INTER-FACULTY MASTER PROGRAM on
COMPLEX SYSTEMS and NETWORKS
SCHOOL of MATHEMATICS
SCHOOL of BIOLOGY
SCHOOL of GEOLOGY
SCHOOL of ECONOMICS
ARISTOTLE UNIVERSITY of THESSALONIKI

Master Thesis

Title:

Causality networks using information measures on non-stationary time series with application to finance.

Δίκτυα αιτιότητας με μέτρα πληροφορίας σε μη-στάσιμες χρονοσειρές με εφαρμογή στα χρηματοοικονομικά.

Gravanis Marios

Supervisor: Dimitris Kugiumtzis, Professor AUTH
Co-supervisors: Ioannis Antoniou, Professor AUTH
Kostantinos Papadopoulos, Assistant Professor AUTH

Thessaloniki, December 2017

INTER-FACULTY MASTER PROGRAM on
COMPLEX SYSTEMS and NETWORKS
SCHOOL of MATHEMATICS
SCHOOL of BIOLOGY
SCHOOL of GEOLOGY
SCHOOL of ECONOMICS
ARISTOTLE UNIVERSITY of THESSALONIKI

Master Thesis

Title :

Causality networks using information measures on non-stationary time series with application to finance.

Δίκτυα αιτιότητας με μέτρα πληροφορίας σε μη-στάσιμες χρονοσειρές με εφαρμογή στα χρηματοοικονομικά.

Γραβάνης Μάριος

Επιβλέπων : Δημήτριος Κουγιουμτζής , Καθηγητής Α.Π.Θ.

Εγκρίθηκε από την Τριμελή Εξεταστική Επιτροπή την _ η Νοεμβρίου 2017.

.....	.....	.....
Δ.Κουγιουμτζής	Ι.Αντωνίου	Κ.Παπαδόπουλος
Καθηγητής Α.Π.Θ.	Καθηγητής Α.Π.Θ.	Επίκουρος Καθηγητής Α.Π.Θ

Θεσσαλονίκη, Δεκέμβριος 2017

……

Μάριος Γραβάνης

Πτυχιούχος Οικονομολόγος

Πανεπιστήμιο Ιωαννίνων

Abstract

In recent years, there has been an increasing interest in graphical models and their use for the representation of causal relationships between the individual variables of a complex, multivariate system.
In order to study the interdependence between the observed variables of a multi-variable dynamic system or stochastic process, several methods have been developed, with more important, the ones that estimate the dependence of the temporal evolution of a variable on another, a concept known as Granger causality.
In the causality analysis of a system consisting of K variables it is important to also include the interaction with the remaining $K-2$ variables in the estimation of the connection or causality from X to Y, and this is referred to as direct Granger causality.
In the context of information theory, two such methods are those given by the *Partial Transfer Entropy* statistic, which is customary to be applied to stationary time series and *Partial Transfer Entropy on Rank Vectors (PTERV)* which can be applied directly in both stationary and non-stationary time series.
When K is large, there is a need to select the most relevant of the $K-2$ variables with the driving variable X or the response variable Y, in order to have more accurate results. In this work, we focus on this selection process, evaluating some approaches in the effort to propose a new, more effective approach to the Granger causality statistic PTERV. The purpose of the thesis is to create causal networks with K nodes by appropriately applying PTERV to K non-stationary time series. In the application in finance, PTERV is applied directly to stock indices, and not their returns in order to estimate direct causal effects on the financial indices.

Key Words

Graph theory, Financial networks, Time series, Non stationarity, Granger causality, Partial Transfer Entropy, Partial Transfer Entropy on Ranked Vectors, False Detection Rate, Mutual Information, Principal component analysis, Sensitivity, Specificity, ,Accuracy measures.

Περίληψη

Τα τελευταία χρόνια, υπάρχει συνεχώς αυξανόμενο ενδιαφέρον για τα γραφικά μοντέλα και τις χρήσεις τους για την αναπαράσταση σχέσεων αιτιότητας μεταξύ των επιμέρους μεταβλητών ενός πολύπλοκου, πολυμεταβλητού συστήματος.
Για τη μελέτη της αλληλεξάρτησης μεταξύ των παρατηρούμενων μεταβλητών ενός πολύ-μεταβλητού δυναμικού συστήματος ή στοχαστικής διαδικασίας έχουν αναπτυχθεί διάφορες μέθοδοι, με πιο σημαντικές αυτές που εκτιμούν την εξάρτηση που έχει η χρονική εξέλιξη μιας μεταβλητής από κάποια άλλη, έννοια γνωστή ως αιτιότητα κατά Granger.
Στην ανάλυση αιτιότητας χρονοσειρών από K μεταβλητές είναι σημαντικό στην εκτίμηση της σύνδεσης ή αιτιότητας από τη X στη Y να συμπεριληφθεί και η αλληλεπίδραση με τις υπόλοιπες $K-2$ μεταβλητές, δηλαδή η άμεση αιτιότητα κατά Granger.
Στο πλαίσιο των μεθόδων που βασίζονται στη θεωρία πληροφορίας, δύο τέτοιες μέθοδοι είναι αυτές που δίνονται από τα στατιστικά *Partial Transfer Entropy* όπου συνηθίζεται να εφαρμόζεται σε στάσιμες χρονοσειρές και *Partial Transfer Entropy on Rank Vectors (PTERV)* όπου έχει τη δυνατότητα να εφαρμόζεται απευθείας και σε στάσιμες, αλλά και σε μη-στάσιμες χρονοσειρές. Όταν το K είναι μεγάλο, υπάρχει ανάγκη επιλογής των πιο σχετικών από τις $K-2$ μεταβλητές με τη μεταβλητή οδηγό X ή τη μεταβλητή απόκρισης Y.
Στην εργασία αυτή εστιάζουμε σε αυτή τη διαδικασία επιλογής, αξιολογώντας κάποιες προσεγγίσεις στην προσπάθεια να προτείνουμε μια νέα πιο αποτελεσματική προσέγγιση σε σχέση με το στατιστικό εκτίμησης άμεσης αιτιότητας κατά Granger PTERV. Σκοπός της εργασίας είναι η δημιουργία δικτύων αιτιότητας με K κόμβους εφαρμόζοντας κατάλληλα το PTERV σε K μη-στάσιμες χρονοσειρές.
Γίνεται εφαρμογή της PTERV απευθείας σε χρηματιστηριακούς δείκτες και όχι τις αποδόσεις τους, ωστε να εκτιμηθούν οι σχέσεις αιτιότητας των χρηματιστηριακών δεικτών.

Λέξεις κλειδιά

Θεωρία Γράφων, Χρηματοοικονομικά Δίκτυα, Χρονολογικές Σειρές, Μη Στασιμότητα, Αιτιότητα κατά Granger, Εντροπία μερικής μεταφοράς, , Εντροπία μερικής μεταφοράς σε διατεταγμένα διανύσματα, Ψευδές ποσοστό ανίχνευσης, Αμοιβαία Πληροφορία, Ανάλυση κύριων συνιστωσών, Ευαισθησία, Ειδικότητα, Μέτρα ακρίβειας.

Contents

Σύνοψη

Τα τελευταία χρόνια, παρουσιάζεται αυξανόμενο ενδιαφέρον για τη θεωρία γράφων και τις εφαρμογές της σε κάθε επιστημονικό πεδίο. Τα δίκτυα, ένας τρόπος αναπαράστασης της δομής πολύπλοκων συστημάτων χρησιμοποιώντας τη θεωρία γράφων, χρησιμοποιούνται ευρέως και με αυξανόμενη τάση. Μία από τις πολλές εφαρμογές τους είναι η οπτικοποίηση και ανάλυση των σχέσεων αιτιότητας μεταξύ των συστημάτων ή μεταξύ των συνιστωσών ενός ενιαίου συστήματος.

Για να μελετηθεί η αλληλεξάρτηση μεταξύ των παρατηρούμενων μεταβλητών ενός δυναμικού πολυμεταβλητού συστήματος , έχουν αναπτυχθεί αρκετές μέθοδοι, με πιο σημαντικές, εκείνες τις οποίες εκτιμούν την άμεση εξάρτηση της χρονικής εξέλιξης μιας μεταβλητής από την άλλη, μια έννοια γνωστή ως αιτιότητα κατά Granger [9].

Αυτές οι έννοιες έχουν επίσης επεκταθεί στο πλαίσιο της θεωρίας πληροφορίας, στο οποίο έχουν αναπτυχθεί πολλά μέτρα και προτείνονται συνεχώς νέα. Ένα πολύ γνωστό μέτρο που χρησιμοποιείται για την ανίχνευση σχέσεων αιτιότητας από τη μεταβλητή X στη μεταβλητή Y είναι η Εντροπία μεταφοράς [11], που συνηθίζεται η εφαρμογή της σε στάσιμες χρονολογικές σειρές.

Το μέτρο αυτό έχει ένα σημαντικό μειονέκτημα, δεν περιλαμβάνει την αλληλεπίδραση με τις υπόλοιπες $K-2$ μεταβλητές του συστήματος. Για το σκοπό αυτό, προτείνεται από τον Valkorin στο [14] μια πολυμεταβλητή έκδοση της μετρικής αυτής, η Partial Transfer Entropy.

Σε εφαρμογές πραγματικού κόσμου, τα περισσότερα από τα παρατηρούμενα φαινόμενα περιγράφονται από δεδομένα χρονολογικών σειρών που δεν είναι στάσιμα. Αυτή η συνειδητοποίηση, γεννά την ανάγκη για μέτρα ικανά να συλλάβουν τις σχέσεις αιτιότητας σε μη στάσιμες χρονολογικές σειρές.

Στο [17] , οι Staniek and Lehnertz προτείνουν μία διαφορετική εκδοχή της Transfer Entropy, την Symbolic Transfer Entropy (STE), η οποία βασίζεται στη Permutation Entropy των Bandt και Pompe [16]. Συγκεκριμένα, αντί για τη χρήση των τιμών κάθε χρονοσειράς για τον υπολογισμό της Εντροπίας, χρησιμοποιούν τα σύμβολα διάστασης m που δημιουργούνται από τη κατάταξη ανά m των τιμών της χρονοσειράς είτε σε αύξουσα είτε σε φθίνουσα σειρά. Λόγω της ανάγκης για την συμπερίληψη των υπόλοιπων $K-2$ μεταβλητών που προαναφέρθηκε, προτείνεται στο [20] η πολυμεταβλητή εκδοχή της μετρικής αυτής, η Partial Symbolic Transfer Entropy.

Το πρόβλημα με την *STE* βρίσκεται στο γεγονός ότι δεν είναι το ακριβές ανάλογο της *TE.* Η επισήμανση αυτή, γίνεται από τον Κουγιουμτζή στο [18] όπου προτείνεται μία μετρική η οποία αποτελεί αυτό το ακριβές ανάλογο, η *Transfer Entropy on Ranked Vectors (TERV).* Η αντίστοιχη πολυμεταβλητή εκδοχή της, η Partial Transfer Entropy on Ranked Vectors (PTERV) προτείνεται στο [19] από τον Κουγιουμτζή και αποτελεί τη βασική μετρική που χρησιμοποιούμε σε αυτή την εργασία.

Όπως αναφέραμε παραπάνω, σημαντικό στοιχείο στη μελέτη των σχέσεων αιτιότητας από μία μεταβλητή X σε μία μεταβλητή Y αποτελεί η συμπερίληψη στους υπολογισμούς των υπόλοιπων $K-2$ μεταβλητών του συστήματος. Όταν όμως μελετούμε ένα σύστημα με μεγάλο αριθμό μεταβλητών K, υπολογιστικά και εννοιολογικά προβλήματα μπορεί να προκύψουν. Για παράδειγμα, με την παρουσία περιττών μεταβλητών, η εφαρμογή της τυπικής διαδικασίας ανάλυσης στην οποία συμπεριλαμβάνουμε στους υπολογισμούς μας όλες τις υπόλοιπες μεταβλητές μπορεί να οδηγήσει σε χαμηλότερη ακρίβεια στην ανίχνευση των σωστών σχέσεων αιτιότητας.

Διάφορες προσπάθειες έχουν γίνει στο παρελθόν για τη μείωση των μεταβλητών δέσμευσης.

Οι Marinazzo et al [33] αντιμετωπίζουν αυτό το ζήτημα, με μια προσέγγιση στο πλαίσιο της θεωρίας της πληροφορίας. Στο έργο τους, επιλέγουν τις μεταβλητές με τα μεγαλύτερα επίπεδα αμοιβαίας πληροφορίας με την αντίστοιχη οδηγό μεταβλητή σε κάθε ζεύγος του συστήματος. Χρησιμοποιούν έναν αλγόριθμο ο οποίος αφού εντοπίσει τη μεταβλητή αυτή, τη προσθέτει στο υποσύνολο των μεταβλητών δέσμευσης και προχωράει βρίσκοντας τις μεταβλητές με τη μεγαλύτερη τιμή αμοιβαίας πληροφορίας ως προς το σύνολο αυτό. Η διαδικασία αυτή επαναλαμβάνεται έως ότου το υποσύνολο αυτό, φτάσει το προκαθορισμένο μέγεθος από τον ερευνητή. Στη συγκεκριμένη εργασία, χρησιμοποιήθηκε η τεχνική αυτή στον όρο δέσμευσης της μετρικής Granger causality [9] και εξάχθηκε το συμπέρασμα πως η χρήση ενός υποσύνολου των μεταβλητών του συστήματος ως μεταβλητές δέσμευσης, οδηγεί σε αποτελέσματα πολύ κοντά σε εκείνα που λαμβάνουμε χρησιμοποιώντας όλες τις $K-2$ μεταβλητές και ειδικότερα σε μικρά δείγματα.

Μια άλλη προσπάθεια μείωσης του αριθμού των μεταβλητών δέσμευσης, γίνεται από τους Zhou et al [34]. Σε αυτή τη δουλειά γίνεται μια προσπάθεια να μειωθεί η διάσταση των μεταβλητών αυτών χρησιμοποιώντας την Ανάλυση Κύριων Συνιστωσών (Principal Component Analysis, PCA) και χρησιμοποιώντας τους γραμμικούς συνδυασμούς των συντελεστών των Κύριων Συνιστωσών που αντιπροσωπεύουν το μεγαλύτερο μέρος της διακύμανσης με τις υπόλοιπες μεταβλητές του συστήματος, ως μεταβλητές ρύθμισης. Η ανάλυσή του, χρησιμοποιώντας την αιτιότητα του Granger ως μετρική για τον εντοπισμό των σχέσεων αιτιότητας, κατέληξε στο συμπέρασμα ότι η χρήση του PCA οδηγεί σε ελάχιστη απώλεια πληροφορίας που δεν επηρεάζει πραγματικά την ανάλυση συνδεσιμότητας. Επίσης, αυτή η μέθοδος, ανάλογα με το μέγεθος του δείγματος, εκτελείται σε εύλογο χρονικό διάστημα καθιστώντας τη έτσι, ένα πιθανώς πολύτιμο εργαλείο για την εκτίμηση των σχέσεων αιτιότητας.

Οι δικές μας προσεγγίσεις στο πρόβλημα αυτό, είναι παρόμοιες με τις παραπάνω και χρησιμοποιούνται σε συνδυασμό με τις μετρικές *PTERV* και *PTE,* οι οποίες προέρχονται από τη θεωρία πληροφορίας και ο υπολογισμός τους βασίζεται σε όρους Εντροπίας.

Συγκεκριμένα, για να ορίσουμε τα υποσύνολα των μεταβλητών δέσμευσης, χρησιμοποιούμε στη πρώτη μας προσπάθεια την αμοιβαία πληροφορία. Θα δοκιμάσουμε να χρησιμοποιήσουμε μία μονό μεταβλητή δέσμευσης, τη μεταβλητή αυτή η οποία παρουσιάζει τη μεγαλύτερη τιμή αμοιβαίας πληροφορίας με την οδηγό μεταβλητή, σε κάθε ζεύγος μεταβλητών αντίστοιχα.

Στη δεύτερη μας προσπάθεια, θα χρησιμοποιήσουμε την τεχνική της PCA, και θα ορίσουμε το υποσύνολο μεταβλητών δέσμευσης ως τον γραμμικό συνδυασμό των συντελεστών (coefficients) της πρώτης κύριας συνιστώσας (PC1) με τις αντίστοιχες μεταβλητές του συστήματος.

Οι υπολογισμοί αυτοί θα πραγματοποιηθούν και σε προσομοιωμένα συστήματα αλλά και σε πραγματικές μη στάσιμες χρηματοοικονομικές χρονοσειρές.

Μετά το πέρας των υπολογισμών αυτών, συμπεραίνουμε ότι οι μέθοδοι που χρησιμοποιήσαμε για τη μείωση των μεταβλητών δέσμευσης στις συγκεκριμένες μετρικές, δίνουν παρόμοια και σε κάποιες περιπτώσεις καλύτερα αποτελέσματα από την δέσμευση σε όλες τις $K-2$ μεταβλητές.

Ένα άλλο πλεονέκτημα της δέσμευσης σε ένα μικρότερο υποσύνολο μεταβλητών, είναι ότι μειώνεται πολύ το υπολογιστικό κόστος διατηρώντας παράλληλα τη ποιότητα των αποτελεσμάτων. Τον μικρότερο χρόνο εκτέλεσης τον προσομοιώσεων είχε η μέθοδος με τη χρήση της μίας μεταβλητής που παρουσιάζει τη μεγαλύτερη τιμή αμοιβαίας πληροφορίας με την οδηγό μεταβλητή, ως υποσύνολο δέσμευσης. Αμέσως μετά στη κατάταξη βρίσκεται η μέθοδος που κάνει χρήση του PCA και τέλος η μέθοδος με τη δέσμευση σε όλο το σύνολο των υπόλοιπων μεταβλητών του συστήματος.

Στην εφαρμογή σε χρηματοοικονομικά δεδομένα, χρησιμοποιώντας δεδομένα του δείκτη MSCI για 23 ανεπτυγμένες αγορές, προσπαθήσαμε να βρούμε τη μετρική και τις τιμές παραμέτρων οι οποίες περιγράφουν καλύτερα τις σχέσεις που απαρτίζουν τη δομη του συστήματος.

Βασισμένοι στις τιμές των μέτρων ακρίβειας που παρατηρήσαμε στις προσομοιώσεις και στον αριθμό συνδέσεων των δικτύων της εφαρμογής, καταλήγουμε οτι η PTERV για μικρή διάσταση ενσωμάτωσης $m = 2$, καταφέρνει να προσεγγίσει καλύτερα τη δομή του αληθινού αυτού συστήματος.

Acknowledgements

I would like to thank my supervisor, Professor Dimitris Kugiumtzis, for the time he invested in this work and his invaluable help during the setting up of the simulations and the writing of this master thesis. Also, I would like to thank my friends and family for their continuous support.

Introduction

Dynamic networks and time series are scientific fields with a wide range of applications. Quite recently, dynamic networks built from time series have been the subject of research, notably in health sciences, environmental sciences and finance.

A very interesting field of study that combines these two concepts is focused in detecting the interdependencies between a complex system's variables. Specifically, the dependence of the temporal evolution of a variable on another, a concept called Granger causality [9]. On this context, many methods based in Information theory have been developed in order to detect causality relationships between time series. As the interactions with the rest of the variables of the system are of great importance, multivariate versions of these measures have been proposed over time.

A known issue among these multivariate measures is that when the system under study has a great number of variables, computational and conceptual problems can occur.

The aim of this work is to propose some new methods in order for these measures to be able to operate with a smaller number of conditioning variables while still returning relevant and trustworthy results.

In chapter 1, we describe the basic properties and statistics of time series, we define time series stationarity and finally we analyse some stochastic models for time series construction.
In chapter 2, we analyse the correlation measures that are used in time series analysis and we also introduce the measures of causality that are used in this work.
In chapter 3, an introduction to the tests that are used to detect significant relationships between time series is made and then the randomization test is analysed
In chapter 4, we discuss principal component analysis, a dimension reduction method that is used later in this work to reduce the dimension of the conditioning variables subset.
In chapter 5, we introduce the concept of networks and their basic definitions and characteristics that are necessary for the reader in order to understand this work.
In chapter 6, the methodology of the construction of networks derived from time series is discussed. Specifically, the concept of the False Discovery Rate (FDR), a method used to decrease the statistical errors occurred in detecting significant network connections is analysed.
In chapter 7, we describe the accuracy measures that are used in this work to evaluate the efficiency of the metrics to identify the correct network connections. Namely, we discuss specificity and sensitivity.
In chapter 8, the main subject of this work is presented. Previous attempts from the academic world are discussed and our approach is analysed.

In chapter 9, we describe the methodology of the simulations conducted, the system that is used and the results these simulations have generated.
In chapter 10, an application of our methodology to real-world financial time series is presented. We describe the data used, the procedure that was followed and we make comments on the results.
In chapter 11, we conclude with a brief discussion of the overall results and suggestions of future research on the subject.

Chapter 1. Time series

In this first chapter, we will summarize the main features and properties of the time series. We will focus mainly on the elements that are necessary for the reader to understand our work. These elements consist of the definition of the time series, how time series are formed, their basic properties and statistics and ways to measure the relationship between multiple time series.

1.1 Definition

A time series of length n consists of a set of ordered observations $x_{t=1}^{n} = \{x_1, x_{2,\dots}, x_n\}$ of a variable X. Because the values of the observed variable change with some small or large randomness or as we call it in statistical analysis, stochasticity, we consider the observed quantity as a random variable X.

Usually these numbers are taken at equally spaced time steps $t = 1,2, \dots, n$ according to a fixed sampling time. Time series are divided into discrete and continuous, as well as univariate and multivariate. Simply stated, a time series is a set of numbers describing the evolution of a variable (or variables in the case of multivariate time series) in time.

Time series are used in a broad range of scientific fields such as statistics, econometrics, signal processing, electroencephalography, weather forecasting, etc.

In Fig1.1, we present a real-world example of a univariate financial time series.

Figure 1.1 *ECB reference exchange rate, US dollar/Euro (1999-2017 daily data points). Source: European Central Bank statistical warehouse.*

1.2 Time Series Parameters

We present below, the most common parameters used in time series analysis which relate to the stochastic process from which the time series originates as well as the statistics we receive when we only have limited observations of the time series, thus, when we have a sample.

- Mean value

  Parameter: $\mu = E[X_t]$

  Sample statistic: $\bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t$

- Variance

  Parameter: $\sigma^2 = Var[X_t] = E(X_t - \mu)^2$

  Sample statistic: $s^2 = \frac{1}{n-1}\sum_{t=1}^{n}(x_t - \bar{x})^2$

- Autocovariance

  Parameter: $\gamma(\tau) = cov(x_t, x_{t+\tau}) = E(x_t - \mu)(x_{t+\tau} - \mu)$

  Sample statistic: $c(\tau) = \frac{1}{n-|\tau|}\sum_{\tau=1}^{n-|\tau|}(x_t - \bar{x})(x_{t+\tau} - \bar{x})$

- Autocorrelation

  Parameter: $\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} = \frac{\gamma(\tau)}{\sigma^2}$

  Sample statistic: $r(\tau) = \frac{c(\tau)}{c(0)} = \frac{c(\tau)}{s^2}$

  We should also mention that the autocorrelation parameter can only be defined if the time series $\{X_t\}$ is stationary.

1.3 Stationarity

Stationarity is a fundamental concept for time series analysis. In a simple way, we define as stationary, a time series that has no change in its mean value and its variance over time and from which any periodic changes have been removed [1].

Stationarity is divided into two categories, strict and weak stationarity.

A time series is said to be *strictly stationary* if the joint probability distribution of the stochastic process that produces it is unaffected by changes in time [1]. That means the following relation must hold:

$$F\left(X_{t1}, \dots, X_{t_n}\right) = F(X_{t1+\tau}, \dots, X_{tn+\tau}) \; \forall \; \tau \in \mathbb{R}$$

In practice, strict stationarity is a very difficult condition to apply. This is why in most cases we use a simpler version, the weak stationarity.

For a time series to be considered *weakly stationary* two conditions must be in effect. The mean value of the time series $\{X_t\}$ and the autocovariance of $\{X_t\}$ and $\{X_{t+\kappa}\}$ must be independent of time.

In particular, $\{X_t\}$ is weakly stationary if these two relations hold at the same time:

- $E[X_t] = \mu, \; \mu \in \mathbb{R}$
- $cov(X_t, X_{t+\tau}) = \gamma_\kappa, \; \gamma_\kappa \in \mathbb{R}$

It is easy to prove that if a time series is strictly stationary, it is automatically and weakly stationary. In real world, we often deal with time series that are non-stationary, but we are able, using various methods, to identify and remove the cause of non-stationarity.

1.4 Testing time series stationarity

As mentioned above, in real-world applications, many time series exhibit non-stationary behavior. These characteristics are very common in financial time series such as stock price indices. There are numerous approaches used to examine the stationarity of time series data. The most popular among them are the Augmented Dickey-Fuller test (ADF), the Phillips-Peron test (PP) and the Kwiatkowski, Phillips, Schmidt and Shin test (KPSS). In this work we will use the Augmented Dickey-Fuller test [23][24].

1.4.1 Augmented Dickey-Fuller test

The ADF test was developed by Dickey and Fuller and it used to detect the presence of unit roots.

A unit root test, is used to investigate whether a time series contains a unit root and thus, is non-stationary. The null hypothesis is generally defined as non-stationarity of the time series / presence of unit root and the alternative hypothesis suggests that the data under examination are stationary or trend stationary depending on the test used [22].

There are three main versions of this test, each corresponding to different type of data [23], [24]:

a) Test for a unit root

$$\Delta y_t = \varphi^* y_{t-1} + \sum_{i=1}^{p-1} \varphi_i \, \Delta y_{t-1} + u_t$$

b) Test for a unit root with drift:

$$\Delta y_t = \beta_o + \varphi^* y_{t-1} + \sum_{i=1}^{p-1} \varphi_i \, \Delta y_{t-1} + u_t$$

c) Test for a unit root with drift and deterministic trend added:

$$\Delta y_t = \beta_o + \varphi^* y_{t-1} + \sum_{i=1}^{p-1} \varphi_i \, \Delta y_{t-1} + \beta_1 t + u_t$$

Where y_t denotes the log return of the time series value (e.g. stock index) at the time period t and $\Delta y_t = y_t - y_{t-1}$. Considering the other parameters, β is the drift term, t is the linear trend term, p is the lag term and u_t is the error term.

About the test hypotheses, the null hypothesis states that the time series is non-stationary or in other words contains a unit root and the alternative hypothesis states stationarity of the time series.

$$H_0: \varphi^* = 0$$

$$H_0: \varphi^* < 0$$

The *t-statistic* $t = \frac{\varphi^*}{\sqrt{Var(\varphi^*)}}$ must be calculated in order to test if our time series contain a unit root. Then, as the classic testing procedure dictates, it will be compared to the corresponding critical value at a pre-set significance level. If the null hypothesis is rejected, we can conclude that the time series y_t doesn't contain a unit root and thus it is stationary.

Considering the choice of one of the above models, it is suggested [25] that the form of the test can be based upon the graphical inspection of the time series.

Moreover, another important aspect is the selection of the appropriate value of the lag term p. If we assign a small value, we may be lead to over reject the null hypothesis when it is true. If we assign a large value the power of the test to reject the null hypothesis may be reduced. One suggested solution to this problem, is based on information criteria such as Akaike Information Criterion (AIC) [31]. In other words, we identify the proper lag length that minimizes the information criteria. There are also other methods to determine the appropriate time lag, but we only mention AIC because this is the one that is used in this work, as described with more details in a next section, in which we describe the simulations methodology.

1.5 Stochastic models for time series construction

In this section, the most well-known stochastic processes that are used to generate time series are described. In real-world applications, which are usually of great complexity, the following simple procedures can be combined together to form a model.

In each of the procedures described below, we present first the univariate and then the multivariate versions which can be considered as generalizations of the univariate cases in that they are essentially multiple univariate processes combined.

1.5.1 White noise

White noise $WN(\mu, \sigma^2)$ is a sequence $\{e_t\}$ of independent and identically distributed random variables (iid) [2]. Since the above statement applies, they come from the same probability distribution and all are mutually independent [3] and therefore, for each of them apply the following:

- $E\{e_t\} = \mu$
- $Var\{e_t\} = \sigma^2$

Also, due to the independence of the variables, the following applies:

$$Cov(e_t, e_{t+k}) = 0 \; if \; k \neq 0$$

$$Cov(e_t, e_{t+k}) = \sigma^2 \; if \; k = 0$$

A k-dimensional vector white noise can be represented by a dimension vector k as follows:

$$E_t = \{e_{1t}, e_{2t}, \dots, e_{kt}\}$$

In this vector, each of the K elements is a white noise with an average μ and variance σ_i^2.

Also, the mean value and the variance of a K-dimensional white noise can be respectively represented by a vector of μ_i components and a diagonal matrix Σ with σ_i^2 diagonal components.

White noise is a process that occurs very often in real world time series simulations. Also, we can assume by its properties that it seems to be a weakly stationary process.

1.5.2 Random walk

Random walk is a stochastic process that comes from the white noise process.

Let's assume that we have a white noise process $\{e_t\}$ with an average μ and a variance of σ^2.

A process $\{X_i\}$ is called a random walk, if the following apply:

$$X_i = X_{i+1} + e_i \text{ and } X_1 = e_1$$

We can easily prove that for the process $\{X_i\}$ the following statements stand:

- $E[X_i] = t\mu$
- $Var[X_i] = t\sigma^2$

Because the mean value and the variance of this process are dependent on time, the random walk is not considered a stationary process [1].

A k-dimensional random walk can also be described by the same set of equations, with the difference that now we consider X_i and E_i as vectors and not as single elements.

1.5.3 Autoregressive process

An autoregressive process of order p, AR(p), is a process where the variable at time t depends on the variable at p previous times.

In the univariate case we have a time series, which is described by the equation:

$X_t = \varphi_1 X_{t-1} + \ldots + \varphi_p X_{t-p} + e_t$, *where* $e_t \sim WN(0, \sigma_{xi}^2)$ *and* $\varphi_i \in \mathbb{R}$

It can be shown that the AR (p) process is stationary, since the roots of the characteristic equation:

$\lambda^p - \varphi_1 \lambda^{p-1}$-…- $\varphi_{p-1}\lambda$-$\varphi_p = 0$ are all within the unit cycle [4].

VAR models (vector autoregressive models) are used for multivariate time series. The structure that they follow is that each variable is a linear function of past lags of itself and past lags of the other variables.

A VAR process of order p and of dimension k, can be described by the following equations:

$$X_{1,t} = \varphi_{1,0} + \varphi_{1,1} X_{1,t-1} + \ldots + \varphi_{1,p} X_{1,t-p} + e_{1,t}$$

$$\vdots$$

$$X_{k,t} = \varphi_{k,0} + \varphi_{k,1} X_{k,t-1} + \ldots + \varphi_{k,p} X_{k,t-p} + e_{k,t}$$

Where $e_{ij} \sim WN(0, \sigma^2)$ and $\varphi_{i,j} \in \mathbb{R}$, we also consider $\varphi_{i,0} = 0$

1.5.4 Moving average process

A moving average process of order q, or MA(q), is defined by the following equation [1]:

$$X_t = \theta_1 \mathcal{E}_{\tau-1} + \cdots + \theta_q \mathcal{E}_{\tau-\mathrm{q}} + \mathcal{E}_\tau$$

Where $\theta_i \in \mathbb{R}$ and $e_t \sim WN(0, \sigma_\varepsilon^2)$. Moreover, this process is weakly stationary for every value of θ_i [1].

In the multivariate case, the vector moving average process of order q, or VMA(q) is described by the following k equations:

$$X_{1,t} = \theta_{1,1} \mathcal{E}_{1,t-1} + \cdots + \theta_{1,q} \mathcal{E}_{1,t-q} + \mathcal{E}_{1,t}$$

$$\vdots$$

$$X_{k,t} = \theta_{k-1} \mathcal{E}_{k,t-1} + \cdots + \theta_{k,\mathrm{q}}, \mathcal{E}_{k,t-\mathrm{q}} + \mathcal{E}_{k,t}$$

Chapter 2. Measures of Correlation and Causality

A major part of time series analysis focuses in finding relationships between them. Meanwhile, the list of measures proposed for this purpose are constantly being renewed. These measures are categorized in linear if they only detect linear relationships and in non-linear if they have the ability to identify non-linear relationships as well.

They are also divided into direct and indirect measures. Direct measures are these that have the potential to detect and eliminate the correlations that occur due to a common influence from a third or more time series. Indirect measures do not have this ability and ultimately detect all correlations, indirect and direct.

In this section we will describe the following measures: Cross correlation (CRCO), Partial cross correlation, Cross mutual information (CMI), Granger causality index (GCI), Conditional Granger causality index (CGCI), Transfer Entropy (TE), Partial transfer entropy (PTE), Transfer entropy on ranked vectors (TERV), Partial transfer entropy on ranked vectors (PTERV).

2.1 Cross-correlation

Cross-correlation is the simplest linear correlation measure. It is the generalization of the above mentioned auto-correlation function to the multivariate case. Autocorrelation calculates the relationship of a time sequence with itself for various time delays. From the above definition we can conclude that the cross-correlation function calculates the relationship of two time series for various time delays. If the time delay is zero, then we actually get the Pearson Correlation coefficient.

The expression of the cross-correlation measure is [1]:

$$r_{xy(\tau)} = \frac{\sum_{t=1}^{n-|t|}(x_t - \bar{x})(y_{t+\tau} - \bar{y})}{\sqrt{\sum_{t=1}^{n}(x_t - \bar{x})^2 \sum_{t=1}^{n}(y_t - \bar{y})^2}}$$

Where τ denotes the time lag.

The measure of cross-correlation is relatively simple in its calculation, but it also has some drawbacks. It can only detect linear relationships between time series, and cannot separate direct from indirect correlations.

Cross-correlation's possible values range in the interval [-1,1]. The interpretation of these values is given below:

- If $r > 0$, there is a positive linear correlation between the two time series, meaning that the time series are moving similarly to each other.
- If $r < 0$, the linear correlation is negative, meaning a complete reverse relation between the values of the time series.
- If $r = 0$, there is no linear correlation between the time series.

Despite being a classic measure of correlation it can actually detect causal relations too [5], depending on the adjustment of the time lag parameter τ:

- If $r_{xy(0)} \neq 0 \ \& \ \tau = 0$, then $\{x_t\}$ and $\{y_t\}$ are instantaneously related.
- If $r_{xy(\tau)} \neq 0 \ \& \ \tau > 0$, then $\{x_t\}$ causes $\{y_t\}$.
- If $r_{xy(\tau)} \neq 0 \ \& \ \tau < 0$, then $\{y_t\}$ causes $\{x_t\}$.

2.2 Cross Mutual Information

Mutual information is an indirect, non-linear correlation measure based on the concept of Shannon's Entropy [6].

Shannon's Entropy is given from the formula:

$$H(X) = -\sum_{x} p_x(x) log_2 p_x(x)$$

and it represents the number of bits required on average to describe the variable X. In other words, it measures the uncertainty of a random variable X [7].

The joint Shannon's entropy of two random variables X and Y is a generalization of the above formula [7]:

$$H(X,Y) = -\sum_{x}\sum_{y} p_{xy}(x,y) log_2 \, p_{xy}(x,y)$$

The mutual information of the random variables X and Y can now be defined as [7]:

$$I(X,Y) = H(X) + H(Y) - H(X,Y) = \sum_{x}\sum_{y} p_{xy}(x,y) log_2 \frac{p_{xy}(x,y)}{p_x(x)p_y(y)}$$

Mutual information is a measure of how much the uncertainty of variable X has been reduced due to the knowledge of variable Y and vice versa. In other words, it calculates how much information the two variables share [7].

In terms of time series, mutual information can be expressed as:

$$I_{XY(\tau)} = \sum_{x_t} \sum_{y_{t+\tau}} p_{XY}(x_t, y_{t+\tau}) log_2 \frac{p_{XY}(x_t, y_{t+\tau})}{p_x(x_t) p_y(y_{t+\tau})}$$

where τ denotes the time lag parameter. As far as the range of values is concerned, mutual information is always a positive quantity, while it is zero if and only if the two variables are independent.

If we wish to include a third variable, or a set of variables to our calculations, the Conditional mutual information between X and Y conditioning on Z can be calculated using the following formula [8]:

$$I(X;Y|Z) = \sum_{x,y,z} p(x,y,z)\, log \frac{p(x,y|z)}{p(x|z)p(y|z)}$$

2.3 Granger causality index

The Granger causality index is a linear measure proposed by Granger in 1969 [9]. The idea behind this measure is extremely simple: We are looking at whether we can have a better prediction of the future values of a time series $\{Y_t\}$ if we also include in our model the time series $\{X_t\}$, compared to only including exclusively previous values of $\{Y_t\}$. In the case that the prediction is improved, we assume that the time series $\{X_t\}$ affects $\{Y_t\}$.

This measure can be described in terms of mathematics in a very simple manner:

Suppose we have two time series, $\{X_t\}$ and $\{Y_t\}$. We construct two models to predict $\{Y_t\}$, one of which contains only past values of $\{Y_t\}$ and hence it can be described by the following expression:

$$y_t = \sum_{i=1}^{k} a_i y_{t-i} + e_t$$

The second model also contains past values of the time series $\{X_t\}$ and can be formed as:

$$y_t = \sum_{i=1}^{k} a_i y_{t-i} + \sum_{i=1}^{k} b_i x_{t-i} + e'_t$$

The coefficients a_i, b_i that appear in the two above models are estimated by a method that minimizes the dispersion of the terms e_t and e_t'. Such methods are the maximum likelihood method and the least squares method.

The answer to whether or not the time series $\{X_t\}$ affects $\{Y_t\}$ will be given to us by the variances of the terms e_t and e_t' [9]. For this purpose, the two variances are compared. If the second term shows less variance, this means that we make better predictions if we also include the past values of $\{X_t\}$ in the model and therefore we assume that the time series $\{X_t\}$ affects $\{Y_t\}$. The formula from which we get the Granger index is as follows [10]:

$$GCI_{X\to Y} = \ln \frac{Var_{(e_t)}}{Var_{(e_t')}}$$

For this measure to be successfully applied, weak stationarity is required. This issue can be addressed by applying the measure to the parts of the time series that are considered to satisfy this property. It is also obvious, that the influence of other variables in the outcome is ignored. Finally, although this measure was originally developed to identify linear relationships, several methods have been developed recently based on Granger's philosophy of detecting nonlinear relationships.

2.4 Conditional Granger Causality Index

This measure constitutes an enhanced version of the Granger causality index in sense that it is able to detect and eliminate indirect relationships. There is exactly the same logic in the methodology of its calculations, except that the new term $\{Z_t\}$ that accounts for the possible influence of all other variables in the system on $\{Y_t\}$ is now added to both models. If we assume that our system consists of a total of K time series, including $\{X_t\}$ and $\{Y_t\}$, the new term regards the multivariate time series:

$$Z_t = \{z_{1,t}, \dots, z_{K-2,t}\}$$

The corresponding models for calculating the index will be configured as follows.

The first model for the prediction of $\{Y_t\}$ now becomes:

$$y_t = \sum_{i=1}^{k} a_i y_{t-i} + \sum_{i=1}^{k} C_i Z_i + e_t$$

The second model that contains $\{X_t\}$ becomes:

$$y_t = \sum_{i=1}^{k} a_i y_{t-i} + \sum_{i=1}^{k} b_i x_{t-i} + \sum_{i=1}^{k} C_i Z_i + e_t'$$

The calculation of the index and the conclusions drawn are similar to the simple version of GCI. We calculate the dispersions of the two terms e_t, e_t' and if the dispersion in the second model is less than the first, then we assume that the time series $\{X_t\}$ affects $\{Y_t\}$. The formula from which we get the CGCI value is the following [10]:

$$CGCI_{X\to Y} = \ln \frac{Var_{(e^t)}}{Var_{(e^{t\prime})}}$$

2.5 Transfer Entropy

Transfer entropy is another measure that is based on entropy. It was first proposed by Schreiber in [10]. It captures the effect that $\{X_t\}$ exerts on $\{Y_t\}$, at T time steps ahead, while accounting for the current state of $\{Y_t\}$. In order to quantify it we have to construct three vectors. The first two vectors contain current and past values of $\{X_t\}$ and $\{Y_t\}$.

$$\boldsymbol{x_t} = \left[x_t, x_{t-\tau_\chi}, \dots, x_{t-(m_x-1)\tau\chi}\right]$$

$$\boldsymbol{y_t} = \left[y_t, y_{t-\tau_y}, \dots, y_{t-(m_y-1)\tau y}\right]$$

And the third, is the future response vector of dimension T,

$$y_t^T = [y_{t+1}, y_{t+2}, \dots, y_{t+T}]$$

Where m_x and m_y are the embedding parameters, τ_x and τ_y denote the time lag, and T is the time step ahead. An intuitive way to express transfer entropy is given below [12]:

$$TE_{X\to Y} = H(y_t^T|\boldsymbol{x_t}) - H(y_t^T|\boldsymbol{x_t}, \boldsymbol{y_t})$$

This can be translated as the number of bits required on average to describe y_{t+T} while $\boldsymbol{x_t}$ is known, less the number of bits required on average to describe y_{t+T} while both $\boldsymbol{x_t}$ and $\boldsymbol{y_t}$ are known. Another representation of transfer entropy that resembles the mutual information formula is the following [11]:

$$TE_{X\to Y} = \sum_{y_{t+T}} p(y_t^T, \boldsymbol{x_t}, \boldsymbol{y_t}) log_2 \frac{p(y_t^T|\boldsymbol{x_t}, \boldsymbol{y_t})}{p((y_t^T|\boldsymbol{y_t}))}$$

As far as the method of estimation is concerned, discretization of the observed variables $\boldsymbol{x_t}$ and $\boldsymbol{y_t}$ can be assumed for the discrete variable x, where the Shannon entropy sum is over the possible bins of x and $p(x)$ is the probability mass function of x [18].

Although more recent research [15] suggests that the use of the k-nearest neighbor's' method may be more appropriate when dealing with systems that contain a large number of variables. The use of the k- nearest neighbor's approach is described in [13].

2.6 Partial Transfer Entropy

This measure was proposed by Valkorin [14] and it is basically an extension of transfer entropy that enables it to be able to identify direct relations among time series. Its approach in eliminating indirect relations is similar to the one used in CGCI in the sense that a vector $Z_t = \{z_{1,t}, \dots, z_{K-2,t}\}$, containing every other variable in the system, is used to nullify the effect of other variables on $\{X_t\}$ and $\{Y_t\}$. An intuitive way to express partial transfer entropy is the following [12]:

$$PTE_{X \to Y|Z} = H(y_t^T | \boldsymbol{x_t}, \boldsymbol{z_t}) - H(y_t^T | \boldsymbol{x_t}, \boldsymbol{y_t}, \boldsymbol{z_t})$$

The method used by Valkorin in order to estimate the entropies is the correlation sums, but as mentioned above, the use of the k-nearest neighbor's' method, may be more appropriate if certain conditions are met.

2.7 Symbolic Transfer Entropy

Bandt and Pompe in [16] suggest a different discretization method that produces far less bins for the high dimensional variables. This is possible by the rank ordering of the components of vector variables [18]. By substituting the sample vectors with rank vectors in Shannon Entropy we can get Bandt's and Pompe's Permutation Entropy [16]:

$$H(n) = -\sum p(\pi) \log p(\pi)$$

where the sum runs over all $m!$ permutations π of order m.

The ranked vectors are formed as following [18]: For each point $\boldsymbol{y_t}$, the ranks of its components, either in ascending or descending order, form a rank vector

$\widehat{y_t} = [r_{t,1}, r_{t,2}, \dots, r_{t,m}]$ where $r_{t,j} \in \{1,2, \dots, m\}$ for $j = \{1,2, \dots, m\}$ is the rank order of the component $y_{t-(j-1)\tau}$. If two components of $\boldsymbol{y_t}$ are equal, the smallest rank is assigned to the component appearing first in $\boldsymbol{y_t}$.

In [17], the same conversion is also proposed for Transfer Entropy. The arguments in the equation of TE are modified as follows:

The sample vectors $\boldsymbol{x_t}$ and $\boldsymbol{y_t}$ are replaced by the rank vectors $\widehat{y_t}$ and $\widehat{x_t}$ respectively and the future response vector y_{t+T} is replaced by the future response rank vector $\widehat{y}_{t+T}$.

The result, called Symbolic Transfer Entropy, can be defined as:

$$STE_{X\to Y} = I(\widehat{y}_{t+T}; \widehat{x_t} | \widehat{y_t}) = -H(\widehat{y}_{t+T}, \widehat{x_t}, \widehat{y_t}) + H(\widehat{x_t}, \widehat{y_t}) + H(\widehat{y}_{t+T}, \widehat{y_t}) - H(\widehat{y_t})$$

2.8 Partial Symbolic Transfer Entropy

This measure was proposed in [20] and it is basically an extension of symbolic transfer entropy that enables it to include the effect of the current state of Z on the future of the response Y and the current state of X. In a similar way as the abovementioned Partial Transfer Entropy, we can define PSTE as:

$$\begin{aligned} PSTE_{X\to Y} &= I(\widehat{y}_{t+T}; \widehat{x_t} | \widehat{y_t}, \widehat{z_t}) \\ &= -H(\widehat{y}_{t+T}, \widehat{x_t}, \widehat{y_t}, \widehat{z_t}) + H(\widehat{x_t}, \widehat{y_t}, \widehat{z_t}) + H(\widehat{y}_{t+T}, \widehat{y_t}, \widehat{z_t}) - H(\widehat{y_t}, \widehat{z_t}) \end{aligned}$$

2.9 Transfer Entropy on Ranked Vectors

Transfer entropy on ranked vectors (TERV) is a correction of STE proposed by Kugiumtzis in [18]. The issue with STE, is that while sample vectors $\boldsymbol{x_t}$ and $\boldsymbol{y_t}$ correspond directly to their respective ranked vectors $\widehat{x_t}$ and $\widehat{y_t}$ and preserve the vector dimension, the vector $y_t^T = [y_{t+1}, y_{t+2}, \dots, y_{t+T}]$ of dimension T is mapped to the ranked vector of $\boldsymbol{y}_{t+\tau} = [y_{t+T}, \dots, y_{t+T-(m-1)\tau}]$ which is the following vector, of dimension m :

$$\widehat{\boldsymbol{y}}_{t+T} = [r_{t+T,1}, \dots, r_{t+T,m}]$$

This difference in the dimension of the sample and the respective ranked vector has implications in the computation of entropy terms [18]. For the future response sample vector of TE, y_t^T, TERV assigns the future response rank vector, of dimension T:

$$\widehat{\boldsymbol{y}}_t{}^T = [r_{t,m+1}, \dots, r_{t,m+T}]$$

Thus, Transfer entropy on ranked vectors can be defined as following [18]:

$$TERV_{X\to Y} = I(\hat{y}_t^T; \hat{x}_t|\hat{y}_t) = -H(\hat{y}_t^T, \hat{x}_t, \hat{y}_t) + H(\hat{x}_t, \hat{y}_t) + H(\hat{y}_t^T, \hat{y}_t) - H(\hat{y}_t)$$

2.10 Partial Transfer Entropy on Ranked Vectors

If we suppose that we have a system containing K multivariate time series, we can extend TERV to be able to capture the effect of the current state of the $K-2$ remaining variables on the future of the response variable y_t and the current state of x_t [19].

If we denote $Z_t = \{z_{1,t}, \dots, z_{K-2,t}\}$ as a vector containing every other variable in the system, we can define PTERV in a similar way as PTE and PSTE as following [19]:

$$PTERV_{X\to Y} = I(\hat{y}_t^T; \hat{x}_t|\hat{y}_t, \hat{z}_t) = -H(\hat{y}_t^T, \hat{x}_t, \hat{y}_t, \hat{z}_t) + H(\hat{x}_t, \hat{y}_t, \hat{z}_t) + H(\hat{y}_t^T, \hat{y}_t, \hat{z}_t) - H(\hat{y}_t, \hat{z}_t)$$

For PTERV, the possible rank permutations of Z_t are $(m!)^{K-2}$, which clearly shows that the demand for largest time series lengths increases with both K and m [19].

Chapter 3. Testing for relationships between time series

In this chapter we describe the ways in which we can check whether two time series have a relationship of correlation or causality. Initially, we summarize all kinds of tests that can be used to separate whether the connection between the time series is important or not. Then, we proceed by analyzing the test we used in this work.

3.1 The different types of tests

The simplest way to define the validity of a connection is by setting a cut-off threshold. If the measure used to detect the connection has a value greater than that, we think it is important and we consider it significant, otherwise we reject it. This way, though, is extremely simple to implement, is arbitrary and is clearly based on the researcher's experience.

Another way of separating the significant from the non-significant links, is *parametric tests*. These tests require specific assumptions about the distribution of the population so that they can be implemented. After testing these assumptions, we then define the two hypotheses:

H_0: the connection is not significant

H_1: the connection is significant

We then calculate a statistic metric from the data and then we check the location in which this statistic is found in relation to the distribution it is supposed to follow. If it is at the edges of the distribution, we consider it to be an extreme value and therefore the zero hypothesis is rejected. If the zero hypothesis was valid, our statistic would appear in the distribution as a standard value.

When we collect data and we can't be certain about the distribution that they follow, we resort to *non-parametric tests*. This type of tests, have a similar methodology to the parametric ones, except that they do not assume any statistical assumptions and do not require normality in the distributions of variables.

Randomization tests differ from the previous two types in the procedure that they follow. In order to decide whether the link is important or not, they compare the statistic formed by the measure of the two time series, with a distribution that is randomly formed by the same measure. In order to obtain this distribution, we retain one time series unchanged in each iteration, while we shift the second, in a process that does not alter its internal consistency. As before, if our statistics are at the edges of the distribution, we consider the link to be important.

3.2 The randomization test

The randomization test can be applied to all measures used to find relationships between time series. The idea is this: We initially take the measure value from the original time series. Then moving continuously one of our time series, while keeping the second one intact, we take the value of the measure for the new couples that arise and shape the distribution of these measures. Because these shifted time series have been randomly generated, if our original pair actually had a relationship, the value of the measure would be an extreme value in relation to all the rest, and therefore we expect it to appear at the edges of the distribution. If this happens, we consider the value of the measure and thus the relation of the time series, as significant. More specifically, this shifting procedure is as follows:

Let us assume that we have applied a measure c to our network connections and now we must examine whether the connection of the time series $\{x_t\}$ and $\{y_t\}$ is significant. In the test that we will apply, the zero hypothesis is that the value we got for the connection is due to random factors and therefore we must consider it zero.

The alternative hypothesis is obviously that this connection is significant and therefore different from zero. If we denote the value of the measure as m_{xy}^0 the two hypotheses can be defined as follows:

$$H_0: c_{xy}^0 = 0$$

$$H_1: c_{xy}^0 \neq 0$$

We will then form the randomized distribution, with the help of which we will decide whether the value of the measure is significant or not. This distribution is formed by the values of the same measure in M different pairs of time series originating from the original, after some random shifting. The new pairs of time series are produced as follows: We maintain the time series $\{x_t\}$ and move the time series $\{y_t\}$ by one random step p, different in each iteration. The resulting time series takes the following form:

$$\{y_t^i\} = \{y_{p+1}, y_{p+2}, \dots, y_n, y_1, \dots, y_p\}$$

The way in which the shift is made is such that the inner bonds and dynamics of the original time series are maintained. This is because some measures need those bonds to deliver the right results. It is also suggested that M is a large enough number so that we can get results at the desired level of significance. However, it is not necessarily good to get all the n-1 possible shifts, because then computational costs will incur, especially for a big length n.

Then we calculate the value of the measure for each of the above M couples of the form $\{x_t\}$, $\{y_t^i\}$ that will occur for each of the shifts and we denote the M + 1 results as $m_{xy}^0, m_{xy^1}^1, \dots m_{xy^M}^M$, where m_{xy}^0 is the value of the original time series couple. We sort these results in ascending order and check if the value is at the edges of the list.

The statistic we use to test our hypothesis, is the following [21]:

$$p-value=\begin{cases} 2\dfrac{rank\left(m_{xy}^{0}\right)-0.326}{M+1+0.348}, & if\ rank\left(m_{xy}^{0}\right)<\dfrac{M+1}{2} \\ 2\left(1-\dfrac{rank\left(m_{xy}^{0}\right)-0.326}{M+1+0.348}\right), & if\ rank\left(m_{xy}^{0}\right)\geq\dfrac{M+1}{2} \end{cases}$$

Then we conduct the following test: If the *p-value* is smaller than *α*, where *α* is the significance level, then the zero hypothesis is rejected and therefore the connection we have is important.

Chapter 4. Dimension reduction methods

In statistics, we refer to dimensionality reduction as the methods that help us reduce the number of random variables of our system, into a smaller number of so-called principal variables. In general, there are many processes that help us achieve a dimension reduction of our system. In this work, we focus in Principal component analysis (PCA) which is considered as the main linear technique for dimensionality reduction.

4.1 Principal Component Analysis

PCA is a multivariate dimension reduction technique that has its aim in the explanation of relationships among several difficult-to-interpret, correlated variables in terms of a few components which are uncorrelated with each other and are capable of accounting for nearly all the variation present in the observed data. PCA therefore finds a linear transformation of original variables into a new set of variables called principal components which are uncorrelated with each other and are capable of accounting for the variation of the obtained data and are derived in such a way that the first few of them can often provide enough information about the data and so the dimensionality of the problem can considerably be reduced [26].

The calculation of the principal components and the final form of the data is a simple and straightforward process that requires understanding of basic mathematical concepts. The steps followed in this procedure are described below [27]:

After the collection of our data, the first step is to subtract the mean value from them. This means subtracting the mean from each time series of our system separately. By following this procedure, we will end up with a dataset with zero mean. Next, we compute the covariance matrix, which is a square matrix of dimensions $(k \times k)$ supposing that our system is of dimension K. Because the covariance matrix is square, it is possible now to calculate the eigenvectors and eigenvalues of it. These, will help us choose the components that we find more suitable for use. This choice of components is the most important step, because the final form of our data will be derived by them.

The most important relationship between our systems' variables is given by the eigenvector with the highest eigenvalue [27]. In other words, this is the principal component of our data. If we rank the eigenvectors based on their corresponding eigenvalues, we have a ranking of components in which the higher a component is ranked, the more of the variance of our system it is able to explain. In our work, we will experiment using only the first principal component (PC1) to transform our data, as this component holds the most of the information of our system. Now, we construct a new matrix of vectors containing the eigenvectors we want to keep, placing them based on the ranking we mentioned above. In our case, this matrix will only contain the eigenvectors that were ranked first.

As a final step, we can derive our final data by multiplying the transposed vector of matrices we have created (the eigenvectors are now in rows) with the transposed matrix of our mean-adjusted data, in which each data dimension corresponds to a row.

There is also a procedure of getting the initial data back from the final dataset, but it doesn't concern this work.

Chapter 5. Networks

In this chapter we will summarize the main aspects of the networks, the categories in which they are separated and some important concepts related to them and are necessary for the understanding of our work.

5.1 Definition

A Network (or graph) is a method of representing a set of entities in which some pairs of entities are linked together. The entities that form the network are called vertices and we denote their total number with n, while the links formed between them are called edges and their total count is denoted as m. The number n is called *order* of the network while the number m is the *size* of the network. In summary we can say that a graph is an ordered pair $G\ (V, E)$, where the set E contains the network's edges and V is the set of its nodes.

5.2 Categories of Networks

Networks are divided into directed and undirected, and also in weighted and non-weighted.

In undirected networks, the edges are not oriented. This means, that the existence of the edge $A - B$ automatically results in the existence of the edge $B - A$ and the edge is represented by a straight line. In directed networks, there may be the edge $A - B$ but this doesn't necessarily means that the edge $B - A$ will exist too. To make sense of the direction of an edge in these networks, the links are represented by arrows.

In undirected networks, the edge set E in consists of bipartite sets of nodes $\{k, \lambda\}$ while in directed networks, it consists of ordered pairs $\{\kappa, \lambda\}$. A directed network is called symmetric, if for each pair $\{k, \lambda\}$ belonging to E, there is the corresponding pair $\{\lambda, \kappa\}$.

An example of a directed and an undirected network is presented in figures 3.1.

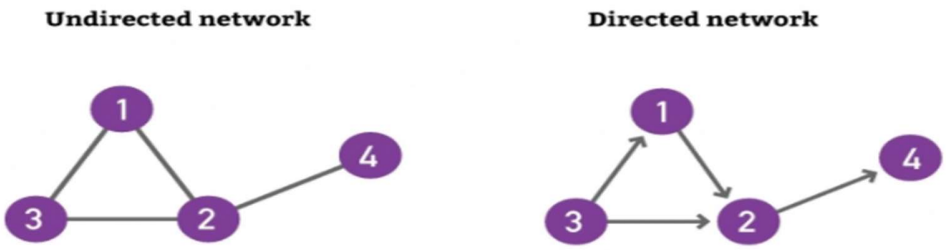

Figure 3.1 An example of directed and undirected networks. From "Network Science". Albert Laszlo Barabasi.

In weighted networks, a weight w_{ij} is assigned to each of the edges, which represents the magnitude of the corresponding link. On the other hand, non-weighted networks handle all the connections between the nodes as equal.

Figure 3.2 presents an example of a weighted and a non-weighted (binary) network:

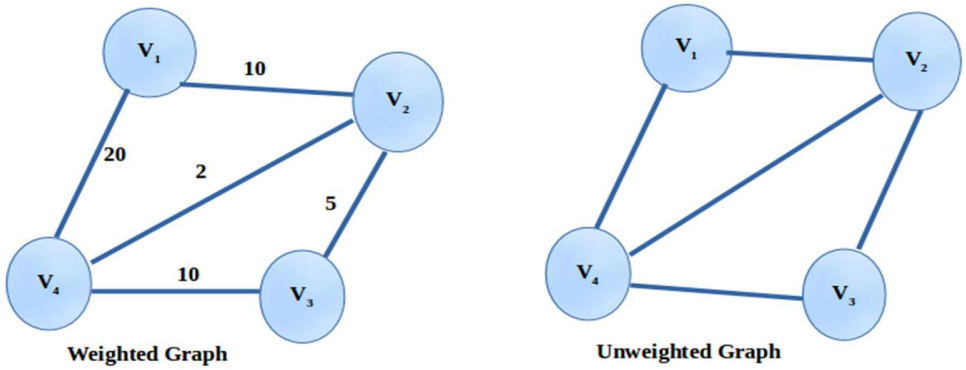

Figure 3.2. An example of a weighted and a non-weighted network.

In our work, we will use directed, non-weighted networks to study causality relationships using the information measures described in a previous chapter.

5.3 The Adjacency matrix

The adjacency matrix is a mathematical representation of the network's structure. In other words, it is a matrix that contains the list of edges between the nodes of the network. The adjacency matrix of a network of N nodes is of size $N * N$ [28].

Below there is an example of an adjacency matrix which represents a network with 4 nodes:

$$A_{ij} = \begin{matrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{matrix}$$

Figure 3.3. The structure of an adjacency matrix. From "Network Science". Albert Laszlo

About the elements of the adjacency matrix, for directed networks each cell is as follows:

$A_{ij} = 1$, if a link exists pointing from node j to node i.

$A_{ij} = 0$, if the nodes i, j are not connected.

For undirected networks, we have 2 cells corresponding to each link, for example if nodes i and j are connected the link is represented both in A_{ij} and in A_{ji} ($A_{ij} = 1$, $A_{ji} = 1$).

The adjacency matrix of weighted networks is a little different. Its elements are populated by the weight of the connection rather than a binary indicator:

$$A_{ij} = w_{ij}$$

Chapter 6. Networks from time series

As described in the previous chapter, a network is a set of entities, some of which are linked to each other by some relationships. When referring to time series networks, these entities are the time series we are concerned with, and the links between them are the relationships that may exist between each pair of these time series.

In order for a network to exist, we must have at least two time series as nodes. Then we have to decide which measure we will use to discover relationships between these time series and with which method we will test if we accept the existence or not of any relationship between them. Depending on the measure of our choice, the network will be formed as undirected, if we use a correlation measure or a directed, if the measure gives us causality relationships, for example if one time series affects another one. We also have to decide if our network will have weights, or if the edges will be binary.

The method that we chose to use in this work to test whether a link is considered significant or not, have been described in detail in Section 1.7.

6.1 False Discovery Rate (FDR)

In our attempt to identify existing connections in a network of n time series, we simultaneously perform $n\ (n-1)$ hypothesis tests. When performing such tests, there are 2 types of errors that may occur:

Type I errors, where the null hypothesis is rejected while it is true, and *type II errors* where the null hypothesis is not rejected while it is actually false. In this paragraph we will be concerned with *type I errors*. In a test, *type I errors* are controlled by the significance level α, usually equal to 0.05. This means that in a simple hypothesis test there is a 5% probability of rejecting the null hypothesis while it is true. The problem lies in the fact that if we carry out multiple tests at the same time, the likelihood of such an error increases considerably. In m simultaneous checks the probability of error is equivalent to $1-0.95^m$. Indicatively, we mention that in a network consisting only of 5 time series , in which $m = 5(5-1) = 20$ tests will be conducted, the probability of error is: $p = 1-0.95^{20} = 1-0.3585 = 0.6415 = 64.15\%$ and therefore on a relatively small network the probability of an error occurrence is greater than the probability of non-occurrence.

False discovery rate is a statistical method used in multiple hypotheses testing to correct the above problem. In a set of results in which the null hypothesis was rejected, the FDR process is designed to control the expected percentage of null hypotheses that were falsely rejected.

The concept of the false discovery rate was first proposed in 1995 by Y.Benjamini and Y.Hochberg [29] and was established as a commonly accepted way of controlling the rate of errors in hypothesis rejection.

Table 6.1. Number of errors realized in m hypotheses tests.

	True Null Hypotheses	True Alternative Hypotheses	Total
Rejected Null Hypotheses	V	S	R
Accepted Null Hypotheses	U	T	$m - R$
Total	m_0	$m - m_0$	m

Let's suppose we have multiple tests for null hypotheses from which the m_0 are true, R is the number of rejected cases (discoveries) and V is the number of the hypotheses that while true, were incorrectly rejected. In our tests, R is a random observed variable, while the variables S, T, U, V as shown in Table 3.1 are non-observable random variables. If we set Q as the percentage of false discoveries, then the false discovery rate is given by the formula:

$$FDR = E(Q) = E\left\{\frac{V}{V+S}\right\} = E(\frac{V}{R})$$

We want to keep this value below a certain threshold a. The procedure proposed by Benjamini and Hochberg is as follows:

Let $p_1, \dots, p_m$ be the observed p-values of the m hypotheses tests, sorted in ascending order. We find the highest k value for which:

$$p_k \leq \frac{k}{m} a$$

Where a denotes the significance level of the FDR procedure. We reject every null hypothesis H_i for $i = 1,2, \dots, k$. By following this procedure, $a\%$ *type I errors* can occur in our multiple tests.

Chapter 7. Accuracy measures

In this paragraph we describe the concepts of sensitivity and specificity as statistical measures used to evaluate the performance of a binary classification test.

7.1 Sensitivity and specificity

Sensitivity and specificity are statistical measures for the performance of a binary categorization test, which have their roots in diagnostic tests [30]. These two measures are closely related to Type I and II errors. For each test, there is usually an inverse relationship between these two measures, through which we try to find the optimal equilibrium point that suits our research. In some cases we need a larger value of sensitivity and a smaller value of specificity, while in other cases we want the opposite.

In the table 7.1 we present every possible outcome of a classification test and then we define the concepts of sensitivity and specificity based on these outcomes.

- A *true positive* test result is one that detects the condition when the condition is present.
- A *true negative* test result is one that does not detect the condition when the condition is absent.
- A *false positive* test result is one that detects the condition when the condition is absent.
- A *false negative* test result is one that does not detect the condition when the condition is present.

Now that we have a clear picture of the possible outcomes, we can define the 2 accuracy measures.

Sensitivity measures the ability of a test to detect the condition when the condition is present. Thus,

$$Sensitivity = TP/(TP + FN).$$

Specificity measures the ability of a test to correctly exclude the condition (not detect the condition) when the condition is absent. Thus,

$$Specificity = TN/(TN + FP).$$

Table 7.1. Possible outcomes of a classification test

	True	False
Test Positive	True Positives (TP)	False Positive (FP)
Test Negative	False Negative (FN)	True Negative (TN)

Chapter 8. Reduction of the number of conditioning variables

While searching for causality relationships between time series, it is important to take into account the conditioning effect from the rest of the variables included in the system. When the number of the remaining $K-2$ variables is large, computational and conceptual problems can occur. For example, in the presence of redundant variables, the application of the standard analysis using full conditioning, may result in a lower accuracy in the detection of the correct causalities.

Marianazzo *et al [33]* addresses this issue, from an information theory approach. In his work, he selects the k most informative variables to the driver variable using the following algorithm:

First, the Mutual Information of the driver and the other variables is calculated in order to select the first variable of the conditioning subsystem. The second variable of the subsystem is selected as the variable that jointly with the first variable, maximizes the mutual information with the driver variable. This procedure continues until the subset of the conditioning variables reaches a size of k.

Finally, by using this technique in the conditioning term of Granger causality, Marianazzo concludes that conditioning to a smaller number of variables, chosen as the most informative ones for the driver node, leads to results very close to the ones that are obtained via full conditioning.

Another attempt to reduce the number of conditioning variables, is made by *Zhou et a*l [34]. In this work, an attempt is made to reduce the dimension of the conditioning variables using Principal Components Analysis and using the linear combinations of the Principal Components that account for the most of the variance as the conditioning variables. His analysis, using conditional Granger causality, concluded that the use of PCA leads to a minimum loss of information which doesn't really affect the connectivity analysis. Also this method, depending on the dataset, runs in a reasonable time frame making it a potentially valuable tool in the estimation of casual relationships.

8.1 Our approach

In our work, we address this issue using variations of the above methods in an effort to propose an efficient approach in reducing the number of conditioning variables. Specifically, we are using two Entropy based measures. These are, *Partial Transfer Entropy* which is customary to be applied to stationary time series and *Partial Transfer Entropy on Rank Vectors* which can be applied directly in both stationary and non-stationary time series.

8.1.1 Using Mutual Information

Our first attempt is made in terms of information theory. In particular, we choose as a conditioning subset, the variable that shares the maximum value of mutual information with the driver variable in each of the variable pairs contained in the system. We decided to keep only one variable, in an attempt to minimize the dimension of the entropy terms in the computation of *PTE* and *PTERV*, and thus to be able to obtain good results even with small data samples.

The procedure for this selection of the most informative variables to the driver, is the following: First, the mutual information between every pair in the system is calculated. As a second step, these values are sorted for every individual variable in the system and the variable that maximizes the mutual information for each of the individual variables is kept. Finally, these most informative variables to the driver, are used as the conditioning variable every time their corresponding counterpart is in the driver position.

8.1.2 Using Principal Component Analysis

Our second attempt lies in the context of dimension reduction methods. Specifically, Principal components analysis. The method in use here is described below:

First, the principal components of the dataset are calculated. Then, we only keep the first principal component (PC1) as this holds the most information about our system.

Finally, we use as the conditioning variable, the linear combination between the coefficients of the first principal component and the remaining $K-2$ variables of our system.

Chapter 9. Simulations in a generated system

In this chapter, we describe the simulations performed to evaluate the time series causality measures described in Chapter 2 and to assess the capability of the new methods, used to define the set of conditioning variables in each of the metrics, to improve their results.

In order to be led in safe conclusions, we generated a system of which the links are known and the evaluation of each combination of measure and dimension reduction method was based on whether the results we get at each different simulation setup go hand in hand with the previously known connections of the system.

In each of the simulations we performed, we kept certain parameters unchanged, while changing some others in order to understand the behavior of the measures on different system settings.

9.1 Coupled Henon Maps

Coupled Henon Maps is a non-linear system that can be generated by choosing the number of time series, their length and the coupling strength between them.

In our work we generate a system with $K = 5$ variables and a coupling strength of $c = 0.4$. The equations describing this system are the following [31], [32]:

$$X_{1,t} = 1.4 - X_{1,t}^2 + 0.3X_{1,t-2}$$

$$X_{2,t} = 1.4 - 0.5c(X_{1,t-1} + X_{3,t-1}) + (1-c)X_{2,t-1}^2 + 0.3X_{2,t-2}$$

$$X_{3,t} = 1.4 - 0.5c(X_{2,t-1} + X_{4,t-1}) + (1-c)X_{3,t-1}^2 + 0.3X_{3,t-2}$$

$$X_{4,t} = 1.4 - 0.5c(X_{3,t-1} + X_{5,t-1}) + (1-c)X_{4,t-1}^2 + 0.3X_{4,t-2}$$

$$X_{5,t} = 1.4 - X_{5,t}^2 + 0.3X_{5,t-2}$$

In this system, each map is coupled to the next apart to the first and the last map, the first only drives the second and the last only drives the second last, with the first and last map uncoupled. These relationships in our 5 variable system are shown below as a network representation:

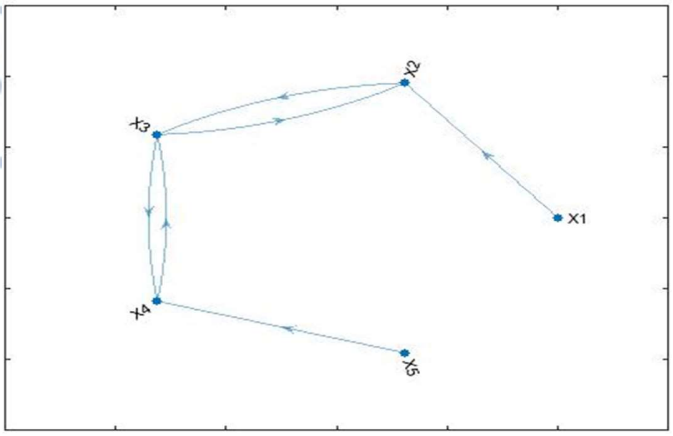

Figure 8.1. The causal relationships between a 5 variable Coupled Henon maps system.

9.2 Setting-up the simulations

Starting from the time series length, we decided to generate 3 systems of different size, $N = 1000, 4000, 12000$, so that we will be able to capture the changes in measure performance while the time series length increases.

In term of stationarity, we apply PTERV in both stationary and non-stationary time series and PTE only in the stationary data, as it is not fit to detect causality relationships in non-stationary time series [17]. Specifically, the time series generated from the set of equations describing Coupled Henon maps are classified as stationary according to the Augmented Dickey-Fuller test which rejects the null hypothesis of the existence of a unit root, for all the $K = 5$ time series of our system and for every length N

The generated coupled Henon systems for each value of N are shown in figures 9.1, 9.2 and 9.3.

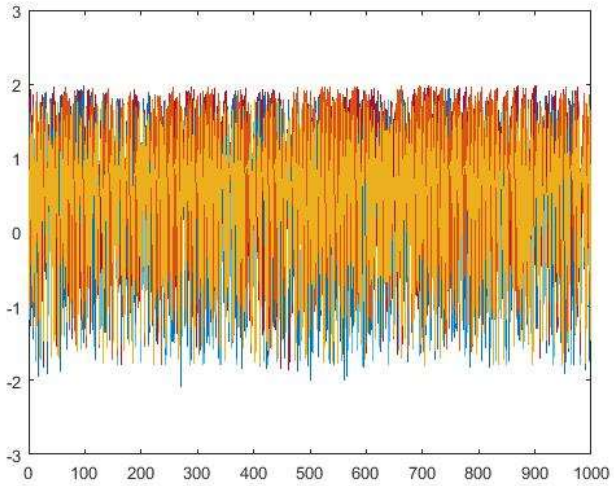

Figure 9.1. The coupled Henon maps system for N=1000.

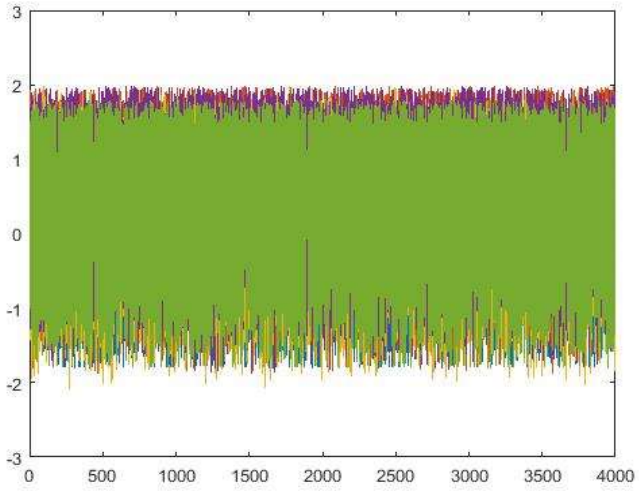

Figure 9.2. The coupled Henon maps system for N=4000.

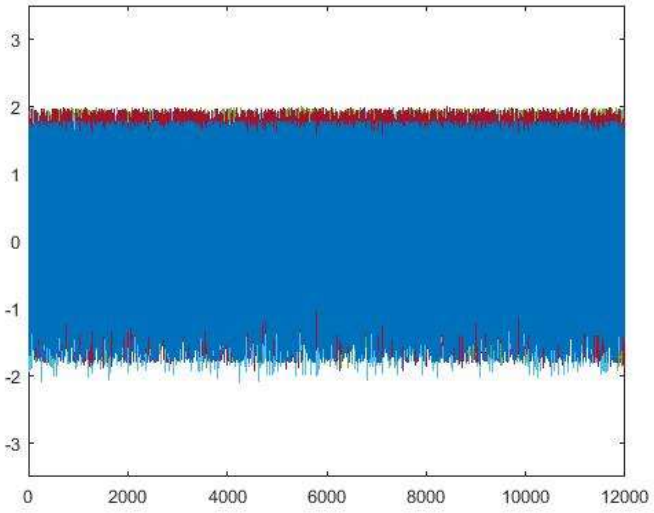

Figure 9.3. The coupled Henon maps system for N=12000.

To create non-stationary time series, we followed a method described by D.Kugiumtzis *et al [*18]. Below we describe the exact procedure.

First a Gaussian random walk time series of the same length as the original time series is generated, where the standard deviation of the random steps is the same as this of the coupled Henon maps. Then, a moving average smoothing of order 100 is applied to it. Then, this smoothed stochastic trend is added to the time series of the first variable of the coupled Henon maps. Then, we repeat this process for the 4 remaining time series. The resulting time series are the following:

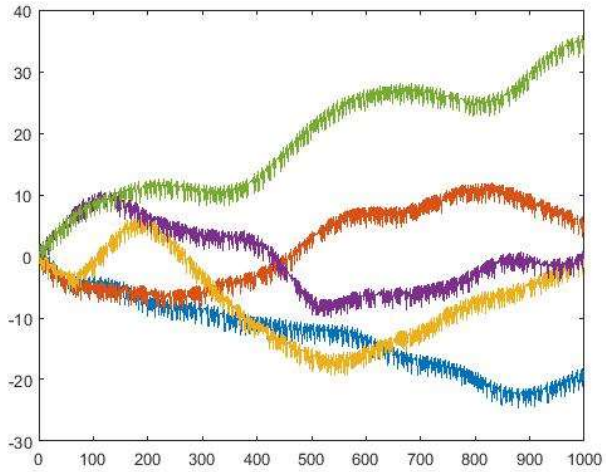

Figure 9.4. The non-stationary coupled Henon maps system for N=1000.

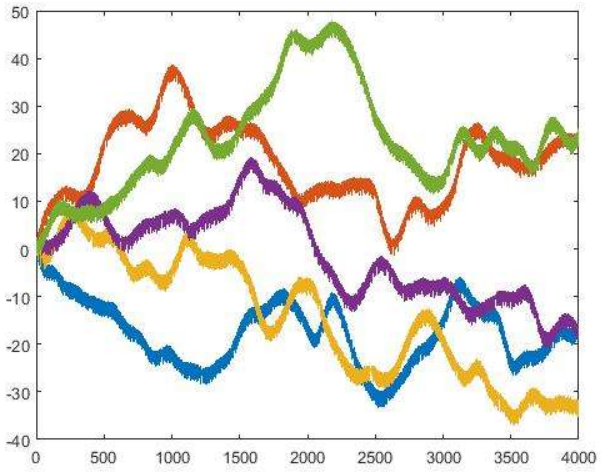

Figure 9.5. The non-stationary coupled Henon maps system for N=4000.

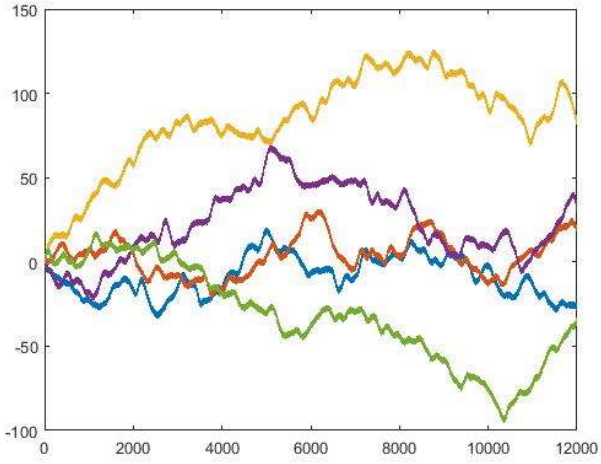

Figure 9.6. The non-stationary coupled Henon maps system for N=12000.

Indeed, the Augmented Dickey-Fuller test accepts the null hypothesis, and thus a unit root exists, making these time series non-stationary.

The coupling strength of the coupled Henon maps is set to $C = 0.4$ which is generally considered a strong coupling.

The embedding dimensions m_x, m_y, m_z are set to 2 and later at 3 for both *Partial Transfer Entropy* and *Partial Transfer Entropy on Ranked Vectors*. The time lag parameters τ_x, τ_y and τ_z are set to 1 for both the measures. The time steps ahead for the future response sample vector y_t^T of PTE and the future response rank vector $\hat{\boldsymbol{y}}_t{}^T$ of PTERV is set to $T = 1$. Finally, the number of nearest neighbors for the computation of the entropy terms in PTE is set to 10

9.3 Evaluating measure performance

For the evaluation of the measures, the randomization test described in a previous chapter is used. The number of surrogates is set to 100. Also, 100 realizations of these computations are generated, with the exception of *Partial Transfer Entropy* in the $N = 12000$ set-ups that due to computational costs the realizations are decreased to 10, in order to calculate the FDR rejections. That is, the rejections of the null hypothesis of no causality in every realization of the system.

Finally, we use the accuracy measures of specificity and sensitivity to evaluate the quantities of FDR rejections in each simulation set up. In terms of networks, sensitivity and specificity can be defined as following [33]:

Sensitivity denotes the percentage of the existing links that are detected and *Specificity* denotes the percentage of missing links correctly recognized as non-existing.

9.4 Results

In this section, we present the results of the simulations described above. The FDR rejections for each coupling of the system and also values of sensitivity and specificity are presented in the following tables.

Partial Transfer Entropy

As mentioned in the previous section, this measure is applied only in the stationary form of the data.

Table 9.1. The FDR rejections and accuracy measures for Partial Transfer Entropy for m=2 and N=1000 using all three ways of defining the conditioning variables.

$K = 5$	Normal Conditioning	Conditioning with MI	Conditioning with PCA
$X_1 \to X_2$	**55**	**10**	**91**
$X_2 \to X_1$	19	0	16
$X_1 \to X_3$	6	1	14
$X_3 \to X_1$	2	2	9
$X_1 \to X_4$	0	3	6
$X_4 \to X_1$	0	1	1
$X_1 \to X_5$	1	1	2
$X_5 \to X_1$	4	0	2
$X_2 \to X_3$	**52**	**10**	**89**
$X_3 \to X_2$	**39**	**1**	**61**
$X_2 \to X_4$	0	2	11
$X_4 \to X_2$	0	0	7
$X_2 \to X_5$	0	1	2
$X_5 \to X_2$	1	1	2
$X_3 \to X_4$	**39**	**5**	**48**
$X_4 \to X_3$	**51**	**6**	**91**
$X_3 \to X_5$	0	5	0
$X_5 \to X_3$	6	5	85
$X_4 \to X_5$	34	0	16
$X_5 \to X_4$	**55**	**10**	**91**
Specificity	1	1	0.89
Sensitivity	1	0	1

Table 9.2. The FDR rejections and accuracy measures for Partial Transfer Entropy for m=2 and N=4000 using all three ways of defining the conditioning variables.

$K = 5$	Normal Conditioning	Conditioning with MI	Conditioning with PCA
$X_1 \to X_2$	**100**	**77**	**100**
$X_2 \to X_1$	94	1	60
$X_1 \to X_3$	56	21	48
$X_3 \to X_1$	18	19	25
$X_1 \to X_4$	0	73	57
$X_4 \to X_1$	7	14	6
$X_1 \to X_5$	9	2	4
$X_5 \to X_1$	9	5	3
$X_2 \to X_3$	**93**	**78**	**100**
$X_3 \to X_2$	**85**	**6**	**94**
$X_2 \to X_4$	0	12	21
$X_4 \to X_2$	0	10	28
$X_2 \to X_5$	2	1	4
$X_5 \to X_2$	0	1	2
$X_3 \to X_4$	**90**	**34**	**88**
$X_4 \to X_3$	**96**	**47**	**100**
$X_3 \to X_5$	10	11	6
$X_5 \to X_3$	62	48	100
$X_4 \to X_5$	89	0	47
$X_5 \to X_4$	**100**	**77**	**100**
Specificity	0.95	0.89	0.89
Sensitivity	1	0.67	1

Table 9.3. The FDR rejections and accuracy measures for Partial Transfer Entropy for m=2 and N=12000 using all three ways of defining the conditioning variables.

$K = 5$	*Normal Conditioning*	*Conditioning with MI*	*Conditioning with PCA*
$X_1 \to X_2$	**10**	**9**	**10**
$X_2 \to X_1$	7	0	8
$X_1 \to X_3$	10	7	8
$X_3 \to X_1$	2	5	1
$X_1 \to X_4$	0	10	10
$X_4 \to X_1$	0	0	0
$X_1 \to X_5$	5	0	0
$X_5 \to X_1$	0	0	3
$X_2 \to X_3$	**10**	**10**	**10**
$X_3 \to X_2$	**10**	**0**	**10**
$X_2 \to X_4$	0	7	10
$X_4 \to X_2$	0	7	10
$X_2 \to X_5$	0	0	0
$X_5 \to X_2$	0	1	1
$X_3 \to X_4$	**10**	**4**	**10**
$X_4 \to X_3$	**10**	**6**	**10**
$X_3 \to X_5$	8	2	4
$X_5 \to X_3$	5	9	10
$X_4 \to X_5$	10	0	10
$X_5 \to X_4$	**10**	**9**	**10**
Specificity	0.73	0.84	0.68
Sensitivity	1	0.67	1

Table 9.4. The FDR rejections and accuracy measures for Partial Transfer Entropy for m=3 and N=1000 using all three ways of defining the conditioning variables.

$K = 5$	*Normal Conditioning*	*Conditioning with MI*	*Conditioning with PCA*
$X_1 \to X_2$	**23**	**7**	**86**
$X_2 \to X_1$	0	0	3
$X_1 \to X_3$	9	0	21
$X_3 \to X_1$	0	0	1
$X_1 \to X_4$	1	2	6
$X_4 \to X_1$	0	0	2
$X_1 \to X_5$	6	0	4
$X_5 \to X_1$	0	1	3
$X_2 \to X_3$	**20**	**7**	**76**
$X_3 \to X_2$	**13**	**3**	**56**
$X_2 \to X_4$	1	2	17
$X_4 \to X_2$	0	1	10
$X_2 \to X_5$	0	1	2
$X_5 \to X_2$	1	1	3
$X_3 \to X_4$	**17**	**4**	**44**
$X_4 \to X_3$	**20**	**4**	**86**
$X_3 \to X_5$	0	0	0
$X_5 \to X_3$	13	5	84
$X_4 \to X_5$	1	0	1
$X_5 \to X_4$	**23**	**7**	**86**
Specificity	1	1	0.89
Sensitivity	0	0	0.83

Table 9.5. The FDR rejections and accuracy measures for Partial Transfer Entropy for m=3 and N=4000 using all three ways of defining the conditioning variables.

$K = 5$	Normal Conditioning	Conditioning with MI	Conditioning with PCA
$X_1 \to X_2$	**100**	**44**	**99**
$X_2 \to X_1$	7	1	51
$X_1 \to X_3$	92	3	49
$X_3 \to X_1$	0	2	4
$X_1 \to X_4$	0	38	48
$X_4 \to X_1$	0	2	1
$X_1 \to X_5$	7	3	4
$X_5 \to X_1$	15	5	2
$X_2 \to X_3$	**87**	**43**	**98**
$X_3 \to X_2$	**86**	**4**	**85**
$X_2 \to X_4$	0	2	23
$X_4 \to X_2$	0	2	8
$X_2 \to X_5$	0	0	1
$X_5 \to X_2$	2	1	4
$X_3 \to X_4$	**91**	**14**	**87**
$X_4 \to X_3$	**87**	**35**	**99**
$X_3 \to X_5$	5	2	1
$X_5 \to X_3$	94	32	99
$X_4 \to X_5$	8	2	37
$X_5 \to X_4$	**100**	**44**	**99**
Specificity	0.94	1	0.79
Sensitivity	1	0	1

Table 9.6. The FDR rejections and accuracy measures for Partial Transfer Entropy for m=3 and N=12000 using all three ways of defining the conditioning variables.

$K = 5$	Normal Conditioning	Conditioning with MI	Conditioning with PCA
$X_1 \to X_2$	**10**	**10**	**10**
$X_2 \to X_1$	10	0	10
$X_1 \to X_3$	10	0	10
$X_3 \to X_1$	0	0	1
$X_1 \to X_4$	0	10	10
$X_4 \to X_1$	0	0	0
$X_1 \to X_5$	5	2	0
$X_5 \to X_1$	5	1	0
$X_2 \to X_3$	**10**	**10**	**10**
$X_3 \to X_2$	**10**	**1**	**10**
$X_2 \to X_4$	0	5	3
$X_4 \to X_2$	0	4	3
$X_2 \to X_5$	0	0	0
$X_5 \to X_2$	0	0	0
$X_3 \to X_4$	**10**	**6**	**10**
$X_4 \to X_3$	**10**	**3**	**10**
$X_3 \to X_5$	0	3	1
$X_5 \to X_3$	10	4	10
$X_4 \to X_5$	5	2	10
$X_5 \to X_4$	**10**	**10**	**10**
Specificity	0.79	0.84	0.73
Sensitivity	1	0.67	1

Partial Transfer Entropy on Ranked Vectors

On the contrary, PTERV is applied to both stationary and non-stationary data. Below we present the result tables first for the stationary and then for the non-stationary time series.

Table 9.7. The FDR rejections and accuracy measures for Partial Transfer Entropy on Ranked Vectors for m=2 and N=1000 using all three ways of defining the conditioning variables.

$K = 5$	*Normal Conditioning*	*Conditioning with MI*	*Conditioning with PCA*
$X_1 \to X_2$	**35**	**4**	**27**
$X_2 \to X_1$	8	0	10
$X_1 \to X_3$	0	0	1
$X_3 \to X_1$	0	1	6
$X_1 \to X_4$	3	1	4
$X_4 \to X_1$	10	2	16
$X_1 \to X_5$	1	1	2
$X_5 \to X_1$	3	1	0
$X_2 \to X_3$	**29**	**4**	**22**
$X_3 \to X_2$	**24**	**0**	**8**
$X_2 \to X_4$	9	0	1
$X_4 \to X_2$	8	0	3
$X_2 \to X_5$	5	0	4
$X_5 \to X_2$	3	0	2
$X_3 \to X_4$	**19**	**2**	**5**
$X_4 \to X_3$	**28**	**1**	**25**
$X_3 \to X_5$	5	1	1
$X_5 \to X_3$	2	2	10
$X_4 \to X_5$	7	0	4
$X_5 \to X_4$	**35**	**4**	**27**
Specificity	0.89	1	1
Sensitivity	1	0	0

Table 9.8. The FDR rejections and accuracy measures for Partial Transfer Entropy on Ranked Vectors for m=2 and N=4000 using all three ways of defining the conditioning variables.

$K = 5$	Normal Conditioning	Conditioning with MI	Conditioning with PCA
$X_1 \to X_2$	**98**	**61**	**92**
$X_2 \to X_1$	82	0	54
$X_1 \to X_3$	17	20	30
$X_3 \to X_1$	51	43	60
$X_1 \to X_4$	6	12	20
$X_4 \to X_1$	51	36	47
$X_1 \to X_5$	7	3	5
$X_5 \to X_1$	9	4	6
$X_2 \to X_3$	**86**	**52**	**81**
$X_3 \to X_2$	**86**	**0**	**39**
$X_2 \to X_4$	56	11	15
$X_4 \to X_2$	50	9	11
$X_2 \to X_5$	43	8	10
$X_5 \to X_2$	4	1	7
$X_3 \to X_4$	**93**	**27**	**23**
$X_4 \to X_3$	**86**	**24**	**82**
$X_3 \to X_5$	48	21	6
$X_5 \to X_3$	26	34	75
$X_4 \to X_5$	69	0	60
$X_5 \to X_4$	**98**	**62**	**92**
Specificity	0.79	1	0.73
Sensitivity	1	0	0.67

Table 9.9. The FDR rejections and accuracy measures for Partial Transfer Entropy on Ranked Vectors for m=2 and N=12000 using all three ways of defining the conditioning variables.

$K = 5$	Normal Conditioning	Conditioning with MI	Conditioning with PCA
$X_1 \to X_2$	**100**	**100**	**100**
$X_2 \to X_1$	100	0	99
$X_1 \to X_3$	98	95	97
$X_3 \to X_1$	100	97	99
$X_1 \to X_4$	25	84	67
$X_4 \to X_1$	98	96	78
$X_1 \to X_5$	13	5	12
$X_5 \to X_1$	14	4	7
$X_2 \to X_3$	**100**	**86**	**99**
$X_3 \to X_2$	**100**	**0**	**89**
$X_2 \to X_4$	95	56	82
$X_4 \to X_2$	96	52	86
$X_2 \to X_5$	94	11	23
$X_5 \to X_2$	20	2	10
$X_3 \to X_4$	**100**	**38**	**81**
$X_4 \to X_3$	**100**	**48**	**100**
$X_3 \to X_5$	100	42	46
$X_5 \to X_3$	99	98	100
$X_4 \to X_5$	99	0	96
$X_5 \to X_4$	**100**	**100**	**100**
Specificity	0.47	0.68	0.52
Sensitivity	1	0.67	1

Table 9.10. The FDR rejections and accuracy measures for Partial Transfer Entropy on Ranked Vectors for m=3 and N=1000 using all three ways of defining the conditioning variables.

$K = 5$	Normal Conditioning	Conditioning with MI	Conditioning with PCA
$X_1 \to X_2$	**5**	**77**	**86**
$X_2 \to X_1$	99	0	10
$X_1 \to X_3$	100	44	59
$X_3 \to X_1$	100	29	40
$X_1 \to X_4$	100	23	50
$X_4 \to X_1$	100	7	27
$X_1 \to X_5$	100	4	5
$X_5 \to X_1$	100	5	11
$X_2 \to X_3$	**99**	**63**	**42**
$X_3 \to X_2$	**100**	**0**	**16**
$X_2 \to X_4$	100	68	76
$X_4 \to X_2$	100	70	71
$X_2 \to X_5$	100	28	38
$X_5 \to X_2$	100	53	71
$X_3 \to X_4$	**100**	**31**	**20**
$X_4 \to X_3$	**99**	**26**	**14**
$X_3 \to X_5$	100	9	37
$X_5 \to X_3$	100	42	35
$X_4 \to X_5$	100	0	2
$X_5 \to X_4$	**7**	**78**	**84**
Specificity	0.26	0.79	0.84
Sensitivity	0.66	0.67	0.5

Table 9.11. The FDR rejections and accuracy measures for Partial Transfer Entropy on Ranked Vectors for m=3 and N=4000 using all three ways of defining the conditioning variables.

$K = 5$	Normal Conditioning	Conditioning with MI	Conditioning with PCA
$X_1 \to X_2$	**1**	**84**	**100**
$X_2 \to X_1$	100	0	90
$X_1 \to X_3$	100	70	54
$X_3 \to X_1$	100	58	22
$X_1 \to X_4$	100	12	42
$X_4 \to X_1$	100	4	27
$X_1 \to X_5$	100	5	11
$X_5 \to X_1$	100	5	14
$X_2 \to X_3$	**100**	**84**	**88**
$X_3 \to X_2$	**100**	**0**	**78**
$X_2 \to X_4$	100	1	14
$X_4 \to X_2$	100	2	10
$X_2 \to X_5$	100	10	17
$X_5 \to X_2$	100	35	48
$X_3 \to X_4$	**100**	**42**	**74**
$X_4 \to X_3$	**100**	**42**	**91**
$X_3 \to X_5$	100	29	15
$X_5 \to X_3$	100	77	77
$X_4 \to X_5$	100	0	89
$X_5 \to X_4$	**2**	**86**	**100**
Specificity	0.26	0.89	0.89
Sensitivity	0.66	0.67	1

Table 9.12. The FDR rejections and accuracy measures for Partial Transfer Entropy on Ranked Vectors for m=3 and N=12000 using all three ways of defining the conditioning variables.

$K = 5$	Normal Conditioning	Conditioning with MI	Conditioning with PCA
$X_1 \to X_2$	**100**	**99**	**100**
$X_2 \to X_1$	100	0	98
$X_1 \to X_3$	100	98	100
$X_3 \to X_1$	100	100	98
$X_1 \to X_4$	100	100	57
$X_4 \to X_1$	100	95	50
$X_1 \to X_5$	100	1	12
$X_5 \to X_1$	100	2	6
$X_2 \to X_3$	**98**	**97**	**100**
$X_3 \to X_2$	**100**	**0**	**99**
$X_2 \to X_4$	100	64	82
$X_4 \to X_2$	100	59	82
$X_2 \to X_5$	100	4	9
$X_5 \to X_2$	100	8	21
$X_3 \to X_4$	**100**	**42**	**97**
$X_4 \to X_3$	**100**	**53**	**98**
$X_3 \to X_5$	100	43	9
$X_5 \to X_3$	100	100	98
$X_4 \to X_5$	100	0	100
$X_5 \to X_4$	**100**	**99**	**100**
Specificity	0.26	0.63	0.47
Sensitivity	1	0.67	1

Table 9.13. The FDR rejections and accuracy measures for Partial Transfer Entropy on Ranked Vectors for m=2 and N=1000 using all three ways of defining the conditioning variables.

$K = 5$	Normal Conditioning	Conditioning with MI	Conditioning with PCA
$X_1 \to X_2$	**26**	**6**	**43**
$X_2 \to X_1$	6	0	5
$X_1 \to X_3$	1	1	6
$X_3 \to X_1$	4	0	10
$X_1 \to X_4$	3	0	2
$X_4 \to X_1$	4	0	9
$X_1 \to X_5$	0	1	4
$X_5 \to X_1$	0	1	6
$X_2 \to X_3$	**22**	**5**	**28**
$X_3 \to X_2$	**17**	**3**	**21**
$X_2 \to X_4$	5	6	23
$X_4 \to X_2$	6	5	23
$X_2 \to X_5$	3	2	9
$X_5 \to X_2$	5	0	3
$X_3 \to X_4$	**13**	**3**	20
$X_4 \to X_3$	**20**	**5**	**28**
$X_3 \to X_5$	6	1	10
$X_5 \to X_3$	0	0	5
$X_4 \to X_5$	6	0	4
$X_5 \to X_4$	**26**	**4**	**43**
Specificity	0.84	1	0.89
Sensitivity	1	0	0.83

Table 9.14. The FDR rejections and accuracy measures for Partial Transfer Entropy on Ranked Vectors for m=2 and N=4000 using all three ways of defining the conditioning variables.

$K = 5$	Normal Conditioning	Conditioning with MI	Conditioning with PCA
$X_1 \to X_2$	**97**	**61**	**99**
$X_2 \to X_1$	65	23	30
$X_1 \to X_3$	16	35	54
$X_3 \to X_1$	39	10	29
$X_1 \to X_4$	4	4	10
$X_4 \to X_1$	30	5	27
$X_1 \to X_5$	6	4	2
$X_5 \to X_1$	4	2	4
$X_2 \to X_3$	**80**	**47**	**81**
$X_3 \to X_2$	**82**	**33**	**72**
$X_2 \to X_4$	51	40	71
$X_4 \to X_2$	47	23	75
$X_2 \to X_5$	28	11	33
$X_5 \to X_2$	1	2	7
$X_3 \to X_4$	**86**	**29**	**74**
$X_4 \to X_3$	**84**	**35**	**81**
$X_3 \to X_5$	27	9	44
$X_5 \to X_3$	25	25	49
$X_4 \to X_5$	58	2	25
$X_5 \to X_4$	**97**	**49**	**99**
Specificity	**0.89**	**1**	**0.73**
Sensitivity	**1**	**0**	**0.83**

Table 9.15. The FDR rejections and accuracy measures for Partial Transfer Entropy on Ranked Vectors for m=2 and N=12000 using all three ways of defining the conditioning variables.

$K = 5$	Normal Conditioning	Conditioning with MI	Conditioning with PCA
$X_1 \to X_2$	**100**	**97**	**100**
$X_2 \to X_1$	100	51	82
$X_1 \to X_3$	97	92	98
$X_3 \to X_1$	84	21	60
$X_1 \to X_4$	26	22	37
$X_4 \to X_1$	79	25	49
$X_1 \to X_5$	7	2	5
$X_5 \to X_1$	7	1	8
$X_2 \to X_3$	**100**	**70**	**100**
$X_3 \to X_2$	**100**	**47**	**91**
$X_2 \to X_4$	95	64	97
$X_4 \to X_2$	95	60	96
$X_2 \to X_5$	82	16	48
$X_5 \to X_2$	22	13	38
$X_3 \to X_4$	**100**	**41**	**96**
$X_4 \to X_3$	**100**	**69**	**96**
$X_3 \to X_5$	90	24	66
$X_5 \to X_3$	94	69	97
$X_4 \to X_5$	99	12	72
$X_5 \to X_4$	**100**	**77**	**100**
Specificity	**0.47**	**0.78**	**0.57**
Sensitivity	**1**	**0.83**	**1**

Table 9.16. The FDR rejections and accuracy measures for Partial Transfer Entropy on Ranked Vectors for m=3 and N=1000 using all three ways of defining the conditioning variables.

$K = 5$	Normal Conditioning	Conditioning with MI	Conditioning with PCA
$X_1 \to X_2$	**6**	**53**	**66**
$X_2 \to X_1$	100	0	2
$X_1 \to X_3$	100	23	24
$X_3 \to X_1$	100	12	19
$X_1 \to X_4$	100	29	42
$X_4 \to X_1$	100	11	17
$X_1 \to X_5$	100	2	7
$X_5 \to X_1$	98	2	7
$X_2 \to X_3$	**100**	**22**	**29**
$X_3 \to X_2$	**100**	**17**	**25**
$X_2 \to X_4$	100	35	28
$X_4 \to X_2$	100	37	22
$X_2 \to X_5$	100	14	21
$X_5 \to X_2$	100	22	50
$X_3 \to X_4$	**100**	**23**	**31**
$X_4 \to X_3$	**100**	**32**	**27**
$X_3 \to X_5$	100	8	24
$X_5 \to X_3$	100	23	26
$X_4 \to X_5$	99	1	6
$X_5 \to X_4$	**4**	**45**	**67**
Specificity	**0.26**	**0.84**	**0.78**
Sensitivity	**0.66**	**0.67**	**0.5**

Table 9.17. The FDR rejections and accuracy measures for Partial Transfer Entropy on Ranked Vectors for m=3 and N=4000 using all three ways of defining the conditioning variables.

$K = 5$	Normal Conditioning	Conditioning with MI	Conditioning with PCA
$X_1 \to X_2$	**6**	**91**	**99**
$X_2 \to X_1$	100	50	88
$X_1 \to X_3$	100	78	76
$X_3 \to X_1$	100	12	18
$X_1 \to X_4$	100	13	40
$X_4 \to X_1$	100	5	16
$X_1 \to X_5$	100	4	15
$X_5 \to X_1$	100	2	18
$X_2 \to X_3$	**100**	**64**	**93**
$X_3 \to X_2$	**100**	**58**	**87**
$X_2 \to X_4$	100	52	81
$X_4 \to X_2$	100	38	83
$X_2 \to X_5$	100	7	10
$X_5 \to X_2$	100	9	42
$X_3 \to X_4$	**100**	**47**	**90**
$X_4 \to X_3$	**100**	**55**	**91**
$X_3 \to X_5$	100	14	22
$X_5 \to X_3$	100	66	77
$X_4 \to X_5$	100	17	84
$X_5 \to X_4$	**4**	**73**	**97**
Specificity	**0.26**	**1**	**0.52**
Sensitivity	**0.66**	**0**	**1**

Table 9.18. The FDR rejections and accuracy measures for Partial Transfer Entropy on Ranked Vectors for m=3 and N=12000 using all three ways of defining the conditioning variables.

$K = 5$	*Normal Conditioning*	*Conditioning with MI*	*Conditioning with PCA*
$X_1 \to X_2$	**100**	**99**	**100**
$X_2 \to X_1$	100	71	99
$X_1 \to X_3$	100	98	99
$X_3 \to X_1$	100	31	33
$X_1 \to X_4$	100	33	58
$X_4 \to X_1$	100	20	20
$X_1 \to X_5$	100	3	11
$X_5 \to X_1$	100	5	15
$X_2 \to X_3$	**98**	**79**	**100**
$X_3 \to X_2$	**100**	**57**	**99**
$X_2 \to X_4$	100	66	98
$X_4 \to X_2$	100	65	97
$X_2 \to X_5$	100	22	23
$X_5 \to X_2$	100	23	66
$X_3 \to X_4$	**100**	**45**	**98**
$X_4 \to X_3$	**100**	**74**	**98**
$X_3 \to X_5$	100	27	39
$X_5 \to X_3$	100	77	99
$X_4 \to X_5$	100	20	98
$X_5 \to X_4$	**100**	**77**	**100**
Specificity	**0.26**	**0.73**	**0.57**
Sensitivity	**1**	**0.83**	**1**

9.5 Summary

For the *Partial Transfer Entropy* applied to stationary data, the best results based on the accuracy measures used are for an embedding dimension of $m = 2$ and a sample length of 1000 (sensitivity =1, specificity =1). In the same metric settings, also the method using PCA produces very decent results with a value of sensitivity equal to 0.89 and specificity =1. For the different conditioning methods separately, the best results for full conditioning were obtained for the setup mentioned above. For the method using PCA the maximum values of the accuracy measures were obtained for m=2 and $N = 1000, 4000$ (sensitivity = 0.89, specificity =1). By conditioning to the most informative variable to the driver, we obtain the best results for $m = 2$ and $N = 4000$ (sensitivity = 0.89 and specificity = 0.67). For *PTE* applied in all these different parameter value combinations, we can observe that superior results to the full conditioning are given by the PCA method for $m = 3$ and $N = 1000$ and almost the same values were observed between the PCA conditioning and the full conditioning for $m = 3$ and $N = 12000$.

When *Partial Transfer Entropy* on Ranked Vectors is applied to stationary data, the best results are obtained for the full conditioning for $m = 2$ and $N = 1000$ (sensitivity = 0.89, specificity = 1) and for the method using PCA for $m = 3$ and $N = 4000$ with the same accuracy measure values. When the information based method is used, it produces the biggest values of the accuracy metrics for $m = 3$ and $N = 4000$ (sensitivity = 0.89, specificity = 0.67). In these simulations, the PCA based method gives superior results to the full conditioning for $m = 2, N = 12000$ and for an embedding dimension of 3 for every sample length.

Moving to the non-stationary time series in which *PTERV* is applied, the most accurate results are given for the full conditioning in a sample length of 4000 and $m = 2$ (sensitivity = 0.89, specificity = 1). We can observe that in the non-stationary data simulations, the method based on mutual information is dominating in 3 of the 6 different parameter value sets. Specifically, this method has superior results to the others for $m = 2$ and $N = 12000$ (0.78, 0.83) and for $m = 3$ and $N = 1000$ and 12000 while it comes second to the PCA based method for $N = 4000$, with a perfect sensitivity of 1 but with a zero specificity.

The superiority of the partial conditioning methods in a higher embedding dimension $(m = 3)$ and in long time series lengths $(N = 12000)$ can be well explained. As the dimension of the entropy terms in the equations of *PTERV* and *PTE* increases both with time series length N and the embedding dimension m [18], we can conclude that reducing the dimension of the system by keeping a smaller number of variables in the conditioning terms can lead us to more accurate results.

Concluding the simulations chapter, we must also mention another advantage of partial conditioning. Both methods outperformed the full conditioning one in the time needed for the simulations to run. Because all simulations were implemented in the same computer, we are able to compare the computational cost of each method. The faster

method is the one using the most informative variable to the driver, that needed for its completion about the 40% of the time that the full conditioning method took to complete. The PCA method comes second with approximately 70%. This leads us to the conclusion that partial conditioning reduces greatly the computational costs while simultaneously maintaining a decent level of accuracy and actually overpowering the full conditioning in some cases.

All simulations were implemented using MATLAB 2016a on a HP 15-ac126nv laptop with a RAM of 8GB and an Intel Core i7-4510U CPU of 2.0 GHz while MATLAB being the only active application at the moment.

Chapter 10. Application in Financial time series

10.1 Data description

The dataset is the Morgan Stanley Capital International's (MSCI) market capitalization weighted index of 23 developed markets in North America, Europe, and the Asia/Pacific Region. It is calculated with the help of the equities values of companies representative of the market structure. The dataset comprises 1300 daily values for each market in the period 5 of March 2004- 5 for March 2009, excluding weekends and holidays.

The list of countries included in this dataset is given below: Australia, New Zealand, Japan, Hong Kong, Singapore, Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, UK, Canada and USA.

10.2 Methodology

As mentioned above, the time series length is $N = 1300$. The 23 time series of the dataset are Non-stationary according to the Augmented Dickey-Fuller test. To transform them to stationary, in order to compare the measure performance, we take their first differences.

The original non-stationary system of 23 time series and the system derived from their first differences, are shown in the figures below:

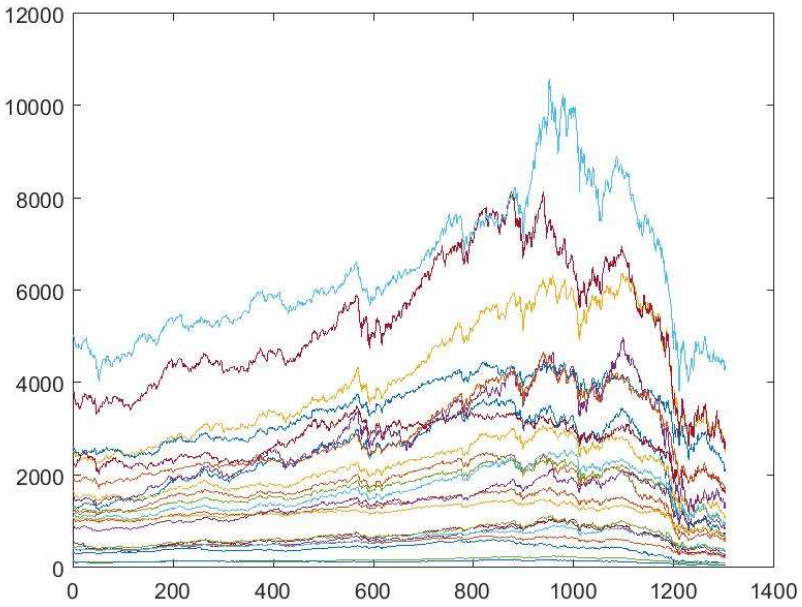

Figure 10.1. The original system of the 23 variables of the MSCI index

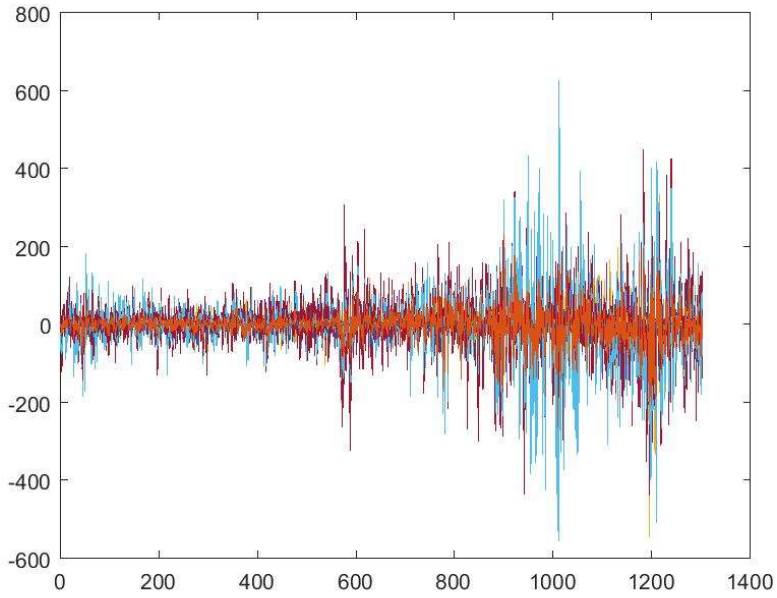

Figure 10.2. The system of the 23 variables of the MSCI index, after taking the first differences.

The 2 measures, are set in the same way as described in the simulations chapter above, with the exception that the embedding dimensions m_x, m_y, m_z are set in the interval from 2 to 6.

The time lag parameters τ_x, τ_y and τ_z are set to 1 for both the measures. The time steps ahead for the future response sample vector y_t^T of PTE and the future response rank vector $\hat{\boldsymbol{y}}_t{}^T$ of PTERV is set to $T = 1$. Finally, the number of nearest neighbors for the computation of the entropy terms in PTE is set to 10

10.3 Results

In this section, we present the networks constructed from the adjacency matrices that were produced following the method described above. First we have to comment here, that *Partial Transfer Entropy* failed to deliver any statistically significant causality relationships in all different combinations of conditioning methods and metric parameter settings and thus, networks couldn't be constructed.

On the other hand, *Partial Transfer Entropy on Ranked Vectors* was able to identify significant connections both in stationary and non-stationary data in almost every combination of parameter values. The networks constructed using *PTERV* differ in their number of connections across the different embedding dimensions and different conditioning sets.

Below these networks will be presented and some comments will be made in the following section.

Starting with the original non-stationary time series, we will compare the networks constructed using the three different methods of selecting the subset of variables that will be used in the conditioning term of *PTERV*.

For an embedding dimension of $m = 2$ the following networks are constructed.

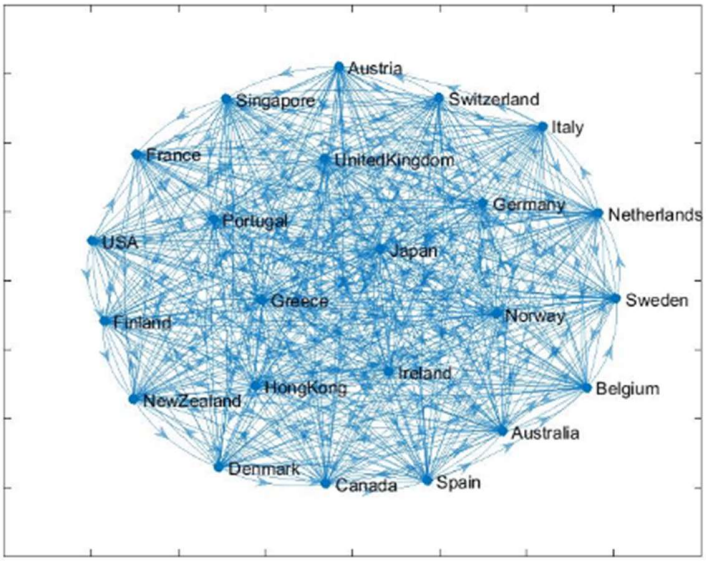

Figure 10.3. The network constructed using PTERV for m=2 and conditioning to all the remaining $K - 2$ variables in the original non-stationary data. Number of connections = 495.

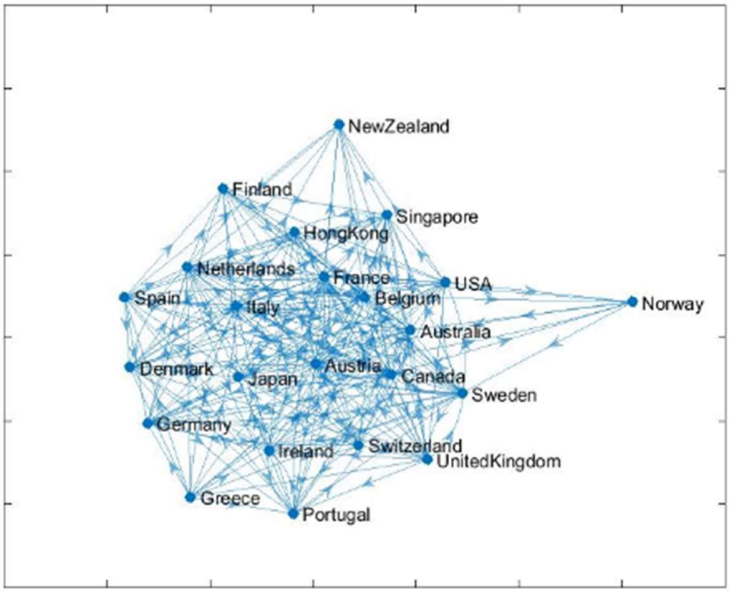

Figure 10.4. The network constructed using PTERV for m=2 and conditioning using PCA in non-stationary data. Number of connections = 231

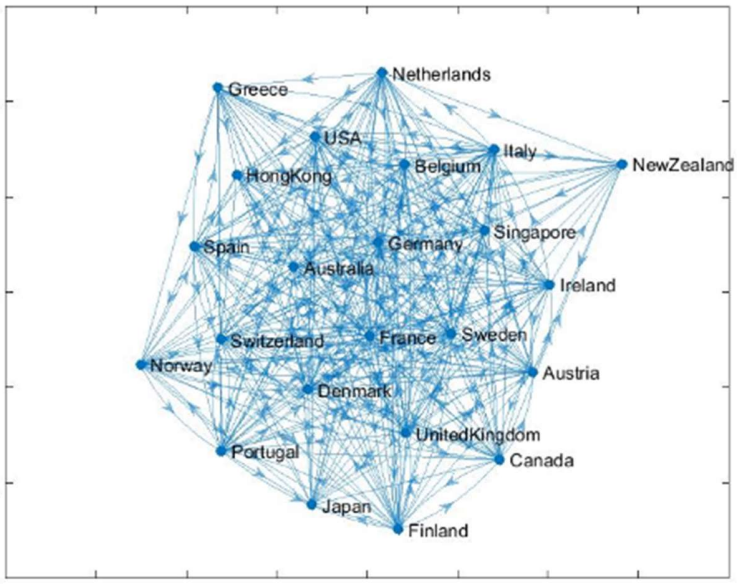

Figure 10.5. The network constructed using PTERV for m=2 and conditioning using the most informative variable to the driver in non-stationary data. Number of connections = 338

For an embedding dimension of $m = 3$ the networks are the following.

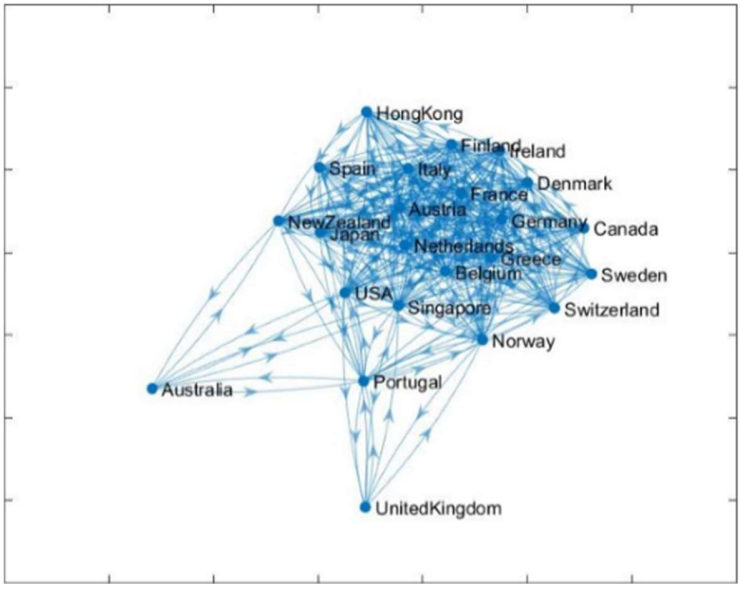

Figure 10.6. The network constructed using PTERV for m=3 and conditioning to all the remaining $K - 2$ variables in the original non-stationary data. Number of connections = 362

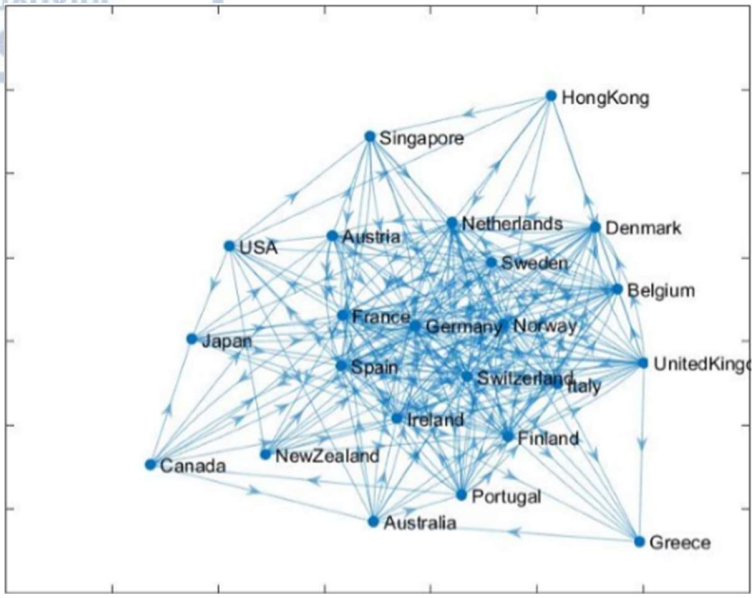

Figure 10.7. The network constructed using PTERV for m=3 and conditioning using PCA in non-stationary data. Number of connections = 259

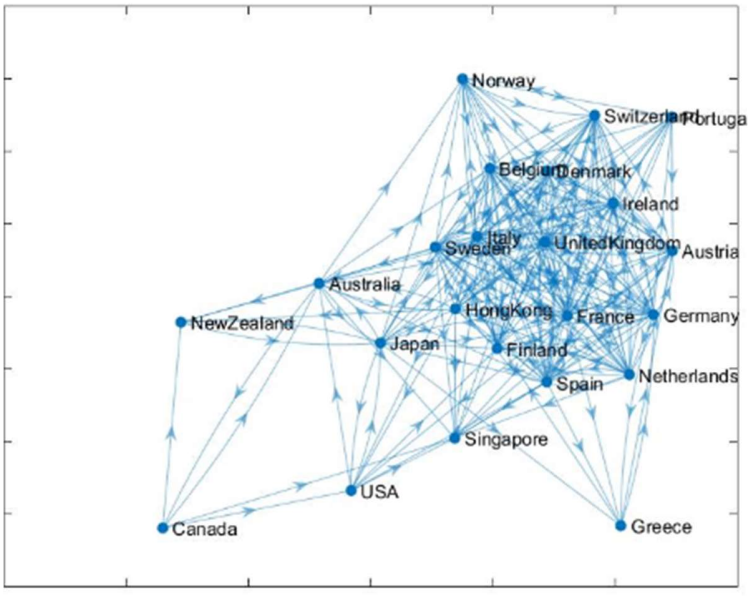

Figure 10.8. The network constructed using PTERV for m=3 and conditioning using the most informative variable to the driver in non-stationary data. Number of connections = 237

Next, the networks constructed by applying *PTERV* in the stationary data are presented.

For $m = 2$ we have extracted the following networks.

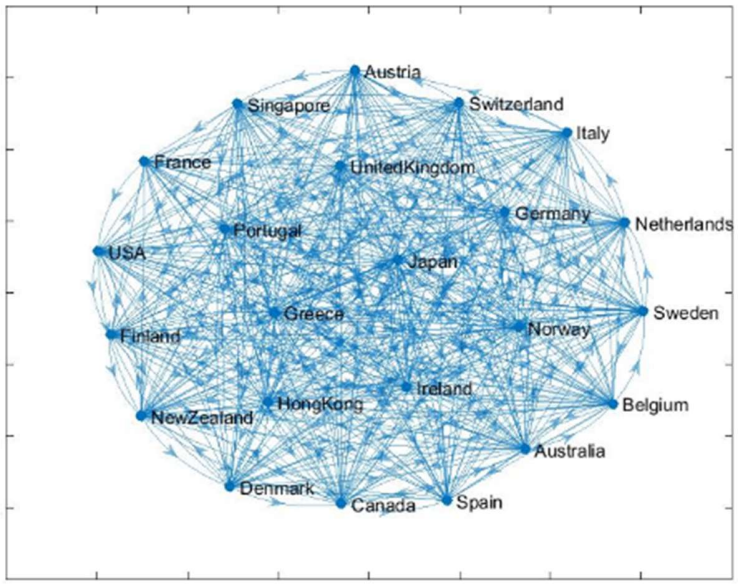

Figure 10.9. The network constructed using PTERV for m=2 and conditioning to all the remaining $K - 2$ variables in stationary data. Number of connections = 463

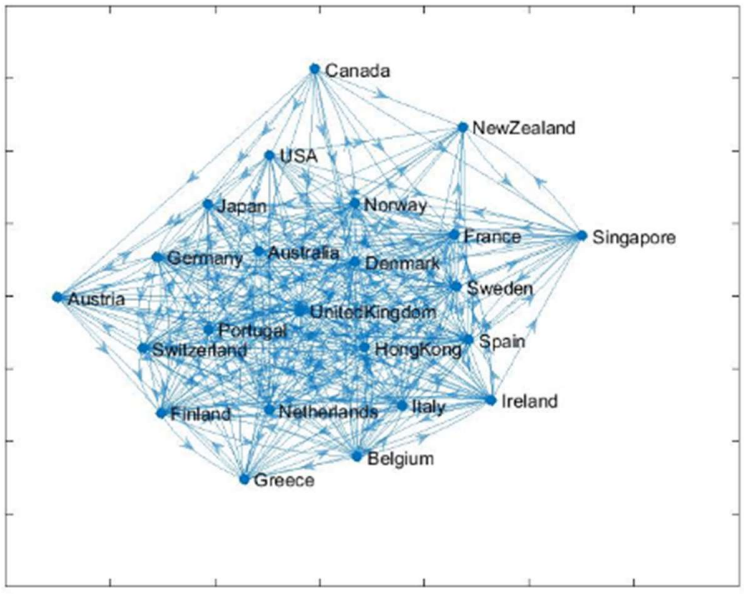

Figure 10.10. The network constructed using PTERV for m=2 and conditioning using PCA in stationary data. Number of connections = 336

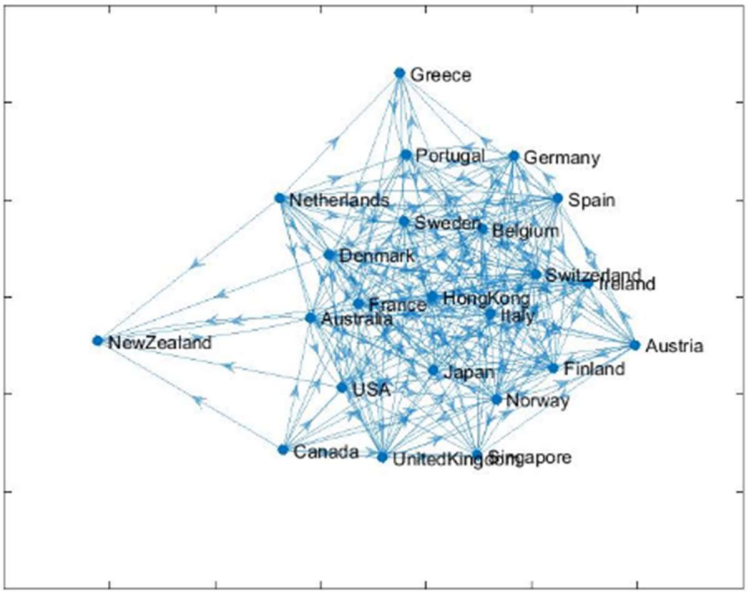

Figure 10.11. The network constructed using PTERV for m=2 and conditioning using the most informative variable to the driver in stationary data. Number of connections = 235

For $m = 3$ we have extracted the following networks.

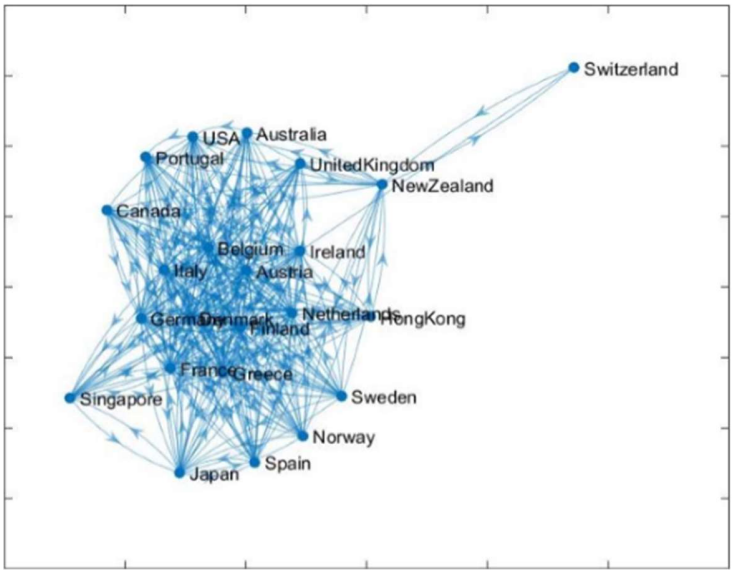

Figure 10.12. The network constructed using PTERV for m=3 and conditioning to all the remaining $K - 2$ variables in stationary data. Number of connections = 343

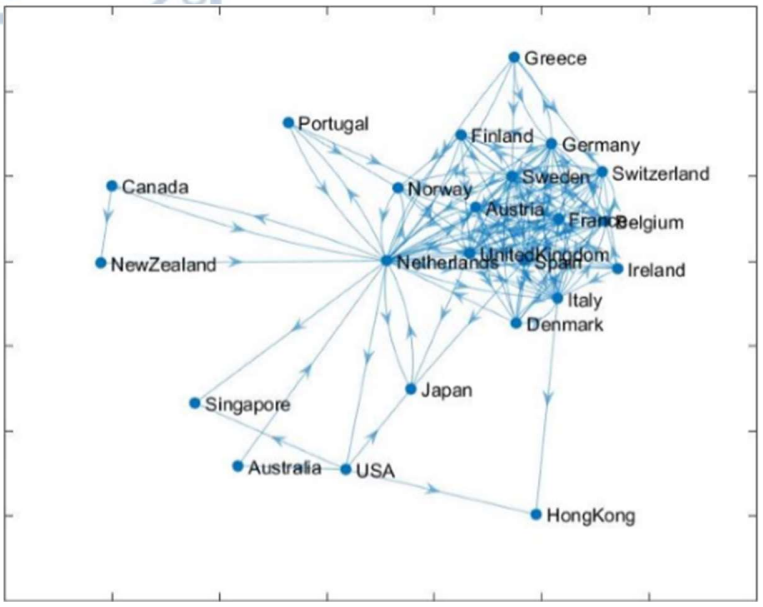

Figure 10.13. The network constructed using PTERV for m=3 and conditioning using PCA in stationary data. Number of connections = 150

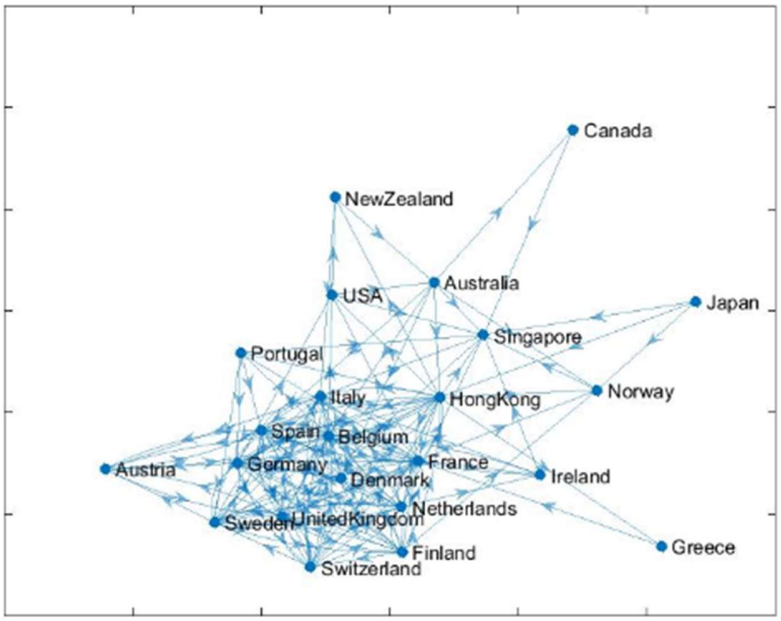

Figure 10.14. The network constructed using PTERV for m=3 and conditioning using the most informative variable to the driver in stationary data. Number of connections = 529

10.4 Summary

To evaluate the networks constructed from the methodology described in a previous section, we have to look into the simulations of the previous chapter for guidance.

The most relevant simulations based on the parameter values are these for a sample length of N=1000 which is relevantly close to the N=1300 of the financial time series used in this application.

For this sample length, the optimal results in the simulations were obtained when full conditioning was applied together with a small embedding dimension ($m = 2$), giving almost identical values of sensitivity and specificity. In the application, we can observe that using these parameter settings the number of connections is similar for the non-stationary and the stationary data, 495 and 463 respectively.

This fact leads us to the conclusion that there is a high probability that these parameter settings are the optimal for the detection of causal relationships in this real world system.

Chapter 11. Conclusions

In this work we have studied the effect of partial conditioning to a limited subset of the confounding variables while estimating casual connectivity between the variables of a system, as an alternative of full conditioning which can sometimes lead to computational and conceptional issues. By conducting our analysis in a simulated system and in a real world financial dataset, we have shown that conditioning on a small number of variables, chosen either as the most informative ones for the driving variable or as a linear combination between our systems' confounding variables with the coefficients of the PC1, can lead to results very close and in some cases even better to those obtained using full conditioning.

The superiority of the results given by partial conditioning in the cases that the embedding dimension or the sample length are higher, is explained by the dimension reduction that these methods offer. As mentioned in a previous chapter, the dimension of the entropy terms in the equations of PTE and $PTERV$ increases both with the time series length and the embedding dimension. The partial conditioning methods come to reduce this dimension, thus leading to more accurate results.

Another advantage of partial conditioning is the reduction of the computational cost. Having a smaller conditioning subset, reduces substantially the time that is needed to complete a set of simulations.

We can conclude that partial conditioning using the above methods is a powerful tool that produces trustworthy results with an accuracy level very close to the one obtained by full conditioning while substantially reducing the computational cost.

References

1) C. Chatfield, (1996), “The Analysis of Time Series”, 5th edition. Chapman and Hall / CRC.

2) en.wikipedia.org/wiki/Independent_and_identically_distributed_random_variables

3) Brockwell, P.J., Davis, R.A. (2002). Introduction to Time Series and Forecasting. 2nd edition. Springer.

4) G. Kirchgässner, J. Wolters, (2008), “Introduction to Modern Time Series Analysis”, Springer.

5) Pierce, D.A., Haugh, L.D. (1977). Causality in Temporal Systems: Characterization and a Survey. Journal of Econometrics. 5 265-293.

6) Shannon, C.E. (1948). A Mathematical Theory of Communication. The Bell System Technical Journal. 27 379-423; 623-656.

7) Cover, T.M., Thomas, J.A. (2006). Elements of Information Theory. 2nd edition. Wiley-Interscience.

8) Yeung, R.w (2008) Information Theory and Network Coding. Springer.

9) Granger, C.W.J. (1969). Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. Econometrica. 37 424-438.

10) Seth, A.K. (2010). A MATLAB Toolbox for Granger Causal Connectivity Analysis. Journal of Neuroscience Methods. 186 262-273.

11) Schreiber, T. (2000). Measuring Information Transfer. Physical Review Letters. 85 461-464.

12) Papana, A., Kugiumtzis, D., Larsson, P.G. (2012). Detection of Direct Causal Effects among Parts of the Brain of Epileptic Patients. International Journal of Bifurcation and Chaos. (to appear)

13) Kraskov, A., Stögbauer, H., Grassberger, P. (2004). Estimating Mutual Information. Physical Review E. 69 066138.

14) Vakorin, V.A., Krakovska, O.A., McIntosh, A.R. (2009). Confounding Effects of Indirect Connections on Causality Estimation. Journal of Neuroscience Methods. 184 152-160.

15) Vlachos, I., Kugiumtzis, D. (2010). Non-uniform State Space Reconstruction and Coupling Detection. Physical Review E. 82 016207.

16) C. Bandt and B. Pompe. Permutation entropy: a natural complexity measure for time series. Physical Review Letters, 88:174102, 2002.

17) M. Staniek and K. Lehnertz. Symbolic transfer entropy. Physical Review Letters, 100(15):158101, 2008.

18) Dimitris Kugiumtzis(2013), Partial Transfer Entropy on Rank Vectors

19) Dimitris Kugiumtzis(2012), Transfer Entropy on Rank Vectors, Journal of Nonlinear Systems and Applications (2012) 73-81

20) A. Papana, C. Kyrtsou, D. Kugiumtzis, and C. Diks. Partial symbolic transfer entropy.

21) Yu, G.H., Huang, C.C. (2001), “A Distribution Free Plotting Position”, Stochastic Environmental Research and Risk Assessment. 15 462-476.

22) https://en.wikipedia.org/wiki/Unit_root_test

23) Dickey, D. A.; Fuller, W. A. (1979). "Distribution of the estimators for autoregressive time series with a unit root". Journal of the American Statistical Association. 74 (366a): 427–431.

24) Y.Xu, Y.Sun. Dynamic linkages between China and US equity markets under two recent financial crises. University of Lund, 2010.

25) M.Veerbek. A Guide to modern Econometrics. Erasmus University Rotterdam. 2004.

26) Ejaz S. Ahmed, Enayetur Raheem, Shakhawat Hossain (auth.), Miodrag Lovric (eds.)-International Encyclopedia of Statistical Science-Springer-Verlag Berlin Heidelberg (2011)

27) L.I. Smith. A tutorial on Principal Components Analysis,2002.

28) A.L.Barabasi. Network Science.

29) Y.Benjamini and Y.Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society, Series B (Methodological)*,Vol. 57, No. 1 (1995), pp. 289-300

30) D.G.Altman, J.M.Bland. Diagnostic tests 1: sensitivity and specificity. BMJ, Volume 308, 1994.

31) A new look at the Statistical model identification. H.Akaike.1974.

32) Periodic orbits in coupled Hénon maps: Lyapunov and multifractal analysis. Antonio Politi and Alessandro Torcini An Interdisciplinary Journal of Nonlinear Science 2, 293 (1992); doi: 10.1063/1.165871.

33) D.Marinazzo, M. Pellicoro and S. Stramaglia. Causal Information Approach to Partial Conditioning in Multivariate Data Sets. Hindawi Publishing Corporation,

Computational and Mathematical Methods in Medicine, Volume 2012, Article ID 303601, 8 pages, doi:10.1155/2012/303601.

34) Zhenyu Zhou, Yonghong Chen, Mingzhou Ding, Paul Wright, Zuhong Lu and Yijun Liu. Analyzing Brain Networks With PCA and Conditional Granger Causality. Human Brain Mapping 30:2197–2206 (2009).

.