόγω κοινής σχέσης με τρίτη μεταβλητή	77
3-3	Υποφαινομενικές παλινδρομήσεις υποδείγματος (3.9) για $\alpha = 5\%$	81
3-4	Υποφαινομενικές παλινδρομήσεις υποδείγματος (3.10) για $\alpha = 5\%$	82
3-5	Ποσοστά απόρριψης ελέγχων μη γραμμικότητας για $\alpha = 5\%$	86
4-1	Υποφαινομενικές συσχετίσεις	98

Κατάλογος Διαγραμμάτων

1-1	AR(1) διαδικασία	9
1-2	Αυτοσυσχετίσεις AR(1) διαδικασίας για $f = 0.8$	9
1-3	Μερικές αυτοσυσχετίσεις AR(1) διαδικασίας για $f = 0.8$	9
1-4	MA(1) διαδικασία	13
1-5	Αυτοσυσχετίσεις MA(1) διαδικασίας για $q = -0.8$	13
1-6	Μερικές αυτοσυσχετίσεις MA(1) διαδικασίας για $q = -0.8$	13
1-7	APMA (1, 1) διαδικασία	15
1-8	Αυτοσυσχετίσεις APMA (1, 1) διαδικασίας για $f = 0.8$ και $q = 0.5$	15
1-9	Μερικές αυτοσυσχετίσεις APMA (1, 1) διαδικασίας για $f = 0.8$ και $q = 0.5$	15
1-10	Στάσιμη σε τάση χρονοσειρά	21
1-11	Στάσιμη σε διαφορές χρονοσειρά	22
2-1	Κατανομή του \hat{b} για $g_y = 400$ και $g_x = 200$, $T = 100$	53
2-2	Κατανομή του \hat{b} για $g_y = 0$ και $g_x = 200$, $T = 100$	56
2-3	Κατανομή του \hat{b} για $g_y = 400$ και $g_x = 0$, $T = 100$	56
4-1	Δύο ημιτονοειδής καμπύλες που διαφέρουν κατά ένα τέταρτο στη φάση	94
4-2	Κατανομή των συσχετίσεων όταν $f_y = 0$ και $f_x = 0$, $T = 100$	99
4-3	Κατανομή της στατιστικής έλεγχου της συσχέτισης όταν $f_y = 0$ και $f_x = 0$, $T = 100$	100
4-4	Κατανομή των συσχετίσεων όταν $Y_t \sim I(1)$ και $X_t \sim I(1)$, $T = 100$	101
4-5	Κατανομή της στατιστικής έλεγχου της συσχέτισης όταν $Y_t \sim I(1)$ και $X_t \sim I(1)$, $T = 100$	101
4-6	Κατανομή των συσχετίσεων όταν $Y_t \sim I(2)$ και $X_t \sim I(2)$, $T = 100$	102

4-7	Κατανομή της στατιστικής έλεγχου της συσχέτισης όταν $Y_t \sim I(2)$ και $X_t \sim I(2)$, $T = 100$	102
4-8	Κατανομή των συσχετίσεων όταν $Y_t \sim I(2)$ και $X_t \sim I(1)$, $T = 100$	103
4-9	Κατανομή της στατιστικής έλεγχου της συσχέτισης όταν $Y_t \sim I(2)$ και $X_t \sim I(1)$, $T = 100$	104
4-10	Κατανομή των συσχετίσεων όταν $f_Y = 0.5$ και $f_X = 0.5$, $T = 100$	105
4-11	Κατανομή της στατιστικής έλεγχου της συσχέτισης όταν $f_Y = 0.5$ και $f_X = 0.5$, $T = 100$	105
4-12	Κατανομή των συσχετίσεων όταν $f_Y = 0.8$ και $f_X = 0.8$, $T = 100$	106
4-13	Κατανομή της στατιστικής έλεγχου της συσχέτισης όταν $f_Y = 0.8$ και $f_X = 0.8$, $T = 100$	107
4-14	Κατανομή των συσχετίσεων όταν $Y_t \sim I(1)$ και $X_t \sim I(1)$ με μέσους όρους 200, $T = 100$	108
4-15	Κατανομή της στατιστικής έλεγχου της συσχέτισης όταν $Y_t \sim I(1)$ και $X_t \sim I(1)$ με μέσους όρους 200, $T = 100$	108
4-16	Κατανομή των συσχετίσεων όταν $Y_t \sim I(1)$ και $X_t \sim I(1)$ με μέσους όρους 200 και -200, $T = 100$	109
4-17	Κατανομή της στατιστικής έλεγχου της συσχέτισης όταν $Y_t \sim I(1)$ και $X_t \sim I(1)$ με μέσους όρους 200 και -200, $T = 100$	110

Κατάλογος Συντομογραφιών

AIC	Akaike's Information Criterion
AR	Autoregressive Process
ARMA	Autoregressive Moving Average Process
ARIMA	Autoregressive Integrated Moving Average Process
BIC	Bayesian Information Criterion
CDF	Cumulative Distribution Function
DSP	Difference Stationary Process
<i>DW</i>	Durbin and Watson Statistic
GLS	Generalized Least Squares
<i>I</i>	Integrated Process
i.i.d.	Independently Identically Distributed
MA	Moving Average Process
SSE	Sum of Squares of Errors
SSR	Sum of Squares from Regression
TSP	Trend Stationary Process

Equation Chapter 1 Section 1

ΚΕΦΑΛΑΙΟ 1

Η ΜΕΘΟΔΟΛΟΓΙΑ ΤΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ ΚΑΙ Η ΧΡΗΣΙΜΟΠΟΙΗΣΗ ΧΡΟΝΟΣΕΙΡΩΝ

1.1 Εισαγωγή

Στη στατιστική θεωρία η ανάλυση παλινδρόμησης αποσκοπεί στη διερεύνηση της σχέσης ανάμεσα σε μια εξαρτημένη μεταβλητή Y και σε ένα σύνολο ανεξαρτήτων μεταβλητών X_1, X_2, \dots, X_k . Συνήθως, θεωρείται ότι η σχέση αυτή είναι γραμμική, οπότε ένα υπόδειγμα που θα μπορούσε να χρησιμοποιηθεί για μια τέτοια ανάλυση είναι το ακόλουθο:

$$Y_t = b_0 + b_1 X_{1t} + b_2 X_{2t} + \dots + b_k X_{kt} + e_t \quad \text{για } t = 1, 2, \dots, T \quad (1.1)$$

και το οποίο χρησιμοποιώντας μήτρες γράφεται ως εξής:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (1.2)$$

όπου \mathbf{Y} είναι το διάνυσμα των παρατηρήσεων της εξαρτημένης μεταβλητής, $\boldsymbol{\beta}$ το διάνυσμα των συντελεστών του υποδείγματος, \mathbf{X} η μήτρα των παρατηρήσεων των ανεξαρτήτων μεταβλητών έχοντας στη πρώτη στήλη μονάδες που χρειάζονται για την εκτίμηση του σταθερού όρου και $\boldsymbol{\varepsilon}$ είναι το διάνυσμα των τυχαίων σφαλμάτων που αντιπροσωπεύουν όλους τους παράγοντες που επηρεάζουν την Y και δε μπορούν να συμπεριληφθούν στις ανεξάρτητες μεταβλητές.

Επιπλέον, για την εκτίμηση του παραπάνω υποδείγματος χρησιμοποιούνται οι ακόλουθες υποθέσεις:

- 1) $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, δηλαδή τα σφάλματα έχουν μηδενικό μέσο όρο.
- 2) $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = S^2\mathbf{I}$, δηλαδή τα σφάλματα έχουν σταθερή διακύμανση και είναι ανεξάρτητα μεταξύ τους, όπου \mathbf{I} η μοναδιαία μήτρα.
- 3) $E(\mathbf{X}'\boldsymbol{\varepsilon}) = \mathbf{0}$, δηλαδή τα σφάλματα δεν συσχετίζονται με τις ανεξάρτητες μεταβλητές, η μήτρα \mathbf{X} δεν είναι στοχαστική.

- 4) $r(\mathbf{X}) = k \Rightarrow |\mathbf{X}'\mathbf{X}| \neq 0$, δηλαδή η μήτρα \mathbf{X} είναι πλήρους τάξεως, που σημαίνει ότι οι ανεξάρτητες μεταβλητές δε συσχετίζονται γραμμικά μεταξύ τους.
- 5) $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$, δηλαδή τα σφάλματα ακολουθούν κανονική κατανομή με μέσο όρο 0 και σταθερή διακύμανση.

Όταν παραβιάζεται η υπόθεση της σταθερής διακύμανσης των σφαλμάτων προκύπτει το πρόβλημα της ετεροσκεδαστικότητας, όταν παραβιάζεται η υπόθεση της ανεξαρτησίας των σφαλμάτων προκύπτει το πρόβλημα της αυτοσυσχέτισης, ενώ όταν υπάρχει γραμμική εξάρτηση των ανεξαρτητών μεταβλητών υπάρχει το πρόβλημα της πολυσυγγραμμικότητας.

Αν τηρούνται όλες οι προηγούμενες υποθέσεις, τότε η εκτίμηση του υποδείγματος (1.2) με τη μέθοδο των ελαχίστων τετραγώνων δίνει για τους συντελεστές του τους εκτιμητές $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$, οι οποίοι είναι γραμμικοί, αμερόληπτοι και άριστοι.

Η χρησιμοποίηση χρονοσειρών ως δείγμα καθιστά τη μεθοδολογία της ανάλυσης παλινδρόμησης περισσότερο περίπλοκη. Ειδικότερα, εκτός από τις παραπάνω υποθέσεις ο ερευνητής θα πρέπει να ελέγχει την ύπαρξη ή μη στασιμότητας των εμπλεκόμενων χρονοσειρών, καθώς και τις ιδιότητες αυτοσυσχέτισής τους. Ενώ στη χρησιμοποίηση διαστρωματικών στοιχείων, δηλαδή στοιχείων που αφορούν τη συμπεριφορά των μεταβλητών κατά τη διάρκεια μιας συγκεκριμένης χρονικής περιόδου, η σειρά παρουσίας των δεδομένων δεν επηρεάζει το αποτέλεσμα, στις χρονοσειρές τα δεδομένα είναι αυστηρά χρονικά ιεραρχημένα. Για τον ίδιο λόγο, ενώ στα διαστρωματικά δεδομένα η έλλειψη μίας η περισσότερων παρατηρήσεων από το δείγμα μπορεί να αγνοηθεί και να αντικατασταθεί με άλλη παρατήρηση, όταν χρησιμοποιούνται χρονοσειρές το πρόβλημα αυτό δεν είναι εύκολο να αντιμετωπιστεί. Τέλος, η εμπειρία έχει δείξει ότι όταν το δείγμα προέρχεται από χρονοσειρές, το πρόβλημα της αυτοσυσχέτισης έχει μεγάλη πιθανότητα να εμφανιστεί και ο ερευνητής θα πρέπει να είναι σε θέση να το αντιμετωπίσει.

Ο κύριος σκοπός του κεφαλαίου αυτού είναι να εξηγήσει έννοιες που αφορούν το χώρο των χρονοσειρών, καθώς και να αναλύσει περιληπτικά διάφορα θέματα που προκύπτουν στην ανάλυση παλινδρόμησης όταν το χρησιμοποιούμενο δείγμα προέρχεται από χρονοσειρές, θέτοντας τις βάσεις για τη κατανόηση των επόμενων κεφαλαίων.

1.2 Οι χρονοσειρές ως στοχαστικές διαδικασίες

Με τον όρο χρονοσειρά ονομάζεται μια ιεραρχημένη ακολουθία παρατηρήσεων που λαμβάνονται σε τακτά ίσα χρονικά διαστήματα. Ο μηχανισμός παραγωγής της χρονοσειράς μπορεί να θεωρηθεί ότι είναι κάποια στοχαστική διαδικασία, δηλαδή ένα σύνολο τυχαίων μεταβλητών $\{X_t, t \in T\}$ ορισμένων στον ίδιο χώρο πιθανότητας. Το σύνολο T είναι το σύνολο των τιμών που παίρνει ο δείκτης t της στοχαστικής διαδικασίας, ο οποίος συνήθως συμβολίζει το χρόνο. Ανάλογα με τις τιμές που παίρνει ο δείκτης t έχουμε στοχαστικές διαδικασίες διακριτού χρόνου και στοχαστικές διαδικασίες συνεχούς χρόνου. Για παράδειγμα, εάν $T = (0, 1, 2, \mathbf{K})$ η στοχαστική διαδικασία είναι διακριτού χρόνου, ενώ εάν $T = (-\infty, +\infty)$ είναι συνεχούς χρόνου.

Μια στοχαστική διαδικασία μπορεί να θεωρηθεί ότι είναι μια συνάρτηση δύο μεταβλητών, του δείκτη της t και ενός σημείου w του δειγματικού χώρου Ω . Κατά συνέπεια, η στοχαστική διαδικασία μπορεί να γραφεί ως $\{X(w, t), w \in \Omega, t \in T\}$. Για δεδομένη τιμή του t η στοχαστική διαδικασία γίνεται τυχαία μεταβλητή, ενώ για δεδομένη τιμή του w γίνεται συνάρτηση του t που ονομάζεται δειγματική πραγματοποίηση της στοχαστικής διαδικασίας και είναι το χαρακτηριστικό της στοχαστικής διαδικασίας που παρατηρείται και μπορεί να αξιοποιηθεί για ανάλυση. Η σχέση ανάμεσα στη στοχαστική διαδικασία και στην πραγματοποίησή της είναι παρόμοια με τη σχέση ανάμεσα στο πληθυσμό και στο δείγμα στην κλασική στατιστική θεωρία.

Σαν στοχαστική διαδικασία μια χρονοσειρά μπορεί να χαρακτηριστεί από τη συνάρτηση κατανομής της. Επειδή όμως η συνάρτηση κατανομής είναι ουσιαστικά αδύνατο να προσδιορισθεί, γι' αυτό ορίζουμε τη διαδικασία με βάση τις πρώτες και δεύτερες ροπές της, δηλαδή χρησιμοποιώντας για κάθε χρονική στιγμή τους μέσους όρους, $m_t = E(X_t)$, τις διακυμάνσεις, $s_t^2 = Var(X_t)$ και τις αυτοσυνδιακυμάνσεις, $g_{t,s} = Cov(X_t, X_s)$ για $t \neq s$. Εάν μπορούσε να εξασφαλισθεί ότι η στοχαστική διαδικασία είναι κανονική, οι προηγούμενες ροπές θα αρκούσαν για να είναι γνωστές πλήρως οι ιδιότητές της. Κάτι τέτοιο δύσκολα επιβεβαιώνεται στη πράξη. Θεωρώντας όμως ότι η διαδικασία είναι γραμμική, δηλαδή, ότι οι τιμές της παράγονται από γραμμικό συνδυασμό παρελθουσών τιμών της ίδιας της

διαδικασίας και παρούσων και παρελθουσών τιμών άλλων διαδικασιών, οι παραπάνω ροπές αρκούν για να είναι γνωστές οι κύριες ιδιότητές της. Το πρόβλημα όμως που προκύπτει είναι ότι οι ροπές δεν παραμένουν σταθερές αλλά αλλάζουν σε κάθε χρονική στιγμή. Έτσι, απαιτείται να εκτιμήσουμε T μέσους όρους, T διακυμάνσεις και $T(T-1)/2$ αυτοσυνδιακυμάνσεις έχοντας μόνο μια πραγματοποίηση της στοχαστικής διαδικασίας, δηλαδή μόνο μια παρατήρηση για κάθε χρονική περίοδο. Για το λόγο αυτό, πρέπει με κάποιο τρόπο να μειωθεί ο αριθμός των αγνώστων παραμέτρων. Σε αυτό συμβάλλουν οι υποθέσεις της εργοδικότητας (*Ergodic*) και της στασιμότητας (*Stationarity*).

Εργοδικότητα σημαίνει ότι οι ροπές από μία πραγματοποίηση της διαδικασίας συγκλίνουν στις πληθυσμιακές τιμές τους, καθώς το μήκος της πραγματοποίησης γίνεται άπειρο. Για παράδειγμα, για το μέσο όρο $\bar{X}_T = \frac{1}{T} \sum_{t=1}^T X_t$, τη χρονική στιγμή T , εφόσον $E(X_t) = m$ θα ισχύει $E(\bar{X}_T) \rightarrow m$ και συνεπώς $Var(\bar{X}_T) = 0$ καθώς $T \rightarrow \infty$, δηλαδή ο \bar{X}_T να είναι συνεπής και αμερόληπτος εκτιμητής του μέσου όρου της στοχαστικής διαδικασίας m . Δεν αρκεί όμως μόνο η ιδιότητα της συνέπειας. Θα πρέπει επιπλέον οι παρατηρήσεις της διαδικασίας που απέχουν χρονικά αρκετά μεταξύ τους να είναι ασυσχέτιστες, έτσι ώστε κάθε νέα παρατήρηση να προσθέτει νέα χρήσιμη πληροφορία στην εκτίμηση της ροπής. Απαραίτητη προϋπόθεση για να ισχύει η εργοδικότητα είναι $Cov(X_t, X_{t-T}) \rightarrow 0$ με ένα αρκετά γρήγορο ρυθμό, καθώς το T αυξάνει. Αν ισχύει η ιδιότητα της εργοδικότητας, τότε μπορούν να εκτιμηθούν οι άγνωστες ροπές χρησιμοποιώντας μία μόνο πραγματοποίηση της διαδικασίας.

Η δεύτερη αναγκαία υπόθεση είναι αυτή της στασιμότητας. Διακρίνουμε δύο βασικά είδη στασιμότητας, την ασθενή στασιμότητα ή στασιμότητα β τάξης (*Weak stationarity* ή *Second order stationarity*) και την αυστηρή στασιμότητα (*Strict stationarity*). Μια στοχαστική διαδικασία λέγεται ασθενώς στάσιμη όταν σε κάθε χρονική στιγμή έχει σταθερό μέσο όρο:

$$E(X_t) = m$$

σταθερή πεπερασμένη διακύμανση:

$$Var(X_t) = s_x^2 < \infty$$

και οι αυτοσυνδιακυμάνσεις της για δύο διαφορετικές χρονικές στιγμές εξαρτώνται μόνο από τη μεταξύ τους απόσταση, δηλαδή:

$$g_k = Cov(X_t, X_{t-k}) = Cov(X_t, X_{t+k}).$$

Χρησιμοποιώντας τις αυτοσυνδιακυμάνσεις διαπιστώνουμε ότι όταν η στοχαστική διαδικασία είναι ασθενώς στάσιμη θα πρέπει για τις αυτοσυσχετίσεις της:

$$r_k = \frac{Cov(X_t, X_{t-k})}{\sqrt{V(X_t)V(X_{t-k})}} = \frac{g_k}{g_0}$$

να ισχύει $r_k = r_{-k}$. Οι αυτοσυσχετίσεις έχουν μεγάλη σημασία όταν εξετάζονται υποδείγματα χρονοσειρών γιατί μετράνε το βαθμό όπου μια τιμή της χρονοσειράς συσχετίζεται με προηγούμενες τιμές της και αποτελούν κατά κάποιο τρόπο τη μνήμη της διαδικασίας.

Αυστηρά στάσιμη λέγεται η στοχαστική διαδικασία για την οποία η από κοινού κατανομή για τις χρονικές στιγμές $t_1, t_2, \mathbf{K}, t_n$ είναι ίδια με την από κοινού κατανομή για τις χρονικές στιγμές $t_{1+k}, t_{2+k}, \mathbf{K}, t_{n+k}$, όπου k θετικός ακέραιος. Δηλαδή ισχύει:

$$f(X_{t_1}, X_{t_2}, \mathbf{K}, X_{t_n}) = f(X_{t_1+k}, X_{t_2+k}, \mathbf{K}, X_{t_n+k})$$

Μια αυστηρά στάσιμη στοχαστική διαδικασία είναι υποχρεωτικά και ασθενώς στάσιμη. Το αντίθετο δεν ισχύει πάντα. Εξάιρεση αποτελεί η κανονική κατανομή, όπου η αυστηρή στασιμότητα ταυτίζεται με την ασθενή στασιμότητα. Η αυστηρή στασιμότητα είναι πολύ δύσκολο να επιβεβαιωθεί στην πράξη και για αυτό συνήθως χρησιμοποιείται η ασθενή στασιμότητα.

Η υπόθεση της εργοδικότητας που έχει αναφερθεί αποτελεί περιορισμό ασυμπτωτικής ανεξαρτησίας για μία στοχαστική διαδικασία. Μια στοχαστική διαδικασία $\{X_t\}$ ονομάζεται ασυμπτωτικά ανεξάρτητη εάν ισχύει:

$$|f(X_1, \mathbf{K}, X_n, X_{h+1}, \mathbf{K}, X_{h+n}) - f(X_1, \mathbf{K}, X_n)f(X_{h+1}, \mathbf{K}, X_{h+n})| \rightarrow 0$$

όταν $h \rightarrow \infty$, που σημαίνει ότι η από κοινού συνάρτηση κατανομής δύο υποακολουθιών της $\{X_t\}$ προσεγγίζει το γινόμενο των κατανομών της κάθε μίας υποακολουθίας, καθώς η απόσταση ανάμεσα στις δύο υποακολουθίες τείνει στο άπειρο (Banerjee, Dolado, Galbraith και Hendry (1993)).

Εκτός από την εργοδικότητα, μια άλλη μορφή ασυμπτωτικής ανεξαρτησίας που συναντιέται συχνά στην ανάλυση χρονοσειρών είναι οι διαδικασίες μίξης (*mixing processes*). Περισσότερες λεπτομέρειες σχετικά με την ασυμπτωτική ανεξαρτησία, την εργοδικότητα και τις διαδικασίες μίξης μπορούν να βρεθούν στους Spanos (1986) και Banerjee, Dolado, Galbraith, και Hendry (1993). Στη συνέχεια θα εξηγηθούν συνοπτικά οι δύο βασικοί τύποι

μίξης που είναι η ομοιόμορφη μίξη (*uniform mixing*) και η ισχυρή μίξη (*strong mixing*).

Η ομοιόμορφη μίξη που ονομάζεται και Φ – μίξη (Φ – *mixing*) βασίζεται στην συνθήκη ότι για δύο ανεξάρτητα ενδεχόμενα A και B ισχύει:

$$P(A|B) - P(A) = 0$$

και η ισχυρή μίξη που ονομάζεται και α – μίξη (α – *mixing*) στην συνθήκη ότι

$$P(A \mathbf{I} B) - P(A)P(B) = 0$$

Κατά συνέπεια, η έννοια της μίξης απαιτεί ότι οι σχέσεις που θα περιέχουν τις προηγούμενες ποσότητες θα ισούνται ασυμπτωτικά με το μηδέν.

Έστω οι δεσμευμένες αντιστοιχίσεις (*bounded mappings*) $(G_1(X_t, \mathbf{K}, X_{t+h}))$ και $G_2(X_t, \mathbf{K}, X_{t+k})$ στον άξονα των πραγματικών αριθμών. Η ακολουθία $\{X_t\}$ ονομάζεται ομοιόμορφη μίξη εάν υπάρχει μια ακολουθία $\{\Phi_n\}$ με $\Phi_n > 0 \forall n$ και $\lim_{n \rightarrow \infty} \Phi_n = 0$ τέτοια ώστε για $n \gg h$ να ισχύει:

$$|E[G_1(X_t, \mathbf{K}, X_{t+h})|G_2(X_{t+n}, \mathbf{K}, X_{t+n+k})] - E[G_1(X_t, \mathbf{K}, X_{t+h})]| < \Phi_n$$

Με την ίδια λογική και χρησιμοποιώντας την ιδιότητα της τομής ανεξαρτήτων ενδεχομένων η ακολουθία $\{X_t\}$ θα ονομάζεται ισχυρή μίξη εάν υπάρχει μια ακολουθία $\{a_n\}$ με $a_n > 0 \forall n$ και $\lim_{n \rightarrow \infty} a_n = 0$ τέτοια ώστε:

$$|E[G_1(X_t, \mathbf{K}, X_{t+h})G_2(X_{t+n}, \mathbf{K}, X_{t+n+k})] - E[G_1(X_t, \mathbf{K}, X_{t+h})]E[G_2(X_{t+n}, \mathbf{K}, X_{t+n+k})]| < a_n$$

Σημειώνεται ότι οι συνθήκες μίξης είναι μορφές ασυμπτωτικής ανεξαρτησίας ισχυρότερες από την εργοδικότητα. Για το λόγο αυτό, αν σε μια στάσιμη στοχαστική διαδικασία ισχύει κάποια από τις δύο προηγούμενες μίξεις συμπεραίνεται ότι θα είναι και εργοδική. Επίσης, η ομοιόμορφη μίξη είναι πιο δυνατή από την ισχυρή μίξη, δηλαδή αν ισχύει η πρώτη τότε συνεπάγεται και η δεύτερη.

1.3 Βασικές στοχαστικές διαδικασίες στην ανάλυση χρονοσειρών

Στη συνέχεια θα αναφερθούν μερικές βασικές στοχαστικές διαδικασίες που συναντιούνται συχνά στην ανάλυση χρονοσειρών.

1.3.1 Η i.i.d. διαδικασία

Μια στοχαστική διαδικασία $\{X_t\}_1^\infty$ ονομάζεται ανεξάρτητα, ομοιόμορφα, κατανεμημένη (*Independently Identically Distributed*) εάν αποτελείται από μια ακολουθία ανεξαρτήτων ταυτόνομα κατανεμημένων μεταβλητών.

1.3.2 Η διαδικασία λευκού θορύβου

Μια στοχαστική διαδικασία $\{X_t\}_1^\infty$ ονομάζεται λευκός θόρυβος (*white noise process*) εάν:

$$E(X_t) = 0$$

και

$$\text{Cov}(X_t, X_s) = \begin{cases} \sigma^2 & \text{για } t = s \\ 0 & \text{για } t \neq s \end{cases}$$

Ο λευκός θόρυβος χαρακτηρίζεται από στασιμότητα δεύτερης τάξης. Αν επιπλέον μπορούσε να εξασφαλιστεί κανονικότητα, τότε είναι και αυστηρά στάσιμη διαδικασία.

1.3.3 Η διαδικασία Innovations

Μια στοχαστική διαδικασία $\{X_t\}_1^\infty$ λέγεται διαδικασία innovation εάν:

$$E(X_t | X_{t-1}, X_{t-2}, \mathbf{K}) = 0$$

και

$$\text{Cov}(X_t, X_s | X_{t-1}, X_{t-2}, \mathbf{K}) = \begin{cases} \sigma^2 & \text{για } t = s \\ 0 & \text{για } t \neq s \end{cases}.$$

1.3.4 Η αυτοπαλίνδρομη διαδικασία p τάξης $\text{AR}(p)$

Έστω e_t μια διαδικασία λευκού θορύβου με μηδενικό μέσο όρο και σταθερή διακύμανση σ^2 . Η διαδικασία που παράγεται από το υπόδειγμα:

$$X_t = f_1 X_{t-1} + f_2 X_{t-2} + \mathbf{K} + f_p X_{t-p} + e_t \quad (1.3)$$

ονομάζεται αυτοπαλίνδρομη διαδικασία (*Autoregressive Process*) p τάξης και συμβολίζεται με $AR(p)$. Χρησιμοποιώντας τον τελεστή χρονικής υστέρησης B (*Backshift Operator*), ο οποίος ορίζεται ως $B^k X_t = X_{t-k}$, η σχέση (1.3) μπορεί να γραφτεί ισοδύναμα ως εξής:

$$X_t - f_1 X_{t-1} - f_2 X_{t-2} - \mathbf{K} - f_p X_{t-p} = e_t$$

$$(1 - f_1 B - f_2 B^2 - \mathbf{K} - f_p B^p) X_t = e_t$$

$$\Phi(B) X_t = e_t$$

και τελικά:

$$X_t = \frac{1}{1 - (f_1 B + f_2 B^2 + \mathbf{K} + f_p B^p)} e_t$$

$$X_t = \left[1 + (f_1 B + f_2 B^2 + \mathbf{K} + f_p B^p) + (f_1 B + f_2 B^2 + \mathbf{K} + f_p B^p)^2 + \mathbf{K} \right] e_t$$

$$X_t = \Psi(B) e_t$$

δηλαδή μακροχρόνια οι τιμές της X_t εξαρτώνται αποκλειστικά από τις τιμές του λευκού θορύβου.

Για να συγκλίνει η παραπάνω σχέση και να υπάρχει στασιμότητα θα πρέπει οι ρίζες του πολυωνύμου:

$$(1 - f_1 B - f_2 B^2 - \mathbf{K} - f_p B^p) = (1 - g_1 B)(1 - g_2 B) \mathbf{K} (1 - g_p B) = 0$$

να βρίσκονται εκτός του μοναδιαίου κύκλου ή διαφορετικά οι ρίζες του πολυωνύμου:

$$\frac{1}{B^p} - f_1 \frac{1}{B^{p-1}} - f_2 \frac{1}{B^{p-2}} - \mathbf{K} - f_p = Z^p - f_1 Z^{p-1} - f_2 Z^{p-2} - \mathbf{K} - f_p = 0 \text{ όπου } Z = \frac{1}{B}$$

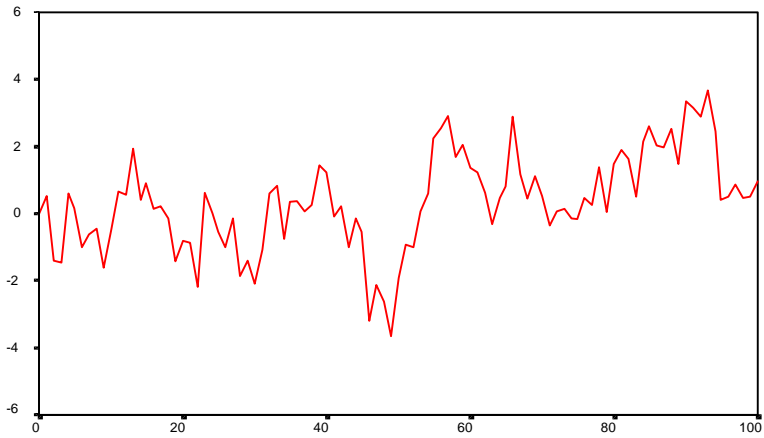
να είναι μικρότερες της μονάδας, δηλαδή να βρίσκονται εντός του μοναδιαίου κύκλου. Αναγκαία προϋπόθεση είναι πάντα $|f_p| < 1$. Όταν τηρούνται οι παραπάνω περιορισμοί η αυτοπαλίνδρομη διαδικασία είναι στάσιμη και μπορεί να εκφραστεί μακροχρόνια από τις τιμές του λευκού θορύβου.

Αποδεικνύεται ότι η αυτοπαλίνδρομη διαδικασία p τάξης έχει διακύμανση που δίνεται από τη σχέση:

$$g_0 = f_1 g_1 + f_2 g_2 + \mathbf{K} + f_p g_p + s^2$$

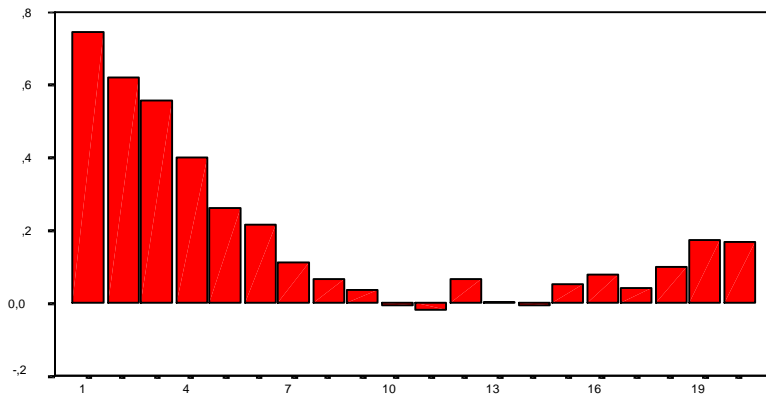
ενώ για τις αυτοσυνδιακυμάνσεις της και τις αυτοσυσχετίσεις της ισχύουν αντίστοιχα:

$$g_k = f_1 g_{k-1} + f_2 g_{k-2} + \mathbf{K} + f_p g_{k-p}$$

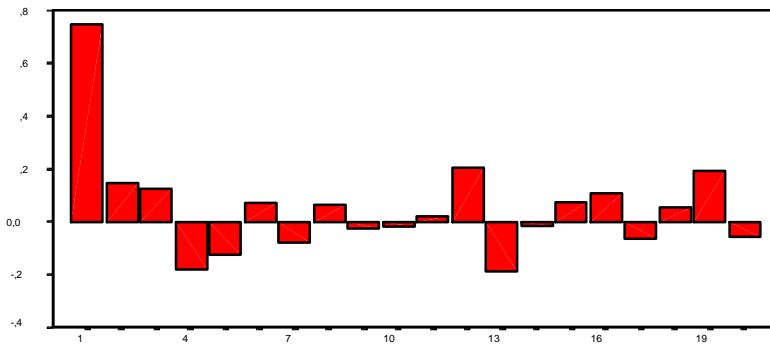


$f = 0.8$, $X_0 = 0$, $t = 100$, $e_t \sim N(0, 1)$, $\mu = 0$

Διάγραμμα 1-1
AR(1) διαδικασία



Διάγραμμα 1-2
Αυτοσυσχετίσεις AR(1) διαδικασίας για $f = 0.8$



Διάγραμμα 1-3
Μερικές αυτοσυσχετίσεις AR(1) διαδικασίας για $f = 0.8$

και

$$r_k = f_1 r_{k-1} + f_2 r_{k-2} + \mathbf{K} + f_p r_{k-p}, \quad r_0 = 0.$$

Ειδικότερα, για την αυτοπαλίνδρομη διαδικασία πρώτης τάξης ισχύουν

$$g_k = f_1^k g_0 \quad \text{με} \quad g_0 = \frac{s^2}{1-f_1^2} \quad \text{και} \quad r_k = f_1^k, \quad \forall k \geq 0.$$

Στην εύρεση της σωστής τάξης μιας αυτοπαλίνδρομης διαδικασίας εκτός από τις αυτοσυσχετίσεις, σημαντική βοήθεια προσφέρουν και οι μερικές αυτοσυσχετίσεις. Η μερική αυτοσυσχέτιση f_{kk} ανάμεσα στις X_t και X_{t-k} μετράει τη συσχέτιση μεταξύ των τιμών της χρονοσειράς, αφού έχει αφαιρεθεί η ταυτόχρονη γραμμική επιρροή όλων των X_2, \mathbf{K}, X_{k-1} , ενδιάμεσων τιμών της χρονοσειράς στις τιμές X_t και X_{t-k} . Η k μερική αυτοσυσχέτιση είναι ο συντελεστής f_{kk} στην εξίσωση:

$$X_t = f_{k1} X_{t-1} + f_{k2} X_{t-2} + \mathbf{K} + f_{kk} X_{t-k} + e_t$$

και δίνεται από τον τύπο:

$$f_{kk} = \frac{r_k - \sum_{j=1}^{k-1} f_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} f_{k-1,j} r_j} \quad \text{με} \quad f_{k,j} = f_{k-1,j} - f_{kk} f_{k-1,k-j}.$$

Η αυτοπαλίνδρομη διαδικασία p τάξης έχει αυτοσυσχετίσεις άπειρες στον αριθμό που φθίνουν στο μηδέν εκθετικά αν είναι όλες θετικές ή με ημιτονοειδή μορφή αν παίρνουν και αρνητικές τιμές και μερικές αυτοσυσχετίσεις που μηδενίζονται για βαθμό χρονικών υστερήσεων μεγαλύτερο από p .

Στο αρχικό υπόδειγμα (1.3) η χρονοσειρά έχει μέσο όρο μηδέν, $m=0$. Αν όμως η χρονοσειρά έχει μη μηδενικό μέσο όρο, $m \neq 0$, τότε το υπόδειγμα τροποποιείται και γράφεται σε αποκλίσεις από το μέσο όρο, ως εξής:

$$X_t - m = f_1 (X_{t-1} - m) + f_2 (X_{t-2} - m) + \mathbf{K} + f_p (X_{t-p} - m) + e_t$$

$$X_t = (1 - f_1 - f_2 - \mathbf{K} - f_p) m + f_1 X_{t-1} + f_2 X_{t-2} + \mathbf{K} + f_p X_{t-p} + e_t$$

$$\Phi(B)(X_t - m) = e_t.$$

Στα Διαγράμματα 1-1, 1-2 και 1-3 απεικονίζονται μια AR(1) διαδικασία με τις αυτοσυσχετίσεις της και τις μερικές αυτοσυσχετίσεις της για $f = 0.8$.

1.3.5 Η διαδικασία κινητού μέσου q τάξης $MA(q)$

Έστω e_t μια διαδικασία λευκού θορύβου με μέσο όρο μηδέν και σταθερή διακύμανση σ^2 .

Η διαδικασία που παράγεται από το υπόδειγμα:

$$X_t = e_t - q_1 e_{t-1} - q_2 e_{t-2} - \mathbf{K} - q_q e_{t-q} \quad (1.4)$$

ονομάζεται διαδικασία κινητού μέσου (*Moving Average Process*) q τάξης και συμβολίζεται με $MA(q)$. Όπως και στην περίπτωση της αυτοπαλίνδρομης διαδικασίας, το υπόδειγμα στην αρχική του μορφή εκφράζει τη βραχυχρόνια συμπεριφορά μιας χρονοσειράς. Για να εξεταστεί η μακροχρόνια συμπεριφορά, το υπόδειγμα μετασχηματίζεται με τη βοήθεια του τελεστή χρονικής υστέρησης και γράφεται ως εξής:

$$X_t = (1 - q_1 B - q_2 B^2 - \mathbf{K} - q_q B^q) e_t$$

$$X_t = \Theta(B) e_t$$

$$e_t = \frac{1}{1 - (q_1 B + q_2 B^2 + \mathbf{K} + q_q B^q)} X_t$$

Τα υποδείγματα κινητού μέσου είναι όλα στάσιμα αφού αποτελούνται από πεπερασμένο αριθμό όρων λευκού θορύβου που είναι πάντα στάσιμος. Για να συγκλίνει η παραπάνω σχέση θα πρέπει οι ρίζες του πολυωνύμου:

$$(1 - q_1 B - q_2 B^2 - \mathbf{K} - q_q B^q) = (1 - h_1 B)(1 - h_2 B) \mathbf{K} (1 - h_q B) = 0$$

να βρίσκονται εκτός του μοναδιαίου κύκλου ή διαφορετικά οι ρίζες του πολυωνύμου:

$$\frac{1}{B^q} - q_1 \frac{1}{B^{q-1}} - q_2 \frac{1}{B^{q-2}} - \mathbf{K} - q_q = Z^q - q_1 Z^{q-1} - q_2 Z^{q-2} - \mathbf{K} - q_q = 0 \text{ όπου } Z = \frac{1}{B}$$

να είναι μικρότερες της μονάδας, δηλαδή να βρίσκονται εντός του μοναδιαίου κύκλου.

Αναγκαία προϋπόθεση είναι πάντα $|q_q| < 1$. Στη περίπτωση που ισχύουν οι προηγούμενοι περιορισμοί η διαδικασία κινητού μέσου ονομάζεται αντιστρέψιμη (*Invertible*) και γράφεται ισοδύναμα ως:

$$e_t = \left[1 + (q_1 B + q_2 B^2 + \mathbf{K} + q_q B^q) + (q_1 B + q_2 B^2 + \mathbf{K} + q_q B^q)^2 + \mathbf{K} \right] X_t$$

$$e_t = \Pi(B) X_t .$$

Επομένως όταν έχει εξασφαλιστεί η αντιστρεψιμότητα η διαδικασία κινητού μέσου q τάξης μπορεί να παρουσιαστεί μακροχρόνια από μια αυτοπαλίνδρομη διαδικασία απείρων όρων, $AR(\infty)$. Επιπροσθέτως, η αντιστρεψιμότητα οδηγεί στην επιλογή του υποδείγματος με τη

μικρότερη διακύμανση και δεν επιτρέπει την ύπαρξη ίδιων υποδειγμάτων για διαφορετικές τιμές του q .

Η διακύμανση μια διαδικασίας κινητού μέσου q τάξης δίνεται από τη σχέση:

$$g_0 = s^2 (1 + q_1^2 + q_2^2 + \mathbf{K} + q_q^2)$$

Οι αυτοσυνδιακυμάνσεις της μπορούν εύκολα να υπολογιστούν με βάση τον τύπο:

$$g_k = \begin{cases} s^2 \sum_{i=0}^{q-k} q_i q_{i+k} & \text{για } q_0 = -1 \text{ για } k = 1, 2, \mathbf{K}, q \\ 0 & \text{για } k > q \end{cases}$$

Άμεση συνέπεια των προηγούμενων σχέσεων είναι ότι για τις αυτοσυσχετίσεις ισχύουν:

$$r_k = \begin{cases} \frac{\sum_{i=0}^{q-k} q_i q_{i+k}}{1 + q_1^2 + q_2^2 + \mathbf{K} + q_q^2} & \text{για } q_0 = -1 \text{ για } k = 1, 2, \mathbf{K}, q \\ 0 & \text{για } k > q \end{cases}$$

Ειδικότερα για τη διαδικασία κινητού μέσου πρώτης τάξης ισχύουν:

$$g_0 = s^2 (1 + q_1^2)$$

$$g_k = \begin{cases} -s^2 q_1 & \text{για } k = 1 \\ 0 & \text{για } k > 1 \end{cases}$$

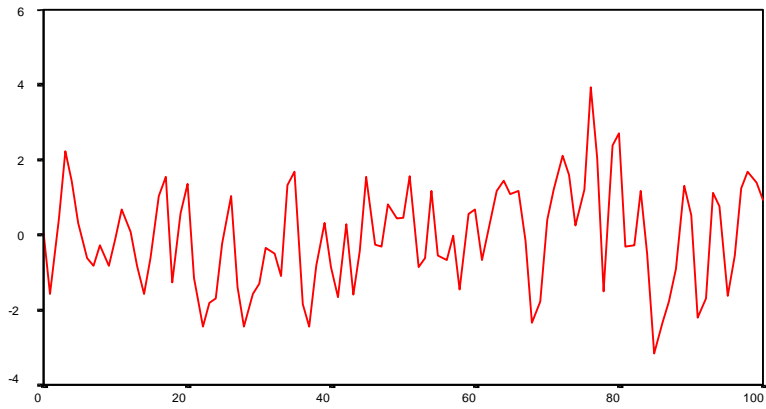
$$r_k = \begin{cases} \frac{-q_1}{1 + q_1^2} & \text{για } k = 1 \\ 0 & \text{για } k > 1 \end{cases}$$

Οι αυτοσυσχετίσεις και μερικές αυτοσυσχετίσεις μίας διαδικασίας κινητού μέσου έχουν αντίθετη συμπεριφορά από τις αυτοσυσχετίσεις και μερικές αυτοσυσχετίσεις μίας αυτοπαλίνδρομης διαδικασίας. Δηλαδή, μια διαδικασία κινητού μέσου έχει μερικές αυτοσυσχετίσεις άπειρες στο αριθμό που φθίνουν εκθετικά η ημιτονοειδώς και αυτοσυσχετίσεις που μηδενίζονται για τιμές του k μεγαλύτερες από q .

Όπως και στη περίπτωση της αυτοπαλίνδρομης διαδικασίας το αρχικό υπόδειγμα (1.4) μπορεί να τροποποιηθεί εάν η χρονοσειρά έχει μη μηδενικό μέσο όρο, $m \neq 0$, και να γραφτεί, σε αποκλίσεις από το μέσο όρο, ως εξής:

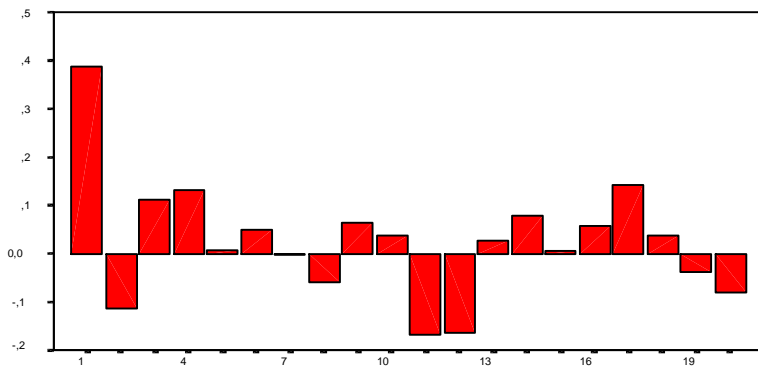
$$X_t - m = e_t - q_1 e_{t-1} - q_2 e_{t-2} - \mathbf{K} - q_q e_{t-q}$$

$$X_t - m = \Theta(B)e_t$$

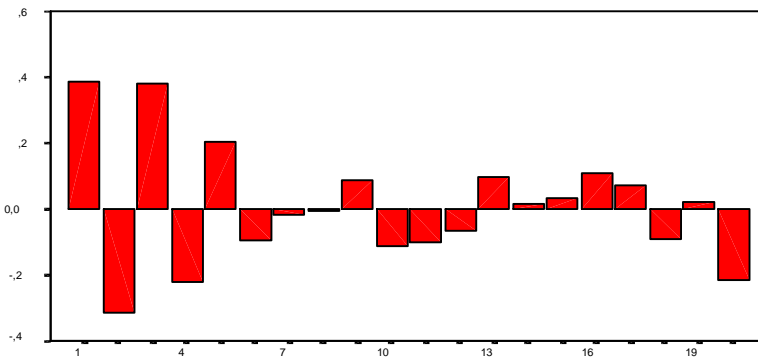


$q = -0.8$, $t = 100$, $e_t \sim N(0, 1)$, $\mu = 0$

Διάγραμμα 1-4
MA(1) διαδικασία



Διάγραμμα 1-5
Αυτοσυσχετίσεις MA(1) διαδικασίας για $q = -0.8$



Διάγραμμα 1-6
Μερικές αυτοσυσχετίσεις MA(1) διαδικασίας για $q = -0.8$

Η ουσιαστική διαφορά που υπάρχει από την αυτοπαλίνδρομη διαδικασία είναι ότι στη περίπτωση αυτή ο μέσος όρος της χρονοσειράς ταυτίζεται με το σταθερό όρο του υποδείγματος κάτι που δεν ισχύει στην πρώτη. Στα Διαγράμματα 1-4, 1-5 και 1-6 απεικονίζονται μια MA(1) διαδικασία με τις αυτοσυσχετίσεις της και τις μερικές αυτοσυσχετίσεις της για $q = -0.8$.

1.3.6 Η αυτοπαλίνδρομη κινητού μέσου διαδικασία ARMA(p, q)

Θεωρώντας μια διαδικασία λευκού θορύβου e_t με μηδενικό μέσο όρο και σταθερή διακύμανση σ^2 και συνδυάζοντας p αυτοπαλίνδρομους όρους και q όρους κινητού μέσου προκύπτει μια αυτοπαλίνδρομη κινητού μέσου διαδικασία ARMA(p, q):

$$X_t = f_1 X_{t-1} + f_2 X_{t-2} + \mathbf{K} + f_p X_{t-p} + e_t - q_1 e_{t-1} - q_2 e_{t-2} - \mathbf{K} - q_q e_{t-q} \quad (1.5)$$

η οποία μπορεί να γραφτεί ισοδύναμα ως:

$$(1 - f_1 B - f_2 B^2 - \mathbf{K} - f_p B^p) X_t = (1 - q_1 B - q_2 B^2 - \mathbf{K} - q_q B^q) e_t$$

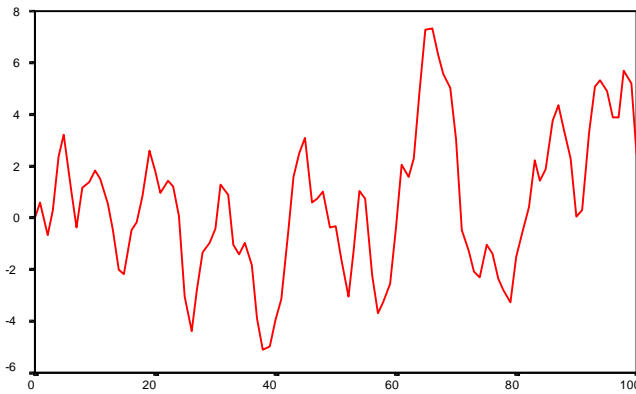
και τελικά

$$\Phi(B) X_t = \Theta(B) e_t$$

Για το αυτοπαλίνδρομο μέρος και το μέρος που περιλαμβάνει τους όρους κινητού μέσου ισχύουν αντίστοιχα οι συνθήκες στασιμότητας και αντιστρεψιμότητας που έχουν αναφερθεί. Εφόσον ισχύουν αυτές οι συνθήκες, η ARMA(p, q) διαδικασία μπορεί να γραφτεί μακροχρόνια ως:

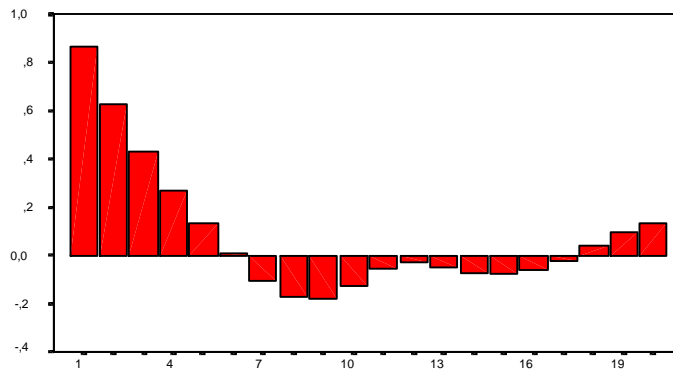
$$X_t = \frac{\Theta(B)}{\Phi(B)} e_t \quad \text{ή} \quad e_t = \frac{\Phi(B)}{\Theta(B)} X_t$$

Οι αυτοσυσχετίσεις και μερικές αυτοσυσχετίσεις μιας ARMA(p, q) διαδικασίας είναι περισσότερο περίπλοκες. Οι πρώτες q αυτοσυσχετίσεις επηρεάζονται τόσο από τους αυτοπαλίνδρομους όρους όσο και από τους όρους του κινητού μέσου με αποτέλεσμα να περιλαμβάνουν συνδυασμό τιμών των f και q και να μην είναι εύκολη η εύρεση κάποιου αναδρομικού τύπου. Όμως, μετά το βαθμό του κινητού μέσου οι αυτοσυσχετίσεις δεν περιλαμβάνουν καθόλου τιμές των q και συμπεριφέρονται ακριβώς όπως ένα AR(p) υπόδειγμα.

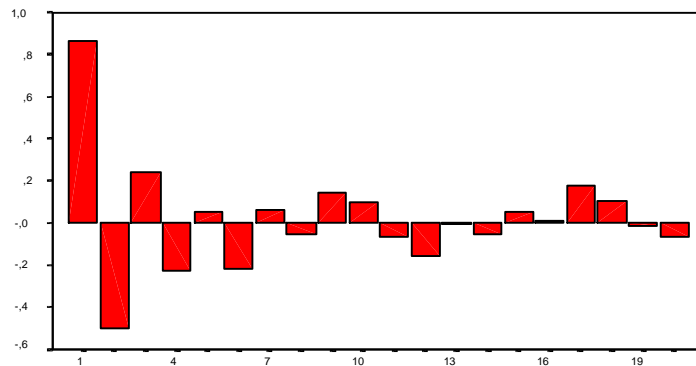


$f = 0.8$, $q = 0.5$, $X_0 = 0$, $t = 100$, $e_t \sim N(0, 1)$, $\mu = 0$

Διάγραμμα 1-7
ARMA(1, 1) διαδικασία



Διάγραμμα 1-8
Αυτοσυσχετίσεις ARMA(1, 1) διαδικασίας για $f = 0.8$ και $q = 0.5$



Διάγραμμα 1-9
Μερικές αυτοσυσχετίσεις ARMA(1, 1) διαδικασίας για $f = 0.8$ και $q = 0.5$

Επομένως, για $k \geq q+1$ οι αυτοσυνδιακυμάνσεις και αυτοσυσχετίσεις μιας ARMA διαδικασίας μπορούν να υπολογιστούν με βάση τις σχέσεις:

$$\mathbf{g}_k = f_1 \mathbf{g}_{k-1} + f_2 \mathbf{g}_{k-2} + \mathbf{K} + f_p \mathbf{g}_{k-p}$$

και

$$\mathbf{r}_k = f_1 \mathbf{r}_{k-1} + f_2 \mathbf{r}_{k-2} + \mathbf{K} + f_p \mathbf{r}_{k-p}$$

ενώ για τις μερικές αυτοσυσχετίσεις ισχύει ότι για $k > p - q$ συμπεριφέρονται όπως μια MA(q) διαδικασία.

Τέλος, εάν η χρονοσειρά έχει μέσο όρο $m \neq 0$, τότε το αρχικό υπόδειγμα (1.5), όπως και στις προηγούμενες περιπτώσεις, γράφεται σε αποκλίσεις από το μέσο όρο, δηλαδή:

$$\begin{aligned} X_t - m &= f_1 (X_{t-1} - m) + f_2 (X_{t-2} - m) + \mathbf{K} + f_p (X_{t-p} - m) + e_t - q_1 e_{t-1} - q_2 e_{t-2} - \mathbf{K} - q_q e_{t-q} \\ X_t &= (1 - f_1 - f_2 - \mathbf{K} - f_p) m + f_1 X_{t-1} + f_2 X_{t-2} + \mathbf{K} + f_p X_{t-p} + e_t - q_1 e_{t-1} - q_2 e_{t-2} - \mathbf{K} - q_q e_{t-q} \\ \Phi(\mathbf{B})(X_t - m) &= \Theta(\mathbf{B})e_t \end{aligned}$$

Στα Διαγράμματα 1-7, 1-8 και 1-9 απεικονίζονται μια ARMA(1, 1) διαδικασία με τις αυτοσυσχετίσεις της και τις μερικές αυτοσυσχετίσεις της για $f = 0.8$ και $q = 0.5$.

1.3.7 Η διαδικασία μερικού αθροίσματος

Έστω $\{X_{ii}\}_{i=1}^{\infty}$ μια ακολουθία από i.i.d. τυχαίες μεταβλητές με μέσο όρο m και διακύμανση s^2 . Η διαδικασία $\{S_k\}_{k=1}^{\infty}$ με $S_k = \sum_{i=1}^k X_{ii}$, η οποία αποτελείται από τα διαδοχικά αθροίσματα της X_t για $0 < t_1 < t_2 < \mathbf{K} < t_k < \mathbf{K} < \infty$ με $0 < t_k < \infty$ και $k = 1, 2, 3, \mathbf{K}$ ονομάζεται διαδικασία μερικού αθροίσματος.

Για το μέσο όρο, τη διακύμανση και την αυτοσυνδιακύμανση μια τέτοιας διαδικασίας ισχύουν αντίστοιχα, Spanos (1986):

$$E(S_k) = km \quad \text{για } k = 1, 2, 3, \mathbf{K}$$

$$\text{Var}(S_k) = ks^2 \quad \text{για } k = 1, 2, 3, \mathbf{K}$$

$$\text{Cov}(S_k, S_m) = \min(k, m)s^2 \quad \text{για } k, m = 1, 2, 3, \mathbf{K}$$

Δηλαδή δεν είναι στάσιμη διαδικασία αλλά οι ροπές της μεταβάλλονται με το k .

Μια διαδικασία μερικού αθροίσματος αποτελείται από τιμές που είναι όλες αμοιβαία εξαρτημένες ανεξάρτητα από τη μεταξύ τους απόσταση. Όμως, η βασική ιδιότητα μιας

διαδικασίας μερικού αθροίσματος είναι η Μαρκοβιανή εξάρτηση που χαρακτηρίζει την υπό συνθήκη κατανομή της. Πιο συγκεκριμένα, η υπό συνθήκη κατανομή της S_k όταν είναι γνωστές όλες οι παρελθούσες τιμές της, $S_{k-1}, S_{k-2}, S_{k-3}, \mathbf{K}, S_1$ εξαρτάται μόνο από την πιο πρόσφατη γνωστή τιμή S_{k-1} δηλαδή:

$$f(S_k | S_{k-1}, S_{k-2}, S_{k-3}, \mathbf{K}, S_1) = f(S_k | S_{k-1}).$$

Επομένως, το είδος της εξάρτησης ανάμεσα στη S_k και στις παρελθούσες τιμές της, περιγράφεται πλήρως από την υπό συνθήκη κατανομή της, γνωρίζοντας την πιο πρόσφατη τιμή της S_{k-1} .

1.3.8 Ο τυχαίος περίπατος

Στην προηγούμενη ενότητα αναφέρθηκε η διαδικασία μερικού αθροίσματος. Μια διαδικασία $\{S_t\}_{t=1}^{\infty}$, η οποία αποτελείται από τα μερικά αθροίσματα κάποιων i.i.d. τυχαίων μεταβλητών $\{X_t\}_1^{\infty}$ ονομάζεται και τυχαίος περίπατος (*random walk*).

Η διαδικασία μερικού αθροίσματος $S_t = \sum_{i=1}^t X_i$ μπορεί να γραφτεί ισοδύναμα:

$$S_t = X_0 + X_1 + X_2 + \mathbf{K} + X_{t-1} + X_t \text{ με } X_0 = 0$$

και τελικά:

$$S_t = S_{t-1} + X_t.$$

Δηλαδή, η τιμή της διαδικασίας $\{S_t\}$ τη χρονική στιγμή t ισούται με την τιμή της τη χρονική στιγμή $t-1$ συν την τιμή της i.i.d. τυχαίας μεταβλητής X_t τη χρονική στιγμή t .

Θεωρητικά, ο τυχαίος περίπατος θα μπορούσε να περιγράψει την κίνηση ενός σφαιριδίου το οποίο κινείται με σταθερό βήμα είτε δεξιά είτε αριστερά. Στη περίπτωση αυτή η τυχαία μεταβλητή X_t ακολουθεί διωνυμική κατανομή και συμβολίζει την κίνηση του σφαιριδίου κατά τη χρονική στιγμή t . Επομένως, η $\{S_t\}$ δείχνει τη θέση που βρίσκεται το σφαιρίδιο τη χρονική στιγμή t και αποτελείται από το άθροισμα όλων των κινήσεων του $X_0, X_1, X_2, \mathbf{K}, X_t$ κατά τις χρονικές στιγμές $0, 1, 2, \mathbf{K}, t$ αντίστοιχα. Έτσι, η θέση S_t του σφαιριδίου τη χρονική στιγμή t είναι η θέση S_{t-1} που κατείχε τη χρονική στιγμή $t-1$ συν

την τυχαία κίνηση X_t είτε δεξιά είτε αριστερά που θα πραγματοποιήσει κατά τη χρονική στιγμή t .

Έστω τώρα ότι μια χρονοσειρά ακολουθεί την αυτοπαλίνδρομη διαδικασία πρώτης τάξης $X_t = f_1 X_{t-1} + e_t$, όπου e_t είναι λευκός θόρυβος με μέσο όρο μηδέν και διακύμανση s^2 . Για να είναι στάσιμο το υπόδειγμα αυτό θα πρέπει $-1 < f_1 < 1$. Αν αναιρεθεί η υπόθεση αυτή και θεωρηθεί ότι $f_1 = 1$ το υπόδειγμα γίνεται $X_t = X_{t-1} + e_t$. Επομένως, ένα AR(1) υπόδειγμα για $f_1 = 1$ γίνεται τυχαίος περίπατος και η τιμή της χρονοσειράς X_t τη περίοδο t ισούται με τη τιμή της X_{t-1} της χρονικής περιόδου $t-1$ συν τη τιμή του λευκού θορύβου.

Για τις ροπές μιας χρονοσειράς που ακολουθεί μια διαδικασία τυχαίου περιπάτου ισχύουν;

$$E(X_t) = X_0 + tm \quad \text{μακροχρόνια}$$

$$E(X_t | t) = X_{t-1} + m \quad \text{βραχυχρόνια}$$

$$Var(X_t) = ts^2$$

$$Cov(X_t, X_{t-k}) = (t-k)s^2 \quad \text{για } k \geq 0$$

Ο τυχαίος περίπατος συχνά ονομάζεται και ολοκληρωμένη διαδικασία πρώτης τάξης (*Integrated Process*) και έχει την ιδιότητα οι πρώτες διαφορές του να είναι στάσιμη διαδικασία.

1.3.9 Η διαδικασία Wiener

Μια διαδικασία $\{W(t)\}_0^\infty$ ονομάζεται διαδικασία Wiener ή κίνηση Brown εάν:

- I) $E(W(t)) = 0 \quad \forall t \geq 0$.
- II) Η $W(t)$ έχει κανονική κατανομή με μέσο όρο 0 και διακύμανση $s^2 t$ για κάθε $t > 0$.
- III) $W(0) = 0$.
- IV) Οι προσαυξήσεις της $W(t)$:

$$W(t_n) - W(t_{n-1}), W(t_{n-1}) - W(t_{n-2}), \mathbf{K}, W(t_1) - W(t_2), W(t_1)$$

είναι ανεξάρτητες στάσιμες και ακολουθούν την κανονική κατανομή με μέσο όρο 0 και διακύμανση $s^2 t$.

Η διαδικασία Wiener που μπορεί να θεωρηθεί το συνεχές αντίστοιχο του τυχαίου περιπάτου $S_T = S_{T-1} + e_T$ ουσιαστικά αποτελεί το όριο ενός τυχαίου περιπάτου όταν το διάστημα ανάμεσα σε δύο διαδοχικές τιμές του τείνει στο μηδέν. Δηλαδή, η διαδικασία Wiener μοιάζει με έναν συνεχή τυχαίο περίπατο ορισμένο στο διάστημα $[0, 1]$, έτσι ώστε σε κάθε υποδιάστημα $[a, b]$ του διαστήματος $[0, 1]$, η διαδικασία $W(r)$ με $r \in [a, b]$ να παρουσιάζει ακανόνιστη συμπεριφορά. Αυτό επιτυγχάνεται με το να αντιστοιχηθεί (*mapped*) ο δείκτης t του τυχαίου περιπάτου από το διάστημα $[0, T]$ στο σταθερό διάστημα $[0, 1]$ χρησιμοποιώντας το δείκτη r και να οριστεί μια διαδικασία $W(r)$ σε αυτό το διάστημα. Τότε η $W(r)$ είναι μια διαδικασία Wiener και για σταθερό r ισχύει $W(r) \sim N(0, r)$, $0 \leq r \leq 1$. Κατά συνέπεια, η $W(r)$ αποτελεί συνάρτηση του r . Μια συνάρτηση αυτής της μορφής που κατασκευάζεται όπως περιγράφηκε ονομάζεται τυχαία συνάρτηση.

1.4 Υποδείγματα παραγωγής μη στάσιμων χρονοσειρών

Ένα σημαντικό πρόβλημα που προκύπτει όταν εξετάζεται η συμπεριφορά μη στάσιμων χρονοσειρών είναι η εύρεση του σωστού μηχανισμού παραγωγής των παρατηρήσεων της χρονοσειράς. Οι Nelson και Plosser (1982) ταξινομούν τις μη στάσιμες χρονοσειρές σε δύο διαφορετικές κατηγορίες με βάση το μηχανισμό παραγωγής τους.

Η πρώτη κατηγορία περιλαμβάνει τις χρονοσειρές που μπορούν να εκφραστούν ως μια ντιτερμινιστική συνάρτηση του χρόνου που ονομάζεται τάση συν μια στάσιμη στοχαστική διαδικασία με μηδενικό μέσο όρο. Συνήθως θεωρείται ότι υπάρχει γραμμική τάση, οπότε ένα υπόδειγμα που εκφράζει μια χρονοσειρά αυτής της κατηγορίας είναι το:

$$X_t = a + bt + e_t$$

με

$$\Phi(B)e_t = \Theta(B)a_t, \quad a_t \sim iid(0, s_a^2) \quad \text{και} \quad E(X_t) = a + bt$$

όπου X_t είναι η παρατηρούμενη μη στάσιμη χρονοσειρά, a και b είναι σταθερές παράμετροι και e_t είναι η στάσιμη και αντιστρέψιμη ARMA σειρά των αποκλίσεων γύρω

από την τάση. Επομένως, η σειρά των αποκλίσεων e_t μπορεί να θεωρηθεί ότι είναι οι τιμές της χρονοσειράς απαλλαγμένες από την τάση. Οι Nelson και Plosser (1982) ονομάζουν μια χρονοσειρά που ανήκει στη παραπάνω κατηγορία ως στάσιμη σε τάση (*Trend Stationary Process – TSP*).

Η δεύτερη κατηγορία περιλαμβάνει εκείνες τις χρονοσειρές των οποίων οι πρώτες ή οι μεγαλύτερες διαφορές είναι στάσιμη και αντιστρέψιμη ARMA διαδικασία. Οι Nelson και Plosser (1982) ονομάζουν μια χρονοσειρά που ανήκει σε αυτή την κατηγορία ως στάσιμη σε διαφορές (*Difference Stationary Process – DSP*). Η πιο απλή περίπτωση αυτής της κατηγορίας είναι αυτή η χρονοσειρά όπου οι πρώτες διαφορές της είναι στάσιμη διαδικασία και η οποία μπορεί να παρουσιαστεί με ένα υπόδειγμα της μορφής

$$X_t = X_{t-1} + b + e_t$$

με

$$\Phi(B)e_t = \Theta(B)a_t \quad \text{και} \quad a_t \sim iid(0, s_a^2)$$

όπου η e_t είναι μια στάσιμη και αντιστρέψιμη ARMA διαδικασία και όπου b είναι ο σταθερός μέσος όρος (drift) των πρώτων διαφορών. Ως πιο απλή υποπερίπτωση αυτής της περίπτωσης μπορεί να θεωρηθεί ο τυχαίος περίπατος του οποίου οι πρώτες διαφορές είναι λευκός θόρυβος.

Γράφοντας τη X_t σαν την τιμή σε κάποιο χρονικό σημείο αναφοράς στο παρελθόν που συμβολίζεται μηδέν συν όλες τις μετέπειτα αλλαγές προκύπτει:

$$X_t = X_0 + bt + \sum_{i=1}^t e_i$$

Από τη προηγούμενη σχέση συμπεραίνεται ότι και οι δύο κατηγορίες χρονοσειρών μπορούν να γραφτούν ως μια γραμμική συνάρτηση του χρόνου συν την απόκλιση από αυτή. Όμως, ο σταθερός όρος στη περίπτωση της στάσιμης σε τάση διαδικασίας είναι αμετάβλητη παράμετρος, ενώ στη περίπτωση της στάσιμης σε διαφορές διαδικασίας είναι συνάρτηση ιστορικών γεγονότων, δηλαδή εξαρτάται από την αρχική τιμή X_0 . Επίσης, στην πρώτη κατηγορία οι αποκλίσεις από την τάση είναι στάσιμη διαδικασία, ενώ στη δεύτερη κατηγορία είναι συσσωρεύσεις (άθροισμα) στάσιμων αλλαγών. Η συσσώρευση αυτή δεν είναι στάσιμη διαδικασία αλλά η διακύμανσή της αυξάνει χωρίς όριο καθώς το t μεγαλώνει. Κατά συνέπεια, η μακροχρόνια πρόβλεψη μιας στάσιμης σε διαφορές χρονοσειράς θα επηρεάζεται

πάντοτε από ιστορικά γεγονότα και η διακύμανση του σφάλματος πρόβλεψης θα αυξάνεται χωρίς όριο.

Θα είχε ενδιαφέρον να εξετάζονταν λίγο περισσότερο η κατηγορία των στάσιμων σε διαφορές χρονοσειρών και συγκεκριμένα η περίπτωση του τυχαίου περιπάτου με μέσο όρο (*drift*) και το πώς μπορεί να προκύψει. Έστω η στάσιμη σε τάση διαδικασία:

$$X_t = a + bt + u'_t.$$

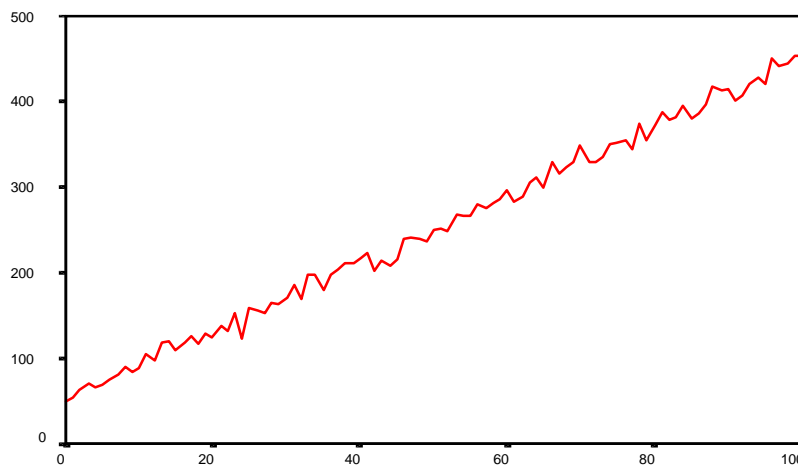
Γράφοντάς την για τη χρονική στιγμή $t-1$ και αφαιρώντας τις δύο σχέσεις προκύπτει:

$$X_{t-1} = a + b(t-1) + u'_{t-1}$$

$$\Delta X_t = b + u'_t - u'_{t-1}$$

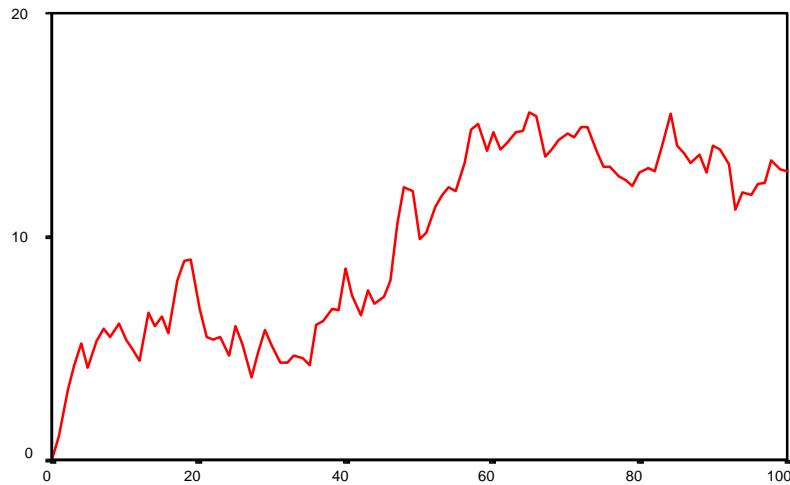
Θέτοντας $e_t = u'_t - u'_{t-1}$ καταλήγουμε στο τυχαίο περίπατο με μέσο όρο $X_t = X_{t-1} + b + e_t$. Το συμπέρασμα δηλαδή είναι ότι ένας τυχαίος περίπατος με μέσο όρο μπορεί να προκύψει παίρνοντας τις πρώτες διαφορές μιας στάσιμης σε τάση χρονοσειράς και ότι ο συντελεστής της τάσης b γίνεται ο μέσος όρος της καινούριας χρονοσειράς.

Στα Διαγράμματα 1 – 10 και 1 – 11 απεικονίζονται μια TSP και μια DSP αντίστοιχα χρονοσειρά.



$$X_t = 50 + 4t + e_t, \quad e_t \sim i.i.d.N(0, 8^2)$$

Διάγραμμα 1-10
Στάσιμη σε τάση χρονοσειρά



$$X_t = X_{t-1} + b + e_t, \quad b = 0, \quad e_t \sim i.i.d. N(0, 1)$$

Διάγραμμα 1-11
Στάσιμη σε διαφορές χρονοσειρά

Μια στάσιμη σε τάση χρονοσειρά μοιάζει με μια στάσιμη σε διαφορές χρονοσειρά καθώς οι αυτοσυσχετίσεις τους έχουν παρόμοια συμπεριφορά. Αυτό έχει αποτέλεσμα ότι οι δύο διαδικασίες πολλές φορές είναι δύσκολο να διακριθούν στην πράξη.

1.5 Η μεθοδολογία των Box και Jenkins

Η μέθοδος των Box και Jenkins (1970) αποτελεί ίσως την πιο διαδεδομένη μεθοδολογία στην ανάλυση χρονοσειρών. Αποτελείται από τέσσερα στάδια βάση των οποίων μπορεί να αναλυθεί οποιοδήποτε χρονοσειρά και να ταυτοποιηθεί σαν μια $ARMA(p, q)$ διαδικασία.

Στάδιο 1: Ταυτοποίηση της χρονοσειράς.

Απαραίτητη προϋπόθεση για να αναλυθεί μια χρονοσειράς είναι η στασιμότητά της. Μια στάσιμη χρονοσειρά έχει αυτοσυσχετίσεις που σβήνουν με γρήγορο ρυθμό και μηδενίζονται σε αντίθεση με μια μη στάσιμη χρονοσειρά όπου οι αυτοσυσχετίσεις της σβήνουν αργά και παραμένουν σημαντικές ακόμη και για μεγάλο αριθμό χρονικών υστερήσεων. Ελέγχοντας τις δειγματικές αυτοσυσχετίσεις αποφασίζεται εάν η χρονοσειρά είναι στάσιμη. Εάν δεν είναι στάσιμη είναι αναγκαίο να μετασχηματιστεί σε στάσιμη με τη χρήση λογαρίθμων ή

διαφορών. Η διαδικασία των διαφορών επαναλαμβάνεται μέχρι οι δειγματικές αυτοσυσχετίσεις να έχουν συμπεριφορά στάσιμης χρονοσειράς.

Οι αυτοσυσχετίσεις και οι μερικές αυτοσυσχετίσεις από κάθε ARMA(p, q) διαδικασία ακολουθούν κάποιο συγκεκριμένο πρότυπο ανάλογα με τις τιμές των p και q . Με βάση τις δειγματικές αυτοσυσχετίσεις και τις μερικές αυτοσυσχετίσεις της αρχικής ή της μετασχηματισμένης χρονοσειράς αν έπρεπε να διαφοροποιηθεί, αποφασίζονται οι κατάλληλες τιμές των p και q . Εάν κρίνονται περισσότερα του ενός υποδείγματα κατάλληλα, τότε η απόφαση για το πιο θα επιλεγεί μπορεί να επιτευχθεί με τη βοήθεια κάποιου πληροφοριακού κριτηρίου. Δύο διαδεδομένα πληροφοριακά κριτήρια είναι το AIC του Akaike (*Akaike's Information Criterion*) και το BIC (*Bayesian Information Criterion*) του Schwartz τα οποία ορίζονται ως:

$$AIC = \ln \hat{S}^2 + \frac{2k}{T} \quad \text{και} \quad BIC = \ln \hat{S}^2 + \frac{k \ln T}{T}$$

αντίστοιχα, όπου k ο αριθμός των παραμέτρων, T το μέγεθος του δείγματος και \hat{S}^2 η εκτίμηση της διακύμανσης του σφάλματος του υποδείγματος.

Επιλέγεται το υπόδειγμα εκείνο που έχει τη μικρότερη τιμή στα δύο παραπάνω κριτήρια. Σημειώνεται ότι τα δύο κριτήρια συνήθως δεν καταλήγουν στα ίδια αποτελέσματα. Έτσι, το AIC επιλέγει μεγάλα σε αριθμό παραμέτρων υποδείγματα ενώ το BIC μικρά.

Εναλλακτικά, εάν δεν γίνεται με βάση τις δειγματικές αυτοσυσχετίσεις και μερικές αυτοσυσχετίσεις να αποφασιστούν οι τιμές των p και q , προτείνεται να εκτιμηθούν όλα τα δυνατά υποδείγματα για κάποιο σταθερό άθροισμα των p και q συνήθως για $p + q \leq 5$ και να επιλεγεί το υπόδειγμα με τη μικρότερη τιμή στα προηγούμενα κριτήρια, προϋποθέτοντας βέβαια ότι έχει στατιστικά σημαντικούς όλους τους συντελεστές του.

Στάδιο 2: Εκτίμηση των παραμέτρων του υποδείγματος.

Εφόσον έχουν αποφασιστεί οι τιμές των p και q εκτιμώνται οι συντελεστές του υποδείγματος. Αυτό μπορεί να επιτευχθεί με αρκετούς τρόπους. Η μέθοδος των ελαχίστων τετραγώνων παρουσιάζει αρκετά προβλήματα στην εκτίμηση των παραμέτρων για αυτό χρησιμοποιείται κάποια άλλη μέθοδος, όπως η μη γραμμική μέθοδος ελαχίστων τετραγώνων και η μέθοδος της μέγιστης πιθανοφάνειας. Η τελευταία μέθοδος παρέχει τις ακριβέστερες εκτιμήσεις.

Στάδιο 3: Έλεγχος του υποδείγματος.

Στο στάδιο αυτό ελέγχεται η επάρκεια του εκτιμηθέντος υποδείγματος. Αρχικά μπορούν

να εκτιμηθούν τα διευρυμένα υποδείγματα είτε στο αυτοπαλίνδρομο μέρος είτε στο μέρος του κινητού μέσου. Δηλαδή, εάν το αρχικό υπόδειγμα είναι ένα $ARMA(p, q)$ να εκτιμηθούν τα $ARMA(p+I, q)$ και $ARMA(p, q+I)$ στα οποία για να είναι το υπόδειγμα ικανοποιητικό θα πρέπει οι επιπλέον παράμετροι να μην είναι στατιστικά σημαντικοί. Στη συνέχεια πρέπει να ελεγχθούν τα σφάλματα από το εκτιμηθέν υπόδειγμα, τα οποία θα πρέπει να είναι λευκός θόρυβος αν το υπόδειγμα είναι ικανοποιητικό. Θα εξηγηθεί σε επόμενη ενότητα πώς μπορεί να πραγματοποιηθεί αυτός ο έλεγχος.

Στάδιο 4: Πραγματοποίηση προβλέψεων.

Αφού έχει επιλεγθεί το κατάλληλο υπόδειγμα και έχει ελεγχθεί η επάρκειά του μπορεί να χρησιμοποιηθεί για τη διενέργεια προβλέψεων για τις μελλοντικές τιμές της χρονοσειράς. Περισσότερες λεπτομέρειες επάνω στη μεθοδολογία των Box και Jenkins και στη πραγματοποίηση προβλέψεων μπορούν να βρεθούν στο βιβλίο των Box και Jenkins (1970), καθώς και σε οποιοδήποτε βιβλίο ανάλυσης χρονοσειρών.

1.6 Η αυτοπαλίνδρομη ολοκληρωμένη κινητού μέσου διαδικασία $ARIMA(p, d, q)$

Στην προηγούμενη ενότητα αναφέρθηκε ότι σύμφωνα με τη μεθοδολογία των Box και Jenkins προτού μια χρονοσειρά μοντελοποιηθεί ως μια $ARMA$ διαδικασία είναι αναγκαία η εξασφάλιση της στασιμότητάς της κάτι που μπορεί να πραγματοποιηθεί με τη διαφοροποίησή της. Μια χρονοσειρά η οποία μπορεί να μετατραπεί σε στάσιμη με τη χρήση διαφορών ονομάζεται ολοκληρωμένη διαδικασία (*Integrated Process*). Εάν, για να γίνει στάσιμη πρέπει να διαφοροποιηθεί μια φορά, δηλαδή εάν ισχύει:

$$\Delta X_t = X_t - X_{t-1} = (1-B) X_t$$

τότε ονομάζεται ολοκληρωμένη διαδικασία πρώτης τάξης (*Integrated Process of order 1*) και συμβολίζεται ως $X_t \square I(1)$. Όμοια, εάν για να γίνει στάσιμη απαιτείται να διαφοροποιηθεί δύο φορές, δηλαδή εάν ισχύει:

$$\Delta^2 X_t = \Delta X_t - \Delta X_{t-1} = X_t - 2X_{t-1} + X_{t-2} = (1-B)^2 X_t$$

τότε ονομάζεται ολοκληρωμένη διαδικασία δεύτερης τάξης και συμβολίζεται ως $X_t \square I(2)$.

Γενικά, εάν για να γίνει στάσιμη πρέπει να διαφοροποιηθεί d φορές, δηλαδή εάν ισχύει:

$$\Delta^d X_t = (1-B)^d X_t$$

τότε ονομάζεται ολοκληρωμένη διαδικασία d τάξης και συμβολίζεται ως $X_t \sim I(d)$.

Από την εμπειρία προκύπτει ότι για τις περισσότερες οικονομικές χρονοσειρές αρκεί η χρήση διαφορών μία ή δύο φορές για να επιτευχθεί η στασιμότητά τους. Στη συνέχεια επιλέγεται το ARMA(p, q) υπόδειγμα που είναι πιο κατάλληλο για τη χρονοσειρά των διαφορών, δηλαδή χρησιμοποιείται για την αρχική χρονοσειρά ένα υπόδειγμα της μορφής:

$$(1 - f_1 B - f_2 B^2 - \dots - f_p B^p) \Delta^d X_t = (1 - q_1 B - q_2 B^2 - \dots - q_q B^q) e_t$$

$$\Phi(B) \Delta^d X_t = \Theta(B) e_t$$

που τελικά γράφεται:

$$\Phi(B)(1-B)^d X_t = \Theta(B) e_t$$

ή ως:

$$\Phi(B)(1-B)^d (X_t - m) = \Theta(B) e_t$$

εάν η χρονοσειρά έχει μη μηδενικό μέσο όρο. Στη περίπτωση αυτή η X_t είναι μια αυτοπαλινδρομη ολοκληρωμένη με d βαθμό ολοκλήρωσης κινητού μέσου διαδικασία (*Autoregressive Integrated Moving Average Process*) και συμβολίζεται ως ARIMA(p, d, q).

1.7 Το πρόβλημα της αυτοσυσχέτισης των σφαλμάτων

Ένα συχνά εμφανιζόμενο πρόβλημα στην ανάλυση παλινδρόμησης, ιδιαίτερα όταν μελετώνται σχέσεις μεταξύ χρονοσειρών, είναι η παραβίαση της υπόθεσης της ανεξαρτησίας των τιμών του τυχαίου σφάλματος του υποδείγματος που ονομάζεται αυτοσυσχέτιση. Η ύπαρξη εξάρτησης στις τιμές του σφάλματος έχει αποτέλεσμα την υποεκτίμηση των διακυμάνσεων των εκτιμητών των συντελεστών του υποδείγματος. Επομένως, όταν πραγματοποιούνται οι έλεγχοι για τη στατιστική σημαντικότητα των συντελεστών, οι τιμές

των στατιστικών $t = \frac{\hat{b}_i}{s(\hat{b}_i)}$ θα είναι μεγαλύτερες από αυτές που θα υπολογίζονταν εάν δεν

υπήρχε το πρόβλημα της αυτοσυσχέτισης και οι μηδενικές υποθέσεις θα απορρίπτονται πιο συχνά, κάτι που μπορεί να οδηγήσει στην εσφαλμένη αποδοχή στο υπόδειγμα ανεξάρτητων μεταβλητών που δεν επηρεάζουν στατιστικά σημαντικά τη συμπεριφορά της εξαρτημένης

μεταβλητής. Για το λόγο αυτό, είναι αναγκαίο να ελέγχεται το πρόβλημα της αυτοσυσχέτισης και να αντιμετωπίζεται. Στη συνέχεια θα αναφερθούν δύο κοινά εφαρμοζόμενοι ελέγχοι για την ύπαρξη αυτοσυσχέτισης που θα συναντηθούν αρκετές φορές στην ανάλυση των επόμενων κεφαλαίων.

1.7.1 Ο έλεγχος των Durbin – Watson

Ο έλεγχος των Durbin – Watson εξετάζει την περίπτωση αυτοσυσχέτισης πρώτου βαθμού των σφαλμάτων, δηλαδή την περίπτωση όπου τα σφάλματα ακολουθούν ένα υπόδειγμα AR(1):

$$e_t = r e_{t-1} + u_t \text{ με } -1 < r < 1, \quad e_t \sim N(0, s_e^2) \text{ και } u_t \sim N(0, s_u^2).$$

Για την πραγματοποίηση του ελέγχου οι Durbin και Watson ορίζουν τη στατιστική:

$$DW = \frac{\sum_{t=2}^T (\hat{e}_t - \hat{e}_{t-1})^2}{\sum_{t=1}^T \hat{e}_t^2} = \frac{\sum_{t=2}^T \hat{e}_t^2 - 2 \sum_{t=2}^T \hat{e}_t \hat{e}_{t-1} + \sum_{t=2}^T \hat{e}_{t-1}^2}{\sum_{t=1}^T \hat{e}_t^2}$$

όπου \hat{e}_t είναι οι τιμές των σφαλμάτων που προκύπτουν από την εκτίμηση του υποδείγματος με τη μέθοδο των ελαχίστων τετραγώνων. Εάν ως εκτίμηση για το συντελεστή r του αυτοπαλίνδρομου υποδείγματος των σφαλμάτων χρησιμοποιηθεί η:

$$\hat{r} = \frac{\sum_{t=2}^T \hat{e}_t \hat{e}_{t-1}}{\sum_{t=1}^T \hat{e}_t^2}$$

και επειδή για μεγάλα δείγματα ισχύει:

$$\sum_{t=2}^T \hat{e}_t^2 \approx \sum_{t=1}^T \hat{e}_{t-1}^2$$

η στατιστική DW μπορεί να γραφτεί ισοδύναμα ως $DW \approx 2(1 - \hat{r})$. Από τη σχέση αυτή συμπεραίνεται ότι εάν δεν υπάρχει αυτοσυσχέτιση, δηλαδή εάν $\hat{r} = 0$, τότε θα πρέπει $DW = 2$. Εάν υπάρχει τέλεια θετική αυτοσυσχέτιση, δηλαδή εάν $\hat{r} = 1$, τότε $DW = 0$. Τέλος εάν υπάρχει τέλεια αρνητική αυτοσυσχέτιση, δηλαδή εάν $\hat{r} = -1$, τότε $DW = 4$. Επομένως, για την στατιστική DW ισχύει:

$$0 \leq DW \leq 4$$

Όσο πιο κοντά είναι η τιμή της DW στο μηδέν τόσο μεγαλύτερη είναι η πιθανότητα να υπάρχει θετική αυτοσυσχέτιση στα σφάλματα. Από την άλλη πλευρά, όσο πιο κοντά στο τέσσερα είναι η τιμή της DW τόσο μεγαλύτερη είναι η πιθανότητα να υπάρχει αρνητική αυτοσυσχέτιση. Μια τιμή της DW κοντά στο δύο αποτελεί ισχυρή ένδειξη ότι δεν υπάρχει πρόβλημα αυτοσυσχέτισης.

Για την πραγματοποίηση του ελέγχου οι Durbin και Watson έχουν υπολογίσει δύο κριτικές τιμές, μια ανώτερη dU και μια κατώτερη dL , οι οποίες εξαρτώνται από τον αριθμό των ανεξαρτήτων μεταβλητών που έχουν συμπεριληφθεί στο υπόδειγμα, το μέγεθος του δείγματος και το επίπεδο σημαντικότητας που επιδιώκεται και που μπορούν να βρεθούν σε οποιοδήποτε βιβλίο οικονομετρίας. Ειδικότερα, έλεγχος για τη περίπτωση της ύπαρξης θετικής αυτοσυσχέτισης ορίζεται ως:

$$H_0 : r = 0$$

$$H_1 : r > 0$$

και απορρίπτεται η μηδενική υπόθεση εάν $DW < dL$, ενώ γίνεται αποδεκτή εάν $DW > dU$. Εάν η τιμή της DW βρίσκεται ανάμεσα στις κριτικές τιμές, δηλαδή $dL \leq DW \leq dU$, τότε δεν μπορεί να εξαχθεί κάποιο συμπέρασμα για την ύπαρξη ή όχι αυτοσυσχέτισης πρώτου βαθμού. Η περιοχή αυτή ονομάζεται αβέβαιη περιοχή και αποτελεί σοβαρό μειονέκτημα του ελέγχου. Εάν επιδιώκεται να ελεγχθεί η ύπαρξη αρνητικής αυτοσυσχέτισης, τότε ο έλεγχος ορίζεται ως:

$$H_0 : r = 0$$

$$H_1 : r < 0$$

και απορρίπτεται η μηδενική υπόθεση εάν $DW > 4 - dL$, ενώ γίνεται αποδεκτή εάν $DW < 4 - dU$. Όπως και στην περίπτωση του ελέγχου για θετική αυτοσυσχέτιση εάν η τιμή της DW βρίσκεται στην αβέβαιη περιοχή, δηλαδή $4 - dU \leq DW \leq 4 - dL$, τότε ο έλεγχος δεν μπορεί να πραγματοποιηθεί. Ο έλεγχος των Durbin και Watson είναι πολύ χρήσιμος γιατί η εμπειρία έχει δείξει ότι όταν εξετάζονται οικονομικές χρονοσειρές τις περισσότερες φορές τα σφάλματα ακολουθούν ένα $AR(1)$ υπόδειγμα. Τελειώνοντας, σημειώνεται ότι ο έλεγχος αυτός αφορά αποκλειστικά την περίπτωση αυτοσυσχέτισης πρώτου βαθμού και για να έχει ισχύ θα πρέπει στο υπόδειγμά να μην υπάρχουν χρονικές υστερήσεις της εξαρτημένης μεταβλητής και να περιλαμβάνεται απαραίτητα ο σταθερός όρος.

1.7.2 Ο έλεγχος για λευκό θόρυβο των Ljung και Box

Ο έλεγχος των Durbin και Watson που αναφέρθηκε στη προηγούμενη παράγραφο είναι δεσμευτικός ως προς τις διαφορετικές μορφές αυτοσυσχέτισης που μπορεί να συναντηθούν στα σφάλματα, καθώς ελέγχει μόνο την περίπτωση αυτοσυσχέτισης πρώτου βαθμού. Όμως, μερικές φορές είναι σημαντικό να γενικευτεί ο έλεγχος καλύπτοντας και άλλες μορφές αυτοσυσχέτισης. Κάτι τέτοιο θα μπορούσε να επιτευχθεί με το να ελεγχθούν εάν τα σφάλματα είναι λευκός θόρυβος, δηλαδή εάν είναι τελείως ασυσχέτιστα μεταξύ τους και έχουν μηδενικό μέσο όρο και σταθερή διακύμανση.

Ένας έλεγχος αυτού του είδους, που θα μπορούσε να επεκταθεί και να εφαρμοστεί και για οποιαδήποτε χρονοσειρά για την οποία επιθυμείται να ελεγχθεί εάν είναι λευκός θόρυβος, γίνεται με τη χρησιμοποίηση της Portmanteau στατιστικής Q που κατασκευάζεται με βάση τις τιμές των δειγματικών συντελεστών αυτοσυσχέτισης των σφαλμάτων. Στην αρχική της μορφή η στατιστική Q προτάθηκε από τους Box και Pierce (1970) και ορίζεται ως:

$$Q = T \sum_{i=1}^m r_i^2$$

όπου

$$r_i = \frac{\sum_{t=i+1}^T \hat{e}_t \hat{e}_{t-i}}{\sum_{t=1}^T \hat{e}_t^2}$$

και m είναι ο αριθμός των δειγματικών αυτοσυσχετίσεων που χρησιμοποιούνται. Αργότερα, οι Ljung και Box (1978) διαπίστωσαν ότι η ασυμπτωτική συμπεριφορά της Q δεν είναι ικανοποιητική και πρότειναν τη τροποποίησή της σε:

$$Q' = T(T+2) \sum_{i=1}^m \frac{r_i^2}{T-i}$$

η οποία ακολουθεί κατανομή X^2 με $m-k$ βαθμούς ελευθερίας, όπου k ο αριθμός των ανεξαρτήτων μεταβλητών που περιλαμβάνει το υπόδειγμα. Στην περίπτωση του ελέγχου μιας δεδομένης χρονοσειρά, για το εάν είναι λευκός θόρυβος, τότε η Q' ακολουθεί την κατανομή X^2 με m βαθμούς ελευθερίας.

Ο έλεγχος ορίζεται ως:

H_0 : Τα σφάλματα είναι λευκός θόρυβος

H_1 : Τα σφάλματα δεν είναι λευκός θόρυβος

και γίνεται αποδεκτή η H_0 όταν $Q' < X_{m,a}^2$.

Όσον αφορά τον αριθμό των δειγματικών αυτοσυσχετίσεων που πρέπει να χρησιμοποιηθούν στον υπολογισμό της Q' προτείνεται ένας αριθμός m ίσος με το ένα τέταρτο του δείγματος αλλά όχι πάνω από πενήντα.

1.8 Βασικές έννοιες ασυμπτωτικής θεωρίας

Στην ανάλυση των επόμενων κεφαλαίων θα γίνει αρκετές φορές χρήση εννοιών που εξετάζουν την ασυμπτωτική συμπεριφορά εκτιμητών. Για το λόγο αυτό κρίνεται αναγκαίο να αναφερθούν περιληπτικά η βασική ορολογία και τα αναγκαία θεωρήματα της ασυμπτωτικής θεωρίας. Οι επόμενοι ορισμοί προέρχονται κυρίως από τον Greene (1993).

1.8.1 Σύγκλιση κατά πιθανότητα

Έστω ότι η X_T είναι μια ακολουθία τυχαίων μεταβλητών με το δείκτη T να συμβολίζει το μέγεθος του δείγματος τότε:

α. Σύγκλιση κατά πιθανότητα σε κάποια σταθερά c

Η τυχαία μεταβλητή X_T συγκλίνει κατά πιθανότητα (*Convergence in Probability*) σε μια σταθερά c εάν:

$$\lim_{T \rightarrow \infty} \Pr(|X_T - c| > e) = 0$$

για κάθε $e > 0$. Εάν η X_T συγκλίνει κατά πιθανότητα στη σταθερά c γράφεται ισοδύναμα:

$$\text{Plim } X_T = c \quad \text{ή} \quad X_T \xrightarrow{P} c.$$

Με τη βοήθεια της σύγκλισης κατά πιθανότητα μπορεί να οριστεί η έννοια του συνεπή εκτιμητή. Έτσι, ένας εκτιμητής \hat{q}_T μιας παραμέτρου q είναι συνεπής εκτιμητής της q εάν $\text{Plim } \hat{q}_T = q$.

β. Σύγκλιση κατά πιθανότητα σε τυχαία μεταβλητή

Η τυχαία μεταβλητή X_T συγκλίνει κατά πιθανότητα στην τυχαία μεταβλητή X εάν:

$$\lim_{T \rightarrow \infty} \Pr(|X_T - X| > e) = 0$$

για κάθε $\epsilon > 0$. Εάν η X_T συγκλίνει κατά πιθανότητα στην τυχαία μεταβλητή X γράφεται ισοδύναμα:

$$P\lim X_T = X \quad \text{ή} \quad X_T \xrightarrow{P} X.$$

1.8.2 Σύγκλιση κατά κατανομή

Έστω πάλι ότι η X_T είναι μια ακολουθία τυχαίων μεταβλητών με το δείκτη T να συμβολίζει το μέγεθος του δείγματος και έστω ότι η X_T έχει αθροιστική συνάρτηση κατανομής (CDF) την $F_T(X)$ τότε η τυχαία μεταβλητή X_T συγκλίνει κατά κατανομή (Convergence in Distribution) στην τυχαία μεταβλητή X με αθροιστική συνάρτηση κατανομής την $F(X)$ εάν ισχύει:

$$\lim_{T \rightarrow \infty} |F_T(X) - F(X)| = 0$$

σε όλα τα συνεχόμενα σημεία της $F(X)$. Εάν η X_T συγκλίνει κατά κατανομή στην τυχαία μεταβλητή X γράφεται ισοδύναμα:

$$\lim_{T \rightarrow \infty} F_T(X) = F(X) \quad \text{ή} \quad X_T \xrightarrow{d} X$$

και η $F(X)$ αποτελεί την οριακή κατανομή της X .

1.8.3 Σύγκλιση τυχαίων διανυσμάτων και πινάκων

Η έννοια της σύγκλισης κατά πιθανότητα σε σταθερά μπορεί να επεκταθεί και για την περίπτωση τυχαίων διανυσμάτων και τυχαίων πινάκων. Έτσι, το τυχαίο διάνυσμα \mathbf{x}_T συγκλίνει κατά πιθανότητα στο διάνυσμα \mathbf{c} που αποτελείται από σταθερές και έχει την ίδια διάσταση με το \mathbf{x}_T , εάν $P\lim \mathbf{x}_T = \mathbf{c}$. Όμοια ο τυχαίος πίνακας \mathbf{X}_T συγκλίνει κατά πιθανότητα στον πίνακα των σταθερών \mathbf{C} που είναι ιδίων διαστάσεων με το \mathbf{X}_T εάν $P\lim \mathbf{X}_T = \mathbf{C}$.

Για την περίπτωση της σύγκλισης κατά κατανομή τυχαίων διανυσμάτων και πινάκων η διαδικασία είναι πιο σύνθετη. Και αυτό γιατί θα πρέπει να χρησιμοποιηθεί η από κοινού αθροιστική συνάρτηση κατανομής των τυχαίων μεταβλητών που περιλαμβάνονται στα διανύσματα ή στους πίνακες. Κατά συνέπεια, το τυχαίο διάνυσμα \mathbf{x}_T με από κοινού

αθροιστική συνάρτηση κατανομής $F_T(\mathbf{x}_T)$ συγκλίνει κατά κατανομή στο διάνυσμα των τυχαίων μεταβλητών \mathbf{x} με από κοινού αθροιστική συνάρτηση κατανομής $F(\mathbf{x})$ εάν:

$$\lim_{T \rightarrow \infty} |F_T(\mathbf{x}_T) - F(\mathbf{x})| = 0.$$

Όμοια, ο τυχαίος πίνακας \mathbf{X}_T με από κοινού αθροιστική συνάρτηση κατανομής $F_T(\mathbf{X}_T)$ συγκλίνει κατά κατανομή στο διάνυσμα των τυχαίων μεταβλητών \mathbf{X} με από κοινού αθροιστική συνάρτηση κατανομής $F(\mathbf{X})$ εάν:

$$\lim_{T \rightarrow \infty} |F_T(\mathbf{X}_T) - F(\mathbf{X})| = 0$$

δηλαδή όταν το προηγούμενο όριο ισούται με μηδέν.

1.8.4 Τάξεις σύγκλισης μιας ακολουθίας

Οι τάξεις σύγκλισης (*Order of a Sequence*) αποτελούν μέτρο του ρυθμού με τον οποίο μια ακολουθία συγκλίνει ή αποκλίνει. Οι επόμενοι ορισμοί προέρχονται από τον White (2001).

Έστω ότι η b_n είναι μια ακολουθία πραγματικών αριθμών τότε:

1) Η ακολουθία b_n λέγεται ότι είναι μεγαλύτερης τάξης από n^l (*at most of order n^l*) και συμβολίζεται $O(n^l)$, εάν για κάθε πεπερασμένο πραγματικό αριθμό $\Delta > 0$ υπάρχει κάποιος πεπερασμένος ακέραιος N , έτσι ώστε για όλα τα $n \geq N$ να ισχύει $|n^{-l} b_n| < \Delta$. Δηλαδή, το όριο στο άπειρο της b_n διαιρεμένης με το n^l να είναι κάποια μη μηδενική πεπερασμένη σταθερά. Εάν $b_n = O(n^0) = O(1)$, η b_n είναι φραγμένη (*bounded*) και μπορεί να έχει ή και να μην έχει όριο.

2) Η ακολουθία b_n λέγεται ότι είναι τάξης μικρότερης από n^l (*order smaller than n^l*) και συμβολίζεται $o(n^l)$, εάν για κάθε πραγματικό αριθμό $d > 0$ υπάρχει κάποιος πεπερασμένος ακέραιος $N(d)$, τέτοιος ώστε για όλα τα $n \geq N(d)$ να ισχύει $|n^{-l} b_n| < d$. Δηλαδή, $n^{-l} b_n \rightarrow 0$.

Εάν $b_n = o(n^0) = o(1)$, τότε $b_n \rightarrow 0$. Εάν κάθε στοιχείο ενός διανύσματος ή ενός πίνακα είναι $O(n^l)$ ή $o(n^l)$ τότε το διάνυσμα ή ο πίνακας είναι αντίστοιχα $O(n^l)$ ή $o(n^l)$.

Έστω τώρα ότι η b_n είναι μια ακολουθία τυχαίων μεταβλητών τότε:

3) Η ακολουθία b_n λέγεται ότι είναι μεγαλύτερης τάξης από n^l κατά πιθανότητα (*at most of order n^l in probability*) και συμβολίζεται $O_p(n^l)$, εάν για κάθε $e > 0$ υπάρχει κάποιος

πεπερασμένος $\Delta_e > 0$ και κάποιος $N_e \in \mathbb{N}$, τέτοιοι ώστε $P\{w: |n^{-1}b_n(w)| > \Delta_e\} < e$ για όλα τα $n \geq N_e$. Δηλαδή, εάν $P\text{lim} \frac{1}{n^l} b_n$ έχει ως αποτέλεσμα κάποια μη μηδενική σταθερά.

4) Η ακολουθία b_n λέγεται ότι είναι τάξης μικρότερης από n^l κατά πιθανότητα (*order smaller than n^l in probability*) και συμβολίζεται $o_p(n^l)$, εάν $n^{-l}b_n \xrightarrow{p} 0$.

1.8.5 Ασθενής σύγκλιση

Η έννοια της ασθενούς σύγκλισης (*Weak Convergence*) αποτελεί επέκταση της σύγκλισης σε κατανομή και μπορεί να εφαρμοστεί για τυχαίες συναρτήσεις. Ένας ορισμός της ασθενούς σύγκλισης μπορεί να βρεθεί στον White (2001):

Έστω ότι m και m_n για $n = 1, 2, \mathbf{K}$ είναι μέτρα πιθανότητας στο μετρικό μετρήσιμο χώρο (*metrized measurable space*) (S, \mathcal{S}) . Τότε το m_n συγκλίνει ασθενώς στο m και συμβολίζεται $m_n \Rightarrow m$ ή $m_n \xrightarrow{d} m$, εάν $m_n(A) \rightarrow m(A)$ καθώς $n \rightarrow \infty$ για όλα τα m -συνεχή (*m-continuity*) σύνολα A του \mathcal{S} . Εναλλακτικά, η έννοια μπορεί να οριστεί χρησιμοποιώντας κατάλληλα ορισμένα διαστήματα συναρτήσεων πραγματικών τιμών. Δηλαδή, $m_n \Rightarrow m$ εάν:

$$\int f dm_n \rightarrow \int f dm \text{ καθώς } n \rightarrow \infty$$

για όλες τις ομοιόμορφα ορισμένες (*bounded*) συνεχείς συναρτήσεις $f: S \rightarrow \mathbb{R}$.

1.8.6 Το κεντρικό οριακό θεώρημα

Όταν μελετώνται οι ασυμπτωτικές ιδιότητες κάποιου εκτιμητή συχνά γίνεται χρήση του κεντρικού οριακού θεωρήματος (*Central Limit Theorem*) με τη βοήθεια του οποίου μια τυχαία μεταβλητή μπορεί με κατάλληλη τυποποίηση να μετασχηματιστεί, ώστε να έχει οριακά κάποια καλά καθορισμένη κατανομή. Με αυτόν τον τρόπο μπορεί να επιτευχθεί ασυμπτωτική κανονικότητα για κάποια τυχαία μεταβλητή.

Το κεντρικό οριακό θεώρημα υπάρχει σε αρκετές παραλλαγές ανάλογα με τους κάθε φορά περιορισμούς. Στη συνέχεια θα αναφερθούν μερικές συχνά χρησιμοποιούμενες μορφές του.

1) Εάν $X_1, X_2, \mathbf{K}, X_n$ είναι ένα τυχαίο δείγμα προερχόμενο από πληθυσμό με πεπερασμένο μέσο όρο m και πεπερασμένη διακύμανση s^2 τότε:

$$\sqrt{n}(\bar{X}_n - m) \xrightarrow{d} N(0, s^2)$$

όπου $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$, (Greene, 1993).

2) Για δεδομένους περιορισμούς για τις ροπές, την εξάρτηση και την ετερογένεια μιας βαθμωτής ακολουθίας $\{X_t\}$ ισχύει:

$$\frac{\bar{X}_n - \bar{m}_n}{\bar{S}_n / \sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \bar{m}_n)}{\bar{S}_n} \square N(0, 1)$$

όπου $E(\bar{X}_n) = \bar{m}_n$ και $Var(\bar{X}_n) = \frac{\bar{S}_n^2}{n}$, όπως δίνεται στον White (2001).

3) Εάν οι $X_1, X_2, \mathbf{K}, X_n$ είναι ανεξάρτητες τυχαίες μεταβλητές με κοινό μέσο όρο m και κοινή διακύμανση s^2 τότε το άθροισμά τους:

$$S_n = X_1 + X_2 + \mathbf{K} + X_n \square N(nm, ns^2)$$

και κατά συνέπεια:

$$\frac{S_n - nm}{s\sqrt{n}} \square N(0, 1)$$

δηλαδή το προηγούμενο πηλίκο ακολουθεί την τυποποιημένη κανονική κατανομή.

1.8.7 Το συναρτησιακό κεντρικό οριακό θεώρημα

Το συναρτησιακό κεντρικό οριακό θεώρημα (*Functional Central Limit Theorem*) παρέχει τη δυνατότητα εφόσον τηρούνται οι αναγκαίοι περιορισμοί μια τυχαία συνάρτηση $W(r)$ να συγκλίνει σε μια διαδικασία Wiener ορισμένη στο διάστημα $[0, 1]$. Δηλαδή, το όριο της $W(r)$ αποτελεί συνάρτηση του r , για αυτό και το οριακό θεώρημα αυτό του είδους ονομάζεται συναρτησιακό. Στην απλούστερή του μορφή το συναρτησιακό οριακό θεώρημα ορίζεται ως εξής:

Έστω $\{X_t\}$ μια ακολουθία τυχαίων i.i.d. βαθμωτών με μηδενικό μέσο όρο. Εάν $s^2 \equiv Var(X_t) < \infty$ για $s^2 \neq 0$ τότε $W(r) \Rightarrow W$, (White, 2001). Η μορφή αυτή του συναρτησιακού κεντρικού οριακού θεωρήματος ονομάζεται θεώρημα του Donsker και είναι γνωστή και ως *invariance principle*.

1.8.8 Το θεώρημα της συνεχούς αντιστοίχισης

Έστω ότι το ζεύγος (S, \mathcal{S}) είναι ένας μετρικός μετρήσιμος χώρος (*metrized measurable space*) και έστω ότι m και m_n είναι μέτρα πιθανότητας στο (S, \mathcal{S}) που αντιστοιχούν στα V, V_n τυχαία στοιχεία του S , για $n = 1, 2, \mathbf{K}$. Τότε, White (2001):

1) Έστω ότι $h: S \rightarrow \square$ είναι μια συνεχής αντιστοίχιση (*continuous mapping*). Εάν $V_n \Rightarrow V$, τότε $h(V_n) \Rightarrow h(V)$.

2) Έστω $h: S \rightarrow \square$ είναι μια αντιστοίχιση τέτοια ώστε το σύνολο των ασυνεχειών του h , $D_h \equiv \left\{ \mathbf{s} \in S : \lim_{r \rightarrow \mathbf{s}} h(r) \neq h(\mathbf{s}) \right\}$ να έχει $m(D_h) = 0$. Εάν $V_n \Rightarrow V$, τότε $h(V_n) \Rightarrow h(V)$.

1.9 Ανακεφαλαίωση

Στα πλαίσια του κεφαλαίου αυτού έγινε προσπάθεια να τεθεί το θεωρητικό υπόβαθρο για την ανάλυση των επόμενων κεφαλαίων. Ειδικότερα, εξηγήθηκαν βασικά θέματα που προέρχονται από το χώρο της ανάλυσης χρονοσειρών, όπως η έννοια της στασιμότητας, της ταυτοποίησης της χρονοσειράς και των ολοκληρωμένων διαδικασιών, τα προβλήματα που προκύπτουν στην οικονομετρική μεθοδολογία όταν χρησιμοποιούνται χρονοσειρές ως δείγμα, καθώς και θεμελιώδη ζητήματα από την ασυμπτωτική θεωρία. Φυσικά, η ανάλυση που πραγματοποιήθηκε ήταν συνοπτική και δεν αποσκοπούσε σε καμία περίπτωση στην εξάντληση όλων των θεμάτων της οικονομετρίας ούτε των εννοιών των χρονοσειρών και της ασυμπτωτικής θεωρίας. Περισσότερες λεπτομέρειες για θέματα οικονομετρίας και χρονοσειρών μπορούν να βρεθούν στους Mills (2002) και Greene (1993) και για θέματα ασυμπτωτικής θεωρίας στους Spanos (1986) και White (2001).

ΚΕΦΑΛΑΙΟ 2

ΤΟ ΠΡΟΒΛΗΜΑ ΤΗΣ ΥΠΟΦΑΙΝΟΜΕΝΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

2.1 Εισαγωγή

Το πρόβλημα της υποφαινομενικής παλινδρόμησης (*spurious regression*) προκύπτει στην οικονομετρία όταν δύο ή περισσότερες ανεξάρτητες χρονοσειρές βρίσκονται να συσχετίζονται σύμφωνα με τη στατιστική συμπερασματολογία της μεθόδου των ελαχίστων τετραγώνων. Δηλαδή, ο συντελεστής προσδιορισμού R^2 του εκτιμηθέντος υποδείγματος είναι πάρα πολύ υψηλός και οι έλεγχοι για τη στατιστική σημαντικότητα των συντελεστών του με τις t και F στατιστικές απορρίπτουν τη μηδενική υπόθεση ότι οι συντελεστές είναι μηδέν αλλά, στην πραγματικότητα, δεν υπάρχει αληθινή σχέση ανάμεσα στην εξαρτημένη και στην ή στις ανεξάρτητες μεταβλητές. Στην περίπτωση αυτή, η στατιστική σημαντικότητα είναι υποφαινομενική και ο ερευνητής μπορεί να παραπλανηθεί και να οδηγηθεί σε λανθασμένα συμπεράσματα.

Κρίνεται αναγκαίο να αποσαφηνιστεί η διαφορά ανάμεσα στην υποφαινομενική παλινδρόμηση (*spurious regression*) και στην παλινδρόμηση χωρίς νόημα (*no sense regression*). Οι δύο έννοιες παρόλο που πολλές φορές χρησιμοποιούνται στη βιβλιογραφία ως ταυτόσημες, καθώς οι αιτίες που τις προκαλούν είναι κοινές, εντούτοις έχουν σημαντική διαφορά. Ο όρος υποφαινομενική παλινδρόμηση αναφέρεται σε μεταβλητές οι οποίες έχει νόημα να συσχετισθούν, όπως για παράδειγμα δύο οικονομικές μεταβλητές. Αντίθετα, ο όρος παλινδρόμηση χωρίς νόημα αναφέρεται σε μεταβλητές οι οποίες δεν είναι λογικό να συσχετίζονται. Χαρακτηριστικό παράδειγμα παλινδρόμησης χωρίς νόημα αποτελεί η μεγάλη ερμηνευτική ικανότητα του υποδείγματος με χαμηλή όμως τιμή για την στατιστική DW που παρατήρησε ο Hendry (1980) ανάμεσα στον τριμηνιαίο πληθωρισμό και την αντίστοιχη συσσωρευμένη βροχόπτωση για τη Μεγάλη Βρετανία. Μια παλινδρόμηση χωρίς νόημα είναι υποχρεωτικά και υποφαινομενική. Το αντίθετο δεν ισχύει πάντα. Είναι φανερό ότι η υποφαινομενική παλινδρόμηση είναι αρκετά πιο δύσκολο να εντοπιστεί και να

αντιμετωπίζεται από τον ερευνητή σε σχέση με την παλινδρόμηση χωρίς νόημα που απλά αγνοείται.

2.2 Υποφαινομενικές Παλινδρομήσεις με ολοκληρωμένες διαδικασίες

Οι Granger και Newbold (1974) παρατήρησαν ότι συχνά στην οικονομετρία προκύπτουν μεταξύ χρονοσειρών εξισώσεις παλινδρόμησης που ενώ φαίνονται ότι έχουν καλό βαθμό προσαρμογής, όπως συμπεραίνεται από την πολύ υψηλή τιμή του R^2 , παρουσιάζουν πολύ χαμηλή τιμή (κοντά στο μηδέν) για τη στατιστική DW των Durbin και Watson. Χρησιμοποιώντας προσομοίωση κατέληξαν στο συμπέρασμα ότι το φαινόμενο συμβαίνει όταν οι εμπλεκόμενες στην παλινδρόμηση χρονοσειρές είναι ανεξάρτητοι τυχαίοι περίπατοι. Τότε, οι συνηθισμένοι έλεγχοι για τη στατιστική σημαντικότητα των συντελεστών του υποδείγματος δεν είναι έγκυροι και τα συμπεράσματα σχετικά με την ύπαρξη σχέσεως μεταξύ των μεταβλητών είναι λανθασμένα.

Έστω το υπόδειγμα της πολλαπλής γραμμικής παλινδρόμησης:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.1)$$

για το οποίο ισχύουν οι υποθέσεις που έχουν αναφερθεί στην ενότητα 1.1. Όπως είναι γνωστό από τη στατιστική συμπεραματολογία της ανάλυσης παλινδρόμησης, ο έλεγχος για τη στατιστική σημαντικότητα των συντελεστών του υποδείγματος μπορεί να πραγματοποιηθεί με δυο διαφορετικούς τρόπους. Ο πρώτος τρόπος είναι να ελεγχθεί η στατιστική σημαντικότητα κάθε συντελεστή ξεχωριστά χρησιμοποιώντας τη στατιστική t , δηλαδή:

$$t = \frac{\hat{b}_i}{s(\hat{b}_i)} \quad (2.2)$$

όπου \hat{b}_i είναι η εκτίμηση της τιμής του συντελεστή b_i και $s(\hat{b}_i)$ το τυπικό σφάλμα του εκτιμητή \hat{b}_i . Ο έλεγχος ορίζεται ως:

$$H_0 : b_i = 0$$

$$H_1 : b_i \neq 0$$

και απορρίπτεται η μηδενική υπόθεση όταν $|t| > |t_{T-k-1, a/2}|$ όπου a το επιθυμητό επίπεδο σημαντικότητας και k ο αριθμός των ανεξαρτήτων μεταβλητών.

Ο δεύτερος τρόπος είναι να ελεγχθεί η ταυτόχρονη στατιστική σημαντικότητα των συντελεστών του υποδείγματος χρησιμοποιώντας τη στατιστική:

$$F = \frac{SSR/k}{SSE/T-k-1} = \frac{SSR}{SSE} \frac{T-k-1}{k} = \frac{R^2}{1-R^2} \frac{T-k-1}{k} \quad (2.3)$$

όπου $SSR = \sum_{i=1}^T (\hat{Y}_i - \bar{Y})^2$ και $SSE = \sum_{i=1}^T \hat{e}_i^2$. Ο έλεγχος ορίζεται ως:

$$H_0 : b_1 = b_2 = \dots = b_k = 0$$

$$H_1 : \text{τουλάχιστον ένα } b_j \neq 0 \text{ για } j = 1, 2, \dots, k$$

ή εάν χρησιμοποιηθεί ο συντελεστής προσδιορισμού ως:

$$H_0 : R^2 = 0$$

$$H_1 : R^2 \neq 0$$

και απορρίπτεται η μηδενική υπόθεση όταν $F > F_{k, T-k-1, \alpha}$ ενώ διαφορετικά γίνεται αποδεκτή. Όταν οι ανεξάρτητες μεταβλητές που εμπλέκονται στην παλινδρόμηση δεν συσχετίζονται με την εξαρτημένη μεταβλητή, δηλαδή η μηδενική υπόθεση είναι αληθινή, το ποσοστό απόρριψής της H_0 και για τους δύο ελέγχους που έχουν αναφερθεί θα πρέπει να ισούται με το επίπεδο σημαντικότητας α . Για να εξετάσουν πως προκύπτουν υποφαινομενικές παλινδρομήσεις οι Granger και Newbold (1974) μελέτησαν πως επηρεάζονται οι έλεγχοι για τη στατιστική σημαντικότητα των συντελεστών του υποδείγματος και η τιμή του R^2 στο απλό γραμμικό υπόδειγμα όταν οι χρησιμοποιούμενες χρονοσειρές είναι μη στάσιμες.

Ανεξάρτητα από τις ιδιότητες κάθε χρονοσειράς υπάρχει κάποιο β , έτσι ώστε τα σφάλματα στο υπόδειγμα (2.1) να έχουν μέσο όρο μηδέν, σταθερή διακύμανση και να είναι ανεξάρτητα μεταξύ τους. Όμως, όταν η εξαρτημένη μεταβλητή δεν είναι λευκός θόρυβος, η μηδενική υπόθεση δεν μπορεί να είναι αληθινή και οι έλεγχοι αυτής κρίνονται ακατάλληλοι. Εάν η μηδενική υπόθεση είναι σωστή και εκτιμηθεί ένα υπόδειγμα παλινδρόμησης της μορφής του υποδείγματος (2.1) στα επίπεδα οικονομικών χρονοσειρών, ενώ οι χρησιμοποιούμενες χρονοσειρές είναι μη στάσιμες ή υψηλά αυτοσυσχετιζόμενες διαδικασίες, κάτι που συμβαίνει συχνά με οικονομικές χρονοσειρές, η διαδικασία ελέγχου με τη στατιστική (2.3) καταρρέει επειδή αυτή δεν θα ακολουθεί την κατανομή F υπό τη μηδενική υπόθεση και τα σφάλματα ε θα έχουν τις ίδιες ιδιότητες αυτοσυσχέτισης όπως η εξαρτημένη μεταβλητή.

Έστω το υπόδειγμα της απλής γραμμικής παλινδρόμησης $Y_t = b_0 + b_1 X_t + u_t$ με τις Y_t και X_t να παράγονται σύμφωνα με τις ανεξάρτητες αυτοπαλινδρομες διαδικασίες πρώτης τάξης $Y_t = f_Y Y_{t-1} + e_{Yt}$ και $X_t = f_X X_{t-1} + e_{Xt}$ αντίστοιχα. Σε αυτή την περίπτωση το R^2 είναι το τετράγωνο του συντελεστή δειγματικής συσχέτισης r ανάμεσα στις Y_t και X_t που για τη διακύμανση του ισχύει σύμφωνα με τον Kendall (1954) $Var(r) = T^{-1}(1 + f_Y f_X)/(1 - f_Y f_X)$. Ο συντελεστής δειγματικής συσχέτισης r εκ κατασκευής παίρνει τιμές στο διάστημα $(-1, 1)$ που συνεπάγεται ότι όταν η διακύμανση του είναι μεγαλύτερη από $\frac{1}{3}$, δηλαδή όταν $f_X f_Y > \frac{T-3}{T+3}$, η κατανομή του δεν μπορεί να έχει μοναδική κορυφή στο μηδέν. Από τα παραπάνω συμπεραίνεται ότι μια υψηλή τιμή του R^2 δεν υποδεικνύει απαραίτητα σημαντική σχέση ανάμεσα σε χρονοσειρές όταν ο μηχανισμός παραγωγής τους είναι αυτοπαλινδρομες διαδικασίες πρώτης τάξης με συντελεστές κοντά στη μονάδα, ενώ μια χαμηλή τιμή για τη στατιστική των Durbin και Watson είναι ένδειξη ότι τα σφάλματα από την εκτίμηση του υποδείγματος (2.1) αυτοσυσχετίζονται. Κατά συνέπεια, οι Granger και Newbold (1974) υποστηρίζουν ότι το φαινόμενο που έχει περιγραφεί μπορεί να προκύψει όταν εκτιμηθούν υποδείγματα παλινδρόμησης που συσχετίζουν τα επίπεδα ανεξαρτήτων μη στάσιμων χρονοσειρών.

Οι προσομοιώσεις που πραγματοποίησαν οι Granger και Newbold (1974) επιβεβαίωσαν τα προηγούμενα συμπεράσματα. Αρχικά χρησιμοποίησαν το υπόδειγμα της απλής γραμμικής παλινδρόμησης:

$$Y_t = a + b X_t + u_t \quad (2.4)$$

όπου οι Y_t και X_t παρήχθησαν ως ανεξάρτητοι τυχαίοι περίπατοι 50 παρατηρήσεων ο καθένας και μελέτησαν την κατανομή συχνότητας της στατιστικής (2.2) που χρησιμοποιείται για τον έλεγχο της στατιστικής σημαντικότητας του συντελεστή της κλίσης b για 100 πραγματοποιήσεις. Στο Πίνακα 2-1 συνοψίζονται τα διαστήματα τιμών της στατιστικής (2.2) με τις αντίστοιχες συχνότητές τους που προέκυψαν από την προσομοίωση των Granger και Newbold (1974).

Ο έλεγχος με τη στατιστική (2.2) σε επίπεδο σημαντικότητας 5% απορρίπτει τη μηδενική υπόθεση της μη ύπαρξης σχέσεως μεταξύ των δύο ανεξάρτητων χρονοσειρών εσφαλμένα περίπου στα τρία τέταρτα των περιπτώσεων. Εάν η (2.2) κατανεμόταν ως $N(0, 1)$ η

προσδοκώμενη μέση τιμή της θα ήταν $\sqrt{2/p} \approx 0.8$ που διαφέρει σημαντικά από τη μέση παρατηρούμενη τιμή που είναι 4.5, κάτι που σημαίνει ότι η τυπική απόκλιση του εκτιμητή \hat{b} υποεκτιμάτε κατά 5.6 φορές. Επομένως, για να εκτελείται σωστά ο έλεγχος και να είναι ένας συντελεστής σημαντικός σε επίπεδο 5% θα πρέπει η κριτική τιμή που θα χρησιμοποιηθεί να είναι 11,2 και όχι κάποια κοντά στο 2.

Πίνακας 2-1
Παλινδρόμηση δύο ανεξαρτήτων τυχαίων περιπάτων

$ t $	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8
Συχνότητα	13	10	11	13	18	8	8	5
$ t $	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16
Συχνότητα	3	3	1	5	0	1	0	1

Πηγή: Granger και Newbold (1974)

Στη δεύτερη προσομοίωση επεκτάθηκαν σε υπόδειγμα πολλαπλής παλινδρόμησης εμπλέκοντας από μία έως πέντε ανεξάρτητες χρονοσειρές. Όλες οι χρονοσειρές που χρησιμοποιήσαν ακολουθούσαν το ίδιο υπόδειγμα. Τα υποδείγματα που μελετήθηκαν ήταν:

- I. Τυχαίοι περίπατοι (*random walks*).
- II. Λευκοί θόρυβοι (*white noises*), δηλαδή διαφορές των τυχαίων περιπάτων.
- III. ARIMA(0, 1, 1).
- IV. Διαφορές των ARIMA(0, 1, 1), δηλαδή MA(1) διαδικασίες.

Το μέγεθος των χρονοσειρών και ο αριθμός των επαναλήψεων παρέμειναν τα ίδια όπως και στην πρώτη προσομοίωση, ενώ τα σφάλματα προέρχονταν από $N(0, 1)$ κατανομή. Ο Πίνακας 2-2 περιέχει τα ποσοστά απόρριψης της μηδενικής υπόθεσης για τον έλεγχο της ταυτόχρονης στατιστικής σημαντικότητας των συντελεστών του υποδείγματος με τη στατιστική (2.3), τη μέση τιμή της στατιστικής των Durbin και Watson, τη μέση τιμή του προσαρμοσμένου συντελεστή προσδιορισμού και το ποσοστό των τιμών του προσαρμοσμένου συντελεστή προσδιορισμού που είναι μεγαλύτερες από 0.7.

Πίνακας 2-2

Πολλαπλές παλινδρομήσεις με μία έως πέντε ανεξάρτητες μεταβλητές

	Αριθμός ανεξάρτητων μεταβλητών	Ποσοστό απόρριψης της H_0 , ($\alpha = 5\%$)	Μέση τιμή στατιστικής των Durbin - Watson	Μέσο \bar{R}^2	Ποσοστό $\bar{R}^2 > 0.7$
Τυχαίοι περίπατοι					
Επίπεδα	1	76	0.32	0.26	5
	2	78	0.46	0.34	8
	3	93	0.55	0.46	25
	4	95	0.74	0.55	34
	5	96	0.88	0.59	37
Διαφορές	1	8	2.00	0.004	0
	2	4	1.99	0.001	0
	3	2	1.91	-0.007	0
	4	10	2.01	0.006	0
	5	6	1.99	0.012	0
ARIMA(0, 1, 1)					
Επίπεδα	1	64	0.73	0.20	3
	2	81	0.96	0.30	7
	3	82	1.09	0.37	11
	4	90	1.14	0.44	9
	5	90	1.26	0.45	19
Διαφορές	1	8	2.58	0.003	0
	2	12	2.57	0.01	0
	3	7	2.53	0.005	0
	4	9	2.53	0.025	0
	5	13	2.54	0.027	0

Πηγή: Granger και Newbold (1974)

Από τα αποτελέσματα της προσομοίωσης εξάγεται το συμπέρασμα ότι όταν χρησιμοποιούνται ανεξάρτητοι τυχαίοι περίπατοι, το ποσοστό απόρριψης της υπόθεσης της

μη ύπαρξης σχέσης αυξάνεται καθώς αυξάνεται ο αριθμός των ανεξαρτήτων μεταβλητών που περιλαμβάνονται στην παλινδρόμηση. Ειδικότερα, όταν περιλαμβάνονται περισσότερες από τρεις ανεξάρτητες μεταβλητές το ποσοστό αυτό γίνεται πολύ μεγάλο και ουσιαστικά προσεγγίζει το 100%. Η μέση τιμή του προσαρμοσμένου συντελεστή προσδιορισμού \bar{R}^2 αυξάνει σταθερά, καθώς αυξάνεται ο αριθμός των ανεξαρτήτων μεταβλητών, κάτι που συμβαίνει και με τη μέση τιμή της στατιστικής των Durbin και Watson. Παρόμοια συμπεράσματα προκύπτουν και όταν χρησιμοποιούνται ARIMA(0, 1, 1) διαδικασίες. Αντίθετα, όταν συσχετίζονται οι διαφορές των τυχαίων περιπάτων, τα σφάλματα είναι λευκός θόρυβος και το ποσοστό απόρριψης είναι κοντά στο 5% που προβλέπεται από τη στατιστική θεωρία. Όμως, όταν συσχετίζονται οι διαφορές των ARIMA(0, 1, 1) διαδικασιών, δηλαδή MA(1) διαδικασίες, το πρόβλημα, αν και σε μικρότερη ένταση, παραμένει γιατί τα σφάλματα δεν είναι λευκός θόρυβος και η μηδενική υπόθεση απορρίπτεται περίπου με διπλάσιο ποσοστό από ότι θα έπρεπε.

Επομένως, το συμπέρασμα που κατέληξαν οι Granger και Newbold (1974) είναι ότι όταν οι μεταβλητές είναι τυχαίοι περίπατοι ή κοντά σε τυχαίους περιπάτους και χρησιμοποιηθούν σε εξισώσεις παλινδρόμησης, ενώ στην πραγματικότητα δεν θα έπρεπε, τότε θα είναι ο κανόνας παρά η εξαίρεση η εύρεση υποφαινομενικών σχέσεων. Κατά συνέπεια, ο συνδυασμός ενός υψηλού R^2 ή \bar{R}^2 και μίας πολύ χαμηλής τιμής της στατιστικής των Durbin και Watson είναι ένδειξη μη αληθινής σχέσης μεταξύ των μεταβλητών.

Η προσομοίωση των Granger και Newbold για το απλό γραμμικό υπόδειγμα επαναλήφθηκε για διαφορετικά μεγέθη χρονοσειρών με σκοπό να εξαχθούν περαιτέρω συμπεράσματα και να μελετηθεί η μεταβολή των παραμέτρων καθώς αυξάνεται το μέγεθος του δείγματος. Οι χρονοσειρές που εξετάστηκαν ήταν ανεξάρτητοι τυχαίοι περίπατοι 15, 30, 50, 100, 500 και 1000 παρατηρήσεων με τους διαταρακτικούς όρους να προέρχονται από $N(0, 1)$ κατανομή. Η προσομοίωση πραγματοποιήθηκε σε Mathematica 4.0 και σε κάθε περίπτωση εκτελέστηκαν 10000 επαναλήψεις. Οι παράμετροι που καταγράφηκαν ήταν η μέση απόλυτη τιμή της στατιστικής (2.2) με το αντίστοιχο ποσοστό απόρριψης της μηδενικής υπόθεσης της μη στατιστικής σημαντικότητας του συντελεστή b , καθώς και η μέση τιμή του συντελεστή προσδιορισμού με την ελάχιστη και μέγιστη τιμή του. Επίσης, εξετάστηκαν οι παρενέργειες που προκαλούνται από την αυτοσυσχέτιση στα σφάλματα καταγράφοντας την ελάχιστη, μέγιστη και μέση τιμή της στατιστικής των Durbin και Watson που παρατηρήθηκαν και το ποσοστό των περιπτώσεων όπου ταυτόχρονα απορρίπτεται η μηδενική

υπόθεση και τα σφάλματα με βάση τον έλεγχο των Durbin και Watson αυτοσυσχετίζονται. Τέλος, εξετάστηκε και η περίπτωση όπου ταυτόχρονα απορρίπτεται η μηδενική υπόθεση, ο συντελεστής προσδιορισμού είναι μεγαλύτερος από 0.7 και τα σφάλματα αυτοσυσχετίζονται. Το πρόβλημα της αβέβαιης περιοχής της κατανομής των Durbin και Watson αντιμετωπίστηκε χρησιμοποιώντας ως μοναδική κριτική τιμή για την πραγματοποίηση του ελέγχου την άνω κριτική τιμή d_U κάτι που ισχύει και για τις επόμενες προσομοιώσεις που θα πραγματοποιηθούν σε αυτή την εργασία. Τα αποτελέσματα των προσομοιώσεων συνοψίζονται στο Πίνακα 2-3.

Πίνακας 2-3
Υποφαινόμενικές παλινδρομήσεις για διάφορα μεγέθη δείγματος

Μέγεθος δείγματος	15	30	50	100	500	1000
Ποσοστό απόρριψης H_0 , ($\alpha = 5\%$)	39.4	56.87	65.21	76.2	89.16	92.49
Μέση τιμή $ t $	2.14	3.04	3.87	5.68	12.64	18.16
Μέσο R^2	0.25	0.24	0.23	0.24	0.24	0.24
Ελάχιστο R^2	0	0	0	0	0	0
Μέγιστο R^2	0.95	0.95	0.94	0.95	0.95	0.94
Μέση τιμή DW	0.91	0.523	0.332	0.176	0.036	0.018
Ελάχιστο DW	0.08	0.029	0.015	0.010	0.002	0.001
Μέγιστο DW	3.195	2.184	1.521	0.774	0.181	0.097
Ποσοστό απόρριψης H_0 και $DW < d_U$	28.54	56.31	65.21	76.2	89.16	92.49
Ποσοστό απόρριψης H_0 και $DW < d_U$ και $R^2 > 0.7$	2.6	4.64	4.35	4.58	4.34	4.86

Τα συμπεράσματα που προκύπτουν από τις προσομοιώσεις είναι:

- 1) Η μέση απόλυτη τιμή της στατιστικής (2.2) και κατά συνέπεια το ποσοστό απόρριψης της μηδενικής υπόθεσης του ελέγχου της στατιστικής σημαντικότητας του συντελεστή b , αυξάνεται σημαντικά καθώς αυξάνεται το μέγεθος του

δείγματος. Για χρονοσειρές λίγων παρατηρήσεων το ποσοστό αυτό είναι κοντά στο 40% ενώ για πολλές παρατηρήσεις ξεπερνάει το 90%.

- 2) Η μέση τιμή του συντελεστή προσδιορισμού παραμένει σταθερή κοντά στο 0.24 ανεξάρτητα του μεγέθους του δείγματος. Τα ίδια ισχύουν και για την ελάχιστη και μέγιστη τιμή του που παραμένουν για όλες τις περιπτώσεις κοντά στο 0 και στο 0.95 αντίστοιχα.
- 3) Η μέση τιμή της στατιστικής των Durbin και Watson καθώς και η ελάχιστη και μέγιστη τιμή της μειώνονται καθώς αυξάνεται το μέγεθος του δείγματος. Όταν το μέγεθος του δείγματος γίνεται πολύ μεγάλο οι τιμές αυτές ουσιαστικά μηδενίζονται.
- 4) Το ποσοστό των περιπτώσεων όπου ταυτόχρονα απορρίπτεται η μηδενική υπόθεση και τα σφάλματα αυτοσυσχετίζονται ταυτίζεται με το ποσοστό απόρριψης της μηδενικής υπόθεσης. Από αυτό συμπεραίνεται ότι το πρόβλημα της αυτοσυσχέτισης των σφαλμάτων είναι που οδηγεί στην απόρριψη της μηδενικής υπόθεσης. Μικρή εξαίρεση εντοπίζεται για μεγέθη δειγμάτων 15 και 30 κάτι που προφανώς οφείλεται στην αδυναμία του ελέγχου των Durbin και Watson στα μικρά σε μέγεθος δείγματα.
- 5) Το ποσοστό των περιπτώσεων όπου ταυτόχρονα απορρίπτεται η μηδενική υπόθεση, τα σφάλματα αυτοσυσχετίζονται και ο συντελεστής προσδιορισμού είναι μεγαλύτερος από 0.7 είναι για όλες τις περιπτώσεις κοντά στο 5%.

Τα αποτελέσματα των προσομοιώσεων εξηγήθηκαν μαθηματικά από τον Phillips (1986) ο οποίος ανέπτυξε μια ασυμπτωτική θεωρία για την παλινδρόμηση όταν οι συσχετιζόμενες χρονοσειρές είναι ολοκληρωμένες διαδικασίες. Έστω πάλι το υπόδειγμα της απλής παλινδρόμησης (2.4) με τις Y_t και X_t να παράγονται από τους ανεξάρτητους τυχαίους περιπάτους:

$$Y_t = Y_{t-1} + v_t \text{ και } X_t = X_{t-1} + w_t \text{ για } t = 1, 2, \mathbf{KT} \quad (2.5)$$

όπου η v_t είναι $iid(0, \mathbf{s}_v^2)$, η w_t είναι $iid(0, \mathbf{s}_w^2)$ και είναι ανεξάρτητες μεταξύ τους.

Επιπροσθέτως, $\{\mathbf{x}_j\}_1^\infty$ είναι μια ακολουθία από τυχαία διανύσματα ορισμένα στο χώρο

πιθανότητας (Ω, B, P) και $S_t = \sum_{j=1}^t \mathbf{x}_j$ είναι μια διαδικασία μερικού αθροίσματος με αρχική

τιμή $S_0 = 0$. Η βασική υπόθεση που έγινε από τον Phillips (1986) είναι ότι:

- 1) $E(x_t) = 0$ για όλα τα t .
- 2) $\sup_{i,t} E|x_{i,t}|^{b+e} < \infty$ για κάποια $b > 2$ και $e > 0$.
- 3) Το $\Sigma = \lim_{T \rightarrow \infty} T^{-1} E(S_T S_T')$ υπάρχει και είναι θετικά ορισμένο.
- 4) Η $\{x_t\}_1^\infty$ είναι μια ισχυρή μίξη με ακολουθία μίξης a_m που ικανοποιεί τη συνθήκη

$$\sum_1^\infty a_m^{1-2/b} < \infty.$$

Εάν $n = 2$ και $x'_t = (v_t, w_t)$, οι συνθήκες που συνεπάγονται από τη παραπάνω υπόθεση για τα innovations της (2.5) είναι ασθενείς και επιτρέπουν οι Y_t και οι X_t να είναι γενικές ολοκληρωμένες διαδικασίες πρώτης τάξεως με τις διαφορές τους να είναι ασθενώς εξαρτημένα innovations. Έτσι, το σύνολο των δυνατών μηχανισμών παραγωγής των παρατηρήσεων είναι μεγάλο και μπορεί να συμπεριλάβει και το ARIMA(p, I, q) υπόδειγμα. Με τη συνθήκη (2) της υπόθεσης ελέγχεται η επιτρεπόμενη ετερογένεια της διαδικασίας, ενώ η συνθήκη αθροιστικότητας (4) ελέγχει το βαθμό ανοχής προσωρινής εξάρτησης στη διαδικασία σε σχέση με τη πιθανότητα ύπαρξης απομονωμένων τιμών και ικανοποιείται όταν ο ρυθμός μείωσης (*decay*) της μίξης είναι $a_m = O(m^{-l})$ για κάποιο $l > b/(b-2)$. Καθώς το b προσεγγίζει το δύο και αυξάνει η πιθανότητα εμφάνισης απομονωμένων παρατηρήσεων, αυξάνει και ο ρυθμός ελάττωσης της μίξης και για το λόγο αυτό, εξαιτίας της συνθήκης της μίξης (4), μειώνεται με γρηγορότερο ρυθμό η επίδραση των απομονωμένων παρατηρήσεων. Επιπλέον, εάν η $\{x_t\}$ είναι ασθενώς στάσιμη τότε ισχύει $\Sigma = E(x_t x_t') + \sum_{k=1}^\infty (x_t x_{t-k}' + x_{t-k} x_t')$ και η σύγκλιση αυτής της σειράς συνεπάγεται από τη συνθήκη μίξης (4) ενώ, όταν $x'_t = (v_t, w_t)$ και τα v_t, w_t είναι ανεξάρτητα, όπως στη περίπτωση της υποφαινομενικής παλινδρόμησης, ισχύει

$$\Sigma = \begin{pmatrix} S_v^2 & 0 \\ 0 & S_w^2 \end{pmatrix} \text{ με } S_v^2 = \lim_{T \rightarrow \infty} T^{-1} E(P_T^2), S_w^2 = \lim_{T \rightarrow \infty} T^{-1} E(Q_T^2) \text{ και } P_t = \sum_1^t v_j, Q_t = \sum_1^t w_j.$$

Στη συνέχεια ο Phillips διατυπώνει το επόμενο λήμμα, το οποίο είναι χρήσιμο για την ανάπτυξη της σωστής ασυμπτωτικής θεωρίας. Οι αποδείξεις του λήμματος αλλά και των αποτελεσμάτων που ακολουθούν και που δίνουν την ασυμπτωτική θεωρία βασίζονται στο συναρτησιακό κεντρικό οριακό θεώρημα και βρίσκονται στην εργασία του Phillips (1986).

Λήμμα. Έστω ότι οι $\{Y_t\}_1^\infty$ και $\{X_t\}_1^\infty$ παράγονται από τους ανεξάρτητους τυχαίους περιπάτους (2.5). Αν οι ακολουθίες των innovations $\{v_t\}_1^\infty$ και $\{w_t\}_1^\infty$ είναι ανεξάρτητες και αν η $\{(v_t, w_t)\}_1^\infty$ ικανοποιεί τη βασική υπόθεση, τότε καθώς $T \uparrow \infty$ ισχύουν:

$$\begin{aligned}
1) \quad & T^{-3/2} \sum_1^T X_t \Rightarrow s_w \int_0^1 W(t) dt \quad \text{και} \quad T^{-3/2} \sum_1^T Y_t \Rightarrow s_v \int_0^1 V(t) dt \\
2) \quad & T^{-2} \sum_1^T X_t^2 \Rightarrow s_w^2 \int_0^1 W(t)^2 dt \quad \text{και} \quad T^{-2} \sum_1^T Y_t^2 \Rightarrow s_v^2 \int_0^1 V(t)^2 dt \\
3) \quad & T^{-2} \sum_1^T (X_t - \bar{X})^2 \Rightarrow s_w^2 \left[\int_0^1 W(t)^2 dt - \left\{ \int_0^1 W(t) dt \right\}^2 \right] \\
& T^{-2} \sum_1^T (Y_t - \bar{Y})^2 \Rightarrow s_v^2 \left[\int_0^1 V(t)^2 dt - \left\{ \int_0^1 V(t) dt \right\}^2 \right] \\
4) \quad & T^{-2} \sum_1^T Y_t X_t \Rightarrow s_v s_w \int_0^1 V(t) W(t) dt \\
5) \quad & T^{-1} \sum_r^T Y_t (Y_t - Y_{t-r}) \Rightarrow \left(\frac{r}{2} \right) \{ s_v^2 V(1)^2 + \Omega_{v0} \} + \sum_{j=1}^r (r-j) \Omega_{vj} \\
& T^{-1} \sum_r^T X_t (X_t - X_{t-r}) \Rightarrow \left(\frac{r}{2} \right) \{ s_w^2 W(1)^2 + \Omega_{w0} \} + \sum_{j=1}^r (r-j) \Omega_{wj} \\
6) \quad & T^{-1} \sum_r^T Y_t (X_t - X_{t-r}) + T^{-1} \sum_r^T X_t (Y_t - Y_{t-r}) \Rightarrow r s_v s_w V(1) W(1)
\end{aligned}$$

όπου οι $W(t)$ και $V(t)$ είναι ανεξάρτητες διαδικασίες Wiener στο $C[0, 1]$ και όπου:

$$\Omega_{vj} = \lim_{T \rightarrow \infty} T^{-1} \sum_{j+1}^T E(v_t v_{t-j}) \quad \text{για} \quad j = 0, 1, \mathbf{K}$$

και

$$\Omega_{wj} = \lim_{T \rightarrow \infty} T^{-1} \sum_{j+1}^T E(w_t w_{t-j}) \quad \text{για} \quad j = 0, 1, \mathbf{K}$$

αντίστοιχα για τα Ω_{vj} και Ω_{wj} .

Τα παραπάνω αποτελέσματα ισχύουν ανεξάρτητα από τις αρχικές συνθήκες Y_0 και X_0 που ορίζονται για τις Y και X . Με $C[0, 1]$ συμβολίζεται το διάστημα όλων των συνεχών συναρτήσεων πραγματικών τιμών στο διάστημα $[0, 1]$. Οι διαδικασίες $W(t)$ και $V(t)$ είναι στοχαστικά ανεξάρτητες με τα δειγματικά τους μονοπάτια να βρίσκονται στο $C[0, 1]$. Με

βάση τα προηγούμενα αποτελέσματα εξασφαλίζεται ότι οι κατάλληλα τυποποιημένες δειγματικές ροπές των ακολουθιών $\{Y_t\}_1^\infty$ και $\{X_t\}_1^\infty$ συγκλίνουν ασθενώς σε σωστά ορισμένες συναρτησιακές διαδικασίες των διαδικασιών Wiener $W(t)$ και $V(t)$ που έχουν καλά ορισμένες μη εκφυλισμένες κατανομές. Για παράδειγμα, η $T^{-3/2} \sum_1^T X_t$ συγκλίνει κατά κατανομή στη κατανομή της συναρτησιακής $s_w \int_0^1 W(t) dt$ της διαδικασίας Wiener στο $C[0, 1]$. Επίσης, επειδή η $W(t)$ είναι Gaussian με μέσο όρο 0 και ανεξάρτητες προσυζήσεις η οριακή κατανομή της $T^{-3/2} \sum_1^T X_t$ είναι κανονική με μέσο όρο 0 και διακύμανση που δίνεται από τη σχέση:

$$s_w^2 E \left\{ \int_0^1 \int_0^1 W(t)W(s) dt ds \right\} = 2s_w^2 \int_0^1 \int_0^r E \{W(r)W(s)\} ds dr = 2s_w^2 \int_0^1 \int_0^r s ds dr = s_w^2 / 3 .$$

Εάν υποτεθεί ότι η (2.4) εκτιμάται με τη μέθοδο των ελαχίστων τετραγώνων και ικανοποιούνται οι συνθήκες του λήμματος τότε καθώς $T \uparrow \infty$ ο Phillips αποδεικνύει ότι ισχύουν:

$$1) \hat{b} = \frac{s_v \left\{ \int_0^1 V(t)W(t) dt - \int_0^1 V(t) dt \int_0^1 W(t) dt \right\}}{s_w \left\{ \int_0^1 W(t)^2 dt - \left(\int_0^1 W(t) dt \right)^2 \right\}} = \left(\frac{s_v}{s_w} \right) z$$

και

$$T^{-1/2} \hat{a} \Rightarrow s_v \left\{ \int_0^1 V(t) dt - z \int_0^1 W(t) dt \right\} .$$

Δηλαδή, σε αντίθεση με τη συνηθισμένη ασυμπτωτική θεωρία της παλινδρόμησης, οι συντελεστές \hat{a} και \hat{b} δεν συγκλίνουν κατά πιθανότητα σε σταθερές καθώς $T \uparrow \infty$ αλλά ο \hat{b} έχει μια μη εκφυλισμένη οριακή κατανομή και η κατανομή του \hat{a} αποκλίνει καθώς $T \uparrow \infty$. Κατά συνέπεια η αβεβαιότητα για το υπόδειγμα (2.4) που οφείλεται στο γεγονός ότι οι χρησιμοποιούμενες χρονοσειρές είναι τυχαίοι περίπατοι, παραμένει ασυμπτωτικά στις οριακές κατανομές. Αντίθεση με τη συνηθισμένη θεωρία παλινδρόμησης υπάρχει και όταν οι Y_t και X_t παράγονται από ανεξάρτητες διαρκείς

(stable) αυτοπαλίνδρομες διαδικασίες όπου τα \hat{a} και \hat{b} συγκλίνουν κατά πιθανότητα στο μηδέν. Αυτό οφείλεται στο γεγονός ότι οι διαδικασίες Y_t και X_t δεν είναι εργοδικές (non – ergodic). Οι δειγματικές ροπές των Y_t και X_t και οι από κοινού δειγματικές ροπές τους, δεν συγκλίνουν σε σταθερές, όπως συμβαίνει με εργοδικές διαδικασίες, αλλά με κατάλληλη τυποποίηση συγκλίνουν ασθενώς σε τυχαίες μεταβλητές.

$$2) T^{-1/2}t_b \Rightarrow \frac{m}{n^{1/2}}$$

και

$$T^{-1/2}t_a \Rightarrow \frac{\left\{ \int_0^1 V(t)dt - z \int_0^1 W(t)dt \right\} \left\{ \int_0^1 W(t)^2 dt - \left(\int_0^1 W(t)dt \right)^2 \right\}}{\left[n \int_0^1 W(t)^2 dt \right]^{1/2}}$$

όπου:

$$m = \int_0^1 V(t)W(t)dt - \int_0^1 V(t)dt \int_0^1 W(t)dt,$$

$$n = \left\{ \int_0^1 V(t)^2 dt - \left(\int_0^1 V(t)dt \right)^2 \right\} \left\{ \int_0^1 W(t)^2 dt - \left(\int_0^1 W(t)dt \right)^2 \right\} - \left\{ \int_0^1 V(t)W(t)dt - \int_0^1 V(t)dt \int_0^1 W(t)dt \right\}^2.$$

Δηλαδή, οι στατιστικές συναρτήσεις (2.2) για τους ελέγχους της στατιστικής σημαντικότητας των συντελεστών του σταθερού όρου a και της κλίσης b του υποδείγματος (2.4) δεν έχουν οριακές κατανομές αλλά οι κατανομές τους αποκλίνουν καθώς $T \uparrow \infty$, έτσι ώστε να μην υπάρχουν ασυμπτωτικά σωστές τιμές για αυτούς τους ελέγχους. Το ποσοστό απόρριψης αυτών των ελέγχων, όταν οι κριτικές τους τιμές προέρχονται από τη συμβατική ασυμπτωτική θεωρία, συνεχίζει να αυξάνει καθώς αυξάνει το μέγεθος του δείγματος. Αυτό βρίσκεται σε αρμονία με τα αποτελέσματα των παραπάνω προσομοιώσεων. Οι Granger και Newbold είχαν προτείνει τη χρησιμοποίηση μιας κριτικής τιμής 11.2 αντί για τη 1.96, όταν εκτιμάται η σημαντικότητα του συντελεστή b σε επίπεδο 5%. Ο Phillips διαφωνεί με τη

χρησιμοποίηση της προηγούμενης κριτικής τιμής. Σύμφωνα με την ασυμπτωτική

θεωρία που ο ίδιος ανέπτυξε, η σωστά τυποποιημένη τιμή είναι $t'_b = \frac{t_b}{\sqrt{T}}$.

$$3) R^2 \Rightarrow \frac{z^2 \left\{ \int_0^1 W(t)^2 dt - \left(\int_0^1 W(t) dt \right)^2 \right\}}{\int_0^1 V(t)^2 dt - \left(\int_0^1 V(t) dt \right)^2},$$

$$DW \xrightarrow{p} 0,$$

$$TDW \Rightarrow \left\{ \left(\Omega_{v0}/s_v^2 \right) + z^2 \left(\Omega_{w0}/s_w^2 \right) \right\} \left[\int_0^1 V(t)^2 dt - \left(\int_0^1 V(t) dt \right)^2 \right] - z^2 \left\{ \int_0^1 W(t)^2 dt - \left(\int_0^1 W(t) dt \right)^2 \right\}.$$

Δηλαδή, ο συντελεστής προσδιορισμού R^2 έχει μη εκφυλισμένη οριακή κατανομή, ενώ η στατιστική των Durbin και Watson συγκλίνει κατά πιθανότητα στο μηδέν. Επίσης, ο Phillips κατασκευάζει και την ασυμπτωτική κατανομή της τυποποιημένης στατιστικής των Durbin και Watson, TDW. Επομένως, σε περιπτώσεις υποφαινομενικών παλινδρομήσεων πρέπει να αναμένονται χαμηλές τιμές για τη στατιστική των Durbin και Watson και μέτριες τιμές για το συντελεστή προσδιορισμού R^2 .

4) Για όλα τα σταθερά $s \geq 1$ ισχύουν $T(r_s - 1) \Rightarrow -\frac{A_s}{B}$ και $r_s = 1 + O_p(T^{-1})$ όπου:

$$A_s = \left(\frac{s}{2} \right) \left[\left\{ V(1) - zW(1) \right\} - \left\{ \int_0^1 V(t) dt - z \int_0^1 W(t) dt \right\} \right]^2 + \left(\frac{s}{2} \right) \left\{ \int_0^1 V(t) dt - z \int_0^1 W(t) dt \right\}^2 + \left\{ s \Omega_{v0}/2s_v^2 + \sum_{j=1}^s (s-j) \Omega_{vj}/s_v^2 \right\} + z^2 \left\{ s \Omega_{w0}/2s_w^2 + \sum_{j=1}^s (s-j) \Omega_{wj}/s_w^2 \right\}$$

και

$$B = \int_0^1 V(t)^2 dt - \left(\int_0^1 V(t) dt \right)^2 - z^2 \left\{ \int_0^1 W(t)^2 dt - \left(\int_0^1 W(t) dt \right)^2 \right\},$$

$$T^{-1}Q_k = \sum_{s=1}^k r_s^2 \xrightarrow{p} k.$$

Δηλαδή, οι συντελεστές σειριακής συσχέτισης για τα κατάλοιπα της παλινδρόμησης συγκλίνουν κατά πιθανότητα στη μονάδα, ενώ η κατανομή της στατιστικής Q_K των Box και Pierce αποκλίνει καθώς $T \uparrow \infty$. Τέλος, ο Phillips υπολογίζει και την οριακή κατανομή του τυποποιημένου συντελεστού $T(r_s - 1)$.

Ο Phillips επεκτείνει τα προηγούμενα αποτελέσματα και για τη περίπτωση του υποδείγματος της πολλαπλής γραμμικής παλινδρόμησης της μορφής:

$$Y_t = \hat{a} + \hat{b}'X_t + u_t, \quad t=1, \mathbf{K}, T \quad (2.6)$$

όπου X_t είναι ένα m – διάνυσμα από ολοκληρωμένες διαδικασίες πρώτης τάξεως. Δεν είναι αναγκαίο τα Y_t και X_t να είναι ανεξάρτητα. Απαραίτητη προϋπόθεση είναι να μην υπάρχει ένας γραμμικός συνδυασμός των διανυσμάτων χρονοσειρών (Y_t, X_t') ο οποίος να είναι ολοκληρωμένος μηδενικής τάξης δηλαδή, να είναι μια στάσιμη διαδικασία. Για την παρακάτω ανάλυση, ο Phillips έθεσε $Z_t' = (Y_t, X_t')$ και θεώρησε ότι η Z_t' είναι μια διανυσματική ολοκληρωμένη διαδικασία διαστάσεως $n = m+1$ της οποίας ο μηχανισμός παραγωγής είναι η:

$$Z_t = Z_{t-1} + x_t, \quad t=1, 2, \mathbf{K} \quad (2.7)$$

Η διαδικασία $\{x_t\}_1^\infty$ στη σχέση (2.7) απαιτείται να ικανοποιεί τη βασική υπόθεση και θεωρούνται για αρχικές συνθήκες είτε ότι το Z_0 είναι σταθερό με πιθανότητα μονάδα είτε ότι είναι τυχαίο με συγκεκριμένη κατανομή.

Εάν υποθεθεί ότι η (2.6) εκτιμάται με τη μέθοδο των ελαχίστων τετραγώνων, η Z_t παράγεται από την (2.7) και διαδικασία $\{x_t\}_1^\infty$ ικανοποιεί τη βασική υπόθεση τότε, καθώς $T \uparrow \infty$ ο Phillips αποδεικνύει ότι ισχύουν:

$$1) \hat{b} \Rightarrow A_{22}^{-1}a_{21}$$

και

$$T^{-1/2}\hat{a} \Rightarrow b_1 - b_2'A_{22}^{-1}a_{21}.$$

Δηλαδή, οι συντελεστές \hat{a} και \hat{b} δεν συγκλίνουν κατά πιθανότητα σε σταθερές καθώς $T \uparrow \infty$ αλλά ο \hat{b} έχει μια μη εκφυλισμένη οριακή κατανομή και η κατανομή του \hat{a} αποκλίνει καθώς $T \uparrow \infty$. Τα συμπεράσματα αυτό είναι παρόμοια με τη περίπτωση της απλής γραμμικής παλινδρόμησης.

$$2) R^2 \Rightarrow a_{21}'A_{22}^{-1}a_{21}/a_{11}$$

και

$$TDW \Rightarrow h'\Sigma_x h / \{a_{11} - a_{21}'A_{22}^{-1}a_{21}\} \text{ όπου}$$

$$A = \begin{bmatrix} a_{11} & a_{21}' \\ a_{21} & A_{22} \end{bmatrix} \begin{matrix} 1 \\ m \end{matrix} = \Sigma^{1/2} \left\{ \int_0^1 Z(t)Z(t)'dt - \int_0^1 Z(t)dt \int_0^1 Z(t)'dt \right\} \Sigma^{1/2}$$

$b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \frac{1}{m} = \Sigma^{1/2} \int_0^1 Z(t) dt$ και $h' = (1, -a'_{21} A_{22}^{-1})$ με τη $Z(t)$ να είναι μια διανυσματική

διαδικασία Wiener στο C^n , το Σ ορίζεται στη βασική υπόθεση και

$$\Sigma_x = \lim_{T \rightarrow \infty} T^{-1} \sum_1^T E(x_t x_t').$$

Δηλαδή, όπως στην περίπτωση της απλής γραμμικής παλινδρόμησης, η στατιστική των Durbin και Watson συγκλίνει κατά πιθανότητα στο μηδέν και ο συντελεστής προσδιορισμού R^2 έχει μη εκφυλισμένη οριακή κατανομή. Επομένως, αναμένονται και πάλι χαμηλές τιμές για τη στατιστική των Durbin και Watson και μέτριες για το R^2 .

$$3) T^{-1} F_b \Rightarrow \left(\frac{1}{m} \right) a'_{21} A_{22}^{-1} a_{21} / \{ a_{11} - a'_{21} A_{22}^{-1} a_{21} \}$$

και

$$T^{-1/2} t_{bi} \Rightarrow (A_{22}^{-1} a_{21})_i / \left\{ (a_{11} - a'_{21} A_{22}^{-1} a_{21})^{1/2} ([A_{22}^{-1}]_{ii})^{1/2} \right\}.$$

Δηλαδή, τόσο η κατανομή της F_b που χρησιμοποιείται για τον έλεγχο της ταυτόχρονης στατιστικής σημαντικότητας των συντελεστών του υποδείγματος όσο και της t_{bi} που χρησιμοποιείται για τον έλεγχο της στατιστικής σημαντικότητας κάθε συντελεστή ξεχωριστά αποκλίνουν καθώς $T \uparrow \infty$ και επομένως, όπως και στην περίπτωση της απλής γραμμικής παλινδρόμησης, δεν υπάρχουν ασυμπτωτικά σωστές κριτικές τιμές για την πραγματοποίηση αυτών των ελέγχων. Η χρήση κριτικών τιμών για αυτούς τους ελέγχους που προέρχονται από τη συμβατική ασυμπτωτική θεωρία, οδηγεί σε αυξανόμενο με το μέγεθος του δείγματος ποσοστό απόρριψης. Ο ρυθμός απόκλισης της κατανομής της F_b είναι $O(T)$ που είναι μεγαλύτερος από το ρυθμό απόκλισης $O(T^{1/2})$ της κατανομής της t_{bi} . Κατά συνέπεια, σε μια παλινδρόμηση με πολλές ανεξάρτητες μεταβλητές το ποσοστό απόρριψης του ελέγχου με τη στατιστική (2.3) πρέπει να είναι μεγαλύτερο από αυτό του ελέγχου με τη στατιστική (2.2) καθώς και από το ποσοστό απόρριψης του ελέγχου από ένα υπόδειγμα με λιγότερες μεταβλητές, κάτι που είναι σύμφωνο με τα αποτελέσματα της προσομοίωσης των Granger και Newbold (1974) όπου για παλινδρομήσεις με ανεξάρτητους τυχαίους περιπάτους παρατήρησαν ποσοστό απόρριψης 76% για μία ανεξάρτητη μεταβλητή

και ποσοστό απόρριψης 96% για πέντε ανεξάρτητες μεταβλητές ενώ για παλινδρομήσεις με ανεξάρτητες ARIMA(0, 1, 1) διαδικασίες τα αντίστοιχα ποσοστά ήταν 64% και 90%.

Τα προηγούμενα συμπεράσματα ισχύουν ανεξάρτητα από το αν οι χρονοσειρές Y_t και X_t είναι ανεξάρτητες ή όχι. Βέβαια, οι ιδιότητες συσχέτισης αυτών των χρονοσειρών έχουν ποσοτικές επιδράσεις στις οριακές κατανομές, οι οποίες εισάγονται μέσω των παραμέτρων των οριακών πινάκων συνδιακυμάνσεων Σ και Σ_x . Όμως, εφόσον ισχύουν οι προϋποθέσεις, οι επιδράσεις αυτές δεν επηρεάζουν τα κύρια ποιοτικά αποτελέσματα. Υπάρχει όμως μια σημαντική περίπτωση όπου οι ιδιότητες συσχέτισης των Y_t και X_t επηρεάζουν τα αποτελέσματα. Στη βασική υπόθεση η συνθήκη (3) απαιτεί ο οριακός πίνακας συνδιακύμανσης Σ να είναι μη ιδιάζων. Αν όμως ο πίνακας Σ είναι ιδιάζων, δηλαδή η ορίζουσά του ισούται με το μηδέν, τότε τα παραπάνω ασυμπτωτικά αποτελέσματα παύουν να ισχύουν και εμφανίζεται το φαινόμενο της συνολοκλήρωσης που έχουν περιγράψει οι Granger και Engle (1985). Επομένως, σύμφωνα με την εργασία του Phillips, απαραίτητη προϋπόθεση για να εμφανίζεται το φαινόμενο της υποφαινομενικής παλινδρόμησης είναι ο πίνακας Σ να είναι μη ιδιάζων, δηλαδή να έχει μη μηδενική ορίζουσα.

2.3 Υποφαινομενικές Παλινδρομήσεις με ολοκληρωμένες διαδικασίες με μη μηδενικό μέσο όρο

Στη προηγούμενη ενότητα εξετάστηκε το πρόβλημα της υποφαινομενικής παλινδρόμησης μεταξύ χρονοσειρών όπου ο μηχανισμός παραγωγής τους ήταν ολοκληρωμένες διαδικασίες με μηδενικό μέσο όρο (*drift*). Ενδιαφέρον έχει να εξεταστεί και η περίπτωση όπου οι εμπλεκόμενες στη παλινδρόμηση χρονοσειρές είναι ολοκληρωμένες διαδικασίες που έχουν μέσο όρο διάφορο του μηδενός.

Έστω ότι οι Y_t και X_t παράγονται σύμφωνα με τους τυχαίους περιπάτους με μέσο όρο (*drift*):

$$Y_t = g_y + Y_{t-1} + v_t \text{ και } X_t = g_x + X_{t-1} + w_t \text{ για } t = 1, 2, \mathbf{K}T \quad (2.8)$$

όπου η v_t είναι $iid N(0, s_v^2)$ και η w_t $iid N(0, s_w^2)$. Μια σειρά προσομοιώσεων πραγματοποιήθηκαν στοχεύοντας στη εξέταση του φαινομένου για τη περίπτωση των

ολοκληρωμένων διαδικασιών με μέσο όρο και τη σύγκρισή του με την περίπτωση των μηδενικών μέσων όρων. Στην πρώτη προσομοίωση και οι δύο μεταβλητές έχουν μη μηδενικό μέσο όρο. Για τις ανάγκες της προσομοίωσης θεωρήθηκαν ότι $g_Y = 400$ και $g_X = 200$. Ο Πίνακας 2–4 περιέχει τα αποτελέσματα της προσομοίωσης.

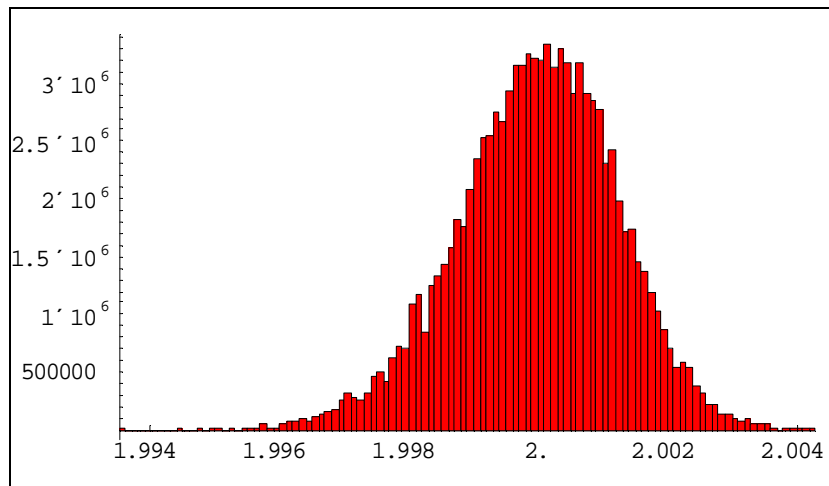
Πίνακας 2-4
Υποφαινομενικές παλινδρομήσεις μεταξύ δύο τυχαίων περιπάτων με μέσο όρο για
διάφορα μεγέθη δείγματος, $\gamma_Y = 400$ και $\gamma_X = 200$

Μέγεθος δείγματος	15	30	50	100	500	1000
Ποσοστό απόρριψης $H_0, (a = 5\%)$	100	100	100	100	100	100
Μέση τιμή $ t $	3224.08	6578.13	11073.7	22325.1	112693	226409
Μέσο \hat{b}	2	2	2	2	2	2
Μέσο R^2	1	1	1	1	1	1
Ελάχιστο R^2	1	1	1	1	1	1
Μέγιστο R^2	1	1	1	1	1	1
Μέση τιμή DW	1.090	0.609	0.381	0.197	0.041	0.021
Ελάχιστο DW	0.219	0.072	0.048	0.021	0.004	0.002
Μέγιστο DW	3.114	2.079	1.495	0.909	0.194	0.088
Ποσοστό απόρριψης H_0 και $DW < d_L$	74.05	98.93	100	100	100	100
Ποσοστό απόρριψης H_0 και $DW < d_L$ και $R^2 > 0.7$	74.05	98.93	100	100	100	100

Τα κύρια συμπεράσματα που προκύπτουν είναι:

- 1) Το ποσοστό απόρριψης του ελέγχου της στατιστικής σημαντικότητας του συντελεστή της κλίσης του υποδείγματος (2.4), που εκτιμάται με τη μέθοδο των ελαχίστων τετραγώνων, με τη στατιστική (2.2) είναι 100% ανεξάρτητα από το μέγεθος των χρονοσειρών.

- 2) Η μέση απόλυτη τιμή της στατιστικής (2.2) αυξάνεται γραμμικά καθώς αυξάνεται το μέγεθος του δείγματος.
- 3) Η μέση, ελάχιστη και μέγιστη τιμή του συντελεστή προσδιορισμού R^2 είναι 1 ανεξάρτητα από το μέγεθος του δείγματος.
- 4) Η μέση, ελάχιστη και μέγιστη τιμή της στατιστικής των Durbin και Watson προσεγγίζουν το 0, καθώς αυξάνεται το μέγεθος του δείγματος.
- 5) Η κατανομή των εκτιμημένων τιμών του \hat{b} είναι συμμετρική με μέση τιμή το 2 δηλαδή, το λόγο $\frac{g_Y}{g_X}$, όπως φαίνεται στο Διάγραμμα 2-1.



Διάγραμμα 2-1
Κατανομή του $\hat{\beta}$ για $\gamma_Y = 400$ και $\gamma_X = 200$, $T = 100$

Όταν ο μέσος όρος υπάρχει μόνο στη μια από τις δυο χρονοσειρές δηλαδή, όταν $g_Y = 0$ και $g_X = 200$ ή $g_Y = 400$ και $g_X = 0$ τα αποτελέσματα αλλάζουν. Οι Πίνακες 2-5 και 2-6 περιέχουν τα αποτελέσματα της προσομοίωσης από όπου συμπεραίνεται πως:

- 1) Το ποσοστό απόρριψης του ελέγχου της στατιστικής σημαντικότητας του συντελεστή της κλίσης και η μέση απόλυτη τιμή της στατιστικής (2.2) είναι παρόμοια και για τις δύο περιπτώσεις και αυξάνονται καθώς αυξάνεται το μέγεθος του δείγματος.
- 2) Η μέση τιμή του συντελεστή προσδιορισμού είναι 0.44, η ελάχιστη τιμή του 0 και η μέγιστη τιμή του 0.99 και για τις δύο περιπτώσεις.

- 3) Η στατιστική των Durbin και Watson προσεγγίζει το 0 καθώς αυξάνεται το μέγεθος του δείγματος και είναι ελαφρά μικρότερη στη δεύτερη περίπτωση.
- 4) Όταν $g_y = 0$ και $g_x = 200$, όπως φαίνεται στο Διάγραμμα 2-2, η κατανομή του \hat{b} παραμένει συμμετρική με μέση τιμή το 0, δηλαδή το λόγο $\frac{g_y}{g_x}$. Όταν $g_y = 400$ και $g_x = 0$, όπως φαίνεται από το Διάγραμμα 2-3, η συμμετρία παραμένει αλλά η κατανομή έχει δυο κορυφές.

Πίνακας 2-5
Υποφαινομενικές παλινδρομήσεις μεταξύ δύο τυχαίων περιπάτων με μέσο όρο για
διάφορα μεγέθη δείγματος, $\gamma_y = 0$ και $\gamma_x = 200$

Μέγεθος δείγματος	15	30	50	100	500	1000
Ποσοστό απόρριψης H_0 , ($\alpha = 5\%$)	64.67	76.16	81.42	87.75	94.31	96.47
Μέση τιμή $ t $	4.10	5.78	7.54	10.80	24.33	34.80
Μέσο \hat{b}	0	0	0	0	0	0
Μέσο R^2	0.44	0.43	0.43	0.43	0.43	0.44
Ελάχιστο R^2	0	0	0	0	0	0
Μέγιστο R^2	0.99	0.98	0.98	0.98	0.99	0.98
Μέση τιμή DW	1.10	0.60	0.38	0.20	0.04	0.02
Ελάχιστο DW	0.200	0.080	0.043	0.021	0.003	0.002
Μέγιστο DW	2.954	2.710	1.656	0.918	0.200	0.102
Ποσοστό απόρριψης H_0 και $DW < d_L$	45.57	75.42	81.4	87.75	94.31	96.47
Ποσοστό απόρριψης H_0 και $DW < d_L$ και $R^2 > 0.7$	16.61	24.66	25.13	25.18	25.39	26.01

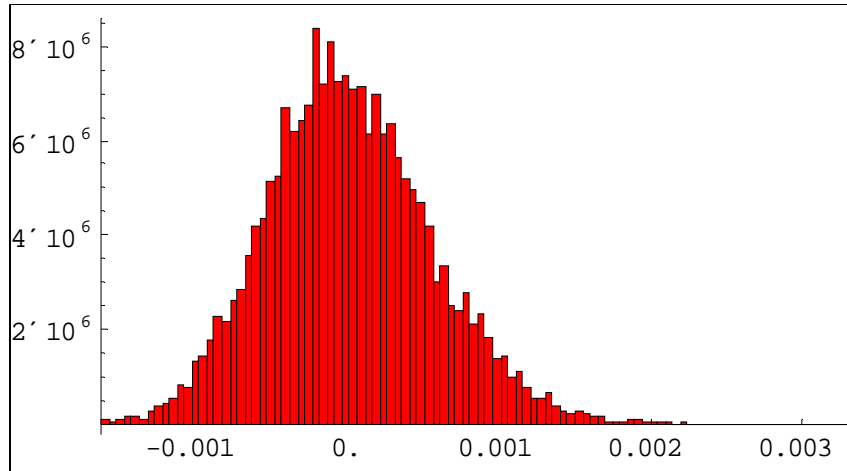
Πίνακας 2-6
Υποφαινομενικές παλινδρομήσεις μεταξύ δύο τυχαίων περιπάτων με μέσο όρο για
διάφορα μεγέθη δείγματος, $\gamma_Y = 400$ και $\gamma_X = 0$

Μέγεθος δείγματος	15	30	50	100	500	1000
Ποσοστό απόρριψης H_0 , ($\alpha = 5\%$)	64.6	76.6	81.79	87.41	94.53	96.04
Μέση τιμή $ t $	4.04	5.91	7.45	10.81	24.59	34.07
Μέσο \hat{b}	-1.58	-45.05	-129.73	-191.42	-240.92	-163.57
Μέσο R^2	0.44	0.44	0.43	0.43	0.44	0.43
Ελάχιστο R^2	0	0	0	0	0	0
Μέγιστο R^2	0.99	0.99	0.99	0.98	0.99	0.98
Μέση τιμή DW	0.540	0.297	0.180	0.095	0.020	0.010
Ελάχιστο DW	0.046	0.012	0.005	0.001	0.000	0.000
Μέγιστο DW	2.868	2.143	1.320	0.838	0.168	0.099
Ποσοστό απόρριψης H_0 και $DW < d_L$	58.39	76.38	81.79	87.41	94.53	96.04
Ποσοστό απόρριψης H_0 και $DW < d_L$ και $R^2 > 0.7$	20.6	26.1	24.81	25.52	25.83	24.45

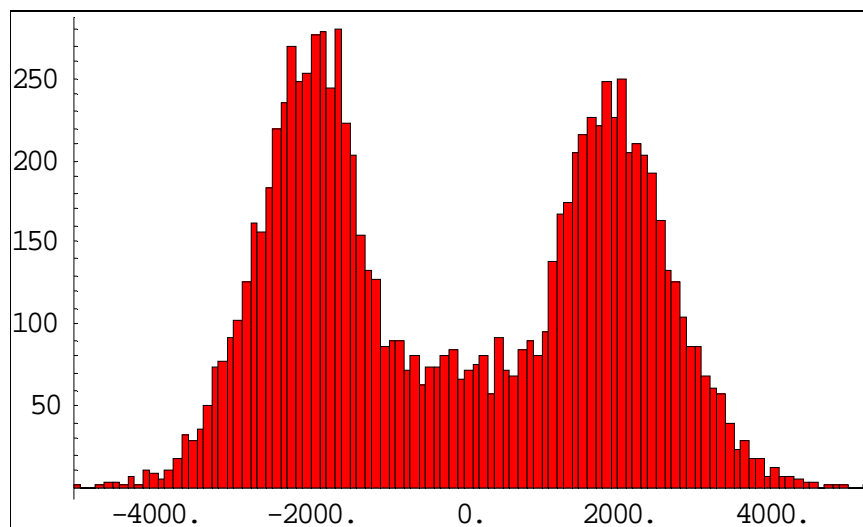
Επίσης, ένα κοινό συμπέρασμα που προέκυψε από τις προσομοιώσεις και για τις τρεις περιπτώσεις που εξετάστηκαν είναι η ισότητα του ποσοστού απόρριψης της μηδενικής υπόθεσης με το ποσοστό όπου ταυτόχρονα απορρίπτεται η μηδενική υπόθεση και τα σφάλματα αυτοσυσχετίζονται. Δηλαδή, όπως και στην περίπτωση των τυχαίων περιπάτων χωρίς μέσους όρους, η αυτοσυσχέτιση στα σφάλματα είναι που οδηγεί στο φαινόμενο της υποφαινομενικής παλινδρόμησης.

Επιπρόσθετα, εξετάστηκε η περίπτωση όπου οι τιμές για τα g_Y και g_X είναι τα ζεύγη 4 και 2, 40 και 20, 400 και 200, 4000 και 2000, 40000 και 20000. Για χρονοσειρές 100 παρατηρήσεων διαπιστώνεται ότι τα αποτελέσματα είναι τα ίδια με μόνη διαφορά στη μέση απόλυτη τιμή της στατιστικής (2.2) όπου αυξάνεται αναλογικά με την αύξηση των g_Y και g_X . Όταν το g_Y είναι 0 και το g_X παίρνει τις τιμές 100, 200, 300, 400, 1000 και 10000

καθώς και όταν ισχύει το αντίθετο τα συμπεράσματα δεν αλλάζουν. Επομένως, όταν μόνο μία από τις δύο χρονοσειρές έχει μη μηδενικό μέσο όρο το μέγεθός του δεν επηρεάζει. Τα αποτελέσματα αυτών των προσομοιώσεων περιλαμβάνονται στο Παράρτημα 1.



Διάγραμμα 2-2
Κατανομή του $\hat{\beta}$ για $\gamma_Y = 0$ και $\gamma_X = 200$, $T = 100$



Διάγραμμα 2-3
Κατανομή του $\hat{\beta}$ για $\gamma_Y = 400$ και $\gamma_X = 0$, $T = 100$

Ο Entorf (1997) εξέτασε τις ασυμπτωτικές ιδιότητες των εκτιμητών της παλινδρόμησης και έδειξε ότι εφόσον ο μηχανισμός παραγωγής των δύο χρονοσειρών είναι η (2.8) με $g_Y \neq 0$ και $g_X \neq 0$ και η (2.4) εκτιμάται με τη μέθοδο των ελαχίστων τετραγώνων τότε:

1) Για την εκτίμηση του συντελεστή της κλίσης \hat{b} ισχύει:

$$\hat{b} \Rightarrow \frac{g_Y}{g_X} \left(1 + \frac{12}{T^{1/2}} \left(\frac{x_Y}{g_Y} - \frac{x_X}{g_X} \right) \right) + \frac{12}{T} \left(\frac{B_{XY}}{g_Y g_X} - \frac{B_{XX}}{g_X^2} \right) + O_p(T^{-3/2})$$

όπου:

- $x_X = \int_0^1 t S_w W(t) dt - \frac{1}{2} \int_0^1 S_w W(t) dt$
- $x_Y = \int_0^1 t S_v V(t) dt - \frac{1}{2} \int_0^1 S_v V(t) dt$
- $B_{XY} = \int_0^1 S_v S_w W(t) V(t) dt - \int_0^1 S_v V(t) dt \int_0^1 S_w W(t) dt$
- $B_{XX} = \int_0^1 S_w^2 W^2(t) dt - \left[\int_0^1 S_w W(t) dt \right]^2$
- $B_{YY} = \int_0^1 S_v^2 V^2(t) dt - \left[\int_0^1 S_v V(t) dt \right]^2$

με τις x_Y και x_X να ακολουθούν την κανονική κατανομή με μηδενικούς μέσους όρους και $W(t)$, $V(t)$ να είναι οι ανελίξεις Wiener που αφορούν τη X_t και τη Y_t αντίστοιχα (χρησιμοποιείται διαφορετικός συμβολισμός από ότι στην ενότητα 2.2 για λόγους διάκρισης αφού τώρα οι τυχαίοι περίπατοι περιέχουν και μέσο όρο), ενώ \Rightarrow σημαίνει σύγκλιση σε κατανομή. Υπολογίζοντας τις διακυμάνσεις των x_X , και x_Y και απλοποιώντας προκύπτει:

$$T^{1/2} \left(\hat{b} - \frac{g_Y}{g_X} \right) \Rightarrow N \left(0, \frac{6}{5} \left(\frac{g_Y}{g_X} \right)^2 \left(\left(\frac{S_v}{g_Y} \right)^2 + \left(\frac{S_w}{g_X} \right)^2 \right) \right).$$

Δηλαδή, σε αντίθεση με την περίπτωση της υποφαινομενικής παλινδρόμησης χωρίς μέσους όρους, ο εκτιμητής ελαχίστων τετραγώνων \hat{b} συγκλίνει στο λόγο του μέσου όρου της εξαρτημένης μεταβλητής g_Y προς το μέσο όρο της ανεξάρτητης μεταβλητής g_X , παρέχοντας συνεπή εκτίμηση του λόγου αυτού. Επίσης, όπως και στη

καθιερωμένη ασυμπτωτική θεωρία υπάρχει ασυμπτωτική κανονικότητα με τη συνηθισμένη ταχύτητα σύγκλισης $T^{-1/2}$.

2) Για την εκτίμηση του σταθερού όρου \hat{a} ισχύει:

$$T^{-1/2}\hat{a} \Rightarrow 4 \left(s_v \int_0^1 V(t) dt - \frac{g_Y}{g_X} s_w \int_0^1 W(t) dt \right) + 6 \left(\frac{g_Y}{g_X} \int_0^1 s_w t W(t) dt - \int_0^1 s_v t V(t) dt \right)$$

και τελικά:

$$T^{-1/2}\hat{a} \Rightarrow N \left(0, \frac{2}{15} \left(s_v^2 + s_w^2 \left(\frac{g_Y}{g_X} \right)^2 \right) \right).$$

Δηλαδή, ο εκτιμητής ελαχίστων τετραγώνων \hat{a} αποκλίνει με ρυθμό $T^{1/2}$ όπως και στην περίπτωση της υποφαινομενικής παλινδρόμησης χωρίς μέσους όρους.

3) Για τις στατιστικές (2.2) του συντελεστή της κλίσης και του σταθερού όρου ισχύουν αντίστοιχα:

$$t_b = O_p(T)$$

και

$$t_a = O_p(T^{1/2}).$$

Δηλαδή, η στατιστική t_b αυξάνει γραμμικά, ενώ η στατιστική t_a αποκλίνει με την ίδια ταχύτητα $T^{1/2}$, όπως και στην περίπτωση της υποφαινομενικής παλινδρόμησης χωρίς μέσους όρους.

4) Για το συντελεστή προσδιορισμού ισχύει:

$$R^2 \Rightarrow 1 - T^{-1} (B_{YY} + O_p(1))$$

και τελικά:

$$1 - R^2 = O_p(T^{-1})$$

Κατά συνέπεια, ο συντελεστής προσδιορισμού συγκλίνει με ταχύτητα T στη μονάδα.

5) Για τη στατιστική των Durbin και Watson προκύπτει ότι:

$$DW = O_p(T^{-1}).$$

Δηλαδή, η στατιστική των Durbin και Watson συγκλίνει στο μηδέν με την ίδια ταχύτητα όπως και στη περίπτωση των απλών τυχαίων περιπάτων.

Οι προηγούμενες αποδείξεις εξηγούν σε μεγάλο βαθμό τα αποτελέσματα των προσομοιώσεων. Η σημαντική διαφορά που υπάρχει σε σχέση με την υποφαινομενική παλινδρόμηση, όπου οι τυχαίοι περίπατοι δεν περιλαμβάνουν μέσους όρους, είναι ότι η ύπαρξη μη μηδενικών μέσων όρων συνεπάγεται τη σύγκλιση του εκτιμητή ελαχίστων τετραγώνων του συντελεστή της κλίσης στο λόγο των δύο μέσων όρων. Ο συντελεστής προσδιορισμού δεν συγκλίνει σε μη εκφυλισμένη τυχαία μεταβλητή αλλά στη μονάδα, ενώ η στατιστική t_b αυξάνει γραμμικά. Οι δύο προαναφερθείσες ιδιότητες προτείνονται από τον Entorf (1997) να χρησιμοποιούνται σαν διαγνωστικό κριτήριο για την διάκριση των υποφαινομενικών παλινδρομήσεων με μέσο όρο από αυτές χωρίς μέσο όρο.

2.4 Υποφαινομενικές Παλινδρομήσεις με Στάσιμες Αυτοσυσχετιζόμενες διαδικασίες

Οι Granger, Hyung και Jeon (2001) επέκτειναν τη προσομοίωση των Granger και Newbold (1974) χρησιμοποιώντας στάσιμες διαδικασίες. Το συμπέρασμα που κατέληξαν είναι ότι το φαινόμενο της υποφαινομενικής παλινδρόμησης εμφανίζεται και στην περίπτωση όπου οι εμπλεκόμενες στην παλινδρόμηση μεταβλητές είναι θετικά στάσιμες αυτοπαλίνδρομες διαδικασίες ή πολλών όρων διαδικασίες κινητού μέσου όρου.

Έστω το υπόδειγμα της απλής γραμμικής παλινδρόμησης (2.4) όπου όμως οι Y_t και X_t δεν είναι ανεξάρτητοι τυχαίοι περίπατοι αλλά οι ανεξάρτητες στάσιμες αυτοπαλίνδρομες διαδικασίες πρώτης τάξης:

$$Y_t = f_Y Y_{t-1} + e_{Yt} \text{ για } |f_Y| < 1$$

και:

$$X_t = f_X X_{t-1} + e_{Xt} \text{ για } |f_X| < 1 \quad (2.9)$$

με τα e_{Yt} και e_{Xt} να είναι iid διαδικασίες με μηδενικούς μέσους όρους.

Οι Granger, Hyung και Jeon (2001) κατέγραψαν το ποσοστό απόρριψης της μηδενικής υπόθεσης που προκύπτει από τη χρήση της στατιστικής (2.2) για τον έλεγχο της στατιστικής σημαντικότητας του συντελεστή της κλίσης του υποδείματος (2.4) όταν οι εμπλεκόμενες στη παλινδρόμηση χρονοσειρές είναι οι AR(1) διαδικασίες (2.9) και η εκτίμηση γίνεται με τη μέθοδο των ελαχίστων τετραγώνων. Τα μεγέθη των παραγόμενων χρονοσειρών ήταν από 100

έως 10000 παρατηρήσεις με τις τιμές για τα e_{y_t} και e_{x_t} να προέρχονται από τυποποιημένη κανονική κατανομή. Επιπλέον, χρησιμοποίησαν και μια ασυμπτωτική τιμή που αποδεικνύεται στην εργασία τους. Στους Πίνακες 2-7 και 2-8 συνοψίζονται τα αποτελέσματα των προσομοιώσεων όταν $f_Y = f_X = f$ και $f_Y \neq f_X$ αντίστοιχα.

Πίνακας 2-7
Υποφαινομενικές παλινδρομήσεις μεταξύ 2 ανεξαρτήτων AR(1) χρονοσειρών
($f_Y = f_X = f$), ποσοστό $|t| > 1.96$

Μέθοδος	T	$f = 0$	$f = 0.25$	$f = 0.5$	$f = 0.75$	$f = 0.9$	$f = 1$
OLS	100	5.3	7.6	13.3	29.1	51.5	77.0
	500	5.8	7.5	16.3	31.5	51.6	90.0
	2000	5.8	7.1	13.5	29.4	52.5	94.5
	10000	4.3	6.6	12.2	30.6	52.3	97.6
	∞	5.0	7.0	13.0	30.0	53.0	100.0
BART	100	7.6	7.7	9.9	16.5	30.6	62.0
	500	6.4	6.8	9.0	14.1	23.9	79.6
	2000	6.0	5.9	6.1	10.3	16.3	86.4
	10000	4.6	5.2	5.5	7.7	12.8	92.5
	∞	5.0	5.0	5.0	5.0	5.0	100.0

Πηγή: Granger, Hyung και Jeon (2001)

Το συμπέρασμα που προκύπτει είναι ότι το φαινόμενο της υποφαινομενικής παλινδρόμησης παρουσιάζεται σε αξιοσημείωτο ποσοστό όταν το f παίρνει τιμές μικρότερες της μονάδας. Το ποσοστό αυτό είναι χαμηλότερο για μικρές τιμές του f , αυξάνει σημαντικά καθώς το f προσεγγίζει τη μονάδα και δεν εξαρτάται από το μέγεθος του δείγματος. Τα ποσοστά των υποφαινομενικών παλινδρομήσεων για τα ζευγάρια ($f_X = m$ και $f_Y = n$) και ($f_X = n$ και $f_Y = m$) με $m \neq n$ είναι τα ίδια. Υπάρχει δηλαδή συμμετρικότητα.

Οι προσομοιώσεις επαναλήφθηκαν με τιμές για τα e_{y_t} και e_{x_t} από κατανομές Cauchy, Laplace, ομοιόμορφη, εκθετική και Student. Τα συμπεράσματα που κατέληξαν ήταν παρόμοια με εξαίρεση τη κατανομή Cauchy, όπου για $T = 100$ το πρόβλημα μειώνονταν με την αύξηση του δείγματος.

Πίνακας 2-8
Υποφαινομενικές Παλινδρομήσεις μεταξύ δύο ανεξαρτήτων AR(1) χρονοσειρών
($f_Y \neq f_X$), ποσοστό $|t| > 1.96$

$f_X = 0$	T	$f_Y = 0$	$f_Y = 0.25$	$f_Y = 0.5$	$f_Y = 0.75$	$f_Y = 0.9$	$f_Y = 1$
OLS	100	5.3	5.1	5.8	6.0	4.4	5.3
	500	5.8	5.8	5.8	6.8	6.6	5.3
	2000	5.8	5.6	5.7	6.2	6.6	4.4
	∞	5.0	5.0	5.0	5.0	5.0	5.0
BART	100	7.6	7.3	7.2	7.5	6.8	6.5
	500	6.4	6.8	7.4	7.1	6.1	5.5
	2000	6.0	5.6	5.7	6.4	6.3	4.1
$f_Y = 0$	T	$f_X = 0$	$f_X = 0.25$	$f_X = 0.5$	$f_X = 0.75$	$f_X = 0.9$	$f_X = 1$
OLS	100	5.3	5.6	5.1	4.6	5.5	5.4
	500	5.8	6.1	7.1	4.6	5.0	5.2
	2000	5.8	5.6	5.6	5.7	6.0	4.6
	∞	5.0	5.0	5.0	5.0	5.0	5.0
BART	100	7.6	7.3	7.6	7.4	9.1	9.4
	500	6.4	6.6	7.1	6.1	5.8	5.7
	2000	6.0	6.0	5.5	5.6	6.0	5.0
$f_X = 0.5$	T	$f_Y = 0$	$f_Y = 0.25$	$f_Y = 0.5$	$f_Y = 0.75$	$f_Y = 0.9$	$f_Y = 1$
OLS	100	5.1	8.8	13.3	19.6	21.9	24.8
	500	7.1	9.9	16.3	21.2	23.8	26.2
	2000	5.6	8.4	13.5	20.3	22.5	22.8
	∞	5.0	8.0	13.0	19.0	23.0	26.0
BART	100	7.6	8.7	9.9	12.8	13.9	14.7
	500	7.1	8.1	9.0	10.0	9.8	9.3
	2000	5.5	5.8	6.1	7.3	7.8	6.3
$f_Y = 0.5$	T	$f_X = 0$	$f_X = 0.25$	$f_X = 0.5$	$f_X = 0.75$	$f_X = 0.9$	$f_X = 1$
OLS	100	5.8	9.0	13.3	18.1	21.8	26.4
	500	5.8	10.2	16.3	19.8	22.1	26.4
	2000	5.7	10.2	13.5	20.0	23.3	24.3
	∞	5.0	8.0	13.0	19.0	23.0	26.0
BART	100	7.2	7.7	9.9	12.2	13.4	16.0
	500	7.4	8.2	9.0	9.3	9.7	10.2
	2000	5.7	6.3	6.1	8.3	7.9	6.9

Πηγή: Granger, Hyung και Jeon (2001)

Έστω τώρα ότι οι Y_t και X_t παράγονται από τις ανεξάρτητες μη αντιστρέψιμες διαδικασίες κινητών μέσων k τάξης:

$$Y_t = \sum_{j=0}^k e_{Y_{t-j}} \text{ και } X_t = \sum_{j=0}^k e_{X_{t-j}} \text{ για } k = 1, 2, 5, 10, 20, 50$$

όπου e_{Y_t} και e_{X_t} είναι iid διαδικασίες με μηδενικούς μέσους όρους. Από το Πίνακα 2-9 που περιέχει τα αποτελέσματα της προσομοίωσης συμπεραίνεται ότι υποφαινομενικές παλινδρομήσεις εμφανίζονται αρκετά συχνά. Το ποσοστό τους αυξάνεται καθώς το k μεγαλώνει και δεν εξαρτάται από το μέγεθος του δείγματος.

Πίνακας 2-9
Υποφαινομενικές παλινδρομήσεις μεταξύ 2 ανεξαρτήτων MA χρονοσειρών,
ποσοστό $|t| > 1.96$

Μέθοδος	T	w.m.	MA(1)	MA(2)	MA(5)	MA(10)	MA(20)	MA(50)
OLS	100	6.4	9.3	17.1	30.6	45.3	60.2	76.0
	500	5.8	12.8	20.7	35.4	48.0	59.4	71.9
	2000	6.3	12.1	20.2	33.8	46.1	59.4	75.7
	10000	4.4	9.9	17.4	33.5	46.3	59.4	74.1
BART	100	7.1	8.1	10.6	13.8	22.5	34.4	58.2
	500	6.1	8.2	10.6	11.7	11.5	12.1	19.7
	2000	6.3	8.3	9.0	10.1	11.4	10.3	14.3
	10000	4.6	6.8	8.2	9.2	9.9	10.1	10.4

Πηγή: Granger, Hyung και Jeon (2001)

Το γεγονός ότι οι εμπλεκόμενες στην παλινδρόμηση χρονοσειρές είναι αυτοσυσχετιζόμενες διαδικασίες επηρεάζει τα σφάλματα τα οποία δεν μπορούν να είναι λευκός θόρυβος. Αυτό συνεπάγεται την απώλεια της συνέπειας του εκτιμητή του τυπικού σφάλματος του \hat{b} , $s(\hat{b})$ και τη μη σύγκλιση της στατιστικής (2.2) σε τυπική κανονική κατανομή με την αύξηση του μεγέθους του δείγματος. Για το λόγο αυτό, οι Granger, Hyung και Jeon (2001) προτείνουν τη διόρθωση της διακύμανσης του εκτιμητή \hat{b} με τη χρήση της:

$$S_{\hat{b}}^2 = M^{-1}VM^{-1} \quad (2.10)$$

όπου:

$$M \equiv E \left[(X_t - \bar{X})^2 \right]$$

και

$$V \equiv \text{Var} \left[T^{-1/2} \sum_{t=1}^T (X_t - \bar{X}) u_t \right].$$

Η συνεπής εκτιμήτρια της (2.10) είναι η:

$$\hat{S}_b^2 = \hat{M}^{-1} \hat{V} \hat{M}^{-1}$$

με:

$$\hat{V} = \frac{1}{T} \sum_{t=1}^T (X_t - \bar{X})^2 \hat{u}_t^2 + \frac{2}{T} \sum_{s=1}^l w(s, l) \times \sum_{t=s+1}^T (X_t - \bar{X}) \hat{u}_t \hat{u}_{t-s} (X_{t-s} - \bar{X}) \xrightarrow{P} V$$

και

$$\hat{M} = T^{-1} \sum_{t=1}^T (X_t - \bar{X})^2 \xrightarrow{P} M \text{ καθώς } T \rightarrow \infty$$

ενώ $w(s, l)$ είναι η βέλτιστη σταθμισμένη συνάρτηση:

$$t_b = \frac{\hat{b}}{\hat{S}_b} \xrightarrow{d} N(0, 1)$$

που αντιστοιχεί στην επιλογή του φασματικού παραθύρου Bartlett:

$$w(s, l) = 1 - \frac{s}{l+1}$$

το οποίο εγγυάται ότι η \hat{S}_b^2 δεν παίρνει αρνητικές τιμές.

Οι προσομοιώσεις επαναλήφθηκαν χρησιμοποιώντας το συνεπή εκτιμητή (2.10). Γενικά, διαπιστώνεται ότι το πρόβλημα βελτιώνεται σημαντικά εκτός για τιμές των f πολύ κοντά στη μονάδα. Επίσης, όταν μία από τις δύο χρονοσειρές ή και οι δύο είναι λευκός θόρυβος τα αποτελέσματα χειροτερεύουν. Κατά συνέπεια, ο συνεπής εκτιμητής (2.10) πρέπει να χρησιμοποιείται αποκλειστικά όταν και οι δύο χρονοσειρές είναι αυτοσυσχετιζόμενες διαδικασίες.

2.5 Ανακεφαλαίωση

Η εμφάνιση του φαινόμενου της υποφαινομενικής παλινδρόμησης είναι ο κανόνας παρά η εξαίρεση όταν διερευνώνται σχέσεις μεταξύ χρονοσειρών που είναι ολοκληρωμένες διαδικασίες. Όμως, παρόλο που αρχικά πιστεύονταν ότι οφείλεται αποκλειστικά στην έλλειψη στασιμότητας των εμπλεκόμενων στην παλινδρόμηση χρονοσειρών, εντούτοις συναντάται σε αξιοσημείωτο ποσοστό και με στάσιμες αυτοσυσχετιζόμενες διαδικασίες, ιδιαίτερα όταν αυτές πλησιάζουν στα όρια της μη στασιμότητας. Αυτό συμβαίνει γιατί οι ιδιότητες αυτοσυσχέτισης των χρονοσειρών επηρεάζουν τα σφάλματα τα οποία παύουν να είναι λευκός θόρυβος. Δηλαδή, μπορεί να θεωρηθεί ότι το πρόβλημα της υποφαινομενικής παλινδρόμησης ουσιαστικά είναι πρόβλημα αυτοσυσχέτισης των σφαλμάτων. Άμεση συνέπεια είναι ότι οι συμβατικοί έλεγχοι που εφαρμόζονται στη μεθοδολογία της παλινδρόμησης για τον έλεγχο της ύπαρξης σχέσεως μεταξύ των μεταβλητών δεν λειτουργούν σωστά και οδηγούν σε εσφαλμένα συμπεράσματα.

Ο ερευνητής θα πρέπει να είναι ιδιαίτερα προσεχτικός όταν εξετάζει υποδείγματα παλινδρόμησης που περιέχουν χρονοσειρές γιατί μπορεί να οδηγηθεί σε υποφαινομενικές σχέσεις. Κρίνεται αναγκαία η μελέτη των ιδιοτήτων αυτοσυσχέτισης κάθε χρονοσειράς και η εξέταση της στασιμότητάς της, προτού αυτή περιληφθεί σε ένα υπόδειγμα παλινδρόμησης. Ακόμα καλύτερα θα ήταν η ταυτοποίηση κάθε χρονοσειράς σύμφωνα με τη μεθοδολογία των Box και Jenkins. Με αυτόν τον τρόπο, ο ερευνητής θα είναι σε θέση να δίνει τη σωστή βαρύτητα στα αποτελέσματα της παλινδρόμησης και να απορρίπτει μεταβλητές που δεν έχουν ουσιαστική συνεισφορά στην ερμηνεία του υποδείγματος. Στην περίπτωση των ολοκληρωμένων διαδικασιών, η εκτίμηση του υποδείγματος σε πρώτες διαφορές ίσως συμβάλλει στην επίλυση του προβλήματος.

Equation Chapter 3 Section 3

ΚΕΦΑΛΑΙΟ 3

ΑΛΛΕΣ ΜΟΡΦΕΣ ΥΠΟΦΑΙΝΟΜΕΝΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

3.1 Εισαγωγή

Το πρόβλημα της υποφαινομενικής παλινδρόμησης που αναπτύχθηκε στο προηγούμενο κεφάλαιο αποτελεί μια ιδιόζουσα περίπτωση που μπορεί να εμφανιστεί κατά την εκτίμηση ενός γραμμικού υποδείγματος με χρονοσειρές. Το φαινόμενο εμφανίζεται όταν δύο οι περισσότερες χρονοσειρές βρίσκονται να συσχετίζονται γραμμικά μεταξύ τους, σύμφωνα με τη στατιστική συμπερασματολογία της μεθόδου των ελαχίστων τετραγώνων, ενώ στην πραγματικότητα δεν υφίσταται γραμμική τους σχέση. Κύριο χαρακτηριστικό της εμφάνισης του προβλήματος είναι ο υψηλός βαθμός αυτοσυσχέτισης των τιμών του τυχαίου σφάλματος που προκύπτει από το εκτιμηθέν υπόδειγμα. Κάθε προσπάθεια διόρθωσης του προβλήματος με τις υπάρχουσες μεθόδους αποβαίνει άκαρπη, γεγονός που φανερώνει ότι το πρόβλημα της αυτοσυσχέτισης στα σφάλματα δεν είναι πραγματικό αλλά εικονικό.

Ειδικότερα, στο προηγούμενο κεφάλαιο αναλύθηκε η συμπεριφορά του φαινομένου αυτού για ορισμένες κατηγορίες χρονοσειρών στάσιμων και μη στάσιμων. Σε όλες αυτές τις περιπτώσεις το κοινό χαρακτηριστικό των χρονοσειρών είναι ότι ήταν αυτοσυσχετιζόμενες διαδικασίες, δηλαδή αυτοπαλίνδρομες διαδικασίες πρώτης τάξης, πολλών όρων διαδικασίες κινητών μέσων όρων και τυχαίοι περίπατοι με η χωρίς μέσο όρο. Το γεγονός αυτό επηρεάζει τα σφάλματα τα οποία δεν μπορούν να είναι λευκός θόρυβος. Όμως, αρκετές φορές το φαινόμενο εμφανίζεται και με χρονοσειρές που δεν αποτελούν αυτοσυσχετιζόμενες διαδικασίες καθώς και με άλλες μορφές.

Στο κεφάλαιο αυτό θα εξεταστούν περιπτώσεις υποφαινομενικής παλινδρόμησης με χρονοσειρές που δεν είναι αυτοσυσχετιζόμενες διαδικασίες, δηλαδή με χρονοσειρές που είναι στάσιμες σε τάση διαδικασίες. Επιπλέον, θα δειχθεί πως μπορεί να προκύψει το φαινόμενο όταν γίνει λανθασμένη ταυτοποίηση της κατηγορίας στην οποία ανήκουν οι χρονοσειρές και

όταν περιληφθούν στο υπόδειγμα άσχετες χρονοσειρές που όμως έχουν κοινή σχέση με κάποια τρίτη χρονοσειρά. Επίσης, θα εξεταστούν πώς επιδρά ο μη γραμμικός μετασχηματισμός των εμπλεκόμενων στην παλινδρόμηση χρονοσειρών στο πρόβλημα και η πιθανότητα εμφάνισης του φαινομένου σε μη γραμμικά υποδείγματα.

3.2 Υποφαινομενικές παλινδρομήσεις λόγω λανθασμένης ταυτοποίησης των χρονοσειρών

Οι τυχαίοι περίπατοι που εξετάστηκαν στο προηγούμενο κεφάλαιο ανήκουν στην κατηγορία των στάσιμων σε διαφορές διαδικασιών (*DSP*). Όμως, πολλές φορές συναντιούνται χρονοσειρές που περιέχουν τάση, ανήκουν δηλαδή στην κατηγορία των στάσιμων σε τάση διαδικασιών (*TSP*). Οι χρονοσειρές αυτές ενώ είναι στάσιμες ως προς το χρόνο δεν είναι και ως προς το μέσο όρο για αυτό προτού χρησιμοποιηθούν για ανάλυση πρέπει να ληφθεί μέριμνα για την απαλοιφή της τάσης. Εάν υποθεθεί ότι η τάση οφείλεται σε γραμμική εξάρτηση από το χρόνο τα σφάλματα από την εκτίμηση ενός υποδείγματος παλινδρόμησης όπου έχει ως εξαρτημένη μεταβλητή την χρονοσειρά και ως ανεξάρτητη το χρόνο αποτελούν στάσιμη διαδικασία, απαλλαγμένη από την τάση, η οποία μπορεί να χρησιμοποιηθεί για περαιτέρω ανάλυση. Εναλλακτικά, όταν εκτιμώνται υποδείγματα παλινδρόμησης μεταξύ χρονοσειρών που περιέχουν τάση, η προσθήκη του χρόνου ως επιπλέον ανεξάρτητης μεταβλητής οδηγεί στην απαλοιφή της τάσης.

Μια μορφή υποφαινομενικής παλινδρόμησης, διαφορετική από αυτές που έχουν εξεταστεί μέχρι στιγμής, οφείλεται στη λανθασμένη ταυτοποίηση της κατηγορίας στην οποία ανήκει η χρονοσειρά. Αυτό σημαίνει ότι ενώ στην πραγματικότητα μια χρονοσειρά είναι στάσιμη σε διαφορές κάποιος την αντιλαμβάνεται εσφαλμένα σαν στάσιμη σε τάση και τη χειρίζεται ανάλογα. Στην περίπτωση αυτή το φαινόμενο της υποφαινομενικής παλινδρόμησης εμφανίζεται και οι στατιστικοί έλεγχοι συνηγορούν υπέρ της στατιστικής σημαντικότητας του συντελεστή της τάσης στα υποδείγματα που εμπλέκεται αυτή η χρονοσειρά.

Η πρώτη περίπτωση που θα εξεταστεί αφορά την εσφαλμένη εκτίμηση του υποδείγματος παλινδρόμησης με τη χρονοσειρά ως εξαρτημένη μεταβλητή και το χρόνο ως ανεξάρτητη. Έστω ότι ο μηχανισμός παραγωγής της στάσιμης σε διαφορές διαδικασίας Y_t είναι ο τυχαίος περίπατος:

$$Y_t = Y_{t-1} + u_t \quad \text{όπου } u_t \sim iid N(0, \sigma_u^2) \quad (3.1)$$

αλλά εσφαλμένα θεωρείται ότι ανήκει στην κατηγορία των στάσιμων σε τάση διαδικασιών. Για τον λόγο αυτό εκτιμάται το υπόδειγμα:

$$Y_t = a + bt + e_t \quad (3.2)$$

με στόχο την απαλοιφή της τάσης από τη χρονοσειρά ή εφόσον διορθωθεί το πρόβλημα της αυτοσυσχέτισης στα σφάλματα τον έλεγχο της στατιστικής σημαντικότητας του συντελεστή b που δηλώνει την παρουσία της τάσης.

Οι Nelson και Kang (1984) εξέτασαν την προηγούμενη περίπτωση διερευνώντας με τη βοήθεια προσομοίωσης τις τιμές των βασικών δεικτών της παλινδρόμησης όταν ο μηχανισμός παραγωγής μιας χρονοσειράς είναι ο τυχαίος περίπατος (3.1) και εκτιμηθεί το υπόδειγμα (3.2) με τη μέθοδο των ελαχίστων τετραγώνων για μεγέθη χρονοσειρών 20 και 100 παρατηρήσεων. Τα συμπεράσματα στα οποία κατέληξαν είναι ότι όταν το μέγεθος του δείγματος είναι 100 παρατηρήσεις ο συντελεστής της τάσης b είναι υποφαινομενικά σημαντικός στο 87% των περιπτώσεων, ενώ ο σταθερός όρος a στο 80% των περιπτώσεων για επίπεδο σημαντικότητας 5%. Η μέση τιμή του συντελεστή προσδιορισμού είναι 0.443, ενώ η μέση τιμή της στατιστικής των Durbin και Watson 0.198. Τα αποτελέσματα αυτά είναι παρόμοια με τα συμπεράσματα των προσομοιώσεων της Ενότητας 2.3 για τις παλινδρομήσεις μεταξύ δύο τυχαίων περιπάτων όταν ο ένας έχει μέσο όρο (*drift*) διαφορετικό από το μηδέν. Για χρονοσειρές μεγέθους 20 παρατηρήσεων ο συντελεστής προσδιορισμού παραμένει σταθερός κοντά στο 0.44 αλλά ως συνέπεια του μικρότερου δείγματος μειώνονται τα ποσοστά που οι συντελεστές a και b είναι στατιστικά σημαντικοί σε 54% και 73% αντίστοιχα. Όταν οι τυχαίοι περίπατοι περιέχουν και μέσο όρο (*drift*) ο συντελεστής προσδιορισμού είναι μεγαλύτερος και προσεγγίζει τη μονάδα πολύ γρήγορα καθώς αυξάνεται το μέγεθος του δείγματος.

Όμως, το σημαντικότερο συμπέρασμα από την προσομοίωση των Nelson και Kang (1984) είναι η συμπεριφορά των δειγματικών αυτοσυσχετίσεων των σφαλμάτων από την εκτίμηση του υποδείγματος (3.2). Η συνάρτηση αυτοσυσχέτισης των σφαλμάτων παίρνει τιμή για δεδομένη χρονική υστέρηση που εξαρτάται από το συγκεκριμένο μέγεθος του δείγματος και έχει σχήμα που μοιάζει με φθίνων ημιτονοειδές κύμα που ταλαντεύεται με περίοδο $\frac{2}{3}T$, κάτι που συνεπάγεται ψευδοπεριοδική συμπεριφορά για την απαλλαγμένη από την τάση

χρονοσειρά παρόλο που δεν υπήρχε τέτοια ένδειξη στην αρχική χρονοσειρά. Για μικρό αριθμό χρονικών υστερήσεων σε σχέση με το μέγεθος του δείγματος η συνάρτηση φθίνει εκθετικά. Οι μέσες τιμές των τριών πρώτων δειγματικών αυτοσυσχετίσεων των σφαλμάτων που προέκυψαν από την προσομοίωση για μέγεθος χρονοσειρών 100 παρατηρήσεων ήταν $r_1 = 0.88$, $r_2 = 0.77$ και $r_3 = 0.68$ ενώ οι μέσες τιμές των μερικών αυτοσυσχετίσεων ήταν $\hat{f}_{22} = -0.04$ και $\hat{f}_{33} = -0.02$. Κατά συνέπεια, ένας ερευνητής κρίνοντας με βάση τις δειγματικές αυτοσυσχετίσεις και μερικές αυτοσυσχετίσεις της απαλλαγμένης από την τάση χρονοσειράς των σφαλμάτων μπορεί να θεωρήσει ότι αυτά αντιπροσωπεύονται από μια στάσιμη AR(1) διαδικασία παρόλο που αρχική χρονοσειρά δεν αποτελεί στάσιμη σε τάση διαδικασία. Για μέγεθος χρονοσειρών 20 παρατηρήσεων οι μέσες δειγματικές αυτοσυσχετίσεις των σφαλμάτων ήταν $r_1 = 0.5$, $r_2 = 0.17$ και $r_3 = -0.04$ και οι μέσες μερικές αυτοσυσχετίσεις $\hat{f}_{22} = -0.18$ και $\hat{f}_{33} = -0.13$. Δηλαδή, όταν η χρονοσειρά έχει μόνο 20 παρατηρήσεις ο ταλαντευόμενος χαρακτήρας της συνάρτησης αυτοσυσχέτισης γίνεται περισσότερο έντονος στις μικρές χρονικές υστερήσεις.

Εάν ο αντικειμενικός σκοπός είναι ο στατιστικός έλεγχος της παρουσίας της τάσης και εφόσον θεωρηθεί ότι τα σφάλματα είναι μια AR(1) διαδικασία το υπόδειγμα (3.2) μπορεί να μετασχηματιστεί και να γραφεί ως εξής:

$$(Y_t - r_1 Y_{t-1}) = a(1 - r_1) + b[t - r_1(t-1)] + (e_t - r_1 e_{t-1})$$

όπου r_1 είναι ο συντελεστής του αυτοπαλίνδρομου υποδείγματος των σφαλμάτων. Αυτό ουσιαστικά αντιστοιχεί με μια διαδικασία Cochrane και Orcutt δύο επαναλήψεων. Ο προηγούμενος μετασχηματισμός θα ήταν σωστός μόνο εάν η τιμή του r_1 θέτονταν ίση με τη μονάδα, που σημαίνει την εκτίμηση του υποδείγματος (3.2) σε πρώτες διαφορές καθώς η Y_t είναι ολοκληρωμένη διαδικασία. Όμως, όταν χρησιμοποιηθεί η τιμή r_1 από το αυτοπαλίνδρομο υπόδειγμα των σφαλμάτων το πρόβλημα της αυτοσυσχέτισης στα σφάλματα υποδιορθώνεται (*undercorrected*) και το φαινόμενο της υποφαινομενικής παλινδρόμησης παραμένει. Η προσομοίωση των Nelson και Kang (1984) για μέγεθος χρονοσειρών 100 παρατηρήσεων έδειξε πως όταν εφαρμόζεται η προηγούμενη διόρθωση οι συντελεστές a και b συνεχίζουν εσφαλμένα να παραμένουν στατιστικά σημαντικοί και σε πολύ υψηλά ποσοστά, 45% και 58% αντίστοιχα για επίπεδο σημαντικότητας 0.05, ενώ η μέση τιμή της στατιστικής των Durbin και Watson αυξάνεται σε 1.88 που είναι παραπλανητικό γιατί

υπονοεί ότι το πρόβλημα της αυτοσυσχέτισης στα σφάλματα έχει διορθωθεί. Οι Nelson και Kang (1984) αποδεικνύουν πως όταν εφαρμόζεται η μέθοδος των Cochrane και Orcutt σε υποδείγματα που περιέχουν τάση ή αυτοπαλίνδρομους όρους η σύγκλιση του r_1 επέρχεται από την πρώτη επανάληψη.

Έστω τώρα ότι διερευνάται η σχέση μεταξύ δύο στάσιμων σε διαφορές χρονοσειρών Y_t και X_t και εσφαλμένα θεωρούνται ότι ανήκουν στην κατηγορία των στάσιμων σε τάση διαδικασιών. Αυτό συνεπάγεται την εκτίμηση του υποδείματος πολλαπλής παλινδρόμησης:

$$Y_t = a + bt + gX_t + e_t \quad (3.3)$$

το οποίο λαμβάνει υπ' όψιν την παρουσία της τάσης στις χρονοσειρές. Το προηγούμενο υπόδειγμα για να εκτιμηθεί σωστά θα έπρεπε να μετασχηματιστεί και να εκτιμηθεί σε πρώτες διαφορές δηλαδή να γραφεί ως:

$$\Delta Y_t = b + g\Delta X_t + w_t \quad \text{όπου } w_t = e_t - e_{t-1}$$

γιατί διαφορετικά η εκτίμηση του συντελεστή g θα υπόκειται στο φαινόμενο της υποφαινομενικής παλινδρόμησης, όπως έχουν περιγράψει οι Granger και Newbold (1974). Η προσομοίωση των Nelson και Kang (1984) για μέγεθος χρονοσειρών 100 παρατηρήσεων έδειξε πως η εκτίμηση του υποδείματος (3.3) με τη μέθοδο των ελαχίστων τετραγώνων, χωρίς την χρήση πρώτων διαφορών, αποδίδει μια μέση τιμή για τον συντελεστή προσδιορισμού κοντά στο 0.501 που είναι ελαφρά μεγαλύτερη από την τιμή 0.443 του υποδείματος (3.2) κάτι που είναι λογικό επειδή το υπόδειγμα τώρα περιλαμβάνει δύο ανεξάρτητες μεταβλητές. Ο συντελεστής g είναι υποφαινομενικά στατιστικά σημαντικός σε ποσοστό 64% για επίπεδο σημαντικότητας 5%. Η υποφαινομενική σχέση με το χρόνο παραμένει αν και το πρόβλημα βελτιώνεται ελαφρά με την προσθήκη της ανεξάρτητης μεταβλητής X_t . Το ποσοστό των περιπτώσεων όπου ο συντελεστής b είναι στατιστικά σημαντικός μειώνεται σε 83% από 87% που ήταν όταν το υπόδειγμα περιελάμβανε μόνο το χρόνο. Η μέση τιμή της στατιστικής των Durbin και Watson είναι 0.260, ελαφρά αυξημένη σε σχέση με το υπόδειγμα (3.2). Οι μέσες δειγματικές αυτοσυσχετίσεις των σφαλμάτων είναι μικρότερες σε σχέση με το υπόδειγμα (3.2), $r_1 = 0.852$, $r_2 = 0.7210$ και $r_3 = 0.604$.

Κρίνοντας από τις τιμές των δειγματικών αυτοσυσχετίσεων των σφαλμάτων και της στατιστικής των Durbin και Watson μπορεί να θεωρηθεί ότι τα σφάλματα ακολουθούν μια στάσιμη AR(1) διαδικασία και το υπόδειγμα (3.3) να μετασχηματιστεί σε:

$$(Y_t - r_1 Y_{t-1}) = a(1 - r_1) + b[t - r_1(t-1)] + g(X_t - r_1 X_{t-1}) + (e_t - r_1 e_{t-1})$$

το οποίο θα ήταν σωστό μόνο εάν η τιμή του r_1 θέτονταν ίση με τη μονάδα. Από την προσομοίωση προέκυψε πως όταν εφαρμοστεί ο προηγούμενος μετασχηματισμός το πρόβλημα της αυτοσυσχέτισης στα σφάλματα ελαφρά μόνο βελτιώνεται και το φαινόμενο της υποφαινομενικής παλινδρόμησης συνεχίζει να εμφανίζεται σε αξιοσημείωτο ποσοστό, 60% και 11,3% για τους συντελεστές b και g αντίστοιχα, ενώ η υψηλή τιμή της στατιστικής των Durbin και Watson 1.820 και οι χαμηλές τιμές των τριών πρώτων δειγματικών αυτοσυσχετίσεων $r_1 = 0.079$, $r_2 = 0.064$ και $r_3 = 0.052$ παραπλανούν γιατί υποδεικνύουν πως τα σφάλματα είναι στάσιμα. Σε αντίθεση με το υπόδειγμα (3.2) η μέθοδος των Cochrane και Orcutt επηρεάζει τα αποτελέσματα αλλά μόνο για το συντελεστή g για τον οποίο το ποσοστό των υποφαινομενικών σχέσεων που προκύπτει με βάση τις τελικές εκτιμήσεις της μεθόδου βελτιώνεται σε 6.7%. Αντίθετα για το συντελεστή της τάσης b η βελτίωση δεν είναι σημαντική και το ποσοστό παραμένει υψηλό στο 51.5%.

Οι Durlauf και Phillips (1988) κάνοντας χρήση του συναρτησιακού κεντρικού οριακού θεωρήματος και της μεθοδολογίας που περιγράφεται στους Phillips (1986 και 1987) και η οποία έχει αναφερθεί συνοπτικά στην ενότητα 2.2, εξήγησαν τα αποτελέσματα των προηγούμενων προσομοιώσεων αναπτύσσοντας μια ασυμπτωτική θεωρία για την παλινδρόμηση όταν χρησιμοποιούνται τυχαίοι περίπατοι από τους οποίους εσφαλμένα απαλείφεται η τάση. Έστω το υπόδειγμα της απλής γραμμικής παλινδρόμησης (3.1) στο οποίο η Y_t υποτίθεται ότι είναι μια στάσιμη σε τάση διαδικασία γύρω από λευκό θόρυβο, ενώ στην πραγματικότητα είναι ο τυχαίος περίπατος (3.2).

Εάν υποτεθεί ότι η (3.1) εκτιμάται με τη μέθοδο των ελαχίστων τετραγώνων, τότε οι Durlauf και Phillips (1988) αποδεικνύουν ότι καθώς $T \uparrow \infty$ ισχύουν:

$$1) \quad T^{-1/2} \hat{a} \Rightarrow 4 \int_0^1 B(r) - 6 \int_0^1 rB(r) \equiv N(0, 2S^2/15)$$

και

$$T^{1/2} \hat{b} \Rightarrow 12 \left[\int_0^1 rB(r) - \frac{1}{2} \int_0^1 B(r) \right] \equiv N(0, 6S^2/5).$$

Δηλαδή, ο εκτιμητής της τάσης \hat{b} είναι συνεπής και συγκλίνει στο μηδέν ενώ ο εκτιμητής του σταθερού όρου \hat{a} δεν είναι συνεπής και αποκλίνει καθώς $T \uparrow \infty$. Υπενθυμίζεται ότι στην περίπτωση της υποφαινομενικής παλινδρόμησης με

ολοκληρωμένες διαδικασίες, ο συντελεστής της κλίσης έχει μη εκφυλισμένη οριακή κατανομή, ενώ η κατανομή του σταθερού όρου \hat{a} αποκλίνει καθώς $T \uparrow \infty$.

2) Η $F_{a=0}$ και η $F_{b=0}$ αποκλίνουν.

Δηλαδή, οι F κατανομές όταν το a είναι μηδέν και όταν το b είναι μηδέν αποκλίνουν. Για το συντελεστή a αυτό είναι λογικό αφού αυτός δεν έχει κάποια ασυμπτωτική κατανομή. Για το συντελεστή b διαπιστώνεται ότι παρόλο που αυτός συγκλίνει στο μηδέν η F στατιστική για τον έλεγχο της στατιστικής του σημαντικότητας αποκλίνει στο άπειρο. Παρόμοια συμπεριφορά για το συντελεστή b υπάρχει στην περίπτωση της υποφαινομενικής παλινδρόμησης με ολοκληρωμένες διαδικασίες, όπως έχειδειχθεί στην ενότητα 2.2.

3) $DW \xrightarrow{p} 0$.

Δηλαδή, η στατιστική των Durbin και Watson συγκλίνει κατά πιθανότητα στο μηδέν. Επομένως, μια πολύ χαμηλή τιμή για τη στατιστική των Durbin και Watson μπορεί να αποβεί χρήσιμη για την αναγνώριση του σφάλματος ταυτοποίησης του υποδείγματος (3.1) παρόλο που δεν υποδεικνύει απαραίτητα ότι η Y_t είναι ολοκληρωμένη διαδικασία.

4) Ο συντελεστής προσδιορισμού συγκλίνει ασθενώς σε μη εκφυλισμένη τυχαία μεταβλητή καθώς $T \uparrow \infty$.

Στην περίπτωση που η Y_t είναι τυχαίος περίπατος με μέσο όρο τα παραπάνω αποτελέσματα δεν αλλάζουν. Αυτό οφείλεται στο ότι μια ολοκληρωμένη διαδικασία με μέσο όρο μπορεί να γραφεί ως μια διαδικασία που περιέχει τάση. Δηλαδή, όπως έχειδειχθεί στην ενότητα 1.4:

$$Y_t = m + Y_{t-1} + u_t = mt + Y_0 + \sum_{i=1}^t u_i = mt + Y_t$$

όπου η $Y_t = Y_0 + \sum_{i=1}^t u_i$ είναι ένας τυχαίος περίπατος χωρίς τάση. Επομένως, η εκτίμηση του υποδείγματος (3.1) όταν η Y_t είναι τυχαίος περίπατος με μέσο όρο διαφορετικό από το μηδέν θα αποφέρει παρόμοια αποτελέσματα εκτός από τον όρο mt ο οποίος θα εκτιμηθεί από το συντελεστή της τάσης.

Έστω το υπόδειγμα της πολλαπλής γραμμικής παλινδρόμησης (3.3) στο οποίο οι Y_t και X_t θεωρούνται εσφαλμένα ότι είναι στάσιμες σε τάση διαδικασίες, ενώ στην

πραγματικότητα είναι ανεξάρτητοι τυχαίοι περίπατοι. Εάν γίνει η εκτίμηση του υποδείγματος με τη μέθοδο των ελαχίστων τετραγώνων σκοπεύοντας στην απαλοιφή της τάσης, τότε καθώς $T \uparrow \infty$ οι Durlauf και Phillips (1988) αποδεικνύουν ότι ισχύουν:

1) Ο εκτιμητής του σταθερού όρου \hat{a} αποκλίνει.

2) $\hat{b} \xrightarrow{p} 0$.

Δηλαδή, ο εκτιμητής της τάσης \hat{b} είναι συνεπής και συγκλίνει κατά πιθανότητα στο μηδέν.

3) Ο εκτιμητής της μεταβλητής X_t , \hat{g} συγκλίνει ασθενώς σε μη εκφυλισμένη τυχαία μεταβλητή.

4) Η $F_{g=0}$ αποκλίνει.

Δηλαδή ο έλεγχος με την F κατανομή αποκλίνει όπως στην περίπτωση της υποφαινομενικής παλινδρόμησης με ολοκληρωμένες διαδικασίες. Η κατάλληλα κανονικοποιημένη ασυμπτωτική κατανομή του ελέγχου επηρεάζεται από την παρουσία της τάσης αλλά το τελικό αποτέλεσμα παραμένει αμετάβλητο. Αυτό πραγματικά έχει σημασία είναι η έλλειψη στασιμότητας των εμπλεκόμενων στην παλινδρόμηση χρονοσειρών και όχι η απαλοιφή της τάσης.

5) $DW \xrightarrow{p} 0$.

Δηλαδή, η στατιστική των Durbin και Watson συγκλίνει κατά πιθανότητα στο μηδέν αποτελώντας και πάλι χρήσιμη ένδειξη για την εμφάνιση του προβλήματος της υποφαινομενικής παλινδρόμησης.

Από τα προηγούμενα ασυμπτωτικά αποτελέσματα, για τα οποία οι ακριβείς μορφές βρίσκονται στους Durlauf και Phillips (1988), συμπεραίνεται πως το πρόβλημα της υποφαινομενικής παλινδρόμησης παραμένει όταν οι χρησιμοποιούμενες στο υπόδειγμα χρονοσειρές ταυτοποιηθούν λανθασμένα και θεωρηθούν ότι είναι στάσιμες σε τάση διαδικασίες, ενώ στην πραγματικότητα είναι ολοκληρωμένες διαδικασίες. Αυτό συμβαίνει τόσο όταν απαλείφεται η τάση από έναν μεμονωμένο τυχαίο περίπατο όσο και από ένα υπόδειγμα παλινδρόμησης μεταξύ δύο τυχαίων περιπάτων. Όμως, στη δεύτερη περίπτωση το φαινόμενο της υποφαινομενικής παλινδρόμησης θα εμφανιστεί ανεξάρτητα από την εμπλοκή της τάσης ως ανεξάρτητης μεταβλητής. Δηλαδή, ο κρίσιμος παράγοντας είναι η έλλειψη στασιμότητας των χρονοσειρών και όχι η απαλοιφή της τάσης καθαυτή.

3.3 Υποφαινομενικές παλινδρομήσεις με στάσιμες σε τάση διαδικασίες

Οι Kim, Lee και Newbold (2004) εξέτασαν την περίπτωση της απλής γραμμικής παλινδρόμησης όπου οι εμπλεκόμενες στο υπόδειγμα χρονοσειρές δεν είναι τυχαίοι περίπατοι αλλά στάσιμες σε τάση διαδικασίες από τις οποίες δεν έχει απαλοιφεί η τάση. Το συμπέρασμα στο οποίο κατέληξαν είναι ότι το φαινόμενο της υποφαινομενικής παλινδρόμησης κάνει την εμφάνισή του και σε αυτή την περίπτωση.

Έστω το υπόδειγμα της απλής γραμμικής παλινδρόμησης:

$$Y_t = a + bX_t + e_t \quad (3.4)$$

όπου οι Y_t και X_t είναι οι στάσιμες σε τάση διαδικασίες:

$$\begin{aligned} Y_t &= m_Y + g_Y t + u_{Yt} & \text{για } u_{Yt} &= f_Y u_{Yt-1} + a_{Yt} & \text{με } |f_Y| < 1 \\ X_t &= m_X + g_X t + u_{Xt} & \text{για } u_{Xt} &= f_X u_{Xt-1} + a_{Xt} & \text{με } |f_X| < 1 \end{aligned} \quad (3.5)$$

ενώ υποτίθεται ότι το a_{Yt} είναι $iid(0, S_Y^2)$, το a_{Xt} είναι $iid(0, S_X^2)$ και είναι ανεξάρτητα μεταξύ τους. Εφόσον ο μηχανισμός παραγωγής των Y_t και X_t είναι η (3.5) και το υπόδειγμα (3.4) εκτιμάται με τη μέθοδο των ελαχίστων τετραγώνων οι Kim, Lee και Newbold (2004) έδειξαν ότι ισχύουν τα εξής:

- 1) Όταν και οι δύο χρονοσειρές έχουν τάση, δηλαδή ισχύουν $g_Y \neq 0$ και $g_X \neq 0$, τότε:

$$\hat{b} \xrightarrow{P} \frac{g_Y}{g_X}$$

και

$$T^{-3/2} t_{\hat{b}} \xrightarrow{P} \left[\frac{S_Y^2}{1-f_Y^2} + \left(\frac{g_Y}{g_X} \right)^2 \frac{S_X^2}{1-f_X^2} \right]^{1/2} \left(\frac{g_X^2}{12} \right)^{1/2} \frac{g_Y}{g_X}.$$

Δηλαδή, ο εκτιμητής \hat{b} του απλού γραμμικού που προκύπτει με τη μέθοδο των ελαχίστων τετραγώνων συγκλίνει κατά πιθανότητα στο λόγο του συντελεστή της τάσης της εξαρτημένης μεταβλητής, g_Y προς το συντελεστή της τάσης της ανεξάρτητης μεταβλητής, g_X και η στατιστική (2.2) που χρησιμοποιείται για τον έλεγχο της στατιστικής του σημαντικότητας αποκλίνει κάτι που δικαιολογεί και την παρουσία του φαινομένου της υποφαινομενικής παλινδρόμησης. Τα αποτελέσματα αυτά είναι παρόμοια με αυτά του Entorf (1997) όπου είχε αποδείξει πως όταν η

εξαρτημένη και η ανεξάρτητη μεταβλητή είναι τυχαίοι περίπατοι με μέσο όρο ο εκτιμητής \hat{b} του υποδείγματος συγκλίνει στο λόγο των δύο μέσων όρων. Η διαφορά που υπάρχει είναι ότι η στατιστική (2.2) του \hat{b} αποκλίνει με μεγαλύτερο ρυθμό $T^{3/2}$ από το ρυθμό $T^{1/2}$ που ισχύει για την περίπτωση που εξέτασε ο Entorf. Η αιτία για την εμφάνιση του προβλήματος της υποφαινομενικής παλινδρόμησης σε αυτή την περίπτωση είναι ότι οι δύο χρονοσειρές παρόλο που είναι ανεξάρτητες οδηγούνται μαζί εξαιτίας της παρουσίας της κοινής γραμμικής τάσης.

- 2) Όταν η ανεξάρτητη μεταβλητή X_t έχει τάση αλλά η εξαρτημένη Y_t δεν έχει, δηλαδή όταν $g_Y = 0$ και $g_X \neq 0$, για τον εκτιμητή ελαχίστων τετραγώνων \hat{b} ισχύει:

$$T^{3/2} \hat{b} \xrightarrow{d} N \left(0, \frac{12s_Y^2}{g_X^2 (1-f_Y)^2} \right)$$

και για την αντίστοιχη στατιστική συνάρτηση (3.2):

$$t_{\hat{b}} \xrightarrow{d} N \left(0, \frac{1-s_Y^2}{(1-f_Y)^2} \right).$$

Αντίθετα, όταν η ανεξάρτητη μεταβλητή X_t δεν έχει τάση αλλά η εξαρτημένη Y_t έχει, δηλαδή όταν $g_Y \neq 0$ και $g_X = 0$, ισχύουν:

$$T^{-1/2} \hat{b} \xrightarrow{d} N \left(0, \frac{g_Y^2 (1-f_X^2)^2}{12s_X^2 (1-f_X)^2} \right)$$

και

$$t_{\hat{b}} \xrightarrow{d} N \left(0, \frac{1-f_X^2}{(1-f_X)^2} \right).$$

Δηλαδή, ο εκτιμητής \hat{b} αποκλίνει αλλά η στατιστική του συνάρτηση (2.2), $t_{\hat{b}}$, συγκλίνει και πάλι σε κανονική κατανομή. Παρόλη τη διαφορετική συμπεριφορά των δύο εκτιμητών και στις δύο περιπτώσεις οι στατιστικές συναρτήσεις για τον έλεγχο της στατιστικής τους σημαντικότητας συγκλίνουν σε παραπλήσιες κανονικές κατανομές με μέσο όρο μηδέν και διακυμάνσεις $\frac{1-s_Y^2}{(1-f_Y)^2}$ και $\frac{1-f_X^2}{(1-f_X)^2}$ αντίστοιχα.

Επιπρόσθετα, από την ασυμπτωτική συμπεριφορά της στατιστικής $t_{\hat{b}}$ διαπιστώνεται ότι η διακύμανση είναι μονότονα αύξουσα συνάρτηση του συντελεστή του αυτοπαλίνδρομου όρου. Αυτό συνεπάγεται ότι καθώς ο συντελεστής f προσεγγίζει τη μονάδα ασυμπτωτικά το ποσοστό απόρριψης της μηδενικής υπόθεσης για τον έλεγχο της στατιστικής σημαντικότητας του συντελεστή \hat{b} θα αυξάνεται. Επομένως, για στάσιμες AR(1) διαδικασίες το φαινόμενο της υποφαινομενικής παλινδρόμησης μπορεί να εμφανιστεί και η συχνότητά του εξαρτάται από το πόσο κοντά στη μονάδα είναι η τιμή του συντελεστή του αυτοπαλίνδρομου όρου.

Πίνακας 3-1

Υποφαινομενικές Παλινδρομήσεις με στάσιμες σε τάση διαδικασίες, ποσοστό $|t| > 1.96$

T	g_Y	g_X	$f_Y = 0$ $f_X = 0$	$f_Y = 0.3$ $f_X = 0.3$	$f_Y = 0.9$ $f_X = 0.9$	$f_Y = 0$ $f_X = 0.9$	$f_Y = 0.9$ $f_X = 0$
50	0	0	5.2	7.3	47.8	5.9	5.3
	0	0.2	5.3	14.5	61.7	6.6	66.2
	0.2	0	6.5	15	61.4	65.2	4.8
	0.2	0.2	100	100	96.2	98.8	99.5
100	0	0	5.2	7.4	49.2	4.8	4.7
	0	0.2	5.9	15.9	67	4.8	68.2
	0.2	0	5.9	16	65.2	65.6	6.7
	0.2	0.2	100	100	100	100	100
10000	0	0	5.4	8.7	54.4	5.1	5.4
	0	0.2	5.3	14.9	68.3	5.3	68.3
	0.2	0	3.6	12.9	65.6	65.4	3.6
	0.2	0.2	100	100	100	100	100

Πηγή: Kim, Lee και Newbold (2004)

Οι Kim, Lee και Newbold (2004) ενίσχυσαν τα παραπάνω ασυμπτωτικά αποτελέσματα χρησιμοποιώντας προσομοίωση, για τρία μεγέθη χρονοσειρών 50, 100 και 10000 παρατηρήσεων και με τους διαταρακτικούς όρους a_{Yt} και a_{Xt} να προέρχονται από $N(0, 1)$. Ο Πίνακας 3-1 περιέχει τα ποσοστά που προέκυψαν από την προσομοίωση των Kim, Lee και

Newbold (2004) για τα οποία ο συντελεστής b του υποδείγματος (3.4) είναι υποφαινομενικά στατιστικά σημαντικός. Διαπιστώνεται πως όταν τα u_{y_t} και u_{x_t} είναι αυτοπαλίνδρομες διαδικασίες πρώτου βαθμού και ταυτόχρονα $g_y \neq 0$ και $g_x \neq 0$, δηλαδή και οι δύο χρονοσειρές έχουν τάση, το ποσοστό απόρριψης είναι 100% ανεξάρτητα από τις τιμές των άλλων παραμέτρων. Εξαιρέση υπάρχει μόνο για μέγεθος χρονοσειρών 50 παρατηρήσεων και για τα ζεύγη τιμών $f_y = 0.9$ & $f_x = 0.9$, $f_y = 0$ & $f_x = 0.9$ και $f_y = 0.9$ & $f_x = 0$ όπου τα ποσοστά απόρριψης είναι 96.2%, 98.8% και 99.5% αντίστοιχα, τα οποία όμως είναι πολύ κοντά στο 100%. Καθώς το μέγεθος του δείγματος αυξάνει ή αυξάνει το f_y ή το f_x , το μέγεθος της στατιστικής (2.2) για τον έλεγχο της σημαντικότητας του b αυξάνει ακόμα πιο πολύ.

3.4 Υποφαινομενικές παλινδρομήσεις λόγω κοινής σχέσης με τρίτη μεταβλητή

Η περίπτωση της υποφαινομενικής παλινδρόμησης που εξέτασαν οι Kim, Lee και Newbold (2004) μπορεί να επεκταθεί περαιτέρω θεωρώντας ότι το χαρακτηριστικό των χρονοσειρών Y_t και X_t δεν είναι η κοινή επίδραση της τάσης, δηλαδή δεν είναι στάσιμες σε τάση διαδικασίες, αλλά η κοινή τους σχέση με κάποια τρίτη χρονοσειρά. Αυτό συμβαίνει γιατί από τα πρώτα βήματα της στατιστικής επιστήμης οι ερευνητές πίστευαν πως η βασική αιτία της δημιουργίας υποφαινομενικών σχέσεων μεταξύ χρονοσειρών ήταν η κοινή τους εξάρτηση με κάποια τρίτη χρονοσειρά.

Έστω το υπόδειγμα της απλής γραμμικής παλινδρόμησης (3.4) όπου οι Y_t και X_t είναι ανεξάρτητες μεταξύ τους αλλά έχουν κοινή γραμμική σχέση με την Z_t , δηλαδή:

$$\begin{aligned} Y_t &= m_y + g_y Z_t + u_{y_t} \\ X_t &= m_x + g_x Z_t + u_{x_t} \end{aligned} \quad (3.6)$$

όπου η Z_t παράγεται σύμφωνα με το AR(1) υπόδειγμα:

$$Z_t = f_z Z_{t-1} + u_{z_t} \quad \text{me} \quad |f_z| \leq 1 \quad (3.7)$$

ενώ το u_{y_t} είναι $iid(0, S_y^2)$, το u_{x_t} είναι $iid(0, S_x^2)$, το u_{z_t} είναι $iid(0, S_z^2)$ και επιπλέον υποτίθεται ότι είναι ανεξάρτητα μεταξύ τους.

Πίνακας 3-2

Υποφαινομενικές παλινδρομήσεις λόγω κοινής σχέσης με τρίτη μεταβλητή

f_z	0	0.1	0.2	0.3	0.4	0.5
Ποσοστό απόρριψης H_0 , ($\alpha = 5\%$)	100	100	100	100	100	100
Μέση τιμή $ t $	17.47	17.54	17.75	18.26	18.93	20.06
Μέσο \hat{b}	1.60	1.60	1.61	1.63	1.65	1.68
Μέσο R^2	0.75	0.75	0.76	0.77	0.78	0.80
Ελάχιστο R^2	0.52	0.56	0.54	0.56	0.47	0.60
Μέγιστο R^2	0.89	0.87	0.89	0.89	0.91	0.91
Μέση τιμή DW	2.000	1.974	1.948	1.920	1.904	1.889
Ελάχιστο DW	1.315	1.241	1.216	1.184	1.130	1.109
Μέγιστο DW	2.841	2.745	2.686	2.708	2.655	2.666
Ποσοστό απόρριψης H_0 και $DW < d_L$	5.97	7.24	9.73	12.07	13.55	16.19
Ποσοστό απόρριψης H_0 και $DW < d_L$ και $R^2 > 0.7$	5.22	6.3	8.62	10.98	12.92	15.83
f_z	0.6	0.7	0.8	0.9	0.95	1
Ποσοστό απόρριψης H_0 , ($\alpha = 5\%$)	100	100	100	100	100	100
Μέση τιμή $ t $	21.60	23.98	28.12	36.61	45.90	67.51
Μέσο \hat{b}	1.71	1.76	1.81	1.88	1.91	1.95
Μέσο R^2	0.82	0.85	0.88	0.92	0.95	0.97
Ελάχιστο R^2	0.61	0.61	0.66	0.72	0.72	0.79
Μέγιστο R^2	0.92	0.96	0.97	0.98	0.99	1.00
Μέση τιμή DW	1.884	1.884	1.902	1.934	1.950	1.981
Ελάχιστο DW	1.185	1.124	1.157	1.195	1.261	1.192
Μέγιστο DW	2.547	2.518	2.683	2.694	2.675	2.738
Ποσοστό απόρριψης H_0 και $DW < d_L$	15.14	15.9	13.23	10.45	8.71	7.17
Ποσοστό απόρριψης H_0 και $DW < d_L$ και $R^2 > 0.7$	15.01	15.84	13.23	10.45	8.71	7.17

Για να διερευνηθεί η περίπτωση αυτή πραγματοποιήθηκαν μια σειρά προσομοιώσεων θεωρώντας ότι $m_y = 100$, $m_x = 100$, $g_y = 4$ και $g_x = 2$ και με τιμές για τα u_{y_i} , u_{x_i} και u_{z_i} προερχόμενες από $N(0, 1)$ κατανομή. Ο συντελεστής f_z του υποδείγματος (3.7) της Z_i θεωρήθηκε ότι παίρνει τις τιμές 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, καθώς και την τιμή 1 για να εξεταστεί το φαινόμενο όταν η Z_i είναι τυχαίος περίπατος. Τα αποτελέσματα των προσομοιώσεων για χρονοσειρές 100 παρατηρήσεων παρουσιάζονται στον Πίνακα 3-2.

Το βασικότερο συμπέρασμα που προκύπτει από αυτή τη διαδικασία είναι ότι το φαινόμενο της υποφαινομενικής παλινδρόμησης κάνει την εμφάνισή του σε όλες τις περιπτώσεις. Ο συντελεστής b του υποδείγματος (3.4) είναι στατιστικά σημαντικός στο 100% των περιπτώσεων ανεξάρτητα από τις τιμές που παίρνει ο συντελεστής f_z . Η μέση τιμή της στατιστικής (2.2) για τον έλεγχο της σημαντικότητας του συντελεστή b αυξάνει συνεχώς με πολύ μικρό ρυθμό για τιμές του f_z μέχρι το 0.7 και στη συνέχεια με μεγαλύτερο ρυθμό, καθώς η Z_i πλησιάζει στα όρια της μη στασιμότητας και τελικά γίνεται τυχαίος περίπατος.

Καθώς το f_z πλησιάζει στη μονάδα, η τιμή του μέσου \hat{b} αυξάνει προσεγγίζοντας το 2, δηλαδή το λόγο $\frac{g_y}{g_x}$. Όταν το f_z ισούται με τη μονάδα το μέσο \hat{b} ισούται με 1.95.

Επομένως, δεν επέρχεται πλήρη σύγκλιση.

Καθώς το f_z αυξάνει, αυξάνεται και ο συντελεστής προσδιορισμού πλησιάζοντας το 1. Η μέση τιμή της στατιστικής των Durbin και Watson ακολουθεί παράξενη πορεία. Μειώνεται μέχρι το f_z να πάρει την τιμή 0.6 και στη συνέχεια αρχίζει πάλι να αυξάνει. Άμεση συνέπεια είναι και η συμπεριφορά του ποσοστού των περιπτώσεων όπου ταυτόχρονα ο συντελεστής b είναι σημαντικός και τα σφάλματα αυτοσυσχετίζονται. Σημειώνεται πως αφού ο συντελεστής b είναι σημαντικός για όλες τις περιπτώσεις, το προηγούμενο ποσοστό ουσιαστικά είναι το ποσοστό των περιπτώσεων όπου εντοπίζεται πρόβλημα αυτοσυσχετίσης στα σφάλματα. Επομένως, συμπεραίνεται πως το ποσοστό των περιπτώσεων όπου υπάρχει πρόβλημα αυτοσυσχετίσης αυξάνεται μέχρι το f_z να πάρει την τιμή 0.6 και στην συνέχεια αρχίζει να μειώνεται. Όταν το f_z ισούται με τη μονάδα, δηλαδή όταν η Z_i γίνεται τυχαίος περίπατος, το ποσοστό αυτό είναι πολύ χαμηλό, 7.17% σχεδόν το ίδιο με το 5.97% που καταγράφηκε όταν η Z_i είναι λευκός θόρυβος. Μια λογική εξήγηση για αυτό είναι ότι καθώς το f_z αυξάνεται

εμφανίζεται πρόβλημα αυτοσυσχέτισης. Όμως, από ένα σημείο και μετά οι Y_t και X_t έχουν συμπεριφορά τυχαίων περιπάτων, με αποτέλεσμα να εμφανίζεται το φαινόμενο της συνολοκλήρωσης που δικαιολογεί την ελάττωση του προβλήματος της αυτοσυσχέτισης. Τέλος, ανάλογη συμπεριφορά παρατηρείται και για το ποσοστό των περιπτώσεων όπου ταυτόχρονα ο συντελεστής b είναι σημαντικός, τα σφάλματα αυτοσυσχετίζονται και ο συντελεστής προσδιορισμού είναι μεγαλύτερος από 0.7.

Από τα παραπάνω αποτελέσματα προκύπτει μια διαφορετική μορφή υποφαινομενικής παλινδρόμησης που δεν έχει γενεσιουργό αίτια την αυτοσυσχέτιση στα σφάλματα. Η αιτία εμφάνισης του φαινομένου είναι ότι ενώ οι Y_t και X_t είναι ασυσχέτιστες, εντούτοις δεν είναι και ανεξάρτητες αφού έχουν κοινή σχέση με την Z_t , η οποία κινεί και τις δύο προς την ίδια κατεύθυνση.

Η μορφή αυτή υποφαινομενικής παλινδρόμησης μπορεί να δώσει εξήγηση και στη δημιουργία παλινδρομήσεων χωρίς νόημα. Ένα τέτοιο παράδειγμα είναι η υψηλή σχέση που έχει εντοπιστεί κατά καιρούς και για διάφορες πόλεις ανάμεσα στον αριθμό των γεννήσεων και των αριθμό των πελαργών. Ο αριθμός των γεννήσεων συσχετίζεται με το βαθμό αστικοποίησης. Από την άλλη πλευρά, ο αριθμός των πελαργών συσχετίζεται και αυτός με το βαθμό αστικοποίησης αφού όσο μεγαλύτερος είναι τόσο περισσότερα κτίρια υπάρχουν που ευνοούν (ιδιαίτερα σε κάποιες πόλεις) την κατασκευή φωλιών και την εγκατάσταση πελαργών. Επομένως, σύμφωνα με τα αποτελέσματα της προσομοίωσης ο αριθμός των πελαργών θα συσχετίζεται με τις γεννήσεις.

3.5 Μη γραμμικές υποφαινομενικές Παλινδρομήσεις

Πολλές φορές εκτιμώνται υποδείγματα παλινδρόμησης που περιλαμβάνουν μεταβλητές στις οποίες έχει εφαρμοστεί κάποιος μη γραμμικός μετασχηματισμός. Στα υποδείγματα αυτά η σχέση των μεταβλητών με τους συντελεστές παραμένει γραμμική κάτι που επιτρέπει την εκτίμησή τους με τη μέθοδο των ελαχίστων τετραγώνων. Αυτό σημαίνει πως, όπως και στη συνηθισμένη γραμμική παλινδρόμηση, η στατιστική συμπερασματολογία σχετικά με τη σημαντικότητα των συντελεστών του υποδείγματος θα επηρεάζεται από τυχόν ύπαρξη προβλήματος αυτοσυσχέτισης στα σφάλματα. Στην ενότητα αυτή θα εξεταστεί η δυνατότητα

ύπαρξης μη γραμμικών υποφαινομενικών παλινδρομήσεων αλλά και μη γραμμικών υποφαινομενικών σχέσεων μεταξύ χρονοσειρών.

3.5.1 Η επίδραση του τετραγωνικού μετασχηματισμού της ανεξάρτητης μεταβλητής

Έστω ότι ο μηχανισμός παραγωγής των χρονοσειρών Y_t και X_t είναι τα αυτοπαλίνδρομα υποδείγματα:

$$\begin{aligned} Y_t &= f_Y Y_{t-1} + u_{Yt} \quad \text{me} \quad |f_Y| \leq 1 \\ X_t &= f_X X_{t-1} + u_{Xt} \quad \text{me} \quad |f_X| \leq 1 \end{aligned} \quad (3.8)$$

όπου το u_{Yt} είναι $iid(0, S_Y^2)$, το u_{Xt} είναι $iid(0, S_X^2)$ και είναι ανεξάρτητα μεταξύ τους. Με τη βοήθεια προσομοίωσης εξετάστηκαν οι επιπτώσεις στο φαινόμενο της υποφαινομενικής παλινδρόμησης όταν εφαρμόζεται τετραγωνικός μετασχηματισμός στην ανεξάρτητη μεταβλητή X_t . Ειδικότερα μελετήθηκαν δύο υποδείγματα παλινδρόμησης. Το υπόδειγμα της απλής γραμμικής παλινδρόμησης:

$$Y_t = a + gX_t^2 + e_t \quad (3.9)$$

στο οποίο η ανεξάρτητη μεταβλητή έχει υψωθεί στο τετράγωνο και το υπόδειγμα της πολλαπλής γραμμικής παλινδρόμησης:

$$Y_t = a + bX_t + gX_t^2 + e_t \quad (3.10)$$

το οποίο περιλαμβάνει ως πρόσθετη ανεξάρτητη μεταβλητή τον τετραγωνικό μετασχηματισμό της X_t . Δηλαδή, πρόκειται για μία καμπύλη ελαχίστων τετραγώνων δευτέρου βαθμού (παραβολή). Οι διαταρακτικοί όροι προήλθαν και πάλι από $N(0, 1)$ κατανομή. Για λόγους σύγκρισης των υποδειγμάτων (3.9) και (3.10) καταγράφηκαν, επιπλέον σε σχέση με τις προηγούμενες προσομοιώσεις, ο έλεγχος της ταυτόχρονης στατιστικής σημαντικότητας των συντελεστών με την F κατανομή και οι τιμές του προσαρμοσμένου συντελεστή προσδιορισμού. Οι τιμές των f_Y και f_X που χρησιμοποιήθηκαν ήταν 0, 0.2, 0.5, 0.8, 0.9 και 1 και θεωρήθηκαν ίσες και για τις δύο χρονοσειρές, ενώ το μέγεθος του δείγματος ήταν 100 παρατηρήσεις. Τα αποτελέσματα των προσομοιώσεων για το υπόδειγμα (3.9) περιλαμβάνονται στον Πίνακα 3-3 και για το υπόδειγμα (3.10) στον Πίνακα 3-4. Η διαπίστωση είναι ότι το φαινόμενο της υποφαινομενικής παλινδρόμησης εμφανίζεται σε σημαντικό ποσοστό και για τα δύο υποδείγματα, ακόμη και όταν η X_t είναι στάσιμη.

Δηλαδή, ο τετραγωνικός μετασχηματισμός της ανεξάρτητης μεταβλητής διατηρεί το πρόβλημα.

Πίνακας 3-3
Υποφαινομενικές παλινδρομήσεις υποδείγματος (3.9) για $\alpha = 5\%$

$f_Y = f_X$	0	0.2	0.5	0.8	0.9	1
Ποσοστό απόρριψης g	4.61	5.54	7.89	23.22	38.14	70.99
Μέση τιμή F	0.99	1.05	1.29	2.79	5.40	39.48
Μέσο R^2	0.01	0.01	0.01	0.03	0.05	0.20
Ελάχιστο R^2	0	0	0	0	0	0
Μέγιστο R^2	0.12	0.16	0.20	0.34	0.57	0.91
Μέσο \bar{R}^2	0.00	0.00	0.00	0.02	0.04	0.19
Ελάχιστο \bar{R}^2	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
Μέγιστο \bar{R}^2	0.11	0.15	0.19	0.33	0.56	0.91
Μέση τιμή DW	2.002	1.617	1.055	0.496	0.314	0.172
Ελάχιστο DW	1.189	0.973	0.505	0.143	0.047	0.007
Μέγιστο DW	2.680	2.364	1.865	1.215	0.997	0.859
Ποσοστό απόρριψης g και $DW < d_L$	0.25	3.3	7.89	23.22	38.14	70.99

Αναλυτικότερα τα συμπεράσματα που προέκυψαν από τις προσομοιώσεις είναι:

- 1) Το ποσοστό απόρριψης του ελέγχου της ταυτόχρονης στατιστικής σημαντικότητας των συντελεστών με την F στατιστική (2.3), και για τα δύο υποδείγματα, αυξάνεται καθώς τα f_Y και f_X αυξάνονται προσεγγίζοντας τη μονάδα. Όταν οι χρονοσειρές γίνονται ολοκληρωμένες διαδικασίες το ποσοστό αυτό γίνεται πολύ μεγάλο, 71% για το υπόδειγμα (3.9). Υπενθυμίζεται πως το ανάλογο ποσοστό για το απλό γραμμικό υπόδειγμα όταν δεν είχε εφαρμοστεί ο τετραγωνικός μετασχηματισμός ήταν 76.2%. Δηλαδή, με τον τετραγωνικό μετασχηματισμό το πρόβλημα ελαφρά βελτιώνεται. Το αντίστοιχο ποσοστό για το υπόδειγμα (3.10) είναι 86.52%. Ανάλογη συμπεριφορά διαπιστώθηκε και για τους ελέγχους της στατιστικής σημαντικότητας κάθε συντελεστή ξεχωριστά με την στατιστική (2.2).

Πίνακας 3-4
Υποφαινομενικές παλινδρομήσεις υποδείγματος (3.10) για $\alpha = 5\%$

$f_y = f_x$	0	0.2	0.5	0.8	0.9	1
Ποσοστό απόρριψης b	5.37	6.19	11.53	32.77	45.76	63.29
Ποσοστό απόρριψης g	4.96	5.48	7.12	20.01	31.68	54.58
Ποσοστό απόρριψης F	5.11	6.09	11.52	39.9	59.53	86.52
Μέση τιμή F	1.03	1.09	1.45	3.56	6.69	35.94
Μέσο R^2	0.02	0.02	0.03	0.06	0.11	0.31
Ελάχιστο R^2	0	0	0	0	0	0
Μέγιστο R^2	0.19	0.17	0.22	0.59	0.71	0.94
Μέσο \bar{R}^2	0.00	0.00	0.01	0.04	0.09	0.29
Ελάχιστο \bar{R}^2	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02
Μέγιστο \bar{R}^2	0.17	0.15	0.20	0.58	0.71	0.94
Μέση τιμή DW	2.003	1.628	1.071	0.527	0.352	0.223
Ελάχιστο DW	1.227	0.929	0.471	0.135	0.067	0.010
Μέγιστο DW	2.700	2.284	1.785	1.157	1.165	0.909
Ποσοστό απόρριψης g και $DW < d_L$	0.45	3.94	11.51	39.9	59.53	86.52

- 2) Η μέση τιμή του συντελεστή προσδιορισμού για τιμές των f_y και f_x μέχρι το 0.8 είναι κοντά στο 0 και στα δύο υποδείγματα. Όταν τα f_y και f_x πάρουν μεγαλύτερες τιμές, η μέση τιμή του συντελεστή προσδιορισμού αυξάνεται. Συγκρίνοντας τα δύο υποδείγματα με τον προσαρμοσμένο συντελεστή προσδιορισμού παρατηρείται ότι το υπόδειγμα (3.10) έχει μεγαλύτερο μέσο προσαρμοσμένο συντελεστή προσδιορισμού από το (3.9) αλλά μόνο για τιμές των f_y και f_x 0.8, 0.9 και 1. Για τις άλλες τιμές και στα δύο υποδείγματα είναι 0.
- 3) Η μέση τιμή της στατιστικής των Durbin και Watson έχει φθίνουσα πορεία και στα δύο υποδείγματα.

- 4) Το ποσοστό των περιπτώσεων όπου ταυτόχρονα οι συντελεστές είναι σημαντικοί και τα σφάλματα αυτοσυσχετίζονται είναι το ίδιο με το ποσοστό των περιπτώσεων που οι συντελεστές είναι σημαντικοί.

Από την τελευταία παρατήρηση συμπεραίνεται πως και σε αυτή την περίπτωση η αίτια της εμφάνισης του φαινομένου της υποφαινομενικής παλινδρόμησης είναι η αυτοσυσχέτιση στα σφάλματα. Επομένως, μπορεί να θεωρηθεί πως ο τετραγωνικός μετασχηματισμός της ανεξάρτητης μεταβλητής, δεν επηρεάζει τις ιδιότητες αυτοσυσχέτισης των χρονοσειρών που περνάνε και πάλι στα σφάλματα οδηγώντας στην εμφάνιση του φαινομένου.

3.5.2 Λογαριθμικές υποφαινομενικές παλινδρομήσεις

Συνηθισμένη πρακτική στην ανάλυση παλινδρόμησης αποτελεί ο λογαριθμικός μετασχηματισμός των χρησιμοποιούμενων στο υπόδειγμα μεταβλητών. Ο de Jong (2003) μελέτησε τις ασυμπτωτικές ιδιότητες της παλινδρόμησης όταν το υπόδειγμα της απλής γραμμικής παλινδρόμησης:

$$Y_t = a + bX_t + e_t$$

μετασχηματίζεται σε:

$$\log|Y_t| = a + b \log|X_t| + e_t \quad (3.11)$$

όπου οι Y_t και X_t είναι οι τυχαίοι περίπατοι:

$$\begin{aligned} Y_t &= Y_{t-1} + v_t \\ X_t &= X_{t-1} + w_t. \end{aligned}$$

Επίσης υποτίθεται πως:

- $v_t = \sum_{k=0}^{\infty} f_k^y h_{t-k}$ και $w_t = \sum_{k=0}^{\infty} f_k^x \epsilon_{t-k}$ όπου (ϵ_t, h_t) είναι μια ακολουθία από i.i.d. τυχαία διανύσματα με μηδενικό μέσο και μη ιδιάζων πίνακα συνδιακύμανσης Σ .
- $\sum_{k=0}^{\infty} f_k^x \neq 0$ και $\sum_{k=0}^{\infty} f_k^y \neq 0$.
- Το (x_0, y_0) είναι ένα αυθαίρετο τυχαίο διάνυσμα το οποίο είναι ανεξάρτητο από όλα τα (w_t, v_t) με $t \geq 1$.
- $\sum_{k=0}^{\infty} k \left(|f_k^x| + |f_k^y| \right) < \infty$ και $E|\epsilon_t|^p + E|h_t|^p < \infty$ για κάποιο $p > 2$.

- Οι κατανομές των ϵ_t και h_t είναι απολύτως συνεχείς σε σχέση με το μέτρο Lebesgue και έχουν χαρακτηριστικές συναρτήσεις $y_x(s)$ και $y_y(s)$ για τις οποίες ισχύει ότι

$$\lim_{s \rightarrow \infty} s^h (y_x(s) + y_y(s)) = 0 \text{ για κάποιο } h > 0.$$

Από τις προηγούμενες υποθέσεις συνεπάγεται ότι όταν $T \geq M$ για κάποια σταθερή τιμή του M , οι πυκνότητες των $T^{-1/2}X_T$ και $T^{-1/2}Y_T$ υπάρχουν και είναι ομοιόμορφα ορισμένες στο $T \geq M$. Επίσης προκύπτει ότι:

$$(T^{-1/2}X_{[T_1]}, T^{-1/2}Y_{[T_2]}) \Rightarrow (S_X W(r_1), S_Y V(r_2))$$

όπου \Rightarrow σημαίνει ασθενής σύγκλιση όπου $W(\cdot)$ και $V(\cdot)$ είναι οι κινήσεις Brown που σχετίζονται με τις X_t και Y_t ενώ $S_X^2 = \lim_{T \rightarrow \infty} E(T^{-1/2}X_T)^2$ και $S_Y^2 = \lim_{T \rightarrow \infty} E(T^{-1/2}Y_T)^2$.

Εφόσον ισχύουν οι παραπάνω υποθέσεις και εκτιμηθεί το υπόδειγμα (3.11) με τη μέθοδο των ελαχίστων τετραγώνων ο de Jong (2003) αποδεικνύει ότι:

- 1) Για τον εκτιμητή του συντελεστή της κλίσης ισχύει:

$$\hat{b} \xrightarrow{d} \frac{\int_0^1 (\log |S_X W(s)| - \int_0^1 \log |S_X W(r)| dr) (\log |S_Y V(s)| - \int_0^1 \log |S_X W(r)| dr) ds}{\int_0^1 (\log |S_X W(s)| - \int_0^1 \log |S_X W(r)| dr)^2 ds} = B$$

και για τον εκτιμητή του σταθερού όρου

$$(\log(T))^{-1} \hat{a} \xrightarrow{d} \frac{1}{2} - \frac{1}{2} B.$$

Δηλαδή, ο εκτιμητής του συντελεστή της κλίσης \hat{b} του υποδείγματος (3.11) είναι ασυμπτωτικά τυχαίος.

- 2) Για τις στατιστικές (2.2) του συντελεστή της κλίσης και του σταθερού όρου ισχύουν αντίστοιχα:

$$T^{-1/2} \hat{t}_a \xrightarrow{d} T_1$$

και

$$T^{-1/2} \hat{t}_b \xrightarrow{d} T_2$$

όπου T_1 και T_2 είναι μη εκφυλισμένες τυχαίες μεταβλητές. Δηλαδή, οι τιμές των στατιστικών (2.2) θα τείνουν ασυμπτωτικά στο άπειρο καθώς το μέγεθος του δείγματος αυξάνει με ρυθμό $T^{1/2}$.

Τα αποτελέσματα αυτά είναι παρόμοια με την περίπτωση της υποφαινομενικής παλινδρόμησης που μελέτησε ο Phillips (1986). Επομένως, σύμφωνα με τον de Jong το

συμπέρασμα που προκύπτει είναι ότι το φαινόμενο της υποφαινομενικής παλινδρόμησης μπορεί να προκύψει και στην περίπτωση όπου έχει εφαρμοστεί λογαριθμικός μετασχηματισμός των μεταβλητών του υποδείγματος, εφόσον αυτές είναι ολοκληρωμένες διαδικασίες.

3.5.3 Η συμπεριφορά των ελέγχων μη γραμμικότητας για ολοκληρωμένες διαδικασίες

Για να διερευνηθεί η ύπαρξη μη γραμμικής σχέσης μεταξύ δύο μεταβλητών έχουν προταθεί μια σειρά μη γραμμικών στατιστικών ελέγχων, οι περισσότεροι από τους οποίους βασίζονται σε ένα υπόδειγμα της μορφής:

$$Y_t = q_0 + q_1 X_t + g(X_t, \mathbf{a}, \mathbf{b}) + e_t$$

όπου η μη γραμμική συνάρτηση $g(X_t, \mathbf{a}, \mathbf{b})$ εξετάζει την πιθανότητα μη γραμμικής συμβολής της X_t , τα \mathbf{a} και \mathbf{b} είναι παράμετροι που ελέγχουν τη μη γραμμική συμπεριφορά της, ενώ υποτίθεται ότι $g(X_t, \mathbf{a}, \mathbf{b} = 0) = 0$, έτσι ώστε η μηδενική υπόθεση του ελέγχου να είναι η $H_0 : \mathbf{b} = 0$.

Οι Lee, Kim και Newbold (2004) χρησιμοποιώντας προσομοίωση μελέτησαν τη συμπεριφορά έξι ελέγχων για την ύπαρξη μη γραμμικότητας όταν οι ελεγχόμενες μεταβλητές είναι ανεξάρτητοι τυχαίοι περίπατοι. Οι έλεγχοι αυτοί είναι ο έλεγχος RESET του Ramsey, ο έλεγχος των McLeod και Li, ο έλεγχος του Keenan, ο έλεγχος νευρονικού δικτύου του White, ο έλεγχος της δυναμικής μήτρας πληροφοριών του White με τρεις διαφορετικές στατιστικές συναρτήσεις ελέγχου και ο εύκαμπτος έλεγχος του Hamilton. Δεν θα δοθούν λεπτομέρειες για τους τρόπους που εφαρμόζονται αυτοί οι έλεγχοι καθώς και για το πώς ορίζονται οι στατιστικές τους συναρτήσεις παραπέμποντας για περισσότερες πληροφορίες στην εργασία των Lee, Kim και Newbold (2004) αλλά θα αναφερθούν τα συμπεράσματα στα οποία κατέληξαν από τη προσομοίωση.

Από τον Πίνακα 3-5 που περιέχει τα αποτελέσματα της προσομοίωσης των Lee, Kim και Newbold (2004) προκύπτει πώς ακόμα και για πολύ μικρό μέγεθος δείγματος, όπως για $T=50$, τα ποσοστά απόρριψης όλων των ελέγχων είναι σημαντικά μεγαλύτερα από το επίπεδο σημαντικότητας του 5%. Το σοβαρότερο πρόβλημα υποφαινομενικής μη γραμμικότητας εντοπίζεται στους ελέγχους του White, οι οποίοι απορρίπτουν τη μηδενική υπόθεση της γραμμικότητας σχεδόν στο σύνολο των περιπτώσεων και ακολουθούν οι έλεγχοι του McLeod

(67%) και του Hamilton (51%). Οι έλεγχοι του Reset (40%), του Keenan (38%) και του νευρονικού

Πίνακας 3-5
Ποσοστά απόρριψης ελέγχων μη γραμμικότητας για $\alpha = 5\%$

T	50	100	500	1000	10000
RESET	40	63	89	96	100
McLeod	67	99	100	100	100
Keenan	38	55	80	85	95
Neural	38	56	80	85	96
White 1	100	100	100	100	100
White 2	100	100	100	100	100
White 3	100	100	100	100	100
Hamilton	51	73	98	100	100

Πηγή: Lee, Kim και Newbold (2004)

δικτύου (38%) είναι οι πιο ανθεκτικοί από όλους που εξετάστηκαν. Ένα άλλο συμπέρασμα που προκύπτει είναι ότι καθώς αυξάνεται το μέγεθος του δείγματος, τα ποσοστά απόρριψης των ελέγχων των White, McLeod και Hamilton συγκλίνουν γρήγορα στη μονάδα, ενώ ο ρυθμός σύγκλισης είναι αρκετά μικρότερος για τους άλλους ελέγχους. Το τελικό συμπέρασμα στο οποίο κατέληξαν οι Lee, Kim και Newbold (2004) είναι ότι για κανέναν από τους προηγούμενους ελέγχους η στατιστική συνάρτηση δεν συγκλίνει κατά κατανομή στην ασυμπτωτική της κατανομή αλλά στην πραγματικότητα τείνει να αποκλίνει στο άπειρο.

3.6 Ανακεφαλαίωση

Από την ανάπτυξη του κεφαλαίου αυτού γίνεται κατανοητό ότι το φαινόμενο της υποφαινομενικής παλινδρόμησης μπορεί να συναντηθεί εκτός με τις συνηθισμένες περιπτώσεις των ολοκληρωμένων και αυτοσυσχετιζομένων διαδικασιών και με άλλου είδους χρονοσειρές. Επίσης, το φαινόμενο αυτό μπορεί να προκύψει και σε αρκετές άλλες περιπτώσεις πέραν του υποδείγματος της απλής γραμμικής παλινδρόμησης. Τέτοια παραδείγματα είναι τα υποδείγματα που γίνεται η απαλοιφή της τάσης όταν οι εμπλεκόμενες

χρονοσειρές έχουν εσφαλμένα ταυτοποιηθεί. Ακόμη, διαπιστώθηκε πως το φαινόμενο της υποφαινομενικής παλινδρόμησης διατηρείται σε μετασχηματισμούς των χρονοσειρών όταν αυτές είναι αυτοσυσχετιζόμενες διαδικασίες.

ΚΕΦΑΛΑΙΟ 4

ΥΠΟΦΑΙΝΟΜΕΝΙΚΕΣ ΣΥΣΧΕΤΙΣΕΙΣ

4.1 Εισαγωγή

Το φαινόμενο της ύπαρξης υποφαινομενικής σχέσης μεταξύ δύο ανεξαρτήτων χρονοσειρών, εκτός από το υπόδειγμα της απλής γραμμικής παλινδρόμησης, μπορεί να προκύψει και όταν εξετάζεται ο συντελεστής γραμμικής συσχέτισης για τις δύο χρονοσειρές. Ο δειγματικός συντελεστής γραμμικής συσχέτισης μετράει το βαθμό της γραμμικής εξάρτησης μεταξύ δύο χρονοσειρών Y_t και X_t και υπολογίζεται από τη σχέση:

$$r = \frac{Cov(Y_t, X_t)}{\sqrt{Var(Y_t)}\sqrt{VarX_t}} = \frac{\sum_{t=1}^T (X_t - \bar{X})(Y_t - \bar{Y})}{\sqrt{\sum_{t=1}^T (X_t - \bar{X})^2} \sqrt{\sum_{t=1}^T (Y_t - \bar{Y})^2}} \quad (4.1)$$

όπου $Cov(Y_t, X_t)$ είναι η συνδιακύμανση των δύο χρονοσειρών ενώ $Var(Y_t)$ και $VarX_t$ οι διακυμάνσεις τους.

Το εύρος τιμών του συντελεστή συσχέτισης είναι:

$$-1 \leq r \leq 1$$

Όσο πιο κοντά στο 1 είναι η τιμή του συντελεστή συσχέτισης τόσο μεγαλύτερη θετική σχέση υπάρχει ανάμεσα στις δύο χρονοσειρές. Αντίθετα, όσο πιο κοντά είναι η τιμή του συντελεστή συσχέτισης στο -1 τόσο μεγαλύτερη είναι η αρνητική σχέση ανάμεσα στις δύο χρονοσειρές. Όταν ο συντελεστής συσχέτισης πάρει την τιμή 1 σημαίνει ότι υπάρχει πλήρης θετική συσχέτιση, ενώ όταν πάρει την τιμή -1 πλήρης αρνητική. Μια μηδενική τιμή του συντελεστή συσχέτισης σημαίνει ότι οι δύο χρονοσειρές δεν συσχετίζονται γραμμικά.

Για να ελεγχθεί η στατιστική σημαντικότητα του συντελεστή συσχέτισης χρησιμοποιείται η στατιστική:

$$t = \frac{r}{\sqrt{\frac{1-r^2}{T-2}}} \quad (4.2)$$

όπου r είναι η τιμή του δειγματικού συντελεστή συσχέτισης και $\sqrt{\frac{1-r^2}{T-2}}$ το τυπικό του σφάλμα. Ο έλεγχος ορίζεται ως:

$$H_0 : r = 0$$

$$H_1 : r \neq 0$$

και απορρίπτεται η μηδενική υπόθεση όταν $|t| > |t_{T-2, a/2}|$ όπου a το επιθυμητό επίπεδο σημαντικότητας.

Έστω τώρα το υπόδειγμα της απλής γραμμικής παλινδρόμησης:

$$Y_t = a + bX_t + e_t$$

Ο εκτιμητής του συντελεστή της κλίσης του προηγούμενου υποδείγματος μπορεί να γραφεί:

$$\hat{b} = \frac{Cov(Y_t, X_t)}{Var(X_t)} = \frac{\sum_{t=1}^T (X_t - \bar{X})(Y_t - \bar{Y})}{\sum_{t=1}^T (X_t - \bar{X})^2} = \frac{\sum_{t=1}^T (X_t - \bar{X})(Y_t - \bar{Y})}{\sqrt{\sum_{t=1}^T (X_t - \bar{X})^2} \sqrt{\sum_{t=1}^T (X_t - \bar{X})^2}} \frac{\sqrt{\sum_{t=1}^T (Y_t - \bar{Y})^2}}{\sqrt{\sum_{t=1}^T (Y_t - \bar{Y})^2}}$$

και τελικά:

$$\hat{b} = r \frac{\sqrt{\sum_{t=1}^T (Y_t - \bar{Y})^2}}{\sqrt{\sum_{t=1}^T (X_t - \bar{X})^2}} \quad (4.3)$$

Ο συντελεστής προσδιορισμού μπορεί να γραφεί:

$$\begin{aligned} R^2 &= \frac{SSR}{SST} = \frac{\sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2}{\sum_{t=1}^T (Y_t - \bar{Y})^2} = \frac{\sum_{t=1}^T (\hat{a} + \hat{b}X_t - \bar{Y})^2}{\sum_{t=1}^T (Y_t - \bar{Y})^2} = \frac{\sum_{t=1}^T (\bar{Y} - \hat{b}\bar{X} + \hat{b}X_t - \bar{Y})^2}{\sum_{t=1}^T (Y_t - \bar{Y})^2} = \frac{\sum_{t=1}^T [\hat{b}(X_t - \bar{X})]^2}{\sum_{t=1}^T (Y_t - \bar{Y})^2} \\ &= \frac{\hat{b}^2 \sum_{t=1}^T (X_t - \bar{X})^2}{\sum_{t=1}^T (Y_t - \bar{Y})^2} = \left[\frac{\sum_{t=1}^T (X_t - \bar{X})(Y_t - \bar{Y})}{\sum_{t=1}^T (X_t - \bar{X})^2} \right] \frac{\sum_{t=1}^T (X_t - \bar{X})^2}{\sum_{t=1}^T (Y_t - \bar{Y})^2} = \left[\frac{\sum_{t=1}^T (X_t - \bar{X})(Y_t - \bar{Y})}{\sqrt{\sum_{t=1}^T (X_t - \bar{X})^2} \sqrt{\sum_{t=1}^T (Y_t - \bar{Y})^2}} \right]^2 \end{aligned}$$

και τελικά:

$$R^2 = r^2 \quad (4.4)$$

Δηλαδή, το τετράγωνο του συντελεστή προσδιορισμού ισούται με το τετράγωνο του συντελεστή συσχέτισης.

Από τις (4.3) και (4.4) είναι φανερό ότι υπάρχει στενή σχέση ανάμεσα στην παλινδρόμηση και στο συντελεστή γραμμικής συσχέτισης. Κατά συνέπεια, μια υποφαινομενική σχέση μεταξύ δύο χρονοσειρών που προκύπτει από ένα υπόδειγμα απλής παλινδρόμησης θα πρέπει να εκφράζεται και μέσω του συντελεστή συσχέτισης. Δηλαδή, οι ιδιότητες στασιμότητας και αυτοσυσχέτισης των χρονοσειρών που, όπως έχει αναλυθεί στα προηγούμενα κεφάλαια, επηρεάζουν τη μεθοδολογία της παλινδρόμησης θα πρέπει να επηρεάζουν και το συντελεστή συσχέτισης (4.1) καθώς και τον έλεγχο (4.2) για τη στατιστική του σημαντικότητας. Όμως, στην περίπτωση του συντελεστή συσχέτισης η υποφαινομενική σχέση μεταξύ δύο χρονοσειρών είναι πιο εύκολο να γίνει εσφαλμένα αποδεκτή. Αυτό συμβαίνει γιατί δεν υπάρχει κάποιο άλλο κριτήριο που θα μπορούσε να χρησιμοποιηθεί για να κριθεί η βασιμότητα της σχέσης, ενώ στην παλινδρόμηση συμβουλευτικό ρόλο έχει το κριτήριο των Durbin και Watson.

Επιστήμονες όπως ο Yule, ο Pearson και ο Neyman παρατήρησαν από την αρχή της στατιστικής επιστήμης την ύπαρξη υποφαινομενικών συσχετίσεων μεταξύ μεταβλητών και είχαν προσπαθήσει να δώσουν λογικές εξηγήσεις στο φαινόμενο μελετώντας τις ιδιότητες του συντελεστή συσχέτισης. Αρκετά χρόνια αργότερα οι Granger και Newbold (1974) αναφέρθηκαν στην εμφάνιση του φαινομένου με υποδείγματα παλινδρόμησης και διαπίστωσαν ως αιτία την έλλειψη στασιμότητας των χρονοσειρών και κατά συνέπεια την αυτοσυσχέτιση στα σφάλματα. Στο κεφάλαιο αυτό θα εξεταστεί το φαινόμενο της ύπαρξης υποφαινομενικής σχέσης μεταξύ δύο ανεξαρτήτων χρονοσειρών μέσω του συντελεστή συσχέτισης.

4.2 Η συσχέτιση δύο απλών αρμονικών συναρτήσεων του χρόνου που διαφέρουν κατά ένα τέταρτο στη φάση

Ο Yule (1926) παρατήρησε ότι συχνά προκύπτουν μεταξύ ποσοτήτων που μεταβάλλονται με το χρόνο (*time – variables*) πολύ υψηλές συσχετίσεις στις οποίες όμως δεν μπορεί να δοθεί κάποια φυσική σημασία, μολονότι με τους συνηθισμένους ελέγχους η συσχέτιση θεωρείται σημαντική. Ένα τέτοιο παράδειγμα αποτελούσε η σχέση ανάμεσα στην πτώση της αναλογίας

των γάμων της εκκλησίας της Αγγλίας ως προς όλους τους γάμους και στην πτώση του ποσοστού θνησιμότητας για τη χρονική περίοδο 1866 – 1911. Η συσχέτιση ανάμεσα στις δύο μεταβλητές υπολογίζονταν σε 0.9512, με το τυπικό σφάλμα του συντελεστή συσχέτισης να είναι πολύ μικρό 0.0140

Η πτώση στην αναλογία των γάμων της εκκλησίας της Αγγλίας οφειλόταν στη διάδοση της επιστημονικής σκέψης από το 1866 και η πτώση στο ποσοστό θνησιμότητας στην πρόοδο της επιστήμης. Επομένως, αφού και οι δύο μεταβλητές επηρεάζονταν από κάποιον κοινό παράγοντα είναι λογικό να παρουσιάζουν υψηλή συσχέτιση. Ένας άλλος τρόπος για να ερμηνευτεί η συσχέτιση θα ήταν να θεωρηθεί σαν αιτία που την προκαλεί η κοινή επίδραση του παράγοντα χρόνου. Ουσιαστικά και η άποψη αυτή λέει ότι υπάρχει κάποια τρίτη κοινή ποσότητα που μεταβάλλεται με το χρόνο. Όμως, ο Yule τονίζει ότι μια τέτοια συσχέτιση δεν πρέπει να ερμηνευτεί με κάποιο τρόπο αλλά δεν έχει καθόλου νόημα. Είναι απλά μια απάτη και αν υπήρχαν στοιχεία για τις μεταβλητές για μια αρκετά μεγαλύτερη χρονική περίοδο δεν θα προέκυπτε καμία αξιοσημείωτη συσχέτιση, ενώ ο τύπος που χρησιμοποιείται για τον

υπολογισμό του τυπικού σφάλματος $\frac{1-r^2}{\sqrt{n}}$ δίνει εντελώς λανθασμένο αποτέλεσμα. Ο Yule

αποδεικνύει ότι όταν οι διαδοχικοί όροι δύο χρονοσειρών σε ένα δείγμα δεν αποτελούν τυχαία σειρά (λευκό θόρυβο) αλλά μια σειρά στην οποία οι διαδοχικοί όροι σχετίζονται στενά ο ένας με τον άλλο, τότε η συνηθισμένη μεθοδολογία δεν μπορεί να εφαρμοστεί και ο τύπος δίνει λανθασμένα αποτελέσματα.

Για να μελετήσει ο Yule το φαινόμενο θεωρεί ότι κάθε μια από τις δύο μεταβλητές προκύπτει ως μια απλή αρμονική συνάρτηση του χρόνου. Επιπλέον, θεωρεί ότι οι δύο συναρτήσεις διαφέρουν μεταξύ τους κατά ένα τέταρτο στη φάση. Δύο τέτοιες συναρτήσεις έχουν την ακόλουθη μορφή:

$$y_1 = \sin 2p \frac{t}{T}$$

και

$$y_2 = \sin 2p \frac{t+t}{T} \text{ για } \frac{t}{T} = \frac{1}{4}$$

όπου T η περίοδος, t η διαφορά φάσης, το πλάτος θεωρείται μονάδα και εξετάζεται η συσχέτιση μεταξύ ταυτόχρονων τιμών των y_1 και y_2 εντός ενός διαστήματος $\pm h$ γύρω από τον χρόνο u .

Ο Yule (1926) αποδεικνύει ότι οι μέσοι όροι M_1 και M_2 για δύο τέτοιες μεταβλητές δίνονται από τις σχέσεις:

$$M_1 = \frac{T}{2ph} \sin 2p \frac{u}{T} \sin 2p \frac{h}{T}$$

και

$$M_2 = \frac{T}{2ph} \cos 2p \frac{u}{T} \sin 2p \frac{h}{T}$$

Για τα τυπικά τους σφάλματα Σ_1^2 και Σ_2^2 ισχύουν οι σχέσεις:

$$\Sigma_1^2 = \frac{1}{2} - \frac{T}{8ph} \cos 4p \frac{u}{T} \sin 4p \frac{h}{T}$$

και

$$\Sigma_2^2 = \frac{1}{2} + \frac{T}{8ph} \cos 4p \frac{u}{T} \sin 4p \frac{h}{T}$$

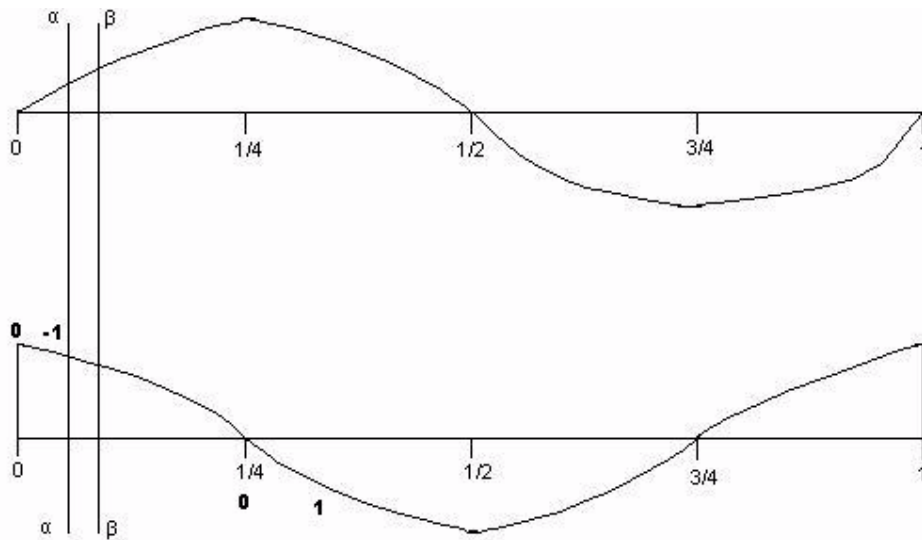
ενώ το μέσο γινόμενο (mean product) r' των y_1 και y_2 είναι:

$$r' = \frac{T}{8ph} \sin 4p \frac{u}{T} \sin 4p \frac{h}{T}$$

Δύο απλές αρμονικές συναρτήσεις του χρόνου που διαφέρουν μεταξύ τους κατά ένα τέταρτο στη φάση έχουν το χαρακτηριστικό ότι σχηματίζουν καμπύλες οι οποίες για μια ολόκληρη περίοδο 0–1 έχουν συσχέτιση 0. Αλλά, στην πραγματικότητα εάν η ολόκληρη περίοδος 0–1 αντιπροσωπεύει πολλούς αιώνες το δείγμα καλύπτει όχι περισσότερο από κάποιο μικρό διάστημα της όλης περιόδου, σαν αυτό που περιλαμβάνεται μεταξύ των δύο καθέτων $\alpha\alpha$, $\beta\beta$ στο Διάγραμμα 4-1. Αυτό το διάστημα είναι τόσο μικρό έτσι ώστε τα τμήματα των δύο καμπυλών που περιλαμβάνονται σε αυτό να είναι σχεδόν ίσιες γραμμές, η μία αυξανόμενη και η άλλη μειωμένη. Έτσι, η συσχέτιση ανάμεσα στις αντίστοιχες παρατηρήσεις θα είναι πολύ κοντά στο –1.

Έστω ότι το διάστημα γίνεται απειροελάχιστα μικρό έτσι ώστε τα τμήματα των δύο καμπυλών να μπορούν να θεωρηθούν αυστηρά γραμμικά. Αν το κέντρο του διαστήματος τοποθετηθεί στο 0, η συσχέτιση πρέπει να είναι 0, επειδή το τμήμα από τη μία καμπύλη είναι οριζόντιο και οι τιμές της δεύτερης μεταβλητής είναι οι ίδιες για όλες τις τιμές της πρώτης. Αλλά, μόλις το κέντρο του διαστήματος μετακινηθεί απλά δεξιά από το 0, το τμήμα της επάνω καμπύλης αυξάνει, ενώ αυτό της κάτω πέφτει. Έτσι, η συσχέτιση γίνεται –1. Η τιμή αυτή παραμένει μέχρι το κέντρο του διαστήματος να περάσει πέρα του σημείου $t = \frac{1}{4}$, όταν η συσχέτιση γίνεται απότομα ξανά 0. Μόλις το κέντρο του διαστήματος περάσει αυτό το σημείο τα τμήματα και από τις δύο καμπύλες πέφτουν και κατά συνέπεια η συσχέτιση γίνεται

+1. Αυτή η τιμή παραμένει μέχρι το κέντρο του διαστήματος να φτάσει τη μέση περίοδο όπου ο κύκλος ξανά επαναλαμβάνεται.



Διάγραμμα 4-1

Δύο ημιτονοειδείς καμπύλες που διαφέρουν κατά ένα τέταρτο στη φάση

Όμως, το δείγμα καλύπτει όχι περισσότερο από ένα πρακτικά απειροελάχιστο διάστημα της όλης περιόδου και το κέντρο αυτού του διαστήματος είναι ισοδύναμα πιθανό να πέσει σε οποιοδήποτε σημείο μεταξύ των χρόνων 0 και 1. Σκοπός είναι να βρεθεί η κατανομή συχνοτήτων των συσχετίσεων για μια τέτοια σειρά. Οι τιμές +1 και -1 είναι οι μοναδικές τιμές της συσχέτισης που συμβαίνουν με πεπερασμένη συχνότητα και κάθε μία από αυτές παραμένει για πάνω από το μισό του ολοκλήρου εύρους στο οποίο το κέντρο του διαστήματος μπορεί να πέσει. Έτσι, η κατανομή συχνοτήτων έχει ξεσπάσει εξωτερικά σε μια τεταγμένη στο -1 και σε μία ίση στο +1. Κατά συνέπεια, δεν είναι δυνατόν να υπάρχουν άμεσες τιμές για τη συσχέτιση. Αν το διάστημα για το οποίο υπάρχουν διαθέσιμα στοιχεία, αντί να είναι απειροελάχιστο, κάλυπτε μια ολόκληρη περίοδο η συσχέτιση θα ήταν μηδέν. Η κατανομή συχνοτήτων των τιμών της συσχέτισης παίρνοντας μια σειρά από τυχαία δείγματα, το καθένα με μήκος μιας ολόκληρης περιόδου, θα ήταν απλά μια τεταγμένη στο μηδέν. Ο Yule εξετάζει περαιτέρω το σχήμα που πρέπει να έχει η κατανομή των συσχετίσεων μεταξύ δύο εμπειρικών σειρών ώστε να δίνουν συσχετίσεις χωρίς νόημα. Αποδεικνύει ότι η κατανομή των συσχετίσεων πρέπει να είναι σχήματος U. Το φαινόμενο ότι η κατανομή των συσχετίσεων έχει σχήμα U οφείλεται καθαρά στο γεγονός ότι ένα μικρό τμήμα από μια

ημιτονοειδή καμπύλη που παίρνεται τυχαία, τείνει να είναι είτε αυξανόμενο είτε μειούμενο, όχι επίπεδο, και κατά συνέπεια κάθε σημείο τείνει να δίνει υψηλές συσχετίσεις με άλλα τμήματα που παίρνονται τυχαία.

Έστω r_1 η σειριακή συσχέτιση για μια σειρά μεταξύ των διαδοχικών της όρων U_s και U_{s+1} , όπου U_s η τιμή της μεταβλητής το έτος S και U_{s+1} η τιμή της μεταβλητής το έτος S+1, r_2 η σειριακή συσχέτιση μεταξύ των όρων U_s και U_{s+2} κ.τ.λ. Ο Yule αποδεικνύει ότι ισχύει:

$$r = \cos 2p \frac{t}{T}$$

Για τυχαία δείγματα μεγέθους n από δύο εμπειρικές σειρές για να προκύψει μια κατανομή συσχετίσεων σχήματος U, κάθε σειρά θα πρέπει να παρουσιάζει θετικές τιμές για τις σειριακές της συσχετίσεις μέχρι το r_{n-1} αλλά επιπροσθέτως και η σειρά των διαφορών της θα πρέπει να παρουσιάζει θετικές σειριακές συσχετίσεις μέχρι το όριο του δείγματος. Οι σειριακές συσχετίσεις της σειράς των διαφορών είναι ο πραγματικά σημαντικός παράγοντας. Με βάση τις σειριακές συσχετίσεις ο Yule ταξινομεί τις στατιστικές σειρές στις ακόλουθες κατηγορίες:

- Τυχαίες σειρές (*Random Series*): είναι οι σειρές για τις οποίες όλες οι σειριακές συσχετίσεις σε μία αόριστα επιμήκης σειρά είναι μηδέν.
- Συναφείς σειρές (*Conjunct Series*): είναι οι σειρές για τις οποίες όλες οι σειριακές συσχετίσεις είναι θετικές.
- Μη συναφείς σειρές (*Disjunct Series*): είναι οι σειρές για τις οποίες όλες οι σειριακές συσχετίσεις είναι αρνητικές.
- Ταλαντευόμενες σειρές (*Oscillatory Series*): είναι οι σειρές για τις οποίες οι σειριακές συσχετίσεις αλλάζουν και γίνονται εναλλασσόμενα θετικές και αρνητικές.

Οι συναφείς σειρές μπορούν να χωριστούν σε δύο υποκατηγορίες: σε συναφείς σειρές των οποίων οι διαφορές είναι τυχαίες και σε συναφείς σειρές των οποίων οι διαφορές είναι επίσης συναφείς σειρές. Οι σειρές αυτές δεν είναι απαραίτητα συσχετισμένες με το χρόνο.

Το σημαντικό θέμα στη προηγούμενη ταξινόμηση είναι η μορφή της συνάρτησης που συνδέει το r_k με το k , η οποία υποδεικνύει τη φύση των σειριακών συσχετίσεων μεταξύ των διαφορών της χρονοσειράς. Εάν η συνάρτηση αυτή είναι γραμμική η χρονοσειρά έχει τυχαίες

διαφορές, ενώ εάν δίνει μια γραφική παράσταση κοίλη προς τα κάτω, οι συσχετίσεις των διαφορών είναι θετικές.

Το συμπέρασμα στο οποίο τελικά καταλήγει ο Yule και που το επιβεβαιώνει και πειραματικά είναι ότι δείγματα από τυχαίες σειρές δίνουν μια κατανομή συσχετίσεων συμμετρική που μοιάζει με κανονική, δείγματα από συναφείς σειρές με τυχαίες διαφορές δίνουν μια πολύ πλατιά διασκορπισμένη κατανομή συσχετίσεων, ενώ δείγματα από συναφείς σειρές με συναφείς διαφορές δίνουν μια τελείως σχήματος U κατανομή, με πάνω από το 1/3 των συσχετίσεων να υπερβαίνουν το 0.9.

Κατά συνέπεια, σύμφωνα με τον Yule, η απάντηση στο πρόβλημα γιατί μερικές φορές προκύπτουν μεταξύ χρονοσειρών συσχετίσεις χωρίς νόημα, είναι ότι μερικές χρονοσειρές είναι συναφείς σειρές με συναφείς διαφορές και όταν εξετάζονται δείγματα από δύο τέτοιες σειρές η κατανομή των συσχετίσεών τους είναι σχήματος U. Άρα, για δείγματα από δύο τέτοιες σειρές οι συσχετίσεις θα είναι υψηλές αλλά δεν θα έχουν καμία σχέση με την πραγματική τιμή της συσχέτισης μεταξύ των σειρών που θα υπολογίζονταν εάν τα διαθέσιμα στοιχεία κάλυπταν μια πάρα πολύ μεγάλη χρονική περίοδο.

4.3 Τα συμπεράσματα του Yule και οι ολοκληρωμένες διαδικασίες

Η εργασία του Yule (1926) εμπλέκει πολλές από τις έννοιες της σύγχρονης ανάλυσης χρονοσειρών. Συγκεκριμένα, αποτελεί την πρώτη αναφορά στις ολοκληρωμένες διαδικασίες και στις επιπτώσεις που προκύπτουν όταν χρησιμοποιούνται για την διερεύνηση σχέσεων μεταξύ μεταβλητών. Από την κατηγοριοποίηση που εφαρμόζει ο Yule στις χρονοσειρές γίνεται αντιληπτό ότι ο όρος τυχαίες σειρές περιγράφει τις χρονοσειρές που αποτελούνται από λευκό θόρυβο, ο όρος συναφείς σειρές με τυχαίες διαφορές τις ολοκληρωμένες διαδικασίες πρώτης τάξης και ο όρος συναφείς σειρές με συναφείς διαφορές τις ολοκληρωμένες διαδικασίες δεύτερης τάξεως. Υπενθυμίζεται ότι μια ολοκληρωμένη διαδικασία πρώτης τάξης έχει αυτοσυσχετίσεις που σβήνουν με αργό ρυθμό και σε πρώτες διαφορές γίνεται στάσιμη και μάλιστα αρκετές φορές λευκός θόρυβος που σημαίνει ότι οι αυτοσυσχετίσεις των πρώτων διαφορών της θα είναι μηδέν. Επίσης, μια ολοκληρωμένη διαδικασία δεύτερης τάξης σε πρώτες διαφορές παραμένει μη στάσιμη που σημαίνει ότι οι αυτοσυσχετίσεις της σειράς των πρώτων διαφορών θα είναι μη μηδενικές. Ο Yule είχε

καταλάβει ότι το φαινόμενο των υποφαινομενικών σχέσεων εμφανίζεται με χρονοσειρές που έχουν τα χαρακτηριστικά των ολοκληρωμένων διαδικασιών, αλλά φυσικά ο όρος ολοκληρωμένη διαδικασία το 1926 ήταν ακόμη άγνωστος. Για το λόγο αυτό, επινόησε τους όρους των συναφών και τυχαίων σειρών.

Μια άλλη διαπίστωση του Yule ήταν το σχήμα που έχει η κατανομή των συσχετίσεων για δύο ολοκληρωμένες διαδικασίες. Οι Banerjee, Dolado, Galbraith και Hendry (1993) με τη βοήθεια προσομοίωσης μελέτησαν την κατανομή των συσχετίσεων μεταξύ δύο ανεξάρτητων ολοκληρωμένων διαδικασιών για διαφορετικές τάξεις ολοκλήρωσης. Το αποτέλεσμα στο οποίο κατέληξαν είναι ότι όταν οι δύο χρονοσειρές είναι $I(0)$ και συγκεκριμένα λευκός θόρυβος, η κατανομή των συσχετίσεών τους είναι συμμετρική και μοιάζει με κανονική, όταν οι δύο χρονοσειρές είναι $I(1)$ η κατανομή των συσχετίσεων μοιάζει με ημιέλλειψη με τις περισσότερες συσχετίσεις να είναι αρκετά μεγαλύτερες από το μηδέν, ενώ όταν οι δύο χρονοσειρές είναι $I(2)$ η κατανομή των συσχετίσεων έχει σχήμα U και οι τιμές -1 και $+1$ είναι οι πιο πιθανές να εμφανιστούν. Τα αποτελέσματα αυτά επιβεβαιώνουν πλήρως τον Yule για την κατανομή των συσχετίσεων μεταξύ δύο τυχαίων σειρών, δύο συναφών σειρών με τυχαίες διαφορές και δύο συναφών σειρών με συναφείς διαφορές. Επιπλέον, οι Banerjee, Dolado, Galbraith και Hendry (1993) μελέτησαν και την κατανομή των συσχετίσεων για μια $I(1)$ διαδικασία με μια $I(2)$ και έδειξαν πως και σε αυτή την περίπτωση η κατανομή των συσχετίσεων έχει σχήμα U αλλά με πιο μικρό άνοιγμα.

4.4 Η κατανομή των συσχετίσεων μεταξύ δύο χρονοσειρών και η εμφάνιση υποφαινομενικών συσχετίσεων

Στην ενότητα αυτή θα εξεταστεί αναλυτικότερα το φαινόμενο της ύπαρξης υποφαινομενικών συσχετίσεων μεταξύ δύο ανεξάρτητων χρονοσειρών Y_t και X_t , καθώς και η μορφή που παίρνει η κατανομή των συσχετίσεών τους σε κάθε περίπτωση. Οι μηχανισμοί παραγωγής των δύο χρονοσειρών που θα μελετηθούν είναι:

Οι ολοκληρωμένες διαδικασίες πρώτης τάξης $I(1)$:

$$Y_t = Y_{t-1} + u_{Yt}$$

και

$$X_t = X_{t-1} + u_{Xt}$$

Οι ολοκληρωμένες διαδικασίες δεύτερης τάξης $I(2)$:

και

$$Y_t = 2Y_{t-1} - Y_{t-2} + u_{Yt}$$

$$X_t = 2X_{t-1} - X_{t-2} + u_{Xt}$$

Οι αυτοπαλίνδρομες διαδικασίες πρώτης τάξης AR(1):

και

$$Y_t = f_Y Y_{t-1} + u_{Yt}$$

$$X_t = f_X X_{t-1} + u_{Xt}$$

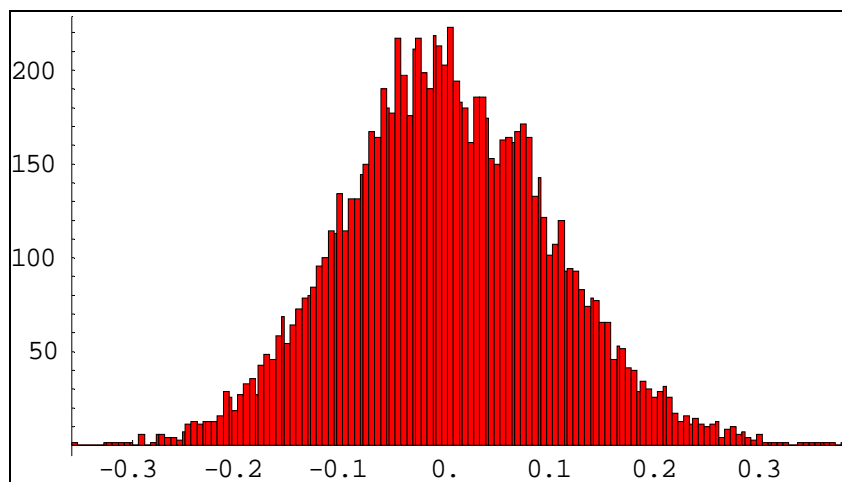
όπου τα f_Y και f_X παίρνουν τις τιμές 0, 0.2, 0.5, 0.8, 0.9 και 0.95. Σε όλες τις περιπτώσεις τα u_{Yt} και u_{Xt} θεωρούνται ότι είναι $iid(0, \sigma_Y^2)$ και $iid(0, \sigma_X^2)$ αντίστοιχα.

Πίνακας 4-1
Υποφαινομενικές συσχετίσεις

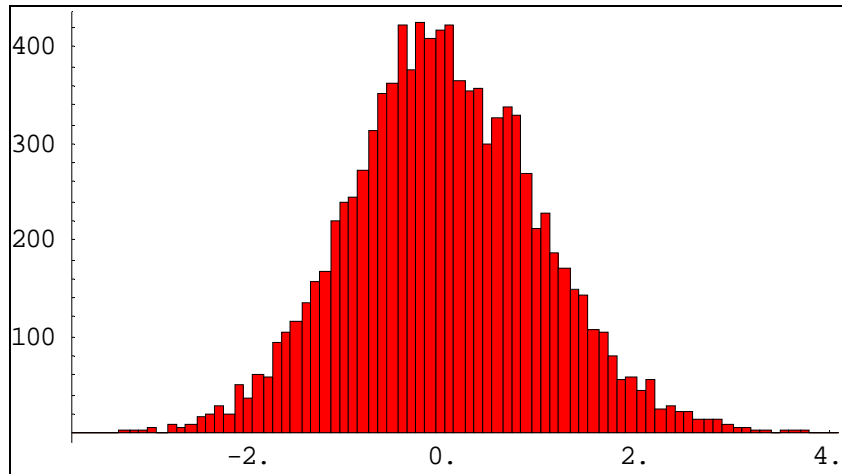
1%	$f_X \backslash f_Y$	0	0.2	0.5	0.8	0.9	0.95	I(1)	I(2)
	0	1.08							
	0.2	0.85	1.36						
	0.5	0.94	1.85	4.36					
	0.8	1.03	3.01	8.86	20.95				
	0.9	1.28	2.9	10.56	28.13	37.31			
	0.95	1.14	3.06	11.4	31.08	42.72	49.63		
	I(1)	1.19	3.28	11.85	34.57	48.02	57.52	68.47	
	I(2)	1.23	3.69	13.41	39.34	54.96	67.35	80.35	93.61
5%	$f_X \backslash f_Y$	0	0.2	0.5	0.8	0.9	0.95	I(1)	I(2)
	0	5.04							
	0.2	4.89	5.92						
	0.5	4.93	7.24	12.68					
	0.8	4.78	9.75	19.29	33.58				
	0.9	5.41	9.56	21.84	41.37	49.85			
	0.95	5.31	9.64	22.43	44.15	55.12	61.02		
	I(1)	4.7	10.14	24.01	47.54	59.21	67.37	75.99	
	I(2)	5.29	10.85	25.11	51.78	65.15	74.99	85.19	95.10
10%	$f_X \backslash f_Y$	0	0.2	0.5	0.8	0.9	0.95	I(1)	I(2)
	0	9.67							
	0.2	10.09	11.53						
	0.5	9.8	13.1	19.94					
	0.8	9.93	16.14	27.68	42.14				
	0.9	10.32	16.41	30.62	49.6	57.04			
	0.95	10.55	16.66	30.98	51.94	61.82	66.85		
	I(1)	9.45	16.85	32.57	54.56	65.43	72.51	79.77	
	I(2)	10.09	17.62	33.89	58.66	70.89	78.85	87.37	96.02

Ο Πίνακας 4-1 συνοψίζει τα ποσοστά απόρριψης του ελέγχου της στατιστικής σημαντικότητας του συντελεστή συσχέτισης (4.1) με τη στατιστική (4.2) για τις ανεξάρτητες χρονοσειρές Y_t και X_t που καταγράφηκαν από μια σειρά προσομοιώσεων για όλους τους δυνατούς συνδυασμούς των μηχανισμών παραγωγής που αναφέρθηκαν. Τα επίπεδα σημαντικότητας που θεωρήθηκαν είναι 1%, 5% και 10%, ενώ η κάθε χρονοσειρά αποτελείται από 100 παρατηρήσεις με τους όρους u_{Y_t} και u_{X_t} να προέρχονται από ανεξάρτητες $N(0, 1)$ κατανομές.

Όταν τα f_Y και f_X ισούνται με το μηδέν, δηλαδή όταν οι δύο χρονοσειρές είναι λευκός θόρυβος, τα ποσοστά απόρριψης της μηδενικής υπόθεσης είναι κοντά στα επίπεδα σημαντικότητας, δηλαδή 1.08%, 5.04% και 9.67% για τα αντίστοιχα επίπεδα σημαντικότητας 1%, 5% και 10%, όπως προβλέπεται από τη στατιστική θεωρία. Το σχήμα της κατανομής των συσχετίσεων απεικονίζεται στο Διάγραμμα 4-2 από όπου φαίνεται πως η κατανομή είναι συμμετρική. Η μικρότερη και η μεγαλύτερη συσχέτιση που κατεγράφησαν είναι -0.358 και $+0.379$ αντίστοιχα, ενώ η μέση συσχέτιση είναι 0 με τυπική απόκλιση 0.0998. Προέκυψαν 5162 αρνητικές και 4838 θετικές συσχετίσεις. Επίσης, από το Διάγραμμα 4-3 που απεικονίζει την κατανομή της στατιστικής (4.2) που χρησιμοποιήθηκε για τον έλεγχο της στατιστικής σημαντικότητας του συντελεστή συσχέτισης συμπεραίνεται ότι η κατανομή είναι κανονική. Η μέση τιμή της υπολογίστηκε σε 0 και η τυπική της απόκλιση σε 1.004. Επομένως, όταν οι δύο χρονοσειρές είναι λευκός θόρυβος η στατιστική (4.2) ακολουθεί ασυμπτωτικά την τυπική κανονική κατανομή κάτι που είναι σύμφωνο με τη θεωρία.



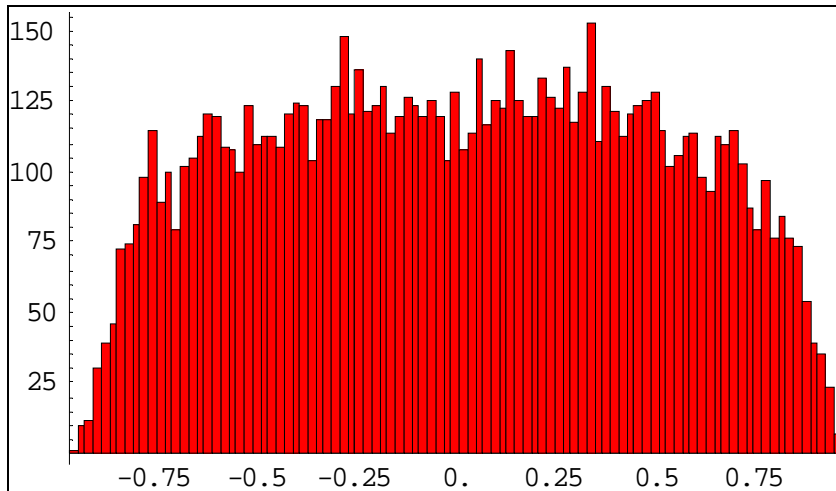
Διάγραμμα 4-2
Κατανομή των συσχετίσεων όταν $f_Y = 0$ και $f_X = 0$, $T = 100$



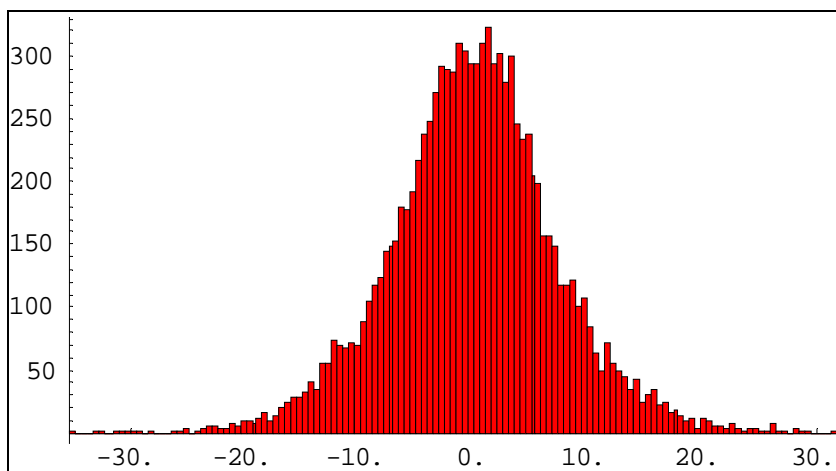
Διάγραμμα 4-3

Κατανομή της στατιστικής ελέγχου της συσχέτισης όταν $f_Y = 0$ και $f_X = 0$, $T = 100$

Όταν οι δύο χρονοσειρές είναι ολοκληρωμένες διαδικασίες πρώτης τάξης το φαινόμενο της υποφαινομενικής συσχέτισης κάνει την εμφάνισή του σε πολύ μεγάλο ποσοστό. Τα ποσοστά όπου ο συντελεστής συσχέτισης είναι εσφαλμένα σημαντικός είναι 68.47 % για επίπεδο σημαντικότητας 1%, 75.99% για επίπεδο σημαντικότητας 5% και 79.77% για επίπεδο σημαντικότητας 10%. Υπενθυμίζεται από την ενότητα 2.2 πως το ποσοστό των υποφαινομενικών παλινδρομήσεων στο επίπεδο σημαντικότητας 5% για τις αντίστοιχες χρονοσειρές ήταν 76.2% που είναι πολύ κοντά στο 75.99% που προέκυψε για το συντελεστή συσχέτισης. Αυτό είναι λογικό αφού, όπως έχει δείχθει, υπάρχει στενή σχέση ανάμεσα στο συντελεστή συσχέτισης και στην παλινδρόμηση. Όπως φαίνεται από το Διάγραμμα 4-4, η κατανομή των συσχετίσεων παραμένει συμμετρική αλλά έχει σχήμα ημιέλλειψης με τις περισσότερες συσχετίσεις να είναι πολύ μεγαλύτερες από το μηδέν. Πιο συγκεκριμένα προέκυψαν 4977 αρνητικές και 5023 θετικές συσχετίσεις, με τη μικρότερη συσχέτιση να είναι -0.962 και την μεγαλύτερη $+0.954$, έχοντας μέση τιμή 0 με τυπική απόκλιση 0.487. Η κατανομή της στατιστικής (4.1) για αυτή την περίπτωση απεικονίζεται στο Διάγραμμα 4-5 από όπου φαίνεται πως παραμένει συμμετρική. Η μέση τιμή της είναι 0 αλλά έχει αυξηθεί πάρα πολύ η τυπική της απόκλιση η οποία υπολογίζεται σε 7.3. Δηλαδή, όταν οι χρονοσειρές είναι ολοκληρωμένες διαδικασίες η στατιστική (4.2) δεν ακολουθεί ασυμπτωτικά την τυπική κανονική κατανομή κάτι που δικαιολογεί και την εμφάνιση του φαινομένου της υποφαινομενικής συσχέτισης.

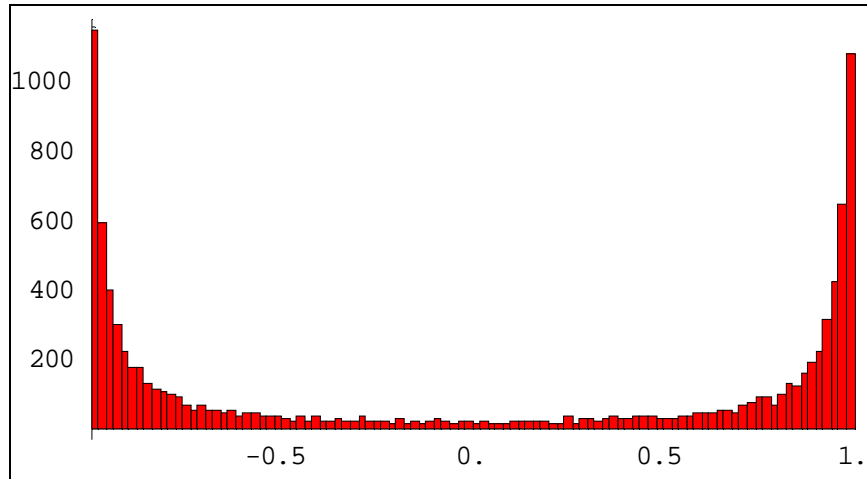


Διάγραμμα 4-4
Κατανομή των συσχετίσεων όταν $Y_t \sim I(1)$ και $X_t \sim I(1)$, $T = 100$

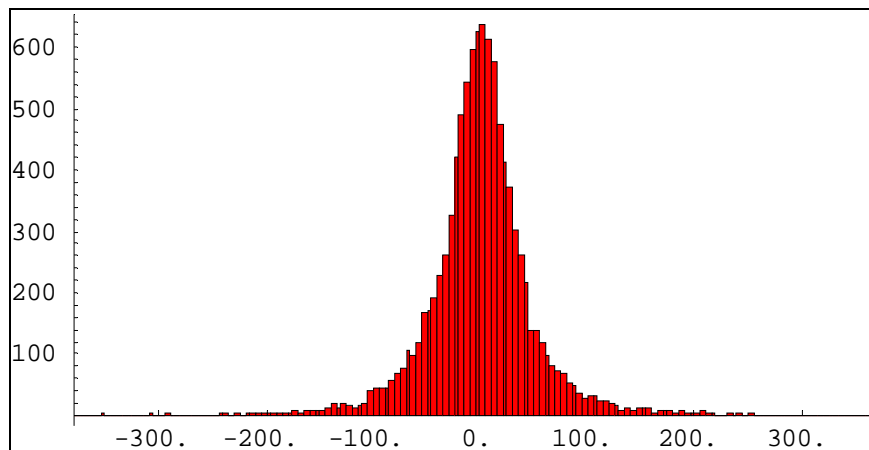


Διάγραμμα 4-5
Κατανομή της στατιστικής ελέγχου της συσχέτισης όταν $Y_t \sim I(1)$ και $X_t \sim I(1)$, $T = 100$

Στην περίπτωση που οι χρονοσειρές είναι ολοκληρωμένες διαδικασίες δεύτερης τάξης είναι σχεδόν σίγουρο ότι θα εμφανιστούν υποφαινομενικές συσχετίσεις. Τα ποσοστά όπου ο συντελεστής συσχέτισης είναι σημαντικός αυξάνονται και ανέρχονται σε 93.61%, 95.1% και 96.02 για τα αντίστοιχα επίπεδα σημαντικότητας 1%, 5% και 10%. Όπως διαπιστώνεται από



Διάγραμμα 4-6
Κατανομή των συσχετίσεων όταν $Y_t \square I(2)$ και $X_t \square I(2)$, $T = 100$

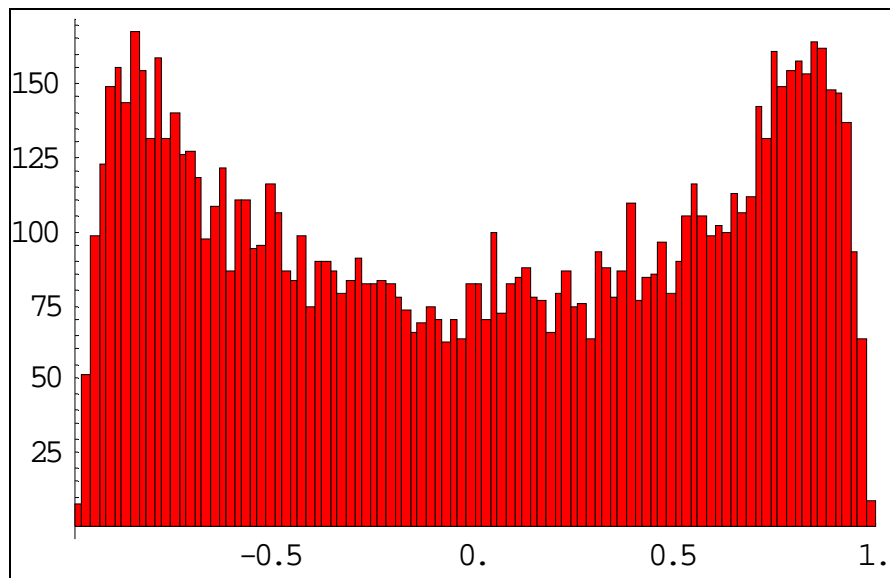


Διάγραμμα 4-7
**Κατανομή της στατιστικής ελέγχου της συσχέτισης όταν $Y_t \square I(2)$ και $X_t \square I(2)$,
 $T = 100$**

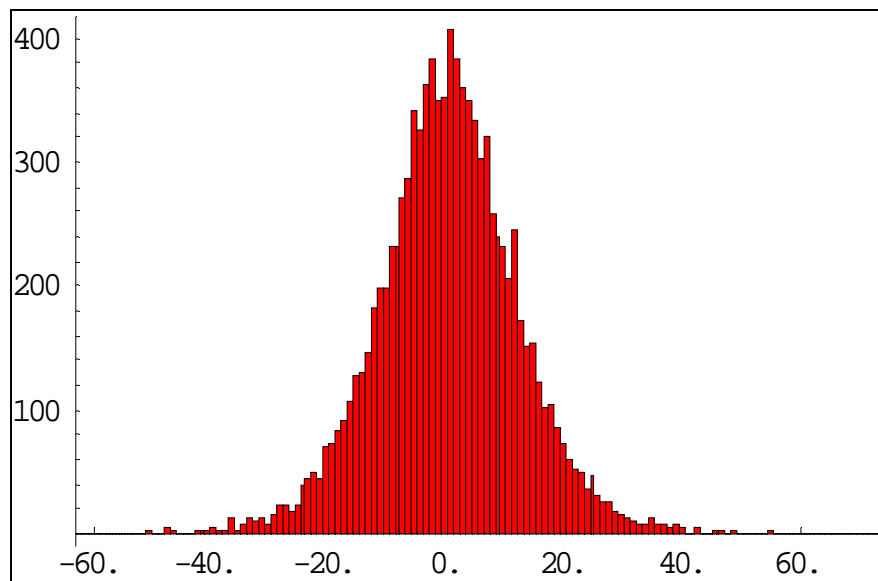
το Διάγραμμα 4-6, το σχήμα της κατανομής των συσχετίσεων, παρόλο που είναι συμμετρικό, αλλάζει και παίρνει την μορφή U με τις τιμές κοντά στο -1 και το +1 να είναι οι επικρατέστερες να εμφανιστούν. Υπάρχουν 4958 αρνητικές και 5042 θετικές συσχετίσεις με μέση τιμή 0 και τυπική απόκλιση 0.84. Το γεγονός ότι η μέση τιμή είναι 0 οφείλεται καθαρά στη συμμετρία της κατανομής. Αυτό αποδεικνύεται από τη μέση απόλυτη τιμή που

υπολογίστηκε σε 0.798 καθώς και με την ελάχιστη και μέγιστη συσχέτιση που καταγράφηκαν αντίστοιχα σε -0.999 και $+0.999$. Η κατανομή της στατιστική (4.2), όπως φαίνεται στο Διάγραμμα 4-7, παραμένει συμμετρική θυμίζοντας κανονική αλλά η μέση τιμή είναι -0.11 και η τυπική της απόκλιση αυξάνεται υπερβολικά σε 51.18.

Όταν η μία χρονοσειρά είναι ολοκληρωμένη διαδικασία πρώτης τάξης και η δεύτερη ολοκληρωμένη διαδικασία δεύτερης τάξης ο αριθμός των υποφαινομενικών συσχετίσεων είναι εξίσου υψηλός. Ο συντελεστής συσχέτισης είναι σημαντικός σε ποσοστά 80.35%, 85.19% και 87.37% αντίστοιχα για τα τρία επίπεδα σημαντικότητας 1%, 5% και 10%. Όπως φαίνεται στο Διάγραμμα 4-8, η κατανομή των συσχετίσεων σε αυτή την περίπτωση έχει σχήμα μια σύνθεση ανάμεσα στην ημιέλλειψη που είχε προκύψει όταν και οι δύο χρονοσειρές ήταν $I(1)$ και στο U που είχε προκύψει για δύο $I(2)$ διαδικασίες. Δηλαδή, η κατανομή ξεκινάει από χαμηλά ως ημιέλλειψη και όσο ανέρχεται σχηματίζει U . Καταγράφηκαν 4927 αρνητικές και 5073 θετικές συσχετίσεις με μέση τιμή 0 και τυπική απόκλιση 0.617. Η ελάχιστη συσχέτιση είναι -0.988 και η μέγιστη $+0.991$. Στο Διάγραμμα 4-9 απεικονίζεται η κατανομή της στατιστική (4.2). Συμπεραίνεται πώς και σε αυτή την περίπτωση η κατανομή μοιάζει με κανονική αλλά έχει μέση τιμή 0.154 και μεγάλη τυπική απόκλιση 12.11.



Διάγραμμα 4-8
Κατανομή των συσχετίσεων όταν $Y_t \sim I(2)$ και $X_t \sim I(1)$, $T = 100$

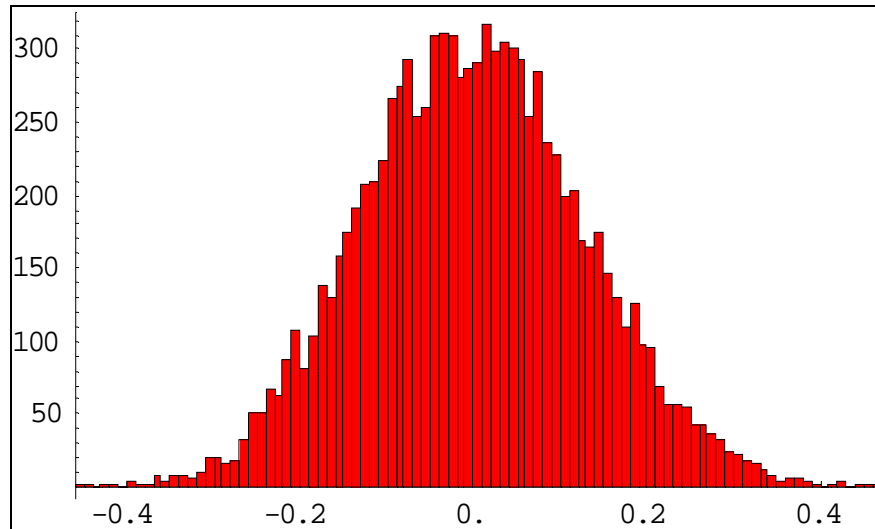


Διάγραμμα 4-9
Κατανομή της στατιστικής ελέγχου της συσχέτισης όταν $Y_t \sim I(2)$ και $X_t \sim I(1)$,
 $T = 100$

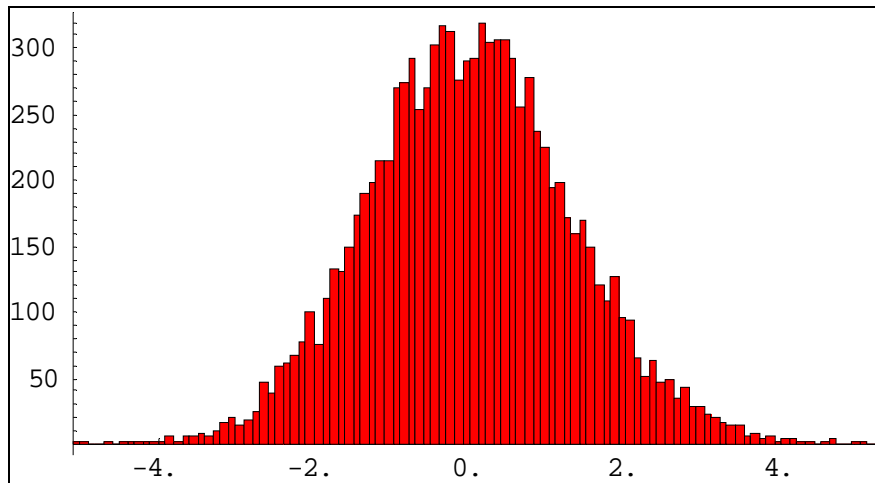
Τα αποτελέσματα των προσομοιώσεων επιβεβαιώνουν πλήρως τα συμπεράσματα των Yule (1926) και Banerjee, Dolado, Galbraith και Hendry (1993). Όμως, από τις προσομοιώσεις προέκυψε και ένα άλλο σημαντικό συμπέρασμα. Οι υποφαινομενικές συσχετίσεις κάνουν την εμφάνισή τους όχι μόνο όταν οι δύο χρονοσειρές είναι ολοκληρωμένες διαδικασίες αλλά και σε όλες τις άλλες περιπτώσεις όπου τα f_Y και f_X έχουν τιμές μεγαλύτερες από το μηδέν. Μόνο όταν η μία από τις δύο χρονοσειρές είναι λευκός θόρυβος, δηλαδή όταν $f_Y \neq 0$ και $f_X = 0$, το φαινόμενο δεν εμφανίζεται και τα αποτελέσματα είναι κοντά σε αυτά που προβλέπονται από τη στατιστική θεωρία. Αντίθετα, καθώς οι τιμές των f_Y και f_X αυξάνονται προσεγγίζοντας τη μονάδα, το ποσοστό των υποφαινομενικών συσχετίσεων αυξάνεται με γρήγορο ρυθμό. Ακόμη και για μέτριες τιμές των f_Y και f_X το ποσοστό απόρριψης του ελέγχου (4.2) είναι αξιοσημείωτο.

Μια επιπρόσθετη διαπίστωση είναι πως ανεξάρτητα από τις τιμές των f_Y και f_X τόσο η κατανομή των συσχετίσεων όσο και η κατανομή της στατιστικής (4.2) παραμένουν συμμετρικές με μέση τιμή το 0. Αυτό που αλλάζει είναι οι τυπικές αποκλίσεις που

αυξάνονται καθώς αυξάνονται τα f_Y και f_X . Δηλαδή, καθώς αυξάνονται τα f_Y και f_X η κατανομή της στατιστικής (4.2) ανοίγει και αποκλίνει από $N(0,1)$ όπως προβλέπει η ασυμπτωτική θεωρία κάτι που δικαιολογεί και την αύξηση του ποσοστού των υποφαινομενικών συσχετίσεων.

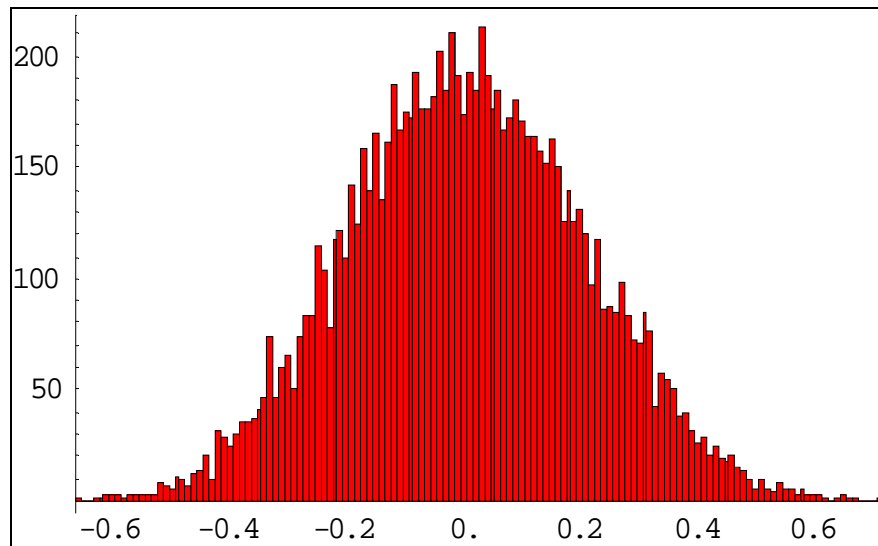


Διάγραμμα 4-10
Κατανομή των συσχετίσεων όταν $f_Y = 0.5$ και $f_X = 0.5$, $T = 100$

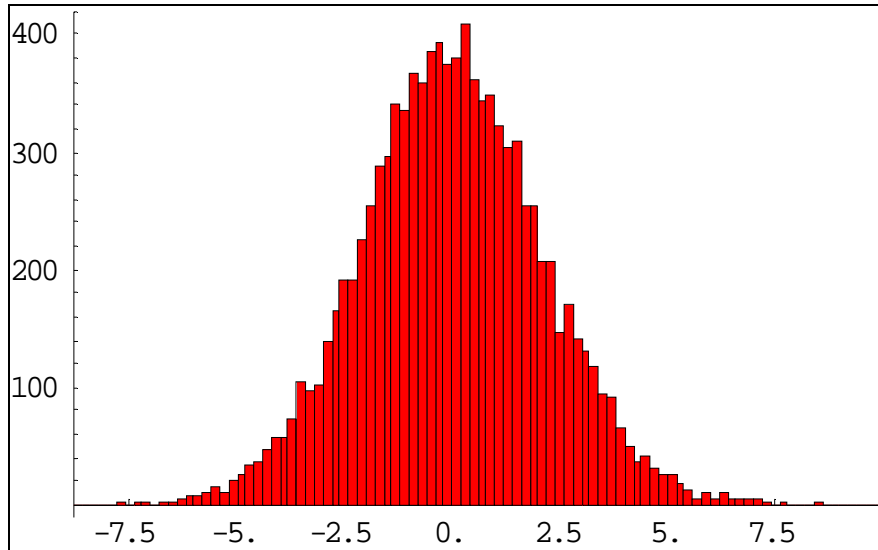


Διάγραμμα 4-11
Κατανομή της στατιστικής ελέγχου της συσχέτισης όταν $f_Y = 0.5$ και $f_X = 0.5$,
 $T = 100$

Για παράδειγμα, όταν $f_y = 0.5$ και $f_x = 0.5$, τα ποσοστά των υποφαινομενικών συσχετίσεων είναι 4.36%, 12.68% και 19.94% για τα αντίστοιχα επίπεδα σημαντικότητας 1%, 5% και 10%. Η ελάχιστη συσχέτιση που παρατηρήθηκε είναι -0.4544 και η μέγιστη $+0.4675$. Από τα Διαγράμματα 4-10 και 4-11 που απεικονίζουν την κατανομή των συσχετίσεων και την κατανομή της στατιστικής (4.2) διαπιστώνεται η συμμετρία αλλά και η αύξηση των τυπικών αποκλίσεων. Η τυπική απόκλιση για την κατανομή των συσχετίσεων υπολογίστηκε σε 0.128 και για την κατανομή της στατιστικής (4.2) σε 1.299 που απέχει από το 1. Επιπλέον, όταν $f_y = 0.8$ και $f_x = 0.8$, τα ποσοστά των υποφαινομενικών συσχετίσεων για τα επίπεδα σημαντικότητας 1%, 5% και 10% είναι 20.95%, 33.58% και 42.14%. Η ελάχιστη συσχέτιση που παρατηρήθηκε είναι -0.6613 και η μέγιστη $+0.7098$. Όπως φαίνεται στα Διαγράμματα 4-12 και 4-13 η συμμετρία στις κατανομές παραμένει. Όμως, αυξάνονται ακόμη πιο πολύ οι τυπικές αποκλίσεις οι οποίες τώρα εκτιμώνται σε 0.20 για την κατανομή των συσχετίσεων και σε 2.124 για την κατανομή της στατιστικής του ελέγχου. Ανάλογα αποτελέσματα παρατηρούνται για οποιονδήποτε συνδυασμό των $f_y \neq 0$ και $f_x = 0$.



Διάγραμμα 4-12
Κατανομή των συσχετίσεων όταν $f_y = 0.8$ και $f_x = 0.8$, $T = 100$

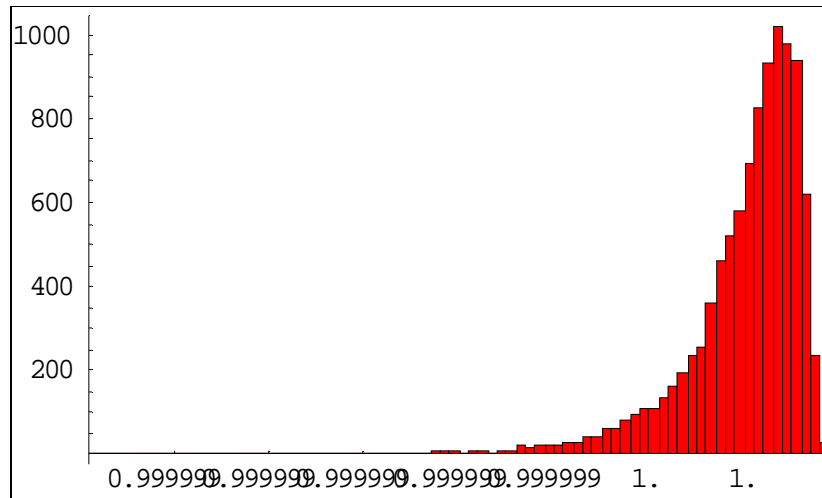


Διάγραμμα 4-13
Κατανομή της στατιστικής ελέγχου της συσχέτισης όταν $f_Y = 0.8$ και $f_X = 0.8$,
 $T = 100$

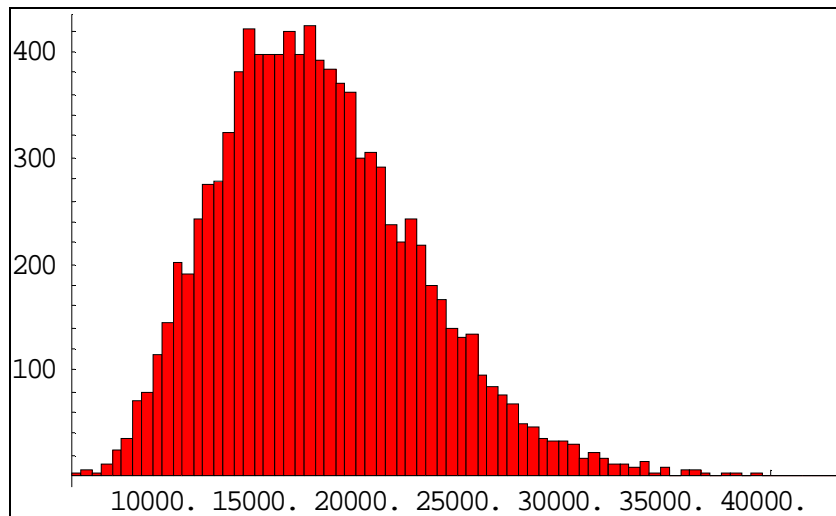
Μια πρόσθετη περίπτωση που αξίζει να ερευνηθεί είναι αυτή όπου οι δύο χρονοσειρές είναι ολοκληρωμένες διαδικασίες με μη μηδενικό μέσο όρο (*drift*). Η περίπτωση αυτή έχει ιδιαίτερο ενδιαφέρον γιατί στην ενότητα 2.3 είχαν δειχτεί τα ιδιαίτερα αποτελέσματα που προκύπτουν στην παλινδρόμηση όταν χρησιμοποιούνται στο απλό γραμμικό υπόδειγμα ολοκληρωμένες διαδικασίες με μη μηδενικό μέσο όρο. Για το λόγο αυτό, πραγματοποιήθηκαν δυο πρόσθετες προσομοιώσεις.

Στην πρώτη προσομοίωση οι χρονοσειρές παρήχθησαν ως ανεξάρτητοι τυχαίοι περίπατοι που όμως έχουν μέσο όρο την τιμή 200. Το βασικότερο συμπέρασμα που προέκυψε από την προσομοίωση είναι ότι το ποσοστό που ο συντελεστής συσχέτισης είναι σημαντικός είναι 100% ανεξάρτητα από το επίπεδο σημαντικότητας. Επίσης, όπως φαίνεται στο Διάγραμμα 4-14, η κατανομή των συσχετίσεων παύει να είναι συμμετρική και έχει σχήμα σαν χωρισμένο στη μέση U με όλες τις συσχετίσεις θετικές και συγκεντρωμένες κοντά στο 1. Πραγματικά η μέση τιμή της συσχέτισης είναι 1, ενώ η ελάχιστη και μέγιστη τιμή που παρατηρήθηκαν 0.999998 και 1. Ανάλογη συμπεριφορά διαπιστώθηκε και με την κατανομή της στατιστικής (4.2) η οποία, όπως φαίνεται στο Διάγραμμα 4-15, έχει μόνο θετικές τιμές,

μέση τιμή 17706 και τεράστια τυπική απόκλιση 4941.6. Επομένως, το ποσοστό απόρριψης 100% που διαπιστώθηκε είναι άμεσο επακόλουθο.

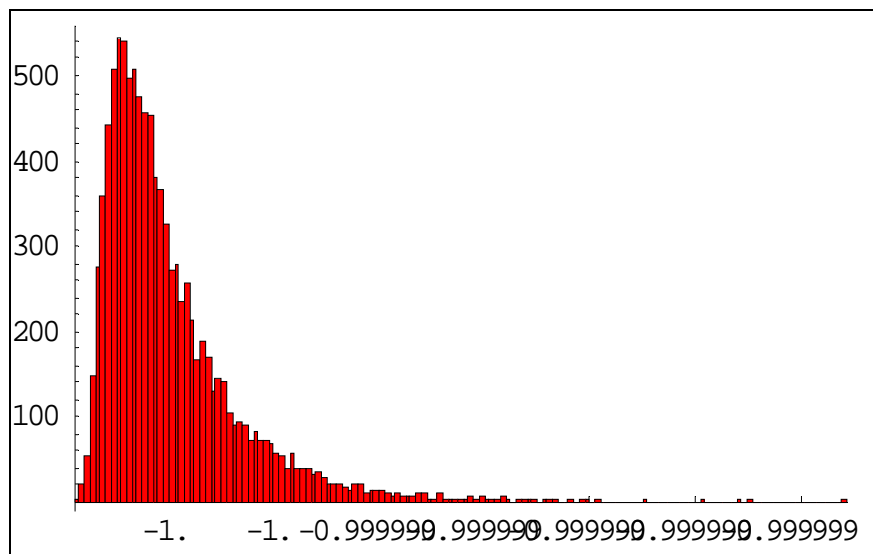


Διάγραμμα 4-14
Κατανομή των συσχετίσεων όταν $Y_t \sim I(1)$ και $X_t \sim I(1)$ με μέσους όρους 200,
 $T = 100$

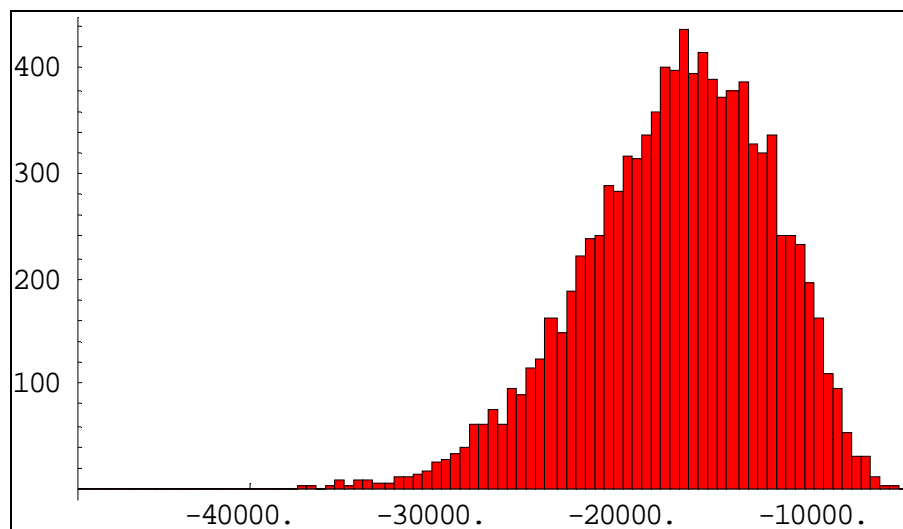


Διάγραμμα 4-15
Κατανομή της στατιστικής ελέγχου της συσχέτισης όταν $Y_t \sim I(1)$ και $X_t \sim I(1)$ με
μέσους όρους 200, $T = 100$

Στη δεύτερη προσομοίωση οι χρονοσειρές παρήχθησαν ως ανεξάρτητοι τυχαίοι περίπατοι με μέσους όρους 200 και -200 . Και σε αυτή την περίπτωση το ποσοστό όπου ο συντελεστής συσχέτισης είναι σημαντικός είναι 100% ανεξάρτητα από το επίπεδο σημαντικότητα. Τα υπόλοιπα αποτελέσματα είναι τα αντίθετα από την περίπτωση όπου και οι δύο χρονοσειρές είχαν μέσο όρο 200. Δηλαδή, όπως φαίνεται στο Διάγραμμα 4-16, η κατανομή των συσχετίσεων δεν είναι πάλι συμμετρική και έχει σχήμα χωρισμένου U από την ανάποδη αυτή τη φορά πλευρά, με όλες τις συσχετίσεις αρνητικές και συγκεντρωμένες κοντά στο -1 . Η μέση τιμή της συσχέτισης είναι -1 ενώ η ελάχιστη και μέγιστη τιμή που παρατηρήθηκαν είναι -1 και -0.99999 . Όμοια, η κατανομή της στατιστικής (4.2), όπως φαίνεται στο Διάγραμμα 4-17, έχει μόνο αρνητικές τιμές με μέση τιμή -17712 και τυπική απόκλιση 4899.2 .



Διάγραμμα 4-16
Κατανομή των συσχετίσεων όταν $Y_t \in I(1)$ και $X_t \in I(1)$
με μέσους όρους 200 και -200 , $T = 100$



Διάγραμμα 4-17

Κατανομή της στατιστικής ελέγχου της συσχέτισης όταν $Y_t \sim I(1)$ και $X_t \sim I(1)$ με μέσους όρους 200 και -200, $T = 100$

Τα αποτελέσματα των δύο τελευταίων προσομοιώσεων δεν πρέπει να προκαλούν εντύπωση αν ερμηνευτούν σε συνδυασμό με τα συμπεράσματα της ανάλογης περίπτωσης για την απλή γραμμική παλινδρόμηση της ενότητας 2.3. Υπενθυμίζεται πως στην απλή γραμμική παλινδρόμηση το ποσοστό απόρριψης του ελέγχου της στατιστικής σημαντικότητας της κλίσης του υποδείγματος ήταν 100%. Επίσης, ο εκτιμητής της κλίσης, το πρόσημο του οποίου καθορίζει εάν η σχέση μεταξύ των δύο χρονοσειρών είναι θετική η αρνητική, σύγκλιε στο λόγο των δύο μέσων όρων. Κατά συνέπεια, είναι λογικό όταν οι δύο μέσοι όροι έχουν θετικό πρόσημο όλες οι συσχετίσεις να είναι θετικές και όταν ο ένας μέσος όρος είναι αρνητικός όλες οι συσχετίσεις να είναι αρνητικές. Επιπρόσθετα, λογικό είναι και το γεγονός ότι ουσιαστικά όλες οι συσχετίσεις παίρνουν την τιμή 1 ή -1, ανάλογα με το εάν η σχέση μεταξύ των δύο χρονοσειρών είναι θετική η αρνητική, αφού ο συντελεστής συσχέτισης είναι η τετραγωνική ρίζα του συντελεστή προσδιορισμού ο οποίος στην περίπτωση της παλινδρόμησης συγκλίνει στη μονάδα.

4.5 Ανακεφαλαίωση

Όπως και στην περίπτωση της υποφαινομενικής παλινδρόμησης, η εμφάνιση υποφαινομενικών συσχετίσεων είναι συνηθισμένο φαινόμενο όταν υπολογίζεται ο συντελεστής συσχέτισης για δύο χρονοσειρές. Η αιτία του φαινομένου είναι πάλι οι ιδιότητες αυτοσυσχέτισης των χρονοσειρών, ανεξάρτητα από το εάν αυτές είναι ή δεν είναι ολοκληρωμένες διαδικασίες, οι οποίες επηρεάζουν τον έλεγχο για την στατιστική σημαντικότητα του συντελεστή συσχέτισης. Πιο συγκεκριμένα, δείχθηκε ότι καθώς οι χρονοσειρές πλησιάζουν στα όρια της μη στασιμότητας, η τυπική απόκλιση της κατανομής της στατιστική που χρησιμοποιείται για τον έλεγχο της σημαντικότητας του συντελεστή συσχέτισης, αυξάνεται συνεχώς. Δηλαδή, ασυμπτωτικά η κατανομή της στατιστικής δεν είναι $N(0, 1)$ όπως προβλέπεται από τη θεωρία με αποτέλεσμα οι κριτικές τιμές να είναι εσφαλμένες.

Η ταυτοποίηση και η εξέταση των ιδιοτήτων αυτοσυσχέτισης κάθε χρονοσειράς προτού χρησιμοποιηθεί στον υπολογισμό του συντελεστή συσχέτισης κρίνεται περισσότερο επιτακτική από ότι στην παλινδρόμηση. Ο λόγος είναι ότι στην παλινδρόμηση εφαρμόζεται και ο έλεγχος των Durbin και Watson που μπορεί να προφυλάξει από την αποδοχή μιας υποφαινομενικής σχέσης, ενώ στην περίπτωση του συντελεστή συσχέτισης δεν υπάρχει κάποιο πρόσθετο κριτήριο διαθέσιμο.

Η ΧΡΗΣΗ ΔΙΑΦΟΡΩΝ ΣΤΗΝ ΕΚΤΙΜΗΣΗ ΓΡΑΜΜΙΚΩΝ ΥΠΟΔΕΙΓΜΑΤΩΝ

5.1 Εισαγωγή

Ένα δίλημμα που τίθεται συχνά όταν εκτιμώνται υποδείγματα παλινδρόμησης με χρονοσειρές αφορά την απόφαση για το εάν η εκτίμηση θα πρέπει να γίνει στα επίπεδα (*levels*) ή στις πρώτες διαφορές (*first differences*) των τιμών των χρονοσειρών. Η διαφωνία των ερευνητών για τη μέθοδο που πρέπει να ακολουθηθεί επηρεάζεται από το γεγονός πως πολλές φορές μια ισχυρή σχέση, όπως συμπεραίνεται από την τιμή του συντελεστή προσδιορισμού και τους στατιστικούς ελέγχους, ενός υποδείματος που έχει εκτιμηθεί στα επίπεδα των χρονοσειρών παύει όταν η εκτίμηση γίνει στις διαφορές.

Όπως αναφέρθηκε στην ενότητα 2.2 οι Granger και Newbold (1974) είχαν διαπιστώσει πραγματοποιώντας προσομοίωση πως η εκτίμηση ενός υποδείματος παλινδρόμησης στα επίπεδα ανεξαρτήτων ολοκληρωμένων διαδικασιών υπόκειται σε σοβαρό κίνδυνο να εμφανίσει το φαινόμενο της υποφαινομενικής παλινδρόμησης. Κάτι τέτοιο δεν ισχύει όταν η εκτίμηση γίνει στις πρώτες διαφορές των χρονοσειρών. Αυτό συμβαίνει γιατί στα επίπεδα οι χρονοσειρές είναι μη στάσιμες διαδικασίες που προκαλεί και την έλλειψη στασιμότητας στα σφάλματα του υποδείματος. Όταν η εκτίμηση γίνει στις διαφορές, οι χρονοσειρές γίνονται λευκός θόρυβος και κατά συνέπεια τα σφάλματα είναι στάσιμα. Το συμπέρασμα είναι πως όταν οι χρονοσειρές είναι μη στάσιμες θα πρέπει να μετασχηματίζονται σε στάσιμες με τη χρήση διαφορών προτού χρησιμοποιηθούν στην εκτίμηση ενός υποδείματος παλινδρόμησης. Επομένως, είναι φανερό ότι πολλές από τις περιπτώσεις που συναντιούνται στην πράξη όπου η χρήση των διαφορών επηρεάζει το αποτέλεσμα της εκτίμησης ανήκουν απλά στην κατηγορία της υποφαινομενικής παλινδρόμησης.

Παρόλα αυτά πολλοί ερευνητές εξακολουθούν να μάχονται για τη χρήση διαφορών στην εκτίμηση του υποδείματος γιατί πιστεύουν πως καταστρέφει χρήσιμες πληροφορίες. Όμως, όπως θα δειχθεί στην συνέχεια, όταν το υπόδειγμα είναι ταυτοποιημένο σωστά, είναι αδιάφορο εάν η εκτίμηση γίνει στα επίπεδα ή στις διαφορές των χρονοσειρών. Στο κεφάλαιο

αυτό θα εξεταστούν κριτήρια που μπορούν να χρησιμοποιηθούν για την επιλογή ανάμεσα στην εκτίμηση ενός υποδείγματος στα επίπεδα και στις διαφορές των χρονοσειρών, καθώς και θέματα που προκύπτουν στη γραμμική παλινδρόμηση όταν εφαρμόζεται ο μετασχηματισμός των χρονοσειρών με τη χρήση διαφορών.

5.2 Οι συνέπειες των διαφορών στο υπόδειγμα της γραμμικής παλινδρόμησης

Έστω το υπόδειγμα της απλής γραμμικής παλινδρόμησης:

$$Y_t = a + bX_t + e_t \quad (5.1)$$

όπου η $e_t \sim iidN(0, \sigma_e^2)$, ενώ οι Y_t και X_t μπορεί να είναι είτε τα επίπεδα είτε οι διαφορές δύο χρονοσειρών. Οι Plosser και Schwert (1978) εξέτασαν τα αποτελέσματα που επέρχονται στο υπόδειγμα (5.1) όταν αυτό εκτιμηθεί στα επίπεδα των χρονοσειρών, ενώ θα έπρεπε στις διαφορές τους και όταν εκτιμηθεί στις διαφορές, ενώ κανονικά θα έπρεπε στα επίπεδα.

Ας υποθεθεί αρχικά ότι η σχέση μεταξύ των χρονοσειρών είναι ορισμένη στα επίπεδα. Γράφοντας την (5.1) για τη χρονική στιγμή $t-1$ και αφαιρώντας προκύπτει:

$$\Delta Y_t = b\Delta X_t + u_t \quad (5.2)$$

όπου

$$u_t = e_t - e_{t-1} \quad (5.3)$$

Το υπόδειγμα αυτό θα μπορούσε να περιλαμβάνει και σταθερό όρο στην περίπτωση όπου το αρχικό υπόδειγμα περιελάμβανε και τάση. Το σημαντικό συμπέρασμα που προκύπτει από το υπόδειγμα (5.2) είναι ότι η σχέση που υπάρχει μεταξύ των επιπέδων των Y_t και X_t εξακολουθεί να υπάρχει και στις διαφορές τους με τον ίδιο τρόπο. Οι μοναδικές διαφορές που παρατηρούνται είναι η απουσία του σταθερού όρου και η διαφορετική δομή των σφαλμάτων. Το υπόδειγμα (5.2) ονομάζεται υπέρδιαφοροποιημένο (*overdifferenced model*) και έχει χαρακτηριστικό ότι τα σφάλματά του ακολουθούν τη μη αντιστρέψιμη κινητού μέσου όρου διαδικασία πρώτης τάξης (5.3). Ο εκτιμητής ελαχίστων τετραγώνων του b στο υπόδειγμα (5.2) θα είναι αμερόληπτος και συνεπής αλλά δεν θα είναι αποδοτικός. Αυτό οφείλεται στο ότι ο συνηθισμένος τύπος για τον υπολογισμό της διακύμανσης του εκτιμητή \hat{b} :

$$S^2(\hat{b}) = (T-2) \frac{\sum_{t=2}^T \hat{u}_t^2}{\sum_{t=2}^T (\Delta X_t - \overline{\Delta X_t})^2} \quad (5.4)$$

είναι μεροληπτικός όταν τα σφάλματα αυτοσυσχετίζονται. Όταν τα σφάλματα ακολουθούν την MA(1) διαδικασία (5.3) αποδεικνύεται πώς ένας αμερόληπτος εκτιμητής της διακύμανσης του \hat{b} δίνεται από την ακόλουθη σχέση:

$$S^2(\hat{b}) = \frac{\hat{S}_u^2(1-\hat{r}_1)}{\sum_{t=2}^T (\Delta \hat{X}_t)^2}$$

όπου \hat{S}_u^2 είναι ένας αμερόληπτος εκτιμητής της διακύμανσης των σφαλμάτων και \hat{r}_1 είναι μια εκτίμηση για την αυτοσυσχέτιση πρώτης τάξης της χρονοσειράς ΔX_t των διαφορών. Στην περίπτωση που η X_t είναι τυχαίος περίπατος, η ΔX_t δεν αυτοσυσχετίζεται και ο εκτιμητής της διακύμανσης (5.4) είναι αμερόληπτος. Όταν όμως η ΔX_t αυτοσυσχετίζεται θετικά ο εκτιμητής (5.4) υπερεκτιμά τη διακύμανση του \hat{b} ενώ όταν αυτοσυσχετίζεται αρνητικά την υποεκτιμά. Η εκτίμηση του υποδείγματος θα μπορούσε να γίνει χρησιμοποιώντας και τις δεύτερες διαφορές, δηλαδή:

$$\Delta(\Delta Y_t) = b\Delta(\Delta X_t) + \Delta u_t$$

Διαπιστώνεται ότι και σε αυτή την περίπτωση ο εκτιμητής του b με τη μέθοδο των ελαχίστων τετραγώνων θα είναι συνεπής αλλά όχι αποδοτικός.

Έστω τώρα ότι η σχέση μεταξύ των Y_t και X_t ορίζεται σωστά στις διαφορές τους, δηλαδή:

$$\Delta Y_t = y_t$$

και

$$\Delta X_t = x_t$$

Αντί όμως να εκτιμηθεί το υπόδειγμα (5.2) εκτιμάται το υπόδειγμα (5.1). Άμεση συνέπεια είναι ότι τα σφάλματα θα ακολουθούν τον τυχαίο περίπατο:

$$e_t = e_{t-1} + u_t$$

Στην περίπτωση αυτή το υπόδειγμα ονομάζεται υποδιαφοροποιημένο (*underdifferenced model*) και έχει χαρακτηριστικό ότι τα σφάλματά του είναι μη στάσιμα. Το σοβαρό πρόβλημα

που εμφανίζεται είναι ότι ο εκτιμητής ελαχίστων τετραγώνων \hat{b} μπορεί να μην είναι συνεπής. Απαραίτητη συνθήκη για να είναι ο \hat{b} συνεπής είναι:

$$p \lim \left[(T-1) \frac{\sum_{t=0}^T \hat{e}_t^2}{\sum_{t=0}^T (X_t - \bar{X})^2} \right] = 0$$

Μια περίπτωση που θα μπορούσε να ισχύει αυτό είναι όταν η X_t είναι απλά γραμμική τάση.

Το τελικό συμπέρασμα στο οποίο κατέληξαν οι Plosser και Schwert (1978) ενισχύοντάς το και με πραγματικά παραδείγματα είναι πως οι παράμετροι ενός σωστά ταυτοποιημένου υποδείγματος μπορούν να εκτιμηθούν σωστά είτε στα επίπεδα είτε στις διαφορές των χρονοσειρών. Επίσης, οι συνέπειες της υποδιαφοροποίησης είναι πολύ πιο σοβαρές από αυτές της υπερδιαφοροποίησης. Ειδικότερα, στην περίπτωση της υπερδιαφοροποίησης οι εκτιμητές είναι συνεπείς και αμερόληπτοι αλλά δεν είναι αποδοτικοί, ενώ στην περίπτωση της υποδιαφοροποίησης οι εκτιμητές δεν είναι συνεπείς και δεν έχουν ροπές. Τότε είναι που εμφανίζεται το φαινόμενο της υποφαινομενικής παλινδρόμησης που έχουν περιγράψει οι Granger και Newbold (1974) και απέδειξε ο Phillips (1986). Επιπρόσθετα, οι Plosser και Schwert (1978) προτείνουν τη χρησιμοποίηση της διαφοροποίησης σαν ένα έλεγχο για την εγκυρότητα του υποδείγματος. Αυτό βασίζεται στο συμπέρασμα πως όταν το υπόδειγμα είναι ταυτοποιημένο σωστά η εκτίμησή του στις διαφορές δίνει παρόμοια αποτελέσματα με αυτά που έδωσε η εκτίμηση στα επίπεδα. Στην αντίθετη περίπτωση, όταν τα αποτελέσματα απέχουν σημαντικά, το υπόδειγμα δεν θα πρέπει να είναι σωστά ταυτοποιημένο.

5.3 Η επιλογή ενός υποδείγματος ανάμεσα στα επίπεδα και στις διαφορές

Πολλές φορές η απόφαση για την επιλογή ενός υποδείγματος παλινδρόμησης με τα επίπεδα των χρονοσειρών από το αντίστοιχο με τις πρώτες διαφορές των χρονοσειρών είναι δύσκολη. Αυτό οφείλεται στην απουσία κάποιου κριτηρίου για τη σύγκριση των δύο υποδειγμάτων. Ο Harvey (1980) κατασκεύασε μια σειρά κριτηρίων τα οποία μπορούν να χρησιμοποιηθούν στην επιλογή του υποδείγματος, όταν οι εναλλακτικές περιπτώσεις είναι τα επίπεδα και οι πρώτες διαφορές των χρονοσειρών. Η φιλοσοφία στην οποία βασίζονται αυτά τα κριτήρια είναι ο έλεγχος της καλής προσαρμογής των υποδειγμάτων, όπως εκφράζεται από

τη λογαριθμική μορφή της συνάρτησης μέγιστης πιθανοφάνειας, συγκρίνοντας τις τιμές των διακυμάνσεων των σφαλμάτων.

Έστω το υπόδειγμα της γραμμικής παλινδρόμησης στα επίπεδα των χρονοσειρών:

$$Y_t = a + X_t' b + e_t \quad \text{για } t = 1, \mathbf{K}, T \quad (5.5)$$

όπου Y_t είναι η t παρατήρηση της εξαρτημένης μεταβλητής, X_t είναι το $K \times 1$ διάνυσμα των ερμηνευτικών μεταβλητών τη χρονική στιγμή t , b είναι το $K \times 1$ διάνυσμα των αγνώστων παραμέτρων και a είναι η άγνωστη παράμετρος του σταθερού όρου, ενώ για τα σφάλματα e_t ισχύει $e_t \sim iidN(0, \mathbf{S}_0^2)$. Το υπόδειγμα (5.5) μπορεί να συγκριθεί με το υπόδειγμα των πρώτων διαφορών:

$$\Delta Y_t = (\Delta X_t)' g + h_t \quad \text{για } t = 2, \mathbf{K}, T \quad (5.6)$$

όπου $\Delta Y_t = Y_t - Y_{t-1}$, ΔX_t είναι το διάνυσμα των πρώτων διαφορών των χρονοσειρών και g είναι το $K \times 1$ διάνυσμα των αγνώστων παραμέτρων, ενώ για τα σφάλματα h_t ισχύει $h_t \sim iidN(0, \mathbf{S}_1^2)$. Όμως, η σύγκριση της συνάρτησης μέγιστης πιθανοφάνειας του (5.5) με αυτή του (5.6) δεν είναι δυνατή γιατί τα δύο υποδείγματα εκτιμώνται με διαφορετικό αριθμό παρατηρήσεων. Επιπλέον, υπάρχει πρόβλημα με την τιμή του Y_1 στην εκτίμηση του υποδείματος (5.6). Για να λυθεί το πρόβλημα αυτό ο Harvey προτείνει το υπόδειγμα (5.6) να γραφεί ως:

$$Y_t = Y_{t-1} + (\Delta X_t)' g + h_t \quad \text{για } t = 2, \mathbf{K}, T \quad (5.7)$$

και να θεωρηθεί το Y_1 σταθερό ώστε να μπορεί να οριστεί μια συνάρτηση μέγιστης πιθανοφάνειας για τα Y_2, \mathbf{K}, Y_T η οποία θα έχει την ίδια μορφή με τη συνάρτηση μέγιστης πιθανοφάνειας για το υπόδειγμα (5.6). Στην συνέχεια για να συγκριθούν τα υποδείγματα μπορούν να κατασκευαστούν τα κριτήρια AIC για το κάθε υπόδειγμα και να προσαρμοσθεί με την χρήση του συντελεστή $\frac{T}{T-1}$ το AIC του υποδείματος των πρώτων διαφορών. Με βάση την προηγούμενη διαδικασία ο Harvey (1980) ορίζει για τη σύγκριση των δύο υποδειγμάτων το κριτήριο:

$$d_1 = \frac{\hat{\mathbf{S}}_0^2}{\hat{\mathbf{S}}_1^2} \exp \left\{ \frac{2(K+2)}{T} - \frac{2(K+1)}{T-1} \right\}$$

όπου \hat{S}_0^2 και \hat{S}_1^2 είναι οι εκτιμήσεις μέγιστης πιθανοφάνειας των διακυμάνσεων S_0^2 και S_1^2 για τα επίπεδα και τις πρώτες διαφορές. Όταν $d_1 < 1$ επιλέγεται το υπόδειγμα στα επίπεδα διαφορετικά επιλέγεται στις πρώτες διαφορές.

Μια διαφορετική προσέγγιση για την τιμή του Y_1 είναι να μην θεωρηθεί σταθερό αλλά ότι παράγεται με βάση το υπόδειγμα στα επίπεδα ως εξής:

$$Y_1 = \bar{a} + X_1'g + h_1 \quad (5.8)$$

όπου $h_1 \sim iidN(0, S_1^2)$. Η σχέση (5.8) μπορεί μαζί με την (5.7) να ορίσει τη συνάρτηση πιθανοφάνειας και για τις T παρατηρήσεις. Ο εκτιμητής μέγιστης πιθανοφάνειας του g και το άθροισμα τετραγώνων των σφαλμάτων παραμένουν αμετάβλητα όταν η Y_1 προστίθεται στα δεδομένα, ενώ το υπόδειγμα στις πρώτες διαφορές περιέχει τον ίδιο αριθμό παραμέτρων με αυτόν του υποδείγματος στα επίπεδα. Επομένως, η σύγκριση μπορεί να γίνει άμεσα με βάση τις συναρτήσεις πιθανοφάνειας, συγκρίνοντας το άθροισμα τετραγώνων των σφαλμάτων του υποδείγματος στα επίπεδα SSE_0 με το άθροισμα τετραγώνων των σφαλμάτων του υποδείγματος στις πρώτες διαφορές SSE_1 ή εναλλακτικά να συγκριθούν οι αμερόληπτοι εκτιμητές S_0^2 και S_1^2 των διακυμάνσεων S_0^2 και S_1^2 αφού και τα δύο υποδείγματα έχουν τους ίδιους, $T - K - 1$, βαθμούς ελευθερίας. Κατά συνέπεια, ο Harvey (1980) ορίζει για τη σύγκριση των δύο υποδειγμάτων το κριτήριο:

$$d_2 = \frac{SSE_0}{SSE_1} = \frac{S_0^2}{S_1^2}$$

Όπως και με το κριτήριο d_1 , όταν $d_2 < 1$ επιλέγεται το υπόδειγμα στα επίπεδα διαφορετικά επιλέγεται αυτό στις πρώτες διαφορές.

Μια διαφορετική μορφή με την οποία μπορεί να συναντηθεί το υπόδειγμα των διαφορών (5.6) είναι η:

$$\Delta Y_t = (\Delta X_t)' b + u_t \quad \text{για } t = 2, \mathbf{K}, T$$

όπου $u_t = e_t - e_{t-1}$. Η μορφή αυτή δεν απαιτεί η πρώτη παρατήρηση να είναι σταθερή και προκύπτει άμεσα από το υπόδειγμα (5.5) όταν διαφοροποιηθεί. Ο Harvey (1980) αποδεικνύει ότι σε αυτή την περίπτωση η σύγκριση μπορεί να γίνει με το κριτήριο:

$$d_3 = \frac{SSE_0}{SSE_1} \exp \left\{ \frac{\ln T}{T-1} \right\}$$

όπου πάλι όταν $d_3 < 1$ επιλέγεται το υπόδειγμα στα επίπεδα.

Μια επέκταση που μπορεί να γίνει στο κριτήριο d αφορά τη δυνατότητα της διάκρισης μεταξύ δύο υποδειγμάτων στα επίπεδα και στις διαφορές όταν τα σφάλματα ακολουθούν μια στάσιμη και αντιστρέψιμη ARMA διαδικασία. Έστω ότι το υπόδειγμα στα επίπεδα είναι:

$$Y_t = a + X_t' b + u_t \quad \text{για } t = 1, \mathbf{K}, T$$

όπου όμως τα σφάλματα u_t ακολουθούν τη στάσιμη και αντιστρέψιμη ARMA(p_0, q_0) διαδικασία:

$$u_t = f_1 u_{t-1} + \mathbf{K} + f_{p_0} u_{t-p_0} + e_t + q_1 e_{t-1} + \mathbf{K} + q_{q_0} e_{t-q_0}$$

με $e_t \sim iidN(0, S_0^2)$ και έχουν πίνακα συνδιακυμάνσεων $S_0^2 V_0$. Εάν $f_0 = (f_1, \mathbf{K}, f_{p_0})'$ και $q_0 = (q_1, \mathbf{K}, q_{q_0})'$ τότε το άθροισμα τετραγώνων των σφαλμάτων που προκύπτει από την εκτίμηση του υποδείγματος με τη μέθοδο των γενικευμένων ελαχίστων τετραγώνων (GLS) είναι:

$$SSE_0 = \hat{u}' V_0^{-1} \hat{u}$$

όπου το t στοιχείο του $T \times 1$ διανύσματος \hat{u} είναι τα σφάλματα:

$$\hat{u}_t = Y_t - \hat{a} - X_t' \hat{b}$$

ενώ \hat{a} και \hat{b} είναι οι εκτιμήσεις για τα a και b που προκύπτουν με τη μέθοδο των γενικευμένων ελαχίστων τετραγώνων. Όμοια το υπόδειγμα στις πρώτες διαφορές (5.6) μπορεί να γραφεί:

$$\Delta Y_t = (\Delta X_t)' g + w_t \quad \text{για } t = 2, \mathbf{K}, T$$

όπου τα σφάλματα w_t ακολουθούν μια στάσιμη και αντιστρέψιμη ARMA(p_1, q_1) διαδικασία, με πίνακα συνδιακυμάνσεων $S_1^2 V_1$ και $SSE_1 = \hat{w}' V_1^{-1} \hat{w}$. Εφόσον τηρούνται οι παραπάνω υποθέσεις ο Harvey (1980) ορίζει το κριτήριο:

$$d_4 = \frac{SSE_0}{SSE_1} \exp \left\{ \frac{1}{T} \ln \left(\frac{|V_0|}{|V_1|} \right) + 2 \frac{(p_0 + q_0 - p_1 - q_1)}{T} \right\}$$

το οποίο, όπως και στις προηγούμενες περιπτώσεις, όταν είναι μικρότερο της μονάδας σημαίνει ότι πρέπει να επιλεγεί το υπόδειγμα στα επίπεδα.

Τα παραπάνω κριτήρια βοηθάνε στην επιλογή μεταξύ δύο υποδειγμάτων όταν το ένα έχει εκτιμηθεί στα επίπεδα και το άλλο στις διαφορές των χρονοσειρών. Όμως, από τις

προσομοιώσεις που πραγματοποίησε ο Harvey (1980) διαπιστώθηκε αδυναμία των κριτηρίων στην επιλογή του σωστού υποδείγματος όταν το δείγμα είναι μικρό. Το σοβαρότερο πρόβλημα εμφανίστηκε όταν το σωστό υπόδειγμα είναι ορισμένο στα επίπεδα αλλά τα σφάλματα ακολουθούν μια AR(1) διαδικασία με τιμές για το f πολύ κοντά στη μονάδα. Στην περίπτωση αυτή, το πιθανότερο είναι να επιλεγεί ένα υπόδειγμα σε πρώτες διαφορές.

5.4 Ο συντελεστής πολλαπλού προσδιορισμού για την σύγκριση με ένα υπόδειγμα σε διαφορές.

Ο συντελεστής πολλαπλού προσδιορισμού αποτελεί το πιο συχνά χρησιμοποιούμενο κριτήριο για την επιλογή του καλύτερου υποδείγματος μεταξύ υποδειγμάτων με τον ίδιο αριθμό ανεξάρτητων μεταβλητών και παρατηρήσεων. Όταν όμως χρειάζεται να γίνει η επιλογή μεταξύ ενός υποδείγματος στα επίπεδα και ενός υποδείγματος στις πρώτες διαφορές, ο συντελεστής προσδιορισμού δεν μπορεί να χρησιμοποιηθεί γιατί τα δύο υποδείγματα δεν έχουν τον ίδιο αριθμό παρατηρήσεων και παραμέτρων. Για το λόγο αυτό είναι αναγκαίο ο συντελεστής προσδιορισμού στα επίπεδα να τροποποιηθεί, ώστε να γίνει συγκρίσιμος με αυτόν που προκύπτει από την εκτίμηση του υποδείγματος στις πρώτες διαφορές.

Ο Harvey (1980) προτείνει την χρήση του συντελεστή πολλαπλού προσδιορισμού για το υπόδειγμα στα επίπεδα:

$$R_D^2 = 1 - \frac{SSE_0}{\sum_{t=2}^T (\Delta Y_t - \overline{\Delta Y_t})^2} = 1 - \frac{SSE_0 (1 - R_1^2)}{SSE_1} \quad (5.9)$$

όπου $\overline{\Delta Y_t}$ είναι ο μέσος όρος της χρονοσειράς των πρώτων διαφορών για την εξαρτημένη μεταβλητή, SSE_0 το άθροισμα των τετραγώνων των σφαλμάτων από την εκτίμηση του υποδείγματος στα επίπεδα, SSE_1 το άθροισμα των τετραγώνων των σφαλμάτων από την εκτίμηση του υποδείγματος σε πρώτες διαφορές και R_1^2 ο συντελεστής πολλαπλού προσδιορισμού για το υπόδειγμα σε πρώτες διαφορές. Ο συντελεστής πολλαπλού προσδιορισμού R_D^2 δεν μπορεί να πάρει τιμές μεγαλύτερες από τη μονάδα αλλά μπορεί να πάρει και αρνητικές τιμές και είναι συγκρίσιμος με το συντελεστή πολλαπλού προσδιορισμού του υποδείγματος σε πρώτες διαφορές.

Ο Maddala (1988) παρατηρεί ότι μειονέκτημα του συντελεστή πολλαπλού προσδιορισμού (5.9) είναι ότι δεν λαμβάνει υπ' όψιν τις διακυμάνσεις των σφαλμάτων στο υπόδειγμα των πρώτων διαφορών οι οποίες δεν είναι ίδιες με αυτές του υποδείματος στα επίπεδα. Εάν τα σφάλματα e_t του υποδείματος στα επίπεδα έχουν διακύμανση s^2 , τότε τα σφάλματα $e_t - e_{t-1}$, του υποδείματος στις πρώτες διαφορές έχουν διακύμανση:

$$\begin{aligned} \text{Var}(e_t - e_{t-1}) &= \text{Var}(e_t) + \text{Var}(e_{t-1}) - 2\text{Cov}(e_t, e_{t-1}) \\ &= s^2 + s^2 - 2s^2 r \\ &= 2s^2(1 - r) = s^2 DW \end{aligned}$$

όπου r είναι ο συντελεστής συσχέτισης μεταξύ των e_t και e_{t-1} και DW η τιμή της στατιστικής των Durbin και Watson για το υπόδειγμα στα επίπεδα. Επομένως, επειδή το άθροισμα τετραγώνων των σφαλμάτων όταν διαιρεθεί με τους βαθμούς ελευθερίας του υποδείματος είναι συνεπής εκτιμητής της διακύμανσης των σφαλμάτων, εάν πολλαπλασιαστεί το άθροισμα των τετραγώνων των σφαλμάτων του υποδείματος στα επίπεδα με την ποσότητα:

$$\left(\frac{T - K - 1}{T - K} \right) DW$$

όπου K είναι ο αριθμός των ανεξαρτήτων μεταβλητών, μπορεί να γίνει συγκρίσιμο με το άθροισμα τετραγώνων των σφαλμάτων του υποδείματος σε πρώτες διαφορές. Με τον τρόπο αυτό ο Maddala βελτιώνει το συντελεστή πολλαπλού προσδιορισμού (5.9) του Harvey σε:

$$R_D^2 = 1 - \frac{SSE_0}{SSE_1} \left(\frac{T - K - 1}{T - K} \right) DW (1 - R_1^2) \quad (5.10)$$

Ο συντελεστής προσδιορισμού (5.10) είναι συγκρίσιμος με το συντελεστή R_1^2 του υποδείματος στις πρώτες διαφορές και λαμβάνει υπ' όψιν τις διαφορετικές διακυμάνσεις των σφαλμάτων.

5.5 Ανακεφαλαίωση

Η απόφαση για την επιλογή μεταξύ ενός υποδείγματος στα επίπεδα των χρονοσειρών και ενός υποδείγματος στις διαφορές των χρονοσειρών μερικές φορές είναι δύσκολη. Τα κριτήρια που αναφέρθηκαν παραπάνω, παρόλο τις αδυναμίες που έχουν, μπορούν να βοηθήσουν στην επιλογή του σωστού υποδείγματος. Όμως, στην περίπτωση που δεν είναι σαφής η επιλογή του υποδείγματος είναι προτιμότερο να προτιμάται ένα υπόδειγμα στις διαφορές των χρονοσειρών. Ο λόγος είναι γιατί οι συνέπειες της εσφαλμένης χρήσης διαφορών στις χρονοσειρές είναι ασήμαντες σε σχέση με την μη χρήση όταν είναι αναγκαία. Δηλαδή, η εκτίμηση ενός υπερδιαφοροποιημένου υποδείγματος θα δώσει συνεπείς και αμερόληπτους εκτιμητές αλλά όχι αποδοτικούς, σε αντίθεση με την εκτίμηση ενός υποδιαφοροποιημένου υποδείγματος που οι εκτιμητές που θα προκύψουν θα είναι μη συνεπείς και δεν θα έχουν ροπές. Σε ένα σωστά ταυτοποιημένο υπόδειγμα όσες φορές και να χρησιμοποιηθούν διαφορές οι εκτιμήσεις των συντελεστών θα είναι πολύ κοντά. Οι διαφορές που θα προκύψουν θα είναι στις τυπικές τους αποκλίσεις και κατά συνέπεια στους ελέγχους για τη στατιστική τους σημαντικότητα, εξαιτίας της παρουσίας υποδειγμάτων κινητών μέσω των όρων για τα σφάλματα. Όταν όμως διορθωθούν οι αυτοσυσχετίσεις στα σφάλματα θα πρέπει και οι έλεγχοι για την στατιστική σημαντικότητα να οδηγούν στα ίδια αποτελέσματα. Με βάση τα προηγούμενα γίνεται εύκολα αντιληπτή η λογική και το αποτέλεσμα της προσομοίωσης των Granger και Newbold (1974) για την εκτίμηση του υποδείγματος στις πρώτες διαφορές των χρονοσειρών όταν αυτές είναι ολοκληρωμένες διαδικασίες που αναφέρθηκε στο κεφάλαιο 2.

ΠΑΡΑΡΤΗΜΑΤΑ

Π1 Αποτελέσματα Προσομοιώσεων

Π2 Προγράμματα Προσομοιώσεων (Mathematica)

Π1 Αποτελέσματα Προσομοιώσεων

ΠΙΝΑΚΑΣ 1

Υποφαινομενικές παλινδρομήσεις μεταξύ δύο τυχαίων περιπάτων με μέσο όρο για $\gamma_Y = 0$ και για διάφορες τιμές του γ_X , $T = 100$

γ_X	100	200	300	400	1000	10000
Ποσοστό απόρριψης H_0 , ($\alpha = 5\%$)	87.28	87.75	87.56	88.08	86.89	86.89
Μέση τιμή $ t $	10.74	10.80	10.75	10.65	10.61	10.61
Μέσο \hat{b}	0.00	0.00	0.00	0.00	0.00	0.00
Μέσο R^2	0.43	0.43	0.43	0.43	0.43	0.43
Ελάχιστο R^2	0	0	0	0	0	0
Μέγιστο R^2	0.99	0.98	0.98	0.98	0.99	0.99
Μέση τιμή DW	0.198	0.200	0.199	0.200	0.198	0.199
Ελάχιστο DW	0.022	0.021	0.014	0.018	0.017	0.015
Μέγιστο DW	0.764	0.918	0.873	0.917	0.827	0.931
Ποσοστό απόρριψης H_0 και $DW < d_L$	87.28	87.75	87.56	88.08	86.89	87.08
Ποσοστό απόρριψης H_0 και $DW < d_L$ και $R^2 > 0.7$	25.45	25.18	25.49	24.91	24.77	24.88

ΠΙΝΑΚΑΣ 2

Υποφαινομενικές παλινδρομήσεις μεταξύ δύο τυχαίων περιπάτων με μέσο όρο για $\gamma_x = 0$ και για διάφορες τιμές του γ_y , $T = 100$

γ_y	100	200	300	400	1000	10000
Ποσοστό απόρριψης H_0 , ($\alpha = 5\%$)	87.54	87.29	87.69	87.41	87.5	87
Μέση τιμή $ t $	10.56	10.74	10.59	10.81	10.76	10.68
Μέσο \hat{b}	-47.99	-92.30	-155.94	-191.42	-439.72	-4879.99
Μέσο R^2	0.43	0.43	0.43	0.43	0.43	0.43
Ελάχιστο R^2	0	0	0	0	0	0
Μέγιστο R^2	0.97	0.99	0.98	0.98	0.98	0.98
Μέση τιμή DW	0.094	0.093	0.094	0.095	0.095	0.093
Ελάχιστο DW	0.001	0.001	0.001	0.001	0.001	0.001
Μέγιστο DW	0.754	0.705	0.847	0.838	0.819	0.721
Ποσοστό απόρριψης H_0 και $DW < d_L$	87.54	87.29	87.69	87.41	87.5	87
Ποσοστό απόρριψης H_0 και $DW < d_L$ και $R^2 > 0.7$	24.59	25.26	24.66	25.52	24.97	24.99

ΠΙΝΑΚΑΣ 3

Υποφαινομενικές παλινδρομήσεις μεταξύ δύο τυχαίων περιπάτων με μέσο όρο για διάφορες τιμές των γ_Y και γ_X , $T = 100$

γ_X	2	20	200	2000	20000
γ_Y	4	40	400	4000	40000
Ποσοστό απόρριψης H_0 , ($\alpha = 5\%$)	100	100	100	100	100
Μέση τιμή $ t $	222.17	2238.26	22325.1	223794	2242720
Μέσο \hat{b}	2	2	2	2	2
Μέσο R^2	0.997	1.000	1.000	1.000	1.000
Ελάχιστο R^2	0.98	1.00	1.00	1.00	1.00
Μέγιστο R^2	1.00	1.00	1.00	1.00	1.00
Μέση τιμή DW	0.197	0.199	0.197	0.199	0.199
Ελάχιστο DW	0.021	0.023	0.021	0.019	0.024
Μέγιστο DW	0.896	0.909	0.909	1.121	0.807
Ποσοστό απόρριψης H_0 και $DW < d_L$	100	100	100	100	100
Ποσοστό απόρριψης H_0 και $DW < d_L$ και $R^2 > 0.7$	100	100	100	100	100

Π2 Προγράμματα Προσομοιώσεων (Mathematica)

α) Υποφαινομενικές παλινδρομήσεις για διάφορα μεγέθη δείγματος (σελ. 42)

```
Needs["Statistics`ContinuousDistributions`"];
Needs["Statistics`MultiDescriptiveStatistics`"];
Needs["Graphics`Graphics`"];
Needs["Statistics`DataManipulation`"];
Needs["Statistics`LinearRegression`"];
n = 124; φα = 1; φβ = 1;
series1 = Table[0, {n + 1}]; series2 = Table[0, {n + 1}];
σταθπαραμετροι = {}; βπαραμ = {};
ταστατιστικες = {}; τβστατ = {};
συντπροσδιορισμου = {}; dwστατιστικες = {}; spurious = {}; spurious2 = {};
Do[
  Do[
    u = Random[NormalDistribution[0, 1]];
    series1[[i]] = φα*series1[[i - 1]] + u;
    , {i, 2, n + 1}];
  Do[
    u = Random[NormalDistribution[0, 1]];
    series2[[i]] = φβ*series2[[i - 1]] + u;
    , {i, 2, n + 1}];
  series3 = Table[series1[[i]], {i, 26, n + 1}];
  series4 = Table[series2[[i]], {i, 26, n + 1}];
  tn = Quantile[StudentTDistribution[Length[series3] - 2, 0.9750] // N;
  (*d[[i]] = {series3[[i]], series4[[i]]};*)
  d = Table[d[[i]] = {series3[[i]], series4[[i]]}, {i, 1, 100}];
  regr = Regress[d, {1, x}, x,
    RegressionReport -> {ParameterTable, RSquared, DurbinWatsonD}];
  parametroi = {estimate, se, tstat, pvalue} =
    Transpose[(ParameterTable /. regr)[[1]]];
  statheros = parametroi[[1, 1]]; metablitia = parametroi[[1, 2]];
  ta = parametroi[[3, 1]]; tb = Abs[parametroi[[3, 2]]];
  prosd = (RSquared /. regr); dw = (DurbinWatsonD /. regr);
  σταθπαραμετροι = Append[σταθπαραμετροι, statheros];
  βπαραμ = Append[βπαραμ, metablitia];
  ταστατιστικες = Append[ταστατιστικες, ta];
  τβστατ = Append[τβστατ, tb];
  συντπροσδιορισμου = Append[συντπροσδιορισμου, prosd];
  dwστατιστικες = Append[dwστατιστικες, dw];
  If[(tb > tn) && (dw < 1.69), spurious = Append[spurious, 1]];
  If[(tb > tn) && (dw < 1.69) && (prosd >= 0.7),
    spurious2 = Append[spurious2, 1]];
  , {10000}];
g = Select[τβστατ, #1 > tn &];
Print["Η μέση τιμή της στατιστικής t σε απόλυτη τιμή για τη κλίση είναι "];
```

```

Mean[tβστατ]];
Print["Το μέσο R^2 είναι ", Mean[συντπροσδιορισμου]];
Print["Το ελάχιστο R^2 είναι ", Min[συντπροσδιορισμου], " και το μέγιστο ",
      Max[συντπροσδιορισμου]];
Print["Το μέσο dw είναι ", Mean[dwστατιστικες]];
Print["Το ελάχιστο dw είναι ", Min[dwστατιστικες], " και το μέγιστο ",
      Max[dwστατιστικες]];
Print["Η κριτική τιμή είναι: ", tn];
Print["Ο συντελεστής της κλίσης είναι σημαντικός ", Length[g], " φορές"];
Print["Αριθμός υποφαινομενικών παλινδρομήσεων ", Length[spurious]];
Print["Αριθμός υποφαινομενικών παλινδρομήσεων και R^2 > 0.7 ",
      Length[spurious2]];

```

β) Υποφαινομενικές παλινδρομήσεις μεταξύ δύο τυχαίων περιπάτων με μέσο όρο για διάφορα μεγέθη δείγματος (σελ. 52, 54 και 55)

```
Needs["Statistics`ContinuousDistributions`"];
Needs["Statistics`MultiDescriptiveStatistics`"];
Needs["Graphics`Graphics`"];
Needs["Statistics`DataManipulation`"];
Needs["Statistics`LinearRegression`"];
n = 124; φα = 1; φβ = 1;
series1 = Table[0, {n + 1}]; series2 = Table[0, {n + 1}];
σταθπαραμετροι = {}; βπαραμ = {};
ταστατιστικες = {}; tβστατ = {};
συντπροσδιορισμου = {}; dωστατιστικες = {}; spurious = {}; spurious2 = {};
Do[
  Do[
    u = Random[NormalDistribution[0, 1]];
    series1[[i]] = 200 + φα*series1[[i - 1]] + u;
    , {i, 2, n + 1}];
  Do[
    u = Random[NormalDistribution[0, 1]];
    series2[[i]] = 400 + φβ*series2[[i - 1]] + u;
    , {i, 2, n + 1}];
  series3 = Table[series1[[i]], {i, 26, n + 1}];
  series4 = Table[series2[[i]], {i, 26, n + 1}];
  tn = Quantile[StudentTDistribution[Length[series3] - 2], 0.9750] // N;
  (*d[[i]] = {series3[[i]], series4[[i]]};*)
  d = Table[d[[i]] = {series3[[i]], series4[[i]]}, {i, 1, 100}];
  regr =
  Regress[d, {1, x}, x,
  RegressionReport -> {ParameterTable, RSquared, DurbinWatsonD}];
parametroi = {estimate, se, tstat, pvalue} =
  Transpose[(ParameterTable /. regr)[[1]]];
statheros = parametroi[[1, 1]]; metablitia = parametroi[[1, 2]];
ta = parametroi[[3, 1]]; tb = Abs[parametroi[[3, 2]]];
prosd = (RSquared /. regr); dw = (DurbinWatsonD /. regr);
σταθπαραμετροι = Append[σταθπαραμετροι, statheros];
βπαραμ = Append[βπαραμ, metablitia];
ταστατιστικες = Append[ταστατιστικες, ta];
tβστατ = Append[tβστατ, tb];
συντπροσδιορισμου = Append[συντπροσδιορισμου, prosd];
dωστατιστικες = Append[dωστατιστικες, dw];
If[(tb > tn) && (dw < 1.69), spurious = Append[spurious, 1]];
If[(tb > tn) && (dw < 1.69) && (prosd >= 0.7),
  spurious2 = Append[spurious2, 1]];
, {10000}];
```

```

g = Select[tβστατ, #1 > tn &];
Print["Η μέση τιμή της στατιστικής t σε απόλυτη τιμή για τη κλίση είναι ",
      Mean[tβστατ]];
Print["Το μέσο R^2 είναι ", Mean[συντπροσδιορισμου]];
Print["Το ελάχιστο R^2 είναι ", Min[συντπροσδιορισμου], " και το μέγιστο ",
      Max[συντπροσδιορισμου]];
Print["Το μέσο dw είναι ", Mean[dwστατιστικες]];
Print["Το ελάχιστο dw είναι ", Min[dwστατιστικες], " και το μέγιστο ",
      Max[dwστατιστικες]];
Print["Η κριτική τιμή είναι: ", tn];
Print["Ο συντελεστής της κλίσης είναι σημαντικός ", Length[g], " φορές"];
Print["Αριθμός υποφαινομενικών παλινδρομήσεων ", Length[spurious]];
Print["Αριθμός υποφαινομενικών παλινδρομήσεων και R^2 > 0.7 ",
      Length[spurious2]];
Print["το μεσο β είναι ", Mean[βπαραμ]];
h = Histogram[βπαραμ];

```

γ) Υποφαινομενικές παλινδρομήσεις λόγω σχέσης με τρίτη μεταβλητή (σελ. 77)

```
Needs["Statistics`ContinuousDistributions`"];
Needs["Statistics`MultiDescriptiveStatistics`"];
Needs["Graphics`Graphics`"];
Needs["Statistics`DataManipulation`"];
Needs["Statistics`LinearRegression`"];
n = 124; φα = 0;
series1 = Table[0, {n + 1}]; series2 = Table[0, {n + 1}]; series3 =
  Table[0, {n + 1}];
σταθπαραμετροι = {}; βπαραμ = {};
ταστατιστικες = {}; τβστατ = {};
συντπροσδιορισμου = {}; δωστατιστικες = {}; spurious = {}; spurious2 = {};
Do[
  Do[
    u = Random[NormalDistribution[0, 1]];
    series3[[i]] = φα*series3[[i - 1]] + u;
    , {i, 2, n + 1}];
  Do[
    u = Random[NormalDistribution[0, 1]];
    series1[[i]] = 100 + 2*series3[[i]] + u;
    , {i, 2, n + 1}];
  Do[
    u = Random[NormalDistribution[0, 1]];
    series2[[i]] = 100 + 4*series3[[i]] + u;
    , {i, 2, n + 1}];
  series4 = Table[series1[[i]], {i, 26, n + 1}];
  series5 = Table[series2[[i]], {i, 26, n + 1}];
  tn = Quantile[StudentTDistribution[Length[series4] - 2], 0.9750] // N;
  (*d[[i]] = {series4[[i]], series5[[i]]};*)
  d = Table[d[[i]] = {series4[[i]], series5[[i]]}, {i, 1, 100}];
  regr =
    Regress[d, {1, x}, x,
      RegressionReport -> {ParameterTable, RSquared, DurbinWatsonD}];
  parametroi = {estimate, se, tstat, pvalue} =
    Transpose[(ParameterTable /. regr)[[1]]];
  statheros = parametroi[[1, 1]]; metablitia = parametroi[[1, 2]];
  ta = parametroi[[3, 1]]; tb = Abs[parametroi[[3, 2]]];
  prosd = (RSquared /. regr); dw = (DurbinWatsonD /. regr);
  σταθπαραμετροι = Append[σταθπαραμετροι, statheros];
  βπαραμ = Append[βπαραμ, metablitia];
  ταστατιστικες = Append[ταστατιστικες, ta];
  τβστατ = Append[τβστατ, tb];
  συντπροσδιορισμου = Append[συντπροσδιορισμου, prosd];
  δωστατιστικες = Append[δωστατιστικες, dw];
```

```

If[(tb > tn) && (dw < 1.69), spurious = Append[spurious, 1]];
If[(tb > tn) && (dw < 1.69) && (prosd >= 0.7),
  spurious2 = Append[spurious2, 1]];
, {10000}];
g = Select[tβστατ, #1 > tn &];
Print["Η μέση τιμή της στατιστικής t σε απόλυτη τιμή για τη κλίση είναι ",
  Mean[tβστατ]];
Print["Το μέσο R^2 είναι ", Mean[συντπροσδιορισμου]];
Print["Το ελάχιστο R^2 είναι ", Min[συντπροσδιορισμου], " και το μέγιστο ",
  Max[συντπροσδιορισμου]];
Print["Το μέσο dw είναι ", Mean[dwστατιστικες]];
Print["Το ελάχιστο dw είναι ", Min[dwστατιστικες], " και το μέγιστο ",
  Max[dwστατιστικες]];
Print["Η κριτική τιμή είναι: ", tn];
Print["Ο συντελεστής της κλίσης είναι σημαντικός ", Length[g], " φορές"];
Print["Αριθμός υποφαινομενικών παλινδρομήσεων ", Length[spurious]];
Print["Αριθμός υποφαινομενικών παλινδρομήσεων και R^2 > 0.7 ",
  Length[spurious2]];
Print["το μέσο β είναι ", Mean[βπαραμ]];
h = Histogram[βπαραμ];

```

δ) Υποφαινομενικές παλινδρομήσεις υποδείγματος (3.9) (σελ. 81)

```
Needs["Statistics`ContinuousDistributions`"];
Needs["Statistics`MultiDescriptiveStatistics`"];
Needs["Graphics`Graphics`"];
Needs["Statistics`DataManipulation`"];
Needs["Statistics`LinearRegression`"];
n = 124; φ $\alpha$  = 1; φ $\beta$  = 1;
series1 = Table[0, {n + 1}]; series2 = Table[0, {n + 1}];
σταθπαραμετροι = {}; βπαραμ = {};
ταστατιστικες = {}; tβστατ = {}; fστατ = {};
συντπροσδιορισμου = {}; διορθσυντπροσδιορισμου = {}; dωστατιστικες = {};
spurious = {}; spurious2 = {}; spurious3 = {}; spurious4 = {};
Do[
  Do[
    u = Random[NormalDistribution[0, 1]];
    series1[[i]] = φ $\alpha$ *series1[[i - 1]] + u;
    , {i, 2, n + 1}];
  Do[
    u = Random[NormalDistribution[0, 1]];
    series2[[i]] = φ $\beta$ *series2[[i - 1]] + u;
    , {i, 2, n + 1}];
  series3 = Table[series1[[i]], {i, 26, n + 1}];
  series4 = Table[series2[[i]], {i, 26, n + 1}];
  tn = Quantile[StudentTDistribution[Length[series3] - 2], 0.9750] // N;
  fn = Quantile[FRatioDistribution[1, Length[series3] - 2], 0.950] // N;
  (*d[[i]] = {series3[[i]], series4[[i]]};*)
  d = Table[d[[i]] = {series3[[i]], series4[[i]]}, {i, 1, 100}];
  regr =
  Regress[d, {1, x^2}, x,
  RegressionReport -> {ParameterTable, RSquared, AdjustedRSquared,
  DurbinWatsonD, ANOVATable}];
parametroi = {estimate, se, tstat, pvalue} =
  Transpose[(ParameterTable /. regr)[[1]]];
anova = (ANOVATable /. regr)[[1]];
statheros = parametroi[[1, 1]]; metablitia = parametroi[[1, 2]];
ta = parametroi[[3, 1]]; tb = Abs[parametroi[[3, 2]]]; f = anova[[1, 4]];
prosd = (RSquared /. regr); diorthprosd = (AdjustedRSquared /. regr);
dw = (DurbinWatsonD /. regr);
σταθπαραμετροι = Append[σταθπαραμετροι, statheros];
βπαραμ = Append[βπαραμ, metablitia];
ταστατιστικες = Append[ταστατιστικες, ta];
tβστατ = Append[tβστατ, tb];
fστατ = Append[fστατ, f];
συντπροσδιορισμου = Append[συντπροσδιορισμου, prosd];
```

```

διορθσυντπροσδιορισμου = Append[διορθσυντπροσδιορισμου, diorthprosd];
dωστατιστικες = Append[dωστατιστικες, dw];
If[(tb > tn) && (dw < 1.69), spurious = Append[spurious, 1]];
If[(tb > tn) && (dw < 1.69) && (prosd >= 0.7),
  spurious2 = Append[spurious2, 1]];
If[(f > fn) && (dw < 1.69), spurious3 = Append[spurious3, 1]];
If[(f > fn) && (dw < 1.69) && (prosd >= 0.7),
  spurious4 = Append[spurious4, 1]];
, {10000}];
g = Select[tβστατ, #1 > tn &];
h = Select[fστατ, #1 > fn &];
Print["Η μέση τιμή της στατιστικής t σε απόλυτη τιμή για τη κλίση είναι ",
  Mean[tβστατ]];
Print["Η μέση τιμή της στατιστικής F είναι ", Mean[fστατ]];
Print["Το μέσο R^2 είναι ", Mean[συντπροσδιορισμου]];
Print["Το ελάχιστο R^2 είναι ", Min[συντπροσδιορισμου], " και το μέγιστο ",
  Max[συντπροσδιορισμου]];
Print["Το μέσο διορθ. R^2 είναι ", Mean[διορθσυντπροσδιορισμου]];
Print["Το ελάχιστο διορθ. R^2 είναι ", Min[διορθσυντπροσδιορισμου],
  " και το μέγιστο ", Max[διορθσυντπροσδιορισμου]];
Print["Το μέσο dw είναι ", Mean[dωστατιστικες]];
Print["Το ελάχιστο dw είναι ", Min[dωστατιστικες], " και το μέγιστο ",
  Max[dωστατιστικες]];
Print["Η κριτική τιμή για τη t είναι: ", tn];
Print["Η κριτική τιμή για τη F είναι: ", fn];
Print["Ο συντελεστής της κλίσης είναι σημαντικός ", Length[g], " φορές"];
Print["Το υπόδειγμα είναι καλό με τη F ", Length[h], " φορές"];
Print["Αριθμός υποφαινομενικών παλινδρομήσεων ", Length[spurious]];
Print["Αριθμός υποφαινομενικών παλινδρομήσεων και R^2 > 0.7 ",
  Length[spurious2]];
Print["Αριθμός υποφαινομενικών παλινδρομήσεων με F ", Length[spurious3]];
Print["Αριθμός υποφαινομενικών παλινδρομήσεων με F και R^2 > 0.7 ",
  Length[spurious4]];

```


ε) Υποφαινομενικές παλινδρομήσεις υποδείγματος (3.10) (σελ. 82)

```
Needs["Statistics`ContinuousDistributions`"];
Needs["Statistics`MultiDescriptiveStatistics`"];
Needs["Graphics`Graphics`"];
Needs["Statistics`DataManipulation`"];
Needs["Statistics`LinearRegression`"];
n = 124; φ $\alpha$  = 1; φ $\beta$  = 1;
series1 = Table[0, {n + 1}]; series2 = Table[0, {n + 1}];
σταθπαραμετροι = {}; βπαραμ = {};
ταστατιστικες = {}; tβστατ = {}; τυστατ = {}; fστατ = {};
συντπροσδιορισμου = {}; διορθσυντπροσδιορισμου = {}; dωστατιστικες = {};
spurious = {}; spurious2 = {}; spurious3 = {}; spurious4 = {};
Do[
  Do[
    u = Random[NormalDistribution[0, 1]];
    series1[[i]] = φ $\alpha$ *series1[[i - 1]] + u;
    , {i, 2, n + 1}];
  Do[
    u = Random[NormalDistribution[0, 1]];
    series2[[i]] = φ $\beta$ *series2[[i - 1]] + u;
    , {i, 2, n + 1}];
  series3 = Table[series1[[i]], {i, 26, n + 1}];
  series4 = Table[series2[[i]], {i, 26, n + 1}];
  tn = Quantile[StudentTDistribution[Length[series3] - 3], 0.9750] // N;
  fn = Quantile[FRatioDistribution[2, Length[series3] - 3], 0.950] // N;
  (*d[[i]] = {series3[[i]], series4[[i]]};*)
  d = Table[d[[i]] = {series3[[i]], series4[[i]]}, {i, 1, 100}];
  regr =
  Regress[d, {1, x, x^2}, x,
  RegressionReport -> {ParameterTable, RSquared, AdjustedRSquared,
  DurbinWatsonD, ANOVATable}];
parametroi = {estimate, se, tstat, pvalue} =
  Transpose[(ParameterTable /. regr)[[1]]];
anova = (ANOVATable /. regr)[[1]];
statheros = parametroi[[1, 1]]; metablitia = parametroi[[1, 2]];
ta = parametroi[[3, 1]]; tb = Abs[parametroi[[3, 2]]];
ty = Abs[parametroi[[3, 3]]]; f = anova[[1, 4]];
prosd = (RSquared /. regr); diorthprosd = (AdjustedRSquared /. regr);
dw = (DurbinWatsonD /. regr);
σταθπαραμετροι = Append[σταθπαραμετροι, statheros];
βπαραμ = Append[βπαραμ, metablitia];
ταστατιστικες = Append[ταστατιστικες, ta];
tβστατ = Append[tβστατ, tb];
τυστατ = Append[τυστατ, ty];
```

```

fστατ = Append[fστατ, f];
συντπροσδιορισμου = Append[συντπροσδιορισμου, prosd];
διορθσυντπροσδιορισμου = Append[διορθσυντπροσδιορισμου, diorthprosd];
dωστατιστικες = Append[dωστατιστικες, dw];

If[(f > fn) && (dw < 1.72), spurious3 = Append[spurious3, 1]];
If[(f > fn) && (dw < 1.72) && (prosd >= 0.7),
  spurious4 = Append[spurious4, 1]];
  , {10000}];
g = Select[tβστατ, #1 > tn &];
e = Select[tγστατ, #1 > tn &];
h = Select[fστατ, #1 > fn &];
Print["Η μέση τιμή της στατιστικής t σε απόλυτη τιμή για τη x είναι ",
  Mean[tβστατ]];
Print["Η μέση τιμή της στατιστικής t σε απόλυτη τιμή για τη x^2 είναι ",
  Mean[tγστατ]];
Print["Η μέση τιμή της στατιστικής F είναι ", Mean[fστατ]];
Print["Το μέσο R^2 είναι ", Mean[συντπροσδιορισμου]];
Print["Το ελάχιστο R^2 είναι ", Min[συντπροσδιορισμου], " και το μέγιστο ",
  Max[συντπροσδιορισμου]];
Print["Το μέσο διορθ. R^2 είναι ", Mean[διορθσυντπροσδιορισμου]];
Print["Το ελάχιστο διορθ. R^2 είναι ", Min[διορθσυντπροσδιορισμου],
  " και το μέγιστο ", Max[διορθσυντπροσδιορισμου]];
Print["Το μέσο dw είναι ", Mean[dωστατιστικες]];
Print["Το ελάχιστο dw είναι ", Min[dωστατιστικες], " και το μέγιστο ",
  Max[dωστατιστικες]];
Print["Η κριτική τιμή για τη t είναι: ", tn];
Print["Η κριτική τιμή για τη F είναι: ", fn];
Print["Ο συντελεστής της x είναι σημαντικός ", Length[g], " φορές"];
Print["Ο συντελεστής της x^2 είναι σημαντικός ", Length[e], " φορές"];
Print["Το υπόδειγμα είναι καλό με τη F ", Length[h], " φορές"];
Print["Αριθμός υποφαινομενικών παλινδρομήσεων με F ", Length[spurious3]];
Print["Αριθμός υποφαινομενικών παλινδρομήσεων με F και R^2 > 0.7 ",
  Length[spurious4]];

```

στ) Υποφαινομενικές συσχετίσεις μεταξύ δυο $I(1)$ διαδικασιών (σελ. 98)

```
Needs["Statistics`ContinuousDistributions`"];
Needs["Statistics`MultiDescriptiveStatistics`"];
Needs["Graphics`Graphics`"];
Needs["Statistics`DataManipulation`"];
n = 124; φ $\alpha$  = 1; φ $\beta$  = 1;
series1 = Table[0, {n + 1}];
series2 = Table[0, {n + 1}];
katanomi = {}; apolytikatanomi = {};
eleghos = {}; katelehg = {};
Do[
  Do[
    u = Random[NormalDistribution[0, 1]];
    series1[[i]] = φ $\alpha$ *series1[[i - 1]] + u;
    , {i, 2, n + 1}];
  Do[
    u = Random[NormalDistribution[0, 1]];
    series2[[i]] = φ $\beta$ *series2[[i - 1]] + u;
    , {i, 2, n + 1}];
  series3 = Table[series1[[i]], {i, 26, n + 1}];
  series4 = Table[series2[[i]], {i, 26, n + 1}];
  r = Correlation[series3, series4];
  katanomi = Append[katanomi, r];
  apolytikatanomi = Append[apolytikatanomi, Abs[r]];
  t = Abs[(r*Sqrt[Length[series3] - 2])/Sqrt[1 - r^2]];
  eleghos = Append[eleghos, t];
  t2 = (r*Sqrt[Length[series3] - 2])/Sqrt[1 - r^2];
  katelehg = Append[katelehg, t2];
  , {10000}];
h = Histogram[katanomi]; f = Histogram[apolytikatanomi]; m =
Histogram[katelehg];
tn $\alpha$  = Quantile[StudentTDistribution[Length[series3] - 2], 0.995] // N;
tn $\beta$  = Quantile[StudentTDistribution[Length[series3] - 2], 0.9750] // N;
tn $\gamma$  = Quantile[StudentTDistribution[Length[series3] - 2], 0.950] // N;
g $\alpha$  = Select[eleghos, #1 > tn $\alpha$  &];
g $\beta$  = Select[eleghos, #1 > tn $\beta$  &];
g $\gamma$  = Select[eleghos, #1 > tn $\gamma$  &];
Print["Η κριτική τιμή 1% είναι: ", tn $\alpha$ ];
Print["Η μηδενική υπόθεση για 1% απορρίπτεται ", Length[g $\alpha$ ], " φορές"];
Print["Η κριτική τιμή 5% είναι: ", tn $\beta$ ];
Print["Η μηδενική υπόθεση για 5% απορρίπτεται ", Length[g $\beta$ ], " φορές"];
Print["Η κριτική τιμή 10% είναι: ", tn $\gamma$ ];
Print["Η μηδενική υπόθεση για 10% απορρίπτεται ", Length[g $\gamma$ ], " φορές"];
αρνητικα = Select[katanomi, #1 < 0 &]; θετικα = Select[katanomi, #1 > 0 &];
```

```
Print["Το μέσο r είναι ", Mean[katanomi], " με τυπική απόκλιση ",  
StandardDeviation[katanomi]]];  
Print["Το μέσο απόλυτο r είναι ", Mean[apolytikatanomi],  
" με τυπική απόκλιση ", StandardDeviation[apolytikatanomi]]];  
Print["Το ελάχιστο r είναι ", Min[katanomi], " και το μέγιστο r είναι ",  
Max[katanomi]]];  
Print["Το ελάχιστο απόλυτο r είναι ", Min[apolytikatanomi],  
" και το μέγιστο απόλυτο r είναι ", Max[apolytikatanomi]]];  
Print["Υπάρχουν ", Length[αρνητικα], " αρνητικες και ", Length[θετικα],  
" θετικές συσχετίσεις"];  
Print["Η μέση τιμή της κατανομής της στατιστικής είναι ", Mean[katelehg],  
" Με τυπική απόκλιση ", StandardDeviation[katelehg]]];
```

ζ) Υποφαινομενικές συσχετίσεις μεταξύ δυο $I(2)$ διαδικασιών (σελ. 98)

```
Needs["Statistics`ContinuousDistributions`"];
Needs["Statistics`MultiDescriptiveStatistics`"];
Needs["Graphics`Graphics`"];
Needs["Statistics`DataManipulation`"];
n = 124; φ $\alpha$  = 1; φ $\beta$  = 1;
series1 = Table[0, {n + 1}];
series2 = Table[0, {n + 1}];
katanomi = {}; apolytikatanomi = {};
eleghos = {}; katelehg = {};
Do[
  Do[
    u = Random[NormalDistribution[0, 1]];
    series1[[i]] = 2*series1[[i - 1]] - series1[[i - 2]] + u;
    , {i, 3, n + 1}];
  Do[
    u = Random[NormalDistribution[0, 1]];
    series2[[i]] = 2*series2[[i - 1]] - series2[[i - 2]] + u;
    , {i, 3, n + 1}];
  series3 = Table[series1[[i]], {i, 26, n + 1}];
  series4 = Table[series2[[i]], {i, 26, n + 1}];
  r = Correlation[series3, series4];
  katanomi = Append[katanomi, r];
  apolytikatanomi = Append[apolytikatanomi, Abs[r]];
  t = Abs[(r*Sqrt[Length[series3] - 2])/Sqrt[1 - r^2]];
  eleghos = Append[eleghos, t];
  t2 = (r*Sqrt[Length[series3] - 2])/Sqrt[1 - r^2];
  katelehg = Append[katelehg, t2];
  , {10000}];
h = Histogram[katanomi]; f = Histogram[apolytikatanomi]; m =
Histogram[katelehg];
tn $\alpha$  = Quantile[StudentTDistribution[Length[series3] - 2], 0.995] // N;
tn $\beta$  = Quantile[StudentTDistribution[Length[series3] - 2], 0.9750] // N;
tn $\gamma$  = Quantile[StudentTDistribution[Length[series3] - 2], 0.950] // N;
g $\alpha$  = Select[eleghos, #1 > tn $\alpha$  &];
g $\beta$  = Select[eleghos, #1 > tn $\beta$  &];
g $\gamma$  = Select[eleghos, #1 > tn $\gamma$  &];
Print["Η κριτική τιμή 1% είναι: ", tn $\alpha$ ];
Print["Η μηδενική υπόθεση για 1% απορρίπτεται ", Length[g $\alpha$ ], " φορές"];
Print["Η κριτική τιμή 5% είναι: ", tn $\beta$ ];
Print["Η μηδενική υπόθεση για 5% απορρίπτεται ", Length[g $\beta$ ], " φορές"];
Print["Η κριτική τιμή 10% είναι: ", tn $\gamma$ ];
Print["Η μηδενική υπόθεση για 10% απορρίπτεται ", Length[g $\gamma$ ], " φορές"];
αρνητικα = Select[katanomi, #1 < 0 &]; θετικα = Select[katanomi, #1 > 0 &];
```

```
Print["Το μέσο r είναι ", Mean[katanomi], " με τυπική απόκλιση ",  
StandardDeviation[katanomi]];  
Print["Το μέσο απόλυτο r είναι ", Mean[apolytikatanomi],  
" με τυπική απόκλιση ", StandardDeviation[apolytikatanomi]];  
Print["Το ελάχιστο r είναι ", Min[katanomi], " και το μέγιστο r είναι ",  
Max[katanomi]];  
Print["Το ελάχιστο απόλυτο r είναι ", Min[apolytikatanomi],  
" και το μέγιστο απόλυτο r είναι ", Max[apolytikatanomi]];  
Print["Υπάρχουν ", Length[αρνητικα], " αρνητικες και ", Length[θετικα],  
" θετικές συσχετίσεις"];  
Print["Η μέση τιμή της κατανομής της στατιστικής είναι ", Mean[katelehg],  
" Με τυπική απόκλιση ", StandardDeviation[katelehg]];
```

η) Υποφαινομενικές συσχετίσεις μεταξύ μιας $I(1)$ διαδικασίας με μια $I(2)$ διαδικασία
(σελ. 98)

```
Needs["Statistics`ContinuousDistributions`"];
Needs["Statistics`MultiDescriptiveStatistics`"];
Needs["Graphics`Graphics`"];
Needs["Statistics`DataManipulation`"];
n = 124; φα = 1; φβ = 1;
series1 = Table[0, {n + 1}];
series2 = Table[0, {n + 1}];
katanomi = {}; apolytikatanomi = {};
eleghos = {}; katelehg = {};
Do[
  Do[
    u = Random[NormalDistribution[0, 1]];
    series1[[i]] = 2*series1[[i - 1]] - series1[[i - 2]] + u;
    , {i, 3, n + 1}];
  Do[
    u = Random[NormalDistribution[0, 1]];
    series2[[i]] = φβ*series2[[i - 1]] + u;
    , {i, 2, n + 1}];
  series3 = Table[series1[[i]], {i, 26, n + 1}];
  series4 = Table[series2[[i]], {i, 26, n + 1}];
  r = Correlation[series3, series4];
  katanomi = Append[katanomi, r];
  apolytikatanomi = Append[apolytikatanomi, Abs[r]];
  t = Abs[(r*Sqrt[Length[series3] - 2])/Sqrt[1 - r^2]];
  eleghos = Append[eleghos, t];
  t2 = (r*Sqrt[Length[series3] - 2])/Sqrt[1 - r^2];
  katelehg = Append[katelehg, t2];
  , {10000}];
h = Histogram[katanomi]; f = Histogram[apolytikatanomi]; m =
Histogram[katelehg];
tnα = Quantile[StudentTDistribution[Length[series3] - 2], 0.995] // N;
tnβ = Quantile[StudentTDistribution[Length[series3] - 2], 0.9750] // N;
tnγ = Quantile[StudentTDistribution[Length[series3] - 2], 0.950] // N;
gα = Select[eleghos, #1 > tnα &];
gβ = Select[eleghos, #1 > tnβ &];
gγ = Select[eleghos, #1 > tnγ &];
Print["Η κριτική τιμή 1% είναι: ", tnα];
Print["Η μηδενική υπόθεση για 1% απορρίπτεται ", Length[gα], " φορές"];
Print["Η κριτική τιμή 5% είναι: ", tnβ];
Print["Η μηδενική υπόθεση για 5% απορρίπτεται ", Length[gβ], " φορές"];
Print["Η κριτική τιμή 10% είναι: ", tnγ];
Print["Η μηδενική υπόθεση για 10% απορρίπτεται ", Length[gγ], " φορές"];
```

```

αρνητικα = Select[katanomi, #1 < 0 &]; θετικα = Select[katanomi, #1 > 0 &];
Print["Το μέσο r είναι ", Mean[katanomi], " με τυπική απόκλιση ",
      StandardDeviation[katanomi]];
Print["Το μέσο απόλυτο r είναι ", Mean[apolytikatanomi],
      " με τυπική απόκλιση ", StandardDeviation[apolytikatanomi]];
Print["Το ελάχιστο r είναι ", Min[katanomi], " και το μέγιστο r είναι ",
      Max[katanomi]];
Print["Το ελάχιστο απόλυτο r είναι ", Min[apolytikatanomi],
      " και το μέγιστο απόλυτο r είναι ", Max[apolytikatanomi]];
Print["Υπάρχουν ", Length[αρνητικα], " αρνητικές και ", Length[θετικα],
      " θετικές συσχετίσεις"];
Print["Η μέση τιμή της κατανομής της στατιστικής είναι ", Mean[katelehg],
      " Με τυπική απόκλιση ", StandardDeviation[katelehg]];

```


ΒΙΒΛΙΟΓΡΑΦΙΑ

- Banerjee, A., Dolado, J., Galbraith, J. W. and Hendry, D. F. (1993). *Co-integration, error correction, and the econometric analysis of non-stationary data*, Oxford University Press, Oxford.
- Box, G. E. P. and Jenkins, G. M. (1970). *Time series analysis, forecasting and control*, Holden-Day, San Francisco.
- Box, G. E. P. and Pierce, A. D. (1970). Distribution of residual autocorrelations in Autoregressive Integrated Moving Average time series models, *Journal of the American Statistical Association*, **65**, 1509-1526.
- de Jong, R. M. (2003). Logarithmic spurious regressions, *Economics Letters*, **81**, 13-21.
- Durlauf, S. N. and Phillips, P. C. B. (1988). Trends versus random walks in time series analysis, *Econometrica*, **56**, 1333-1354.
- Entorf, H. (1997). Random walks with drifts: Nonsense regression and spurious fixed-effect estimation, *Journal of Econometrics*, **80**, 287-296.
- Granger, C. W. J. and Newbold, P. (1974). Spurious regressions in econometrics, *Journal of Econometrics*, **2**, 111-120.
- Granger, C. W. J. and Engle, R. F. (1985). *Dynamic model specification with equilibrium constraints: Cointegration and error correction*, Discussion Paper no. **85-18**, University of California at San Diego, CA.
- Granger, C. W. J., Hyung, N. and Jeon, Y. (2001). Spurious regressions with stationary series, *Applied Economics*, **33**, 899-904.
- Greene, W. H. (2003). *Econometric analysis*, 5th edn, Prentice Hall.
- Harvey, A. C. (1980). On comparing regression models in levels and first differences, *International Economic review*, **21**, 707-720.
- Hendry, D. F. (1980). Econometrics – Alchemy or science?, *Economica*, **47**, 387-406.
- Kendal, M. G. (1954). *Exercises in theoretical statistics*, Griffin, London.
- Kim, T. L., Lee, Y. S. and Newbold P. (2004). Spurious regressions with stationary processes around linear trends, *Economics Letters*, **83**, 257-262.
- Lee, Y. S., Kim, T. L. and Newbold P. (2004). *Spurious nonlinear regressions in econometrics*. <http://www.swan.ac.uk/economics/dpapers/>
- Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models, *Biometrika*, **65**, 297-303.

- Maddala, G. S. (1988). *Introduction to econometrics*, Macmillan Publishing Company, New York.
- Mills, T. C. (2002). *The econometric modeling of financial time series*, 2nd edn, Cambridge University Press.
- Nelson, C. R. and Kang, H. (1984). Pitfalls in the use of time as an explanatory variable in regression, *Journal of Business and Economic Statistics*, **2**, 73-82.
- Nelson, C. R. and Plosser, C. I. (1982). Trends and random walks in macroeconomic time series, *Journal of Monetary Economics*, **10**, 139-162.
- Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics, *Journal of Econometrics*, **33**, 311-340.
- Phillips, P. C. B. (1987). Time series regression with a unit root, *Econometrica*, **55**, 277-301.
- Plosser, C. I. And Schwert, G. W. (1978). Money, Income, and Sunspots: Measuring economic relationship and the effects of differencing, *Journal of Monetary Economics*, **4**, 637-660.
- Spanos, A. (1986). *Statistical foundations of econometric modeling*, Cambridge University Press.
- White, H. (2001). *Asymptotic theory for econometricians*, Revised edn, Academic Press.
- Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time-series, *Journal of the Royal Statistical Society*, **89**, 1-64.