INTER-FACULTY PROGRAMME OF MASTER STUDIES in
COMPLEX SYSTEMS and NETWORKS
DEPARMENT OF MATHEMATICS
DEPARTMENT OF BIOLOGY
DEPARTMENT OF GEOLOGY
DEPARTMENT OF ECONOMIC STUDIES
ARISTOTLE UNIVERSITY OF THESSALONIKI

MASTER THESIS

ANALYSIS OF STOCK MARKET CRISIS

A NETWORK THEORY APPROACH

Michail D. Vamvakaris

Supervisor: Nikolaos Varsakelis

Associate Professor A.U.TH

Thessaloniki, June 2016

ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ στα

ΠΟΛΥΠΛΟΚΑ ΣΥΣΤΗΜΑΤΑ και ΔΙΚΤΥΑ

ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ

ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

ΤΜΗΜΑ ΓΕΩΛΟΓΙΑΣ

ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΑΝΑΛΥΣΗ ΤΩΝ ΧΡΗΜΑΤΗΣΤΗΡΙΑΚΩΝ ΚΡΙΣΕΩΝ

ΜΕΣΩ ΤΗΣ ΘΕΩΡΙΑΣ ΔΙΚΤΥΩΝ

ΜΙΧΑΗΛ Δ. ΒΑΜΒΑΚΑΡΗΣ

ΕΠΙΒΛΕΠΩΝ: Νικόλαος Βαρσακέλης

Αν. Καθηγητής Α.Π.Θ.

Θεσσαλονίκη, Ιούνιος 2016

INTER-FACULTY PROGRAMME OF MASTER STUDIES in
COMPLEX SYSTEMS and NETWORKS
DEPARMENT OF MATHEMATICS
DEPARTMENT OF BIOLOGY
DEPARTMENT OF GEOLOGY
DEPARTMENT OF ECONOMIC STUDIES
ARISTOTLE UNIVERSITY OF THESSALONIKI

MASTER THESIS

ANALYSIS OF STOCK MARKET CRISIS
A NETWORK THEORY APPROACH

Michail D. Vamvakaris

Supervisor: Nikolaos Varsakelis
Associate Professor A.U.TH

Approved by the Three-member Assessment Committee on 16th June 2016.

....
N. Varsakelis
Assotiate Professor
A.U.TH

....
N. Karampetakis
Professor A.U.TH.

....
A. Pantelous
Reader
University of Liverpool

Thessaloniki, June 2016

....

Michail D. Vamvakaris
Graduate of the School of Economics of A.U.TH

ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ στα

ΠΟΛΥΠΛΟΚΑ ΣΥΣΤΗΜΑΤΑ και ΔΙΚΤΥΑ

ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ
ΤΜΗΜΑ ΓΕΩΛΟΓΙΑΣ
ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΑΝΑΛΥΣΗ ΤΩΝ ΧΡΗΜΑΤΗΣΤΗΡΙΑΚΩΝ ΚΡΙΣΕΩΝ

ΜΕΣΩ ΤΗΣ ΘΕΩΡΙΑΣ ΔΙΚΤΥΩΝ

ΜΙΧΑΗΛ Δ. ΒΑΜΒΑΚΑΡΗΣ

ΕΠΙΒΛΕΠΩΝ: Νικόλαος Βαρσακέλης

Αν. Καθηγητής Α.Π.Θ.

Εγκρίθηκε από την Τριμελή Εξεταστική Επιτροπή την 16η Ιουνίου 2016.

....

Ν. Βαρσακέλης
Αν. Καθηγητής Α.Π.Θ.

....

Ν. Καραμπετάκης
Καθηγητής Α.Π.Θ.

....

Α. Παντελούς
Reader
University of Liverpool

Θεσσαλονίκη, Ιούνιος 2016

....
Μιχαήλ Δ. Βαμβακάρης
Πτυχιούχος Οικονομικών Επιστημών Α.Π.Θ.

ABSTRACT

In literature, the issue of economic and financial crisis is well studied and many econometric models have been developed for predicting economic or financial crisis or investigating if a crisis affects significantly the economy and financial markets. However, only a little work has been done on how crisis affects the way that stock markets behave. In this work, we take advantage of the recently introduced Visibility Graph algorithm, a method that maps a time series into a network, and we examine how stock markets responded to specific events. More specifically, using five minutes data from January 1996 to March 2016 for the S&P100, S&P500 and S&P1000 indexes, we make a quantitative analysis of the indexes behaviour during this period. We found that indexes react immediately in almost all events and in some cases there are even some early warnings. However, the magnitude of the reaction depends on the size of the index and all indexes do not react in the same way under the same circumstances. Finally, the time series of the returns of an index do not demonstrate the same properties as the price time series.

Keywords

Economic Crisis, Visibility Graph, S&P500

ΠΕΡΙΛΗΨΗ

Στη βιβλιογραφία, το θέμα των οικονομικών και χρηματοοικονομικών κρίσεων έχει μελετηθεί επαρκώς και πολλά οικονομετρικά μοντέλα έχουν αναπτυχθεί με σκοπό είτε να προβλέψουν μελλοντικές οικονομικές ή χρηματοοικονομικές κρίσεις είτε να μελετήσουν αν οι κρίσεις επιδρούν σημαντικά στον τρόπο λειτουργίας των οικονομιών και των αγορών γενικότερα. Ωστόσο, λίγη έρευνα έχει αναπτυχθεί πάνω στο πως οι κρίσεις επηρεάζουν τον τροπό λειτουργίας των χρηματηστηριακών αγορών. Σε αυτή την εργασία εκμεταλλευόμαστε την θεωρία Γράφων Ορατότητας που αναπτυχθηκε πρόσφατα, μια μέθοδος που μετατρέπει μια χρονοσειρά σε γράφημα, και εξετάζουμε πως οι χρηματιστηριακές αγορές αντιδρούν στα διάφορα γεγονότα. Πιο συγκεκριμένα, χρησιμοποιώντας δεδομένα με συχνότητα ανα πέντε λεπτά απο τον Ιανουάριο του 1996 μέχρι τον Μάρτιο του 2016, πραγματοποιούμε μια ποσοτική ανάλυση της συμπεριφοράς των χρηματιστηριακούς δεικτών $S\&P100$, $S\&P500$ και $S\&P1000$. Τα αποτελέσματα δείχνουν ότι αυτοί οι δείκτες αντέδρασαν τάχυστα σε όλα τα γεγονότα που έλαβαν χώρα αυτή την περίοδο ενώ σε κάποιες περιπτώσεις το χρηματιστήριο αντέδρασε εκ των προτέρων σηματοδοτώντας επικείμενα γεγονότα. Έπισης, φάνηκε ότι το μέγεθος του κάθε δείκτη παιζει σημαντικό ρόλο στον τρόπο που αντιδράει ο κάθε δείκτης κάτω από τις ίδιες συνθήκες καθώς όλοι οι δείκτες δεν αντιδρούν με τον ίδιο τρόπο και το ίδιο έντονα. Τέλος, οι χρονοσειρές των αποδόσεων των δεικτών δεν εμφανίζουν τις ίδες ιδιότητες με τις άρχικές χρονοσειρές των δεικτών.

Λέξεις Κλειδιά

Γράφος Ορατότητας, οικονομική κρίση

Contents

ΣΥΝΟΨΗ

Όπως αναφέρθηκε και στην περίληψη, σε αυτή την εργασία έχουμε ως σκοπό να εξετάσουμε πως οι κρίσεις, οικονομικές είτε χρηματοοικονομικές, επηρεάζουν τον τροπο λειτουργίας των χρηματιστηρίων. Για το σκοπό αυτής της εργασιας θα χρησιμοποιήσουμε δεδομένα για 20 χρόνια, απο το 1996 μέχρι το 2016 για τους χρημαστηριακούς δείκτες *S&P*100, *S&P*500 και *S&P*1000. Για όσους δεν είναι εξοικειωμένοι με τα χρηματοοικονομικά αναφέρουμε ότι αυτοί οι δείκτες απαρτίζονται απο μετοχές που διαπραγματεύονται στο χρηματιστήριο της Νέας Υόρκης και του *Nasdaq*.

Ο *S&P*500 αποτελεί για πολλούς τον ¨αντιπρόσωπο¨ της Αμερικάνικής οικονομιάς καθώς οι μετοχές που τον απαρτίζουν έχουν επιλεγεί με τέτοιο τρόπο ώστε όλες οι βιομηχανίες της χώρες να εκπροσωπόνται αναλόγως. Οι άλλοι δυο δείκτες επιλέχθηκαν για λόγους σύγκρισης. Ο ένας περιλαμβάνει τις 100 πιο σημαντικές μετοχές του δείκτη *S&P*500 από άποψη κεφαιλαιοποίησης, *S&P*100, ενώ ο άλλος απαρτίζεται άπο άλλους δύο δείκτες, τον *S&P*400 και τον *S&P*600.

Αρχικά πρέπει να αναφέρουμε ότι υπάρχουν δύο ειδών κρίσεις, οι οικονομικές που σχετίζονται με περιόδους πτώση της οικονομικής δραστηριότητας και οι χρηματοοικονομικές κρίσης που μπορεί να έχουν τη μορφή νομισματικής κρίσης, κρίσης χρέους, τραπεζικής κρίσης και απότομης παύσης ροής κεφάλαιου, με την τελευταία να σχετίζεται άμεσα με τη παύση της ξέφρενης και αναιτιολόγικης αύξησης των τιμών των μετοχών στο χρηματιστήριο. Η κρίση προϋπολογισμού συναντάται μόνο σε προεδρικά συστήματα όπως των Η.Π.Α., και σχετίζεται με την αδυναμία της κυβέρνησης να περάσει τον προϋπολογισμό από το κοινοβούλιο. Δεν θα εντρυφήσουμε περαιτέρω στο σημείο αυτό καθώς υπάρχει εκτενή αναφορά σε κάθε είδος κρίσεις στην βιβλιογραφική επισκόπιση της εργασίας.

Αυτό που αξίζει όμως να αναφέρουμε σε αυτό το σημείο είναι οι κρίσεις και τα γεγονότα που έλαβαν χώρα στο παραπάνω διάστημα ώστε να γίνει κατανοητό στο αναγνώστη τι ακριβώς επιδιώκουμε μέσα από αυτή την ανάλυση. Οι παρατηρήσεις των χρονοσειρών που έχουμε αρχίζουν στις 2 Ιανουαρίου 1996, μια περίοδο που η Αμερική βίωνε μια έντονη πολιτική αντιπαράθεση με τα σημάδια της να είναι εμφανή σε όλη την οικονομία της χώρας. Από τα μέσα Δεκεμβρίου του 1995 είχε ξεσπάσει κρίση προϋπολογισμού στην Αμερική με τον τότε πρωθυπουργό Κλίντον να μην μπορεί να περάσει τον προϋπολογισμό της επόμενης χρονιάς από το κοινοβούλιο και τη βουλή να ανακοινώνει αναστολή όλων των μη-απαραίτητων υπηρεσίων για 20 μέρες. Οι λειτουργίες της βουλής άρχισαν ξανά στα μέσα Ιανουαρίου 1996 αλλά η αναστάστωση που προκλήθηκε διήρκησε για αρκετό καιρό.

Ήδη από τα μέσα του 1995 οι τιμές των μετοχών στα χρηματιστήρια των Η.Π.Α. κάλπαζαν χωρίς τα μακροοικονομικά μεγέθη της χώρας να αιτιολογούν αυτή τη ραγδαία αύξηση, δημιουργώντας μια χρηματιστηριακή φούσκα. Και ενώ οι μετοχές το χρηματιστήριο των Η.Π.Α. κέρδιζαν συνεχώς αξία, το καλοκαίρι του 1997 ξέσπασε η νομισματική κρίση στην Ασία η οποία σύντομα πήρε την μορφή οικονομικής κρίσης και επεκτάθηκε μέχρι τη Ρωσία προκαλώντας την Ρωσική κρίση ένα χρόνο αργότερα. Παρότι η κρίση δεν πήρε διαστάσεις παγκόσμιας κρίσης, οι παγκόσμιες αγορές ταράχθηκα σημαντικά από αυτό το γεγονός.

Η φούσκα που άρχισε να δημιουργείται στα χρηματιστήρια των Η.Π.Α το 1995 έφτασε στο αποκορύφωμα στις αρχές του 2000 όπου και η φούσκα έσκασε και μια περίοδος οικονομικής κρίση ξεκίνησε. Αυτή η κρίση έμεινε γνωστή στη ιστορία ως *dot.com* λόγω της ραγδαίας ανάπτυξης που γνώρισαν εταιρίες στον κλάδο της πληφορορικής και του κυρίως του διαδικτύου. Οι τιμές των μετοχών έφτασαν σε πολύ χαμηλά επίπεδα σε διάστημα μερικών μηνών και χρειάστηκε πάνω απο δύο χρόνια μέχρι οι δείκτες να επανέλθουν στο φυσιολογικά τους επίπεδα. Το ίδιο συνέβη και στα μέσα του 2007 όπου το χρηματιστήριο έφτασε στα ανώτερα επίπεδα μετά από ένα ράλυ σχεδόν δύο χρόνων. Από τον Οκτώμβριο του 2007 και μέσα σε επτά μήνες οι δεικτες έπεσαν κατά 37%. Η κρίση σύντομα μετατράπηκε σε οικονομική κρίση επηρεάζοντας την πραγματική οικονομιά της χώρας ενώ οι διαστάσεις ξεπέρασαν τα σύνορα της χώρας και προκλήθηκε μια παγκόσμια οικονομικά κρίση.

Κατα την διάρκεια αυτή της κρίσης προκλήθηκε και τραπεζική κρίση με αποκορύφωμα το γεγονός του Σεπτεμβρίου 2008. Μία από τις μεγαλύτερες τράπεζες του αμερικάνικου συστήματος η *LehmanBrothers* δεν άντεξε τις απώλειες που προκλήθηκαν από την κρίση και σε συνδιασμό με την άρνηση της κυβέρνησης για ανακεφαλαιοποίηση, ανακοινώθηκε στις 15 Σεπτεμβρη 2008 η πτώχευση του τραπεζικού ιδρύματος προκαλώντας ντόμινο αντιδράσεων τόσο στον τραπεζικό σύστημα όσο και στην πραγματική οικονομία.

Η έξοδος της χώρας απο την οικονομική κρίση του 2008-2009 σχεδόν συμπίπτει χρονικά με την κρίση χρέους που ξέσπασε στην Ευρωπαϊκή Ένωση (ΕΕ) στις αρχές του 2010 και έπληξε κυρίως τα κράτη που συμμετέχουν στο κοινό νόμισμα. Κράτη όπως η Πορτογαλία, η Ιρλανδία, η Ελλάδα, η Ιταλία, η Ισπανία και η Κύπρος που βίωναν οικονομικές δυσκολίες με το χρέος τους να ξεπερνάει τα ανώτερα επιτρεπτά όρια, αναγκάστηκαν να ζητήσουν οικονομική βοήθεια από τρίτους οργανισμούς όπως η Ευρωπαϊκή Κετρική Τράπεζα (ΕΚΤ) και το Διεθνές Νομισματικό Ταμείο (ΔΝΤ). Η λιτότητα που επήλθε αμφισβήτησε την σταθερότητα της Ευρωζώνης και προκάλεσαι ένα ισχυρό σεισμό στην παγκόσμια οικονομία. Όπως είναι φυσικό, το γεγονός αυτό είχε ισχυρό αντίκτυπο στα χρηματιστήρια όλου του κόσμου ενώ παρολίγο αποφεύχθηκε και μια νομισματική κρίση.

Η τελευταία κρίση που έλαβε χώρα αυτή την περίοδο συνέβη στην Κίνα στις αρχές και τα μέσα του 2015. Όπως συνέβη στη Αμερική στο παρελθόν δύο φορές, έτσι και στην Κίνα δημιουργήθηκε φούσκα στις τιμές των μετοχών στο χρηματιστήριο που κατέληξε σε κατάρρευση των αγορών. Ένας πρώτος πανικός δημιουργήθηκε στα χρηματιστήρια της Κίνας στις αρχές του 2015 με τις αγορές να ακολουθούν μια πτωτική πορεία, να ανακάμπτουν και σύντομα να ξαναμπαίνουν σε μια διαρκή πτωτική πορεία περί τις αρχές Αυγούστου του 2015 όπου κατέληψε με τους χρηματιστηριακούς δείκτες της χώρας να αγγίζουν τα χαμηλότερα επίπεδα της δεκαετίας και να παρασέρνουν μαζί τους στην πτωτική πορεία τα μεγαλύτερα χρηματιστήρια παγκοσμίως.

Πριν συνεχίσουμε στην σύντομη περιγραφή της μεθοδολογίας και των αποτελεσμάτων, θα αναφέρουμε ακόμα ένα γεγονός το οποίο δεν συμπίπτει με καμία κρίση αλλά είχε σημαντική επίδραση στην αγορά όπως έδειξαν τα αποτελέσματα. Η αρχή του 2004 ήταν καθοριστική καθώς συμπτίπτει με την κορύφωση του πολέμου στο Ιρακ όπου συμμετέχουν τα Αμερικάνικα στρατεύματα, την αύξηση των επιτοκίων απο την Εθνική Τράπεζα των Η.Π.Α., την ανάρρωση της οικονομιάς μετά την κρίση του 2000-2001 και την τιμή του πετρέλαιο να φτάνει σε ασυνήθιστα υψηλά επίπεδα.

Μετά από την αναφορά στα γεγονότα της περιόδου 1996-2016 είναι εύκολο να καταλάβει κανείς την σημαντικότητα αυτής της περιόδου από άποψη εξελήξεων. Ωστόσο κρίνεται αναγκαίο να δώσουμε και μια ποσοτική αιτιόλογιση στο γιατί επίλεξαμε να βασίσουμε την έρευνά μας σε αυτό το χρονικό διάστημα.

Κοιτώντας κανείς την πρόσφατη ιστορία, από το 1950 και μετά, η περίοδος 1996-2016 μπορεί να χαρακτηριστεί ως η πιο ασταθής περιόδος με τόσες κρίσεις να λαμβάνουν χώρα ουσιαστικά η μία μετά την άλλη. Για να σας δώσουμε μια ποσοτική διάστασης της αστάθεις στο χρηματιστήριο αυτή την περιόδο, ο δείκτης $S\&P500$ την περιοδο 1964-1984 είχε τυπική απόκλιση 14 μονάδες, ενώ την περίοδο 1996-2016 είχε τυπική απόκλειση 360 μονάδες. Αναφέρουμε την τυπική απόκλιση ως μέτρο σύγκρισης μεταξύ των δύο περιόδων διότι στην θεωρία των χρηματοοικονομικών χρησιμοποιείται ως μέσο ποσοτικοποίησης του ρίσκου μιας επένδυσης.

Ένας από τους λόγους που επιλέχθηκαν χρηματιστηριακοί δείκτες από τις Η.Π.Α. σαν βάση για την έρευνα μας είναι κυρίως διότι το Χρηματιστήριο της Νέας Υόρκης είναι το μεγαλύτερο παγκοσμίως σε όρους κεφαλαιοποίησης. Από την άλλη πλευρα, η Αμερικάνικη οικονομικά είναι ανάμεσα στις δύο πιο ισχυρές οικονομίες παγκοσμίως μαζί με την Κίνα σε όρους Ακαθάριστου Εθνικού Προϊόντος (ΑΕΠ), εξαγωγών κτλ. Επίσης δεν πρέπει να ξέχναμε, ότι εκεί ξεκίνησαν οι τρεις από τις τέσσερις μεγάλες χρηματιστηριακές κρίσεις του αιώνα αν συμπεριλαβουμε την κρίση του 1928-1930 και την Μαύρη Δευτέρα του 1987 όπου και τα δύο γεγονότα ήταν επίσης αποτέλεσμα χρηματιστηριακής φούσκας. Ως τελευταίο λόγο θα αναφέρουμε το γεγονος ότι ειδικά ο $S\&P500$ είναι ένας δεικτης που έχει μελετηθεί πολύ στο παρελθόν οπότε θα ήταν πιο εύκολο να συγκρίνουμε την δουλεία μας με την δουλειά άλλων ερευνητών.

Για να μπορέσουμε να αναλύσουμε με τον καλύτερο τρόπο τα δεδομένα μας και να έχουμε όσο το δυνατόν καλύτερα αποτελέσματα, διαχωρίσαμε την αρχική χρονοσειρά σε υπο-περιόδους (παράθυρα) και υιοθετήσαμε την μέθοδο του κινητού παραθύρου όπου δύο συνεχόμενα παράθυρα έχουν επικάλυψη δύο μήνων ενώ το συνολικό μήκος του κάθε παραθύρου είναι 3 μήνες. Η επικάλυψη μεταξύ των διαδοχικών παραθύρων εξασφαλίζει συνέχεια στα αποτελέσματα μας και μειώνει την πιθανότητα να υπάρξει κάποιο γεγονός το οποίο δεν θα αποτυπωθεί στα αποτελέσματα. Το μήκος του παραθύρου επιλέχθηκε με βάση τα παράθυρα που έχουν χρησιμοποιηθεί από άλλους ερευνητές σε αντίστοιχες έρευνες που αφορούν χρηματοοικονομικές σειρές αλλά και με βάση κάποιους περιορισμούς που επιβλήθηκαν από τον αλγόρυθμο που χρησιμοποιήσαμε. Παράθυρο μηκρότερου μήκους θα μείωνε αισθητά την αξιοπιστία των αποτελεσμάτων.

Ο αλγόρυθμος που χρησιμοποιήθηκε στην ανάλυση των δεδομένων δεν θα παρουσιαστεί σε αυτό το σημείο, ωστόσο επιγραμματικά θα αναφέρουμε ότι αυτό που προσφέρει η μεθοδογία που εφαρμόστικε εδώ είναι ότι μπορεί να διακρίνει αν μια σειρά (ή χρονοσειρα) παρουσιάζει ίδια στοιχεία με μια τυχαία σείρα με αυτοσυσχέτιση μηδέν, ή με στοχαστική σειρά με αυτοσυσχέτιση ή μια σειρά που παρουσιάζει στοιχεία χάους ενώ παράλληλα μπορεί να ποσοτικοποιήσει το βαθμό της αυτοσυσχέτισης σε μια στοχαστική σειρά. Από την άλλη πλευρά, υπάρχει η δυνατότητας διάκρισης μιας ασυσχέτιστα τυχαίας σειράς από μια σειρά χάους ακόμα και αν η δεύτερη έχει ¨μολυνθεί' με θόρυβο. Το άλλο χαρακτηριστικό του αλγορύθμου είναι ότι μπορεί να ποσοτικοποιήσει τη μη-γραμμικότητα σε μια σειρά. Αυτά τα δύο στοιχεία θα αναλύσουμε τόσο στις αρχικές σειρές των δεικτών που προαναφέρθηκαν αλλά και στις σειρές των αποδόσεων. Επιδιώκουμε να εξετάσουμε κατά πόσο οι κρίσεις επηρεάσαν τα χαρακτηριστικά των χρονοσειρών, και αν κατά τη διάρκεια των διαφόρων γεγονότων επηρεάστηκε η μη-γραμμικότητα των χρονοσειρών.

Τα αποτελέσματα παρουσιάζονται στα γραφήματα 5.1 εώς 5.8 στο κυρίως μέρος της εργασίας. Ο εκθέτη γάμμα δείχνει αν μια σειρά παρουσιαζεί τις ίδες ιδιότητες με μια ασυσχέτιστη τυχαία σειρά, με μια συσχετισμένη στοχαστική σειρά ή με μια σειρά χάους. Ο εκθέτη γάμμα ισούται με $-\ln(3/2)$ για ασυσχέτιστες τυχαίες σειρές, είναι με μεγαλύτερος από $-\ln(3/2)$ για σειρές που μοίαζουν με χάος και είναι μικρότερος από $-\ln(3/2)$ για στοχαστικές σειρές με αυτοσυσχέτιση. Στην τελευταία περίπτωση, όσο πιο μικρή είναι η τιμή του εκθέτη, δηλαδή όσο πιο πολύ πλησιάζει στο -1, τόσο πιο ισχυρή είναι η αυτοσυσχέτιση μεταξύ των τιμών, γεγονός που μειώνει την πιθανότητα παρατήρησεις ακραίων τιμών.

Από την άλλη πλευρά, η απόσταση ***Helliger*** μεταξύ της κατανομής των έσω και των έξω βαθμών των κορυφών στη περίπτωση των Γράφων Ορατότητας ποσοτικοποιεί τη μη-γραμμικότητα των υποπεριόδων και την μη αναστρεψιμότητα τους. Όσο μεγαλύτερη η απόσταση μεταξύ των δύο κατανομών τόσο πιο μη-γραμμική είναι μια σειρά ενώ στην περίπτωση που η δύο κατανομές ταυτίζονται τότε πρόκειται για στάσιμες σειρές. Φυσικά αυτό που ξέρουμε από την θεωρία είναι ότι οι χρηματοοικονομικές σειρές είναι μη στάσιμες γεγονός το οποίο επιβεβαιώνεται καθώς για κανένα παράθυρο οι δύο κατανομές δεν ταυτίστηκαν.

Αυτό που είναι προφανές είναι ότι οι κρίσεις φαίνονται να επηρεάζουν τα χαρακτηριστκά των χρονοσειρών. Κάθε κρίση επηρεάζει με τον δικό της μοναδικό τρόπο το χρηματιστήριο, ωστόσο κάποια κοινά χαρακτηριστικά υπάρχουν μεταξύ όλων των κρίσεων. Αρχικά αυτό που γίνεται αντιληπτό με μια πρώτη ματιά είναι ότι το χρηματιστήριο αντιδρά διαφορετικά σε δύο είδη κρίσεων, αυτές που λαμβάνουν χώρα στο εξωτερικό περιβάλλον της αγοράς, είτε άμεσο είτε έμμεσο, και αυτές που δημιουργούνται εγγενώς από την αγορά. Στην υπό μελέτη περίοδο στην δεύτερη κατηγορία ανήκουν οι φούσκες του 2000 και του 2007, ενώ όλες οι άλλες κρίσεις ανήκουν στην πρώτη κατηγορία.

Το γεγονός ότι ο εκθέτης γάμμα δεν εμφανίζει μεγάλη συσχέτιση ανάμεσα στους δείκτες, 0.43 είναι η συσχέτιση μεταξύ του $S\&P100$ και $S\&P500$, σημαίνει ότι τα ίδια γεγονότα δεν επηρεάζουν το ίδιο τους δείκτες. Ωστόσο, όπως φαίνεται στις περιόδους των κρίσεων ο εκθέτης γάμμα ακολουθεί σχεδόν σε μεγάλο ποσοστό κοινή πορεία και για τους τρείς δείκτες. Το ίδιο δεν συμβαίνει όμως ανάμεσα στις σειρές των αποδόσεων των δεικτών και τις αρχικές σειρές των δεικτών. Ο εκθέτης γάμμα της σειράς των αποδόσεων δεν μοιάζει να συμπορεύται με τον αντίστοιχο εκθέτη της κάθε σειράς. Και στις τρεις περιπτώση, η συσχέτιση του εκθέτη γάμμα μεταξύ των αποδοσεών κι των αρχικών σειρών των δεικτών είναι σχεδόν μηδέν. Επίσης, η σειρές των αποδόσεων δεν φαίνεται να παρουσιάζουν διαφορετικά χαρακτηριστικά κατά τη διάρκεια των κρίσεων.

Ο εκθέτης γάμμα δείχνει ότι οι κρίσεις επηρεάζουν τις αγορές. Όταν η κρίση δημιουργείται ενδογενώς, στην περίπτωση που υπήρχει δηλασή φούσκα στο χρηματιστήριο η οποία έσκασε, οι δείκτες συμπεριφέρονται στοχαστικά και με μεγάλο βαθμό αυτοσυσχέτισης ενώ παράλληλα η μη-γραμμίκότητα την χρονοσειρών μειώνεται. Σε κάθε άλλη περίπτωση οι δείκτες συμπεριφέρονται σαν χάος που συνοδεύεται με αύξηση της μη-γραμμικότητας των σειρών γεγονός που είναι πλήρως αναμενόμενο αν λάβει κανείς υπόψην του ότι το χάος είναι άρρηκτα συνδεδεμένο με τα μη γραμμικά συστήτα.

Ένα επίσης σημαντικό αποτέλεσμα είναι ότι το μέγεθος του δείκτη επηρεάζει το πόσο μη-γραμμική είναι μια σειρά. Κατά γενική ομολογία όσο πιο πολλές μετοχές συμπεριλαμβάνονται σε ένα δείκτη τόσο πιο μη-γραμμική τίνει να είναι η χροσοσειρά του δείκτη γεγονώς που μειώνει την προβλεπτική ικανότητα των επενδυτών. Όπως φαίνεται στο γράφημα 5.7 η χρονοσειρά του $S\&P1000$ εμφανίζεται να συμπεριφέρεται αρκετά πιο μη-γραμμικά σε σχέση με τους άλλους δύο δείκτες ενώ η μέση τιμή για τον $S\&P500$ ξεπερνάει την αντίστοιχή του $S\&P100$.

Εντύπωση προκαλεί το γεγονός ότι παρότι τα αποτελέσματα δεν είναι καθόλου αντιφατικά μεταξύ τους, ωστόσο συγκούονται με την θεωρία αποδοτικών αγορών όπου οι τιμές και κατά συνέπεια και οι αποδόσεις των μετοχών πρέπει να εμφανίζουν ίδια χαρακτηριστικά με ασυσχέτιστα τυχαίες σειρές και να μην υπάρχει η δυνατότητα πρόβλεψης από την πλευρά των επένδυτων. Ωστόσο με βάση τα αποτελέσματα ούτε οι αποδόσεις των δεικτών ούτε οι αρχικές σειρές των δεικτών φαίνεται να συμπεριφέρονται τυχαία. Το χάος επιτρέπει την βραχυπρόθεσης πρόβλεψης ενώ η ύπαρξη αυτοσυσχέτισης ενθαρρύνει τόσο την βραχυπρόθεσμη όσο και την μακροπρόθεσμη πρόβλεψη υπό προϋποθέσεις.

Αυτή η εκτενής περίληψη ολοκληρώνεται εδώ και μετά την σύντομη αναφορά στα αποτελέσματα της έρευνα. Το υπόλοιπο της εργασία είναι γραμμένο στα αγγλικά και παρακάτω μπορεί κάποιος να βρει όλες τις τεχνικές λεπτομέριες σχετικά με τη μεθοδολογία ή τα δεδομένα.

INTRODUTION

Financial markets is the place where people buy and sell financial products such as currencies, stocks and derivaties. The average daily trading value was approximately US$169 billion in 2013 for the New York Stock Exchange which is more than the annual Gross Domestic Product of Algeria, the fourth strongest economy in Africa. As many people claim, stock markets are a billion dollar bussiness and milllions of people are invovled in this industry.

Investors try to guess or as they claim to predict future prices of financial products but no one actually can accomplish this consistenly. And this unpredictability of the market challenges more and more people. Many models have been developed in order to predict the price of stock in the future by using either quantitative or qualitative methods. A broad range of tools coming from different fields has been used so as to predict the market. Quantum finance, financial econometrics, mathematical and quantitative finance, neuro-finance and behavioural finance are different branches of finance science that represent tools coming from the corresponding field of research.

Although the main challenge in the field of finance is the forecast and prediction, in this work we do not try to predict anything but rather to investigate if crises affect the way that financial markets work and if all crises have the same impact on the financial markets. We recruite an algorithm who originated from the field of statistical mechanics in order to analyze financial data. More specifically, by using data with a frequency of five minutes and by analysing them by mean of Visibility Graph and Horizontal Visibility Graph, we attempt to find some common patterns among all crises.

The dataset, that we obtain, contains observations for twenty years for the indexes S&P 500, S&P 100 and S&P1000. For the purpose of our analysis we will use the first two indexes and the last one will be used only for comparison.

The most common problem that people analysing financial time seires face is that financial series are characterized as non-stationary series and most models cannot perform well under these circumstances. As a result, most people study the bahaviour of return series instead of the original series. The algorithm that we recruite do not underperform even in that case. This work is the first attempt to analise stock price series and not return series.

The Horizontal Visibility Graph has been developed in a theoritical framework and has never been actually applied to any real data. The Visibility Graph has been used for real life data analysis but even in that cases the data was not related with financial time series. So, this work is original and innovative since we apply for the first time the Horizontal Visibility Graph in real data and the purpose of our research deviate from the traditional reseach in the field of finance.

In the following section there is a brief introduction to economic crisis theory, network theory and existing literature regarding the model applied to data, in section three we talk about the methodology, section four is dedicated to a brief description of data used for the purpose of this work. Results are presented in section five and we close in section six with the conclusions.

ACKNOWLEDGEMENTS

I would like to thank my family from the bottom of my heart for their not only financial but also psychological support during my studies. I also thank Dr. Christos Emmanouilidis who supported me throughout my studies and encouraged me to attend this master program. Last but not least I thank my friends Marianthi Moschou, Prokopis Karadimos and Antonios Meimaris for their priceless support.

Chapter 1

LITERATURE REVIEW

1.1 Economic & Financial Crisis Review

If someone searched in a dictionary for the Complex Systems definition, he might come across the following description:

"A Complex System is composed by many diverse and autonomous parts which are interrelated and inter-depended since they are connected through interconnections/links. Consequently, behaviour of a complex system arises from the interaction of its individual parts and cannot be predicted from the properties of the components."

In our modern societies examples of complex systems can be found in many fields, such as sociology, neuroscience, chemistry, artificial intelligence, economy etc. In this work we only focus on the study of the economy in particular throughout periods of economic crisis. Yet, before proceeding with the main part of this paper, we should define what a crisis is, recognize its different types and study how a crisis can be propagated into the system.

In any system, a crisis is defined as period of very poor performance of the system, which renders an immediate corrective action imperative. Thereafter, an economic crisis, generally speaking, is a period of dismal economic performance. However, since economy is a complex system, a crisis could be triggered by a crisis having occurred in a component of the economic system which then propagates into the whole system. The great range of components is responsible for the different types of possible crises which are analyzed below.

The two major components in the economic system is the real economy and the financial industry. When a crisis is triggered by a poor performance of the real economy then we refer to recession or depression, depending on the magnitude of the crisis, otherwise we refer to a financial crisis. These two major parts can be partitioned into smaller parts and as a result there is a very rich list with names describing different crises regarding where the problem originated from. We will delve into details only for the most common types of crisis which are recession and depression known also as economic crisis,as well as financial crisis such as currency crisis, debt crisis, bank crisis and sudden stop.

As the National Bureau of economic research suggests: *"A recession is a significant decline in economic activity spread across the economy, lasting more than a few* *months, normally visible in real GDP, real income, employment, industrial production, and wholesale-retail sales. A recession begins just after the economy reaches a peak of activity and ends as the economy reaches its trough. Between trough and peak, the economy is in an expansion. Expansion is the normal state of the economy; most recessions are brief and they have been rare in recent decades."* Recession is a part of the business cycle, yet an unpleasant one.

Another description of recession is as *"an important drop in the economic activity for example industrial production, employment, real income and wholesale-retail trade, usually lasting more than a few months. The prevailing opinion dictates that a recession is a drop in quarterly gross domestic product (GDP), after being inflation adjusted, for two consecutive quarters."*

In most cases depression describes an extreme recession which usually lasts more than two years. A remarkable increase in unemployment rate, a dwindling industrial output, bankruptcies and sovereign debt defaults, reduced trade, and a volatility in currency values are some of the symptoms of depression. At such times, the trust of the consumers and investments in real economy diminishes, causing a severe drop in the economic activity.

Despite the existence of definitions describing both recession and depression it is still difficult to pinpoint the exact date at which a crisis initiated. Economists still disagree partially about the definitions; they are indeed really abstract, and many statistical models have been developed in order to trace the beginning and the end of a crisis without being particularly successful. Most of the times we realize that a crisis has begun by observing either the reaction of financial markets or other macroeconomic factors but such methods are prone to delays in observtion.

However, before proceeding with defining financial crisis, we have to define a very special type of crisis that appeared twice in US economy in 1996 and 2013, the budget crisis. The reason we mention this particular type of crisis is that both events of budget crisis coincide with the time interval under study. A budget crisis can happen only in presidential systems and can be described as the situation in which the government provisionally suspends non-essential services because a government budget has failed to passing. Politically speaking, a budget crisis may develop in a situation of disagreement between state and civil society.

On the other hand, as mentioned before, financial crisis can be partitioned into currency, bank, debt or sudden stop crisis. The definitions of these crises are more concrete. However, there is usually an overlap among them since the initiation of one crisis will cause the initiation of another crisis at the same time. At this point it is worth mentioning that real economy crisis can trigger a financial crisis and also a financial crisis can be rapidly transmitted into real economy.

As International Monetary Fund (IMF) claims *"While financial crises have common elements, they do come in many forms"*. Some of the most common elements are essential change in assets prices, large scale balance sheet deficit and financial support of the government either in form of recapitalization of the banking sector or providing liquidity to the economy. On the other hand, financial crisis does not come out of the blue but in most cases there is an event that triggers the crisis. Abrupt bank runs, contagion among financial markets, existence of asset price bubbles, credit insolvency and fire sales are on the top of the list with events that drive the economy out of equilibrium. Events with asset and credit booms that sooner or later burst are linked with almost all type of crisis in the financial sector. According to the report with the [3], published by IMF, *"there are two types of financial crisis those classified using strictly quantitative definitions; and those dependent largely on qualitative and judgmental analysis. The first group mainly includes currency and sudden stop crises and the second group contains debt and banking crises."*

A currency crisis emerges when investors believe that there is a lack of foreign exchange reserves and thus doubt the ability of a central bank to preserve a fixed exchange. This type of crisis is usually followed by a speculative attack resulting in a devaluation/depreciation of the local currency in the foreign currency markets causing the government to respond immediately. The government intervention could take many forms such as increasing the interest rates sharply , forcing capital controls or even expending a large amount of the international reserves. A currency crisis is often correlated with a real economy crisis but it could also be the aftermath of a prolonged period of payments deficit. A strongly quantitative definition of currency crisis is given by [5] as a nominal depreciation of a currency of at least 25%.

A sudden stop is a terminology for describing either a large and often sudden decline in international capital inflows or a dramatically swift reversal in the capital flows to a country. Sudden stops are commonly associated with decrease in production, private spending and consumption as well as a depreciation of the exchange rate. This type of crisis can be classified by a quantitative definition as follows; sudden stops are commonly described as periods that contain at least one observation where the year-on-year fall in capital flows lies at least two standard deviations below its sample mean [2].

We refer to a debt crisis in case a country cannot or for political purposes refuses to service its sovereign or foreign debt obligations. In the case of foreign debt crisis, the government does not complete a payment to its borrowers while a sovereign debt default can take two forms. One option is to refuse to service its domestic fiscal obligations by defaulting explicitly or the other option is to devaluate its currency triggering inflation in its economy. The problem with this type of crisis is that, even though we know the exact day that the government defaulted, we can never be sure when the crisis is over, an issue on which economists still have not expressed a unanimous solution.

The final form of financial crisis that we may face in this work is the banking crisis. It is mostly due to abrupt cash withdrawals caused by a sudden decline in depositor confidence or fear that the bank will be closed by the chartering agency. Since banks save only a small amount of money it is possible that they will run out of liquidity or they go bankrupt.

Since the deeper purpose of this paper is to analyze the reaction of stock exchange markets to different kinds of crisis by means of Visibility Graphs, we thought it would preferable for the reader to know in advance some details about economic crisis.

1.2 Fundamentals of Graph Theory

1.2.1 *Graph Definitions*

Many people in their daily life deal with or participate in one or another kind of networks but only some of them have ever thought about it. From computer networks, to social networks and from road network to financial markets networks, all of us are involved in a kind of a network structure. Graph, on the other hand, is the vehicle for the mathematical modeling of networks. In this work we mostly work with graphs, and for this reason we start by defining what a graph is and then citing some properties.

Simple Graph

A simple definition of Graph is: *"A graph is a collection of vertices that can be connected to each other by means of edges. In particular, each edge of graph joins exactly two vertices* [13] or equivalently *"A graph is a way of specifying relationships* *among a collection of items. A graph consists of a set of objects, called nodes, with certain pairs of these objects connected by links called edges" [4].*

Translating this abstract description into a mathematical definition we obtain:

Definition 1. A graph G consists of a collection V of vertices and a collection of edges E, represented by $G = (V, E)$. Each edge $e \in E$ is said to join two vertices, which are called its end points. If e joins $u, v \in V$, we write $e = (u, v)$. Vertex u and v in this case are said to be adjacent (or neighbours) and degree (κ) is the number of vertices adjacent to any particular vertex $v \in V$. Edge e is said to be incident with vertices u and v, respectively [13].

This definition corresponds to the simplest form of graphs one can meet. But here some questions arise:

1. Is it possible for an edge to start and end at the same vertex?

2. Is it possible for two vertices to be connected by more than one edges?

In both case the answer is "Yes". In the first case we say that there is a loop in the network while in the second case there are multiple-edges. A graph where there is no loops and any multiple edges is called a *simple graph.*

Regular & Complete Graph

A special case of a simple graph is when all vertices are connected with each other and the graph is called *complete.* If this is not true, yet it is possible for all vertices to reach any other vertex, the graph is call *connected*, otherwise it is called *unconnected* and consequently there is at least one vertex in the graph with $\kappa = 0$. Another special form of graph is one where all vertices has the same κ degree and this graph is usually named *κ-regular.*

Directed Graphs

In a simple graph, an edge $e = (u, v)$ is an unordered set and the graph is called *undirected.* However, this is not always the case. Many real life graphs are modeled as *directed graphs*, which means that a vertex A can access vertex B but not the opposite. In this case, an edge connecting two vertices is an ordered set $e = (u, v)$. Directed networks are common in a transportation networks, for instance the route map of a shipping company where the direction shows where the destination of each route it.

Weighted Graph

In addition, in a simple graph an edge indicates that there exists a relation among two vertices; this type of graphs is called *unweighted*. When there is a need for quantifying the relation among vertices then, the most appropriate way of doing this is by means of *weighted graphs* where the weight could be, for instance, the distance among two cities or how many times per month person A and B meet each other or even the work load that moves from one stage of the production process to the next one.

By now, we have defined what a graph is and we have referred to some classification of graphs based on some properties they have. In the rest of this section, we will mention some measurements that uncover details regarding network topology.

1.2.2 *Properties of Graphs*

Degree Distribution

Degree sequence is a list which contains the degree of each vertex of the graph. By normalizing the frequency of each distinct degree κ in the degree sequence, we obtain the degree distribution. As it is mentioned below, random networks have an exponential degree distribution while scale free networks have a power law degree distribution. Degree distribution gives us, indeed, much information regarding the corresponding graph but in most cases it is not enough when one would like to compare the structure of two different graphs. For this reason we examine the most broadly used measures of a network.

Diameter & Radius

Path is an ordered set of edges that connects a sequence of distinct vertices or equivalently, path is a sequence where there are no repeated vertices. If the path starts and ends at the same edge then we refer to a cycle. The shortest path between two nodes is the path among all possible paths with the minimum length and is called geodesic. The length of the longest shortest path in a graph or otherwise the length of the maximum geodesic is called diameter of a graph while the shortest geodesic is the radius.

Clustering Coefficient

Another measure that sheds light into the way that vertices are connected with each other is the clustering coefficient (cc). Clustering Coefficient measures to which degree vertices of a graph intend to cluster together. The basic idea of this measurement is to examine whether two vertices, which are incident to the same vertex, are also incident to each other; in other words if these three vertices form a triangle. There are two different versions of the index, the local and the global one.

In the case of the local clustering coefficient (clustering coefficient of a particular vertex), the index equals the ratio of the number of triangles formed by a particular vertex i and its neighbours to all possible triangle that could be formed. Of course the highest value of this index can be achieved in the complete graph where all nodes are incident with each other. The mathematical definition of this index is given in [13]:

Definition 2. Consider a simple connected, undirected graph G and a vertex $v \in V(G)$ with neighbor set $N(V)$. Let $n_v = |N(v)|$ and m_v be the number of edges in the subgraph induced by $N(v)$, i.e., $m_v = |E(G[N(v)])|$. The clustering coefficient $cc(v)$ for a vertex v with degree $\kappa(v)$ is defined as

$$cc(v) = \begin{cases} m_v/\binom{n_v}{2} = \frac{2*m_v}{n_v(n_v-1)} & \text{if } \kappa(v) > 1 \\ \text{Undefined} & \text{otherwise.} \end{cases}$$

The global clustering coefficient or otherwise known as transitivity, is defined as

$$n_\Delta(G)/n_\Lambda(G),$$

where $n_\Lambda(G)$ is the number of all distinct triples and $n_\Delta(G)$ the number of all distinct triangles.

Density of a Graph

The last metric to be cites here is the density of a graph. In fact the density give us some information regarding how close to the complete graph is the graph under investigation. The mathematical definition is quite intuitively, the number of edges existing in a graph over the number of edges of a complete graph with the same number of vertices $m/\binom{n}{2}$ where m is the number of edges in a graph and n the number of vertices.

1.2.3 *Common types of Graphs*

So far we introduced the most important indicators that can give us an insight into the topology of the network. In the next few pages, there is a reference to some types of networks that always exhibit the same properties: Random networks introduced by Erdos and Renyi in 1959, Small-world network by Watts and Strogatz in 1998 and Preferential Attachment networks introduced by Albert-Barabasi in 1999.

Random Graph

A random network (ER) is a simple graph in which all pairs of nodes are connected with a fixed probability p, where $0 \leq p \leq 1$ and as the probability p increases, it is more and more probable to observe a denser graph. In other words, given a set of n vertices and for all $\binom{n}{2}$ potential edges, each edge will appear with the same probability p. In addition, this means that the total number of edges in a random network will be approximately $\binom{n}{2}p$. For this reason, there are two different ways to represent this networks, $G_{n,p}$ where n is the number of nodes and p the probability of the existence of an edge between, or otherwise $G_{n,e}$ where e is a fixed number of edges existing in a graph.

Intuitively, all graphs with the same number of nodes and the same number of edges or the same probability p are equally likely. A $G_{n,e}$ graph is selected uniformly at random from the set of all possible graphs with number of n nodes and e links. A $G_{n,p}$ graph is created by connecting vertices in an arbitrry manner.

Regarding the topological properties of random networks, it is worth mentioning that the degree distribution follows a binomial distribution. The probability of a particular vertex to have a degree κ equals with $p^{\kappa}(1-p)^{n-1-\kappa}$and as a result for the whole network the probability of obtaining a vertex of degree κ equals with

$$P[K(u) = \kappa] = \binom{n-1}{\kappa} p^{\kappa}(1-p)^{n-1-\kappa}.$$

On the other hand, the clustering coefficient of an $G_{n,p}$ equals the probability p while the average path length of a random network can be calculated by

$$\bar{d}(H) = \frac{\ln n - \gamma}{\ln pn} + 0.5$$

where γ is the Euler constant and $\bar{d}$ is the average distance among all vertices [13].

Small-world Graph

A small-world network incorporates in fact the properties of ER networks but the important difference appears in the clustering coefficient. While ER networks display small average path length and a small clustering coefficient, the small world networks exhibit small average path length and a clustering coefficient significantly higher in comparison with random networks.

This is a type of graphs where, even though most nodes are not connected with each other directly, most nodes can reach any other node in the graph with only a few steps or equivalently the average path length is small. Nodes are usually connected with other nodes by more than one short path. In small- world networks the distance between any two arbitrary nodes increases pro-rata to the logarithm of the number of vertices N in a graph, $L \propto \log N$, and it can be proven that the maximum distance among two vertices cannot be larger than (n/k) where n is the number of vertices and k the mean degree.

In literature, there are some algorithms for constructing a network that exhibits small-world properties, yet the most broadly used is the Watts–Strogatz mechanism (WS). At this point, it is worth mentioning the WS mechanism since it is a simple one and will uncover the underlying similarities between small-world and random networks.

We can start by assuming that there is a set of n vertices and an even number κ such that $n \gg k \gg \ln n \gg 1$. In the first step, we order all vertices into a ring and then, link each vertex with the first $\kappa/2$ nodes on its right side (clockwise) and the first $\kappa/2$ nodes on its left side (counterclockwise). In the second step, we replace each edge (v, h) with an edge (v, u) with probability p, where u is an arbitrary selected node from $V(G)$ other than h, and such that (v, u) does not belong to $E(G)$. Obviously, as p increases from 0 to 1, the network becomes more and more similar with a random network $G_{n,e}$.

It is proven that for any graph G constructed with the algorithm proposed by Watts-Strogatz such that the clustering coefficient is equal to

$$cc(G) = \frac{3}{4}\frac{(\kappa - 2)}{(\kappa - 1)}.$$

In addition, it is also proven that the average path length of every Watts-Strogatz graph can be approximated by

$$\bar{d}(u) \approx \frac{(n-1)(n+\kappa-1)}{2\kappa n}.$$

Regularly, small-world networks are characterized by an over-abundance of hubs and by hubs we mean nodes with a high degree in comparison to the average degree of the network. Existence of hubs can facilitate the minimization of the average path length.

Scale-free network

The last kind of networks that we cite is the scale-free network. A definition of a scale-free network is that its degree distribution follows a power law $P(\kappa) \sim \kappa^{-\alpha}$ distribution regardless of any other property the network has. This means that there are only few nodes with extremely high degree and that the number of nodes with a high degree decreases exponentially. Commonly, α lies in the interval [2, 3] but this is not always the case.

In a scale free network it is common for the clustering coefficient to decrease as the degree of a node increases. This happens because nodes with low degree form dense sub-graphs and those sub-graphs are linked to each other through hubs. If hubs are connected to each other depends on the network structures. There are some networks where hubs tend to be connected to each other, for instance the world network of airports, while under different circumstances hubs tend to be linked only with low-degree nodes for instance in biological networks. This phenomenon discloses a hierarchical structure in the network.

In comparison to other types of network cited above, scale-free ones hve the most robust network structure against failure and this comes from the hierarchical structure of the network. Since the majority are low-degree nodes and a failure occurs randomly, the probability for a hub to be affected is very small. And even in the case that a hub is affected by a failure, the network will not lose its connectedness because of the remaining hubs in the network. However, in case that the failure of nodes is not random but targeted it is much easier to decompose a scale-free network by targeting its hubs.

Albert and Barabasi have proposed an algorithm for constructing scale free networks. Beginning with a network with n_0 nodes, at each time step we add m new nodes. Each of these m newly added nodes j will be connected with $n < n_0$ existing nodes with probability p. The probability p for a new node j to be connected with an already existing node i depends on the number of links that the existing node has.

Mathematically speaking, the probability p that a new node j will be connected with an existing node i is calculated by:

$$p_i = \frac{\kappa_i}{\sum_j \kappa_j}$$

where κ_i is the number of links of node i and the sum is calculated over all pre-existing nodes j. The higher the degree of an existing node, the more probable is for a newly added node to connect with it. This is the reason why highly connected nodes are tend to be become even more connected while nodes with few are not appear to be so attracted. This is the well-known preferential attachment principle described in[1].

Regarding the properties of scale free networks, we know that the average path length increases almost logarithmically as the size of the network increases. The average path length is well approximated by $l \sim \frac{\ln N}{\ln \ln N}$, where N is the number of nodes in a network. On the other hand, the clustering coefficient cannot be calculated analytically, but an empirical approximation suggests that $cc \sim N^{-0.75}$.

1.3 Visibility Graph Theory

1.3.1 *Natural Visibility Graph*

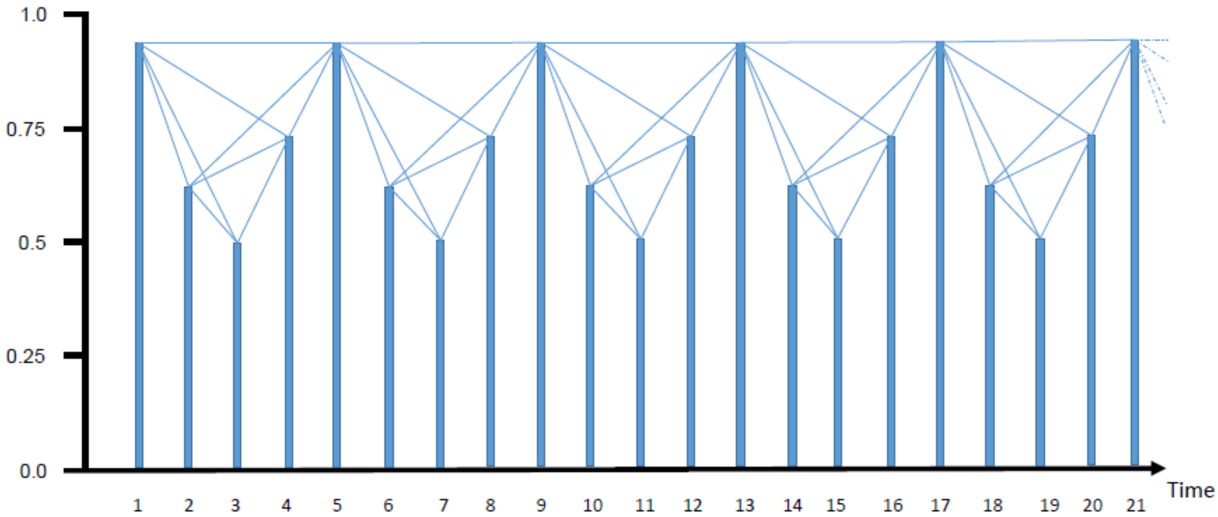
Recently, a novel method has been developed which maps a time series into a network which inherits some properties of the initial time series. As a result one can apply network theory analysis on the Network produced by Visibility algorithm in order to extract useful results regarding the corresponding time series. It has been found in [8] that periodic time series are converted into a regular network with a discrete degree distribution which is plotted with a finite number of peaks related to the period of the initial series, random series are mapped into exponential random networks and fractal series are converted into scale-free networks, supporting recent evidence that power law degree distributions are related to fractality.

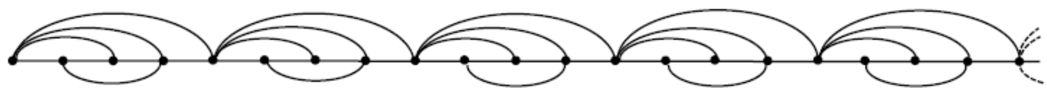
Every series data from the initial series corresponds to a node in a Visibility Graph (VG) and two nodes are connected to each other only if visibility exists between two nodes. Two arbitrary observations (y_a, t_a) and (y_b, t_b) have visibility if for every observation (y_c, t_c) such that $t_a < t_c < t_b$, the following holds:

$$y_c < y_b + (y_a - y_b)\frac{(t_b - t_c)}{(t_b - t_a)}.$$

Figure 2.1 is a graphical representation of this mapping algorithm.

Figure 1.1: A graphical representation of the mapping a time series into a network.





The main properties of a visibility graph are:

1. It is connected since all series data "sees" at least the previous and the following data observation. The degree k of each node lies between
$2 < k < n - 1$ where n is the number of series data.

2. The outcome of the visibility algorithm is by default an undirected graph. However, the graph can easily been transformed into a directed one if the degree of each node is divided into ingoing-degree (k_{in}) and outgoing-degree (k_{out}), which means how many nodes have visibility on a specific node i and how many nodes can been shown by a specific node i ,respectively.

3. It is invariant under affine transformation of the series data.

4. It is unweighted. Nonetheless, this graph can be transformed into a weighed one by using the slope of the visibility line that connects two series data heights as weigths. By applying this method we can achieve reversibility in the process described above.

A lot of attention has been drawn into fractals and self-similarity, it is interested to investigate how visibility graph approach can identify fractal time series and whether it is possible to recognize different types of fractality i.e. stochastic and deterministic fractality. Actually, the proposed methodology captures the hub repulsion phenomenon associated with fractal networks and thus distinguishes scale-free visibility graphs evidencing the small-world effect from those showing scale invariance. As shown in [8], fractal series are transformed into scale-free network that exhibit power law degree distribution while it is claimed that for fractional Brownian motions $\alpha = 2.00 \pm 0.01$ and for Conway series $\alpha = 1.2 \pm 0.1$ where α is the exponent of the tail of the degree distribution. In particular, it was shown also shown that the visibility graph obtained from the well-known Brownian motion has got both the scale-free and the small world properties. In addition, because fractal series are described by its Hurst exponent it is possible to distinguish different types of fractal series through VG since Hurst exponent and the exponent of the power law distribution of the associated are linearly related. It is proven in [9] that the relationship between Hurst exponent and power law distribution exponent is the following:

$$\alpha(H) = 3 - 2H$$

or equally:

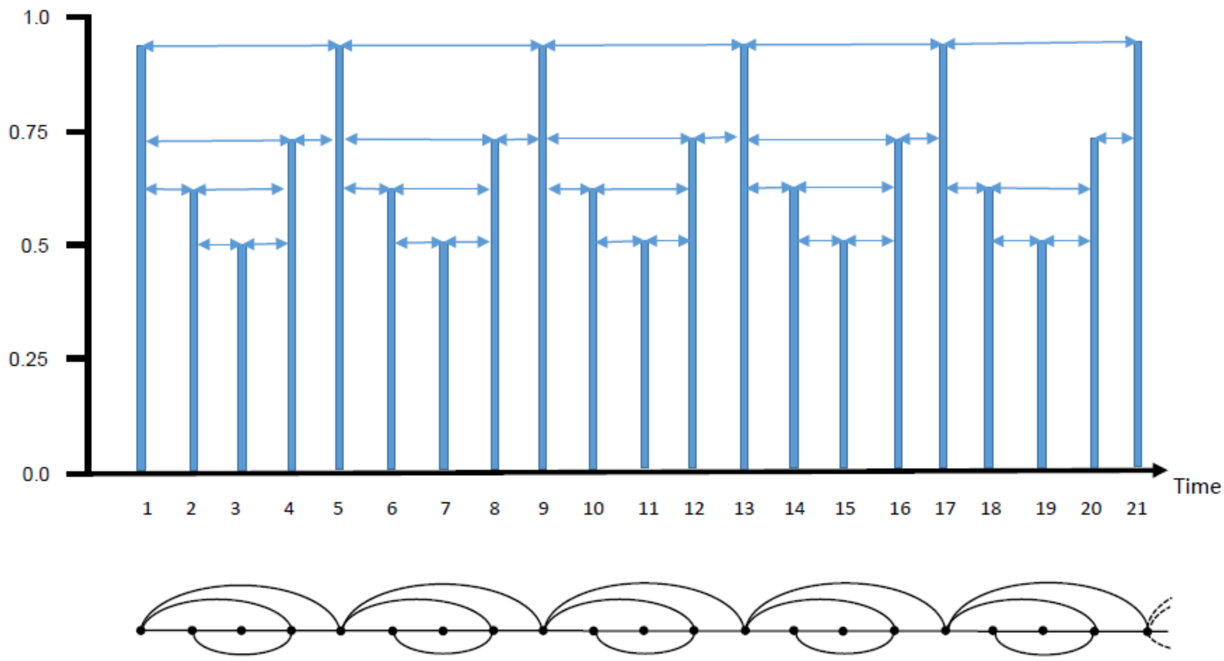
$$H(\alpha) = \frac{(3-\alpha)}{2}$$

where H is Hurst Exponent and α is the exponent from the power law degree distribution $p(k) \sim k^{-\alpha}$ Hurst exponent, known as the index of long-range dependence, can take any value in the interval $0 < H < 1$. If $H = 0.5$ the series is completely uncorrelated, if $H > 0.5$ the series is strongly correlated meaning that a high value is probable to be followed by another high value and also values a long time into the future will also tend to be high. On the contrary, if case of $H < 0.5$ the series is anti-correlated which means that a high value probably will be followed by a low value.

1.3.2 *Horizontal Visibility Graph*

Horizontal Visibility Graph (HVG) is a methodology for mapping time series into a network similar to visibility graph approach proposed above, yet the visibility criterion is much simpler. The HVG is based on the following geometric criterion: each data series is transformed into a node in the associate network and two arbitrary data (y_a, t_a) and (y_b, t_b) are connected each other if for every data (y_c, t_c) such that

Figure 1.2: A graphical representation of the mapping a time series into a network with the Horizontal Visibility Graph methodology.



$t_a < t_c < t_b$:

$$y_a, y_b > y_c.$$

Figure 2.2 is a graphical representation of this mapping algorithm

In practice, HVG is a sub-graph of the visibility graph and as a result inherits all the main property described above. However, it should be noticed that some information is inevitably lost during this process of mapping a time series into a network since two periodic time series with the same period with be mapped in the same network. So, unless we use a weighted network, this process seems irreversible. One can easily transform this process into a reversible one by weighting all edges with weight that is equal to the height difference of the associated data.

Before proceeding to show how efficient is HVG in discriminating randomness and chaos, we should present some important characteristics of the HVG. As proven in [6] , a graph is a Horizontal Visibility Graph if and only if it is outerplanar and has a Hamilton path. So, a HVG is a non-crossing graph. It is also implied that if G is an HVG then minimum $\kappa(G) = 1 or 2$ and maximum $\kappa(G) \leq n - 1$, where k is the degree of a node and n the number of vertices.

Discriminating Randomness and Chaos

By means of Horizontal Visibility Graph we can discriminate between random and chaotic time series. More specifically, as has been shown in [12], the degree distribution of any random time series converted into a HVG is

$$P(\kappa) = \frac{1}{3}\left(\frac{2}{3}\right)^{\kappa-2}$$

regardless of the probability distribution $f(x)$. It is also proven that there is a close relationship between the value of the series data and the degree of the corresponding node in the associated graph.

Since the degree distribution of any random time series has an exponential form $P(\kappa) = \frac{1}{3}\left(\frac{2}{3}\right)^{\kappa-2}$, it is shown that the HVG can discriminate between both low and high dimension chaos and randomness in time series. The performance of the algorithm is undoubtable since even in the extreme case where a chaotic time series was polluted with 100% noise the algorithm could discriminate chaotic series from a random one. It is important to notice that this methodology cannot quantify chaos, instead it discriminates between chaos and uncorrelated randomness effectively.

In figure2.3, it is obvious that the semi-log degree distribution of the noisy chaotic series diverges from the theoretical one of the random process.

Discriminating Correlated stochastic and Chaotic Processes

Based on findings above, it is easy to show that

$$P(\kappa) = \frac{1}{3}\left(\frac{2}{3}\right)^{\kappa-2}$$

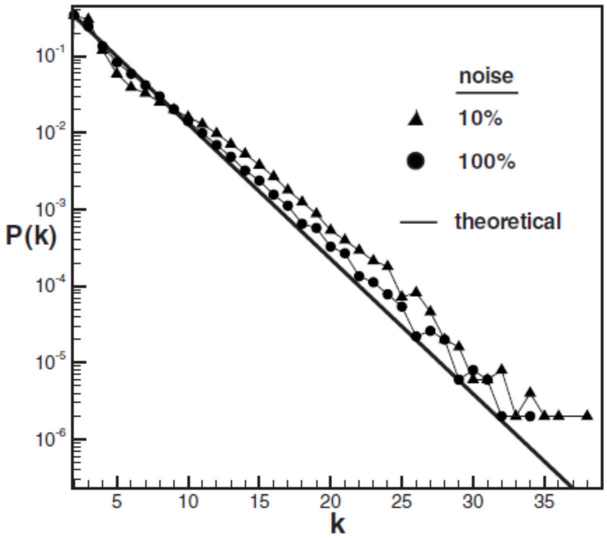
can be rewritten as:

$$P(\kappa) = \exp^{-\gamma_{un}\kappa}$$

where the power law exponent γ of tail of the degree distribution of a uncorrelated random series is equal to $\gamma_{un} = \ln\frac{3}{2}$. In [11], has been shown that time series related to correlated stochastic process has a degree distribution which is exponentially decaying, although with a larger slope than that in case of uncorrelated random series which means that $\gamma_{stoch} > \gamma_{un}$. At this point, we have to point out that in correlated stochastic process the exponent γ of the tail of the degree distribution is in inverse proportion with the correlation strength of a time series.

The explanation is straightforward, the higher the degree of correlation of the

Figure 1.3: The semilogrithmic plot of the degree distribution of a horizontal visibility graph associated with (triangles) noisy chaotic series extracted from logistic map ($\mu = 4$) with a measurement noise level of 10% (by amplitude) and (circles) idem but for noise level of 100%. The solid line corresponds to the theoretical prediction of random series $P(\kappa) = \frac{1}{3}\left(\frac{2}{3}\right)^{\kappa-2}$ [12].



correlation function, the less probable is to obtain an extreme value on the data and as a result the tail of the degree distribution becomes smoother. Consequently, as the correlation strength of a time series decreases, the degree distribution approaches the degree distribution of a random series i.e. $\gamma_{un} = \ln\frac{3}{2}$. Generally speaking, correlated series present lower data variability than uncorrelated ones, and as a result the possibility of a node to reach far visibility is diminishing.

On the other hand, it is found in [11] that in case of chaotic time series the tails of the degree distribution can be well approximated by

$$P(\kappa) = \exp^{-\gamma\kappa}$$

where $\gamma < \gamma_{un}$ because a chaotic process has an HVG whose degree distribution has an exponential tail with smaller slope than that of random process. In addition, there are evidences of a net deviation from the exponential shape for small values of the degree, which is associated to short-range memory effects.

As proposed, the calculation of γ comes straightforwardly from the fitting of the HVG's degree distribution (concretely, the tail) to an exponential function.

Irreversibility of Stationary Time Series

A physical process is time-reversible if the underlying dynamics of the process are preserved when the coherence of time-states is reversed. More specifically, a stationary time series is called time-reversible when the reversible time series and its reversed time series have the same joint probability distribution. If a time series is irreversible, then there are evidences for non-linear underlying dynamics in the process. Whether a time series is reversible one or not is an important issue since irreversibility is also related with entropy production.

For some processes is very important to know if they are reversible one or not and for this reason have been developed many approaches for tracing irreversibility. One method recently introduced is concerned about stationary time series and is based on natural Visibility Graph and Horizontal Visibility Graph.

A directed HVG (DHVG) is constructed by dividing the degree $\kappa(x_i), i = 1, 2, 3 \ldots$ of each node, into an ingoing-degree $\kappa_{in}(x_i)$ and outgoing degree $\kappa_{out}(x_i)$ such that $\kappa(x_i) = \kappa_{in}(x_i) + \kappa_{out}(x_i)$. As $\kappa_{in}(x_i)$ is defined the number of past nodes that have visibility on a specific node N_i while $\kappa_{out}(x_i)$ is the number of future nodes which the node N_i has visibility on.

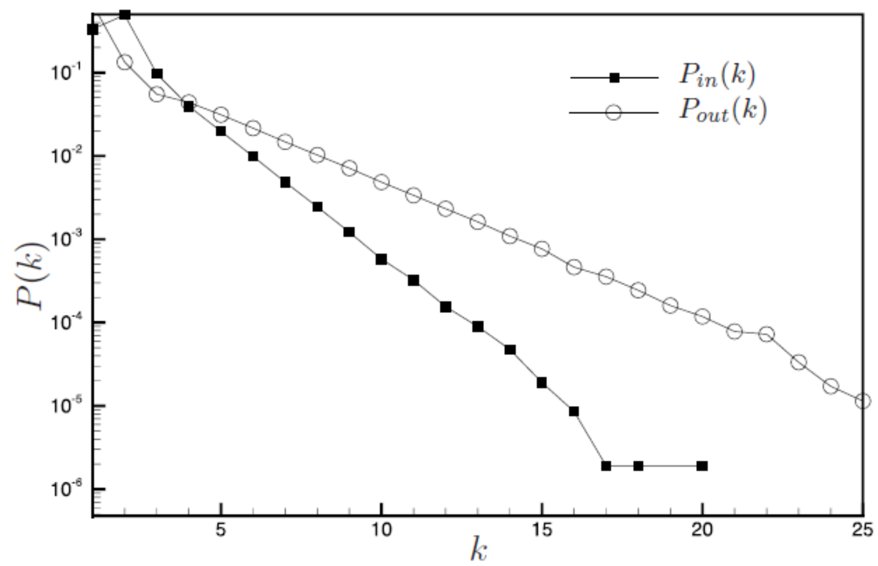
Before proceeding, an assumption has to be made. All information related to the amount of irreversibility time series is stored in the ingoing and outgoing degree distribution. Consequently, the amount of irreversibility of a series can be approximated by the distance between the ingoing, $P_{in}(\kappa)$, and the outgoing, $P_{out}(\kappa)$,degree distribution.

Kullback-Leibler Divergence (KLD) is tool coming from the realm of information theory which can measure the distance between two probability distributions f and q. Usually, probability distribution f is thought as the actual one based on available data while the other one (q) is thought as the theoretical one. Given a random variable x and two probability distributions on x, $f(x)$ and $q(x)$ the KLD is defines as:

$$D(p||q) \equiv \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}.$$

If two distributions are coincided in all points then the KLD equals zero, otherwise it's values is bigger than zero. This means that the higher the value of the KLD the more distant the two distributions are. In [10], has been shown that the information stored in the outgoing degree distribution is sufficient to distinguish between reversible and irreversible stochastic stationary series which are real valued but dis-

Figure 1.4: In this figure it is plotted the $P_{in}(\kappa)$, and the outgoing, $P_{out}(\kappa)$, degree distribution of a chaotic time series, Logistic map where $\mu = 4$. We know that this process is irreversible and it is obvious who the in- and out-degree distribution diverge [10].



crete in time. By using the KLD, it is possible to quantify the divergent between the in- and out- degree distribution, as well. In this particular case the KLD is defined as:

$$D(P_{out}(\kappa)||P_{in}(\kappa)) \equiv \sum_{x} P_{out}(x) \log \frac{P_{out}(x)}{P_{in}(x)}.$$

If KLD=0 the $P_{out}(\kappa) = P_{in}(\kappa)$ and the time series is reversible.

It has also been proven that uncorrelated stochastic, linearly correlated stochastic and both exponentially decaying and power law decaying correlated stochastic series are reversible. On contrary, dissipative chaotic series, even in the case that they are polluted with noise, are irreversible while conservative chaotic series are reversible one. This process not only can effective discriminate between reversible and irreversible series but can also quantify the level of irreversibility.

Although, the author of HVG has proposed the KLD as an apropriate measure of the distance between the in- and out degree distribution of a graph, dealing with real data some times can cause many problems. The KLD requires that there is no $P_{in}(\kappa) = 0$ nd also requires that the domain for both $P_{in}(\kappa)$ and $P_{out}(\kappa)$. In case that maximum $P_{in}(\kappa) >$ maximum $P_{in}(\kappa)$ then, the distacne cannot be difined.

For the purpose of our work, we exploite the Helliger distance (H) as a measure for quantifying the distance of the in- and out degree distribution. The H for two descrete probability distributions $P = (p_1, p_2, p_3, ..., p_n)$ and $Q = (q_1, q_2, q_3, ..., q_n)$ is defined as

$$H(P,Q) = \frac{1}{\sqrt{1}}\sqrt{\sum_{i=1}^{n}(\sqrt{p_i} - \sqrt{q_i})^2}.$$

If the value of H is equal to zero then, the series is reversible and there are not any non-linear dynamics related to this particular process. Otherwise, the higher the value of the H, the more irreversible a series is and the more non-linear dynamics underly the process under study.

Irreversibility of non-Stationary Time Series

Non-stationary time series is a stochastic process whose joint probability distribution does change when shifted in time and as a result regularly non-stationary series are irreversible, with some exceptions. Consequently, parameters such as the mean and variance, if they are present, also do change over time and follow some trends. As a result, non-stationary series are irreversible since statistical properties of a non-stationary process vary over time. However, one could argue that there should exist different degrees of irreversibility in non-stationary processes if we take into consideration the source of irreversibility in close relation to directionality (or, in other words, to underlying sources of memory): for instance, a Markovian random walk should arguably be "less irreversible" than a non-Markovian one, even if both are non-stationary [7].

The methodology for quantifying irreversibility in non-stationary time series is the same the one proposed in the previous section. However, what differs is that for non-stationary series that are actually reversible, specific examples will be cited below, the $KDL(P_{in}||P_{out})$ will have a positive value for finite-size series and approached zero as the size of the series increases.

It is interested to refer some special cases whether even though the series is non-stationary, it is actually reversible. Simple random walks in the form of:

$$X_{t+1} = X_t + \xi$$

where $\xi \in [-1, 1]$, are reversible time series and non-stationary at the same time. There are also some other case where even the series is non-stationary is in practice reversible but there is no point in delving into more details for the purposes of this project.

To summarize, the above described process not only can effective discriminate between reversible and irreversible series but can also quantify the level of irreversibility. It is well known that financial times series are non-stationary and irrevesible process.

Chapter 2

METHODOLOGY

So far, we have cited all available literature regarding the algorithm recruited for the purpose of this research. In this section, we describe the whole process from the data analysis and the algorithm implementation while in the next section there is a particular reference in data used in the context of this paper.

There are some questions we should answer before proceeding into results, in order to give an overview of our reasoning to the reader. Why did we choose to proceed our research with one process out of all proposed methodologies for analysing financial data? This question will be answered in the remaining of this section.

To begin with, even though in the literature review section we refer to both Visibility Graph and Horizontal Visibility Graph method, in this work we focus on the HVG. There are three different reasons why we did select HVG. First of all, since the HVG is a filtration of the VG it is much less computational expensive and if one takes into account the enormous amount of data, we could not have completed this project otherwise. To give you a flavor, to map a time series of 5800 observation into network trough VG method we need more than twenty hours (without using any parallel programming method) while to complete the same work with the HVG it requires on average one minute. Secondly, the VG has been already used in some papers as method for analyzing real data something that is not true for the HVG. Finally, as one can realize by studying the literature review, the only thing that VG offers while HVG does not, is the calculation of the Hurst exponent, an index of memory in fractal Brownian time series. Although in financial markets it is very important to quantify the memory of the market under different circumstances and this is an index of how efficient markets works, this issues has been extensively studied so, there would be no innovation in our work by studying once again this issue.

Since we investigate how different types of crisis have affected the US stock market in the recent history, we have to break the initial data set into smaller series (windows) and study the behaviour of each particular window. But before proceeding, we have to decide what should be the length of the window of our study, if we will use the method of rolling window and if so, what should be the overlap between two consecutive windows or the method of non-overlapping windows.

It is generally accepted that stock market time series are characterized by autocorrelation which means that past prices affect future prices. However, the very long history does not offer any information for the future and so a proper interval should be specified. Bearing this in mind and after looking for the time interval that other researchers used in their paper, we concluded that the proper interval is a window with length three months.

Moreover, looking into historical data of the index S&P 500 for the last two decades, there were six major market crisis and in three cases the index recovered into pre-crash levels within 40 to 50 trading days which is close enough with a window of 71 trading that we selected to use. Because of the aforementioned reasons we choose to work on time intervals of three months.

Given the fact that the HVG is a filtration, part of information is lost during the mapping of a time series into a network. For this reason there is approximately a minimum number of observation that a series should contain in order to obtain robust results. After some empirical experiments, we found that the best length for the window is three months or 5800 obsevations and this number almost coincide with the number of observation that the author of the HVG methodology implicitly set as minimum required number (6000 observations). More details in the next section.

Having specified that the length of windows under study is 3 months, we decided to work with the method of rolling window in order to achieve a continuity in our results. As it is also mentioned in more details below, what we study with the HVG method is the average behaviour of the market over a specific period of time. In case we had selected the method of non-overlapping window, for instance, we would have missed information in case that something have happened within the second half of a window and the first half of the next window. This method is also the prevailing method used by people how make research in the field of finance.

Applying a rolling window to a data set requires to be specified both the length of the window and the overlap between two consecutive windows. In case that the overlap between two consecutive windows would be 70 trading days, that would mean an enormous amount of data that would be difficult to be processed. On contrary, a small overlap would conceal some events or some interesting facts. We found empirically that the most appropriate overlap interval is 48 trading days or equivalently two months.

In this first papers of the section methodology part we have extensively analyze the main points of the methodology of this work. But what are we looking for actually?

For each window with length 3 calendar months or 71 trading days we have mapped the time series into the corresponding Horizontal Visibility Graphs, the directed and the undirected one. For each window we calculated the distance between the in-degree and the out-degree distribution by means of Hellinger Distance which quantifies the degree of irrevesibility of the time series during this period of time and the exponent γ of the tail of the degree distribution.

In this way we can examine how crisis affect the properties of a market series and whether there is any correlation among the variables that we calculated. In the next section there will be a detailed description of the data set we used.

Chapter 3
DATA DESCRIPTION

Data used for the purpose of this research project come from the private data base Sirca. We possess data from 2^{nd} January 1996 to 5^{th} April 2016 for the indexes Standard & Poor 100 (S&P100) and Standard & Poor 500 (S&P500) while for the Standard & Poor 1000 (S&P1000) we have data from 31^{st} December 2001 to 5^{th} April 2016.We have data with a frequncy of five minutes and consequently we obtain 447,709 observation for the first two indexes and 316,467 observation for the latter one. In the rest of this chapter we explain the importance of these indexes as well as the role of the US in the global economy.

What these indexes represent?

The S&P500 index is a stock market index grounded on the market capitalization of 500 largest companies having their common stocks trading either on New York Stock Exchange (NYSE) or on NASDAQ. On contrary with other indexes such as Dow Jones Industrial Average or NASDAQ Composite, Standard & Poor indexes are grounded on a weighting methodology.

For many people working in the field of finance, S&P500 is thought as a representative of the US stock market and a predictor of the US economy. According to National Bureau of Economic Research, *S&P500 has been classified as a leading indicator of business cycles.*

Companies that are included in this index are selected by a committee and every time that there is a proposal for a new company to be added in the S&P500, the committee assesses the potential company using the following criteria: market capitalization, liquidity, domicile, public float, sector classification, financial viability, length of time publicly traded and listing exchange. The selection of the index components are very restrictive and there is also the provision that companies included in S&P500 are representative of all industries of the US economy. Throughout last ten years the composition of the index has been changed dramatically. From 2005 to 2015, 188 components of the index have been substituted by new components.

When professors introduce undergraduate students into modern portfolio theory and more specifically into single factor model, S&P500 is thought as the "market" while in many cases S&P500 is used as a benchmark for the whole US stock market, for instance, many funds manager compare their perfomance with the corresponding performance of a passive portfolio based on S&P 500. For these reasons we selected this particular index.

In order to be sure that the results we extract are robust, in our research we include two other indexes, a narrower one and a broader one. The S&P100 index is a proper subset of S&P500, including 100 leading US stocks. Its components represent almost 57% of the market capitalization of the S&P 500 and about 45% of the U.S. equity market capitalization. Components the S&P 100 are the largest and most established companies of the S&P 500. The average market capitalization of the index is nearly twice that of S&P500 while the correlation among these indexes is very high.

There are not much to say about the S&P 1000 Index rather than it combines the S&P MidCap 400 also known as S&P400 and the S&P SmallCap 600, known as S&P600, to form an investable benchmark for the mid- to small-cap segment of the U.S. equity market. We have to point out that even S&P100 is a proper subset of S&P500, none of these indexes are actually a proper subset of S&P1000. These is the reason why S&P1000 is not closely related with the other indexes.

Usually, in finance we are interested in the behaviour of returns of an investment instead of the price itself. For this reason, we analyse time series of both prices and returns for all indexes. By adapting this procedure, we are able to examine if all indexes respond the same in the same crisis/event or if the size of the index indeed matter. An interesting issue is to investigate if time series of the index returns exhibit the same properties with the time series of prices.

Why do we study the period 1996-2016?

Now it is about time to answer the most important question and specify why we did focus on this particular time interval i.e. in the period of 1996-2016. What is so important and why did not selected, for instance, the time interval between 1973 and 1993?

If one have a look in the historical closing prices of S&P500, he will realize that from 1964 to 1986 the index was almost flat. Between 1986 and 1987 there was an abrupt increase in points of the index which ended up with the black Monday on 19 October 1987. In the mean period until 1995 nothing interesting have happened while in 1995-1996 it is estimated that started the dot.com bubble which finally burst in the beginning of the new millennium.

What we just claimed is also obvious if one check some statistical parameters.

The mean points of the index during the period 1964-1983 was 99.12 points with a standard deviation at 14.9 points while in the period 1996-2016 the mean was 1272.9 points and the standard deviation was 339 points. Obviously, it is much more interesting to investigate what happened in this latter period and examine which events cause this high volatility of the index.

At the end of 1995 and early in 1996 the US experienced a budget crisis. The former Democratic president of US, Bill Clinton, conflicted with the Republican Congress over federal budget of 1996. This event caused a suspension of all non-essential services from November 14 through November 19 of 1995, and from December 16 of 1995 to January 6 of 1996, for a total of 27 days. This triggered a period of high uncertainty in the whole economy and in stock market in particular lasted for several months.

In 1995 the so called doc.com bubble emerged but it was not that obvious until 1996. Between 1996 and 2000 the US stock market was rallying but the effect of a crisis happened thousands miles away blocked this upward slope of the index. In the summer of 1997 the Asian financial crisis started and a year later the Russian crisis took place. Both crisis originated as a currency crisis which then propagated into the real economy causing an economic crisis for the country involved. Even though these crisis did not propagate worldwide, the fear of an upcoming meltdown of the world economy affected also the US stock market.

The US stock market welcome the new millennium with the worst way. Some early warnings for the upcoming crisis appeared in the first months of 2000. The dot.com bubble burst on September 2000 and the S&P500 climb down by 37% within seven months. This crisis lead the economy into a recession lasting from March 2001 to November 2001. In the meantime, the terrorist attack of 9th September in New York happened, an event that lead S&P500 to a deep recrudescence. At the same time Argentina was suffering by its debt crisis which result the default of the country since 1999. The official default happened in the last week of 2001. However, the impact of the crisis in Argetntina is not obvious since this event coincide with a more severe economic crisis in the US economy. The US stock market needed about 1,015 trading days to recover since the initiation of the dot.com crisis.

The next crisis is the well-known global financial crisis of 2007-2008. In fact, many people still disagree when the crisis originated. Some claim that the crisis started in July 2007 when the market reached its maximum points while other support that the crisis started in December of 2009 when the recession actually became obvious. The peak of the crisis was the Lehman Brothers bankruptcy on September 2008. This event escalated the crisis even more. It took more than two years for the index to recover. In March of 2009, S&P 500 index reached its minimum value (629 points) since July 1996. In macroeconomic terms the recession period ended on July 2009 while the stock market started its upward-sloping trajectory a month earlier.

In the end of 2009 another crisis appeared in the European continent. The foreign debt of countries such as Greece, Portugal, Ireland, Spain and Cyprus was huge and countries could not repay or refinance their government debt and bail out over-indebted banks without the intervention of third parties like other Eurozone countries, the European Central Bank (ECB), or the International Monetary Fund (IMF). The S&P500 recovered after 130 trading days after the initial shock. The debt crisis almost ended up since July 2014 when Ireland and Portugal exit successfully their bailout programs.

While European Debt crisis was on process, in 2013 US experienced another budget crisis. The first sixteen days of October 2013, the United States federal government entered a shutdown and curtailed most routine operations. A problem created when the two chambers of Congress did not agree to an appropriations continuing resolution. Political fights over this and other issues between the House on one side and President Barack Obama and the Senate on the other led to a budget impasse which threatened massive disruption.

Finally, the Chinese economy that was expanding rapidly last years starting slowing down and this concerns a lot investor about an upcoming slowdown in the worldwide economy. The stock markets in china exhibit some early warning after having experiencing three market crisis within last 9 months, without triggering a bear market period. These event was propagating all over the world and the world largest stock markets experienced loses following the trend of Chinese stock market. In order to boost again its economy, Chinese government devaluated its currency, yet the investors' concern about the future of the economy is obvious in the market. Oil and other commodities are still trading at a very low price, fact that make it doubtable if the worldwide economy is about to keep expanding for next years.

Why do we choose to study US stock markets?

For all this reason, we believe that is of major interest to examine this period and investigate what the reaction of the market was in any of the aforementioned events. The reason why we did chose to study the US market is still open.

The US economy is ranked as the largest economy in the world in terms of Gross Domestic Product by IMF, World Bank and United Nations. It is second largest economy in terms of exports and it was part of G8 group throughout the whole period under study. Looking through history one can realize that most economic crisis or market crashes has originated in US and this a particular reason for studying this market.

Chapter 4
RESULTS

As we mentioned above, in this section we analyse the results we got after applying the HVG into the dataset. However, before proceeding in a detailed analysis of the results, it would be interesting to have an look in the relationship between three indexes.

The index series of S&P500 and the S&P100 index series have a correlation of 0,689 for the period from 2^{nd} January 1996 to 5^{th} April 2016. On the other hand, for the period from 31^{st} December 2001 to 5^{th} April 2016 the correlation of these indexes is 0.98 while for the same time span the correlation with S&P1000 is 0.95 and 0.91 for the S&P500 and S&P100 respectively. We can infer that these last two indexes are more closely related with each other, which is intuitive, since S&P100 is a proper subset of S&P500, as we mentioned above.

At this point we have to point out that for the rest of the paper, when we make a comparison analyses between S&P100 and S&P500 we refer to the interval from 2^{nd} January 1996 to 5^{th} April 2016, while when we compare S&P100 or S&P500 with S&P1000 we refer to the interval from 31^{st} December 2001 to 5^{th} April 2016.

The method of the rolling window resulted in the construction of 233 distinct periods for the S&P500 and S&P100 and 164 periods for the S&P1000. Each period consists of 5800 observations which is equal with 71 trading days, almost 3 months. To make it clear, for instance, the first window of S&P500 contains the observations of the period 2/1/1996-1/4/1996, the second window contains the period 31/1/1996-6/5/1996 and so on. Obviously between these two consecutive windows there is an overlap of 3900 observations or equally 2 months.

By means of this analysis we try to investigate:

1. If stock markets under study work efficiently.

2. If all crisis affect stock markets in same way.

3. If the size of the index is a factor determining its reaction to crisis

As has already been mentioned, our analysis focuses on the properties of S&P500 index, and we take advantage of the other index for comparison purposes. More specifically, the index S&P1000 will be used only to answer the question if the size of the index affect its response into crisis. This happens because the series of this index has not the same length with the other two indexes and this might cause misleading results.

Since the output of computational calculations is an enormous amount of data which would be meaningless to be presented in form of tables, we base our analysis on the graphical representation of results.

The shadowed area represents periods of turbulence. The red color shadow represents periods where the crisis raised internally from the US stock market, for instance, a stock market bubble of 2007-08, while the gray color represents periods in which turbulence originated outside of the market, either from another country, for instance the EU debt crisis or from the economical/political reality of US, for instance the 1996 budget crisis of US.

Figures 4.1, 4.2 and 4.3 depict the calculated value of $-\gamma$ of the equation $P(\kappa) = \exp^{-\gamma_{un}\kappa}$ for both index and returns time series. The horizontally lying red line on each graph represents the value of $-\gamma_{un} = -\ln(3/2)$, which corresponds with the value of $-\gamma$ for uncorrelated random series. At this point, we have to remind you that values of $-\gamma > -\ln(3/2)$ are related with a stochastic process where the smaller the value of γ, the higher the degree of correlation of the correlation function. On the contrary, values of $-\gamma > -\ln(3/2)$ are related with chaotic process.

In figures 4.4, 4.5 and 4.6, we plot the values of Hellinger Distance for each distinct period. As was the case in the previous graph, values of Hellinger Distance for the index price and returns of the index is plotted in each graph. The higher the value of the Helliger Distance, the more irreversible a series is and the more non-linear the time series is.

4.1 Stock markets and Efficiency Market Hypothesis

With a glimpse on the figures 4.1, 4.2 and 4.3, it is easy to observe that returns series tend to behave more close to uncorrelated random process than index series in all cases, since their values are more concentrated around the red line. For the return series of S&P500, the average γ value is -0.4146 and γ lies between $-0.50 < \gamma < -0.313$, while for the original index series, the average γ is -0.436 and γ lies between $-0.65 < \gamma < -0.043$. The same happens with S&P100. The average value of γ for the return series of S&P100 is -0.417 and for the index series is -0.423.

We observe that both returns and index series have a γ value that is not coincident with $-\ln(3/2)$. Both series behave as an autocorrelated stochastic process with a low degree of correlation of the correlation function for almost half of win-

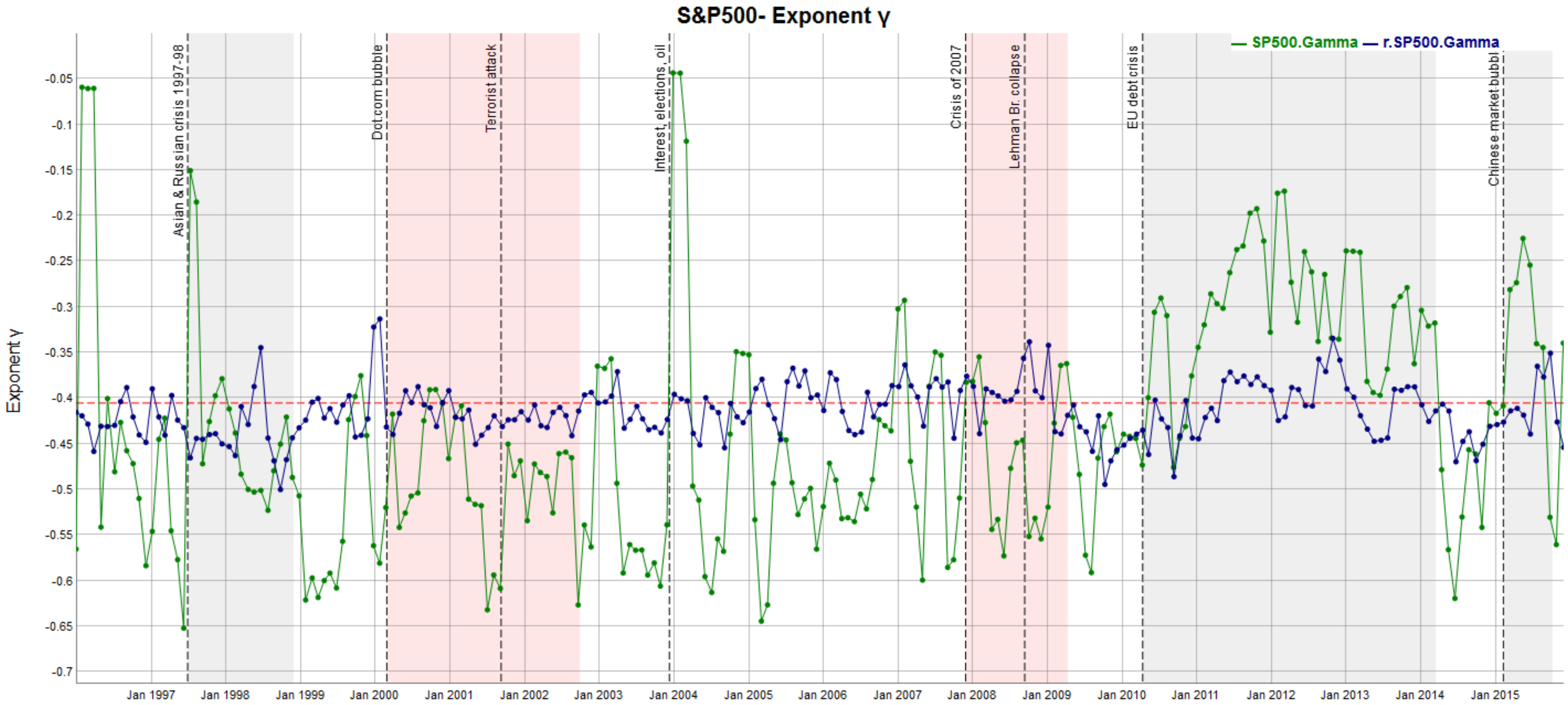


Figure 4.1: Values of exponent γ for all windows of the index (green) and returns (blue) series for S&P 500. In the vertical axis it is plotted the γ values and on the horizontal line we plot the first date of each window. The red line represents the value $\gamma = \ln 3/2$ which correspond to an uncorrelated random process. Red shadow represent periods of market crashes as a consequence of market bubble with grey shadow represents periods of crisis that are not created internally by the market.

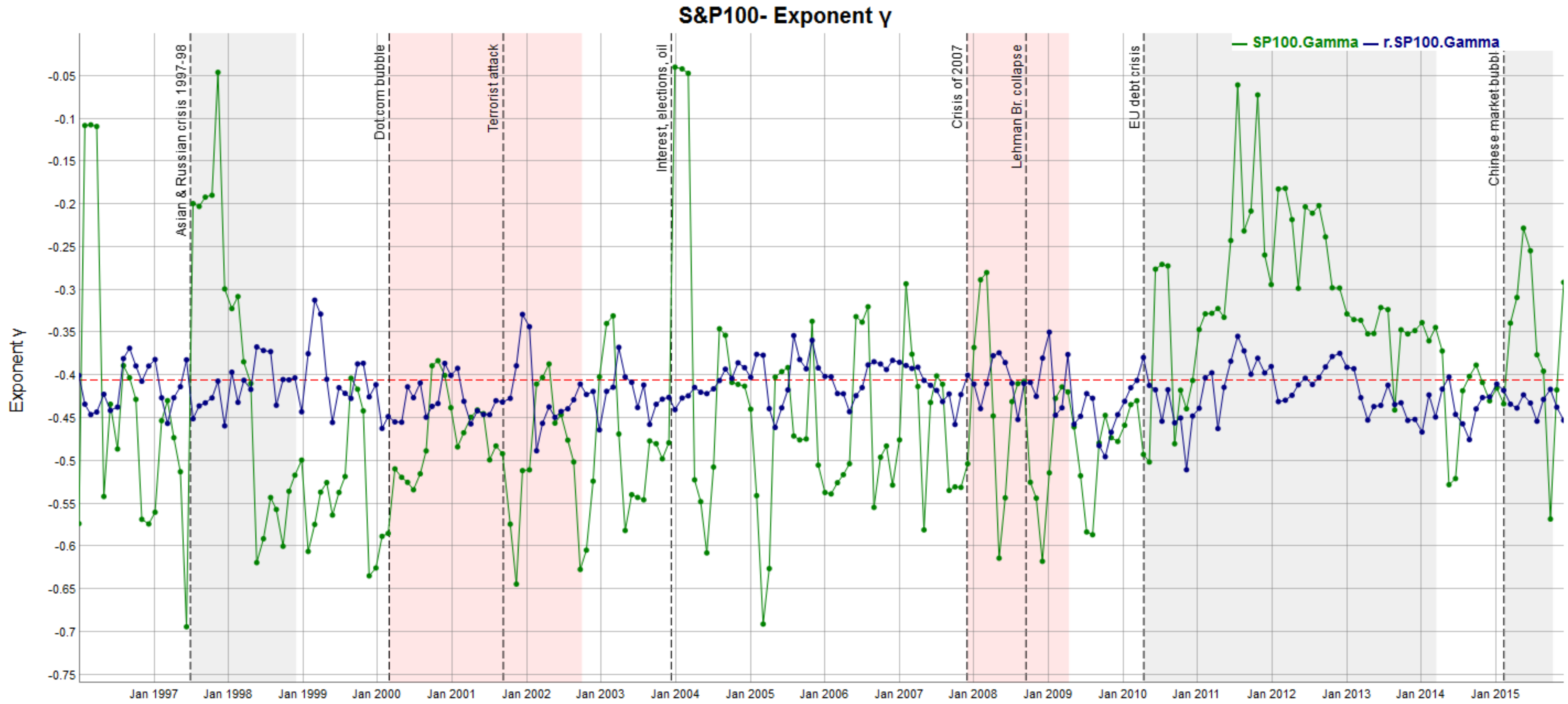


Figure 4.2: Values of exponent γ for all windows of the index (green) and returns (blue) series for S&P 100. In the vertical axis it is plotted the γ values and on the horizontal line we plot the first date of each window. The red line represents the value $\gamma = \ln 3/2$ which correspond to an uncorrelated random process. Red shadow represent periods of market crashes as a consequence of market bubble with grey shadow represents periods of crisis that are not created internally by the market.

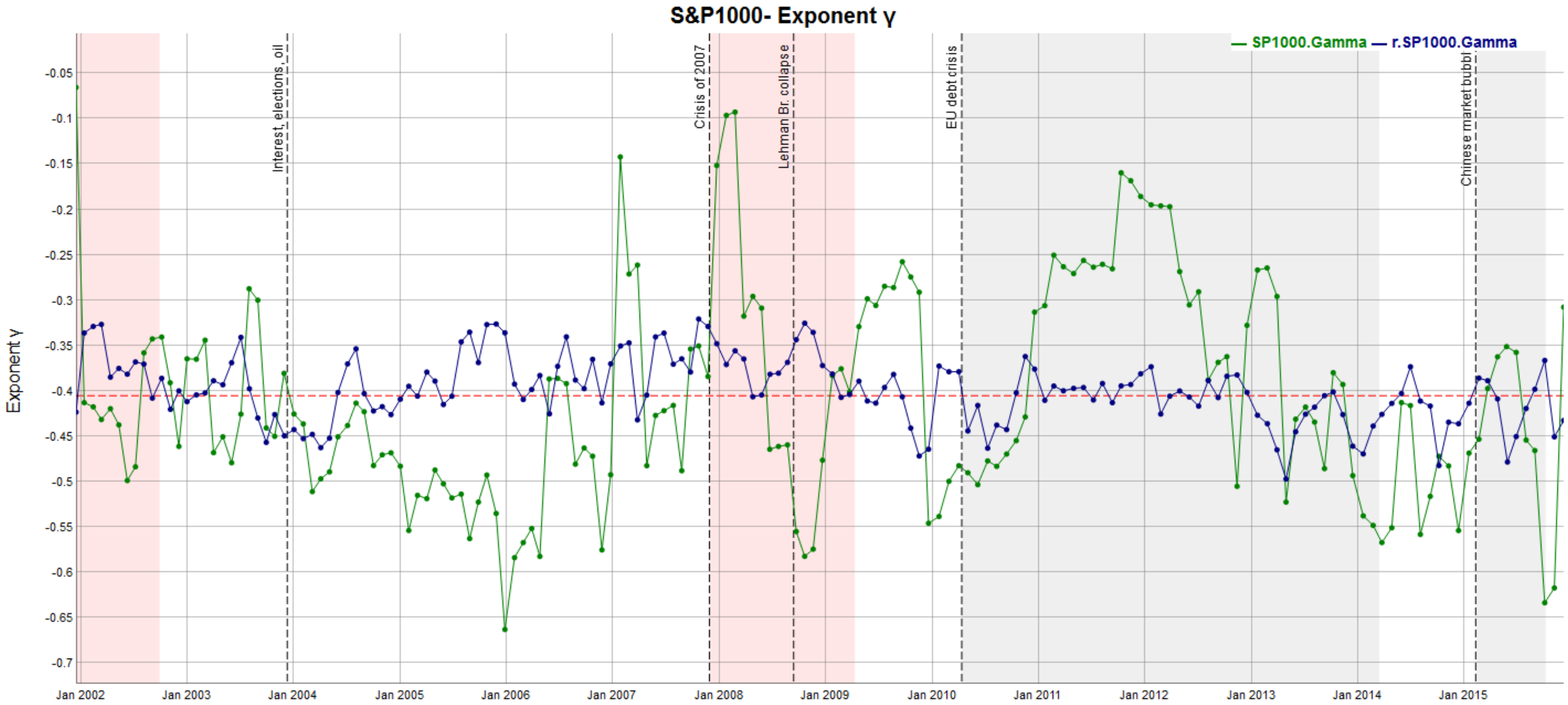


Figure 4.3: Values of exponent γ for all windows of the index (green) and returns (blue) series for S&P 1000. In the vertical axis it is plotted the γ values and on the horizontal line we plot the first date of each window. The red line represents the value $\gamma = \ln 3/2$ which correspond to an uncorrelated random process. Red shadow represent periods of market crashes as a consequence of market bubble with grey shadow represents periods of crisis that are not created internally by the market.

dows,which contradicts the modern financial theory which dictate that price series of stocks have no memory and as a result returns are random. Based on the Efficient Market Hypothesis (EMH) introduced by Professor Eugene Fama, stock prices should reflect instantaneously all available information and thus no one can predict future prices of stocks. Consequently, there is a very close link between EMH and the random walk hypothesis according to which stock market prices evolve forming a random walk and thus cannot be predicted. For those who are not familiar with mathematics, random walk is a path of a successive random steps.

What one should expect is that, returns should behave as a random series. In fact, for S&P500 index returns series, only 29 out of 233 windows' behaviour approaches the one of a random uncorrelated process, 134 windows appear to have the properties of a correlated stochastic process and 72 windows have properties of a process similar to chaos.The corresponding frequencies of returns series of S&P100 are 27 windows that behave close to random process, 145 windows of correlated stochastic behaviour and 61 windows with properties similar to chaos. We know that in both correlated stochastic and chaos processes there is an opportunity for predictions, even short term prediction in case of chaos, which contradicts with EMH.

In addition, we found that there is a strong correlation between the value of γ for the S&P 500 and S&P 100 for the index series. The correlation is 0.82 which is reasonably expected since the index time series are very high correlated. The correlation of γ for the return series of the aforementioned indexes is 0.43 which is much lower than that of the index series, yet considerable one.

4.2 Crisis and index response

In this section, we try to examine if all crises affect stock market the same way. As we already described in section 4, during the period from 1996 to 2016 seven major events have occurred and we would like to examine their effect on the stock market. Two of these events have been created internally from the US stock market i.e. stock market bubble of 2000 and 2007, two events are related with changes in the economic/political reality of US i.e. the budget crisis, and three crises happened outside of the boarders of the country, i.e. the Asian crisis, the EU debt crisis and the Chinese stock market bubble.

What we found is that all crises impact the behaviour of the index even in case that a crisis has been originated in another country. Even though the behaviour of the market is not exactly the same during periods of turbulence, there are many common elements that we can infer on how market react when a crisis occur.

First of all, if we draw our attention to the red and grey shadowed area in figures 4.1, 4.2 and 4.3, it is easy to realize that there are two different patterns in the behaviour of the market series during periods of crises. On the one hand, the market series behave similarly to a chaotic process when an event takes place in the outer environment of the market no matter if this event occurs within the economic/political reality of the country or abroad. On the other hand, when the crisis has been raised internally from the market such as the bubble of 2000 and 2007, market series behave similarly to a correlated stochastic process with a high level of correlation in the correlation function.

In case that an event occurred in the outer environment of the market, we observe that S&P500 appears to behave as a chaotic process. This means that prices become very sensitive since a small change in the initial conditions can cause a very different trajectory in the future. This is the symptom of uncertainty that markets exhibit. Of course, events that take place within the country make prices more sensitive than crises that take place abroad. Notice that the value of exponent γ is much higher early in 1996 and early in 2004 in comparison with the corresponding value of the exponent during European or Chinese crisis.

Detouring for a while of the main topic of this work, we can pay attention to an event that took place in November 2004, the elections in which George Bush was re-elected as president of US. From 08/11/2004 to 12/04/2005 the stock market exhibits a behaviour similar to chaos. This is the period of the first 5 months of the new government and the prices in the stock market seem really fragile. During this period the index is very sensitive and even a small change in the initial condition of the system could trigger a rapid divergence in the system later on. The S&P 500 responded the same early in 2004. At that time, the war in Iraq had just finished, Federal Reserve raised interest rates and the oil hit its maximum price since September 1991. With all these factors changing, the market could not do anything rather than display chaotic dehavior in a very unstable economic environment.

On the other hand, whenever a crisis is created internally by the market, the stock market behaves as a correlated stochastic process. This sounds reasonable if we bear in mind that after a price bubble has burst prices follow a downward-sloping trajectory for a long time. There is no uncertainty during this period because the market itself make a correction in the asset evaluation after a period of inertia. All know that the downward-sloping trajectory will stop as soon as the price of asset represents the real value of equities.

However, in the beginning of 2007 we can observe that there were months in which the market behaves similarly to chaos before the bubble burst. Many people believe that this is related with some early warnings that there was at that time and herald the upcoming crisis.

Series of returns of S&P500 and S&P 100 do not follow the same path as the index series. The correlation between exponent γ of the return series and index series in both cases is 0.165 which is negligible. In fact, a small correlation between return and index series for both indexes arise after the beginning of the EU debt crisis. In total, none of the events under study appear to affect the behaviour of the return series over time.

Regarding the Hellinger Distance which is a quantitative assessment of nonlinearity of a series, it is widely admitted that financial time series are characterized as non-stationary. This is verified also by means of HVG, since the Helliger Distance is higher than zero for all indexes and for the whole period between 1996 and 2016.

In addition, we observe that the original index series tend to be more non-linear in comparison with series of index returns while the most interesting remark is that crisis also do affect the nonlinearity of the index series. Throughout periods of internally created crisis by the market, the nonlinear dynamics of the index series fluctuate at a low level with lower variance. On the contrary, during a crisis that takes place in the external environment of the market, the Helliger Distance fluctuates at higher level than normal while its variance increases considerable.

This outcome is more than expected. The nonlinearities of a correlated stochastic process is much more less than the corresponding of a process that exhibit properties of chaos. We calculated the correlation of γ exponent and Hellinger Distance for all windows for the index series. The correlation is 0.36 and 0.303 for S&P 500 and S&P 100 respectively. The same is not true for returns series where the correlation is less than 0.076 for both indexes.

Most times it is difficult to isolate particular events and examine their effect on the stock market. The investors' behaviour is actually influenced by many different factors that sometimes it is risky to attribute a particular trend of the market to one event. This is the reason why we observe frequently that the behaviour of the index changes remarkably even between two consecutive windows.

4.3 Size of index

This section emphasizes on how the size of an index can affect the behaviour of the index. At this section we refer to the period from 31^{st} December 2001 to 5^{th} April 2016 and we include in this part also S&P 1000.

If we draw our attention on Figure 4.8 we can realize that γ exponent values for S&P 500 and S&P 100 are moving almost together, as we also mentioned above. On the other hand, S&P 1000 is co-moving at a lower degree in comparison with the other indexes, but it fluctuates more than the other indexes. More specifically, observing the trajectory of the γ exponent for each index series it is possible to realize that the smaller the size of the index the less the exponent γ fluctuates. The values of γ of S&P 500 tend to exceeds the corresponding values of S&P 100 and γ value of S&P 1000 exceeds the corresponding value for the other two indexes. By "exceed" we mean that when the value of γ lies in the chaotic band then its values tend to be closer to zero, while in the other case its values tend to be closer to -1.

What we mentioned in the previous paragraph for the value of the exponent γ for the index series can be also applied to the returns series. Nevertheless, this is just a general observation and cannot be formed in the form of a rule. There are windows where this is not true and the value of γ of S&P 100 index series exceed the corresponding value of the other indexes.

The size of the index does not affect only the behaviour of an index but also determines the level of the underlying non-linear dynamics. We found that the higher the size of the index, the higher the nonlinearities on a series for both the index series and the returns series.

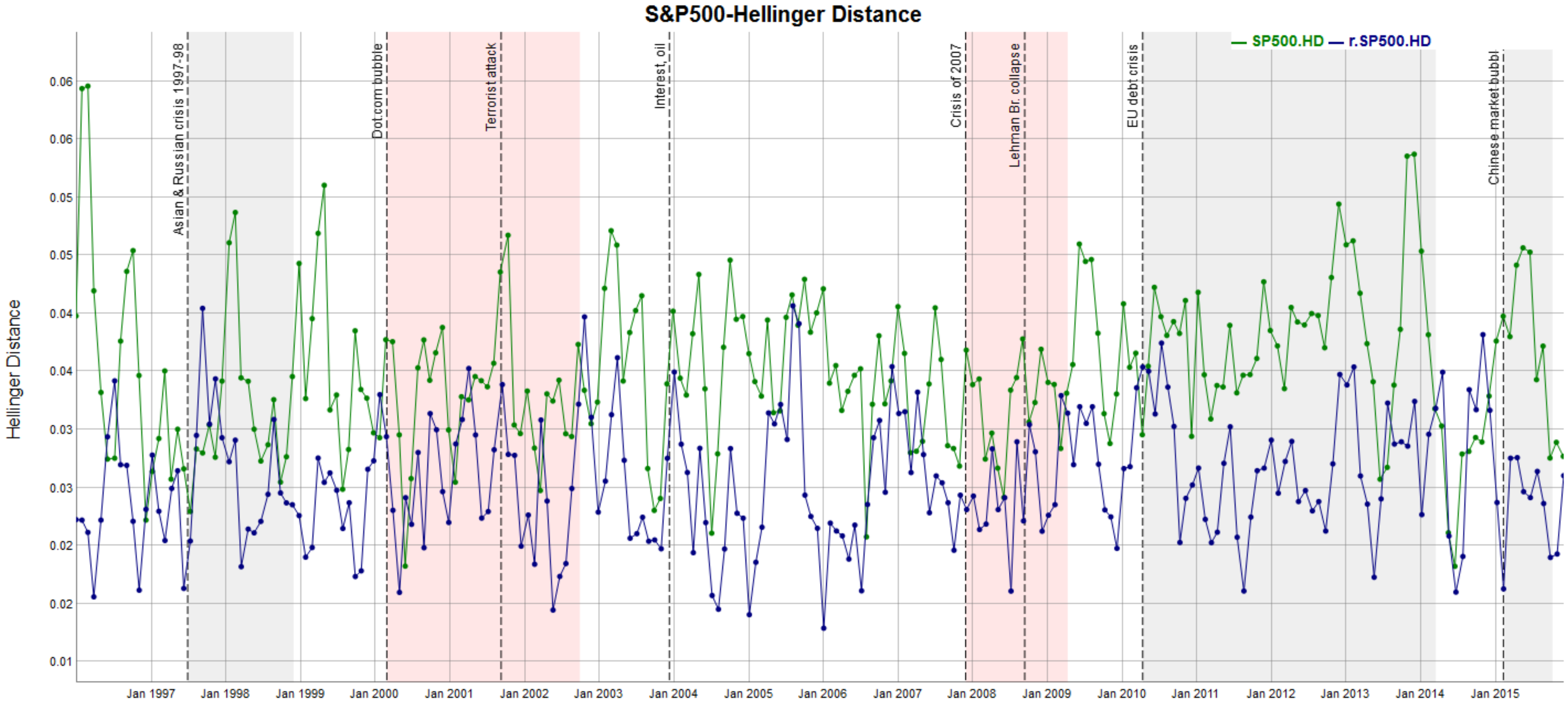


Figure 4.4: Values of Hellinger distance for all windows of the original index (green) and returns (grey) series for S&P 500. In the vertical axis it is plotted the Hellinger distance values and on the horizontal line we plot the first date of each window. Red shadow represent periods of market crashes as a consequence of market bubble with grey shadow represents periods of crisis that are not created internally by the market.

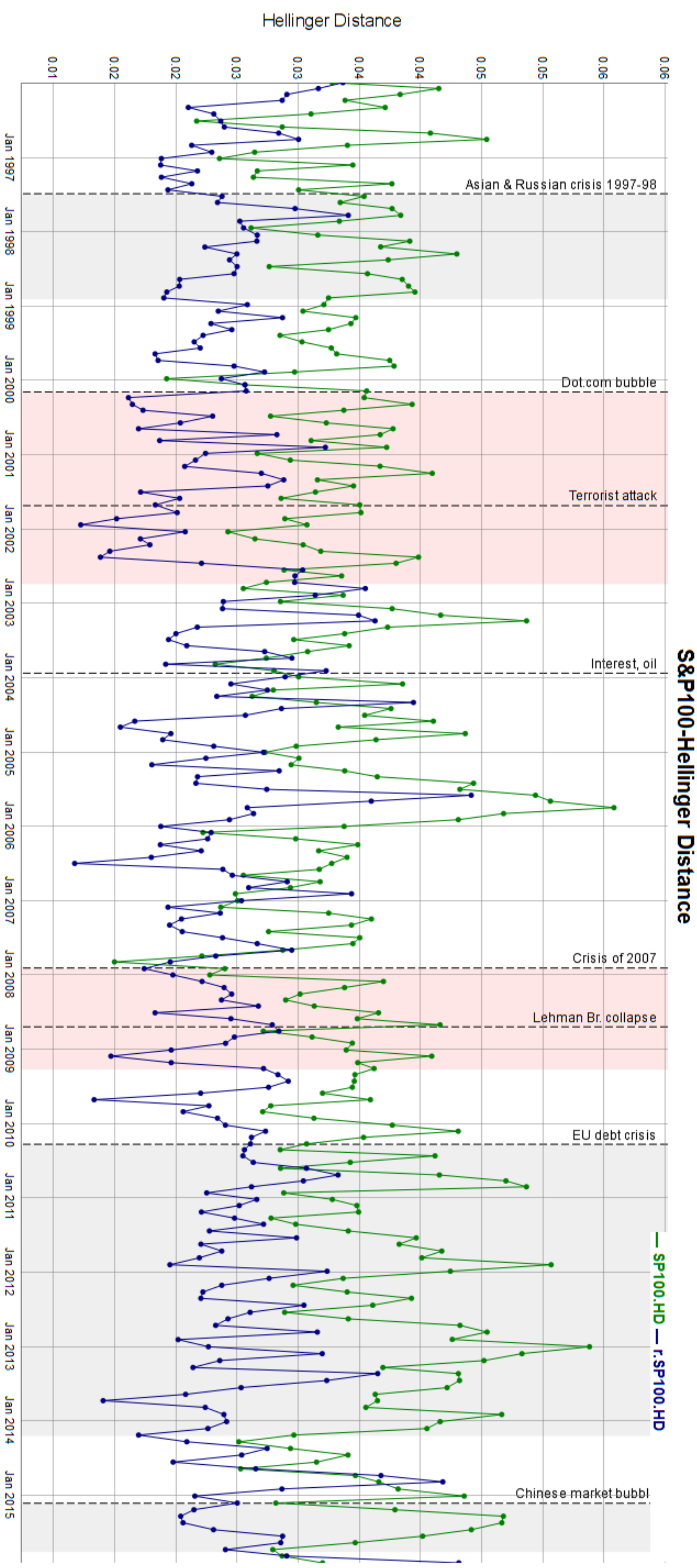


Figure 4.5: Values of Hellinger distance for all windows of the original index (green) and returns(grey) series for S&P 100. In the vertical axis it is plotted the Hellinger distance values and on the horizontal line we plot the first date of each window. Red shadow represents periods of market crashes as a consequence of market bubble with grey shadow represents periods of crisis that are not created internally by the market.

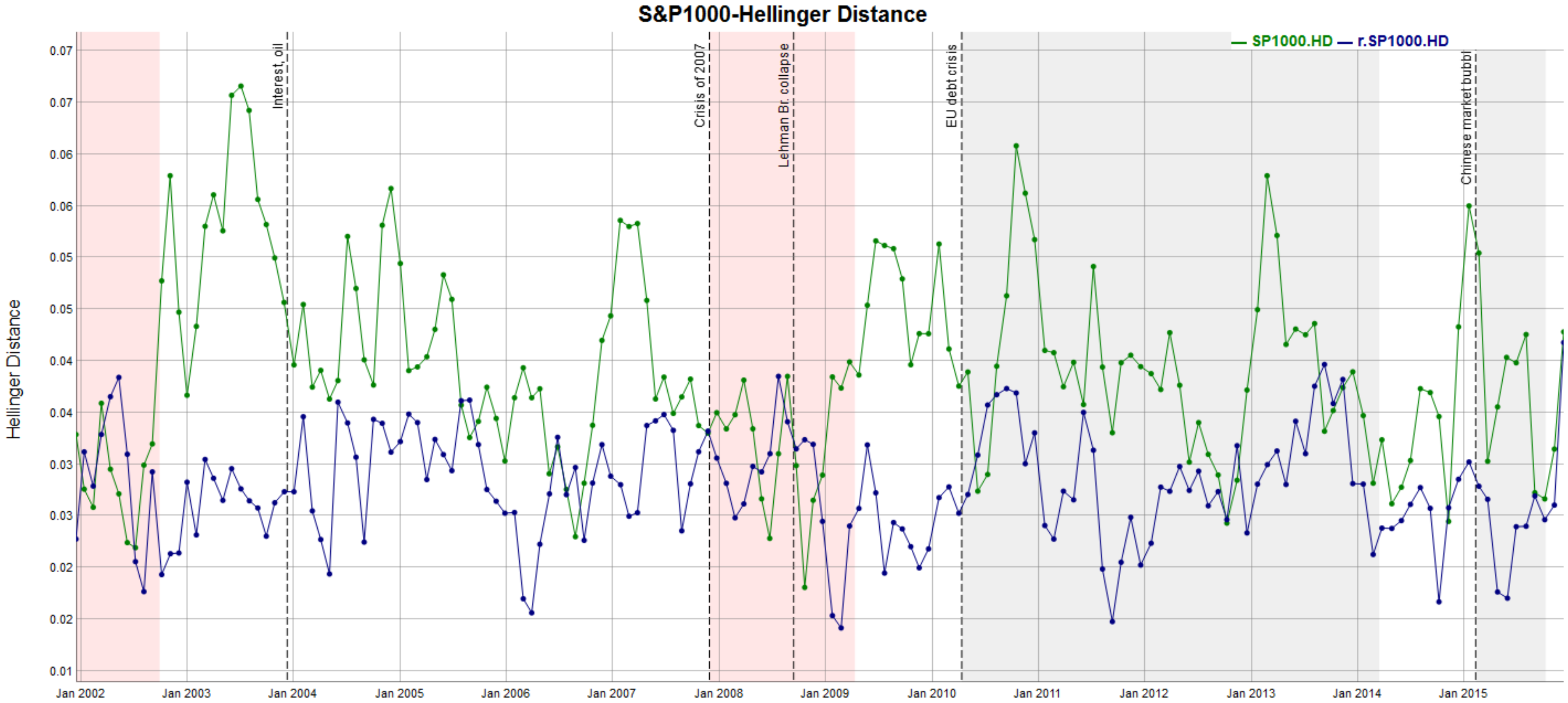


Figure 4.6: Values of Hellinger distance for all windows of the original index(green) and returns (grey) series for S&P 1000. In the vertical axis it is plotted the Hellinger distance values and on the horizontal line we plot the first date of each window. Red shadow represent periods of market crashes as a consequence of market bubble with grey shadow represents periods of crisis that are not created internally by the market.

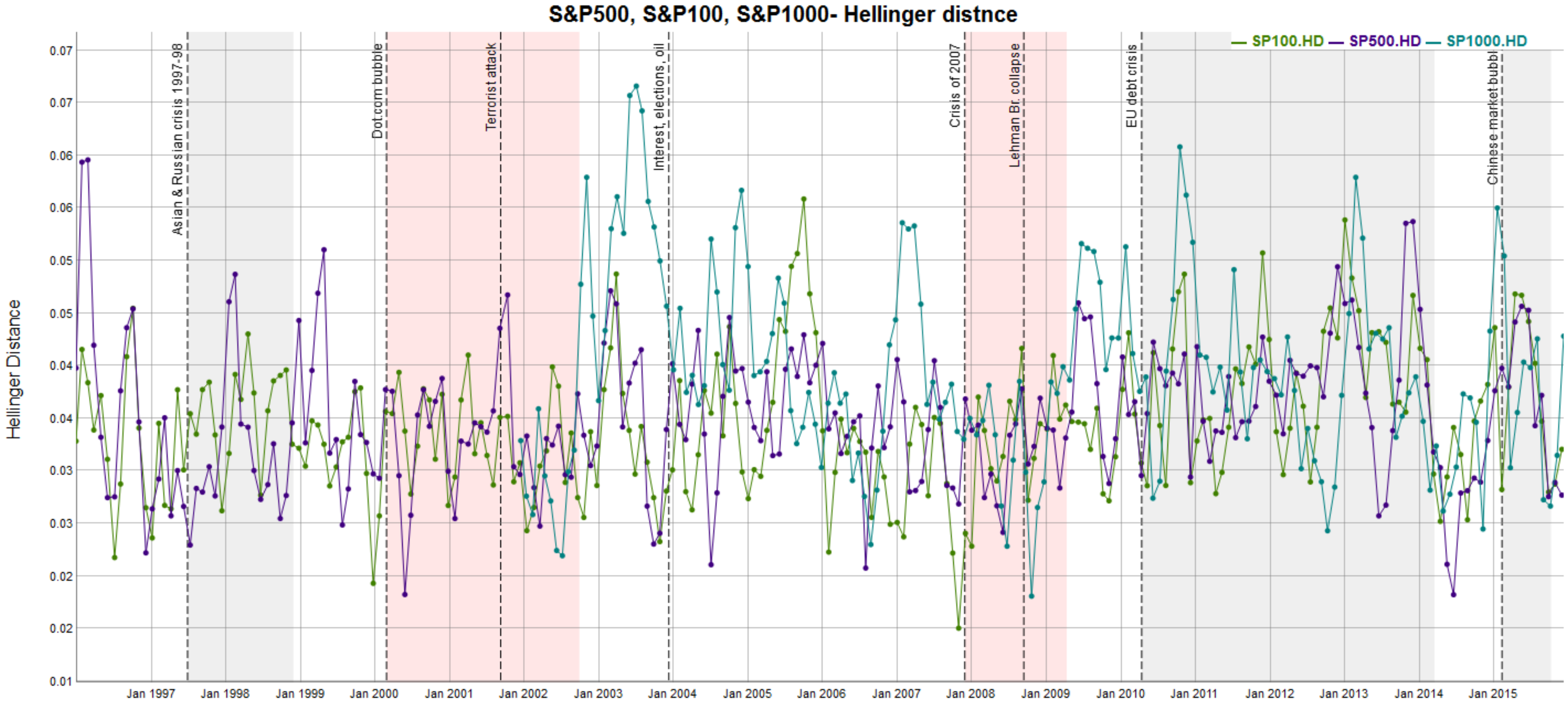


Figure 4.7: A comparative representation of the value of exponent γ for S&P 500 (green), S&P 100 (purple) and S&P 1000 (turqouise) index series for the whole period. Red shadow represent periods of market crashes as a consequence of market bubble with grey shadow represents periods of crisis that are not created internally by the market.

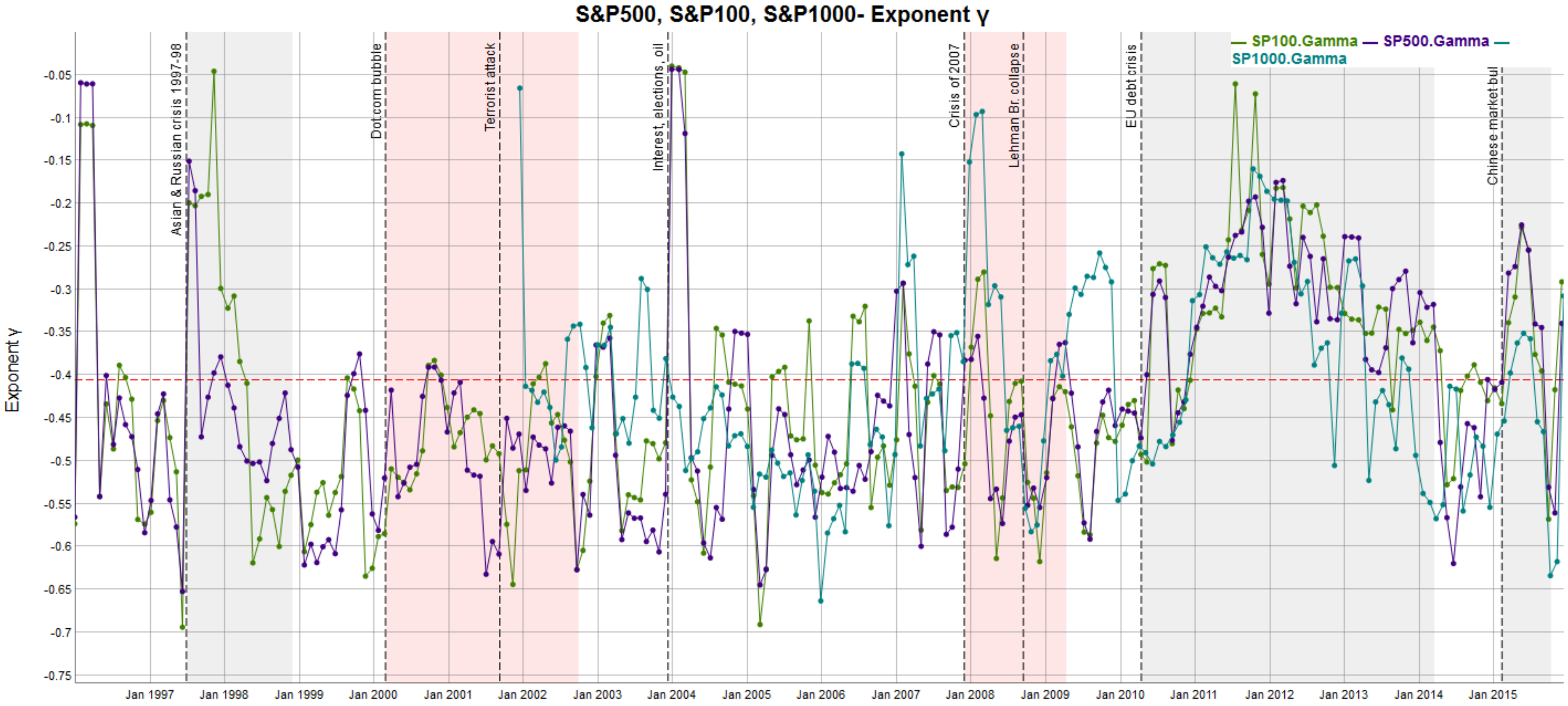


Figure 4.8: A comparative representation of the value of exponent γ for S&P 500 (green), S&P 100 (purple) and S&P 1000 (turqouise) index series for the whole period.Red shadow represent periods of market crashes as a consequence of market bubble with grey shadow represents periods of crisis that are not created internally by the market.

Chapter 5

CONCLUSIONS

In the previous chapters we tried to implement visibility graph methodology in order to analyze how particular periods of crisis affected the US stock markets. More specifically, using observations with a frequency of five minutes for the period 1996-2016 for the indexes S&P 100, S&P 500 and S&P 1000 and by using the method of rolling window, we investigated how the behaviour of the indexes changes over time and how crisis affect these indexes.

We found that all crises that occurred during this period indeed have affected US stock market, regardless if the crisis took place within the boarders of US or abroad. Interesting enough is the fact that all indexes behave as a correlated stochastic process with diminishing non-linearities on the index series during the period after a stock market bubble has burst, while in all other crisis events the behaviour was similar to a chaotic process with non-linearities increasing. On the other hand, returns series of the indexes do not seem to be particularly affected by the existence of a crisis and there is no correlation between the bahavior of returns and index series. The size of the index can explain a different behaviour among all indexes of the same stock market. The more broad the index is, the more extreme values the γ exponent reaches and the more non-linearities appear in the index series.

However, there are still issues that require deeper research. It would be interesting to examine if the results found in this paper are statistically significant ad these issue will be studied in a future paper. An expansion of this research into all US stock indexes would be also beneficial. This process could disclose how each industry in the country has been affected by crisis event and help investor develop different approaches.

Exept for the Visibility Graph theory that introbuced in 2008, all the coding, programming, data selection and analysis has been conducted by the author of this work.

BIBLIOGRAPHY

[1] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

[2] Guillermo A Calvo, Alejandro Izquierdo, and Luis-Fernando Mejia. On the empirics of sudden stops: the relevance of balance-sheet effects. Technical report, National Bureau of Economic Research, 2004.

[3] Stijn Claessens and Mr M Ayhan Kose. *Financial Crises Explanations, Types, and Implications*. Number 13-28. International Monetary Fund, 2013.

[4] David Easley and Jon Kleinberg. *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, 2010.

[5] Jeffrey A Frankel and Andrew K Rose. Currency crashes in emerging markets: An empirical treatment. *Journal of international Economics*, 41(3):351–366, 1996.

[6] Gregory Gutin, Toufik Mansour, and Simone Severini. A characterization of horizontal visibility graphs and combinatorics on words. *Physica A: Statistical Mechanics and its Applications*, 390(12):2421–2428, 2011.

[7] Lucas Lacasa and Ryan Flanagan. Time reversibility from visibility graphs of nonstationary processes. *Physical Review E*, 92(2):022817, 2015.

[8] Lucas Lacasa, Bartolo Luque, Fernando Ballesteros, Jordi Luque, and Juan Carlos Nuno. From time series to complex networks: The visibility graph. *Proceedings of the National Academy of Sciences*, 105(13):4972–4975, 2008.

[9] Lucas Lacasa, Bartolo Luque, Jordi Luque, and Juan Carlos Nuno. The visibility graph: A new method for estimating the hurst exponent of fractional brownian motion. *EPL (Europhysics Letters)*, 86(3):30001, 2009.

[10] Lucas Lacasa, Angel Nunez, Édgar Roldán, Juan MR Parrondo, and Bartolo Luque. Time series irreversibility: a visibility graph approach. *The European Physical Journal B*, 85(6):1–11, 2012.

[11] Lucas Lacasa and Raul Toral. Description of stochastic and chaotic series using visibility graphs. *Physical Review E*, 82(3):036120, 2010.

[12] Bartolo Luque, Lucas Lacasa, Fernando Ballesteros, and Jordi Luque. Horizontal visibility graphs: Exact results for random time series. *Physical Review E*, 80(4):046103, 2009.

[13] Maarten Van Steen. Graph theory and complex networks. *An introduction*, 144, 2010.