DOCTOR OF ECONOMICS AND BUSINESS MANAGEMENT

Essays in Econophysics and Applied Econometrics
Modeling Complexity in Finance

Dahlqvist, Carl-Henrik

Award date:
2018

Awarding institution:
University of Namur
Université Catholique de Louvain

Essays in Econophysics and Applied Econometrics: Modeling Complexity in Finance

Carl-Henrik Dahlqvist

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

Louvain School of Management

Louvain Finance

UNIVERSITÉ DE NAMUR

Faculté des sciences économiques et de gestion

Center for Research in Management and Finance

Thèse présentée en vue de l'obtention du grade de docteur en

Sciences économiques et de gestion

Tuesday 4th September, 2018

Graphisme de couverture : ©Presses universitaires de Namur

Rempart de la Vierge, 13
B - 5000 Namur (Belgique)

Imprimé en Belgique
ISBN : 978-2-39029-015-5
Dépôt légal: D/2018/1881/20

Numéro de thèse Louvain School of Management: 3|2018

Doctoral Committee

Supervisors

Professor Sophie Béreau, Université catholique de Louvain
Professor Jean-Yves Gnabo, Université de Namur

Jury members

Professor Annick Castiaux, Université de Namur
Professor Marco Saerens, Université catholique de Louvain
Dr Renaud Lambiotte, Somerville College, Oxford

President

Professor Isabelle Linden, Université de Namur

In loving memory of my mother, Ghislaine Massin,
who continues to be a major source of inspiration,
like a distant star still shining long after it disappeared.

Contents

Acknowledgments **iii**

Introduction **v**

1 Testing causality in financial time series **1**
1.1 Introduction 1
1.2 On causality detection 4
1.2.1 From Granger representation to entropy and transfer entropy measures 4
1.2.2 Transfer entropy as a nonlinear measure 9
1.2.3 Estimation and statistical inference 10
1.3 Simulations and first conclusions 13
1.3.1 Data generating processes 14
1.3.2 Results 17
1.4 Application to European and U.S. financial sector 25
1.4.1 Motivations 25
1.4.2 Data description 26
1.4.3 Network inference and topological measures estimations 26
1.5 Conclusion 33

2 Effective Network Inference Algorithm **35**
2.1 Introduction 35
2.2 Network structure, information decomposition, and indirect links 38
2.3 Network inference algorithm 40
2.3.1 Extension 46
2.4 Simulation 47
2.5 Application to the US financial sector 54
2.5.1 Topological measures estimation 55
2.5.2 Financial institutions' risk and topological measures 58
2.5.3 Empirical comparison with pairwise Granger causality 61
2.6 Conclusion 62

3 Multichannel Information Transfer Estimation 65
3.1 Introduction . . . 65
3.2 Granger based multi-channel causality detection . . . 67
3.2.1 Model description . . . 68
3.3 Transfer entropy based multi-channel causality detection . . . 72
3.3.1 Model description . . . 73
3.4 Simulation . . . 77
3.5 Application to Financial institutions . . . 87
3.5.1 Empirical data and network inference . . . 89
3.5.2 Mutli-channel causality detection in the frequency domain . . . 89
3.5.3 Quantile mutli-channel causality detection . . . 92
3.5.4 Cross-quantile mutli-channel causality detection . . . 94
3.5.5 Mutli-channel causality detection . . . 95
3.6 Conclusion . . . 98

4 Cross-country Information Transmissions 101
4.1 MultiChannel Markov Switching Granger Causality . . . 104
4.2 Results and Discussion . . . 108
4.2.1 Cross-country information transmissions . . . 108
4.2.2 Data . . . 110
4.2.3 Analysis of Cross-country information transmissions . . . 110
4.2.4 Commodity markets and Cross-country information transfers . . . 118
4.2.5 Analysis of commodity markets and Cross-country information transfers . . . 120
4.3 Conclusion . . . 124

Conclusion 127

Appendix A.Testing causality in financial time series 155

Appendix B. Effective Inference Algorithm 157

Appendix C. Cross-country Information Transmissions 165

Appendix D. Summary of main contributions and publication status 169

Acknowledgments

Lorsque j'ai décidé de quitter mon emploi d'analyste financier chez SWIFT pour débuter ma thèse de doctorat en Finance il y a 5 ans, je ne pensais pas qu'elle me changerait autant. Elle m'a tout d'abord permis d'acquérir de nombreuses connaissances mais également et surtout de développer une véritable approche scientifique via le développement d'un esprit critique mais ouvert et l'affermissement de certaines qualités telles que la rigueur ou la créativité. La formation doctorale permet en effet de développer un large éventail de compétence du aux multiples aspects d'une recherche scientifique: esprit de synthèse et critiques constructives lorsque l'on s'approprie un sujet via une revue de la littérature; créativité, rigueur et humilité lorsque l'on produit de nouvelles connaissances ; pédagogie, discernement et confiance en soi lorsque l'on présente ses résultats via une présentation orale ou la soumission d'un article. Mais la réalisation d'une thèse ne se résume pas seulement au développement de qualités personnelles, il s'agit d'un partage d'idées entre différents acteurs, comme les promoteurs, les lecteurs, les collègues, les personnes assistants aux séminaires ou aux conférences. Tous ces commentaires et suggestions sont nécessaires pour faire évoluer un travail de recherche. Une thèse de doctorat est donc surtout une entreprise collaborative. Il est important de mettre en évidence l'apport de tous ces acteurs à la réalisation de ce travail.

Je tiens donc tout d'abord à remercier mes deux promoteurs Sophie Béreau et Jean-Yves Gnabo d'avoir accepté de m'encadrer durant la réalisation de cette thèse mais également pour leurs nombreux conseils, leur soutien et leur bienveillance. Ce fut un privilège de pouvoir collaborer avec des personnes brillantes mais aussi très humaines. Je remercie tout particulièrement Jean-Yves pour toute l'aide qu'il m'a apportée lors de la réalisation et soumission de certains chapitres qui m'ont beaucoup fait évoluer au niveau de l'élaboration de contenus scientifiques rigoureux. Je le remercie également pour son soutien lorsque j'ai décidé d'entreprendre parallèlement à ma thèse un Master en Physique. Je remercie également les membres de mon jury et de mon comité d'encadrement Marco Saerens, Annick Castiaux, Renaud Lambiotte et Isabelle Linden qui m'ont fait profiter de leur expertise et de leurs précieux conseils.

Je remercie l'Université de Namur qui m'a permis de réaliser cette thèse en me finançant via mon poste d'assistant. Je remercie ainsi l'équipe didactique ainsi que les professeurs Oscar Bernal, Pierre Giot et Dominique Helbois pour leur aide et tout ce que l'encadrement des travaux pratiques m'a apporté. Cette thèse ayant été effectuée en co-tutelle, je tiens à remercier également l'Université catholique de Louvain et Louvain Finance pour leur soutien et la richesse de leur parcours doctoral. Je remercie enfin Sandrine Delhaye et Pierrette Noel pour leur gentillesse et leur aide dans toutes mes démarches administratives.

Je souhaite ensuite remercier l'ensemble de mes proches pour leur soutien durant toutes ces années; tous mes amis, mes collègues Gregory, Nicolas, Marco, Donald, Cyril, Eli, Olivier, Christian, Jolan, Modeste ou encore Alexandre. Mais j'ai surtout une pensée particulière pour ma petite maman qui a vécu le début de cette aventure mais qui malheureusement n'est plus là pour son aboutissement. Je lui dédie ce travail en souvenir de tous les moments de bonheur que nous avons eu la chance de vivre ensemble mais également pour toutes les valeurs qu'elle m'a transmises et qui me guident aujourd'hui. Je remercie également le reste de ma famille, mon frère qui est toujours extrêmement enthousiaste et intéressé par ce que je fais mais également mon papa et mes deux soeurs qui m'ont toujours soutenu. Et comme on garde toujours le meilleur pour la fin, je remercie surtout Virginie qui est la condition de ma réussite. Car c'est elle qui m'a soutenu le plus, subissant les longues journées de travail, les sessions d'examen, mon stress et parfois ma mauvaise humeur. Elle a toujours su me rassurer et me faire voir les choses du bon côté dans les moments difficiles. Merci d'être là!

Introduction

General Introduction

Historical development of the notion of Causality

The concept of causality has been, in the past decades, the subject of many debates in all fields of Sciences, as the development of new tools and theories highlighted the complexity of Nature. As new tools were developed, the increasingly detailed representation of our environment asked for new methodologies able to provide a deeper comprehension of the underlying mechanisms that govern Nature. The concept of Causality and even its existence has become a crucial question in modern Science. We propose in a series of essays to contribute to this growing literature by looking at the question of causal relationships identification and characterization in the field of Finance and Econometrics. Our approach will be both methodological and empirical with the development of new methodologies to infer causal links, and their application to financial time series to better apprehend the relationships between agents in financial markets.

The notion of causality is particularly difficult to grasp as its definition is multiple and evolved in time. The first proper theory of causality dates back to Aristotle who conditioned the knowledge of a thing on the definition and understanding of its causes [1]. In his Physics II, Aristotle highlighted, based on critical examination of the work of his predecessors and the study of natural phenomena, four types of causes that underlie everything that requires an explanation, from natural phenomena to human action (Falcon, 2011):

- the material cause ("that out of which something is done")
- the efficient cause ("the primary source of change or rest of something done")
- the formal cause ("the form of what is to be done")
- the final cause ("that for the sake of which something is done")

[1] Posterior Analytics, *APost.* 71 b 9-11., 94 a 20

The final cause has been introduced to account for the regularities observed in Nature and the adaptations found throughout the fauna and the flora. The Aristotelians were particularly concerned with the goal-directed process which could have produced these regularities and adaptations to the environment. This theological approach has then been central during the Medieval time where God, as the creator of our World, was always considered as the main actor. So the question of causality could often reduce to knowing if God could have created a World in which a specific causal relationship could be observed. A good illustration is given by the work of Thomas Aquinas who proposed a specific hierarchy for Aristotle causes: first the final causes, then the efficient causes followed by the material causes and finally the formal causes (May, 1970).

During the scientific revolution initiated partly by the work of Francis Bacon, the meaning of the word "cause" changed and narrowed to the concept of efficient cause which was more convenient for scientific investigation. David Hume gave the first modern definition of causality using six assertions(Hume, 1738):

- "The cause and effect must be contiguous in space and time."
- "The cause must be prior to the effect."
- "There must be a constant union between the cause and effect"
- "The same cause always produces the same effect, and the same effect never arises but from the same cause. This principle we derive from experience, and is the source of most of our philosophical reasoning."
- "Where several different objects produce the same effect, it must be by means of some quality, which we discover to be common amongst them."
- "The difference in the effects of two resembling objects must proceed from that particular, in which they differ."

Nevertheless, the position of Hume on the ability of human to have certain knowledge of causal relations was that there is nothing more to causality than the regular sequence of phenomena and that such a regular sequence cannot give a necessary connection; in consequence certainty could not be reached (White, 2013). Despite the fact that for Hume neither logic nor experience gives us secure grounds from causality inference, his assertions have inspired the different definitions of causality we use today, such as the two principles behind the definition of Granger causality (Granger, 1969), the cause happens prior to its effect and the cause has unique information about the future evolution of its effect. We will come back to this specific definition of causality, as it will represent the starting point of many of the development proposed in this thesis.

Other approaches have been derived from the theory developed by Hume, such as the counterfactual theory of causation whose main idea was proposed by Hume itself in his Enquiry Concerning Human Understanding (Hume, 1748): "We may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second". Nevertheless, Hume never really explored this alternative counterfactual approach to causation. The main idea behind the counterfactual causation is that causal link may be defined in terms of counterfactual conditionals of the form "If A had not occurred, B would not have occurred". The main difficulty with this theory of causation is the definition of suitable counterfactuals.

Closer to the principles used in this thesis, the probabilistic theory of causation has benefited, in the past decades, from a growing interest among researches in disciplines such as statistic, Economics, machine learning or artificial intelligence (Berzuini et al., 2012), as it eased causality inference by reducing, to a certain extend, causality to probabilities. According to researches such as Suppes (1970), this theory defines a causal link in terms of the ability of a cause to increase the probability of observing its effects, everything else being equal. However, for Pearl (2000a) this vision is too close to the counterfactual approach and shares the same problems. Therefore Pearl (2000a) proposes instead to replace the inequality $P(\text{effect} \mid \text{cause}) > P(\text{effect})$ by $P(\text{effect}|\text{do}(\text{cause})) > P(\text{effect})$. This means that instead of simply comparing the probability of occurrence of the effect when the cause is observed, we should rather consider the probability of observing the effect when the cause in voluntarily applied and not simply passively observed. We can derive from Pearl's vision the famous quote that "correlation does not imply causation" [2], in other word, counting the number of lung cancer cases among smokers and the ones among non-smokers is not enough to infer a causal link. Randomized experiments are needed to be able to infer a proper causal link.

Causality in Sciences

As the earliest definitions of causality have been derived to study the fundamental laws of Nature, we propose to start our investigation by looking at how causality is defined and used in Physics. We will then move to other fields of Science and eventually examine in more details its use in Finance and Econometrics. This broad analysis will help us to understand how these precursory approaches to causality have influenced its conceptualization in Finance and Econometrics.

[2]The phrase is generally attributed to Karl Pearson

The explanatory ambition of physical sciences accounts for the ancient idea that everything may be explained by understanding the underlying laws of Nature. This idea led to the concept of causal determinism which implies that once the laws of Nature have been defined, every event may be explained by antecedent events together with a set of laws of Nature. As stated by Laplace (1820): "We ought to regard the present state of the universe as the effect of its antecedent state and as the cause of the state that is to follow. An intelligence knowing all the forces acting in nature at a given instant, as well as the momentary positions of all things in the universe, would be able to comprehend in one single formula the motions of the largest bodies as well as the lightest atoms in the world, provided that its intellect were sufficiently powerful to subject all data to analysis; to it nothing would be uncertain, the future as well as the past would be present to its eyes. The perfection that the human mind has been able to give to astronomy affords but a feeble outline of such an intelligence."

Nevertheless, beyond theological consideration, 19th and 20th century mathematical studies showed convincingly that neither a finite, nor an infinite but embedded-in-the-world intelligence may have the computing power necessary to predict future states of the entire world (Hoefer, 2016). Moreover, the development of Chaos theory which demonstrates the sensitiveness of certain types of determinist causal structures on initial conditions questioned the possibility of predicting future states for even small systems. Another salient feature of Chaos theory is the fact that the outcome of a chaotic system even if described by a deterministic causal structure could effectively mimics a random or stochastic process. It makes it very difficult, even impossible, to infer from certain systems a proper causal structure as in some instance we cannot formally distinguish a deterministic system from an indeterministic one. This difficulty has been raised by Suppes (1993) who states that: "There are processes which can equally well be analyzed as deterministic systems of classical mechanics or as indeterministic semi-Markov processes, no matter how many observations are made."

It implies that we may never be able to fully explain the behavior of our universe or even detect certain types of causal patterns. This leads to the question of causality legitimacy in Physical theory. Since the famous quote of Russel in 1913 "The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm", issue have been raised on the actual existence of causality in Physics. Indeed, Norton (2003) argues that causation even fails to hold for simple systems describe by Classical mechanics. He considers a ball at the top of a frictionless dome whose motion equation is dependent on the radial distance from the top. He argues then that following the first two laws of Newtonian mechanics that there exists more

than one solution [3], meaning that the system fails to be deterministic or causal. [4]

However, Zinkernagel (2010) demonstrates in his paper that the considered mass does not respect the first law of Newtonian mechanics in the demonstration proposed by Norton and advocate that there exists a clear 'causality content' in Newton's first two laws: "A body in uniform motion continues its motion unless the body is caused (by a force) to change its motion (accelerate). The same causes (forces) acting in the same circumstances will have the same effects." The debate is still open, but many researchers shared the vision of Zinkernagel.

In special relativity, most of Newtonian mechanics is still valid as long as the effect of speed is taken into account via the Lorentz transformation. Therefore, special relativity shares the causal content of Newtonian mechanics as both theories have the same static, unchanging space-time structure. In addition, the fact that no object could travel faster than the speed of light restricts the causes to be considered for a specific effect. Indeed, causality in Special relativity implies that a cause cannot generate an effect which is not in his front light cone (future) in the Minkowski representation. Therefore causal influence cannot travel faster than the speed of light and can occur only from the past to the future.

To get a clearer view of the implication of special relativity on the importance of causality in the comprehension of our environment, let's look in more details at the effect of speed on the relative distance and time between events seen by different observers. As in the Newtonian mechanics, standard rods and clocks may be defined in special relativity. These standard measures permit the definition of time intervals and distances for observers with relative speed. Through the effects described by special relativity, distances and time intervals between the same two events varies from one observer to the other depending on their speed. Even the order between events could change in special cases.

The introduction of the concept of space-time is therefore necessary to reach an objective reality, i.e. a universal agreement about distance and time between events. Indeed, when using the definition of space-time, we get only one position and one time for every events which allows to reach, through the definition of space-time intervals, a universal agreement about the sequence of event. Space-time intervals are defined

[3] The equations possible outcomes are not confined to the ball staying at rest on the top of the dome, the ball falling is also a valid solution of the equations.

[4] Newton first two laws state that: "Every object in a state of uniform motion tends to remain in that state of motion unless an external force is applied to it" and that "The relationship between an object's mass m, its acceleration a, and the applied force F is $F = ma$".

by the following equation: $S.I. = (dx^2 - (cdt)^2)$ which includes a subtraction meaning that we could get a positive, zero or negative space-time interval. If we get a positive space-time interval, this means that no information can go from one event to the other and we cannot reach universal agreement implying that there will always be observers who disagree on which events happens first. If zero or negative, it means that there can be a transfer of information between the events and everyone agree on the sequence between these events[5]. This suggest that even if we do not agree on distance, time, past, future or present, the only thing that we could know for sure about two events is their ability to influence each other. This means that causality which corresponds to negative space-time interval is the only thing that we could agree on. Time is therefore not responsible for causality but causality is responsible for time as we agree on temporal distance only thanks to causality. Causality has therefore a central role in this space-time vision of reality (Winnie, 1977).

The other major theory of the 20th century in physics is Quantum mechanics which raised different questions about causality due to its probabilistic approach. At first sight this theory seems to be acausal as the same cause could lead to different outcomes. But we know from non-relativistic Quantum physics that a physical system is described by wavefunction resulting from the Schrödinger equation which is fully deterministic. The final outcome nevertheless depends on the measurement of the system. Physicists have postulated that this measurement led to an indeterministic collapse of the wavefunction with probabilities for the different outcomes (Hoefer, 2016). The Probabilistic theory of causation tells us then that it should exist a causal structure, as outcomes without any causes could not be predicted even probabilistically. As Pratt (2003) states: "quantum systems certainly behave unpredictably, but if they were not subject to any causal factors whatsoever, it would be difficult to understand why their collective behavior displays statistical regularities". This raises the question of determining the existence or not of a definite causal structure pre-existing the measurement. The theorems developed by Kochen and Specker (1967), and by Bell (1964), the famous Bell inequallities, imply that quantum mechanics is incompatible with physical observables possessing pre-existing hidden values determining the potential outcome of the physical observables measures. But does it hold for causality structure (Brukner, 2014)? Hardy (2005) developed a new framework to address this question by defining causal structures that are both dynamical to account for the effects of general relativity, and indefinite like quantum observables. This new framework is based on the definition of the 'causaloid' which represents a mathematical object regrouping information about the structure of causal relations between two point of space-time.

[5]This result confirm the fact that an effect has to be in the light cone of its cause

Despite the different theories addressing the difficult question of measurement, the question is still open and affects other field of Science such as Neuroscience where it relates to the notion of consciousness. Indeed as mentioned earlier, the standard interpretation of quantum physics implies that all objects in our environment exist in an objective, unambiguous state only when they are being observed, measured (Pratt, 2003). Many scientist and philosopher highlighted the importance of consciousness to solve the measurement problem (Wigner (1962); Eccles and Popper (1977); Toben (1974); Faber (1986) to quote only a few). The specific outcome is causally determined by the measurement done by a conscious observer. A large part of the literature has been devoted to the relationship existing between the concept of causality and processes of consciousness (see among others Kafatos and Nadeau (1990); Corbi et al. (2000); McGinn (2004)). Despite the fact that this standard interpretation appeals to notions too anthropomorphic (i.e. measurement and consciousness) and seems too ad hoc to be a fundamental law of nature; it nevertheless influenced numerous researchers in their quest to define the relationships between mind and brain and how classical and quantum causality work at both levels (see Wurzman and Giordano (2009); Thompson (1990))

Beyond these philosophical aspects of Neuroscience, other researches have been devoted to the understanding of the processes at work to acquire, maintain, and utilize an up-to-date system of causal information and to infer, from this system, effects from potential causes (Patterson and Barbey, 2012). Closer to what we will do in this thesis, a large body of the literature in neuroscience relates to the understanding of the brain functions in terms of transmissions between parts of the brain or even between neurons. The inference of causal link between parts of the brain permits the characterization of the functional circuits which underpin important brain functions such as perception, cognition, behavior, and even consciousness (Seth et al., 2015). The tools used to study such functional circuits are very close to the ones we study and develop in this thesis, as they use mainly time series as raw material to investigate the causality structure. As systems such as brain are very complex in terms of causal structure, different levels of precision or representations of causal structure are possible. Functional connectivity aims at describing the dependencies between a set of variables without making any hypothesis about the underlying causal structure producing the observed patterns. On the contrary, the effective connectivity tries to provide the simplest causal structure able to describe the observed patterns in the set of selected variables (Friston et al., 2013). While functional connectivity relies on simple pairwise causality measures such as the one based on Granger principles, effective connectivity asks for more elaborate strategies to keep only the more relevant causal links. This could be achieved by comparing how well different causal structures describe the time evolution of the set of selected variables or by adopting pruning approaches such as Lizier and

Rubinov (2013); Pearl (2000b); Spirtes et al. (2000) which tries to retain only the most valuable causal links.

These kinds of approaches are also well established in Biology especially in genetics. In a nutshell, two sets of causes may be identified in Biology, the proximate causes which represents responses to the environment and the ultimate causes which are responsible for the genetic heritage via the evolution processes. If we recall Aristotle and his interest in the comprehension of Nature, we may postulate that his definition of final cause derives directly from the incredible adaptively of the living things and that this definition of final cause inspired biologist for their ultimate causes. The main limitation in the identification of ultimate causes comes from the characteristics of evolution such as the randomness of mutations, the uniqueness of every biological entity or the emergence of new qualities through reproduction (Mayr, 1961). The functional approach tries, in contrast, to identify proximate causes, i.e. how a specific genotype causes certain phenotype. This approach shares therefore a lot with the methods developed in Neuroscience . However, the functional approach is not immune to specific problems. Indeed, this approach does not treat simply the question of how genes interact and produce the observable outcomes, a more complete view of how the cell is organized is necessary. Once the structure of the DNA is taken into account, other regulatory elements interact with the DNA sequences to determine which protein has to be produced. Indeed, both the cell and its environment should be considered in addition to the DNA to understand the production of proteins (Noble, 2008). We see clearly here than we cannot simply look at the relationship between genotype and phenotype, a complex environment has to be considered. The same issues arise in Climatology where different types of factors interact to form a very complex system including feedback loop. Beyond the usual simulation approach, new methodologies based on causal theory emerge to analyze these systems from a different perspectives (see van Nes et al. (2015); Hassani et al. (2016))but also to go beyond the usual correlation analysis (van Nes et al., 2015).

Causality in Finance and Economics

We have so far highlighted the central role of causality in many domains of Science, and the questions it raises. Let's now look at its implication in the disciplines which will be at the core of this thesis, the Economics and Finance. If we recall the four types of causes proposed by Aristotle, material and formal causes are the most fundamental for economic ontology for many philosophers of Economics (Mäki, Cambridge University Press). However, the influence of physics is such that in practice the concept of efficient cause is often preferred by practitioners (Bunge, 1963; Hoover, 2008).

As illustrated by the title of the seminal work of Adam Smith "An Inquiry into the Nature and Causes of the Wealth of Nation" (Smith, 1776), causality has always be one of the main preoccupations in Economics. David Hume also emphasized the role of causation in public affairs, i.e. Economics (Hume, 1742), even if as we have seen earlier his vision about the possibility of causal inference was relatively skeptical. Other economists such as John Stuart Mill had a more positive view of causality inference but was more skeptical about its use in Economics as he states in his Political Economy (Mill, 1848) that Economics was an "inexact and separate Science". He considered Economics as a Science where the guiding principles were essentially known "a priori" through the definition of ceteris paribus clauses (see Hausman (1992)). As we will see later, this vision of causality had a profound impact on the way causality will be treated in Economics.

In the late 19th and early 20th century, researchers such as Francis Galton and Karl Pearson set the foundation of modern statistical science with the development of concepts such as correlation (Galton, 1888; Stigler, 1989), standard deviation, regression analysis or the method of moments (Walker, 1975; Pearson, 1900). The notion of correlation or statistical significance in regression analysis (Edgeworth, 1885), relates directly to the notion of causality. However, in contrast with simple correlation the statistical significance of a coefficient in a regression provides further information about the direction of the relationship and is therefore closer to the notion of causality. The development of new theories in Economics, in the early 20th century, led to the vision of a partial or general equilibrium ruling the economy (Marshall, 1930; Walras, 1954). This implied simultaneity in different processes, such as the definition of price and quantity, which questions the possibility of inferring causal relationships. Indeed simultaneity leads to a symmetric relation in term of statistical significance, i.e. a problem of observational equivalence resulting in the now familiar problem of identification in econometric. The identification problem relates for example to the definition of the supply and demand curves of any goods, knowing only the price and quantity, which is impossible without any other knowledge about the underlying process as the price and quantity are defined simultaneously. This identification problem has been addressed by the Cowles Commission in the fifties (Koopmans, 1950; Hood and Koopmans, 1953) which proposed to rely on additional causal determinants that discriminate otherwise simultaneous relationships (Hoover, 2008). These additional parameters could be, in the case of milk, hay price for the supply curve and the period of the year for the demand curve (higher consumption in winter for hot chocolate and in summer for ice cream). In the econometric language, these additional parameters are called exogenous variables while the parameters of interest are called endogenous. Even if, at first sight, causality seems to be pushed at the background in this identification process, we can emphasize that the discrimination is only possible if the endogenous parameters are

causally determined by several exogenous parameters, i.e. the discrimination is done via the identification of causal relationships.

In addition to the approach proposed by the Cowles Commision which is referred as "structural" as it implies an underlying structural model describing the evolution of the endogenous variables, other approaches have been proposed. The "process analysis" approach relies on the asymmetry of causality stressed by the condition of Hume "The cause must be prior to the effect" (Morgan, 1991). This second approach has gained interest only in the nineties, the structuralism being preferred since the Cowles Commision (Hoover, 2004). In addition to the differentiation between 'structural' and 'process' approaches, Hoover proposed two other distinctions in the causal theories developed in Economics and Econometrics, the approaches which relies on an 'a priori' identification of the relationships to be considered and the ones which try to infer causes directly from data.

The 'a priori' structural approach is the one that has been proposed by the Cowles Commision with the term 'a priori' implying that assumptions are made on the underlying causal structure. The inferential structural approach which derived partially from the framework developed by the Cowles commission, has first been proposed by Simon(Simon, 1953) in 1953 and then further developed by Hoover (1990, 2001) to nonlinear systems of equation found in modern rational-expectations models. The main idea is to infer causal structures not on the basis of temporal dependence but by considering recursive process. Let's look at the following bivariate system:

$$x_t = \beta y_t + \epsilon_t^1 \quad (1)$$
$$y_t = \epsilon_t^2 \quad (2)$$

Where ϵ_t^1 and ϵ_t^2 are independent white noise.

Following Simon, in this bivariate system Y causes X, as Y is recursively ordered ahead of X. Indeed, we need to know the value of Y to determine the value of X but we know everything about Y without knowing X. Y could therefore be used to control the outcomes of X. Nevertheless, we still face a problem of discrimination. Indeed if we consider now the related system described by:

$$x_t = \xi_t^1 \quad (3)$$
$$y_t = \alpha x_t + \xi_t^2 \quad (4)$$

Where $\alpha = \beta\sigma^2(\epsilon_t^2)/(\beta^2\sigma^2(\epsilon_t^2) + \sigma^2(\epsilon_t^1))$, $\xi_t^1 = \epsilon_t^1 + \beta\epsilon_t^2$ and $\xi_t^2 = (1 - \alpha\beta)\epsilon_t^2 - \alpha\epsilon_t^1$ where ξ_t^1 and ξ_t^2 are uncorrelated and $\sigma^2(\delta)$ represents the variance of δ.

This second system provides exactly the same form as the first one, once the two expressions are expressed in terms of ϵ_t^1 and ϵ_t^2. However, considering the second system, the previous causality criterion indicates that X causes Y although the reverse is true once we express in term of ϵ_t^1 and ϵ_t^2. How can we determine the true direction of this relationship? This observational equivalence problem which called previously for an 'a priori' knowledge of the considered model structure, is treated by Simon via controlled or natural experiments on the data. If it is possible to modify, via an experiment, the distribution of Y conditioned on X without modifying the marginal distribution of X, we would face the structure of the second system where X causes Y. It would indeed involve a modification of either α or ξ_t^2 without affecting ξ_t^1. The fact that Simon rely on experimental data explains why his approach is referred to 'inferential' rather than 'a priori'.

One of the most used and known causal approaches is due to Granger (1969) who proposed an 'inferential process' approach relying on two assertions we already mentioned: the cause happens prior to its effect and the cause has unique information about the future values of its effect. In contrast with the inferential structural approach, the Granger causality focuses on temporal dependence to infer a causal relationship between two variables. Suppes (1970) consider this approach as a remarkable example of the modern probabilistic approach to causality in line with the theory proposed in the 18th century by Hume. Indeed the main definition of Granger causality implies the following rules: considering two stationary ergodic processes X_t and Y_t, it is said that Y Granger causes X if $\mathbb{P}[x_{t+1} \in A|\Omega_t] \neq \mathbb{P}[x_{t+1} \in A|\Omega_t - y_t]$ with Ω_t representing the amount of data available to describe x_{t+1}. It implies that if X can be better predicted using the past values of Y and X rather than sole past values of X, Y Granger causes X. The operational definition proposed by Granger (1969) relies on the following linear regression model:

$$x_t = \beta_0 + \beta_{11}x_{t-1} + \beta_{12}y_{t-1} + \epsilon_1 \quad (5)$$

$$y_t = \beta_0 + \beta_{21}x_{t-1} + \beta_{22}y_{t-1} + \epsilon_2 \quad (6)$$

With Y Granger-causing X if $\beta_{12} \neq 0$ and X Granger-causing Y if $\beta_{21} \neq 0$. While Granger-causality imply, in general, inferential structural causality the reversed is not true. Indeed taking the structural form of Eq.3 and 4, we get the following system of equations:

$$x_t = \theta y_t + \beta_{11} x_{t-1} + \beta_{12} y_{t-1} + \epsilon_1 \quad (7)$$
$$y_t = \gamma x_t + \beta_{21} x_{t-1} + \beta_{22} y_{t-1} + \epsilon_2 \quad (8)$$

If we are interested in knowing if Y causes X, we see that structural causality implies that either θ or β_{12} must be different from 0 in order to observe a causal link. While for Granger causality, Y is considered as causing X if $\beta_{12} + \theta\beta_{22} \neq 0$ (Hoover, 2008). This means that if $\beta_{12} = \beta_{22} = 0$, Y does not Granger cause X although Y still structurally causes X as long as $\theta \neq 0$.

While the "a priori" structural approach does not make any clear reference to empirical reality when considering causal processes, the "a priori" process approach proposed by Zellner (1979) sides with Granger on the necessity of predictability in causal attribution (process). Zellner's vision of causality is close to the one developed by Feigl (1953) who understand causality in terms of predictability according to law. Therefore Zellner departs from Granger definition and sides with Simons on the importance of an underlying structure in order to discriminate causal laws from false causal generalizations. His definition of law is not as rigid as in physics or mathematics. He considers a law as a probabilistic description of a succession of states of the world (Hoover, 2008) and follows therefore a Bayesian approach of inference on which he apply a specific framework to reduce the possible range of causal relationships to consider (Savage's "small world" assumption (Savage, 1972)).

Not all causality definitions fit in one of these four categories and other approaches have been proposed. One good example of a methodology that lies at the edges of two categories is the structural VAR of Sims (1982, 1986). This approach is an evolution of the VAR (vector auto-regressive) model proposed by Sims, which considers the effect of a shock on several variables to assess a causal structure. As in the Granger causality framework, it relies on the error terms of linear regressions but here the focus is put on the error terms reflecting the shocks. These error terms being rarely independent, an orthogonalization process is needed to obtain a diagonal covariance matrix. This was done by considering a recursive model in the vein of Simon.

$$y_t = \beta_{21} x_{t-1} + \beta_{22} x_{t-1} + \epsilon_2 \quad (9)$$
$$y_t = \gamma x_t + \beta_{21} x_{t-1} + \beta_{22} x_{t-1} + \epsilon_2 \quad (10)$$

This solution implies the arbitrary choice of a recursive order by considering Y in the set of explanatory variables of X, or the reverse. This explain the term 'structural' in the SVAR specification as we have to consider which of θ and γ in Eq. 7-8 have to be set to zero (see Eq.9-10 for the two cases applied on Y). Even if it relies on the VAR approach, this model is therefore closer to the 'a priori' framework than to the inferential one and could be considered as a semi-structural approach as only the structure of contemporaneous variable have to be specified, not the structure of the lagged variables, i.e. the β (Hoover, 2008). Different orthogonalization processes have been proposed, such as the cholesky-factor orthogonalisation where the first variable to be considered is only affected contemporaneously by its own innovations, the second is then affected contemporaneously by its own innovations and by the first variable and so on until the last variables.

Another class of methods that have been developed outside Economics but have been recently extensively used in Finance and Economics relies on Graph-Theory (Pearl, 2000b; Spirtes et al., 2000). If the underlying graph model is undirected, the model is referred as a Markov random field while for directed/causal model we rather refer to as a Bayesian network. The main idea behind the concept of Bayesian network is that any system may be represented by a graph in which each causal relationship is identify by an arrow indicating the direction of the information flow (Selva and Hoover, 2003) and represented by a joint probability distribution. Double-headed arrows represent simultaneity while simple arrows identify causal structure. By considering an entire system instead of pairwise relationships as previously, we get closer to network theory by considering the complex environment around each individual node.

These methods developed in Graph-Theory rely primarily on pruning steps, by considering first a complete graph where every nodes are connected and applying a set of rules to remove non causal links. These rules are based on probabilistic dependence or independence tests which often use conditional correlation measures. Indeed the causal links are not inferred from temporal relationships but by the results of the independence test. For example, if we consider the following relationship $A- > B- > C$, before any test of independence we should see a probabilistic link between C and A, but once conditioned on B this relationship disappear and A and C are considered as independent. The PC algorithm developed by Spirtes et al. (2000) is the most illustrative example of this type of methods. It starts by considering a fully connected graph and applies on it an independence test on each pair of variables via a correlation measure. Once done, it applies on every pair of variables, a series of conditional independence test, using variables in the close environment as conditions, in order to remove the possible redundant links. The different methodologies that have developed in Graph-Theory to treat the question of redundancy are mainly based the definition

of a set of Formal Logic propositions. In this framework, the different nodes represent the alphabet of the logic whereas the existence of a relationship between these nodes depends on a set of function and logic connectors. However, Graph-Theory departs from Formal Logic by its ability to treat relationships in a probabilistic way while Formal Logic considers that a proposition is either true or false. The pruning algorithms developed in Graph-Theory are called DAG which stands for directed acyclical graph meaning that we assume the existence of only recursive structures inside the graph although variables could often present cyclicality in Finance and Economics.

An alternative approach based on the same idea of pruning steps is the general-to-specific method which has its roots in the work of Sargan (1964) but was further developed and popularized by David Hendry and Hans-Martin Krolzig (Hendry and Krolzig, 1999, 2001). As with the PC algorithm, the identification of significant variables stems from the application of a well-designed simplification algorithm which starts with a pre-search stage and then sequentially dropping or adding variables based on a specific testing procedure. In contrast with Graph-Theory, the simplification procedure is often based on a VAR specification. Hendry improved with several co-authors, this procedure and apply it successfully in a series of influential time-series studies treating aggregate demand-for-money and consumption functions (e.g., Davidson et al. (1978); Hendry and Mizon (1978)). Hoover and J. Perez (1999) further explicit the guidelines used in the simplification algorithm and adopted a multi-path strategy by considering all possible initial paths to avoid 'path dependency'. Indeed by considering only one path, such as successively removing the variable with the lowest absolute t-value, the simplification algorithm could result in the suppression of relevant variables. Different paths are therefore confronted to create a terminal model following specific criterion. The simplification is then applied once again on the last terminal model until either every path provides the same result or no more simplification are possible.

Going back to pairwise relationships, despite the high number of approaches that have been proposed in the literature, the framework developed by Clive Granger is probably the most prominent source of inspiration in the recent development of new causality measures in Economics and Finance. Indeed numerous approaches deriving from Granger's definition of causality have been proposed in the past decades to analyze different types of information transmission, by considering for example the volatility rather than the usual mean with methods such as the variance decompositions developed by Pesaran and Shin (1998) and further improved by Diebold and Yilmaz (2014) as a connectedness measure or the LM (Lagrange multiplier) test of Hafner and Herwartz (2006). The LM test has been constructed specifically for causality in variance through the adaptation of the general Lagrange Multiplier misspecification test, introduced by Lundbergh and Terasvirta (2002), to the estimation of univariate GARCH

(1,1) models. Other methods extend the linear regression specification of Granger causality test in order to look at specific causal patterns, such as non-linear causal relationship with the development of regime switching Granger causality (Psaradakis et al., 2005b) based on the Markov regime switching specification, the Multivariate Granger causality proposed by Barrett and Barnett (2010) to cope with relationships between systems composed of more than one variable or the MIDAS Granger causality developed by Ghysels et al. (2016) which is devoted to mixed frequency data.

Outside Finance and Economics, several methods deriving from Granger notion of causality have been designed based on concept borrowed from different field of science such as Information theory or statistical physics. One of these measures, which will be extensively used in this thesis, is the transfer entropy which has been first proposed by Schreiber (2000). The definition of transfer entropy borrows, for the Information theory developed by Shannon (1948), the concept of information entropy which represents a measure of the information content of a system, i.e. the amount of information needed to describe the configuration of a system. Shannon was the first to proposed a physical definition of the concept of information which became a measurable quantity via its entropy. As we will see in more details in the next chapter, transfer entropy is based on the comparison between two conditional information entropies. In the first one, the entropy of the series of occurrences representing a variable X is conditioned on it past observations, while for the second one the entropy of X is conditioned on both the past of X and the past of Y. If we observed a reduction of entropy when knowing the past of Y it implies that there is a transfer of information from the variable Y to the variable X. Thanks to the definition of transfer entropy, the links existing between systems may be seen as transfers of information, a more broad notion than causal relationships as it encompasses a notion of quantity rather than only probability.

Research approach

We have seen so far that causality was a central question in many domains of science and most of the visions that have been developed around this question are inter-related. Indeed, each discipline provided a piece of the causality puzzle. While Newtonian and special relativity highlighted the central role of causality in our understanding of how our environment evolves, quantum mechanics led to a probabilistic approach of causality. This probabilistic vision played a crucial role in the development of measures, in Finance and Economics, to study causal relationships. Other fields of science brought elements to the discussion, such as biology, neuroscience or information theory which rely, as Finance, heavily on time series. In addition, by considering highly complex systems, i.e. the brain or genes, these disciplines had to consider not only trans-

missions between pairs of variables but look at the broader picture and consider the impact of their environment leading to the concept of network which is used nowadays in various fields from the understanding of social networks to the definition of power grid reliability. All these complementary visions of causality ask for the development of inter-disciplinary approaches. This explains why it is so important to consider not only the history of causality and the measures developed so far in Finance and Economics, but also look at what has been done in other fields of Science. We propose in this thesis to adopt this interdisciplinary approach which will allow us to develop new methodologies in order to address some fundamental questions in Finance such as the comprehension of how the current financial system works and how the complex inter-relations existing between its members could affect its stability.

Beyond the question of the definition of valuable sources of inspiration for the development of new methodologies to infer causal links, let's consider the epistemological framework that will be used throughout this thesis. The definition of such an epistemological framework is necessary to determine how knowledge may be produced and how this knowledge may be justified. There are three important dimensions of the Epistemology (Taskin and de Nanteuil, 2001): the validity of the scientific knowledge, the meaning of the knowledge, and the limits of the knowledge. Different paths have been proposed since the Renaissance to develop a theory of knowledge that could answer those three aspects. The positivism seems the most dominant approach in Finance and Economics (Friedman (1953) or Frankfurter (2007) among others) but other trends have affected these fields of Science. The positivist approach is inductive as it uses observed facts to create a new theory and then test it through the experience. Methodologies developed in Finance follow partly this approach, as they often rely on market data in order to proposed new theories about how the economy works. The series of essays proposed in this thesis follows this mainstream current of thought, as the development, the validation and the application of the proposed models will be mainly data driven. However, the proposed approaches goes beyond positivism, considering the importance of the validation processes throughout the different chapters of this thesis. We side rather with Popper, highlighting the importance of the falsifiability of every models which determines their ability to provide useful knowledge about the evolution of financial systems.

However, both the positivism and the falsificationism have several limitations in Finance and Economics. A first limitation is related to the validity of the many assumptions on which financial models are based in order to simplify the underlying complex process and be able to represent it. Indeed, ideological influences act on the assumptions selection process (Frankfurter and McGoun, 1999). We therefore draw away from the positivist approach to get closer to the structuralist one where the

development of new knowledge depends on its own field and on the place one has in this field. As Frankfurter and McGoun (1999) pointed out, modern Finance seems to follow this trend and to be trapped in an epistemological approach impregnated by ideologies which favors specific approaches and limits the innovation. Considering the interdisciplinary approaches that we proposed to follow in this series of essays, we try to draw away from the mainstream approaches proposed so far in Finance to infer causal links. We propose indeed to consider the limitations of the current causality measures used in Finance and look at approaches deriving from other fields of Science. But we cannot entirely rule out the influences of our environment. Other facts tend to influence our research subject and the way we treat it. Indeed, the history affects the way we think and what knowledge we want to create. As Foucault (1982) said, "the system of thought is the result of contingent turns of history". History plays a major role in the development of new knowledge. The recent financial crisis, for example, led to large changes in the way we saw the financial system. It permitted the development of new methodologies related to network theory or causality measures which have inspired the different topics addressed in this thesis.

Indeed, the recent literature on causality is very fruitful in Finance especially the applications of causality measures to investigate either the relationships between financial assets or to understand the topology of a specific market and its impact on the system fragility. The greater interdependence of markets across the globe explained by the financial globalization and its effect on the global economy and social welfare during turmoil periods, as illustrated by the financial crisis of 2008, highlighted the necessity for a rethinking of economic and financial policies. The development of a new vision requires a better understanding of the mechanisms governing information transmission inside the economy. This transmission mechanism was investigated from a number of perspectives, for different market and asset classes such as the equity market (Longin and Solnik, 1995; Hong et al., 2009b; Celik, 2012; Dungey and Gajurel, 2014), the financial market (Diebold and Yilmaz, 2014; Billio et al., 2012; Dungey et al., 2012), the sovereign bond market(Longstaff et al., 2011; Gorea and Radev, 2014; Fernández-Rodrígueza et al., 2016) or the commodities (Bhar and Hammoudeh, 2011; Hegerty, 2016; Zhang et al., 2016).

This literature treating about interrelations focused primarily on the detection of stress transmissions often identified as contagion (Dungey and Gajurel (2014); Pasquariello (2006); Yuan (2005); Longin and Solnik (1995) to quote only a few). This contagion phenomenon may be identified at several levels: inside a specific market, between markets or between countries. The question of stress transmission may be addressed at two different levels, a micro one where we consider individually each transmission channel or at a macro one where the entire system is considered. This second approach

often relies on topological measures coming from network theory to understand the sources of the fragility of a system. A pioneering contribution in that vein has been proposed by Billio et al. (2012) who estimate pairwise Granger causality on a large set of data of financial institutions and apply then different topological measures to understand the evolution of the financial system characteristics before and during the financial crisis to relate the evolution of its topological characteristics to its fragility.

We propose, in this thesis, to follow both strategies by looking at how causal links may be detected but also at how the interrelations organize themselves and how the topology created by these individual connections could give information about the system and the individual entities. We will focus in our applications, primarily on the functioning of the global financial system, but also on the information transmission between countries. We will consider mainly the western countries, especially the U.S., because of its central role in the recent financial crisis. As documented in the literature, several challenges have to be addressed to obtain an accurate representation of causal relationships between financial entities. Among those challenges stands out the absence of comprehensive and reliable information on physical connections and the difficulties to obtain them (Feng et al., 2014). These difficulties are shared by other fields than Finance including the spread of disease (Lipsitch et al., 2003; Wallinga and Teunis, 2004) or human travel (Brockmann et al., 2006), to quote only a few. As an alternative, a strand of the literature has recently proposed to rely on dependencies between the asset prices as a proxy for physical connections. This approach is extremely convenient as it relies on easily available data and as it allows one to remain agnostic on the specific channel through which the variables are connected. We rely therefore, in every application presented in this thesis, on returns estimated from asset prices to recover the possible underlying causal relationships between each variables of the considered financial system.

Main contributions

Each chapter of the thesis shares a similar framework: we first introduce the topic and the different problems it raises and then propose a solution in the form of a new methodology or existing methodologies in the case of the first chapter. The methodologies are then tested in a simulation exercise to validate or characterize them. Once validated, we apply the methods on empirical data to treat a specific topic in Finance often related to the financial system. In the first two chapters, the application to empirical data helps also in confirming the results of the validation obtained via the Monte Carlo simulations.

The main validation strategy in Sciences is to perform multiple times independent experiments and estimate based on the results of the experiment and the prediction of the model, statistics about the uncertainty of the proposed model. However, when dealing with modelization of financial data, we cannot perform independent experiments as we deal with continuous time series. The proposed solution is to rely on simulation of financial time series based on their main characteristics that have been highlighted in the literature. We follow this validation strategy in the four chapter of this thesis. However in the first two chapters, we rely additionally on non experimental data which is not common in Economics of Finance. Indeed, the relationships between financial time series being not know a priori we cannot use supervised approaches which relies on reproducible experimental data. In order to avoid this shortcoming, we propose to identify a proxy allowing the determination of the quality of our models estimates. As we will see in more details, the proposed proxy account for the information content of the causal structures identified by our different models.

The methodological part of this thesis has the objective of understanding the causality in its complexity, by looking at different aspects of causality such as the importance of the environment around a causal link and the different channel through which information can travel. Each chapter will therefore bring a new piece of the causality puzzle. Our work relies mainly on Granger concepts of causality. We start therefore, in the first chapter, by considering two pairwise causality measures based on these concepts: Granger causality which has been extensively used in Finance and Economics and transfer entropy which derives from information theory and has been applied only a few times on financial data.

The objective of this first chapter is to introduce the question of causality by describing the functioning of these two simple methods and the main concepts that led to their development. We compare then, the ability of both measures to detect different types of connections illustrating the main features of financial time series. To achieve this objective, we use several well-known data generating processes (DGP). After having characterized both measures in the context of financial data, we look at their information content in the application. The objective is to determine if there is a difference between the two measures, and if a difference exists, if their information content are complementary.

In the second chapter, we try to go beyond the simple pairwise approach and look at the impact of the environment around the variables transmitting and receiving the information. Indeed the pairwise approach tends to overestimate the number of causal relationships by considering both the direct and indirect links. We mean by indirect

links, the links that are due only to the existence of information transiting by a third variable. Let's take the example of a variable A sending information to a variable B which in turns sends information to a variable C. When using a simple pairwise approach, we should observe a transmission of information from A to B, B to C and A to C. While the first two are considered as direct connections, the third one is an indirect one, only caused by the existence of a transfer from A to B and from B to C. This link does not really exist and represents a redundancy in the global information transfer, as it relates to the portion of the information received from A that B effectively transmitted to C. The elimination of such indirect links is one of the most crucial challenges to be addressed to get an effective representation of a network.

Inspired by Graph-Theory, the proposed algorithm uses conditional information transfers and Formal Logic propositions to get rid of these indirect links. Two types of indirect links are considered, the ones that are caused by lagged relationships such as the example presented before, if we consider a lag of one period between the transfer from A to B and B to C and a lag of two periods between A and C. The other one is caused by instantaneous causal relationships. In this case we consider an instantaneous information transfer either between A en B or between B and C, the other having a lag of one period as well as the transfer between A and C. Both cases are treated by the algorithm with the additional objective of reducing as much as possible the dimension of the conditional information transfer to be estimated. To this extend, we have been inspired by the general-to-specific approach as well as by the directed acyclical graph techniques. Indeed, the proposed framework follows the two steps approach of the Bayesian network models, i.e. an initialization step followed by a pruning step. However, the conditional probability distributions that Bayesian models rely on during the pruning process, are replaced by conditional causality measures but as in Graph-Theory the emphasis is put on the selection of the right conditions during the pruning procedure. Indeed our parsimonious approach relies on a pre-search step where every connection are tested using a simple pairwise approach. In a second step, the algorithm uses conditional measure to prune the network from its non-relevant causal relationships.

Beyond this question of effective links, going back the simpler case of pairwise connections, we may find that more than one channel may be used to transfer information between two systems; either because the system is represented by several variables or that we can decompose each variable into spectrum using frequency or quantile decomposition. For such complex systems, the transfer of information can happen at different levels or in other words through different channels. If we consider the usual pairwise approach, only one channel at a time may be investigated which prevents us from considering in one measure every possible channels. Inspired by Neuroscience

where multivariate connectivity is a common problematic and by papers such as Barrett and Barnett (2010), we will try to address, in the last two chapters, this complex question of multi-channel transmissions. We will get away from the usual problematic of multivariate explanatory variables to account for systems characterized by both multivariate explanatory and dependent variables. The aim of these last chapters is therefore to look at the evolution, in time, of the relative importance of every channels in one global measure.

We propose two different frameworks: the first one will rely on the simplifying assumption that, at each time step, a specific channel dominates the information transfer between the two considered systems. We therefore start by the creation of a measure of the relative importance of each channel at every time steps. Once defined, the objective of the algorithm is to determine, for the considered step, the channel that maximizes the information transfer. These two steps are performed for every time steps and the selected channels form a map of the time evolution of the channels' activity. To avoid instability in the channel selection, a softening procedure is applied. This framework was originally developed for the transfer entropy causality measure but is also applied on the regression based, Granger causality measure.

The last chapter of the thesis treats the same question of multichannel relationships but from a different perspective. Rather than considering the relative importance of every channel at each time step, our methodology relies here on the regime switching Granger causality test (Psaradakis et al., 2005b) which enables us to look at the entire time window in one step. We propose to modify the regime switching model to take into account multiple dependent variables. Instead of considering each explanatory variable as a possible source of information for one specific dependent variable, we consider every possible channel represented by a specific pair of dependent and explanatory variable, as a possible state of the world. The Markov regime switching framework provides comprehensive information about the probability of activity of every channel for each time step.

We have so far presented the different methodological questions which will be raised in the four chapters of this thesis, but not yet the empirical applications. The purpose of these applications will be twofold, first we will try to demonstrate empirically the interest of the different methodologies developed so far, but these methods will also be a way to address some crucial questions in Finance. In the first three chapters we will look more deeply into the functioning of the western financial system. Chapter 1 and 2 will try to highlight the possible link existing between the level of risk of the financial system and its topological characteristics. These topological characteristics will be estimated on networks inferred using the algorithm developed in chapter 2, which

allows us to get rid of the possible redundant causal relationships. We will look at the possible link from two perspectives; we will first consider the system as a whole by looking at the average level of risk in the financial market to assess the existence of a link between the systemic risk and the topology of the financial network. We then propose in chapter 2, to look at this question from a micro perspective by considering the impact of the topological characteristics of the individual financial institutions on their own risk level. We therefore withdraw from the alternative question of which financial institutions contribute the most to the risk of the system (see Brunnermeier et al. (2009); Acharya (2009); Anand et al. (2012); Battiston et al. (2012) among others) and treat the reverse question of how the system and, more specifically, its topology could impact the fragility of a specific institution. Another objective of the application proposed in chapter 2 will be to determine the effect of the dataset frequency on the ability of the topological characteristic to describe the evolution of the risk. We aim therefore at defining the optimal frequency leading to the network representation closest to the true underlying network. This optimal frequency is determined by using the topological characteristics ability to describe the risk level, as a proxy of the distance between the network inferred and the true underlying network. Looking at this proxy, we try to define which frequency leads to the network representation giving the higher explanatory power. The considered frequencies are, daily, hourly and high frequency, i.e. every 15 minutes.

In contrast with chapter 1 and 2, the application developed in chapter 3 does not consider the entire system but look in more details at the way financial institutions communicate to each other. We propose to decompose the asset prices of every considered financial institution into either frequency or quantile spectra. These spectra are then used to estimate a multi-channel information transfer between every connected pair of U.S. financial institutions using the two measures developed in the methodological part. Indeed every institution will be represented by a set of time dependent variables representing the different frequencies or the different quantiles.

As for the chapter 4, we withdraw from the question of the financial system connectivity and look at a higher level by considering the relationships existing between countries around the world. We do not rely on signal decomposition anymore, we rather consider different financial parameters to define the economic health of every countries in our dataset. The regime switching multi-channel causality measure developed earlier helps us to define for each country which financial variables are involved in the information transmissions and which financial variables are affected by other countries. In addition to the inference of information channels between countries, we also look more deeply at the impact of the commodities market on the information transfer dynamics by considering this market as a possible additional channel of information transmissions.

Chapter 1

Testing causality in financial time series

1.1 Introduction

As recalled by Hlavackova-Schindlera et al. (2007), identifying causal relationships between variables has been the fundamental issue of most natural and social sciences over the history of human knowledge. In the aftermath of the recent crisis, the topic has experienced a regain of interest in the finance literature searching for a salient quantification of financial indstitutions' exposure to the so-called "systemic risk" that is the risk that an individual institution's default spreads out to the entire financial system then impacts the real economy by damaging economic growth and social welfare. Indeed, different approaches have been proposed in the literature to assess the level of threat an individual institution represents to the system relying on various techniques (see Acharya, 2009; Brunnermeier et al., 2009; Adrian and Brunnermeier, 2009 or Brownlees and Engle, 2017 among others). A specific strand of the literature considers that the answer lies in the topological properties of the financial system itself (see in particular Battiston et al., 2012; Krause and Giansante, 2012). More specifically, those contributions rely on the conception that the financial system constitutes a complex network of financial institutions (the nodes of the network) caracterized by different features and linked by means of various channels. Those channels can be physical (e.g. cross-exposures as in Boss et al., 2004; Masi and Gallegati, 2012) or inferred from observable stock market data (e.g. prices, as done in Billio et al., 2012; Diebold and Yilmaz, 2014 or Bianchi et al., 2015 using various techniques) by consdering that if features from two nodes are dependent according to a specific measure of correlation or causality there should be an underlying (physical) link between them. As physical linkages are most of the time not directly observable, this last approach has appeared convenient to capture information transfer between financial institutions. In addition, if we believe that prices do reveal all the relevant information available

in the market, it allows to remain agnostic on the specific channels through which information transfer operates.

Although there is no unique definition of causality, it seems that in economics and finance in particular, the approach popularized by Clive W. Granger (see Granger, 1969, Granger, 2003, among other contributions) has concentrated most of the empirical developments within the field. Based on the well known intuition that (i) the cause should occur prior to its effect and that (ii) causal variables should content unique information allowing to better predict the effect, various direct testing procedure based on estimations of linear models have been developed in both time series and panel contexts (see Granger, 1980, Holtz-Eakin et al., 1988, Nair-Reichert and Weinhold, 2001 or more recently Dumitrescu and Hurlin, 2012). While those approaches have been extensively - and in an exclusive manner - used in empirical contributions in economics and finance (see Geweke et al., 1983 and more recently Hoover, 2008 for surveys), other sciences such as cognitive sciences, physics or machine learning have relied on more general concepts, mixing statistical representation, graph theory and counter-factual reasoning (see Pearl, 2009 highly influential contribution for a deep understanding of the concept). A specific branch of this literature that we believe is particularly fruitful, borrows from the developments in information theory (Shannon, 1948) and the concepts of entropy and transfer entropy to reappraise the issue of causality. Indeed, it has been shown that close links exist between the informational transfer entropy measures (Shannon, 1948; Schreiber, 2000) and that of causality developed by Granger (1969), see Barnett et al. (2009) or Hlavackova-Schindler (2011).

Our aim within this chapter is to assess the accuracy of those information-based measures of causality and develop a testing procedure to assess whether they lead to better identification of causal relationships between financial variables than standard Granger procedures. Indeed, the equivalence between transfer entropy and the linear form of Granger causality test highlighted by Barnett et al. (2009), was shown only under the assumption of Gaussianity of the underlying time series. But financial variables have, in general, many specific features incompatible with the simple Gaussian representation. Financial variables are most of the time nonstationary, returns are therefore often preferred when conducting financial analysis. These returns are stationary in mean but typically not in variance as their volatility can evolve in time. On top of this question of stationarity, one has also to take into account the fact that financial variables are typically asymmetrical, leptokurtic and possess fat tails (see Cont (2001) among others). We propose therefore to investigate how these characteristics may affect the ability of Granger based causality measures and transfer entropy to infer causal links. We conduct a simulation exercise based on different data generating processes (DGPs) exhibiting some striking features of financial variables such as their non linearity using

regime switching specification, their fat tails using causal co-jump specification and their heteroskedasticity by considering a causal generalized autoregressive conditional heteroscedastic (GARCH) specification.

Once we have established an alternative methodology to identify causal dependence between financial time series, we propose in an empirical application to implement it to retrieve financial networks and revisit the question of how risk and topological characteristics are interrelated, considering here, additionally the influence of the causality measure used to infer the network representing the financial system. To that aim, we first investigate the differences existing between the informational content of the networks estimated with the usual Granger causality test and the ones estimated with transfer entropy, considering also their potential complementarity. We then discuss the still unexplored question of how sensitive is the network to its environment, i.e. how the knowledge of the environment influences the definition of the network topology. We defined here the environment as a set of relevant components not included in the considered network but that may still impact the relationships between the members of this network. We address, therefore, empirically the difficult question of network optimal sampling (see Krivitsky and Kolaczyk, 2015; Kolaczyk, 2017) considering the following questions: How does the environment affect the network characteristics? How transfer entropy and Granger causality measures deal with this additional information sources? Are we mislead when considering only partial network settings? To achieve this second objective, we rely on the effective network inference algorithm proposed in Dahlqvist and Gnabo (2018). This algorithm is based on conditional causality measures which allow to take care of possible indirect links. The conditions applied on the causal estimators play a crucial role in the definition of the relationships between members of the network. Therefore, the size of the environment which is used as an additional source of information during the inference procedure, should modify to a certain extend the topology of the inferred network.

Empirically, the chapter focuses on the European and U.S. financial system. We consider four different datasets, the first one contains the CDS spreads of 24 European and U.S. Systematically Important Financial Institutions (SIFIs) and represents the network of interest. The three others, which will be incrementally added to the original dataset to account for the surrounding environment of the considered SIFIs, gather the CDS spreads of selected banks, insurance companies and real estate companies. We assess the impact of the surrounding environment of the SIFIs network by looking at the ability of the topological properties of the inferred network to describe the evolution of its own risk level. We incrementally increase the number of considered peripheral financial institutions to infer our effective networks. This allows us to assess the added value of the knowledge of these peripheral financial institutions in the definition of the

links existing between the SIFIs, i.e. the impact of the environment on the pruning procedure of the effective network inference algorithm.

As a whole, our contribution with this chapter is threefold. First we contribute to the methodological literature by proposing a causality testing procedure that relies on transfer entropy measure[1]. More specifically, we propose a new symbolization procedure in the framework of Marschinski and Kantz (2002) effective transfer entropy, that allows to consider different parts of the distribution function of the times series of interest when inferring transfer entropy. Second, we document in a simulation-based exercise the added value of the effective transfer entropy in presence of highly nonlinear systems compared to alternative approaches. Third, we apply this framework along the standard Granger causality measure to revisit systemic risk measures based on network representations estimated from market data, considering both the complementarity of these approaches and the difficult question of network sampling. The remainder of the chapter is the following. Section 1.2 reviews the concept of Granger causality, entropy and transfer entropy and exposes our methodology to estimate transfer entropy and develop a bootstrap-based inference methods. Section 1.3 exposes the various DGPs we assume for our financial processes and exposes the results regarding the competing measures of causality. Section 1.4 then turns to the empirical application and exposes the data, estimation method and results. Finally, Section 1.5 concludes.

1.2 On causality detection

1.2.1 From Granger representation to entropy and transfer entropy measures

As recalled by Guo et al. (2010) and Bressler and Seth (2011), the concept of Granger causality stems from the pioneering work of Wiener (1956), which states that one variable or time series should be called " causal" to another if the ability to predict the second variable or time series is improved by incorporating information about the first one. Granger (1969) later formalized this idea and provided a practical implementation method in the context of linear autoregressive models of stochastic processes.

[1]Note that a concomitant contribution (Diks and Fang, 2017) has proposed a very similar testing framework to ours, see Béreau and Dahlqvist (2014).

More formally and considering two stationary ergodic[2] processes X and Y, it is said that Y Granger causes X if future values of X, i.e. say x_{t+1}, can be better predicted using joint past values of Y and X rather than the sole past values of X, which leads to:

$$\mathbb{P}\left[x_{t+1} \in A|\Omega_t\right] \neq \mathbb{P}\left[x_{t+1} \in A|\Omega_t - y_t\right] \tag{1.1}$$

with Ω_t representing the set of observations available to describe future state of X. As stated by Granger (1980), for that inequality to hold, Y needs to contain unique information about what value X will take in the immediate future.

The operational definition proposed by Granger (1969) relies on the linear regression model expressed in Eq.1.2 as follows:

$$x_t = \beta_0 + \sum_{l=1}^{L} \alpha_l x_{t-l} + \sum_{l=1}^{L} \beta_l y_{t-l} + \varepsilon_t \tag{1.2}$$

where the null hypothesis of Y not Granger causing X corresponds to the joint nullity of $\beta_l,\ \forall l \in \{1, \ldots, L\}$ leading to Eq.1.3:

$$x_t = \beta_0 + \sum_{l=1}^{L} \alpha_l x_{t-l} + \eta_t \tag{1.3}$$

The intuition is the following. Since Eq.1.2 contains more explanatory variables than Eq.1.3, the variance of ε_t is at most as high as that of η_t. In particular, if the variance of the residual component ε_t is significantly less than that of η_t, it implies an improvement in the prediction of future values of X due to Y. Different tests have been proposed to measure how much Y causes X, such as the Wald test:

$$\mathcal{F}^{W}_{Y\to X} = T \cdot \frac{\widehat{\sigma}^2_{\eta_t} - \widehat{\sigma}^2_{\varepsilon_t}}{\widehat{\sigma}^2_{\eta_t}} \tag{1.4}$$

the likelihood-ratio test of Geweke (1982):

$$\mathcal{F}^{LR}_{Y\to X} = T \cdot \mathsf{ln}\frac{\widehat{\sigma}^2_{\varepsilon_t}}{\widehat{\sigma}^2_{\eta_t}} \tag{1.5}$$

[2]The ergodicity is a necessary condition to be able to infer, from a given sample, the statistical properties of the time series while the stationarity insure the stability in time of these properties.This notion of ergodicity is therefore similar to the one uses in the context of Markov chains as it requires a positive probability to pass from any state to any other in one step in order to avoid converging in a given state which would result in the differentiation between the statistical properties estimated over time and over the phase space.

or the Lagrange multiplier test:

$$\mathcal{F}_{Y\to X}^{LM} = T \cdot \frac{\widehat{\sigma}_{\eta_t}^2 - \widehat{\sigma}_{\varepsilon_t}^2}{\widehat{\sigma}_{\varepsilon_t}^2} \tag{1.6}$$

The Wald test performs better for small sample sizes. Statistical inference on that test can be developed according to the distribution of two alternative statistics which are both computed under the null of joint nullity of $beta_l$ as follows:

1. The Granger-Sargent statistic: $\mathcal{F}_{Y\to X} = \dfrac{\widehat{\sigma}_{\eta_t}^2 - \widehat{\sigma}_{\varepsilon_t}^2/L}{\widehat{\sigma}_{\eta_t}^2/(T-2L)} \sim_{H_0} \mathcal{F}_{L,T-2L}$

2. The Granger-Wald statistic: $\mathcal{F}_{Y\to X} = T \cdot \dfrac{\widehat{\sigma}_{\eta_t}^2 - \widehat{\sigma}_{\varepsilon_t}^2}{\widehat{\sigma}_{\eta_t}^2} \sim_{H_0} \chi_L^2$

Both statistics are asymptotically equivalent. Nevertheless, the Granger-Wald test has a higher power in case of small sample sizes. Those developments have been further extended to test causal relationships beyond the linear regression context between two time series of limited dimension. In particular, nonlinear extensions relying on threshold models and Markov-Switching specifications (see Billio and DiSanzo, 2006 for a survey) or alternative non-parametric and semi-parametric methods (see Taamouti et al., 2012 or Jeong et al., 2012 among recent contributions in the field) have been suggested in the literature to cope for potential nonlinear dependence structures while shrinkage methodologies such as Lasso-Granger algorithms have been introduced to address linear causality testing in presence of high-dimensional time series, see Bahadori and Liu (2012) and Liu and Bahadori (2012) for a survey of recent extensions.

More recently, transfer entropy, a concept borrowed from statistical physics and information theory has been suggested in the literature (see Kwon and Yang, 2008; Kim et al., 2012; Zaremba and Aste, 2014 to quote only a few), as an alternative measure to Granger causality. This causality measure presents the advantage to be "model agnostic" in the sense that it only depends on the estimation of multivariate entropies taken from the probability distributions of tested variables, and sensitive to nonlinear signal properties.

The foundation of the concept of entropy may be found in Rudolf Clausius's research in thermodynamics in the 1850s which led to the first definition of entropy as the dissipative energy used during a change of state. This measure can be seen as the thermodynamic quantity representing the inability of a system to transform its thermal energy into mechanical works. In the late 19th century, Ludwig Boltzmann and others gave entropy a more formal statistical definition. In such a context, the entropy measures the multiplicity of a system's microstates and how their probabilities are

spread out. Assuming k_B to be the Boltzmann constant and p_i the probability that the system is in its i^{th} microstate, the entropy takes the following logarithmic form:

$$S = -k_B \sum_{i=1}^{m} p_i \log p_i, \tag{1.7}$$

with k_B, the Boltzmann constant

Shannon (1948) extended this definition in the context of information theory, to create the information entropy, a measure of the information content of a system, i.e. the amount of information needed to describe the configuration of a system. In such a framework, the concept of system must be understood as a series of discrete outcomes. Assuming a discrete variable X with a probability distribution $p(x_i)$ and x_i the different states of X, the Shannon entropy is defined by:

$$S_s = -\sum_{i=1}^{m} p(x_i) \log_2 p(x_i), \tag{1.8}$$

with X a variable with m outcomes $\{x_1, ..., x_m\}$ and $I(p_{x_i}) = -log_2 p(x_i)$ the information content of realization x_i

The Shannon entropy corresponds to the expected value of the information of the distribution of X, in other words, when expressed in logarithm of basis 2, it gives the average number of bits needed to optimally encode all realizations x_i of the distribution X. The higher the uncertainty of the process, the more bits will be needed to encode the realizations. As an extreme case, when one realization is certain, i.e. associated to a probability one, the associated information content of x_i and thus entropy of X, is null. On the contrary, when all outcomes are equally likely, the entropy is maximized. This last result is known as the " Gibbs inequality". For more details on entropy and statistical physics in general, one can refer to Mezard and Montanari (2009).

It has to be noted that although Shannon entropy provides information about specific outcomes, it does not give an explicit description about the contribution of its components. The conditional form of information entropy may help to estimate such contributions of additional variables to the uncertainty of the whole series. The conditional Shannon entropy measures the increase of the system entropy imputable to new observations, in other words the increase in uncertainty due to these new variables. For an entropy of order n, the conditional Shannon entropy linked to an additional variable of length one is given by:

$$CS_s = -\sum_{i=1}^{m} p(x_{i_1}, ..., x_{i_{n+1}}) \log_2 p(x_{i_{n+1}} \mid x_{i_1}, ..., x_{i_n}) \tag{1.9}$$

From another point of view, the conditional entropy determine to which extend the knowledge of a series of outcome may help to predict future values of the variable X.

In his contribution, Schreiber (2000) used this definition of entropy and the theory developed earlier by Granger to define a new approach of causality based on information theory. The main idea of Granger was to test whether the prediction of future values of X knowing its past would be improved by the knowledge of another time series Y. Schreiber followed a similar approach using conditional Shannon entropy rather than linear regression. Its definition of causality or information transfer can be seen as the difference between the information about future observations of X gained from the joint past observations of X and Y, and the information about future observation of X gained from sole past observations of X. Formally, transfer entropy can be seen as the difference between two conditional Shannon entropies. Other definitions of entropy may be used such as the Tsallis entropy (Tsallis, 1988) , the Renyi entropy (Renyi, 1960) or the κ-entropy (Wada and Suyari, 2013). These alternative generalizations of Boltzmann-Shannon information entropy measure rely on additional parameters and emphasize only parts of the probability density function, which could lead to improvements but come to the cost of higher complexity and are beyond the scope of the present chapter.

Assuming X and Y, two stationary Markov processes of order n and m, the transfer entropy determines to which extend the knowledge of the process Y reduces the uncertainty in the future values of the process X given its own past.

$$\begin{aligned}
T_{Y\to X} &= H(x_t \mid x_{t-n}) - H(x_t \mid x_{t-n}, y_{t-m}) && (1.10)\\
&= \sum_{x_t, x_{t-n}, y_{t-m}} p(x_t, x_{t-n}, y_{t-m}) \log_\alpha(p(x_t \mid x_{t-n}, y_{t-m})) && (1.11)\\
&- \sum_{x_t, x_{t-n}, y_{t-m}} p(x_t, x_{t-n}, y_{t-m}) \log_\alpha(p(x_t \mid x_{t-n})) && (1.12)\\
&= \sum_{x_t, x_{t-n}, y_{t-m}} p(x_t, x_{t-n}, y_{t-m}) \log_\alpha \left(\frac{p(x_t \mid x_{t-n}, y_{t-m})}{p(x_t \mid x_{t-n})} \right) && (1.13)
\end{aligned}$$

This definition of transfer entropy can be seen as the deviation of the bivariate system from the assumption of no information flow summarized by the Markov property

$p(x_t \mid x_{t-n}) = p(x_t \mid x_{t-n}, y_{t-m})$. This deviation measures the information flow between Y and X and is called the Kullback-Liebler distance/divergence, see Mezard and Montanari (2009) previously quoted.

1.2.2 Transfer entropy as a nonlinear measure

The origins of transfer entropy capacity to detect nonlinear relationships follows directly from its definition. Unlike Granger causality, no hypotheses are made on the kind of relationship existing between the different series of outcomes. The determination of conditional entropy is based solely on a frequency analysis of specific patterns $[x_t, x_{t-n}, y_{t-m}]$ for which no assumptions are made on the type of dependence structure linking x_t and y_{t-m}.

Following the definition of transfer entropy made in Sec.1.2.1, we know that for a set of conditional probabilities $\big(P(x_t \mid x_{t-n}, y_{t-m})_1, P(x_t \mid x_{t-n}, y_{t-m})_2, ..., P(x_t \mid x_{t-n}, y_{t-m})_k\big)$ with k the number of possible patterns, the conditional entropy $H\left(x_t \mid x_{t-n}, y_{t-m}\right)$ may be reduced if a positive number of conditional probability $P(x_t \mid x_{t-n}, y_{t-m})_j$ differ from the equiprobable case, in other words when the variance of the conditional probabilities set is higher than zero. Once applied to transfer entropy this concept means that in order to observe a directed link between Y and X, the variance of the set of conditional probabilities $P(x_t \mid x_{t-n}, y_{t-m})_j$ should be higher than the one of the set of conditional probabilities $P(x_t \mid x_{t-n})_j$, therefore adding y_{t-m} in the condition should make the original probability distribution less uniform. More than one conditional probability distribution of $P(x_t \mid x_{t-n}, y_{t-m})_j$ complies with this pattern of increased variance, implying that several patterns may describe the link existing between both series of outcomes. The aggregation of these patterns corresponds therefore to a broader class of nonlinear relationships between X and past Y. In this framework, the most interesting patterns are the ones that result in a higher conditional probabilities $P(x_t \mid x_{t-n}, y_{t-m})_i$ compared to the other conditional probabilities $P(x_t \mid x_{t-n}, y_{t-m})_j$ sharing the same x_t and x_{t-n} but with a different y_{t-m}. Indeed they represent the channels use to transfer the information between both systems. As can be seen, transfer entropy determines the causal relationship between two series of outcomes by comparing pairs of conditional probabilities distant from the equiprobable case when conditioning on Y. Transfer entropy may therefore focus only on subsets of the system to directly find the source of the transfer information in contrast with other causality measures such as the linear form of Granger causality test which looks at average and continuous relationships.

1.2.3 Estimation and statistical inference

As showed in Section 1.2.1, the determination of transfer entropy implies the estimation of a set of joint probabilities based on discrete variables. Traditionally, financial series, and returns in particular, are modeled as continuous random variables (usually log-Normally iid series). They thus have to be discretized to allow transfer entropy estimation. There is an extensive literature treating the problem of probability distribution function estimation and more broadly entropy estimation (for an extensive review see Hlavackova-Schindlera et al., 2007). A high number of possible approaches have been explored such as kernel estimators (Moon et al., 1995), maximum likelihood (Paninski (2003)), neural networks (Schraudolph, 2004) or partitioning. Although some of them provided better estimates (Steuer et al., 2002), partitioning has gained in popularity in the empirical literature (Staniek and Lehnertz, 2008; Kwon and Yang, 2008; Dimpfl and Peter, 2013; Wang and Yu, 2012) as the other methods are often computationally intensive which make them difficult to apply to larger dataset. The implementation of partitioning rules appears as a practical solution to limit the number of possible outcomes allowing the determination of joint probabilities using simple frequency analysis. In order to assure an effective partitioning, the stationarity of the sample must be verified. We propose in this chapter to partition the observations into discretized levels following encoding rules based on specific quantiles of the sample. Other partitioning methods have been proposed in the literature (see the generalized binning with B-splines of Daub et al. (2004) or the adaptive partitioning of Darbellay and Vajda (1999) to quote only a few) but considering the specificity of financial variables, the quantile approach is often preferred as it allows to highlight more easily some of the main features of these variables. An alphabet of n symbols implies $n-1$ quantile levels. Assuming an alphabet of three symbols the encoding rules follows:

$$P_t = \begin{cases} 1 & \text{if } o_t \in [0, q_L[\\ 2 & \text{if } o_t \in [q_L, q_U[\\ 3 & \text{if } o_t \in [q_U, 1[\end{cases} \tag{1.14}$$

where q_L and q_U provide the position of the lower and upper quantiles in the distribution function of o_t, the price observed at time t relatively to the maximum value observed over the entire sample.

Every data point is replaced by a specific symbol depending on the above encoding rules. The probabilities are then generated using a simple counting method calculating the number of occurrence of each pattern in a predefined time window.

Inspired by spectral analysis, an extension of the quantile symbolization method is

proposed in this chapter. The new approach intends to differentiate the source of information transfer or causality according to their location on the probability distribution of the time series outcomes. In this framework, we draw away from the equipartioning approach (Butte and Kohane (2000); Daub et al. (2004)) and divide the sample distribution in five regions, two regions of interest positioned on either side of the 50 percent quantile with each a specific symbol and the remaining regions sharing a common symbol. The two regions of interest are characterized by their position relative to the 50 percent quantile and a specific bandwidth. The range of possible position depends directly on the bandwidth of the considered regions, with the number of possible positions decreasing when the bandwidth increases. The aim of this approach is to differentiate what happens in the center or the core of the distribution from what happens in the tails, the extend to which realizations are considered in the core vs. the tail depending of the position and bandwidth parameters. This symbolization process seems especially relevant for financial time series characterized by asymmetric correlation patterns (see Ang and Chen, 2002; Campbell et al., 2002; Bae et al., 2003 or Patton, 2004 among others) and potential nonlinearities.

$$P_t = \begin{cases} 3 & \text{if } o_t \in [0, q_L[\\ 1 & \text{if } o_t \in [q_L, q_L + \delta_L[\\ 3 & \text{if } o_t \in [q_L + \delta_L, q_U - \delta_U[\\ 2 & \text{if } o_t \, [q_U - \delta_U, q_U[\\ 3 & \text{if } o_t \in [q_U, 1] \end{cases} \tag{1.15}$$

where q_L and δ_L give the lower bound and the size of the lower region of interest while q_U and δ_U provide the upper bound and the size of the upper region of interest.

Based on a set of possible bandwidths (δ_L, δ_U), a transfer entropy map is then created by testing every position within the possible range $q_L/q_U = j \cdot sps$ with sps the stepsize and $j \in [0, (0.5 - \delta)/sps]$. In the symmetrical case, same bandwidth and same set of position, this method provides a two dimensional transfer entropy map, the first axis representing the considered bandwidth size and the second one the positions of both regions compared to the 50 percent quantile. Examples of this transfer entropy map are provided in Section 1.3 below when assuming various DGPs for the price time series.

The symbolization methodology used to simplify the complex behaviour of the time series also participates to the ability of transfer entropy measure to detect potential linear or nonlinear causal patterns. By reducing the complexity of the system through categorization, transfer entropy is able to consider x_t sensitivity to specific stimuli

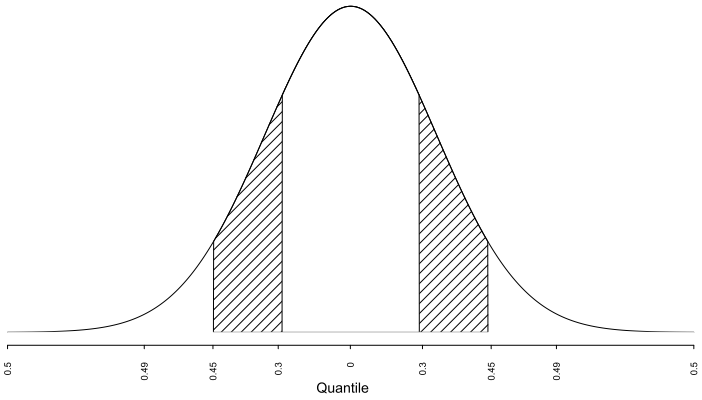


Figure 1.1: Spectral quantile symbolization, symmetrical position 30%, bandwidth 15%

(x_{t-n} and y_{t-m})[3], getting rid of the impact of small changes on the overall relationship as the error terms does to some extent in the Granger causality linear test. Indeed, each region of the probability distribution is considered separately, reducing the potential impact of the noise present in the other regions of the probability distribution. The resulting transfer entropy map should therefore provide more precise information regarding the causal relationships, along the identification of the regions of the probability distribution of time series X and Y having the higher causal dependency. The definition of the bandwidth and the set of positions directly affect the resolution of the analysis.

As recalled in Mezard and Montanari (2009), the determination of the true level of transfer entropy requires processes of infinite length in order to compute the right transition probabilities. In practice however, transition probabilities have to be estimated from a finite sample. Marschinski and Kantz (2002) demonstrated the existence of a bias in the estimation of transfer entropy due to the small sample effect. To limit this bias, those authors proposed an adjustment of the transfer entropy based on a bootstrap methodology. Their unbiased estimation, called " effective transfer entropy", is defined as the difference between the regular transfer entropy and a bootstrapped transfer entropy as follows:

$$ET_{Y\to X} = T_{Y\to X} - T_{Y_{Boot}\to X} \tag{1.16}$$

where $T_{Y_{Boot}\to X}$ represents the transfer entropy with the series Y being bootstrapped.

[3]In our simulation exercises, $n = m = 1$ but results could easily be extended to higher order processes.

Through this bootstrapped Y, all statistical dependencies between the two series are removed. As $T_{Y_{Boot}\to X}$ converges to zero for increasing sample size, all non zero values of $T_{Y_{Boot}\to X}$ demonstrate the presence of a bias due to small sample effects and should therefore be removed. We propose to rely on this methodology to construct our statistical test of the existence of a causal relationship using transfer entropy. The statistical test regroups the following steps:

- Step 1: Estimate n times the transfer entropy measure based on n independent draws from n resampled series Y (bootstrapped Y).
- Step 2: Build a probability distribution function of the bootstrapped transfer entropies, $\mathcal{K}(T_{Y_{Boot}\to X})$ using a Kernel approach (Gaussian Kernel)
- Step 3: Determine the bias at a specific confidence level α using the Kernel distribution $\mathcal{K}^{\alpha}(T_{Y_{Boot}\to X})$
- Step 4: Estimate the effective transfer entropy by removing the estimated bias from the original transfer entropy $ET_{Y\to X} = T_{Y\to X} - \mathcal{K}^{\alpha}(T_{Y_{Boot}\to X})$

The Kernel density gives the probability distribution of transfer entropy levels in the case of no information transfer. A positive effective transfer entropy confirms the existence of a transfer of information, if null or negative no information transfer may be found. Indeed the null hypothesis for the transfer entropy is $T_{Y\to X} = 0$ which is equivalent to $p(x_t \mid x_{t-n}) = p(x_t \mid x_{t-n}, y_{t-m})$ but once the small sample effect is taken into account the null hypothesis becomes: $T_{Y\to X} > T_{Y_{Boot}\to X}$ or every possible values of $T_{Y_{Boot}\to X}$ with the statistical test given by:

$$\mathcal{T}_{Y\to X} \sim_{H_0} \mathcal{K}(T_{Y_{Boot}\to X}) \tag{1.17}$$

Regarding the bootstrap of Y, we rely on a block bootstrap to maintain time dependence of the original data within the bootstrapped time series. We draw from a uniform distribution both the position and size of every blocks. The resulting blocks are concatenated to form a new time series for Y.

1.3 Simulations and first conclusions

In order to explore in more detail the properties of the models presented in Section1.2.1, we rely on Monte Carlo simulations performed on various data generating process allowing for linear vs. nonlinear specification and dependencies in mean and variance between the two processes of interest, Y and X. In particular, we try to disentangle

under which circumstances the two approaches, i.e. Granger causality vs. transfer entropy could lead to significant differentiated results. We proposed four different models that in a simplified way mimic the main behaviors of financial time series following the current literature, a Vector Autoregressive model, a Markov Switching model (Psaradakis et al., 2005a) and a Co-jump model for the causal processes (Andersen et al., 2007) in mean and a Multivariate Stochastic Volatility process with Granger causality for the causal processes in variance (Yu and Meyer, 2006). Indeed Sewell (2011) demonstrates that the distribution of returns in financial time series is often non-stationary due to the volatility clustering and has a positive kurtosis increasing when the time interval between two observations decreases. Non-linearities are also reported in both the mean and variance of returns. All these elements tend to demonstrate the interest of considering the proposed set of data generating processes.

1.3.1 Data generating processes

Simulation 1: VAR specification

The primary test of both methodologies involves the classical VAR approach. This first model allows to compare the behavior of both techniques using the assumption behind the most common definition of Granger Causality. For simplicity, we will consider a linear coupling in only one direction. Let's assume two autoregressive processes X and Y, with X being also influenced by Y in a linear way. The value x and y at time t are determined as follows:

$$x_t = \alpha x_{t-1} + \beta y_{t-1} + (1 - \alpha - \beta)\varepsilon_{t_x} \tag{1.18}$$

$$x_t = \delta y_{t-1} + (1 - \delta)\varepsilon_{t_y} \tag{1.19}$$

The coupling strength is determined by the factor β for which a range of values from 0.5 to 0.05 is used. The other parameters are chosen based on the estimation of a VAR model on a set of financial data regrouping equities, bonds, CDS and foreign exchange leading to an average value of the parameters α and δ of 0.06. The error term is defined by an uncorrelated Gaussian noise with zero mean and unit standard deviation and the factor $(1 - \alpha - \beta)$ is applied in order to get a unitary evolution in the long term. The initial value of X and Y are determined by the same zero mean white noise process.

Simulation 2: Markov Switching specification

We investigate in a second step, the effect of nonlinear relationships on our causality tests. A Markov Switching Granger Causality model is added to the two other causality

measures to assess the ability of Granger causality framework to identify non-linear relationships. This model is a simplified version of the Markov Switching Granger Causality technique proposed by Psaradakis et al. (2005a). The two states characterizing the Markov Switching part of the model determine the presence or absence of a link between the variable X and Y.

$$x_t = \alpha x_{t-1} + \varepsilon_t \quad \text{when } s_t = 0 \tag{1.20}$$
$$x_t = \alpha x_{t-1} + \beta y_{t-1} + \varepsilon_t \quad \text{when } s_t = 1 \tag{1.21}$$

The transition between both states is defined by the following transition probabilities:

$$p_{ij} = \mathbb{P}(s_t = j \mid s_{t-1} = i) \tag{1.22}$$

The statistical inference is performed via the t-ratio of the estimated β.

The first nonlinear DGP to be considered is the Markov Switching VAR with two regimes which defines the behavior of the variable X, the variable Y following a simple VAR. The first state is characterized by a one direction linear coupling between Y and X and the second one by a simple VAR.

$$x_t = \alpha x_{t-1} + \beta y_{t-1} + (1 - \alpha - \beta)\varepsilon_{t_x} \quad \text{when } s_t = 0 \tag{1.23}$$
$$y_t = \delta y_{t-1} + (1 - \delta)\varepsilon_{t_y} \quad \text{when } s_t = 1 \tag{1.24}$$

The transition between both states depends on a simple normal distribution function with zero mean and one unit standard deviation with a probability of observing a transmission from Y to X varying between 80 percent and 1 percent. The parameter β is fixed at a 0.25 level and the α and δ keep their value of 0.06 used in the previous VAR data generating process. The error terms follow the usual zero mean white noise process.

Simulation 3: Co-jump specification

The second nonlinear simulation is based on a causal extension of the co-Jump process developed by Andersen et al. (2007) which allows us to look at the impact of fat tails on the causal link detectability. This causal co-jump process features a diffusion process, a specific jump component for each variable and a co-jump component lagged by one period for the variable X compare to Y to generate the causal relationships from Y to X. Usually written in continuous time, we can represent the co-jump process in

discrete time as follows:

$$x_t = \left(m_x - \lambda_{jx}\left(e^{\frac{m_{jx}+\sigma_{jx}^2}{2}} - 1\right) - \lambda_{cj}\left(e^{\frac{m_{cj}+\sigma_{cj}^2}{2}} - 1\right)\right) + \sigma_x \varepsilon_t^x + jx_t + cj_t \quad (1.25)$$

$$y_t = \left(m_y - \lambda_{jy}\left(e^{\frac{m_{jy}+\sigma_{jy}^2}{2}} - 1\right) - \lambda_{cj}\left(e^{\frac{m_{cj}+\sigma_{cj}^2}{2}} - 1\right)\right) + \sigma_y \varepsilon_t^y + jy_t + cj_t \quad (1.26)$$

with m the mean value, λ_{jx}, λ_{jy} and λ_{cj} the intensity of respectively the jump process of X, Y and the causal co-jump, σ the variance of the different processes and ε a Gaussian white noise.

The occurrence of both types of jump is determined by a Poisson process with the intensity for the causal co-jump ranging from 5 percent to 0.5 percent. The other parameters are once again estimated based on the set of financial data used for the calibration of the VAR model.

$$jx_t = \left(N_t^{\lambda_{jx}} - N_{t-1}^{\lambda_{jx}}\right)\left(e^{m_{jx}+\sigma_{jx}^2\varepsilon_{t_{jx}}} - 1\right) \quad (1.27)$$

$$jy_t = \left(N_t^{\lambda_{jy}} - N_{t-1}^{\lambda_{jy}}\right)\left(e^{m_{jy}+\sigma_{jy}^2\varepsilon_{t_{jy}}} - 1\right) \quad (1.28)$$

$$cj_t = \left(N_t^{\lambda_{cj}} - N_{t-1}^{\lambda_{cj}}\right)\left(e^{m_{cj}+\sigma_{cj}^2\varepsilon_{t_{cj}}} - 1\right) \quad (1.29)$$

where N^λ follows a poisson process with $\mathcal{P}(N^\lambda = k) = e^{-\lambda}\frac{(\lambda)^k}{k!}$ and λ the intensity of the Poisson process.

Simulation 4: Multivariate Stochastic Volatility specification

The fourth type of process, to be considered in this chapter, is a Multivariate Stochastic Volatility process with Granger causality which will benchmark the ability of our causality measures to capture relationship in variance. We rely here on the Lagrange Multiplier test (LM test) developed by Hafner and Herwartz (2006) to compare the results of the three previous causality measures with a test that has been designed specifically for causality in variance. This causality measure is an adaptation of the general Lagrange Multiplier misspecification test, introduced by Lundbergh and Terasvirta (2002), to the estimation of univariate GARCH (1,1) models.

$$h_{t+1_x} = \alpha_x + \beta_x x_t^2 + \delta_x h_{t_x} \quad (1.30)$$

$$x_t = exp\left(\frac{1}{2}h_{t_x}\lambda_t\right)\varepsilon_{t_x} \qquad \lambda_t = 1 + s_{ty}\theta \quad (1.31)$$

where $s_{ty} = (Y_t^2, h_{t_y})'$. The existence of a causal relationship in variance between X and Y is controlled by the parameter θ so that the null and alternative hypothesis of

the LM test are $H_0 : \theta = 0$, $H_1 : \theta \neq 0$. The LM test is constructed by means of the score of the Gaussian log-likelihood function of ε_{t_x} evaluated at $\theta = 0$ and the variance of the set of parameter $\psi_x = (\alpha_x, \beta_x, \delta_x)$ evaluated at $\theta = 0$, noted $\Omega = V(\psi_x)$.

$$LM = (1/4T)(\sum_{t=1}^{T}(\varepsilon_{t_x}^2 - 1)s'_{ty})\Omega^{-1}(\sum_{t=1}^{T}(\varepsilon_{t_x}^2 - 1)s_{ty}) \sim_{H_0} \chi^2(2) \quad (1.32)$$

$$\Omega^{-1} = (\kappa/4T)(\sum_{t=1}^{T} s_{ty}s'_{ty} - \sum_{t=1}^{T} s_{ty}z'_{tx}(\sum_{t=1}^{T} z_{tx}z'_{tx})^{-1}\sum_{t=1}^{T} z_{tx}s'_{ty}) \quad (1.33)$$

where $\kappa = (1/T)\sum_{t=1}^{T}(\varepsilon_{t_x}^2 - 1)^2$, $z_{t_x} = (1/h_{t_x})(\partial h_{t_x}/\partial\psi_x))$

The Multivariate Stochastic Volatility process (Yu and Meyer, 2006), considered for the analysis of causality in variance, is specified as follow:

$$x_t = exp\Big(\frac{1}{2}h_{t_x}\Big)\varepsilon_{t_x} \quad (1.34)$$

$$y_t = exp\Big(\frac{1}{2}h_{t_y}\Big)\varepsilon_{t_y} \quad (1.35)$$

$$h_{t+1_x} = \mu_x + \alpha(h_{t_x} - \mu_x) + \beta(h_{t_y} - \mu_y) + \eta_{t_x} \quad (1.36)$$

$$h_{t+1_y} = \mu_y + \alpha(h_{t_y} - \mu_y) + \eta_{t_y} \quad (1.37)$$

where η_{t_x} and η_{t_y} are Gaussian white noise with zero mean and variance σ_x and σ_y respectively. The parameters μ, α and η have been estimated based on the set of financial data use previously (resp. -8, 0.7 and 0.8).

1.3.2 Results

The Monte Carlo exercise consists in the simulation of times series for the four different DGPs presented in the previous section. We rely on the modulation of two parameters to compare the different causality measures; namely the parameters controlling the strength of the link between both times series as well as the length of the times series used to infer the causal relationship. This second parameter is specifically designed to challenge transfer entropy, as the small sample effect which affects its ability to infer properly causal relationships, increases when the window size reduces. The causal link is tested in both directions, from Y to X and from X to Y, allowing us to compute the true positive and false positive rates for both parametrizations. For each values of the strength parameters, 500 samples are simulated each of them containing 1500 observations. Turning to the times series length, we rely on the same number of samples and test window sizes ranging between 100 and 3000 observations[4].

[4]The strength parameter of each DGP is fixed for the entire set of the window sizes with $\beta = 0.15$ for the VAR specification, a probability of regime switch fixed at 4% for the Markov switching VAR

In addition to the power and size of the considered statistical test for the different DGP, we additionally report the performance of the considered approaches in terms of receiver operating characteristic curve (ROC) and precision-recall curve (PR) (see Marbach et al., 2012). The PR curve measures the balance between the result relevancy, i.e. its precision and the recall which gives the percentage of truly relevant results returned by the test. The ROC curve provides a performance index for binary classifiers, reporting the true positive rate against the false negative rate. The higher the area under the curve the better the classifier with a value of 0.5 for a random classifier. Both curves are estimated based on sample sizes of 1500 observations. For all four comparison tools, the existence of a causal relationship is estimated for each pair of times series by using the statistical inference methodologies presented previously, i.e. a Wald test for Granger causality, a t-ratio for Markov Switching Granger causality, the LM test for the GARCH based causality model and a bootstrapped estimation for the transfer entropy. The average percentage of true and false positive rates are then computed based on the 500 samples generated. A significance level of 5% is selected for all causality measures except for the estimation of the ROC and PR curves where a set of 20 different significance levels are used.

For each DGP, the position and bandwidth used in the transfer entropy estimation are selected in order to maximize its effective value. This optimization is done via the estimation of transfer entropy based on spectral symbolization[5]. For both VAR and Markov Switching DGPs, the position is set to 20 percent and the bandwidth to 30 percent as we face an average relationship. As for the causal co-jump process, the jumps occurring in the tails of the distribution, a position of 49 percent and a bandwidth of 1 percent are preferred. Finally, for the Multivariate Stochastic Volatility process with Granger causality, the highest transfer entropy is found for a position of 5 percent and a bandwidth of 45 percent. Fig.1.1,1.2 and 1.3 summarize the results for the three DGPs designed to simulate causal relationships in mean.

Result 1: VAR specification

As can be seen from Fig.1.1, the Granger causality performs clearly better than the transfer entropy for the VAR specification which is not surprising as the VAR represents the underlying framework of the Granger causality measure. However, when considering larger data sets, both measures seem to perform equivalently. This result tends to confirm the equivalence between Granger causality and transfer entropy,

specification, the causal jump intensity at 0.015 and $\beta = 0.75$ for the multivariate stochastic volatility specification.

[5]Detailed results are available on request from the authors.

highlighted by Barnett et al. (2009), once the issue of small sample effect has been dealt with.

Result 2: Markov Switching specification

When turning to the regime switching VAR specification, the Markov Switching Granger causality framework does not seem to perform better than Granger causality or transfer entropy, with Granger causality being better in term of power for almost the entire sets of strength parameters and window sizes but also a higher PR curve. We also see that the performance of the Markov Switching Granger causality worsen when the window size increases which could be due to the higher complexity of this model and convergence problems in the maximum likelihood estimation procedure. The good results of the Granger causality for the nonlinear Markov Switching specification demonstrate that Granger causality does not need a continuous relationship between the two processes to work properly and that as long as Y brings, at specific moment, information reducing the uncertainty about the outcome of X, the model is able to detect accurately the occurrence of information transmission. We again note the equivalence between Granger causality and transfer entropy for large datasets.

Result 3: Co-jump specification

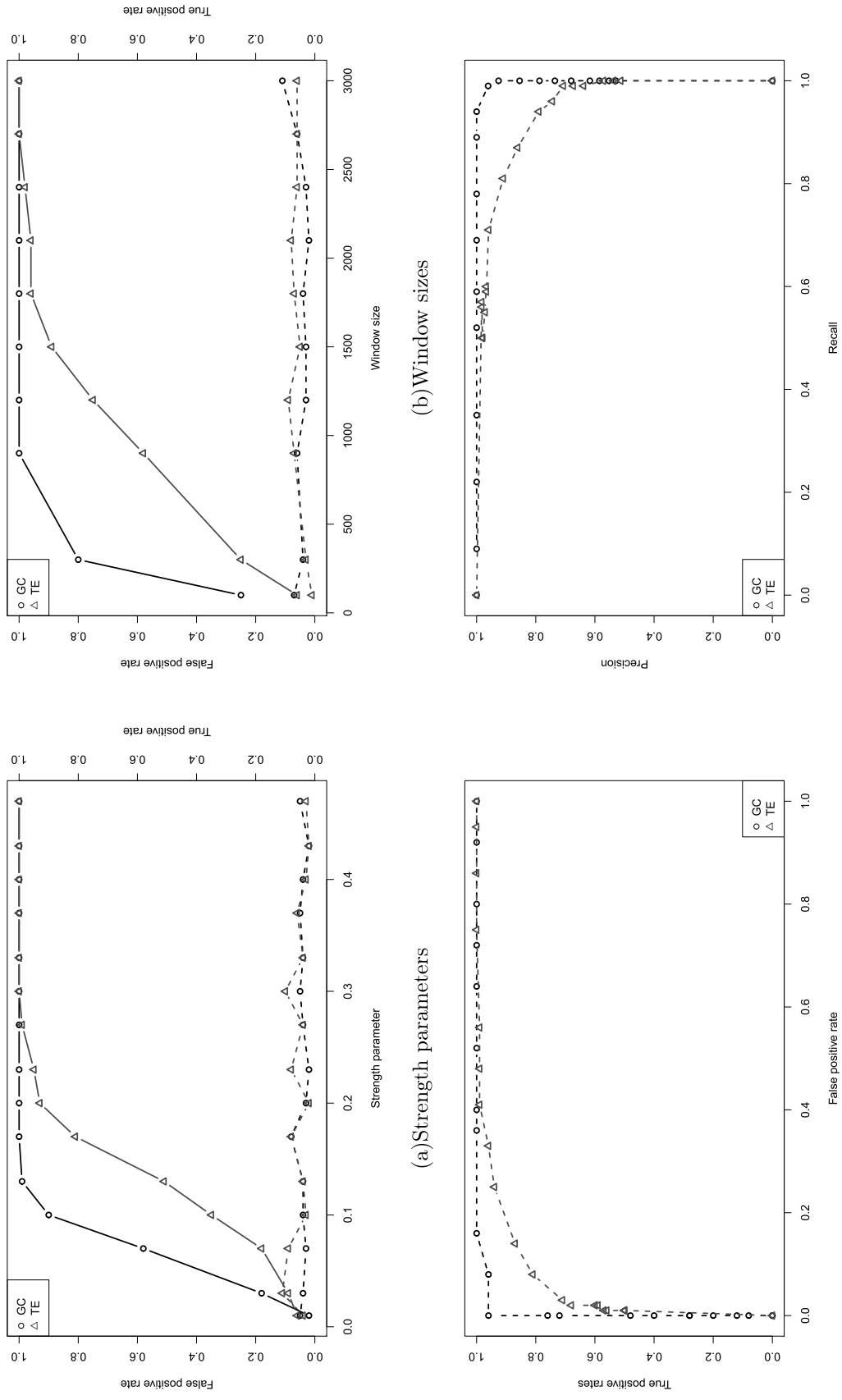
Although Granger causality performed better for the two VAR-based DGPs, the results of Fig.1.3 demonstrate the ability of transfer entropy to outperform Granger causality when facing highly nonlinear processes caused by rare events. Regarding the size, we observe that transfer entropy has, in most cases, a lower rate than the two other causality measures. The results are especially striking when considering the ROC and PR curves. Besides, in the case of rare events transfer entropy seems to suffer less from a small sample bias.

These good results of transfer entropy for the causal co-jump DGP may be partly explained by its ability to focus only on specific patterns to determine the causal relationship. In contrast with transfer entropy, Granger causality tries to define an average relationship using the whole sample, which means in this case, mainly data where no information transfer occurs. The Granger causality test being based on the comparison between the variance of the error terms from both the restricted and unrestricted models, the few observations helping to predict future states of X are not numerous enough to modify significantly the variance of the unrestricted model. It seems therefore that Granger causality needs more occurrences to be able to infer a causal link in presence of rare causal events.

On the contrary, in the case of continuous linear causal relationships, the transfer entropy loses this advantage because the whole set of probabilities become a source of relevant information. In this situation, the higher resolution of Granger causality leads to a better performance. Indeed the quantile symbolization limits the capacity of transfer entropy to observe information transfer when facing low transfer coefficient. Increasing the number of categories may seem at first sight a convenient solution but it usually takes place alongside with an increase of the small sample effect.

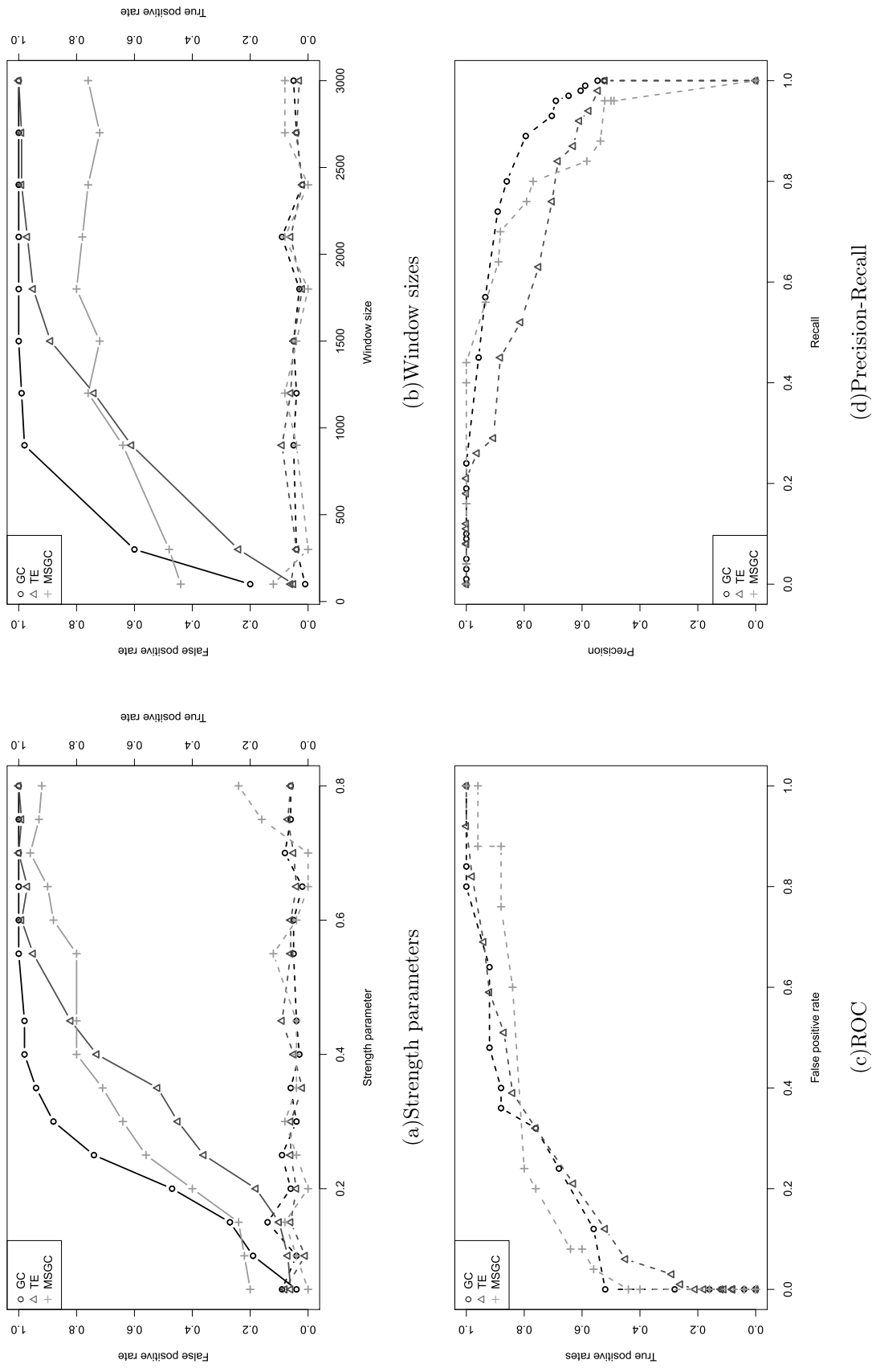
Result 4: Multivariate Stochastic Volatility specification

Finally, turning to the results of the Multivariate Stochastic Volatility process, it shows that transfer entropy is much more effective in detecting causal relationships in variance than simple Granger based models. Transfer Entropy even performs better than the dedicated Lagrange Multiplier test for the higher transfer rates. These results tend to prove that transfer entropy may be used as a broader causality measure, taking into account relationships in linear as well as nonlinear models and both in mean and variance. These results constitute only an overview of the capabilities of transfer entropy. However, they show that for some form of nonlinearity and skewness in the data, transfer entropy leads to better identification of causal relationships. Those features are common for financial time series (see Jondeau et al. (2007)). As such, transfer entropy should be seen as a good complementary causality measure to be used along the usual Granger causality when dealing with financial data. The question of their complementarity will be further investigated in the empirical application detailed in Section 1.4. Eventually, when considering the complexity of the considered models in terms of number of parameters to be estimated, transfer entropy being non parametric, it presents a clear advantage in term of parameter parsimony. The results of our simulations tend indeed to confirm the negative impact of the number of parameters to be estimated on the models ability to infer causal relationships, looking at the results of the Markov switching specification when considering large datasets or at the LM specification for higher transfer rates. However, this advantages is partly compensated by the small sample effect which affects more the transfer entropy.



(a)Strength parameters (b)Window sizes (c)ROC (d)Precision-Recall

Table 1.1: Figures a-d report for the VAR data generating process respectively the evolution of size and power for the set of strength parameters, for the set of window size, the ROC curve and the Precision-Recall curve for the Granger causality and transfer entropy. For Figures a-b, the dotted and solid lines reports respectively the size and power.



(a)Strength parameters (b)Window sizes (c)ROC (d)Precision-Recall

Table 1.2: Figures a-d report for the regime switching VAR data generating process respectively the evolution of size and power for the set of strength parameters, for the set of window size, the ROC curve and the Precision-Recall curve for the Granger causality, Markov Switching Granger causality and transfer entropy. For Figures a-b, the dotted and solid lines reports respectively the size and power.

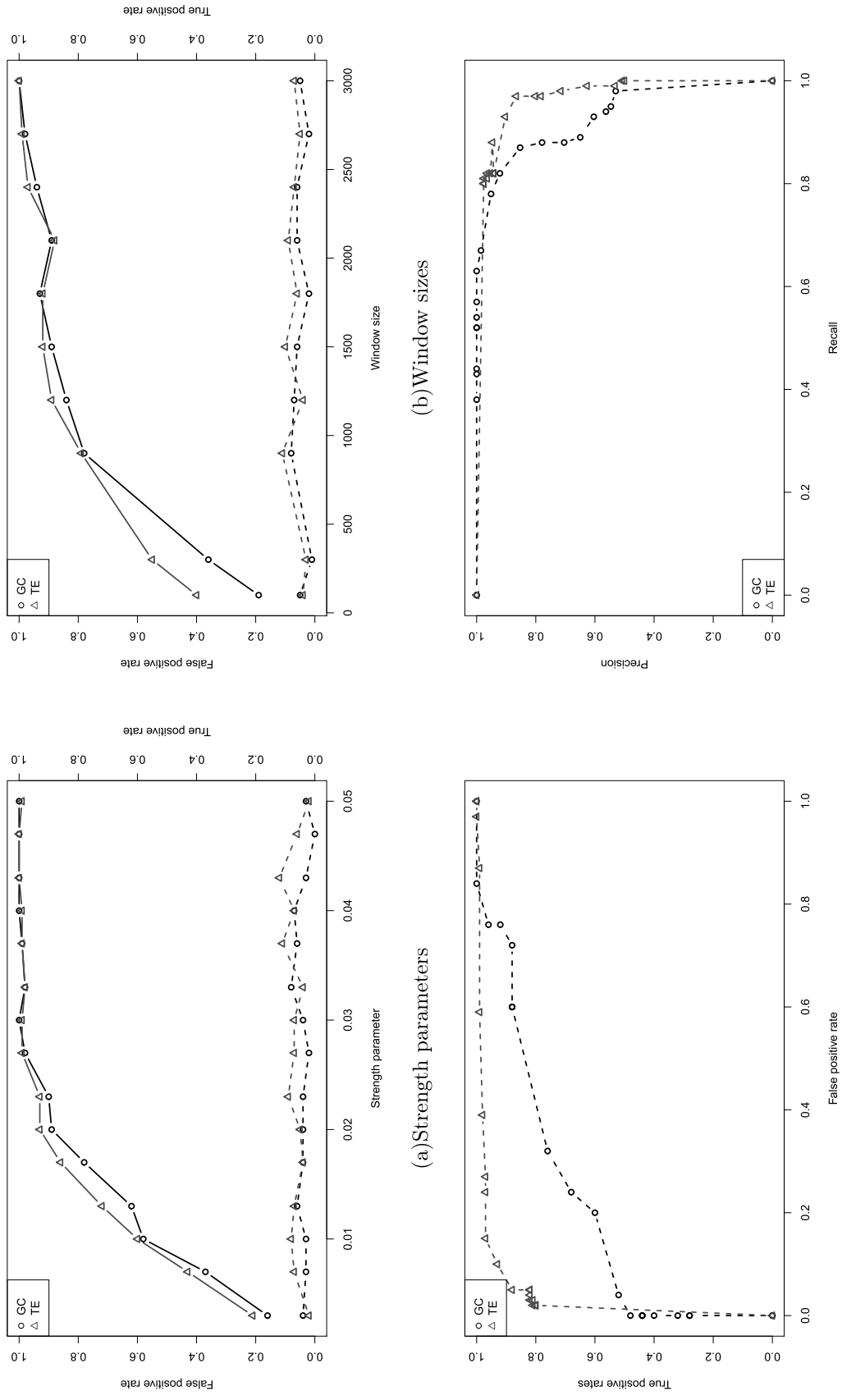


Table 1.3: Figures a-d report for the causal co-jump data generating process respectively the evolution of size and power for the set of strength parameters, for the set of window size, the ROC curve and the Precision-Recall curve for the Granger causality and transfer entropy. For Figures a-b, the dotted and solid lines reports respectively the size and power.

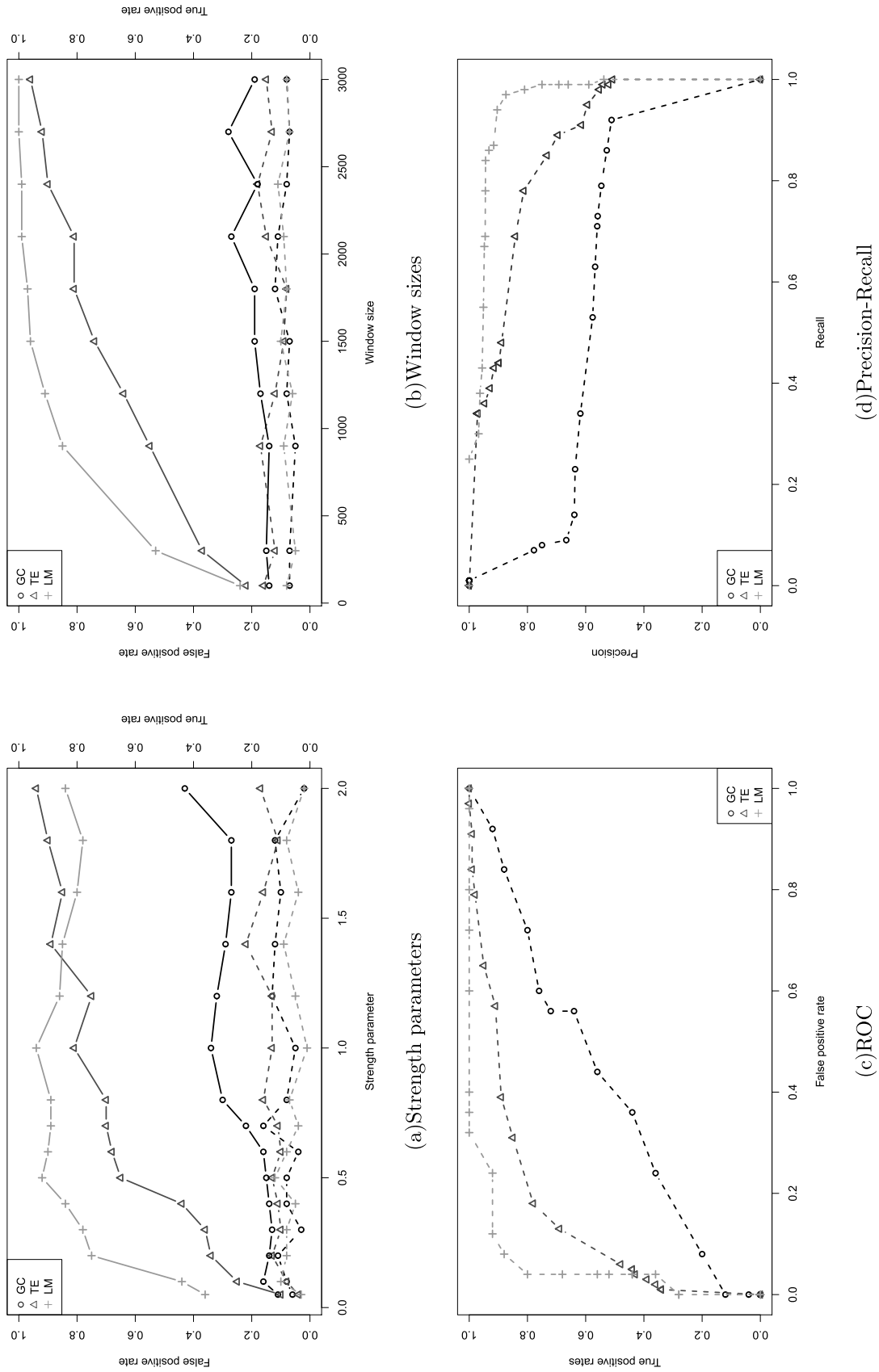


Table 1.4: Figures a-d report for the Multivariate Stochastic Volatility data generating process respectively the evolution of size and power for the set of strength parameters, for the set of window size, the ROC curve and the Precision-Recall curve for the Granger causality, the LM test and transfer entropy. For Figures a-b, the dotted and solid lines reports respectively the size and power.

1.4 Application to European and U.S. financial sector

1.4.1 Motivations

We propose in this section an empirical application on a selected set of European and U.S. financial institutions to document the complex relationship that exists between the network characteristics of the financial system and its overall systemic risk level. The network of interest for this application is composed of 24 Systematically Important Financial Institutions, the so-called "SIFIs", active in the U.S. and in Europe. We believe as Battiston et al. (2012) or Diebold and Yilmaz (2014), that focusing on a small group of financial institutions is sufficient to capture the overall level of systemic risk within the global financial system as long as those institutions represent the main sources of distress dissemination due to their central position within the system. Our endogenous variable consists thus in the overall risk measure of the system, defined as the average level of risk of the financial institutions composing our network, and is approximated using the Max%Loss approach of Ballaa et al. (2014). The network is generated by relying alternatively on the linear form of Granger causality (GC) as well as the transfer entropy (TE) measures. We then compute various topological metrics in the vein of Diebold and Yilmaz (2014), such as average connectivity, eigenvector centrality or betweenness centrality measures to describe the evolution of risk over time. Beyond the determination of the explanatory power of those alternative topological characteristics on risk, which has already been addressed to some extent in the empirical finance literature, a second objective of this application is to understand to what extend the selected causality measure (GC vs. TE) may affect the information content of these topological characteristics with regard to systemic risk prediction. In other words we aim to study whether: (i) one specific causality measure allows to convey more relevant information than the other with regard to systemic risk prediction and in case both measures do not convey the same type of information, (ii) whether they are complementary or substitutes. We aim to address those questions by considering first separately the explanatory power of each measure, then jointly.

The algorithm we use to infer our SIFI network is designed to partially address the problem of omitted variables by conditioning the information transfer between institutions on the relevant environment surrounding both the emitter and receiver of information (for more details see Dahlqvist and Gnabo (2018)). But this relevant environment is not necessarily limited to the institutions included in the network of interest. We thus propose while focusing our attention on the SIFI network, to consider additionnally external sources of information that would come outside this group of 24 financial institutions by including incrementally in the network inference process,

three additional groups of financial institutions composed of banks, insurance companies and real estate companies respectively. This allows us to address partially the difficult question of network sampling, that is the fact that network topology may differ substancially when assessed at the sample level (when nodes and/or links are not fully observed, or selected randomly) compared to the (generally unobserved) population level. Here we limit the discussion to the exploration of the accurate network sample size, that is the number of units (nodes and links between them) that are relevant in our context to infer stable and salient characteristics with regard to systemic risk prediction, see Krivitsky and Kolaczyk (2015) or more recently Kolaczyk (2017) for an extensive discussion on the topic. More concretely, we study the possible added value of including the surrounding environment (vs. restricting the analysis to the sole network of SIFIs) in the definition of the relationship between the SIFIs and determine how it could improve the explanatory power of the network topological characteristics of the SIFI sub-network in the prediction of the overall systemic risk. We also propose to compare how different causality measures deal with increasing sample sizes.

1.4.2 Data description

Our dataset consists in the corporate CDS spreads of 104 financial institutions from both Europe and the United States. Theses financial institutions may be labelled in four different categories: (i) the SIFIs on which our study focuses[6], that we disentangle from the other (ii) banks, (iii) insurance companies or (iv) real estate companies from the sample. The CDS spreads represent the daily midpoints of bid-ask prices of 5-year contracts, which is the most widely used measure in the literature due to their higher liquidity. The data cover the period from January 2005 to January 2016 which corresponds to 2896 daily observations for each financial institution. We included companies who have filed for bankruptcy, have been restructured or merged with other financial institutions mainly during the turmoil of the last financial crisis, including all the available data. The selected institutions provide observations for at least 60 percent of the considered time frame. To cope with potential non-stationarity, we estimate the daily variations of the CDS spreads.

1.4.3 Network inference and topological measures estimations

The estimation of the network using TE starts by the definition of the optimal quantiles used for the symbolization of the times series. To that aim, we rely on the spectral quantile symbolization method developed earlier, to obtain for each pair of financial

[6]We identify the SIFIs according to the G-SIB ranking established in 2015 by the Financial Stability Board, see **http://www.fsb.org/2015/11/fsb-publishes-the-2015-update-of-the-g-sib-list/.**

institutions the optimal bandwidth and position in order to define precisely the regions of the probability distribution where the main information transfer occurs. As can be seen in Fig.1.2, we begin by estimating a TE map and then select the bandwidth and position giving the highest value of TE. We use a stepsize of 0.05 for the position and select six different bandwidths (0.05, 0.1, 0.2, 0.3, 0.4, 0.45 resp.). We then estimate the TE at a significance level of 0.05 and use the whole sample of 2895 observations for the estimation in order to get an average relationship. With 104 financial institutions, we get $104 \cdot (104 - 1) = 10712$ optimal bandwidths and positions which are regrouped in two adjacency matrices further used to generate the networks. We see in Fig.1.2 that on average both regions of interest are located in the tails of the probability distribution rather than in the core. The slope of the curve being relatively small, it implies that the size of the bandwidth is usually important with on average a value of 0.18, which tends to prove that we are not looking only at extreme events.

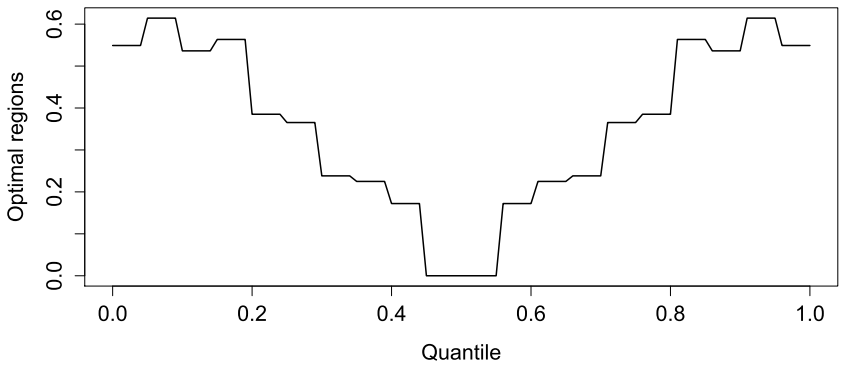


Figure 1.2: We see here the average optimal regions on either side of the 50 percent quantile. The value reported on the Y-axis, represents the number of time the quantiles reported on the X-axis have been selected to be part of the range of quantile included in the two regions of interest.

Equipped with the optimal positions and bandwidths for the TE estimation, we apply the effective network inference algorithm on the four different nested datasets, the first one including the sole 24 SIFIs, the second adding 32 banks to the previous, then, 33 insurance companies in the third and finally, adding an additional set of 15 real estate companies to complete the environment. To account for network dynamicity, we follow the literature by applying a rolling window approach Dungey et al. (2012); Billio et al. (2012) with a window size of 480 observations corresponding to two years (considering 20 trading days per month on average) and a stepsize of 5 observations corresponding to a week. The size of two years has been selected to have sufficient observations to limit the small sample effect for the estimation of TE measures. The effective network inference algorithm is divided in two steps, a first one where a pre-search analysis is

performed, connecting institutions using simple pairwise causality estimation, and a second step where the network is cleaned from indirect links using conditional causality estimation. For the first step, the significance level is set for both causality measures to 0.01 while for the second one a significance level of 0.05 is preferred as the conditioning reduces the detectability of causal links due to the higher dimension of the estimation. We also choose, for the estimation of the network, a time horizon of two lags, meaning that the presence of causal links will be tested for a lag of one and two periods, the resulting estimated networks will provide therefore two adjacency matrices, one for each lag.

Band/Pos	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45
0.05	0.0027	0.0001	0	0	0.0012	0	0	0	0.0072
0.1	0.0024	0.0020	0	0	0	0	0	0.0051	
0.2	0.0064	0.0072	0.0050	0.0007	0	0.00243			
0.3	0.0163	0.0101	0.00204	0.0005					
0.4	0.0139	0.0010							
0.45	0.0026								

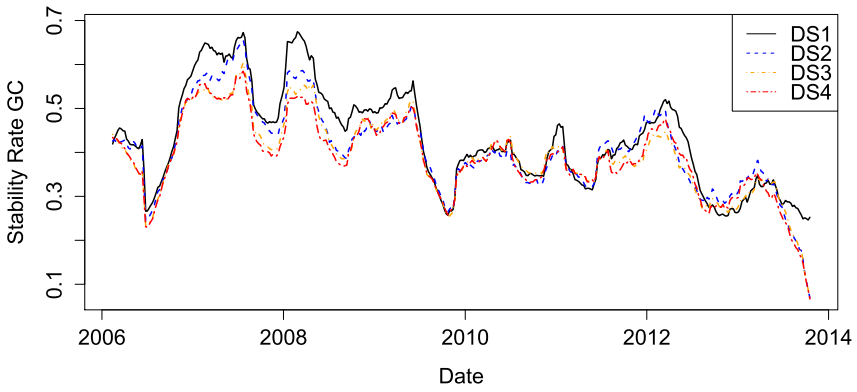
Table 1.5: This table reports the transfer entropy estimation of the information transfer from ING to Bank of America for the different bandwidths and positions. As can be seen from the results, the maximum transfer entropy corresponds to a bandwidth of 0.3 and a position of 0.05.

Once the set of networks have been estimated for the four datasets, we extract the information relative to the sole 24 SIFIs institutions from the various networks considered, which implies to reduce the size of the resulting adjacency matrices and neglect all lines and columns relative to other types of institutions. Doing so, we are able to better capture the impact on systemic risk prediction, of incorporating a wider information set when generating the network. In a second step, we apply on these sub-networks a second set of rolling windows to limit the instability of the observed connections, caused by the sensitiveness of the estimated networks to outliers. This second set of rolling windows have a window size of 20 networks covering therefore a period of 580 observations, and a stepsize of one period corresponding to one week. We then sum, for each window, the adjacency matrices of the 20 networks for both lags and apply a filter to remove the less stable connections. The filter assigns a connection between two financial institutions if the link appears more than a predefined percentage of time in the summed adjacency matrix. This percentage has been optimized for both causality measures to maximize the information content of the estimated network, leading for all datasets to a value of 15 percent for GC and surprisingly a value of 0 percent for the TE, meaning no filtering. Equipped with the filtered networks, we compute a stability measure for both GS and TE, expressing the average percentage of time a connection is observed between two connected institutions. This stability measure provides useful insights about the average strength of the links inferred by our two causality measures

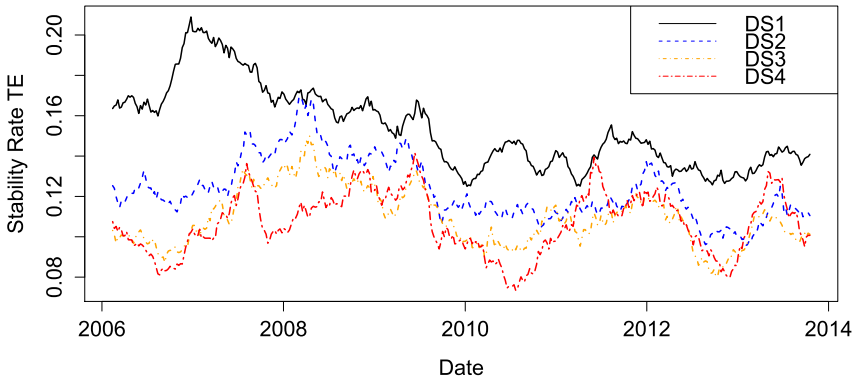
and their ability to detect causal links. Lower values of this stability index would indicate less reliable causal estimations in the case of linear relationships or a higher sensitivity to either rare events or outliers. From Fig.1.3 (a-b), we see that the stability is much higher in the case of GC based networks which could be explained by the higher sensitivity of TE to rare events. There is also a relatively strong difference for the networks estimated with TE between the first dataset (DS1) and the three others (DS2-4) which implies that TE is much more sensitive to the background noise brought by the peripheral institutions. This result tends to demonstrate that (i) the question of the optimal sample size is indeed crucial since network characteristics between core SIFIs institutions vary significantly when considering various information sets, (ii) the optimal network sample size may differ from one causality measure to another. The linear regression results should provide further information regarding the information content of the inferred networks. We see nevertheless for both causality measures a higher stability rate during the financial crisis.

Fig.1.3 (c) reports the average distance between the networks estimated with TE and the ones estimated with GC, with a value of 1 for a perfect match between both networks and 0 for no common connections. We see that apart from the period just before the crisis, the two measures seem to give a quite different view of the relationships between the SIFIs, meaning that they should bring different information about the underlying network.

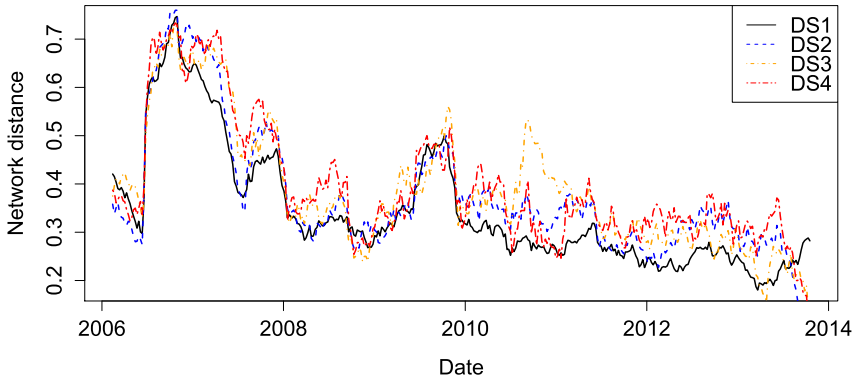
We finally use the filtered adjacency matrices to compute the different topological characteristics allowing us to define how the topology of the network contributes to the risk of the system. We propose here to follow Ballaa et al. (2014) or Billio et al. (2012) and carry out linear regressions. However, in contrast with their approaches, we perform temporal linear regressions on the dependent and explanatory variables averaged for all the 24 SIFIs of our network, highlighting how the sub-network works as a whole and how the evolution of the risk level could be explained by the considered explanatory variables. In line with Ballaa et al. (2014), we define the risk/fragility of an institution as the maximum percentage financial loss during a specific period of time; also called Max%Loss. The considered period of time corresponds to the same period of 580 observations used for the determination of the filtered networks. We then estimate four different topological measures based on these filtered networks, namely: the average number of connections inside the network, the eigenvector centrality, the betweeness centrality as well as the clustering coefficient. The eigenvector centrality measures the influence of each node on the rest of the network based on the idea that connections to highly connected nodes contribute more to risk diffusion. As for the betweeness centrality it assesses the influence of a node by computing the number of shortest paths from all the other nodes of the network that pass through the considered node. Finally



(a) Stability measure of the networks estimated with Granger causality



(b) Stability measure of the networks estimated with transfer entropy



(c) Network distance

Figure 1.3: Graphs a and b display the evolution in time of the stability of the connections existing inside the network estimated with both causality measures. Graph c presents the distance between both network representation

the clustering coefficient estimates the probability that adjacent nodes are connected [7].

As these different measures are based on close concepts, we start by performing a VIF test to look for possible multicolinearity. The result shows no sign of multicolinearity between these measures for almost all datasets which allows us to include those various explanatory variables simultaneously in our regressions as their information content does not seem redundant. The usual preliminary analysis is then performed before each regression to test for heteroskedasticity and autocorrelation. Having detected signs of autocorrelation of the error terms[8], we estimate the heteroskedasticity and autocorrelation consistent (HAC) covariance matrix using the procedure developed by Newey and West (1987) to correctly compute the t-value and significance level of the explanatory variables.

Dataset/Measure		**Transitivity**		**Betweeness**		**EV Centrality**		**Connectivity**		R^2
		coeff	t-stat	coeff	t-stat	coeff	t-stat	coeff	t-stat	
Dataset 1	TE	0,070	0,22	-0,018	-4,61 **	1,007	4,24 **	-0,157	-5,31 **	0,45
	GC	-0,230	-1,25	-0,004	-1,49	0,482	3,75 **	-0,024	-2,89 **	0,41
Dataset 2	TE	-0,198	-2,36 *	-0,001	-0,55	0,327	3,36 **	0,069	3,54 **	0,53
	GC	-0,068	-0,30	-0,004	-2,39 *	0,010	0,11	-0,011	-1,89	0,44
Dataset 3	TE	0,267	2,44 *	0,002	1,78	0,115	0,93	0,100	4,42 **	0,41
	GC	-0,077	-0,36	-0,005	-2,54 *	0,262	2,48 *	-0,020	-3,67 **	0,45
Dataset 4	TE	-0,003	-0,02	0,002	0,84	0,515	3,71 **	0,018	0,77	0,25
	GC	-0,148	-1,16	-0,005	-2,34 *	0,626	4,46 **	-0,026	-2,36 *	0,53

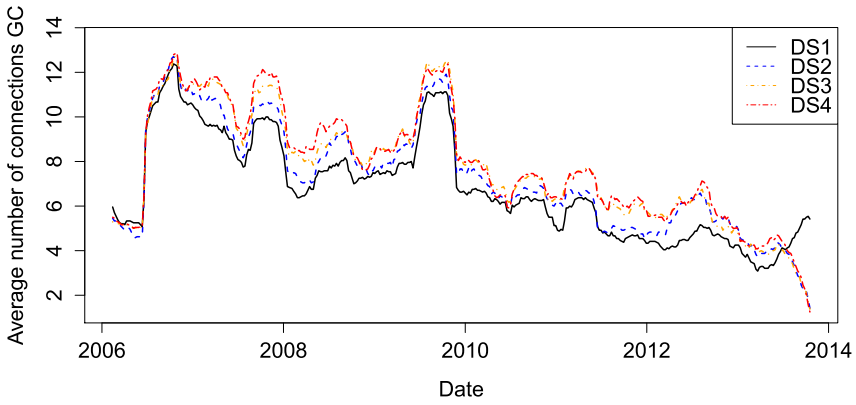
Table 1.6: This table reports the results of linear regressions based on the network of 24 SIFIs with as explanatory variables the topological measures estimated with successively TE and GC. The results are reported for the four datasets. (* and ** stand for statistical significance at the 5% and 1% levels, respectively.)

Table 1.6 overviews the results for the linear regressions using either TE or GC to infer the network. As can be seen from the significance tests, the main drivers of the average risk level are the average connectivity as well as the average eigenvector centrality. As for the coefficients, it seems that the average eigenvector centrality increases the risk of the system, while the connectivity shows less clear results with positive and negative coefficients. We also see from the R-squared values that the additional information brought by the surrounding environment, increases clearly the ability of the topological characteristics to describe properly the evolution of the risk level when using GC. The results are less clear for TE, with a clear increase when including the set of banks, but with a decrease when including all the financial institutions. This may be partly explained by the lower number of connections per node when the size of the surrounding environment increases, as presented in Fig.1.4. We see that the increasing size of the network leads for the TE based sub-network to a strict reduction of the connec-

[7] A more formal description of the different topological measures is provided in Appendix A.2.

[8] Results of autocorrelation and heteroskedasticity tests are available upon request from the authors.

tivity in contrast with the sub-network based on GC where the average connectivity remains stable. This could be explained by the different types of information detected by both measures and/or by the possible difficulties of TE to estimate properly causal relationships when the number of conditions to apply during the inference process increases. This also may confirm that TE reaches more quickly its optimal network sample size when considering relatively small windows sizes while GC is able to used the information content of the entire set of peripheral institutions. The R-squared value shows nevertheless that the information content of the SIFI network is relatively similar when using TE or GC.



(a)Granger causality

(b)Transfer entropy

Figure 1.4: Graphs a and b display the evolution in time of the average number of connections per node for both the network estimated with GC and TE.

At this point, given the equivalence of both measures in terms of information content of their respective networks, we may wonder if their information content are equivalent or if they are complementary. Starting from the filtered networks already estimated,

we propose to include in a single linear regression the four different topological characteristics given by both GC and TE based networks. For most datasets we found no sign of multicolinearity, meaning that these characteristics should provide distinct information. The results from Table 1.7 confirm the complementarity of both causality measures as we get clearly a higher percentage of explained variance when considering both measures than the two separately. We also observe that the percentage of explained variance stays stable for all datasets which may be explained by the decrease of the information content of the topological characteristics provided by the TE based networks, partly compensated by the increase for the Granger based topological characteristics. As for the coefficients, the signs are relatively equivalent to the case where we considered these measures separately.

Dataset/Measure		**Transitivity**		**Betweeness**		**EV Centrality**		**Connectivity**		R^2
Dataset 1	TE	-0,048	-0,67	-0,002	-0,93	0,253	1,55	-0,025	-3,35 **	0,59
	GC	0,143	0,65	-0,015	-3,74 **	0,725	3,24 **	-0,075	-3,27 **	0,59
Dataset 2	TE	0,039	0,34	-0,002	-1,03	0,011	0,27	-0,011	-1,49	0,57
	GC	-0,128	-1,12	-0,003	-2,59 **	0,200	2,30 *	0,037	2,78 **	0,57
Dataset 3	TE	0,037	0,34	-0,003	-2,65 **	0,026	0,13	-0,015	-1,91	0,58
	GC	0,406	1,58	0,001	0,48	0,118	0,52	0,050	1,45	0,58
Dataset 4	TE	-0,003	-0,03	-0,005	-3,56 **	0,539	4,95 **	-0,029	-2,25 *	0,56
	GC	0,139	1,46	0,000	-0,29	0,199	1,18	0,023	2,00	0,56

Table 1.7: This table reports the results of linear regressions based on the network of 24 SIFIs with as explanatory variables the topological measures estimated with both causality measures. The results are reported for the four datasets. (* and ** stand for statistical significance at the 5% and 1% levels, respectively.)

1.5 Conclusion

In this chapter, we have explored the possibility to better identify causal relations between financial series using empirical tools derived from information theory. Based on simulations of possible alternative data generating processes for returns with causal structure in mean and variance, we have studied the properties of those different causality measures based on rigorous testing beyond the Gaussian equivalence case put into lights by Barnett et al. (2009). Results show that for highly nonlinear and/or non Gaussian DGPs incorporating extreme and rare events and for causal relationships in variance, transfer entropy leads to better causality detection than standard Granger causality; the bootstrapped Kernel statistical tests which are associated to this measure, presenting better results in terms of power and size but also in terms of ROC and PR curves.

An empirical application using daily CDS of European and U.S. financial institutions has then been proposed to compare the information content of the networks estimated

with both causality measures. This comparison has been performed by studying the impact of the topological characteristics of our networks on the evolution of the average risk level of a system composed of 24 SIFIs. We additionally considered the influence of the surrounding environment of the network in the definition of its connectivity using an algorithm treating the issue of redundancies in the information transfer process. We found that the information content of the networks estimated with each causality measure explained well the evolution of this risk, considering the high percentage of variance explained by our models. We also showed that both causality measures were complementary. Regarding the effect of the peripheral institutions on the SIFIs network, we proved its positive impact on the information content of the networks estimated with Granger causality for all type of financial institutions and for banks in the case of transfer entropy. These results demonstrate that the ability of a network to benefit from the information content of its surrounding environment depends directly on the ability of the underlying measure to treat high dimensional estimation. In the case of transfer entropy such high dimensional estimation requires longer time series. Apart from that, the results do not provide a clear limit on the network sample size for the considered application. Indeed, for the Granger causality based inference process the entire set of peripheral institutions increased the information content of the inferred networks.

Chapter 2

Effective Network Inference Algorithm

2.1 Introduction

Over the past decades, complex network research has received increasing attention from the literature analyzing the developments in experimental science as well as in human and social sciences. Such developments have become necessary in order to describe systems in which interactions are governed by stochastic phenomena. Along these lines, several tools have been developed in or borrowed from areas such as statistical physics, information theory and computer sciences to explore and understand further complex systems.

Among the fields that have newly embraced the complex system perspective, the fields of economics and finance stand out markedly with contributions such as Haldane (2009); Haldane and May (2011); Battiston et al. (2012); Billio et al. (2012); Cohen-Cole et al. (2011); Matthew et al. (2014); Diebold and Yilmaz (2014); Acemoglu et al. (2015); Brummitt and Kobayashi (2015), to mention a few. In the aftermath of the financial crisis, understanding the causes of such a breakdown and in particular how the failure of a financial institution could propagate within the system and extend to the real economy became a central question in the quest of proper regulatory framework, calling for a shift from micro to macro or system-wide risk perspective. In particular, a growing interest has emerged among academics, policymakers and practitioners for having a deeper knowledge of the connectedness among financial institutions. One of the main issue in this quest however lies in the absence of comprehensive and reliable information on physical links between financial entities, such as interbank loans or syndication via common assets, and the difficulties to obtain them (Feng et al., 2014; ECB, Januray 2011). Against this background, a now common approach in the literature to infer unobserved links and reconstruct the whole financial network consists in making use of the statistical measurement (be it correlation, transfer entropy or Granger causality measures for instance) of temporal dependencies between one or

several observable characteristics associated to the nodes such as their stock market returns. In this framework, financial institutions represent the nodes of the network with the linkages reflecting the relative influences between pairs of firms.

Areas such as epidemiology (Lipsitch et al., 2003; Wallinga and Teunis, 2004), genetics (Zhang et al., 2012; Tung et al., 2007), neuroscience (Vicente et al., 2011) or human travel (Brockmann et al., 2006) follow a similar approach by considering interacting systems that can be modeled in terms of signals propagating over underlying networks (Rodriguez et al., 2014). Given the difficulty of observing the actual network structure, a growing interest exists in the literature for inference procedures based on time series representing a parameter of the underlying network such as fMRI or EEG data in neuroscience, gene expression data in genetics or asset prices in finance (Billio et al., 2012; Vicente et al., 2011; Liao et al., 2011).While convenient and intuitive, the time-series approach to network reconstruction is not immune to specific problems. One of the major problems stems from the inability to accurately separate a direct dependence between a pair of entities from indirect effects coming through the remaining part of the network. We propose to address this issue in this contribution.

In a regression setting, this problem corresponds to the omitted-variable bias. Such a bias, in a network context, could end-up to wrongly associate two nodes that are not directly connected (false positive) or fail to detect a connection that exists (false negative) because the parameters capturing their relationship are not accurately estimated. Traditionally, this is dealt with by extending a set of independent variables to control for the auxiliary effects. However, when the network is large, the solution is not trivial, as one potentially needs to add a very large set of regressors, the so-called control variables or controls, especially if the dependence stems from lagged variables. Without a specific treatment of the problem, statistical tests, whether in the absence of controls or conversely in the presence of an excessive number of controls, are likely to perform poorly.

Different strategies are proposed in the econometric literature to deal with or at least mitigate the aforementioned problem. The least absolute shrinkage and selection operator (LASSO) technique, developed two decades ago (Tibshirani, 1996), has recently received increasing attention for retrieving financial networks (Diebold and Yilmaz, 2014). The method consists in applying least squares estimation subject to some constraint. The constraint is to ease the variable selection so as to obtain a parsimonious model. A well-known alternative to LASSO in the econometric literature is the general-to-specific approach developed by David Hendry and Hans-Martin Krolzig (Hendry and Krolzig, 1999, 2001, 2003). In this case, the identification of significant variables stems from the application of a well-designed simplification algorithm. The

most basic steps of the algorithm involve first estimating the complete model in a pre-search stage and then sequentially dropping or adding variables based on a specific testing procedure.

Alternative strategies have been proposed outside the field of finance and econometrics to eliminate spurious relationships from time series data such as the so-called silencing approach (Barzel and Barabasi, 2013) used to separate the direct and indirect links in biological networks or the decomposed transfer entropy proposed by Runge et al. (2012). Other approaches based on the concept of step algorithm developed in the general-to-specific model have proven capable to infer effective networks (see, among others, Pearl (2000b); Aliferis et al. (2003); Tsamardinos et al. (2003); Tung et al. (2007); Zhang et al. (2012)). The main idea is to consider the environment around individuals transmitting and receiving the information. Conditional measures are used to identify the true path through which information travels and to remove possible redundancies in the information transfer. Two types of dependencies are considered in the literature, the instantaneous dependence estimated using correlation or mutual information and then studied using methodologies such as the directed acyclic graph (DAG) technique (Pearl, 2000b; Spirtes et al., 2000), and the dynamic dependence focusing on lagged relationships and estimated using transfer entropy or Granger causality (Runge et al., 2012; Tung et al., 2007; Lizier and Rubinov, 2013). This second approach, which we aim to further extend, provides more comprehensive information on the network, as it enables to feature the direction of edges between two nodes.

In this study, we propose to follow the literature on step algorithms using both the pre-search and pruning steps and develop a methodology that combines conditional information transfer and greedy algorithm to infer effective networks. The effective network approach Friston (1994) attempts to infer a minimal topological structure, regrouping only the directed relationships that can properly describe the evolution of the system. We further extend this approach and consider the possible spurious connections due to both instantaneous and lagged causal relationships by applying an extended set of possible conditions. As described above, the conditional information transfer approach faces the problem of dimensionality stemming from both the large number of nodes in the system and the fact that the relationship can exist for different lags. We pay particular attention in the algorithm to minimize the number of conditions for each information transfer by considering only the most relevant ones. Overall, this approach enables us to detect redundancies (i.e., lower the false positive rate) as well as to mitigate the loss of efficiency stemming from high dimensionality (i.e., lower the false negative rate).

In a series of simulation exercises using artificial networks, we confirm that our method-

ology performs well compared to a set of state-of-the-art methodologies such as the global silencing approach. Equipped with our approach, we investigate in the application part, the link between the banks' topology and systemic fragility. The past decade shows an increasing interest among the academic community and financial regulators to adopt a network perspective to describe and analyze financial markets. From the regulators' perspective, such an approach aims to identify critical relationships in the market as well as to improve regulatory policies. Following the recent studies of Billio et al. (2012), Sandoval (2014) or Diebold and Yilmaz (2014), who used asset prices to retrieve financial networks and analyze their properties, we use data on large US banks to assess the explanatory power of topological properties such as in-degree and out-degree to describe financial institution fragility. As described in section 1.4, our approach is well designed to deal with several important aspects of network reconstruction in finance.

The rest of the chapter is organized as follows. Section 2.2 introduces the concept of network and the difficulties associated to its determination. Section 2.3 presents our algorithm. In section 2.4, we demonstrate via Monte Carlo simulations the performance of our approach compared to other methodologies. Section 2.5 then turns to the empirical application and exposes the data, estimation method and results. Finally, Section 2.6 concludes.

2.2 Network structure, information decomposition, and indirect links

In a multivariate time series framework, the network is usually represented by an oriented graph made of a set of nodes, each represented by a stationary time series $X = \{x_t\}_{t=1,\cdots,T}$, and a set of oriented edges specifying the relationship between the nodes through the definition of information transfer. Each node X can therefore be characterized by a set of parents $\mathcal{P}_{\boldsymbol{k_p}}(X) \in \mathbb{R}^{P\times T}$, with P the number of parents of X, from which the node receives information, and a set of children $\mathcal{C}_{\boldsymbol{k_c}}(X) \in \mathbb{R}^{C\times T}$, with C the number of children of X, to which the node transfers information. These relationships are characterized by a specific temporal structure, according to which the value of node X at a specified time t is partly determined by the value of its predecessors at a time $t - \boldsymbol{k_p}$, where $\boldsymbol{k_p}$ depends on the lag structure of the relationship. Each term of the index vector $\boldsymbol{k_p}$ represents a specific subset of parents, $\mathcal{P}_{k_p}(X) \in \mathbb{R}^{P_{k_p}\times T}$, with the same lag k_p compared to X, with therefore P_{k_p} defining the number of parents with a lag of k_p. As Eq.2.1 shows, the information content of X can be decomposed into three components: information from the past of X, denoted by $I(x_t, x_{t-\boldsymbol{k_x}})$; information from the set of parents in previous periods $\boldsymbol{k_p}$, denoted

by $I(x_t, \mathcal{P}_{\boldsymbol{k_p}}(X) \mid x_{t-\boldsymbol{k_x}})$; and a stochastic factor $S_t(x_t)$ representing the remaining information from other sources or created by X itself (Lizier and Rubinov, 2013).

$$I(x_t) = I(x_t, x_{t-\boldsymbol{k_x}}) + I(x_t, \mathcal{P}_{\boldsymbol{k_p}}(X) \mid x_{t-\boldsymbol{k_x}}) + S_t(x_t) \tag{2.1}$$

with $\boldsymbol{k_x} = (1, \cdots, k_p^{max})$.

Our main interest in this chapter is to determine the set of parents $\mathcal{P}_{\boldsymbol{k_p}}(X)$ to be included in the information transfer $I(x_t, \mathcal{P}_{\boldsymbol{k_p}}(X) \mid x_{t-\boldsymbol{k_x}})$. To examine which nodes in the network deserve to be included in this set, we need to consider two types of links that exist between a parent node and a child node: the usual direct links where a clear causal relationship can be established from a node to another regardless of the rest of the system, and the indirect links by which the relationship between one node and another depends on at least one auxiliary node in the system. The existence of indirect links leads to redundancies in the information transfer process. When faced with indirect links, the child node seems to receive the same information from different sources. Therefore, the transfer of information cannot be expressed as a simple sum of the contribution of each source, as presented in Eq.2.2, but a pruning process is required to remove the redundant sources. The objective of the algorithm developed in this chapter is therefore to define, for each node, the minimal set of sources $\mathcal{P}_{\boldsymbol{k_p}}(X)$ to be included in $I(x_t, \mathcal{P}_{\boldsymbol{k_p}}(X) \mid x_{t-\boldsymbol{k_x}})$ to allow X to be conditionally independent from the remaining nodes $\boldsymbol{X} \setminus \mathcal{P}_{\boldsymbol{k_p}}(X)$ of the network, where $\boldsymbol{X} = \{X^n\}_{n=1,\cdots,N}$ represents all the nodes inside the considered network.

$$I(x_t, \mathcal{P}_{\boldsymbol{k}}(X) \mid x_{t-\boldsymbol{k_x}}) \neq \sum_{i=1}^{N} I(x_t, \mathcal{P}_{k_p}^i(X) \mid x_{t-\boldsymbol{k_x}}) \tag{2.2}$$

At this stage, we distinguish between two types of indirect links depending on whether the true parent of X receives information from the source of the indirect link or transmits information to the source of the indirect link (cases 2 and 3, respectively, in Fig.1). This definition of indirect links goes beyond the traditional one, which considers only case 2 (Florens and Mouchart, 1982). The elimination of these indirect links relies on conditional information transfer measures and on specific hypotheses for each type of indirect links. We illustrate the main idea behind the elimination of both types of indirect links in Fig.1 as follows. In case 1, information is transferred to node X from its parent W. In case 2, node X receives information from its parent, denoted by Z. In addition, information is transferred from node Y to Z. Thus, a spurious information transfer may occur between Y and X if we disregard the rest of the network, as we could wrongly associate to Y the fraction of information Z transmitted to X. The contribution of the transmitter Y to X will therefore disappear when conditioned on Z, but not the contribution of Z when conditioned on Y, as long as Z

transmits other personal information, which is usually the case for non-deterministic information treatment or transfer. Now, if the transmitter Y is the true parent of X, as displayed in case 3, we rely on the fact that for each transfer of information, a loss occurs between the transmission and reception of information (in this case, between Y and Z). A minimum information transfer should therefore always survive for the transmitter Y when conditioned on Z, but will disappear for Z when conditioned on the transmitter Y.

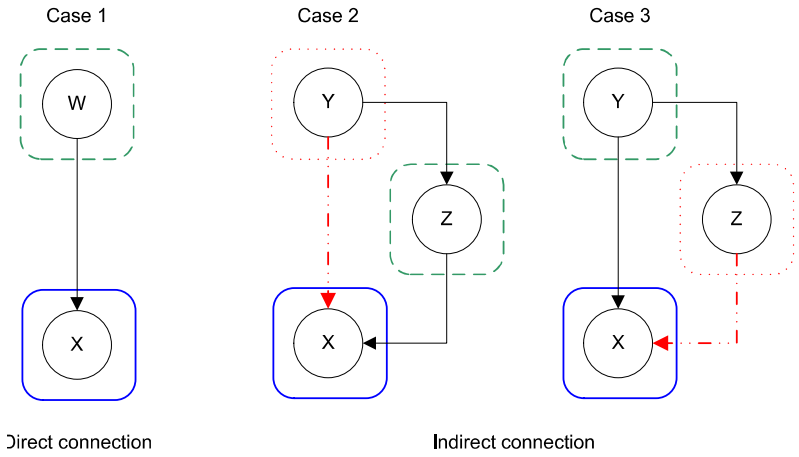


Figure 2.1: The figure displays three types of direct and indirect connections from the upper node to the lower one. Existing directed edges are drawn in bold lines; the non-existing or spuriously detected ones are drawn in dotted lines. Case 2 represents the situation where the true parent receives information from the source of the indirect link, and case 3, the situation where the true parent sends information to the source of the indirect link.

2.3 Network inference algorithm

In what follows, we develop a novel multilag approach that differentiates two levels of relationship, inter-lag and intra-lag information transfers, both of which could lead to the development of indirect links. As regards the indirect links due to inter-lag relationships, the new methodology exploits a two-way algorithm by using both the parent and children nodes of the transmitter to remove the two types of indirect relationships presented in Sec.2.2 . Let $\mathcal{P}^i_{k_p}(X) \in \mathbb{R}^T$ be a vector representing one of the parents of X with lag k_p, where $\mathcal{P}_{k_p}(X)$ is the set of parents of node X with a lag of k_p and i the parent index inside the set $\mathcal{P}_{k_p}(X)$. This means that each node $\mathcal{P}^i_{k_p}(X)$, considered at

first as parent of node X, will be removed from the set of parents of X, $\mathcal{P}_{\boldsymbol{k_p}}(X)$, if it does not transmit information to X while conditioned at the same time on its own set of parents and children connected to node X.

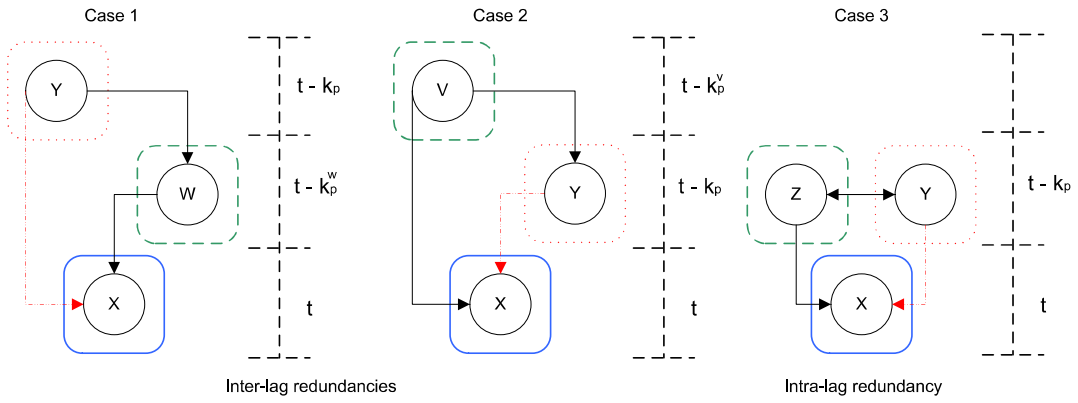


Figure 2.2: In case 1, the transmission from Y to W and from W to X lead to the detection of a spurious link between Y and X, which can be removed by conditioning the transfer from Y to X on its child W. In case 2, the transmission from V to X and from V to Y lead to the detection of a spurious link between Y and X, which can be removed by conditioning the transfer from Y to X on its parent W. As regards case 3, the transmission from Y and X and Y to Z or Z to Y lead to the detection of a spurious link between Y and X which can be remove by conditioning on Z. This last case corresponds simply to the case 1 or 2 with an instantaneous transfer of information between the true parent of X and the source of the spurious link Z which acts either as W or V.

The algorithm also controls for intra-lag redundancies. This form of redundancies stems from relationships that appear instantaneous, considering the frequency of observations used to build the network, but shares the same root causes of inter-lag redundancies (see Fig.2.2). The main objective here is not the estimation of instantaneous links between parents, but rather the elimination of redundancies in the information transfer process due to this kind of relationships. Considering the set of parents $\mathcal{P}_{k_p}(X)$ of a node X at a specific lag k, the method seeks to estimate for each parent $\mathcal{P}^i_{k_p}(X)$ in this set, the minimum information transfer to node X, considering its immediate environment in terms of lag. The resulting information transfer is estimated by conditioning each transfer between a parent $\mathcal{P}^i_{k_p}(X)$ and X on a specific subset of the other parents of X with the same lag k_p. This specific subset is defined using a greedy algorithm that incrementally adds new conditions from the set $\mathcal{P}_{k_p}(X) \setminus \mathcal{P}^i_{k_p}(X)$ to the information transfer between the considered parent $\mathcal{P}^i_{k_p}(X)$ and the node X. At each step, the condition that maximizes the reduction of information transfer between $\mathcal{P}^i_{k_p}(X)$ and X is selected. The greedy algorithm stops when no more reduction is observed in the conditional information transfer. The intuition behind this greedy approach follows the same idea as that for inter-lag relationships considering the two

types of possible indirect links, but here the connection between the true transmitter of information and the source of the indirect link appears instantaneous considering the sampling frequency. By removing the effect of indirect links due to both inter- and intra-lag relationships, we define in our algorithm a minimum information transfer for each parent. Therefore, the sum of all specific information transfers will never sum up to $I(x_t, \mathcal{P}_{\boldsymbol{k_p}}(X) \mid x_{t-\boldsymbol{k_x}})$; however, our objective is the determination of all oriented edges in our network considering redundancies.

The elimination of inter-lag or intra-lag redundancies as described above requires us to condition the information transfer on parents and children of each pair of nodes. Without any prior on the set of existing connections in the network, this would require us to condition on all nodes for each time lag. Against this background, the second objective of the algorithm is to reduce the dimension of the condition set to apply in order to limit the loss of efficiency due to small-sample effect (Runge et al., 2012). To this end, we initialize the algorithm by estimating pairwise information transfers for the entire set of nodes. Using pairwise information transfers as prior allows us to drastically reduce the number of conditions used at every further step to clean the network. Indeed, the pruning steps can use this temporary network to simultaneously define the set of parents to be considered for each node and the conditions to be applied to these parents so as to remove the possible indirect links. We will therefore not consider the entire set of nodes as other approaches do (Lizier and Rubinov, 2013; Tibshirani, 1996), but rely on the fact that the local environment (i.e., direct parents and children) of each node will regroup all the information on the evolution of the network relevant to the considered node. We focus our attention on this local environment to define the true sources of information for each node inside the network.

At this stage, considering the limited knowledge owing to pairwise estimation, we define a temporary set of parents and children that will be adjusted in the pruning process. For each node, the temporary sets of parents and children are represented as follows (Runge, 92):

$$\mathcal{P}_{\boldsymbol{k_p}}(X) = \left\{v_{t-k_p} : V \in \boldsymbol{X} \setminus X, k_p > 0, I(x_t, v_{t-k_p} \mid x_{t-\boldsymbol{k_x}}) > 0\right\}, \tag{2.3}$$

$$\mathcal{C}_{\boldsymbol{k_c}}(X) = \left\{w_t \in \boldsymbol{X} \setminus X, k_c > 0, I(w_t, x_{t-k_c} \mid w_{t-\boldsymbol{k_w}}) > 0\right\}, \tag{2.4}$$

The different steps of the algorithm are as follows. To begin with, we estimate a temporary network by removing the most distinct non-causal relationships using a simple pairwise information transfer method. We then use this information in the pruning

step to apply parsimonious conditions based on the set of parents and children of each node, represented respectively by Eqs.2.3 and 2.4. In this pruning step, we first eliminate the indirect links due to lagged relationships and then remove the ones caused by instantaneous relationships. We first eliminate the inter-lag indirect links because part of the indirect links due to intra-lag relationships are in fact imputable to inter-lag relationships. Indeed, if two nodes Y and Z have a common parent V with a lag of k and this parent sends information to a third node X with a higher lag, a spurious connection could appear between the two nodes Y/Z and the node X. This indirect link can be removed in just one step by conditioning on the common parent V, whereas two steps and a complementary condition would be necessary if we first considered the intra-lag connections (i.e., Z and V to remove the indirect link between Y and X, and Y and V for Z). Therefore, by considering first the nodes responsible for the inter-lag indirect links, the set of conditions at the end of this first substep is such that some intra-lag redundancies are already treated, leading to a reduction in number of conditions needed in the second substep, and therefore to a reduction in number of conditions to be included in the estimation of the effective transfer of information. As regards the order in which the possible transmitters are treated, we choose a simple bottom-up approach, starting with the transmitter with the lowest lag to the one with the highest lag relative to the receiver. This approach leads to similar results as for the top-down approach, but decreases the computing speed. Indeed, most of the inter-lag indirect links are parent driven; that is, the true parent transmits information to the source of the indirect link, and since this type of indirect links are removed faster with the bottom-up approach, they do not appear as conditions for parents with higher lags. The detailed steps of our iterative algorithm are as follows:

- Pre-search step: *This first step applies a filter to the considered complete network and eliminates the most distinct non-causal relationships through the detection of information transfers:*

  For a given number of lags k_{max} called horizons, estimate the N(N-1) possible connections inside the network, with N the number of nodes, using the unconditioned information transfer

  $$Y \overset{IT}{\rightarrow} X : I(x_t, y_{t-k} \mid x_{t-k_x}) > 0. \tag{2.5}$$

  For each connected pair of nodes, compare the amount of information transferred for $k = 1, ..., k_{max}$ and keep the highest information transfer[1].

[1]The choice of keeping the highest information transfer seems more natural, as it is reasonable to assume in a number of cases that the transfer mainly happens at a specific lag. However, the algorithm is flexible to alternative choices, such as always keeping the information transfer with the lowest or highest lag.

This first step returns a temporary network of $m \leq k_{max}$ layers consisting of direct and indirect links. In a few instances, the pairwise approach might fail to detect existing links. This may happen when direct and indirect information transfers strictly cancel out each other. As illustrated in the simulation part of this chapter, such cases do not significantly affect the algorithm's performance. On the other hand, using this pairwise network as a prior might help mitigate the problem of dimensionality.

- Pruning step: *This second step aims to remove the indirect links due to lagged (substep A) as well as instantaneous (substep B) causal relationships:*

  Following the bottom-up approach, remove redundant information transfers. For each node $X^i \in \boldsymbol{X} = \{X^n\}_{n=1,\cdots,N}$, begin with the parent node $Y^{i,j}_{k_p} = \mathcal{P}^j_{k_p}(X^i)\ \forall j = 1, \cdots, P^i_{k_p}$, with the lowest lag k_p compared to X^i, and apply the following substeps:

  A. Remove the inter-lag redundancies in the information coming from $Y^{i,j}_{k_p} = \mathcal{P}^j_{k_p}(X^i)$ by conditioning the information transfer on both its parents and children who are also connected to X^i.

  * Regroup both parents and children of the node $Y^{i,j}_{k_p}$ in a unique condition set

  $$\boldsymbol{Z}^{i,j}_{\boldsymbol{k_{pc}}} = \left\{ \boldsymbol{V}^{i,j}_{\boldsymbol{k^x_p}} \cup \boldsymbol{W}^{i,j}_{\boldsymbol{k^{x\prime}_p}} \right\}. \tag{2.6}$$

  With $\boldsymbol{V}^{i,j}_{\boldsymbol{k^x_p}} = \left\{ \mathcal{P}_{\boldsymbol{k^y_p}}(Y^{i,j}_{k_p}) \cap \mathcal{P}_{\boldsymbol{k^x_p}}(X^i) \right\}$, the set of parent nodes of $Y^{i,j}_{k_p}$ connected to X^i and $\boldsymbol{k^x_p}$ their lags compared to X^i.

  With $\boldsymbol{W}^{i,j}_{\boldsymbol{k^{x\prime}_p}} = \left\{ \mathcal{C}_{\boldsymbol{k^y_c}}(Y^{i,j}_{k_p}) \cap \mathcal{P}_{\boldsymbol{k^{x\prime}_p}}(X^i) \right\}$, the set of children nodes of $Y^{i,j}_{k_p}$ connected to X^i and $\boldsymbol{k^{x\prime}_p}$ their lags compared to X^i (see Fig.2.3 for a graphical representation).

  * Now, estimate the information transfer conditioned on this set. The parent node $Y^{i,j}_{k_p}$ is disconnected from X^i if the following condition is not fulfilled:

  $$I(x^i_t, y^{i,j}_{t-k_p} \mid x^i_{t-\boldsymbol{k_x}}, \boldsymbol{z}^{i,j}_{t-\boldsymbol{k_{pc}}}) > 0. \tag{2.7}$$

  The outcome of substep A is the estimated information transfer between $Y^{i,j}_{k_p}$ and X^i, lowered from the part imputable to inter-lag redundancies (see Eq.2.7).

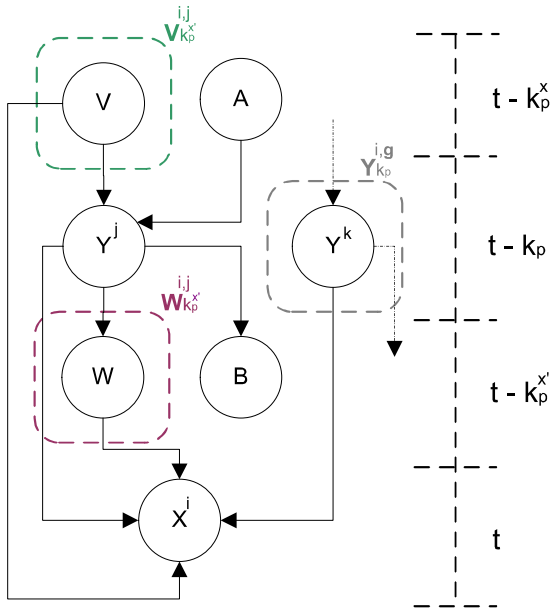


Figure 2.3: The figure provides a graphical definition of the different groups of conditions used in the pruning process: $\boldsymbol{V}^{i,j}_{\boldsymbol{k}^{\boldsymbol{x}}_{\boldsymbol{p}}}$ the group of nodes which are the parents of nodes Y^j and X^i at the same times but with lags of respectively $k^y_p = k^x_p - k_p$ and k^x_p, $\boldsymbol{W}^{i,j}_{\boldsymbol{k}^{\boldsymbol{x}}_{\boldsymbol{p}},\boldsymbol{k}^{\boldsymbol{x}'}_{\boldsymbol{p}}}$ the group of nodes which are simultaneously the parents of X^i with a lag of $k^{x'}_p$ and children of Y^j with a lag of $k^y_c = k_p - k^{x'}_p$ and finally $\mathbf{Y}^{i,\boldsymbol{g}}_{k_p}$ the group of nodes which are the parents of X^i with the same lag k_p as Y^j.

B. Greedy algorithm: If $Y^{i,j}_{k_p}$ stays, after substep A, a valuable source of information for X^i, estimate the minimal information transfer between both nodes using as conditions the other parents of X^i sharing the same lag k_p as $Y^{i,j}_{k_p}$. This set of parents is defined as $\mathbf{Y}^{i,\boldsymbol{g}}_{k_p} = \mathcal{P}_{k_p}(X^i) \backslash Y^{i,j}_{k_p}$, with $\boldsymbol{g} \in \mathbb{R}^{(P^i_k - 1)}$ the indices of the considered parents.

* For each parent $Y^{i,l}_{k_p} \in \mathbf{Y}^{i,\boldsymbol{g}}_{k_p}$, evaluate the difference between the information transfer from $Y^{i,j}_{k_p}$ to X^i conditioned on $\boldsymbol{Z}^{i,j}_{\boldsymbol{k_{pc}}}$ and the one conditioned on both $\boldsymbol{Z}^{i,j}_{\boldsymbol{k_{pc}}}$ and $Y^{i,l}_{k_p}$, as described in Eq.2.8.

$$I(x^i_t, y^{i,j}_{t-k_p} \mid x^i_{t-\boldsymbol{k_x}}, \boldsymbol{z}^{i,j}_{t-\boldsymbol{k_{pc}}}) - I(x^i_t, y^{i,j}_{t-k_p} \mid x^i_{t-\boldsymbol{k_x}}, \boldsymbol{z}^{i,j}_{t-\boldsymbol{k_{pc}}}, y^{i,l}_{t-k_p}) \quad (2.8)$$

* At each iteration, select the parent $Y^{i,l}_{k_p}$ that provides the largest incremental information transfer variation and include the considered parent $Y^{i,l}_{k_p}$ in the set of conditions $\boldsymbol{Z}^{i,j}_{\boldsymbol{k_{pc}}}$ (see Eq.2.9). Remove from $\mathbf{Y}^{i,\boldsymbol{g}}_{k_p}$ the selected par-

ent $Y^{i,l}_{k_p}$ and any other parents included in $\mathbf{Y}^{i,\boldsymbol{g}}_{k_p}$ for which no information transfer variation is observed.

$$\begin{aligned}\max(\ & I(x^i_t, y^{i,j}_{t-k_p} \mid x^i_{t-\boldsymbol{k_x}}, \boldsymbol{z}^{i,j}_{t-\boldsymbol{k_{pc}}}) \\ & -I(x^i_t, y^{i,j}_{t-k_p} \mid x^i_{t-\boldsymbol{k_x}}, \boldsymbol{z}^{i,j}_{t-\boldsymbol{k_{pc}}}, y^{i,l}_{t-k_p}),\ \ l \in \boldsymbol{g}) \end{aligned} \tag{2.9}$$

* Repeat both substeps as long as $\mathbf{Y}^{i,\boldsymbol{g}}_{k_p}$ is not empty.

If non-zero, the resulting value represents the final effective information transfer between nodes $Y^{i,j}_{k_p}$ and X^i at lag k_p, cleaned from the effects of both inter- and intra-lag redundancies. If the resulting value is below or equal to zero, the node $Y^{i,j}_{k_p}$ is removed from the set of parents of X^i, $\mathcal{P}_{\boldsymbol{k_p}}(X^i)$.

* Repeat substeps A and B for the next parent $Y^{i,j'}_{k'_p}$ of X^i with a lag k'_p higher or equivalent to $Y^{i,j}_{k_p}$.

* Reiterate these operations for each node $X^i \in \boldsymbol{X} = \{X^n\}_{n=1,\cdots,N}$ included in the considered network to eventually obtain the effective network cleaned of indirect links.

2.3.1 Extension

When facing a reduced number of observations, two modifications may be applied to the algorithm to decrease the dimension of the information transfer estimation and limit the small-sample effect. First, substep A may be divided into two successive stages by conditioning $Y^{i,j}_{k_p}$ first on the set of parents $\boldsymbol{V}^{i,j}_{\boldsymbol{k_p}}$ and then on the set of children $\boldsymbol{W}^{i,j}_{\boldsymbol{k_c}}$. The set of conditions $\boldsymbol{Z}^{i,j}_{\boldsymbol{k_{pc}}}$ regrouping both parents and children may also be set to zero at the beginning of substep B. These modifications enable us to test each type of condition separately and reduce the number of conditions applied at each step. Nevertheless, the interactions between these types of conditions cannot be identified, leading possibly to a higher false positive rate. This modified algorithm has been tested successfully with the simulated networks of Sec.2.4 and showed results relatively similar to those of the full algorithm with clearly a higher false positive rate.

An extension of the algorithm may be proposed to include links whose effect can only be detected through interactions (e.g., XOR function). A third sub-step, described in Appendix B3, may be included to deal with this type of relationships.

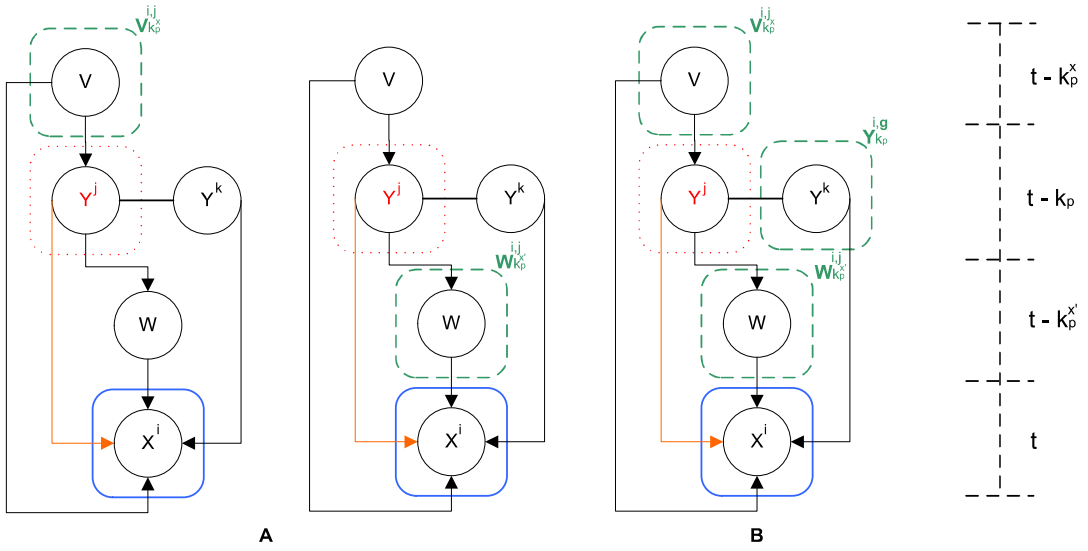


Figure 2.4: The figure regroups the cutting up of substep A into two successive stages (the conditioning of parents on the right and children in the middle) and substep B (on the left).

2.4 Simulation

To explore the performance of our approach, we carry out a Monte-Carlo simulation exercise. The simulated networks include both temporal dependent and contemporaneous relationships (i.e., resp. inter- and intra-lag) in order to assess the behavior of the two main stages of our algorithm (substeps A and B). The topology of the simulated networks is generated randomly and allows cycles, i.e. feedback loops[2]. Scale-free network being the usual representation in finance as well as in other fields of science (Barabasi and Albert, 1999), the inter-lag degree distribution of the generated nodes follows a power law with $\gamma = 2.5$ and up to 6 incoming edges per nodes. The simulator assigns to each edge a lag of 1 with a probability of 70 percent and a lag of 2 with the remaining probability of 30 percent. Regarding the creation of the intra-lag links, two additional rules are applied: the receiver and the transmitter must not be already connected, and either the transmitter or the receiver must have at least one child. The objective of both rules is to increase the number of potential intra-lag indirect links and challenge the inference algorithms ability to prune this type of indirect links. For each pair of node respecting the previous rules, an instantaneous link is created 20 percent of the time. The network generator starts with the creation of inter-lag connections, going from node to node, and then adds intra-lag connections in a second

[2]The number of cycles have been estimated for the different sets of networks with an average number of 8.2 and 22.5 cycles per network for the networks composed of respectively 20 and 40 nodes.

step. All connections are stored in an adjacency matrix with their lag as parameter.

The time series of each node is then estimated based on the connections stored in the adjacency matrix using two different frameworks, a vector auto-regression (VAR) structure to account for linear relationships and a regime switching specification to account for non-linear behavior. Taking the case presented in Fig.2.4, Eq.2.10 provides the evolution of the time series Y^i following the VAR specification. The coupling strengths are set to $\beta \in [0.4, 0.6]$ for inter-lag relationships and $\delta \in [0.5, 0.6]$ for intra-lag relationships. We choose relatively high values for these coefficients because the main idea behind these simulations is to assess the ability of our algorithm to disentangle the direct links from the indirect ones, and not the ability of the selected causality measures to detect information transfers.

$$y_t^i = \alpha y_{t-1}^i + \beta w_{t-k_p^{x'}} + \delta y_{t-k_p}^k + \varepsilon_t \tag{2.10}$$

where the parameters β and δ regulate the coupling strength for respectively the inter-lag and intra-lag relationships, and ε_t represents an independent Gaussian white noise with zero mean and a variance of one.

As regards the regime switching specification, Eq.2.11 gives the general framework used to generate the time series. Each node connected to Y^i transmits its information intermittently depending on the state of the world s_t. The transition between the different states depends on a Poisson process with an intensity of 10 percent. The coupling strengths β and δ are defined in the same way as for the VAR specification.

$$y_t^i = \alpha y_{t-1}^i + \beta w_{t-k_p^{x'}} + \delta y_{t-k_p}^k + \varepsilon_t \quad When\ s_t = 0 \tag{2.11}$$
$$y_t^i = \alpha y_{t-1}^i + \delta y_{t-k_p}^k + \varepsilon_t \quad When\ s_t = 1 \tag{2.12}$$
$$y_t^i = \alpha y_{t-1}^i + \beta w_{t-k_p^{x'}} + \varepsilon_t \quad When\ s_t = 2 \tag{2.13}$$

We compare the results of our technique using both transfer entropy and Granger causality (see Appendix B1) as information transfer measure, denoted respectively by TE and GC, and four other state-of-the-art methodologies, the Greedy methodology (Lizier and Rubinov, 2013), denoted by Greedy; the Minimum Description Length methodology (Zhao et al., 2006) denoted by MDL; the Global Silencing methodology (Barzel and Barabasi, 2013; Feizi et al., 2013); and the Granger LASSO method (Tibshirani, 1996).

The Greedy algorithm uses transfer entropy as the information transfer measure,

whereas the MDL and Global Silencing methodologies use a causal form of mutual information (for more details, see Zhao et al. (2006)). The Global Silencing methodology and MDL use the first step of our algorithm as initialization. The resulting multi-lag matrix is merged for the Global Silencing methodology to form a two-dimensional matrix. As regards the Granger LASSO method, for each variable, all the dataset nodes are included in the model for the selected time horizon of two lags. The main parameter influencing the performance of our algorithm and the Greedy methodology is the p-value used as threshold to remove the non-causal links. The MDL relies on a parameter which accounts for the influence of the network coding length on its cost function while the two last methodologies relies on an ad hoc threshold determining the level at which a link should be considered as indirect.

As commonly done in the literature, we assess the performance of our algorithm and other methodologies by comparing the inferred effective network with the known underlying network created by the simulator. We report the performance of the considered approaches in terms of area under the receiver operating characteristic curve (AUROC) and area under the precision-recall curve (AUPR), as commonly done in the literature (Marbach et al., 2012). Two different ROC curves are proposed to consider both inter and intra-lag indirect links. The first one uses the traditional total number of false positives while the second ROC curve, called instantaneous ROC, relies on the number of false positive imputable to an instantaneous links. We start our simulation exercise by generating 100 simulated networks composed of 20 nodes and represented by time series of 200 observations. We apply on them the five different methodologies.

Fig.2.5(a) shows the performance of the different methodologies using simulated networks including 20 nodes. From the the AUROC graph, we see that our algorithm using both information transfer measures, TE and GC, does well compared to the other methodologies, with an AUROC of respectively 0.96 and 0.93 [3]. In terms of instantaneous ROC curves, the difference is still significant for the GC based model but less pronounced for the TE based model which is closer to the other methodologies. Despite the lower values compared to the previous ROC curves, we see that our algorithm is still able to remove effectively indirect links produced by instantaneous relationships. The Greedy algorithm, which is specifically designed to treat intra-lag indirect links, do not provide better results and seems to suffer more than our algorithm from the curse of dimensionality with an AUROC of 0.87. With regards the precision-recall curves, our algorithm using GC as underlying information transmission measure, pro-

[3]The difference between the results obtained with both information transmission measures should shrink if we consider longer time series as transfer entropy suffers more from the small sample effect. Indeed it has been proven that transfer entropy and Granger causality where equivalent in the Gaussian case (Barnett et al., 2009).

vides good results with an AUPR of 0.68, while the other curves are relatively close although the ones associated with TE and LASSO are a bit higher than the others with an AUPR of respectively 0.60 and 0.57. The other four state of the art methodologies are in general relatively close to each others in terms of AUPR, AUROC and AUIROC.

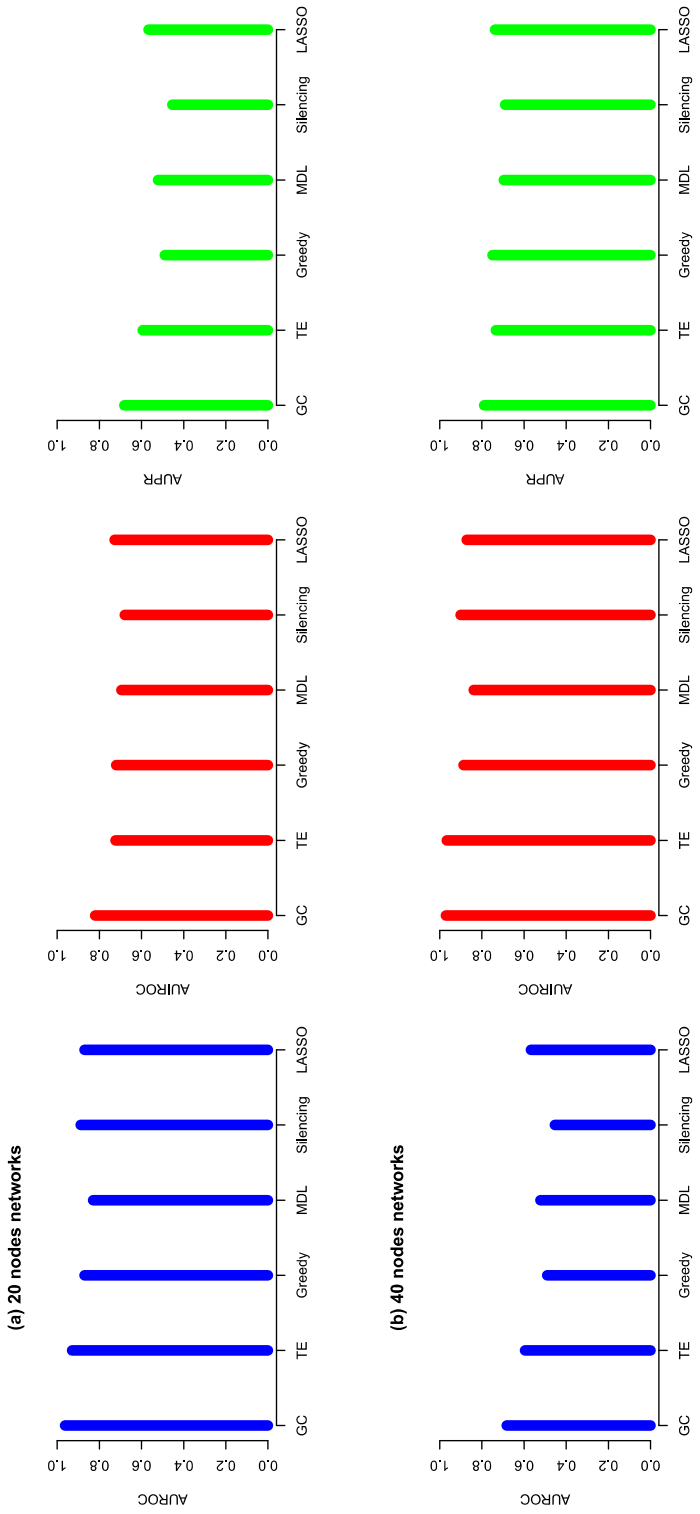


Figure 2.5: The figures (a) and (b) display the AUROC, AUIROC and AUPR values for a set of 100 simulated networks composed of repsectively 20 and 40 nodes with a VAR specification for the information transmission.

The first simulated networks included only 20 nodes to challenge the five considered approaches, but how well do they perform when considering a larger network? To answer this question, we suggest to run a second set of simulated networks composed of 40 nodes and compare the results with the ones obtained in the previous run. Looking at the result from Fig.2.5(b), the proposed algorithm performs better, with an AUROC of around 0.97 for both information transfer measures. The ability of our algorithm to infer the underlying network seems even to increase slightly with the number of nodes, especially for the TE based model. The AUPR values are relatively similar to the ones observed previously. The good results associated with our algorithm, independently from the number of nodes, arise from its ability to reduce the high dimensionality of the network inference puzzle via the pre-search step. The inference procedure can then be applied on smaller parts of the network, conserving a low dimensionality whenever the size of the complete network.

We eventually consider the impact of non linear information transmissions on the ability of the fives methodologies to infer properly the underlying networks. We simulate 100 networks composed of 20 nodes and use the regime switching specification to generate the time series. The results provided in Fig.2.6 show the negative effect of non linearity on all methodologies with lower AUROC, AUIROC and AUPR. Nevertheless, the reduction is more pronounced for the other methodologies. Indeed, our algorithm demonstrates its ability to cope with non linear information transmissions, especially when using transfer entropy, the result being, in this case, very similar to the one obtained with the VAR specification. This results from the higher sensitivity of transfer entropy to non linear patterns. Considering the three simulation sets, the runner-up methodology seem to be mainly the LASSO approach.

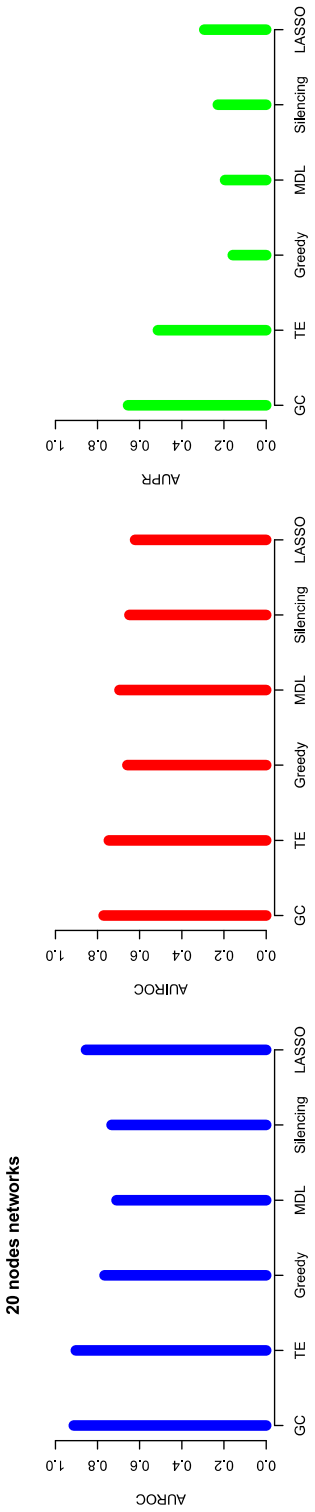


Figure 2.6: The figure displays the AUROC, AUIROC and AUPR values for a set of 100 simulated networks composed of 20 nodes with a regime switching specification for the information transmission.

We have thus shown through these simulations that, regardless of the underlying causality measure selected, the proposed algorithm generally performs better than the other methodologies we considered when examining both intra- and inter-lag relationships. Moreover, in spite of good results in some instances, the other approaches[4] necessitate good calibration of their parameters, making it difficult to use in practical applications on real data without other information on the underlying network.

2.5 Application to the US financial sector

In this section, we propose an empirical application of our network inference algorithm to shed new light on important issues of systemic risk assessment. As documented in the literature, several challenges have to be addressed to obtain an accurate representation of financial networks and in turn characterize their topology. Among those challenges stands out the absence of comprehensive and reliable information on physical connections, such as interbank loans or syndication via common assets (Feng et al., 2014; ECB, Januray 2011). As an alternative, a strand of the literature has recently proposed to rely on dependencies between the asset prices of financial institutions as a measure of connections (Billio et al., 2012; Diebold and Yilmaz, 2014; Barigozzi and Brownlees, 2013). This approach is convenient for different reasons among which our ability to remain agnostic on the specific channel through which two institutions are connected. Another advantage of such an approach lies in that asset prices are easily available data for a large set of world-wide institutions. A pioneering contribution in this line of research has been proposed by Billio et al. (2012), who estimate pairwise Granger causality on a large set of data. While both convenient and intuitive, this approach is not immune to specific problems. First, it has to account for time variation of the networks, as links can appear and disappear over time depending on the banks' contracts with other banks. Second, there is no consensus on the data frequency at which this estimation should be made and whether this choice impacts the retrieved network. Third, the pairwise approach can lead to a misrepresentation of connectivity as it does not allow to separate direct links from the indirect ones. Fourth, when adopting the Granger causality approach, one needs to specify the lag structure of the time series dependence, as the connections might be invisible with a single lag but appear as we go down the past (Barigozzi and Brownlees, 2013).

In what follows, we apply our methodology and analyze the evolution of the financial network characteristics before, during, and after the recent financial crisis and its impact on the fragility of financial institutions. Our algorithm seems particularly well

[4]Apart from the Greedy methodology which also relies on the definition of a p-value.

suited to tackle this issue. As described in the previous sections, it is designed to identify direct links efficiently while accounting for multilag relationships. Following the literature, it can also easily accommodate dynamicity in the networks by applying a rolling window approach (Dungey et al., 2012; Billio et al., 2012). Eventually, we use data at different frequencies, including high-frequency data, to analyze how frequency sensitive the network topology and its impact on bank fragility can be.

For our multi-frequency analysis, we use the stock returns at three different frequencies, daily observations, hourly observations, and high-frequency observations, that is, every 15 minutes. We then rely on measures in the vein of Diebold and Yilmaz (2014), such as the average in- and out-degree or the eigenvector of centrality, to determine the relationship between the topological evolution of the network and the risk level of its components. We therefore withdraw from the alternative question of which financial institutions contribute the most to the risk of the system (see Brunnermeier et al. (2009); Adrian and Brunnermeier (2009); Acharya (2009); Anand et al. (2012); Battiston et al. (2012) among others) and treat the reverse question of how the system and, more specifically, its topology could impact the fragility of a specific institution. Different approaches have been proposed to analyze bank fragility, such as the SRISK measure of Acharya et al. (2012) [5], DebtRank (see Battiston et al. (2012); Thurner and Poledna (2013); Bardoscia et al. (2015) to quote only a few), or closer to what we do Ballaa et al. (2014); Billio et al. (2012), which relates it to the financial institutions' topological characteristics. However, to our knowledge, this is the first study to propose the use of both low-frequency and high-frequency data for this type of analysis. Moreover, the aforementioned studies are not designed to properly deal with the redundancy issues that could strongly affect the final results.

2.5.1 Topological measures estimation

Following Ballaa et al. (2014), we focus on the largest US banks. The selected dataset regroups the stock returns of 22 financial institutions traded on the NYSE and Nasdaq, covering the period from January 2005 to January 2013. As a preliminary step in network estimation, we filter out our dataset from the effect of common global factors using a market index[6]. The filter is a simple linear regression done with stock returns as the dependent variable and the S&P 500 as explanatory variable. The residual of the linear regression is then used to estimate the networks. In order to explore the relationship between financial institutions both in mean and variance, we follow the

[5]The SRISK measure uses a conditional form of capital shortfall, with market risk level as the condition.

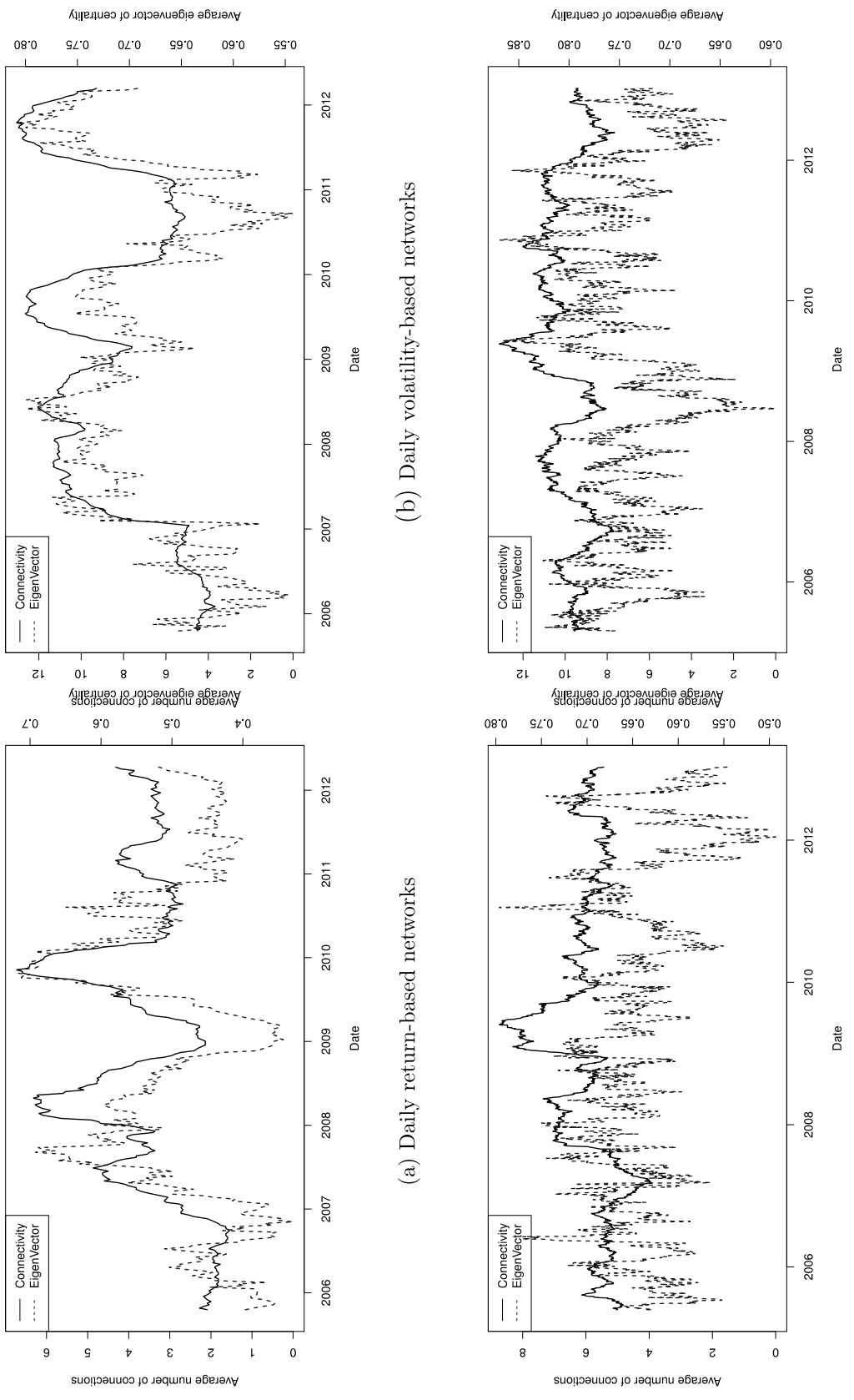
[6]Billio et al. (2012) use this approach for their robustness analysis and do not find qualitative differences in their results.

recent literature on non-parametric volatility measures (Barndorff-Nielsen and Shephard, 2003; Andersen et al., 2005; Mykland and Zhang, 2009; Andersen et al., 2010) by using high frequency data (i.e., 15 minutes) to reconstruct lower-frequency volatility (i.e., daily), which forms a fourth sample. Specifically, we use an approach similar to the one developed in Barndorff-Nielsen and Shephard (2003) to consider the possible outliers in our dataset by computing a jump-robust measure of volatility, namely, the Bipower variation.

We begin our analysis by applying our network inference algorithm on the data using conditional Granger causality as transfer information measure. As shown in the simulation exercise, this information transfer measure performs well with small samples. Therefore, it allows us to reduce the number of observations needed to reconstruct the network. For each frequency, we consider window sizes of 240 observations to estimate the information transfer, which corresponds to one year for daily frequency, since we have 20 trading days per month on average. Since the number of observation is fixed for all frequencies, we have a common level of bias from the small-sample effect. The network dynamics is then characterized using traditional rolling windows[7].

A well-known issue of rolling windows is their sensitiveness to outliers, which can lead to highly volatile indicators of interconnectedness. To deal with this issue and smooth the indicators, we apply a second set of rolling windows on the estimated networks with a window size of six months and stepsize of one period for all frequencies. For each pair of nodes, we compute the number of times a transfer of information is observed in the set of networks included in each six-month window. Each pair of nodes is then considered as connected if the number of connections observed in each six-month window exceeds a threshold of 10 percent. We eventually estimate three topological measures using these filtered network representations: the average number of outgoing and incoming connections per node and the eigenvector of centrality.

[7]For daily frequency, the step size is five observations, corresponding to a week; for the hourly (resp. 15 minutes) frequency, the step size is six observations (resp. 24 observations), corresponding to a trading day.



(a) Daily return-based networks

(b) Daily volatility-based networks

(c) Hourly return-based networks

(d) High-frequency return-based networks

Figure 2.7: The average connectivity and eigenvector of centrality for different data sets covering the period from 2006 to 2012. Graphs a and b, which display the evolution of the characteristics of the networks inferred using low frequencies, show clearer and more pronounced patterns with stiff variation during the financial crisis.

Fig.2.7 displays, for four different samples, the evolution over time of the average connectivity[8] and eigenvector of centrality. The averages computed are based on the 22 financial institutions of the dataset. As the graphs show, the networks inferred from the daily frequencies show more pronounced variation and clearer patterns, especially during the crisis, suggesting a lower level of noise and potentially more information content. We also observe an increase in connections detected with higher frequency data in the crisis period, and, more specifically, in the last part of 2008 and 2009. This peak, which is not visible for daily frequency-based networks, might suggest that the transfer of information accelerates in crisis periods, thereby being better detected with high-frequency data.

2.5.2 Financial institutions' risk and topological measures

Equipped with the retrieved networks and using different measures, we propose to follow Ballaa et al. (2014) and determine how topological characteristics contribute to the risk of financial institutions. In line with Ballaa et al. (2014); Billio et al. (2012), we define the exposure of an institution as the maximum percentage financial loss during a specific period of time; also called Max%Loss. We use the same window size as used previously to compute the Max%Loss series and carry out linear regressions using the Max%Loss series as dependent variable, and the number of incoming edges, the number of outgoing edges, and the eigenvector of centrality as explanatory variables. A preliminary analysis is then performed for each institution; this consists of the usual multi-colinearity, auto-correlation of errors, and heteroscedasticity tests[9]. Since signs of auto-correlation of error terms have been detected for most institutions and frequencies, we estimate the heteroskedasticity and autocorrelation consistent (HAC) covariance matrix using the procedure developed by Newey and West (1987) to correctly compute the covariance matrix of the explanatory variables[10]. Finally, we estimate the percentage of the explained variance attributable to the different predictors as the squared semi-partial correlation.

As evoked earlier, our empirical procedure should shed light on two important questions: (i) Do topological measures such as in-degree and out-degree as well as the eigenvector of centrality provide a good indication of financial institution fragility? (ii) Is the information content of the topological measures constructed from low-frequency data the same as that constructed from high-frequency data? The regression results

[8] As we computed the average values for 22 banks, the in-degree and out-degree give similar results.

[9] The related tests for multi-colinearity, auto-correlation of errors, and heteroscedasticity are respectively the VIF test, the Durbin Watson test, and the Breusch-Pagan test, using the R package CAR of Fox and Weisberg (2011) and LMTEST of Zeileis and Hothorn (2002).

[10] We use the R package Sandwish of Zeileis (2004) to estimate the HAC covariance matrix.

Ticker	Eigenvector			In-degree			Out-degree		
	coeff	t-stat	R^2	coeff	t-stat	R^2	coeff	t-stat	R^2
SNV	-0,352	-1,942	0,121	0,078	4,071 **	0,032	0,042	1,614	0,193
STI	-0,881	-3,599 **	0,140	0,064	4,942 **	0,178	0,075	2,213 *	0,136
ZION	-0,578	-2,835 **	0,278	0,047	1,684	0,138	0,045	3,870 **	0,074
BBT	-0,461	-1,723	0,014	0,030	1,197	0,083	0,021	0,685	0,048
HCBK	-0,457	-3,010 **	0,127	0,044	5,575 **	0,072	0,036	2,585 *	0,208
NYCB	-0,155	-0,985	0,133	-0,037	-2,262 *	0,018	0,022	2,680 **	0,096
RF	-0,677	-4,334 **	0,002	0,122	4,380 **	0,109	0,008	0,670	0,426
JPM	-0,657	-4,214 **	0,192	0,022	1,740	0,322	0,040	3,287 **	0,041
MS	-0,268	-2,062 *	0,230	0,009	0,639	0,037	0,042	7,011 **	0,006
STT	-0,191	-2,384 *	0,008	0,076	5,201 **	0,018	0,011	2,796 **	0,219
WFC	-0,601	-3,610 **	0,089	0,029	1,811	0,122	0,034	6,708 **	0,035
BAC	-0,998	-7,202 **	0,014	0,115	4,022 **	0,192	0,022	1,253	0,229
BK	-0,545	-2,487 *	0,179	0,032	3,426 **	0,233	0,042	3,624 **	0,153
C	-1,535	-7,908 **	0,175	0,067	4,435 **	0,378	0,073	5,595 **	0,156
GS	-0,034	-0,225	0,089	0,019	0,830	0,001	0,021	23,833 **	0,019
FITB	-1,114	-5,118 **	0,466	0,054	3,909 **	0,408	0,086	7,567 **	0,064
HBAN	-0,616	-2,278 *	0,113	0,055	1,635	0,065	0,070	2,679 **	0,132
KEY	-0,462	-6,036 **	0,397	0,065	4,052 **	0,106	0,053	7,800 **	0,270
MTB	-0,527	-9,752 **	0,131	0,051	2,132 *	0,142	0,042	5,302 **	0,122
PBCT	-0,766	-5,068 **	0,194	0,073	7,101 **	0,199	0,069	9,637 **	0,419
PNC	-0,586	-12,041 **	0,108	0,099	10,462 **	0,052	0,085	14,665 **	0,363
CBSH	-0,318	-4,291 **	0,555	0,020	1,953	0,108	0,048	6,579 **	0,048

Table 2.1: This table reports the results of linear regression using the daily returns of 22 US financial institutions selected for our application as base for network estimation.(* and ** stand for statistical significance at the 5% and 1% levels, respectively.)

are reported in Tables 1-4 for the three different frequencies. For the daily frequency, we use both returns and volatility. The tables report the estimated coefficient, the t-statistic, the p-value, and the percentage of the explained variance attached to each variable. We find that the three topological characteristics are significant determinants of the Max%Loss variable for most financial institutions. The number of significant variables is nevertheless lower for the highest frequency, suggesting the presence of noise in the data. Moreover, the volatility data seem to better capture the dynamics of the networks with, on average, a higher percentage of explained variance for all variables. The signs of the coefficients are similar for almost all frequencies and all financial institutions. The coefficients for in-degree and out-degree are clearly positive, indicating that more connected institutions in and out are more prone to losses. The positive sign of the incoming edges seems to support the view that institutions with an increasing number of in-coming links are more subject to shocks from the system and therefore are more fragile. As for the number of outgoing edges, its positive impact on risk may be explained by feedback mechanisms, according to which a stress in highly connected institutions is more likely to be transmitted to the rest of the system. As a result, the worsening of the financial environment deteriorates further the financial condition of institutions under stress. The coefficients of the eigenvector of centrality

Ticker	**Eigenvector**			**In-degree**			**Out-degree**		
	coeff	t-stat	R^2	coeff	t-stat	R^2	coeff	t-stat	R^2
SNV	-0,504	-2,431 *	0,316	0,026	3,015 **	0,158	0,040	4,450 **	0,158
STI	-1,039	-5,024 **	0,121	0,033	11,558 **	0,257	0,229	2,059 *	0,257
ZION	-0,267	-2,063 *	0,021	0,067	8,828 **	0,624	0,014	-1,494	0,624
BBT	-0,611	-5,515 **	0,060	0,030	2,580 *	0,277	0,059	3,988 **	0,277
HCBK	0,370	1,199	0,000	0,005	0,927	0,011	0,042	0,012	0,011
NYCB	-0,644	-16,859 **	0,057	0,023	7,765 **	0,227	0,433	2,932 **	0,227
RF	-0,291	-2,331 *	0,117	0,041	3,397 **	0,462	0,016	5,085 **	0,462
JPM	0,094	2,121 *	0,057	0,016	9,046 **	0,082	0,006	4,545 **	0,082
MS	-0,737	-10,305 **	0,030	0,042	6,987 **	0,429	0,128	2,640 **	0,429
STT	-0,358	-4,863 **	0,372	0,022	6,202 **	0,127	0,077	11,062 **	0,127
WFC	-0,165	-1,176	0,053	0,001	0,353	0,000	0,016	2,782 **	0,000
BAC	-0,997	-18,541 **	0,327	0,039	4,487 **	0,202	0,161	42,988 **	0,202
BK	-0,487	-3,942 **	0,460	0,001	0,340	0,001	0,254	5,512 **	0,001
C	0,177	0,848	0,001	0,062	5,559 **	0,283	0,002	-0,356	0,283
GS	-0,387	-6,830 **	0,098	0,018	5,689 **	0,085	0,050	15,459 **	0,085
FITB	-0,372	-1,460	0,001	0,071	15,906 **	0,668	0,022	0,242	0,668
HBAN	-0,301	-5,047 **	0,028	0,038	8,578 **	0,249	0,012	1,979 *	0,249
KEY	-0,586	-1,661	0,327	0,004	0,301	0,002	0,057	3,294 **	0,002
MTB	0,068	1,431	0,016	0,026	7,238 **	0,337	0,001	-5,720 **	0,337
PBCT	-0,517	-4,778 **	0,080	0,015	3,013 **	0,141	0,131	6,748 **	0,141
PNC	-1,258	-3,043 **	0,218	0,023	2,011 *	0,130	0,498	3,839 **	0,130
CBSH	-0,384	-10,736 **	0,754	0,000	-0,033	0,000	0,160	6,899 **	0,000

Table 2.2: This table reports the results of linear regression using the daily volatilities of 22 US financial institutions selected for our application as base for network estimation. (* and ** stand for statistical significance at the 5% and 1% levels, respectively.)

are usually negative. This measure of centrality is slightly more complex than the previous ones as it combines information on both the number of in- and out-degree links and the importance of the connected nodes. With this in mind, our result indicates that the more the banks are connected to important institutions, once we control for the number of in- and out-degree links, the lower their level of risk. As for the percentage of explained variance, it differs significantly from one institution to another, but tends to prove the importance of the number of outgoing and incoming edges.

Finally, we discuss how sensitive these results are to the choice of data frequency. Several insights can be drawn from our results. First, topological indicators matter for all the considered frequencies. Second, significant differences exist between them. Thus, topological indicators explain financial fragility in more instances with high-frequency (1 hour) information than with low-frequency information (1 day). However, this result reverses for very high (15 min) frequencies. This finding is corroborated by the explanatory power of the different regressions, which is on average 37 percent (R-square) for daily observations, compared to 38 percent for hourly and 25 percent for 15-minute observations. Thus, we can assume a trade-off between information and noise when increasing the frequency of data, leading to an optimal frequency to

Ticker	Eigenvector			In-degree			Out-degree		
	coeff	t-stat	R^2	coeff	t-stat	R^2	coeff	t-stat	R^2
SNV	-0,644	-3,557 **	0,179	0,044	4,932 **	0,150	0,044	7,741 **	0,210
STI	-0,611	-5,554 **	0,118	0,052	4,685 **	0,196	0,029	7,838 **	0,308
ZION	-0,627	-2,413 *	0,037	0,046	5,597 **	0,111	0,025	2,631 **	0,195
BBT	-0,333	-4,796 **	0,323	0,033	2,949 **	0,066	0,035	8,749 **	0,188
HCBK	-0,186	-2,195 *	0,139	0,020	3,211 **	0,051	0,013	5,143 **	0,127
NYCB	-0,375	-2,993 **	0,037	0,022	3,047 **	0,165	0,012	2,535 *	0,107
RF	-0,529	-6,234 **	0,166	0,062	5,745 **	0,094	0,043	9,701 **	0,426
JPM	-0,537	-3,169 **	0,177	0,016	2,962 **	0,282	0,024	1,856	0,043
MS	-0,646	-3,538 **	0,517	0,028	2,275 *	0,312	0,043	6,021 **	0,122
STT	-0,437	-10,128 **	0,344	0,020	6,874 **	0,117	0,035	9,255 **	0,063
WFC	-0,352	-1,925	0,055	0,028	5,649 **	0,036	0,027	2,502 *	0,109
BAC	-0,504	-1,574	0,163	0,042	2,388 *	0,081	0,031	3,831 **	0,177
BK	-0,262	-5,969 **	0,334	0,022	4,975 **	0,111	0,020	12,217 **	0,142
C	-0,836	-3,576 **	0,292	0,044	2,262 *	0,144	0,053	4,779 **	0,088
GS	-0,529	-2,690 **	0,175	0,040	2,814 **	0,285	0,026	2,301 *	0,396
FITB	-0,608	-5,211 **	0,241	0,053	3,024 **	0,127	0,041	2,786 **	0,167
HBAN	-0,573	-4,582 **	0,109	0,052	6,336 **	0,107	0,037	6,088 **	0,239
KEY	-0,480	-3,077 **	0,312	0,039	8,649 **	0,099	0,041	4,392 **	0,167
MTB	-0,317	-4,147 **	0,025	0,027	6,991 **	0,065	0,010	3,237 **	0,201
PBCT	-0,449	-2,545 *	0,292	0,023	2,529 *	0,326	0,028	2,588 **	0,134
PNC	-0,537	-3,798 **	0,179	0,036	4,992 **	0,157	0,043	3,779 **	0,234
CBSH	-0,276	-6,730 **	0,328	0,018	4,736 **	0,221	0,016	10,602 **	0,186

Table 2.3: This table reports the results of linear regression using the hourly returns of 22 US financial institutions selected for our application as base for network estimation. (* and ** stand for statistical significance at the 5% and 1% levels, respectively.)

retrieve financial networks from asset prices for systemic risk assessment.

2.5.3 Empirical comparison with pairwise Granger causality

As already mentioned, pairwise estimations are relatively common in finance to infer networks. Therefore, we propose, as a final exercise to assess the added value of our algorithm, to compare the information content of the network inferred in the previous section with the information content we would obtain by using a simple pairwise approach. Indeed, the more precise representation of the network in our algorithm could increase the information content of the different variables used in the previous section to describe the evolution of financial institution risk.

In order to simplify our analysis, we focus on daily returns and volatility. The network estimations for the pairwise approach use the pre-search step of our algorithm with the same parameters. We follow the approach described in the previous section, again with two levels of rolling windows for topological parameter estimation and linear regressions based on these variables. As evident from Tables 5 and 6, for both data sets, there is a clear decrease in the number of significant variables when using the pairwise approach compared to the results obtained with our algorithm (see Tables

Ticker	Eigenvector			In-degree			Out-degree		
	coeff	t-stat	R^2	coeff	t-stat	R^2	coeff	t-stat	R^2
SNV	-0,491	-3,973 **	0,241	0,014	1,849	0,106	0,023	7,858 **	0,052
STI	-0,444	-1,724	0,237	0,013	4,728 **	0,110	0,021	3,171 **	0,045
ZION	-0,431	-4,398 **	0,123	0,018	8,372 **	0,059	0,024	3,197 **	0,109
BBT	-0,370	-3,266 **	0,172	0,017	2,789 **	0,118	0,016	2,464 *	0,160
HCBK	-0,005	-0,123	0,003	0,006	1,555	0,000	0,001	0,512	0,084
NYCB	-0,292	-2,388 *	0,193	0,006	1,412	0,187	0,010	2,978 **	0,030
RF	-0,707	-3,545 **	0,287	0,021	2,704 **	0,200	0,032	3,502 **	0,110
JPM	-0,368	-2,711 **	0,052	0,013	2,345 *	0,076	0,011	1,728	0,138
MS	-0,353	-1,813	0,062	0,022	2,048 *	0,057	0,015	2,485 *	0,176
STT	-0,444	-3,343 **	0,062	0,023	2,539 *	0,119	0,013	2,844 **	0,240
WFC	-0,608	-2,924 **	0,310	0,002	0,413	0,256	0,027	3,104 **	0,001
BAC	-0,477	-2,160 *	0,102	0,026	2,023 *	0,102	0,016	2,558 *	0,166
BK	-0,337	-2,431 *	0,175	0,010	1,694	0,141	0,014	2,937 **	0,069
C	-0,409	-1,993 *	0,073	0,032	3,687 **	0,061	0,016	1,902	0,196
GS	-0,282	-1,960	0,188	0,009	2,412 *	0,080	0,015	2,711 **	0,048
FITB	-0,392	-4,561 **	0,129	0,024	3,368 **	0,054	0,021	8,592 **	0,153
HBAN	-0,720	-2,615 **	0,144	0,021	2,951 **	0,155	0,029	2,693 **	0,139
KEY	-0,155	-1,050	0,022	0,023	3,966 **	0,009	0,008	1,525	0,216
MTB	-0,337	-2,287 *	0,108	0,019	2,897 **	0,114	0,013	2,712 **	0,202
PBCT	-0,295	-2,934 **	0,143	0,011	1,768	0,184	0,012	3,018 **	0,118
PNC	-0,299	-5,452 **	0,115	0,015	2,634 **	0,076	0,014	2,836 **	0,123
CBSH	-0,236	-2,475 *	0,046	0,010	2,016 *	0,178	0,004	1,911	0,194

Table 2.4: This table reports the linear regression results using the high-frequency returns of 22 US financial institutions selected for our application as base for network estimation. (* and ** stand for statistical significance at the 5% and 1% levels, respectively.)

1 and 2). Apart from the in-degree for the volatility data set, a similar trend can be seen for the percentage of explained variance with, on average, a decrease ranging between 28 and 46 percent for the pairwise approach. As for the coefficients, we obtain, overall, the same signs as in Tables 1 and 2 with a negative impact of the eigenvector of centrality and a positive impact on the firm's risk for the in- and out-degree. We can therefore conclude that our algorithm tends to provide more precise information on interactions within the network.

2.6 Conclusion

In this second chapter, we explored the possibility of improving network inference from temporal data acting on two different levels. We proposed a novel approach taking into account the effects of both redundancies in information transfer processes and increasing dimensionality on information transfer estimation. We tried to reduce the impact of redundancies caused by both inter- and intra-lag relationships and to minimize the number of conditions to apply to each information transfer to discard these redundancies. The proposed approach may be divided into two main steps: the elimination of inter-lag spurious connections by taking as condition both the parents

Ticker	Eigenvector			In-degree			Out-degree		
	coeff	t-stat	R^2	coeff	t-stat	R^2	coeff	t-stat	R^2
SNV	-0,425	-1,858	0,064	0,054	13,613 **	0,241	-0,015	-4,915 **	0,056
STI	-1,184	-5,733 **	0,000	-0,022	-2,339 **	0,099	0,052	6,228 *	0,354
ZION	0,928	1,152	0,250	-0,043	-5,879	0,173	0,003	0,131 **	0,001
BBT	0,204	6,721 **	0,000	-0,001	-0,190	0,001	-0,006	-1,800	0,004
HCBK	0,086	0,292	0,770	0,027	3,090 **	0,111	-0,010	-0,490 **	0,081
NYCB	0,853	0,393 **	0,001	0,012	0,918 **	0,036	-0,031	-4,695	0,244
RF	-0,970	-15,694	0,118	0,005	0,716	0,003	0,021	1,534	0,052
JPM	-0,903	-242,28 **	0,000	0,022	3,002	0,119	-0,014	-1,911 **	0,078
MS	-0,005	-0,014	0,989	-0,017	-2,441 **	0,148	0,032	4,399 *	0,315
STT	-0,016	-0,087	0,931	0,008	1,789 **	0,024	-0,032	-0,826	0,228
WFC	1,393	1,846	0,066	-0,014	-1,764 *	0,048	-0,031	-2,084	0,297
BAC	-1,345	-5,151 **	0,000	0,016	11,981 **	0,022	0,016	4,467 **	0,058
BK	-0,030	-0,043	0,965	-0,020	-1,337	0,099	0,016	0,112	0,053
C	-2,638	-11,848 **	0,001	0,052	3,431 **	0,199	0,044	22,279 **	0,314
GS	0,232	0,948	0,344	-0,015	-3,254 *	0,185	0,009	1,987 **	0,026
FITB	-0,784	-4,933 **	0,001	0,033	5,104	0,177	-0,012	-1,490 **	0,010
HBAN	1,452	4,944 **	0,001	0,028	4,868 **	0,141	-0,073	-17,64 **	0,380
KEY	-1,041	-2,151 *	0,032	-0,013	-1,348 **	0,042	0,037	3,502	0,219
MTB	0,612	7,096 **	0,001	-0,006	-0,610 **	0,009	-0,007	-5,377	0,012
PBCT	-0,668	-2,826 **	0,005	-0,003	-0,806 *	0,004	0,018	2,543	0,094
PNC	1,614	6,543 **	0,001	-0,014	-2,086 **	0,011	-0,040	-9,941 *	0,352
CBSH	1,130	6,503 **	0,001	-0,027	-24,66 **	0,505	-0,012	-3,379 **	0,064

Table 2.5: This table reports the results of linear regression using the daily volatilities of 22 US financial institutions selected for our application as base for network estimation using pairwise Granger causality. (* and ** stand for statistical significance at the 5% and 1% levels, respectively.)

and children nodes of the transmitter, and the greedy algorithm trying to reduce the impact of instantaneous information transfer by conditioning each transmitter on the other parents of the receiver sharing the same lag.

We demonstrated that our two-level approach provides good results compared to state-of-the-art methodologies using Monte Carlo network simulations including inter- and intra-lag connections or only inter-lag connections. We showed that for a small sample, conditional Granger causality is the adequate measure for information transfer estimation. Transfer entropy proves to be as effective for larger samples. Moreover, this approach is relatively simple to implement and is flexible, as it can be used with different causality measures and extended to specific relationships, such as links whose effect can only be detected through interactions.

We then applied the new algorithm to financial data to analyze the evolution of the US banking sector during the financial crisis. We proposed a new approach to examine the impact of topological characteristics on financial institutions risk. While using linear regressions, we showed the relevance of the environment and the position of an institution inside the network to describe the evolution of its risk. The in-degree,

Ticker	**Eigenvector**			**In-degree**			**Out-degree**		
	coeff	t-stat	R^2	coeff	t-stat	R^2	coeff	t-stat	R^2
SNV	-0,443	-1,643	0,101	0,040	1,868	0,117	0,028	1,241	0,203
STI	-0,332	-2,221 *	0,288	-0,030	-1,146	0,072	0,072	2,978 **	0,031
ZION	-0,161	-3,093 **	0,263	-0,005	-0,647	0,020	0,037	4,952 **	0,002
BBT	-0,805	-4,621 **	0,201	0,029	1,303	0,566	0,034	4,854 **	0,052
HCBK	0,157	0,560	0,002	-0,003	-0,348	0,013	0,003	0,200	0,001
NYCB	-0,091	-11,221 **	0,102	0,018	13,500 **	0,015	0,016	5,700 **	0,047
RF	-0,465	-2,750 **	0,091	0,036	2,896 **	0,077	0,032	2,627 **	0,074
JPM	0,037	0,353	0,085	-0,004	-0,316	0,001	0,024	2,771 **	0,005
MS	0,406	2,840 **	0,012	-0,004	-0,274	0,079	0,009	1,633	0,002
STT	0,192	1,165	0,007	0,029	10,592 **	0,040	-0,008	-0,595	0,172
WFC	-0,272	-1,456	0,079	-0,011	-0,596	0,062	0,033	3,583 **	0,005
BAC	0,035	0,170	0,101	0,079	8,837 **	0,000	-0,047	-2,280 *	0,603
BK	-0,050	-0,428	0,032	0,008	1,522	0,004	0,017	1,307	0,025
C	-0,686	-0,910	0,058	0,024	0,730	0,051	0,052	1,686	0,015
GS	0,013	0,633	0,148	-0,001	-0,301	0,000	0,021	27,492 **	0,001
FITB	-0,337	-2,222 *	0,393	-0,024	-1,146	0,043	0,071	7,493 **	0,021
HBAN	0,001	0,013	0,007	0,060	2,228 *	0,000	0,008	0,789	0,144
KEY	-0,297	-1,915	0,102	0,036	2,819 **	0,063	0,025	3,729 **	0,220
MTB	-0,798	-3,382 **	0,319	0,012	0,736	0,260	0,047	6,866 **	0,007
PBCT	0,265	5,993 **	0,029	0,006	2,873 **	0,089	-0,009	-7,777 **	0,031
PNC	0,314	1,023	0,221	-0,025	-1,045	0,044	0,046	3,812 **	0,066
CBSH	-0,348	-3,316 **	0,030	0,018	6,022 **	0,240	0,008	2,517 *	0,246

Table 2.6: This table reports the results of linear regression using the daily returns of 22 US financial institutions selected for our application as base for network estimation using pairwise Granger causality. (* and ** stand for statistical significance at the 5% and 1% levels, respectively.)

out-degree, and eigenvector of centrality have proved to be determinants of this risk and the results for the different frequencies showed that it exists a trade-off between information and noise leading to an optimal frequency to retrieve financial networks. Finally, we compared these results with those obtained using a simple pairwise approach and confirmed that our algorithm gives a more precise view of the interactions inside networks and therefore gives more information on financial institution fragility.

Chapter 3

Multichannel Information Transfer Estimation

3.1 Introduction

The past decades have seen, in many domains of Science, the development of different tools to better understand the behavior of complex systems in terms of dynamical interactions among their components. This strand of the literature has been recently, mainly devoted to the inference and analysis of network topology, as shows by the numerous contributions in this field (see Zhang et al. (2012); Barigozzi and Brownlees (2013); Barzel and Barabasi (2013) to quote only a few). The network approach helps in understanding the interactions, at a micro level, between its different members, but the relevant interactions may occur at different scales within a larger system. We could indeed consider complex systems such as the world financial sector, the human brain or cells as a series of sub-networks, each having their own specific characteristics, but exchanging information at a higher level through the variables describing their components. The analysis of these interactions operating between sub-networks may be achieved by considering each of them as a multivariate sub-system described by a series of time dependent parameters represented by their nodes. The detection of links between such multivariate complex systems asks for the development of new methodologies able to assess the information flow between a series of predictor and predictee variables represented by the system components.

Most of the existing literature related to information transfer treated the question of multivariate causality detection only partially by looking either at a series of pairwise causality measures to define a global relationship as done in panel analysis (Weinhold (1996); Dumitrescu and Hurlin (2012)) or looking at the effect of several predictor variables using only one predictee variable (Stramaglia et al. (2012)). We propose in this third chapter a new multi-channel causality framework in the vein of Barrett

and Barnett (2010) to identify information flow between a series of predictors and a series of predictees. In addition to the definition of a global information flow between two systems, we propose, as done to some extent in regime switching causality models (Psaradakis et al. (2005b)), to define, at each time step, the local driving pair of parameters involved in the information transfer process. Two multi-channel causality measures are developed in the chapter, one based on the linear regression form of Granger causality (Granger (1969); Geweke (1982)) and the other based on the transfer entropy measure proposed by Schreiber (2000) to account for linear as well as non-linear relationships. The main idea behind our measures is the definition of the local explanatory power of each predictor on every predictee variables and compare them to define the pair of parameters which locally dominates the information transfer process. Once selected for each time step, we aggregate all the relevant observations to estimate a global information flow. We propose therefore a single framework from which we derive two different multichannel causality measures.

We have so far demonstrated the interest of this kind of approach to assess, at a macro level, the interactions existing between systems. But at a lower scale, a multi-channel causality measure may also help in understanding in more details the interaction between two individual components. Indeed, the time series of an individual component may be converted in a set of time dependent variables by decomposing the time series in terms of quantiles or in term of frequencies via wavelet transform. The spectral representation of time series for causality detection has been used widely in Finance and Economics (see Lemmens et al. (2008); Hacker et al. (2014); Bodart and B. (2009); Croux and Reusens (2013); Ciner (2011) among other contributions), Neuroscience (Sato et al. (2006); Brovelli et al. (2004); Friston et al. (2014)) and biology. Most on-going research use the framework proposed by Geweke (1982) which allows to test for causality at each frequency. Although the framework has been adapted for multivariate time series analysis (Barrett and Barnett (2010)), the frequency must always be the same for the predictor and predictee variables. Using our new methodology and wavelet transform, would allow us to assess the cross-frequency information transfer in a single measure looking at transfers with similar or different frequencies.

After having presented our two measures and tested their ability to detect multichannel information transfer via Monte Carlo simulations, we propose an empirical application investigating the dynamical interactions existing in the U.S. financial system at both macro and micro scales. We first start by representing the financial system by a directed network using the methodology proposed by the author in Dahlqvist and Gnabo (2018). We then define, at a micro level, the different channels used in the information transfer process between pairs of connected financial institutions. These channels are expressed both in terms of spectral representation and probability distri-

bution. We eventually look, more broadly, at the information transmission between different parts of the financial system, by dissociating the systematically important financial institutions (SIFI) from the rest of the system.

The remainder of this chapter is organized as follows. Section 3.2 discusses briefly the concepts of Granger causality and present our multi-channel Granger causality test (MCGC). In Section 3.3, we introduce the main ideas behind the transfer entropy and adapt the model to multi-channel causality detection. Section 3.4 studies via Monte Carlo simulations the ability of the two measures to detect multi-channel information transfer using both linear and non linear data generating processes (DGP). Then, Section 3.5 applies the proposed causality measures on a data-set regrouping the major U.S. financial institutions. Section 3.6 eventually concludes.

3.2 Granger based multi-channel causality detection

The root of Granger causality definition dates back to the pioneering work of Wiener (1956) which proposed to define this concept of causality as the ability of a times series to improve the prediction of a another times series. Granger (1969) later formalized this idea using two simple principles: the cause occurs prior to its effect and the cause has unique impact on the future values of its effect. In the context of linear autoregressive models, if we consider two stationary ergodic processes $X \in \mathcal{R}^T$ and $Y \in \mathcal{R}^T$, it is said that Y Granger causes X, if predictions of future values of X based on the past value of Y and on its own past are more precise than the predictions of future values of X knowing only its own past. When using linear regressions, the estimation of Granger causality relies on the comparison of the error terms of two linear regressions, an unrestricted one defined by:

$$x_t = \beta_0 + \sum_{l=1}^{L} b_{1l} x_{t-l} + \sum_{l=1}^{L} b_{2l} y_{t-l} + \varepsilon_t \tag{3.1}$$

and a restricted form, including only the past of X and therefore represented by a simple autoregressive model (AR):

$$x_t = \beta_0 + \sum_{l=1}^{L} b_{1l} x_{t-l} + \eta_t \tag{3.2}$$

The testing procedure is then based on the comparison of the variability of the error terms of both regressions. Eq.3.1 containing more explanatory variables than Eq.3.2,

the variability of ε_t should be higher or equal to η_t. If the variability of ε_t is significantly less than the variability of η_t, it means that the time series Y brings information about the future of X leading thus to an improvement in the prediction of future values of X. As such, different statistical tests have been proposed to compare the variance of the error terms of both the unrestricted and restricted models such as the Granger-Sargent statistic or the Granger-Wald statistic. After having tested different significance tests, we rely in this chapter on the Granger-Sargent statistic which gives the best ratio of false positive on true positive rate. The Granger Sargent statistic is estimated via a bootstrap approach which is preferred to the usual Granger-Sargent significance test, as the optimization used for the multichannel causality test modifies the probability distribution of non causality meaning that we cannot rely anymore on a single F-statistic. The Granger-Sargent statistic tests for the null hypothesis of joint nullity of β_k for the conditional and unconditional form of Granger causality. The test statistic is given by:

$$\mathcal{GS}_{Y\to X} = \frac{(SSR_r - SSR_u)/K}{SSR_u/(T-2L)} \tag{3.3}$$

$$\tag{3.4}$$

where $SSR_r = \sum_t^T \eta_t^2$ and $SSR_u = \sum_t^T \varepsilon_t^2$.

In the proposed bootstrap framework, the F-statistic is computed several times with the time series Y bootstrapped to remove any dependencies. The obtained set of values provides an estimation of the F-statistics for non causality. A Gaussian kernel approach allows the estimation of a probability distribution of non causality based on this set of F-statistics. This probability distribution of non causality is finally used to estimate the probability of observing the initial $\mathcal{GS}_{Y\to X}$ in the population of $\mathcal{GS}_{boot(Y)\to X}$ which gives the significance of the causal link. This last step requires to compute the F-statistic for the non bootstrapped time series Y. The obtained value is then compared with the probability distribution of non causality to obtain eventually the probability of the existence of a causal link. The lower false positive rate of our bootstrap approach compared to the usual Granger-Sargent test goes along a loss of computational efficiency as the F statistic has to be estimated several time with the source series bootstrapped in order to construct the probability distribution of non causality. This leads to a higher computation time.

3.2.1 Model description

As mentioned in the introduction, we consider in this chapter, systems that are characterized by more than one times series, either because these systems have different parameters characterizing them or that their main time series may be divided in sub-

series via quantile or spectral decomposition. In the current literature, pairwise approach are usually adopted in a panel framework to estimate the relationships between these kinds of system, by estimating an average relationships between each pair of parameters. We propose in this section to develop a more global approach by considering in one estimation all the channels, i.e. pair of times series, through which information may transit between the considered multivariate systems. Indeed, as can be done to some extend in the Markov regime switching approach, we try with this algorithm to answer three questions: (i) Does the knowledge of the evolution of one particular system improves the prediction of another one? (ii) When does the transfer of information occurs for a specific channel? (iii) Which channels are preponderant in the description of future states of the system receiving the information? But in contrast with the Markov regime switching approach, we are not looking only at how different explanatory variables may explain the evolution of one dependent variable, we considered several explanatory and dependent variables to infer the causal relationships.

The first algorithm, called mutli-channel Granger causality (MCGC), relies on the Granger causality approach we presented before, to define, at each time step, which couple of time series participates the most to the information transfer process. Indeed, the selection of the 'active' channel is done via the comparison of the local explanatory power of the different transmitting time series. This comparison is based on the error terms of the restricted and unrestricted linear regressions, taking as postulate that the channel through which the highest volume of information transits, will show the highest reduction of the error terms when considering the additional explanatory variable. We need therefore to consider only normalized times series to be able to properly compare their respective error terms. The simplest normalization may be achieved by dividing all the observations of a specific time series by its standard deviation.

We starts by estimating for all possible couples of times series representing the different channels, the restricted and unrestricted linear regressions via the usual ordinary least square (OLS). Considering the relationships between two systems $\boldsymbol{Y} \in \mathcal{R}^{N\times T}$ and $\boldsymbol{X} \in \mathcal{R}^{M\times T}$, described respectively by N and M parameters evolving in time, with T the number of observations, this gives us the following $N \times M$ couples of linear regressions to be estimated:

$$x_t^i = \beta_0 + \sum_{l=1}^{L} b_{1l} x_{t-l}^i + \sum_{l=1}^{L} b_{2l} x_{t-l}^j + \varepsilon_t^{i,j} \tag{3.5}$$

$$x_t^i = \beta_0 + \sum_{l=1}^{L} b_{1l} x_{t-l}^i + \eta_t^{i,j}, \tag{3.6}$$

with $i \in I = \{1, \cdots, M\}$ and $j \in J = \{1, \cdots, N\}$.

Once estimated, the absolute difference between the unrestricted and restricted error terms of the different pair of linear regressions are stored in a time dependent adjacency matrix $\boldsymbol{A}_t = \{a_t^{ij} : i \in I, j \in J, t \in \{1, \cdots, T\}\} \in \mathcal{R}^{N \times M}$, with the elements of the matrix given by:

$$a_t^{i,j} = \left|\varepsilon_t^{i,j}\right| - \left|\eta_t^{i,j}\right| \tag{3.7}$$

Two additional time dependent adjacency matrices $\boldsymbol{U}_t = \{u_t^{i,j}\}$ and $\boldsymbol{R}_t = \{r_t^{i,j}\}$ store respectively the error terms of the unrestricted and restricted linear regressions. We compute then for each time step t, the lowest error terms observed in $\boldsymbol{A}_t$ which forms a matrix $\boldsymbol{B}_t = \{b_t^{i,j}\}$ defined by:

$$b_t^{i,j} = \begin{cases} 1 & \text{if } a_t^{i,j} = \min(\boldsymbol{A}_t) \\ 0 & \text{otherwise} \end{cases} \tag{3.8}$$

In its current form, the matrix $\boldsymbol{B}_t$ shows a relatively high temporal instability in term of channels' variations due to the linear representation used in the regression model. To avoid the effect of this instability in the determination of the optimal channel through which the information is locally transmitted, we rely on a rolling window approach to smooth the channels' variations. The rolling windows are applied on the matrix $\boldsymbol{B}_t$ and the true channel is defined as the one showing the highest number of occurrences in $\boldsymbol{B}_t$ for a specific time window. The stepsize is fixed to one step while the size of the window ω will be chosen so as to maximize the resulting information transfer.

$$c_{t_{rw}}^{i,j} = \begin{cases} 1 & \text{if } \sum_{t=t_{rw}-\omega/2}^{t_{rw}+\omega/2} b_t^{i,j} = \max \sum_{t=t_{rw}-\omega/2}^{t_{rw}+\omega/2} \boldsymbol{B}_t \\ 0 & \text{otherwise} \end{cases} \tag{3.9}$$

with $t_{rw} \in [\omega/2, L - \omega/2]$.

The matrix $\boldsymbol{C}_{t_{rw}} \in \mathcal{R}^{N \times M}$, which provides the selected channel for each time step, is then used to reconstructed the two series of error terms on which the bootstrapped Granger-Sargent test will be performed. As can be seen from Eq.3.10, $\boldsymbol{C}_{t_{rw}}$ allows to select the optimal error terms in $\boldsymbol{U}_t$ and the respective terms in $\boldsymbol{R}_t$.

$$z_t^x = \boldsymbol{J}(\boldsymbol{C}_{t_{rw}} \circ \boldsymbol{R}_{t_{rw}})\boldsymbol{J'}^T \tag{3.10}$$

$$z_t^y = \boldsymbol{J}(\boldsymbol{C}_{t_{rw}} \circ \boldsymbol{U}_{t_{rw}})\boldsymbol{J'}^T, \tag{3.11}$$

with $\boldsymbol{J} \in \mathcal{R}^{1\times N}$, $\boldsymbol{J'} \in \mathcal{R}^{1\times M}$, unit vectors composed of one and used to sum all the values contained in the matrix $\boldsymbol{C}_{t_{rw}} \circ \boldsymbol{R}_{t_{rw}}$ and $\boldsymbol{C}_{t_{rw}} \circ \boldsymbol{U}_{t_{rw}}$ providing the term in $\boldsymbol{U}_t$ and in $\boldsymbol{R}_t$ related to the selected optimal channel for the time step t.

The causality test which is performed using both vector of error terms $\boldsymbol{Z}^x$ and $\boldsymbol{Z}^y$, gives the p-value corresponding to the global flow of information between the system $\boldsymbol{Y}$ and $\boldsymbol{X}$ while the matrix $\boldsymbol{C}_{t_{rw}}$ gives for each time step, the channel which participates the most to this information flow. As mentioned earlier, different window sizes are tested to maximize the information transfer. The optimal window size is defined as the one that minimizes the p-value of the causality test. We make here the additional assumption that optimal channel selection represented by $\boldsymbol{C}_{t_{rw}}$, i.e. the one closest to the reality, coincides with the one minimizing the p-value. We therefore consider that the selected window size is optimal for both the minimization of the p-value and the minimization of the distance between $\boldsymbol{C}_{t_{rw}}$ and the true underlying channel dynamics. This distance will be further investigated in the simulation part. We can eventually summarize the main steps of the algorithm as follows:

- Estimation of the restricted and unrestricted linear regressions for each possible channel from which is computed the adjacency matrices $\boldsymbol{R}_t$ and $\boldsymbol{U}_t$ containing their respective error terms.

- Determination of the local optimal channel, for each time step, by selecting in $\boldsymbol{A}_t$ the minimal difference between the absolute error terms of the unrestricted and restricted models

- Smoothing of the resulting vector $\boldsymbol{B}_t$ using a rolling window approach. Different window sizes $\boldsymbol{\omega}$ are tested to maximize the information flow based on the p-value of the causality test.

- Definition of the time series Z^x and Z^y representing the 'restricted' and 'unrestricted' error terms for the optimal channels selection. At each time step the optimal pair of time series are selected based on the matrix $\boldsymbol{C}_{t_{rw}}$

- Estimation of the information flow for the different window sizes $\boldsymbol{\omega}$ using the bootstrapped Granger-Sargent test and selection of the optimal window size through the minimization of the p-value.

An additional step may be applied at the beginning of the algorithm to improve the channels detection. It involves a pre-search step where the usual pairwise Granger causality test is performed for each pair of time series. Once performed, we apply the previous steps on the channels for which an information transfer has been detected in the pre-search step. This allows us to consider only the most relevant time series and therefore potentially reduce the number of wrongly selected channels. Nevertheless, it means that the considered channels should reach a minimal activity level to be detected in this first step. This implies that channels in which information transit only in a few instance will be discarded even if they potentially lead, at some point, the information transfer process between both systems.

3.3 Transfer entropy based multi-channel causality detection

Before introducing the second multi-channel causality measure, we discuss briefly the concept of transfer entropy on which is based this second algorithm. Assuming W and Z, two stationary Markov processes of order h and k, the transfer entropy determines the reduction in uncertainty about Z when learning the past of W, the past of Z being already known until an order h. The formal definition is given by:

$$T_{W\to Z} = H(z_t \mid z_{t-h}) - H(z_t \mid z_{t-h}, w_{t-k}) \tag{3.12}$$

$$= \sum_{z_t, z_{t-h}, w_{t-k}} p(z_t, z_{t-h}, w_{t-k}) log_\alpha(p(z_t \mid z_{t-h}, w_{t-k})) \tag{3.13}$$

$$= \sum_{z_t, z_{t-h}, w_{t-k}} p(z_t, z_{t-h}, w_{t-k}) log_\alpha(\frac{p(z_t \mid z_{t-h}, w_{t-k})}{p(z_t \mid z_{t-h})}) \tag{3.14}$$

The estimation of the different joint probabilities may be done via a Kernel approach (Schreiber (2000)) or more simply using a symbolization process (Staniek and Lehnertz (2008)). We favor here the second approach, as it is computationally more efficient to reduce the complexity of the signal contained in the time series before the estimation of the joint probabilities. We define a specific alphabet for each time series using the spectral quantile symbolization (SQS) proposed in chapter 1, but other approaches may be used to compute the joint probabilities. The complexity of the symbolization derives directly from the number of quantiles considered to separate the different regions, each of them represented by a specific symbol. Once the times series have been symbolized, the different probabilities are computed based on a frequency analysis of the different patterns.

As the transfer entropy suffers from small sample effect, a bias due the finite size of the time series used for the estimation of the joint probabilities, we apply a bootstrapped methodology similar to the one used with the Granger causality setup to remove this bias. This methodology defines the effective transfer entropy as follows:

$$TE^{\alpha}_{W\to Z} = T_{W\to Z} - T^{\alpha}_{W_{Boot}\to Z} \tag{3.15}$$

where $TE^{\alpha}_{W_{Boot}\to Z}$ is the value of transfer entropy, with series W bootstrapped to remove any dependencies and α the specific significance level of the test. This value is obtained via a Gaussian kernel approach using a series of transfer entropy with series W bootstrapped.

3.3.1 Model description

Similarly to the Granger causality based measure, we start by considering two systems $\boldsymbol{X}$ and $\boldsymbol{Y}$ described by several parameters indexed respectively by $i \in I = \{1, \cdots, M\}$ and $j \in J = \{1, \cdots, M\}$ whose evolution in time is represented by time series depending on the index t. The first step of the multi-channel transfer entropy algorithm (MCTE) is to symbolize the different time series. The quantiles used in the SQS are selected based on the maximization of the overall transfer entropy between each pair of time series. We obtain, for each channel, a specific set of quantiles for the time series representing the receiver and the transmitter. It implies a different symbolization for the $N \times M$ pairs of time series. As the symbols of these different channels will be mixed in the maximization process used to determine the overall transfer of information, we have to consider a different alphabet for each channel. We store then the symbolized time series in two separate time dependent adjacency matrices, one related to the transmitter and the other to the receiver, they are denoted respectively $\boldsymbol{S}_t = \{s_t^{i,j}\}$ and $\boldsymbol{R}_t = \{r_t^{i,j}\}$.

We then determine for each pair of time series X^i and Y^j the relative importance of each observation of Y^j in the definition of the overall information flow from $\boldsymbol{Y}$ to $\boldsymbol{X}$. Taking the simple case of two stationary Markov processes of order 1, we look at the impact of removing y^j_{t-1} from the estimation of both joint probabilities, $p(x^i_t, x^i_{t-1}, y^j_{t-1})$ and $p(x^i_{t-1}, y^j_{t-1})$ on the transfer entropy value. We rely here on the fact that only some observations participate to the information transfer process, and that the removal of such observations should lead to an overall reduction of the transfer entropy value. These observations are the ones that increase the asymmetry existing in the condi-

tional probability distribution $p(x_t^i, x_{t-1}^i \mid \boldsymbol{y}_{t-1}^j)$ relative to the possible values taken by $\boldsymbol{y}_{t-1}^j$. By removing this type of observations, we make the conditional probability distribution $p(x_t^i, x_{t-1}^i \mid \boldsymbol{y}_{t-1}^j)$ more symmetric and therefore closer to the case of a conditional probability equivalent for all $\boldsymbol{y}_{t-1}^j$ which should reduce the overall transfer of information. The opposite may be true for observations of Y^j which do not help in the understanding of future values of X^i. In this case, the removal of the observation could lead to an asymmetrization of the conditional probability distribution $p(x_t^i, x_{t-1}^i \mid \boldsymbol{y}_{t-1}^j)$ and thus to an increase of the transfer entropy. This means that if we observed a negative variation, the selected observation Y^j helps in describing future state of X^i and if positive this means that the observation Y^j does not bring any valuable information about the future state of X^i. This transfer entropy variation is given by Eq.3.16.

$$\Delta T_{Y^j \to X^i}^t = T_{Y^j \setminus Y_t^j \to X^i} - T_{Y^j \to X^i} \tag{3.16}$$

Once we have estimated, for each time step and each pair of parameters, the variation of transfer entropy, we define the 'active' channel as the pair of parameters presenting the highest reduction of transfer entropy. This minimal variations are then stored in a time dependent adjacency matrix $\boldsymbol{B}_t = b_t^{i,j} \in \mathcal{R}^{N \times M}$.

$$b_t^{i,j} = \begin{cases} \min \boldsymbol{A}_t & \text{if } a_t^{i,j} = \min \boldsymbol{A}_t \\ 0 & \text{otherwise} \end{cases} \tag{3.17}$$

with here the adjacency matrix $\boldsymbol{A}_t$ defined by:

$$\boldsymbol{A}_t = \begin{pmatrix} \Delta T_{Y^1 \to X^1}^t & \Delta T_{Y^2 \to X^1}^t & \cdots & \Delta T_{Y^j \to X^1}^t \\ \Delta T_{Y^1 \to X^2}^t & \Delta T_{Y^2 \to X^2}^t & \cdots & \Delta T_{Y^j \to X^2}^t \\ \vdots & \vdots & \ddots & \vdots \\ \Delta T_{Y^1 \to X^i}^t & \Delta T_{Y^2 \to X^i}^t & \cdots & \Delta T_{Y^j \to X^i}^t \end{pmatrix} \tag{3.18}$$

As more than one pair of time series may show a negative local variation of transfer entropy, we are again only able to define which channel is dominating at a specific moment the information transfer process. Additionally, thanks to the symbolization process we do not have to normalize the time series before applying our algorithm as we did for the Granger based measure. The symbolization process acts as a normalization due to the quantiles used to assign the different symbols.

The next step involves the rolling windows approach we used in Sect.3.2 to reduce the temporal variability of the selected pair of time series in matrix $\boldsymbol{B}_t$. This step is even more essential for the transfer entropy based measure because of the non-linearity of the information flow we usually try to detect with this causality measure. Indeed, for these types of relationship we often observe a higher sparsity of the patterns representing the transfer of information. We compute then, for each pair of parameters, the average variation observed in the considered rolling window. Based on this average value, we select the pair of parameters showing the lowest average transfer entropy variation and store it in an adjacency matrix $\boldsymbol{C}_{t_{rw}}$ given by:

$$c_{t_{rw}}^{i,j} = \begin{cases} 1 & \text{if } \sum_{t_{rw}=t_{rw}-\omega/2}^{t+\omega/2} b_t^{i,j} = \min \sum_{t=t_{rw}-\omega/2}^{t_{rw}+\omega/2} \boldsymbol{B}_t \\ 0 & \text{otherwise} \end{cases} \tag{3.19}$$

with $t_{rw} \in [\omega/2, L - \omega/2]$.

Once a pair of time series has been selected for each time step, we estimate the final transfer entropy by first defining the composite time series of the system transmitting the information and the one receiving it, based on the symbols stored in the time dependent adjacency matrices $\boldsymbol{S}_t = \{s_t^{i,j}\}$ and $\boldsymbol{R}_t = \{r_t^{i,j}\}$ and the channel activity matrix C_t. These two times series are determined as follows:

$$z_t = \boldsymbol{J}(\boldsymbol{C}_{t_{rw}} \circ \boldsymbol{R}_{t_{rw}})\boldsymbol{J'}^T \tag{3.20}$$

$$w_t = \boldsymbol{J}(\boldsymbol{C}_{t_{rw}} \circ \boldsymbol{S}_{t_{rw}})\boldsymbol{J'}^T \tag{3.21}$$

with $\boldsymbol{J} \in \mathcal{R}^{1\times N}$, $\boldsymbol{J'} \in \mathcal{R}^{1\times M}$, unit vectors composed of one and used to sum all the values contained in the matrix $\boldsymbol{C}_{t_{rw}} \circ \boldsymbol{R}_{t_{rw}}$ and $\boldsymbol{C}_{t_{rw}} \circ \boldsymbol{S}_{t_{rw}}$ providing the term in $\boldsymbol{R}_t$ and in $\boldsymbol{S}_t$ related to the selected optimal channel for the time step t.

After having computed the transfer entropy for different time windows w based on the vectors Z and W, we have to define the effective transfer entropies to get rid of both the impact of our optimization process and the small sample effect. Indeed, as we have selected, at each time step, the pair of parameters showing the lowest transfer entropy variation, we have artificially increased the resulting global transfer entropy. This increase is especially striking for the smallest windows of the stabilization process as can be seen from Fig.3.1. Therefore we cannot simply bootstrapped the resulting W time series; we need to apply the previous steps on a bootstrapped system in which each time series Y^j representing a parameter of the system $\boldsymbol{Y}$ is bootstrapped. Given

the series of bootstrapped transfer entropies for each window size ω, we are able to estimate the effective transfer entropy at a specific level using Eq.3.16. We may also define the probability of observing our underlying transfer entropy in the population of bootstrapped transfer entropies to a get a value comparable to the p-value of the MCGC. In both case a Gaussian kernel is used to define the probability distribution of the the bootstrapped transfer entropies. The different steps of the algorithm may be summarized as follows:

- Symbolization of every pair of time series via SQS and creation of two time dependent adjacency matrices $\boldsymbol{S}_t = \{s_t^{i,j}\}$ and $\boldsymbol{R}_t = \{r_t^{i,j}\}$ regrouping the resulting symbol for respectively the transmitter and the receiver of the information
- Estimation of the transfer entropy variation for each pair of time series and each observations; the resulting variations are stored in the matrix $\boldsymbol{A}_t$.
- Determination of the local optimal channel for each time step, by selecting the minimal transfer entropy variation contained in $\boldsymbol{A}_t$
- Smoothing of the resulting vector $\boldsymbol{B}_t$ using a rolling window approach. For each window, the selected channel is the one reporting the lowest total transfer entropy variation. Different windows sizes $\boldsymbol{\omega}$ are tested to optimize the smoothing by maximizing the resulting effective transfer entropy.
- Definition of the symbolized time series Z and W based on the pair of time series selected in $\boldsymbol{C}_{t_{rw}}$.
- Estimation of the transfer entropy for the different window sizes $\boldsymbol{\omega}$ based on the time series Z and W. Definition of the effective transfer entropy using a bootstrapped approach consisting in applying the previous step on a bootstrapped version of the system $\boldsymbol{Y}$. Use this bootstrapped values to compute either the effective transfer entropy or the probability of information transfer using a Gaussian kernel.

Recalling what has been done in Sect.3.2, we may add a pre-search step to select only the channels for which a transfer information is observed considering the whole sample. This increase the ability of our algorithm to detect the true underlying channels. Another filter may be applied during the selection of the 'active' channel. Indeed, as a positive variation of transfer entropy is considered as a 'non causal' observation, if we face this kind of observation for every channels, we should conclude that no information transfer occur for the selected time step. We could therefore add an additional state indicating that no information transfer is underway during a specific time step.

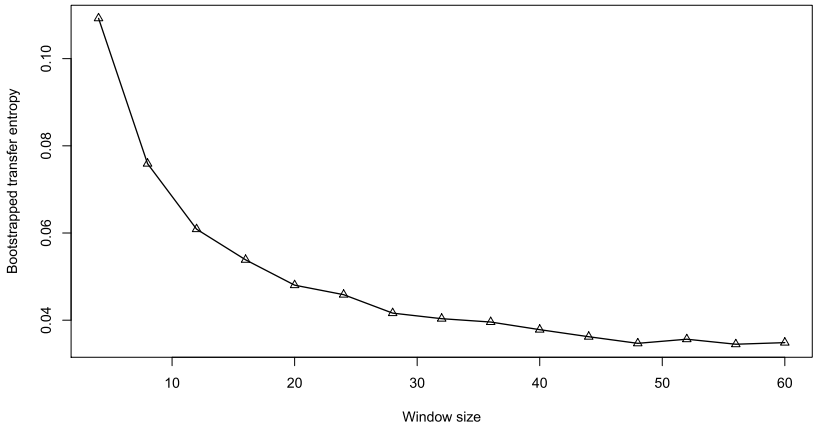


Figure 3.1: The figure shows the evolution of the bootstrapped transfer entropy relative to the window sizes selected during the stabilization process.

3.4 Simulation

We run a Monte Carlo experiment to investigate the ability of both our models to first accurately represent the true information transfer and then to effectively detect the periods during which a channel is dominating the information transfer process. We consider two groups of data generating processes (DGPs) simulating the information transfer either linearly or non-linearly. This allows us to test both causality measures, knowing the Granger based model should normally have a higher detection rate in case of linear relationships while transfer entropy based model should perform better for non-linear ones. The two simulated systems $\boldsymbol{X}$ and $\boldsymbol{Y}$ are represented by three time varying parameters. We consider sequentially two, three and four connected pairs of times series to define the ability of our model to infer causal relationships when the complexity of the information transfer process increases. In each case, the regime switching is done via a Poisson process for which we test ten different intensities ranging from 10 percent to 1 percent to determine the impact of the switching frequency on the causality detection. The connected time series of the system $\boldsymbol{X}$ are simulated as followed when considering a linear information transfer:

$$x_t^1 = \alpha x_{t-1}^1 + \beta y_{t-1}^1 + (1-\alpha-\beta)\varepsilon_t^1 \qquad \text{when } s_t = 0 \tag{3.22}$$

$$x_t^3 = \alpha x_{t-1}^3 + \beta y_{t-1}^3 + (1-\alpha-\beta)\varepsilon_t^3 \qquad \text{when } s_t = 1 \tag{3.23}$$

with ε_t^i a Gaussian white noise with zero mean and unit standard deviation and s_t defined as follows:

$$s_t = \begin{cases} 1 & \text{if } \{s_{t-1} = 1 \ \& \ p_t = 0\} \ or \ \{s_{t-1} = 0 \ \& \ p_t = 1\} \\ 0 & \text{otherwise} \end{cases} \tag{3.24}$$

with $p_t = pois(1, I)$ and I the intensity.

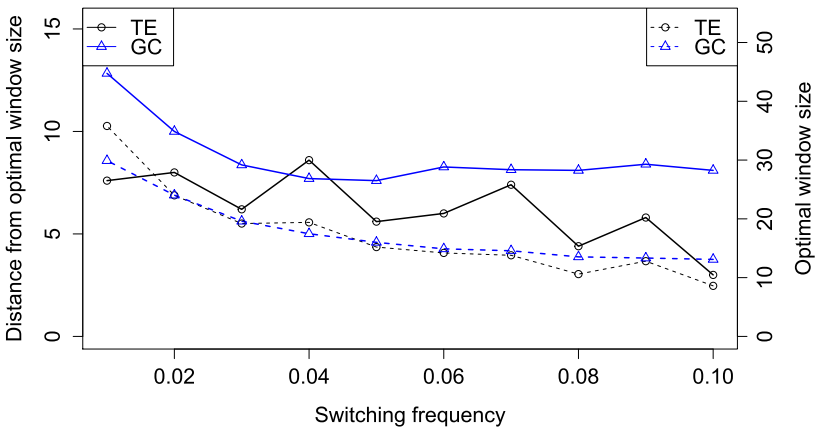
The non connected series X^i and all the time series Y^j follow the usual auto-regressive process (AR):

$$x_t^i = \alpha x_{t-1}^i + (1 - \alpha)\varepsilon_t^i \tag{3.25}$$

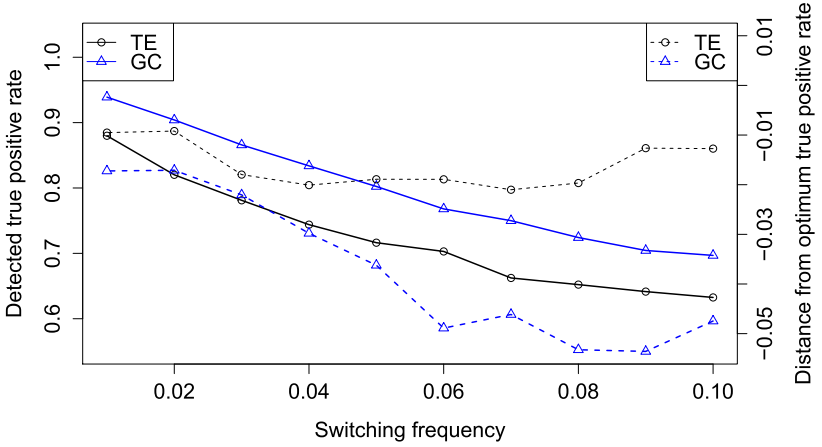
$$y_t^j = \alpha y_{t-1}^j + (1 - \alpha)\eta_t^j \tag{3.26}$$

where $i \in \{1, 2, 3\}$ depending on the period and $j \in \{1, 2, 3\}$.

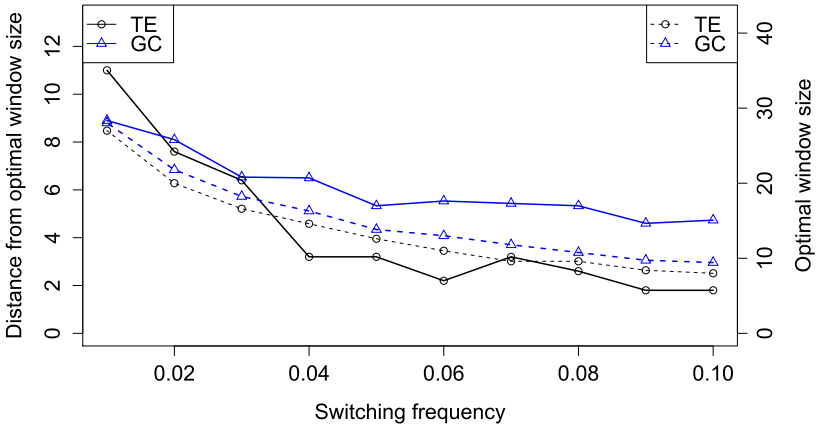
As can be seen from Eq.3.22, we connect for this simulation set, alternatively the time series X^1 and Y^1 and X^1 and Y^3. The two other DGPs with respectively three and four connected pair of nodes are described by similar equations except for the determination of s_t. In those cases, the Poisson process defines at which frequency a switch occurs but then a uniform distribution is used to determine which channel is selected for the next period. The β parameters are set for all simulations to 0.6 while the α is set to 0.2 during information transfer period and then to 0.4 for the AR part. We then simulate 200 systems for each DGP and each intensity. The window sizes used for the definition of the optimal channel selection, range between 4 and 60 observations with a step-size of 4 observations. For the transfer entropy based measure, we start the estimation by defining, for each pair of simulated time series, the position and bandwidth used in the SQS that maximize the resulting transfer entropy. The significance level is then set to 0.05 for both causality measures and the number of bootstrapped time series to 10 for the estimation of the effective transfer entropy.



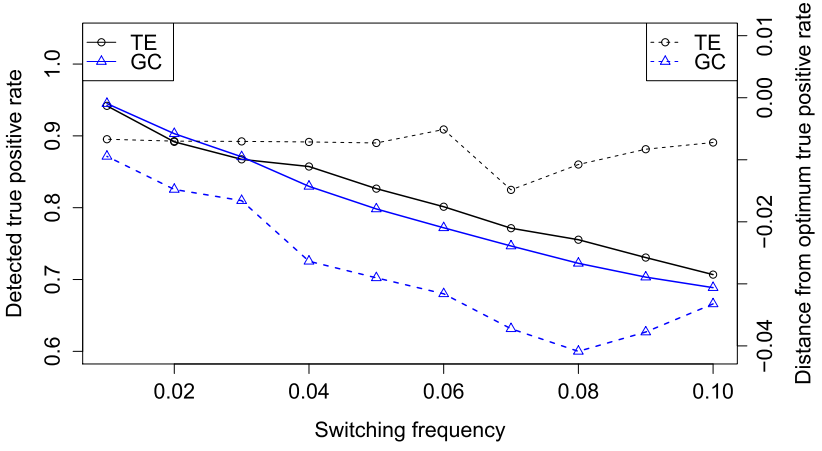
(a)Window size for two channels DGP



(b)True positive rate for two channels DGP

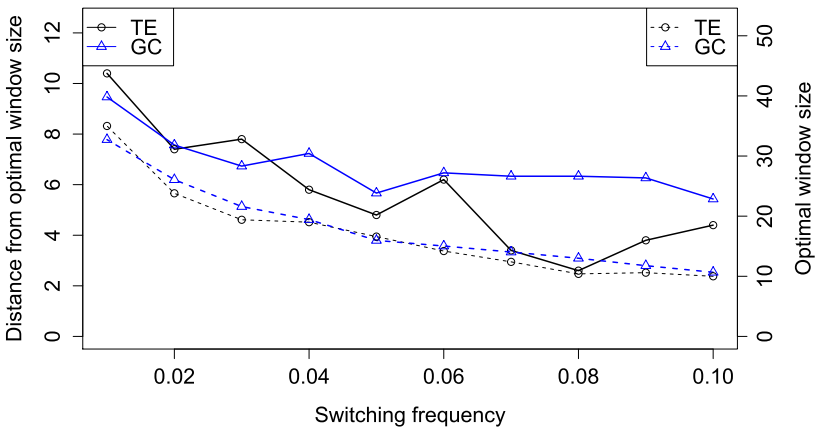


(c)Window size for three channels DGP

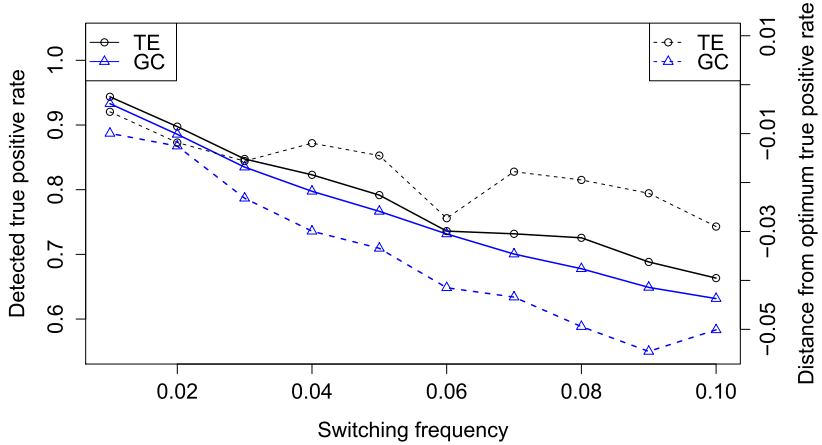


(d)True positive rate for three channels DGP

Figure 3.2: The figures (a) and (c)show, for the linear based DGP, the average distance expressed in number of observations between the window size giving the highest value in term of transfer entropy and Granger causality and the window size leading to the highest true positive rate when comparing the channel selected at each time step and the true underlying channels. They also gives the average window size leading to the highest true positive rate. The figures (b) and (d) report the average true positive rate when using the window size giving the highest value of information transfer and also the average distance between this rate and the maximum true positive rate.



(a)Window size for four channels DGP



(b)True positive rate for four channels DGP

Figure 3.3: The figure (a) shows, for the linear based DGP, the average distance expressed in number of observations between the window size giving the highest value in term of transfer entropy and Granger causality and the window size leading to the highest true positive rate when comparing the channel selected at each time step and the true underlying channels. They also gives the average window size leading to the highest true positive rate. The figure (b) reports the average true positive rate when using the window size giving the highest value of information transfer and also the average distance between this rate and the maximum true positive rate.

Fig.3.2 reports the results of the simulation for the three different DGPs with the x-axis representing the different intensities. We look here at the ability of the two causality measures to detect the true underlying 'active' channels and verify the hypothesis that the window sizes selected via information transfer maximization gives the closest representation of the true channels dynamic. We use therefore two different sets of measures based respectively on the average window sizes and on the average true positive rate which is defined as the number of observations for which the selected channel matches the true underlying one on the total number of observations. The figures on the left show the average window size giving the highest true positive rate, and the distance between this window size and the one selected when maximizing the information transfer, i.e. minimizing the p-value of the MCGC and maximizing the value of the effective MCTE.

As can be seen from Fig.3.2 (a), (c) and Fig.3.3 (a), the optimal window size increases for both MCGC and MCTE when the intensity decreases meaning that the windows size could be an indicator of the underlying switching frequency. We also see that the distance between the optimal window size and the selected one, increases with the average size of the window. As regards the true positive rate, Fig.3.2 (b), (d) and Fig.3.3 (b) provides both the true positive rate obtained when the window size is selected using the information transfer and the distance between this true positive rate and the optimal one. We observe a clear increase of the true positive rate when the switching frequency decreases. We see also that the true positive rate does not seem to suffer from an increase in the underlying model complexity with overall similar values for the three cases. The distance between both true positive rates is also shrinking when the switching frequency decreases. It seems therefore that the curvature of the true positive rate given the window size become less pronounce around the optimum when the intensity reduces, i.e. the switching frequency rises. This could explain the difficulties to select the right window size for low intensity values. It is worth mentioning that the transfer of information is detected with both methods for all intensities.

As regards the simulation using non-linear information transfers, we follow the same approach by looking at the effect of the switching frequency and the complexity of the information transfer dynamic. The simulated connected time series of the system $\boldsymbol{X}$, for a non-linear information transfer, are given by:

$$x_t^1 = \alpha x_{t-1}^1 + \beta (y_{t-1}^1)^2 + (1-\alpha-\beta)\epsilon_t^1 \qquad \text{when } s_t = 0 \tag{3.27}$$

$$x_t^3 = \alpha x_{t-1}^3 + \beta (y_{t-1}^3)^2 + (1-\alpha-\beta)\epsilon_t^3 \qquad \text{when } s_t = 1 \tag{3.28}$$

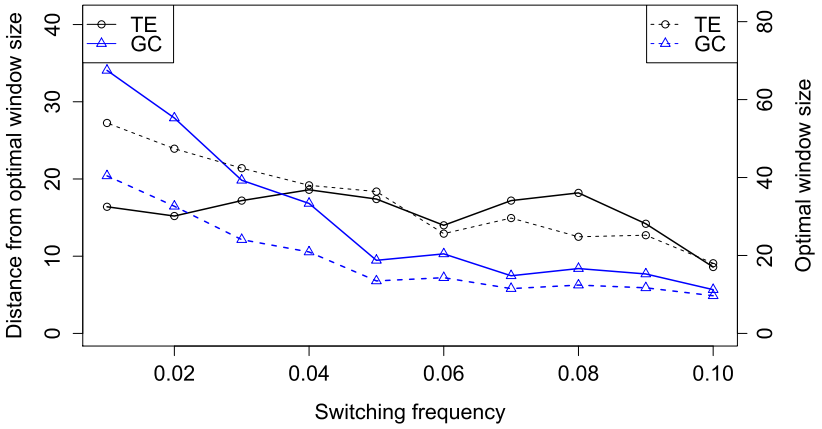
with the non connected pairs X^i and all the time series Y^j following again an AR

process:

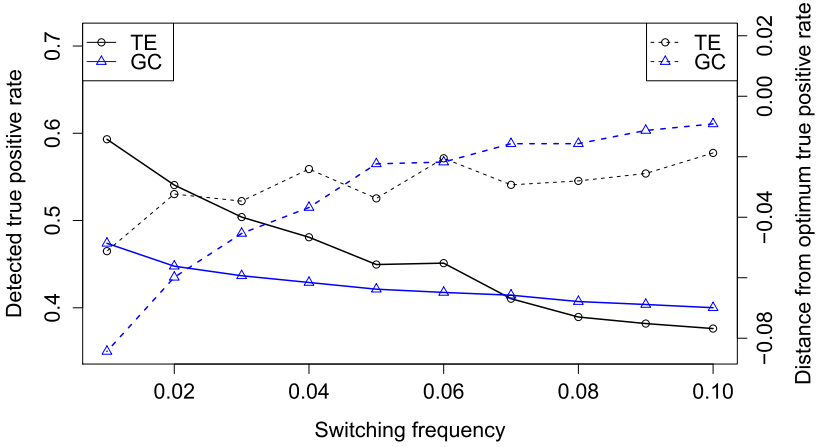
$$x_t^i = \alpha x_{t-1}^i + (1-\alpha)\epsilon_t^i \tag{3.29}$$

$$y_t^j = \alpha y_{t-1}^j + (1-\alpha)\eta_t^j \tag{3.30}$$

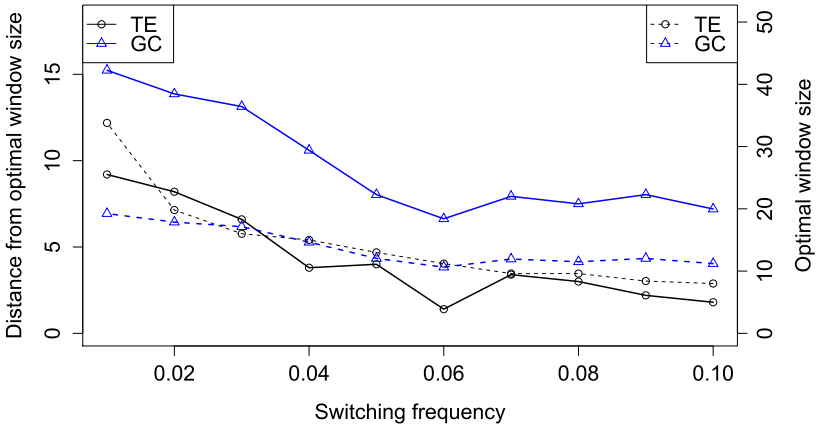
where $i \in \{1,2,3\}$ depending on the period and $j \in \{1,2,3\}$.



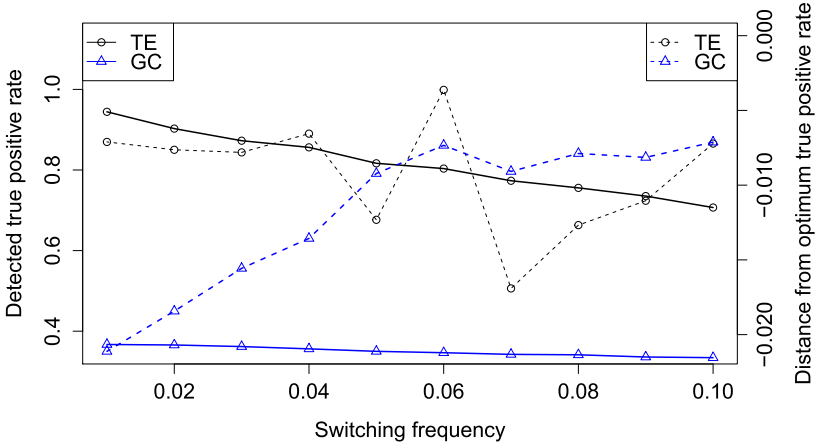
(a)Window size for two channels DGP



(b)True positive rate for two channels DGP

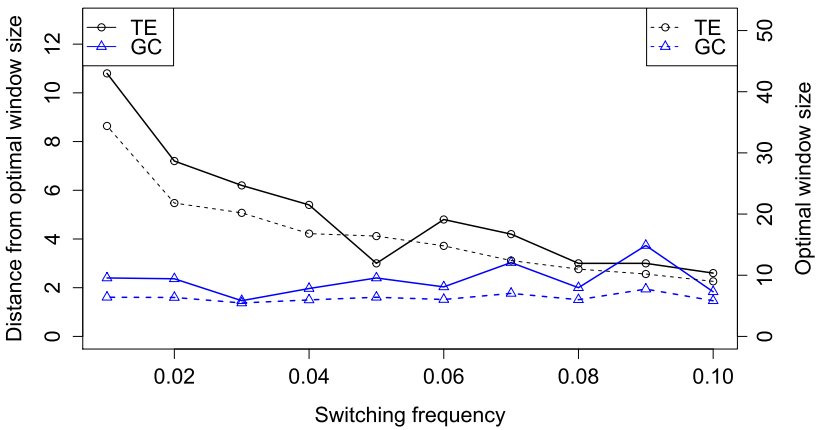


(c)Window size for three channels DGP

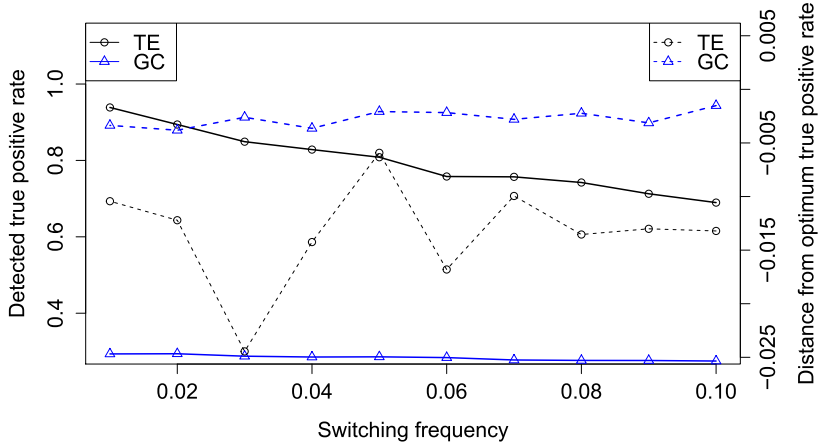


(d)True positive rate for three channels DGP

Figure 3.4: The figures (a), (c) show, for the nonlinear based DGP, the average distance expressed in number of observations between the window size giving the highest value in term of transfer entropy and Granger causality and the window size leading to the highest true positive rate when comparing the channel selected at each time step and the true underlying channels. They also gives the average window size leading to the highest true positive rate. The figures (b), (d) report the average true positive rate when using the window size giving the highest value of information transfer and also the average distance between this rate and the maximum true positive rate.



(a)Window size for four channels DGP

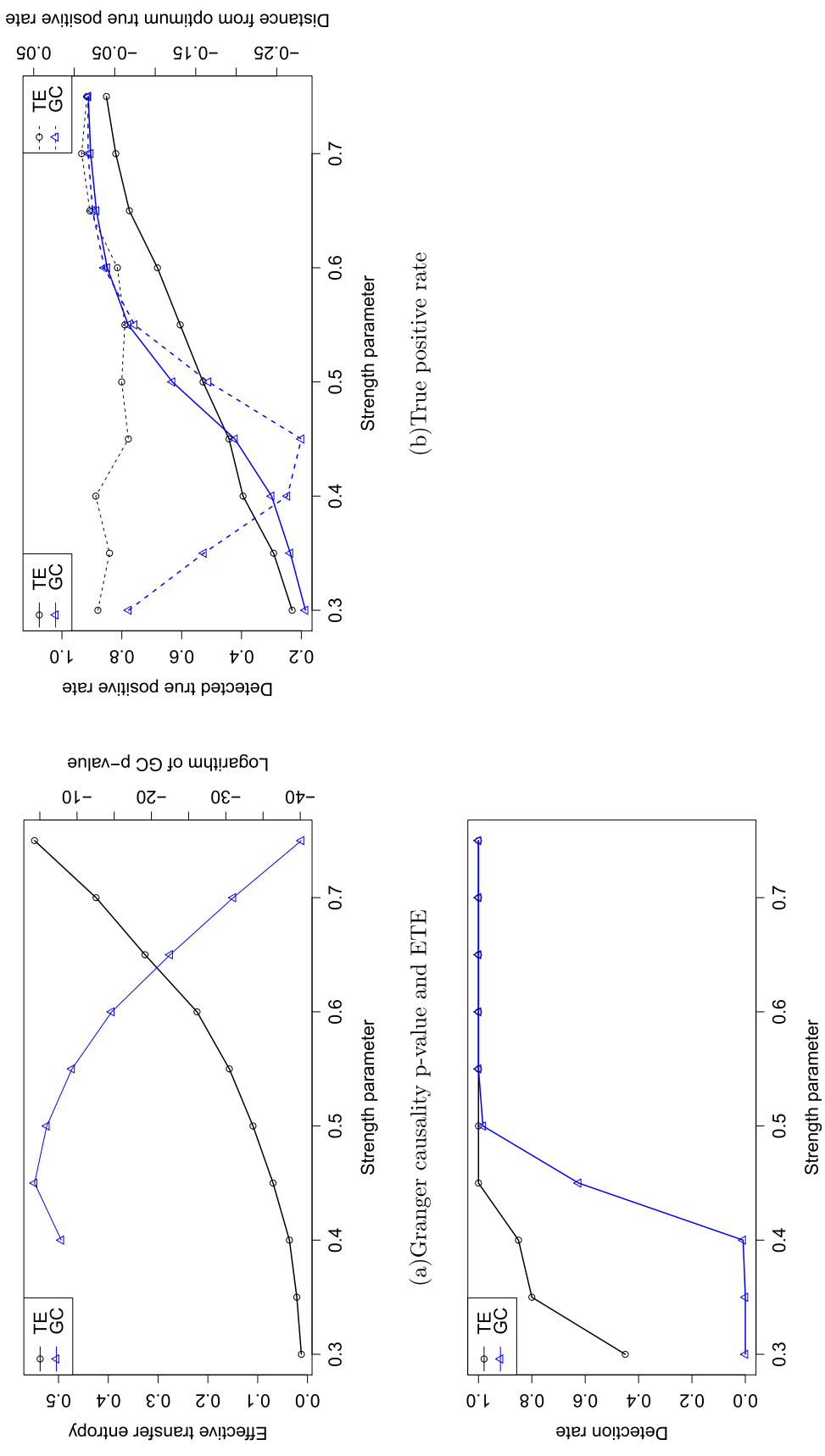


(b)True positive rate for four channels DGP

Figure 3.5: The figure (a) shows, for the nonlinear based DGP, the average distance expressed in number of observations between the window size giving the highest value in term of transfer entropy and Granger causality and the window size leading to the highest true positive rate when comparing the channel selected at each time step and the true underlying channels. They also gives the average window size leading to the highest true positive rate. The figure (b) reports the average true positive rate when using the window size giving the highest value of information transfer and also the average distance between this rate and the maximum true positive rate.

Fig.3.4 and 3.5 gives the same measures as in the linear case. As can be seen from Fig.3.4 (b), (d) and 3.5 (b), the transfer entropy based measure work very well compared to the Granger causality based measure, with results similar to the linear case. These good results are not a surprise as transfer entropy is often used to study non linear relationships. The bad results for the Granger causality based measure come simply from the fact that apart from the case of only two 'active' channels, it does not detect any causal relationship between the two systems. As for the results from Fig.3.4 (a), (c) and 3.5 (a), we have a similar trend compared to the previous linear case.

We have so far assessed the ability of both causality measures to detect the true underlying channels using different switching frequencies and linear as well as non linear relationships. Nevertheless, we have not yet inquired the impact of the strength of these relationships on their detectability. We propose to examine this effect by considering the most simple simulation set which uses linear information transfer and includes only two 'active' channels. We fix the intensity of the Poisson process to 2 percent and test then different value for the β parameter ranging from 0.3 to 0.75. The results are reported in Fig.3.6, with figure (a) presenting the average value of the effective transfer entropy and the logarithm of the average p-value estimated via the bootstrapped approach for the Granger based model. We see from these results and comparing with Fig.3.6(b) that the ability of our models to represent correctly the underlying channel dynamics depends directly from the results of both model in terms of effective transfer entropy and p-value. The results from Fig.3.6(c) shows that the transfer entropy detects surprisingly better the information transfer for the lowest value of the β parameter despite the fact that the underlying model is linear.



(a)Granger causality p-value and ETE

(b)True positive rate

(c)Detection rate

Figure 3.6: The figure (a) reports the value of the effective transfer entropy and the p-value of the Granger causality while figure (c) presents the percentage of detected causal relationship for both multi-channel causality measures. The figure (b) shows the average true positive rate when using the window size giving the highest value of information transfer and also the average distance between this rate and the maximum true positive rate.

We have shown in this series of Monte Carlo simulations that our model could detect in most cases a transfer of information between two multivariate systems, independently of the relationship complexity. We have also proven that the ability of both multi-channel measures to determine the true channel dynamics increases when the frequency of the channel switching decreases. The value of the effective MCTE or the p-value of the MCGC could be a good measure of the distance between the detected channel dynamics and the true underlying one, as these values evolve in a similar fashion compared to the true positive rates. Although the distance between the optimal window size and the one detected can be sometimes relatively high, the result in term of true positive rate does not seem to be significantly affected, especially when the switching frequency decreases. Finally, the MCTE seems to detect more easily a link between two systems when the strength of the relationship decreases. The MCTE should also be the measure to consider when looking at non-linear relationships.

3.5 Application to Financial institutions

In order to demonstrate the usefulness of our new causality measures, we propose an empirical application looking at the interactions existing between individual financial institutions inside the U.S. financial system or between parts of this system. As illustrated by its implication in the financial crisis of 2008, the U.S. financial institutions play a crucial role in the stability of the global financial system. A better understanding of the dynamics of the channels through which information propagates inside the system seems therefore important to solve many outstanding issues in finance. In this application, we look at these dynamics from two different perspectives. We start by defining, at a micro level, the channels through which information are transmitted between pairs of financial institutions by considering different frequencies and different quantiles inside the time series probability distribution. In a second step, we consider at a broader level, the transmission between subgroups of financial institutions.

Looking from the micro perspective, we have seen in Sect.3.1 that the usual framework used for causal analysis in the frequency domain involves the model proposed by Geweke (1982) and further developed by Breitung and Candelon (2006). In contrast with this literature (Lemmens et al. (2008); Bodart and B. (2009); Croux and Reusens (2013); Ciner (2011)), we propose in this application to apply our new method on the wavelet transform of the selected times series to allow us to look at cross-frequency causal relationships. Several studies have already relied on wavelet transforms to consider causal relationships at different frequencies in the context of Economics(see Hacker et al. (2014) or Alzahrani et al. (2014)) but not cross-frequency causal relation-

ships. The rationales behind the investigation of causal relationships in the frequency domain comes from the existence of multiple investment horizons across market players. Indeed, the market participants investing in financial products such as shares have different objectives and therefore different horizons, mutual or hedges funds having longer horizon and market makers or daily traders speculating on movements in the shorter term. It implies that several frequencies may carry different information related to the financial institution fundamentals on longer term or rumors and market volatility on shorter term. The relative importance of the different frequencies may of course evolves in time, with a higher importance of high frequencies in period of turmoil and lower ones in period of growth. Our multichannel model applied on wavelet transforms may accommodate these different features of the financial market.

Regarding the effect of the probability distribution parameters on the causality detection, we draw away from the traditional approach of Granger causality which looks primarily at the relationships existing in mean and variance (see among others Granger (1969); Hafner and Herwartz (2006)) as it considers only parts of the distribution. Inspired by the literature on quantile regressions (Koenker and Bassett, 1978) and the recent development of methods to infer causality over different quantiles (Chuang et al., 2009; Chen et al., 2009) or in the left tail using a Value-at-Risk (VaR) framework (Hong et al., 2009a; Candelon et al., 2013), we adapt our model based on Granger causality to detect dynamically the quantiles responsible for the information transfer between financial institutions. We propose then a complementary approach based on the transfer entropy version of our algorithm to look in more details at cross-quantile transmissions. Indeed the quantile regression framework considers only information transmission and reception at a given quantile. We proposed therefore to benefit the flexibility of the symbolization process used in the transfer entropy estimation to be able to look at information transmission with different quantiles for the transmitter and the receiver time series.

We eventually look, from a macro perspective, at the transmission between parts of the financial system as it is done to some extend in Yang and Zhou (2013) where they use clusters and principal component analysis (PCA) to reduce the complexity of each group and be able to estimate information transfer with the usual Granger causality framework. Our new algorithm allows us to avoid this PCA and to keep the entire set of times series in each group. We define in this application two subgroups and analyze the links existing between the systematically important financial institutions (SIFI) and the rest of the financial network. Indeed, given the systemic importance of these institutions, their relationships with the rest of the system should give us information about their impact on the financial sector fragility. Each connection between financial institutions from both groups is here considered as a possible channel for the transmis-

sion of information between the two sub-systems. We then apply our method to infer dynamically the channels activity between the two subgroups of financial institutions.

3.5.1 Empirical data and network inference

In this application, we employ stock prices of 22 U.S. financial institutions traded on the NYSE and Nasdaq, covering the period from January 2005 to January 2013. Our dataset includes the largest U.S. financial institutions along 8 institutions included in the list of current SIFIs. Two different data frequencies are used in the application with daily and hourly stock prices.

The first step of our analysis is the definition of the causal links to be considered in our model. The determination of the relevant links is done by inferring the effective network representing our financial system using the method proposed by the author in Dahlqvist and Gnabo (2018) which takes care of possible indirect links. In order to properly infer the network, we first estimate the daily and hourly variations of the stock prices in order to avoid the issue of stationarity. Then, the resulting data-set is filtered from the effect of common global factors using linear regressions with as explanatory variable the index S&P 500. The residuals of each time series is then used for the estimation of the effective network which is done in two steps: a pre-search step where a pairwise causality test is performed on every pair of financial institutions to connect the network and a pruning step where the possible indirect links are removed using conditional causality test. Except for the Sect.3.5.4, we use the traditional Granger causality for the estimation of the network. The application is mainly based on the Granger causality version of the proposed algorithm as it suffers less from small sample effect and is computationally more efficient although it gives only a partial view as it is design to tackle only linear relationships.

3.5.2 Mutli-channel causality detection in the frequency domain

Once the effective networks have been estimated for both the daily and hourly data-sets, we may look in more details at how the information circulates inside these networks. We start by considering the different frequencies at which a signal may be sent and received. Each time series of the considered data-set is therefore decomposed into a series of signals representing the different frequencies. This decomposition is done via a wavelet transform with five different wavelengths of respectively 4, 8, 16, 32 and 64 observations. Once estimated, the set of signals obtained for each financial institution is considered as an individual system on which the method presented in Sect.3.2.1 may be applied.

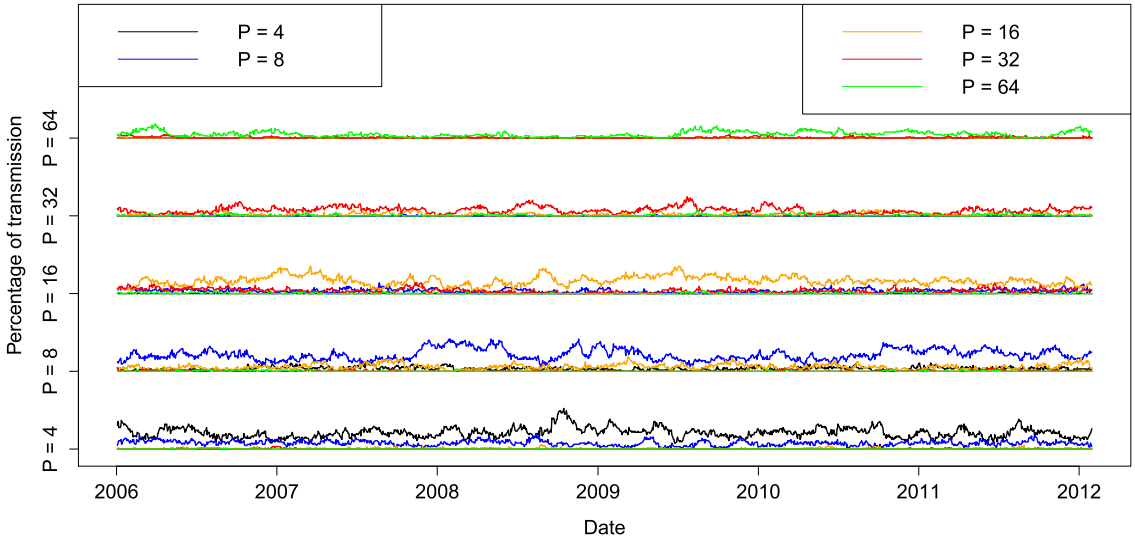


Figure 3.7: This figure gives the results of the MCGC for the dataset using daily stock returns. It provides for each channels, represented by a specific pair of frequencies, the percentage of information traveling inside the network which passes through them. The y-axis reports the receiving frequencies while the different curves give the transmitting frequencies.

The effective network estimated in the previous section allow us to consider only the most relevant connections inside the system by considering only the connected pair of nodes in this effective network. Before applying the multichannel Granger causality test, we first normalize all time series by dividing each observations by the standard deviation of its time series which enables us to compare properly the error terms of every time series. The window sizes tested in the model ranges from 20 to 500 observations with a stepsize of 20 observations in order to scan a large set of possible windows sizes. The MCGC is based on the Granger-Sargent bootstrapped methodology presented in Sect.3.2. Regarding the lag at which the causal relationships are estimated, we have tested for each pair of frequencies, several lags by applying, on the wavelet transforms, a simple Granger causality test for the entire sample. As in Alzahrani et al. (2014), the optimal lags for the different pairs of frequencies, i.e. the ones minimizing the p-value, are very close to each other. We select therefore a lag of one period for every pairs of frequencies as it represents the most frequent lags. However our approach being very flexible, a specific lag could be selected for each pair of frequencies/channels. The results of the MCGC provides three different information for each pair of institutions, the p-value of the Granger-Sargent bootstrapped causality test, the selected optimal rolling window size and a temporal adjacency matrix giving for each time step, the selected transmitting and receiving frequencies. We then aggregate, for the entire network, the information regarding the different frequencies by summing the adjacency

matrices of all connected pair of nodes inside the network. This gives us, for each time step, a global repartition of the channels used for the information transfer.

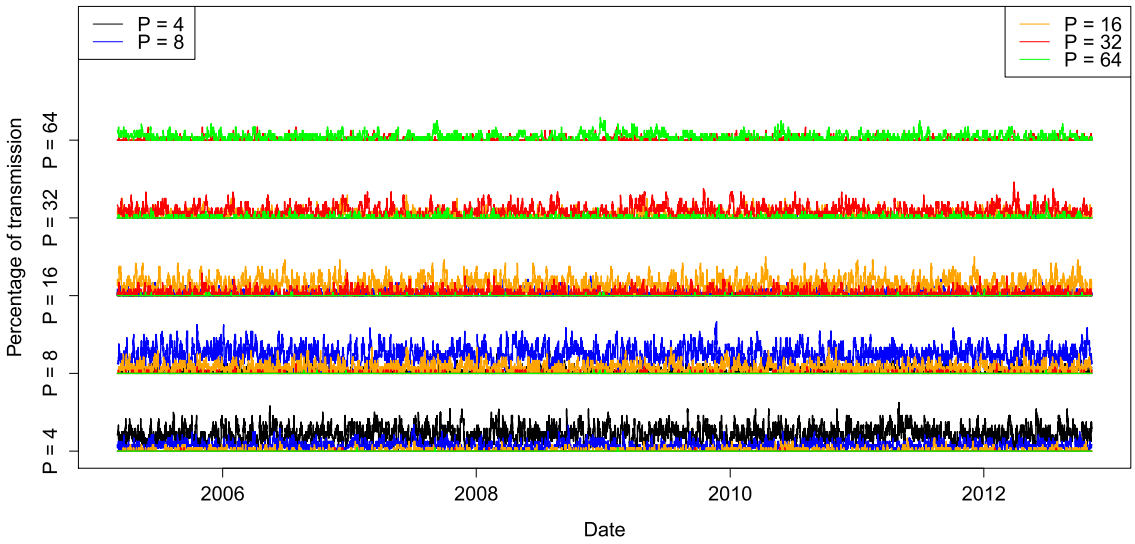


Figure 3.8: This figure gives the results of the MCGC for the dataset using hourly stock returns. It provides for each channels, represented by a specific pair of frequencies, the percentage of information traveling inside the network which passes through them. The y-axis reports the receiving frequencies while the different curves give the transmitting frequencies.

The results are presented in Fig.3.7 and 3.8. As can be seen from both figures, the transfer of information inside the network occurs mainly at the higher frequencies with the wavelength of 4 and 8 observations showing the highest values. They represent together and for both datasets, around 65 percent of the total information circulating inside the network. We observe, for the daily dataset, a clear increase of importance of the highest frequencies in the beginning and during the financial crisis which may be explained by the higher volatility in the market. Both the amount of information considered as relevant and the speed at which these information are treated increase during period of turmoil. The mix of uncertainty and overreaction of the actors inside the market increase therefore the frequency of significant price changes leading to a higher information content of these highest frequencies. We find also that for the two datasets, the cross-frequency transfers occur mainly in the lowest wavelength with a clear information transfer between the wavelength of 8 observations to the one of 4 observations. When looking at the results of the significance test, 78 percent of the connected pair of nodes inside the effective network are significant when considering the MCGC test for the daily dataset and 92 percent for the hourly dataset. The results presented in Fig.3.7 and 3.8 include only the pairs of connected nodes for which a

causal link have been detected with the MCGC at a significance level of 0.05. Finally, we get an average optimal window size of 150 observations for the daily frequency and 29 for the hourly frequency showing that the switching are relatively sparse especially for the daily dataset when considering its total length.

3.5.3 Quantile mutli-channel causality detection

After having documented the information transfer between nodes in terms of frequency, we now look more closely to the probability distribution. We start by adapting the quantile regression framework of Koenker and Bassett (1978) to the multi-channel Granger causality test by replacing the error terms resulting from the usual ordinary least square approach by the ones estimated using quantile regression. As different quantiles may be used for the estimation of the β parameter, we consider each quantile as a specific channel. This framework allows only co-quantile transfers, i.e. transfer of information between two time series at the same quantile. The cross-quantile transfer will be further discussed in the next section when using the multi-channel transfer entropy. As for the analysis in the frequency domain, we apply for each data-set, the adapted MCGC on the pair of connected nodes taken from the inferred network. Five different quantiles are used in the estimation, the 0.5 quantile giving the usual linear regression framework, the 0.3 and 0.7 quantiles to look at the upper part of the distribution for respectively the negative and positive variations and the 0.1 and 0.9 quantiles assessing the extreme variations. The resulting adjacency matrices estimated for each pair of connected nodes are then used to compute the relative importance of each channel in the information transfer process occurring inside the network.

As can be seen from Fig.3.9 and 3.10, the information transfers occur mainly at the tails of the probability distribution, showing that the market reacts more to extreme variations. Indeed, these two channels represent for the daily and the hourly dataset respectively 63 and 73 percent of the total information circulating in the network. Surprisingly, we observed for the daily frequency dataset that during the financial crisis, the percentage of information transfer decreases for the lowest quantile while it increases for the highest one. This means that during such periods, the market pays more attention to positive variations in the environment of financial institutions than to negative ones which contrasts with the common idea that negative variations spread more quickly to the surrounding environment than positive ones. As for the time evolution of the relative importance of each channel when using the hourly frequency dataset, we see that the two tails evolve in opposite phase during the crisis meaning that investors pay alternatively more attention on positive and negative extreme variations. With regards the results of the significance test for the daily and hourly

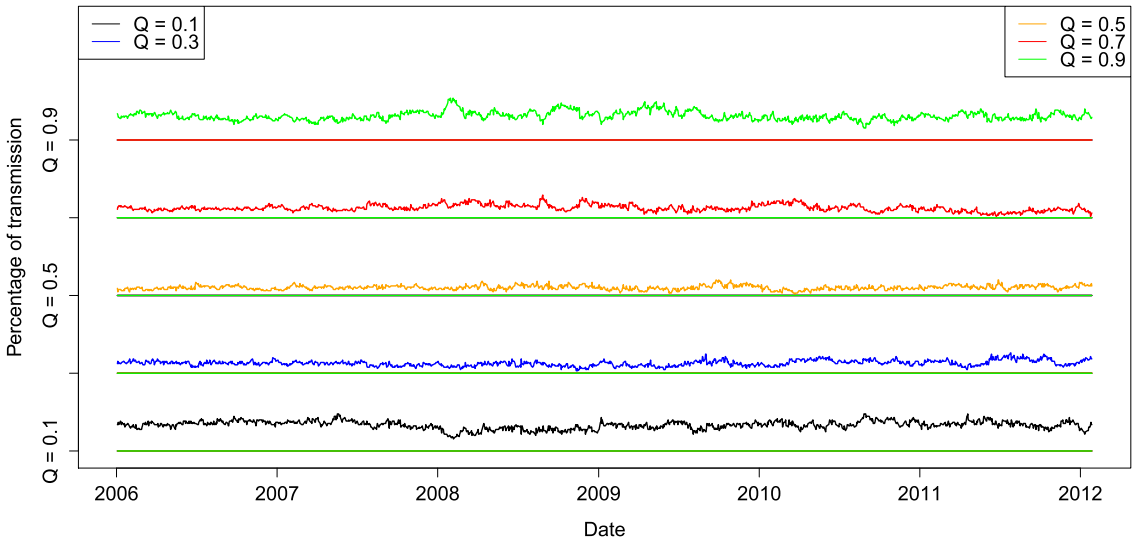


Figure 3.9: This figure gives the results of the MCGC for the dataset using daily stock returns. It provides for each channels, represented by a specific quantile, the percentage of information traveling inside the network which passes through them. The y-axis reports the different quantiles.

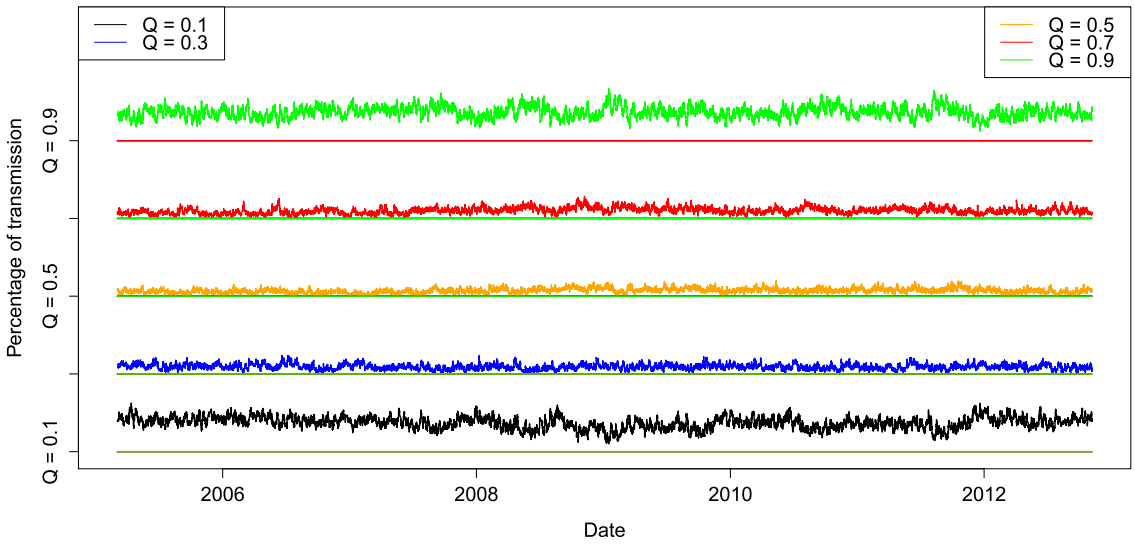


Figure 3.10: This figure gives the results of the MCGC for the dataset using hourly stock returns. It provides for each channels, represented by a specific quantile, the percentage of information traveling inside the network which passes through them. The y-axis reports the different quantiles.

frequency, we have respectively 94 percent and 79 percent of the pairs connected inside the network for which a multi-channel causal link may be found at a significance level of 0.05. In contrast to the frequency analysis, the average optimal window sizes give,

a lower size for the daily dataset with 84 observations and a higher one for the hourly dataset with 150 observations. Nevertheless, when considering the frequency of both datasets, the switching frequency for the daily dataset is still low.

3.5.4 Cross-quantile mutli-channel causality detection

With the objective of extending the previous quantile analysis to cross-quantile information transfer, we propose in this section to use the transfer entropy based multichannel causality measure. Indeed, thanks to the symbolization process used to estimate the transfer entropy, we are able to consider information transfer between times series symbolized using different encoding rules. We may therefore use the spectral quantile symbolization method to look at the interaction between time series with a different set of quantiles for the transmitter and the receiver. This method of symbolization uses two parameters, a position and a bandwidth. The position corresponds to the beginning of the regions of interest on either side of the 50 percent quantile while the bandwidth represents the size of these regions of interest. Both regions get a specific symbol while the remaining parts of the probability distribution share a third symbol. The different pairs of position and bandwidth assigned to each couple of time series represent then a specific channel of communication between the two systems. We have selected 4 different sets of position and bandwidth in order to consider the different parts of the probability distribution with respectively a position and bandwidth of: 0.05-0.45, 0.2-0.3, 0.35-0.15 and 0.45-0.5. These different sets help us to look at both extreme and average relationships and with these 4 sets we get 16 potential channels. We start the analysis by inferring the effective network using the same methodology as before but taking transfer entropy as causality measure instead of Granger causality. Once estimated, the MCTE is applied on every connected pair of nodes inside the effective network. Then, as done in the previous section, we estimate the relative importance of each channel for the different time steps by averaging on all connected pairs inside the networks.

Looking at Fig.3.11, we see clearly the impact of the financial crisis on the channels dynamic inside the financial sector. Although the extreme variation seems to have little impact on the information transfer process before and after the financial crisis, they are clearly leading this process during the period of turmoil between 2008-2010. We see indeed for both the receiver and the transmitter a high percentage of transfer between the two highest quantiles, i.e. 0.35-0.15 and 0.45-0.05, during this period. The reverse is true for the average relationship represented by the two other quantiles, with a leading role during the periods of stability and a decreasing influence during the financial crisis, with even a total disappearance of the channels with parameters 0.05-0.45 for the receiver. This evolution can be explained by the combination of the

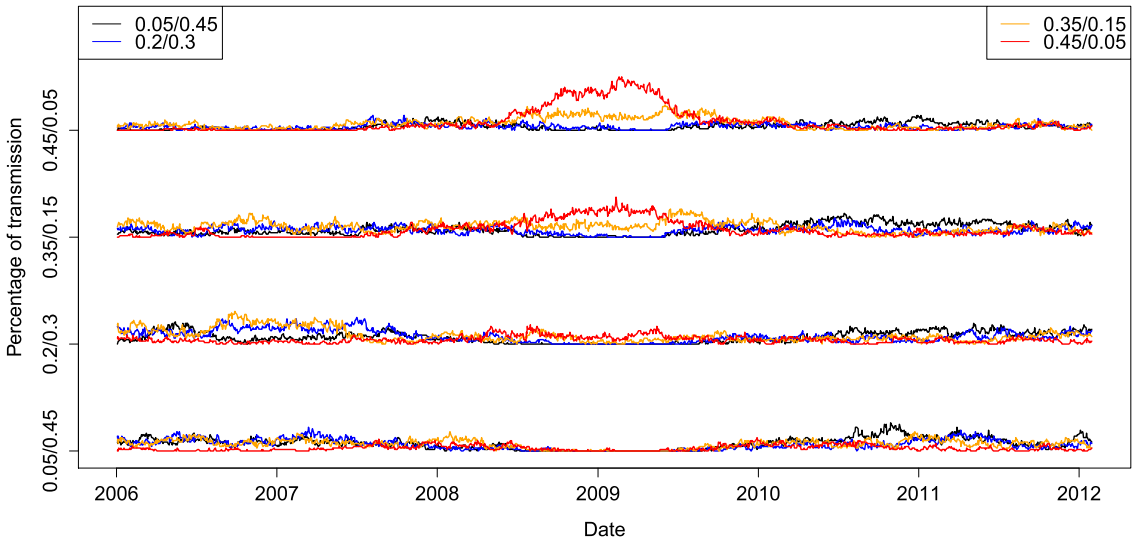


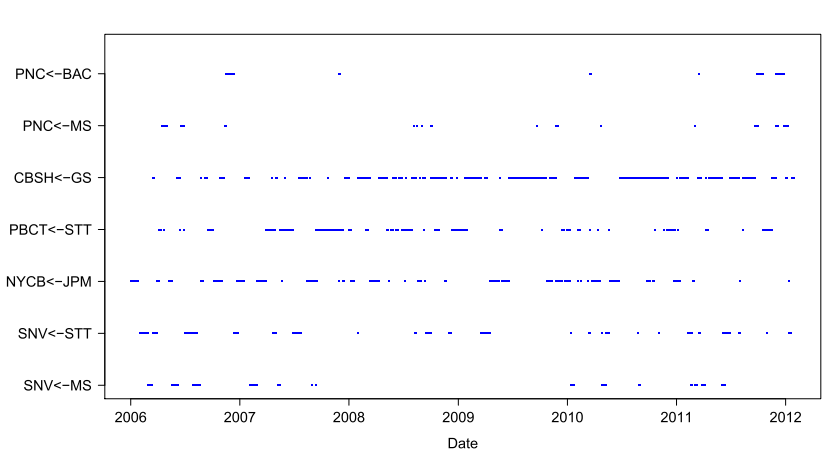
Figure 3.11: This figure gives the results of the MCTE for the dataset using daily stock returns. It provides for each channels, represented by two positions and bandwidth, one for each node, the percentage of information traveling inside the network which passes through them. The y-axis reports the different positions and bandwidth of the receiver while the different curves give these parameters for the transmitter.

higher volatility during period of crisis which leads to an increasing number of extreme variations, and the fact that investor are more concerned by extreme variations during such periods. As for the cross-quantile transfers, we identify a clear evolution between the periods before and after the crisis, with the channels with parameters 0.05-0.45 for the transmitter leading after the crisis while the channels with parameters 0.2-0.3 and 0.35-0.15 were leading before. This means that the investors are looking more broadly at the environment of the financial institutions in term of probability distribution and not only at the highest variations. When looking at the value of the effective transfer entropy, we see that 80 percent of the connected pair of nodes, selected during the network inference process, are still significant for our multichannel analysis. Finally the average window size is 115 observations demonstrating a relatively low switching frequency considering the 2018 observations of our daily dataset.

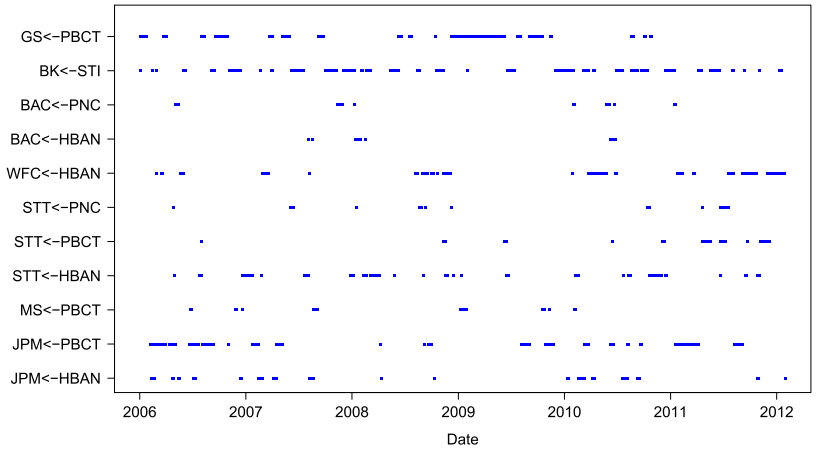
3.5.5 Mutli-channel causality detection

The objective of this last section is to look at the US financial system from a macro perspective by considering groups of financial institutions instead of individual pair of nodes as done previously. We divide our dataset in two subgroups and analyze how these groups interact based on the identification of the main channels through which

the information transits from one group to the other. The first group consists in the financial institutions included in our data-set which are considered as systematically important for the global financial system, while the second group regroups all the other institutions. The connected pairs forming the effective network inferred in Sect.3.5.1 are used as a base to determine the possible channels of communication between the two groups. We estimate then the MCGC both from the SIFI group to peripheral financial institutions and from the peripheral financial institutions to the SIFI group. This estimation is based once again on normalized time series.



(a)SIFI group as source



(b)Non-SIFI group as source

Figure 3.12: The figures (a) and (b) shows, for the daily frequency dataset, the pair of nodes leading the information transfer from the SIFI group to the rest of the network for (a) and reversely for (b). The figures report only the pair of node which have led the information transfer for at least one time step.

Fig.3.12 and 3.13 provide, for each time step, the pair of financial institution leading the information transfer between the two systems for respectively the daily and hourly frequency. We note that the transfer from the SIFIs to the rest of the network seems more stable in time than the transfer of information in the reverse direction, with longer periods of transfer for the different channels. When looking at the financial crisis period, we observe that Goldman Sachs and Barclays lead the information transfer process to the peripheral financial institutions for both data-sets, along with JP Morgan Chase for the daily frequency. Overall Goldman Sachs and Barclays seems to be the main source of information for the peripheral system, leading the information transfer process 36.1 and 28.6 percent of the time for the daily frequency and 21.1 and 19.3 percent of the time for the hourly data-set. We also see from these graphs that some financial institutions have lost their importance during and after the financial crisis such as Citigroup for the hourly dataset and Morgan Stanley for both datasets. As for the reverse causal relationship from non-SIFI to SIFIs, we find that the sources of information are more fragmented for the SIFIs, meaning that this group looks more broadly at the evolution of their environment. With regards to the significance test, we see a clear causal link in both directions for the two datasets. Looking at the high value of the significance test, the channels representation should be close to the true underlying one.

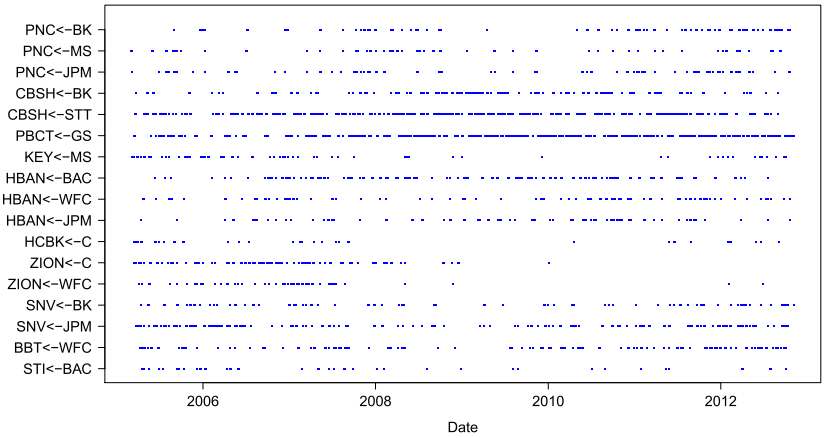
3.6 Conclusion

We proposed in this chapter a new method to estimate the causal relationships existing between complex systems described by several parameters. Additionally to the detection of causal links, the proposed approach gives insights on the underlying dynamics of the information transfer process by defining at each time step through which channel the flux of information is the highest. We developed two different multi-channel causality tests based on different definition of causality, the multi-channel Granger Causality test which accounts mainly for linear causal relationships and the multi-channel transfer entropy test which looks at both linear and non linear information transfer. Both tests are easy to implement and can be used for several purposes, from the analysis of macro systems to the estimation, at a lower level, of cross-spectral and cross-quantile information transfers between individual series.

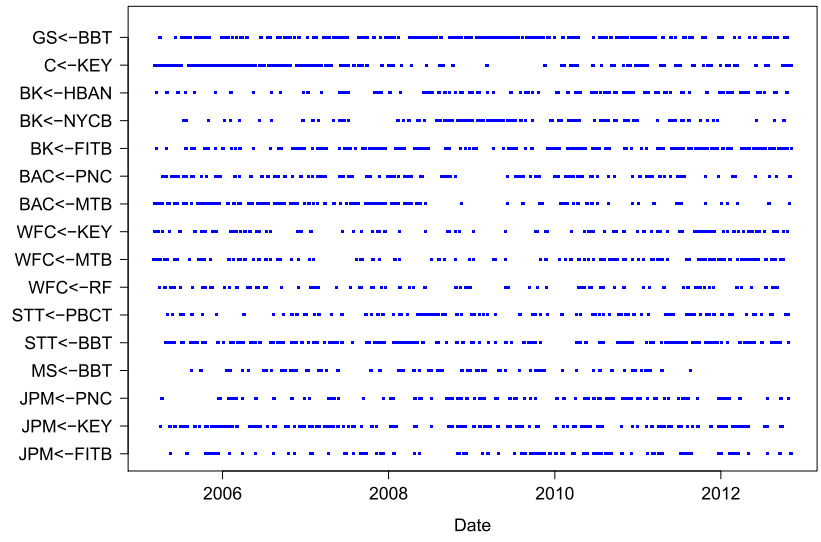
We performed thereafter a Monte Carlo simulation exercise, assessing how our models react to different configurations of coupled systems, looking at the effect of the switching frequency, the strength of the relationships, the type of relationships and the complexity of the channels dynamic involved in the information transfer. We

demonstrated the ability of our models to infer properly this channels dynamic, especially for low switching frequencies which are often observed in the case of financial time series as proven by the empirical application. Both models show equivalent results when the underlying dynamic is linear, while the transfer entropy based measure shows better results for the non linear dynamic.

We eventually proposed an empirical application to illustrate the usefulness of the new framework; looking at the relationships existing between individual financial institutions inside the U.S. financial sector as well as between parts of the financial sector, focusing our attention mainly on the systematically important financial institutions. The results of the multi-channel Granger causality test in the frequency domain shows that the transfer of information occurs mainly at high frequency especially during the financial crisis of 2008. As for the quantile and cross quantile analysis, it provided a clear insight on the importance of the extreme variations during the financial crisis confirming the importance of the tails in periods of turmoil. We finally demonstrated the interest of the methods when looking at groups of financial institutions to analyze the channels dynamic during period of stress or to determine to which extent the sources of information for a specific group were concentrated and stable in time.



(a)SIFI group as source



(b)Non-SIFI group as source

Figure 3.13: The figures (a) and (b) shows, for the hourly frequency dataset, the pair of nodes leading the information transfer from the SIFI group to the rest of the network for (a) and reversely for (b). The figures report only the pair of node which have led the information transfer for at least one time step.

Chapter 4

Cross-country Information Transmissions

Introduction

Throughout the last decade, new methodologies and approaches have been developed to better understand the interconnections existing between parts of the global economic system, especially after the last financial crisis. Indeed, the greater interdependence of markets across the globe explained by the financial globalization, and the financial crisis of 2008 highlighted the necessity for a rethinking of economic and financial policies which requires a better understanding of the mechanisms governing information transmission inside the economy. This transmission mechanism was investigated from a number of perspectives, relying mainly on causality analysis, for different markets and asset classes such as the equity market Longin and Solnik (1995); Hong et al. (2009b); Celik (2012); Dungey and Gajurel (2014), the financial market Diebold and Yilmaz (2014); Billio et al. (2012), the sovereign bond market Longstaff et al. (2011); Gorea and Radev (2014); Fernández-Rodrígueza et al. (2016) or the commodities Smiech et al. (2015); Bhar and Hammoudeh (2011); Hegerty (2016); Zhang et al. (2016).

The literature related to interrelations focuses primarily on the detection of stress transmissions often identified as contagion. Contagion may be defined as cross-country transmissions of shocks or more generally as cross-country spillover effects. As recalled by Dungey and Gajurel (2014), the current literature identifies two possible theories of contagion. The first theory defines the interdependence of economies through real and financial linkages such as flows of goods, service and capital, as the main carrier of information. The second strand of the literature argues that stress transmission occurs from one country to another due to the lack of anticipation of investors having incomplete information. Indeed a shock in one country may trigger a reassessment of the risks in other countries by international investors Pasquariello (2006); Yuan (2005).

This risk reassessment also called "Wake up call" in Bekaert et al. (2014) increases the relative importance of domestic fundamentals in the transmission of stress during period of turmoil.

We propose in this chapter to consider a more broad connectedness measure that does not only consider shocks in order to infer the causal relationships that we identify as information transmissions. We rely on a new framework inspired by the concept of causality developed by Granger (1969) which has already been extensively used to study the interrelations existing inside the economic system (see Ratanapakorn and Sharma (2007); Pradhan et al. (2014); Jammazi et al. (2017) among others). However, in contrast with the current literature which often considers one type of asset or one market at a time, we propose to consider several types of asset at the same time to estimate more globally the transmission of information between countries. Each country is therefore characterized by a set of financial variables on which a novel multichannel causality test is applied.

In our multichannel framework, a channel is defined as a pair of parameters characterizing the state of two different systems. Each system being described in terms of several parameters, different channels may possibly convey causal relationships over time. In the proposed application, the fundamentals represent the different parameters while the considered systems are countries. The set of selected fundamentals has been therefore chosen so as to proxy the state of each country economy. The objective of the chapter is therefore not the identification of contagion periods but rather the analysis of the evolution of the channel through which the information is transmitted depending on the period considered: crisis period vs. non crisis period. We consider thus not only stress transmission but more broadly information transmission, i.e. inter-dependencies affecting either positively or negatively countries, and analyze how the channels through which the information transits evolve in period of turmoil where contagion may be observed.

To empirically assess the relative contribution of several transmission channels over time, our econometric framework should accommodate two important features: (i) multivariate dimension and (ii) a time varying dimension. Inspired by the work of Barrett and Barnett (2010); Geweke (1982) who adapted the univariate Granger causality test to the multivariate case, we propose an extension of the Markov regime switching Granger causality test developed by Psaradakis et al. (2005b) to account for information transmission between distinct multivariate systems. In our model, each regime corresponds to a specific pair of variables including one of the financial variables describing the country receiving the information, i.e. the dependent variable and one of the financial variables describing the country transmitting the information, i.e. the

explanatory variable. The proposed framework enables us to go beyond the usual estimation of cross-country inter-dependencies by considering multiple explanatory and dependent variables describing two different systems. This method allows us to define both the existence of a causal link and the main channels through which the information transit at each time step.

Our empirical investigation starts by the analysis of cross-country information channels. We consider five different measures of economic wealth for each country: an equity index, an index representing the main financial institutions of the country, the sovereign debt returns, the exchange rate of the local currency and a measure of the volatility inside the equity market. These indicators regroup most of the financial variables extensively used in the literature to treat inter-dependencies between countries. Aside from the identification of cross-country connectedness, the aim of the chapter is also to consider the interactions existing between countries and the commodity markets and determine if commodities could be considered as an additional indirect channel used to transmit information from one country to another.

Overall, the chapter contribution to the literature is mainly threefold. First, we propose a multichannel approach which is able to define the existence of causal relationships between distinct complex systems such as countries which could be characterized by several features (e.g. equity indices, sovereign bonds). This approach provides also a clear view on the dominating channels of transmission over time. Secondly, we propose an empirical application of this model to assess the evolution of the relationships existing between countries forming the world major economies. Different financial variables are considered to identify potentially different channels of information transmission, focusing our attention on the global financial crisis and on the European sovereign debt crisis. Third, we include in our investigation the effect of the commodity markets as it plays a key role in both the real economy and the financial system.

The chapter is structured in the following manner. The second section presents the multichannel causality measure deriving for the Markov switching setting. In the third section, we present the data-set used in the empirical study, outline the methodology and present the results for the model including only the countries and the model including the additional effect of the commodity market. We finally concludes on the main contributions of the chapter in the last section.

4.1 MultiChannel Markov Switching Granger Causality

The concept of Markov-switching regressions was first proposed in econometrics by Goldfeld and Quandt (1973) and then Cosslett and Lee (1985) who introduced the estimation via a likelihood function. The formulation we use in this chapter derives directly from the work of Hamilton (1988, 1994) who developed an iterative inference algorithm, namely the Markov switching model. This approach is one of the most popular nonlinear time series models in the literature and many variants have been proposed ever since. The model, we develop in this chapter, is and extension of a Vector Autoregressive (VAR) variant of Hamilton's model proposed by Krolzig (1997) and then extended by Psaradakis et al. (2005b) for causality detection using the approach of Granger (1969). The proposed approach being based on the concept of causality developed by Granger (1969), we draw away from the "a priori structural" approach (e.g. instrumental variables) that have been used extensively in the microeconometrics literature (see Hoover (2004) and Hoover (2008) for surveys). Indeed Granger's approach may be qualified as an "inferential process" approach, as it relies directly on the data ("inferential") and on the asymmetry of causality stressed by the condition of Hume "The cause must be prior to the effect" ("process")Morgan (1991) to estimate the causal relationships. However, the interest for Granger based approaches in economics and finance has increased in recent years Ratanapakorn and Sharma (2007); Billio et al. (2012); Pradhan et al. (2014); Jammazi et al. (2017).

We consider, in our model, two systems $\boldsymbol{X}$ and $\boldsymbol{Y}$ characterized each by a set of time dependent variables $\boldsymbol{X} = \{x_t^1, \cdots, x_t^N\} \in \mathbb{R}^{N \times T}$, with N defining the number of parameters characterizing the system $\boldsymbol{X}$, and $\boldsymbol{Y} = \{y_t^1, \cdots, y_t^M\} \in \mathbb{R}^{M \times T}$, with M defining the number of parameters characterizing the system $\boldsymbol{Y}$. The objective of the framework, we developed hereafter, is the detection of a possible transmission of information from system $\boldsymbol{Y}$ to system $\boldsymbol{X}$ and the definition of the favored channel for each time step. The channels are defined as pairs of time series composed of one time dependent variable of system $\boldsymbol{X}$ and another of system $\boldsymbol{Y}$. If a causal link may be inferred between both variables, the channel is considered as active in the information transfer process. The channels are, nevertheless, not mutually exclusive as for each time step, a probability of activity is provided for each channels.

We start by defining the relationship between our two systems thanks to a multivariate VAR model taking into account both the effects of the past of $\boldsymbol{X}$ and $\boldsymbol{Y}$ on the current value of $\boldsymbol{X}$. We see in Eq.4.1 that both the dependent and explanatory variable are multivariate.

$$X_t = \boldsymbol{\alpha} + \sum_{l=1}^{L} \boldsymbol{\beta}_l X_{t-l} + \boldsymbol{R}_{l,t} \boldsymbol{\Phi} Y_{t-l} + \boldsymbol{\varepsilon}_t \tag{4.1}$$

Here $\boldsymbol{\Phi}$ represents the coefficient characterizing the transfer of information from the system $\boldsymbol{Y}$ to the system $\boldsymbol{X}$ and is defined by:

$$\boldsymbol{\Phi} = \begin{pmatrix} \phi^{1,1} & \phi^{1,2} & \dots & \phi^{1,M} \\ \phi^{2,1} & \phi^{2,2} & \dots & \phi^{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ \phi^{N,1} & \phi^{N,2} & \dots & \phi^{N,M} \end{pmatrix} \tag{4.2}$$

with $\phi^{i,j}$ the coefficient characterizing the relationship between the j^{th} parameter of the system $\boldsymbol{Y}$ and the i^{th} parameter of the system $\boldsymbol{X}$ with a lag of l.

As for the parameter $\boldsymbol{R}_t$, it provides the information about the active channel at each time step and is given by:

$$\boldsymbol{R}_t = \begin{pmatrix} r_t^{1,1} & r_t^{1,2} & \dots & r_t^{1,M} \\ r_t^{2,1} & r_t^{2,2} & \dots & r_t^{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ r_t^{N,1} & r_t^{N,2} & \dots & r_t^{N,M} \end{pmatrix} \tag{4.3}$$

Here, all $r_t^{i,j}$ are latent random variables that reflect the regime of the system for every time step t. Each variable $r_t^{i,j}$ takes its value in the set of $\{0, 1\}$ which implies an information transmission from the variable j of system $\boldsymbol{Y}$ to the variable i of system $\boldsymbol{X}$ for $r_t^{i,j} = 1$ and no transmission for $r_t^{i,j} = 0$. Considering the number of parameters for both systems, there exist $N \times M$ possible states which are represented by a new parameter S_t:

$$S_t = \begin{cases} 1 & \text{if } r_t^{1,1} = 1 \\ 2 & \text{if } r_t^{1,2} = 1 \\ \cdots & \\ M+1 & \text{if } r_t^{2,1} = 1 \\ \cdots & \\ N \times M & \text{if } r_t^{N,M} = 1 \end{cases} \tag{4.4}$$

Assuming that $\boldsymbol{R}_t$ is the realization of a two-state Markov chain, we get $(N \times M)^2$ possible states of nature and the probability for each state of nature is given by:

$$Pr(S_t = r \mid S_{t-1} = q) = p_{q,r} \tag{4.5}$$

with $q, r \in \{1, \cdots, N \times M\}$.

In this framework, the probability of a regime switch depends only on the value of the most recent regime which involves a short term memory, but other specifications may be applied to take into account a longer period of the systems past. Following the definition of S_t, we create two new vectors representing the selected variables in the systems $\boldsymbol{X}$ and $\boldsymbol{Y}$ in each state of nature S_t, $X_t^{S_t}$ and ,$X_t^{S_t}$.

We do not observe S_t directly but only its effect on the behavior of X_t knowing X_{t-1} and Y_{t-1}. The probability density governing X_t can be fully described by determining the set of parameters in Eq.4.1, i.e. the intercept $\boldsymbol{\alpha}$, the auto-regressive coefficient $\boldsymbol{\beta}_l$, the causal parameters $\boldsymbol{\Phi}_l$, the transition probabilities $p_{q,r}$ and the variance of the Gaussian white noise $\boldsymbol{\varepsilon}_t$. Based on this probability law, we can infer the probability of observing each state of nature S_t for every time step using an iterative algorithm. We start by rewriting the probability of observing $\boldsymbol{x}_t$ in the state r in time step t as:

$$\xi_{r,t} = Pr(S_t = r | \Omega_t, \boldsymbol{\theta}) \tag{4.6}$$

Here Ω_t is the set of information available to describe X_t^r , we get therefore $\Omega_t = \{X_{t-1}, \cdots, X_{t-l}, Y_{t-l}\}$. As for $\boldsymbol{\theta}$, it regroups all the parameters of the model $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}_l, \boldsymbol{\Phi}, p_{q,r}, \sigma\}$. With N variables for the system $\boldsymbol{X}$ and M parameters for the system $\boldsymbol{Y}$, we get a total number of $(2 \times N + N \times L + N \times M + (N \times M)^2 + 1)$ parameters to be estimated. The sum of the $\xi_{r,t}$ for the $N \times M$ possible regimes equals unity as they represent all the possible outcomes. We then follow the approach first proposed by Hamilton (1988) and apply an iterative procedure to infer the $\xi_{r,t}$ for all the time

step t. Each iteration involves the estimation of a likelihood function which represents the weighted probability of observing the different possible outcomes considering the set of estimated parameters and the past realizations of the system. The likelihood function is defined as a conditional density of X_t in the following manner:

$$f(X_t|\Omega_{t-1}, \boldsymbol{\theta}) = \sum_{q}^{N\times M} \sum_{r}^{N\times M} p_{q,r}\eta_{r,t}\xi_{q,t-1} \tag{4.7}$$

where $\eta_{r,t}$ gives the Gaussian density of X_t for the $N \times M$ different regimes:

$$\eta_{r,t} = f(X_t|S_t = r, \Omega_{t-1}, \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma} exp\left[-\frac{X_t^r - \sum_{l=1}^{L} \boldsymbol{\beta}_l X_{t-l}^r - \boldsymbol{\phi}^r Y_{t-l}^r - \alpha}{2\sigma^2}\right] \tag{4.8}$$

with α acting as the mean of the probability density.

We are then able to estimate the probability of observing the system in the state r at time t, based on the likelihood function, the transition probabilities $p_{q,r}$, the probability of observing the system in the different possible state q at previous time step t-1, and the Gaussian density of X_t for the different regimes. As can be seen from Eq.4.9, the probability $\xi_{r,t}$ represent simply the sum of the probability of observing the final state r considering all the possible initial state q at time t-1, weighted by their likelihood $\eta_{r,t}$ and then divided by the total weighted probability of observing every possible final states (the likelihood function).

$$\xi_{r,t} = \frac{\sum_{q}^{N\times M} p_{q,r}\eta_{r,t}\xi_{q,t-1}}{f(X_t|\Omega_{t-1}, \boldsymbol{\theta})} \tag{4.9}$$

By performing this iteration from t=1 to t=T, with T the length of the time series characterizing the two systems, we get the following global log-likelihood function:

$$log\left[f(X_1, X_2, \cdots, X_T|X_0)\right] = log\left[\sum_{t=1}^{T} f(X_t|\Omega_{t-1}, \boldsymbol{\theta})\right] \tag{4.10}$$

We then maximize the resulting log-likelihood function to obtain the optimal set of parameters $\boldsymbol{\theta}$, i.e. the one that minimizes the most the error terms of Eq.4.1 weighted by the probabilities of the states of nature S_t instead of the latent factor $\boldsymbol{R}_t$. The maximization of the resulting log-likelihood function is done via a quasi-Newton optimization algorithm based on the Broyden-Fletcher-Goldfarb-Shano secant to update the Hessian but other optimization algorithm may be considered. The value of $\xi_{r,t}$

is then inferred for each time step t using this set of optimal parameters. The statistical inference of a causal link between the system $\boldsymbol{Y}$ and $\boldsymbol{X}$ is finally performed via the F-stat of the estimated $\boldsymbol{\phi}$ which provides information about the existence of a transmission of information for each considered channel. In contrast with Barrett and Barnett (2010), we do not get a global statistic of the existence of a causal link between the two systems but one value per channel which could be considered as more precise. In addition, we are able to define at each time step the probability of a transfer through every possible channels. A multivariate extension of the causal time varying VAR approach of Christopoulos and León-Ledesma (2008) could also be considered as an alternative to this procedure. Nevertheless, our approach is model free and is able to treat more complex transition patterns between regimes compared to the single smooth transition considered in Christopoulos and León-Ledesma (2008). It provides also, in a single estimation, information about the existence of a causal relationship for every channels and their relative activity in terms of information transmission for each time step.

As can be seen from Eq.4.10 and 4.9, we have to initialize the algorithm by defining $\xi_{q,0}$ for the step $t = 1$. If we assume that the considered Markov chain is ergodic, we can simply set $\xi_{q,0} = Pr(S_t = q \mid \Omega_0, \boldsymbol{\theta})$ equal to the unconditional probability $\xi_{q,0} = Pr(S_t = q)$. Following the approach proposed by Hamilton (1994), theses initial probabilities may be estimated from the $(N \times M + 1)$th row of the matrix $(A^t A)^{-1} A^t$ with A defined as:

$$A = \begin{pmatrix} I_{N\times M} - P \\ 1' \end{pmatrix} \tag{4.11}$$

with P the matrix of $p_{q,r}$ and $I_{N\times M}$ a diagonal matrix of dimension $N \times M$.

4.2 Results and Discussion

4.2.1 Cross-country information transmissions

As recalled by the numerous contributions related to risk transmission (see Longin and Solnik (1995); Hong et al. (2009b); Diebold and Yilmaz (2009); Billio et al. (2012); Longstaff et al. (2011); Gorea and Radev (2014) to quote only a few) several channels of information transmission exist between countries, such as their stock market, financial institutions, sovereign bond yield, exchange rate or interest rate among others. We have selected, in this chapter, five different high frequency, i.e. daily, variables reflecting the overall state of a country: the main stock market index, an index regrouping

the main financial institutions of the country, a volatility index, the exchange rate and the 5 years USD denominated sovereign bond yield. Most of the selected financial variables have already been extensively used in the literature to treat interdependencies between countries.

Related studies are Longin and Solnik (1995) or Diebold and Yilmaz (2009) which find empirical evidences of stress transmission between international equity markets or Celik (2012) which finds evidence of contagion for most of the developed and emerging countries during the 2008 financial crisis. They demonstrate the ability of stock markets to reflect the information transmission between countries. This ability to represent cross-country relationships comes primarily from the fact that stock markets reflect the evolution of a country production capacity both in terms of goods and services and could therefore assess to some extend the evolution of the country balance of payment as the exchange rate does also. With regards the exchange rate, we also included it in the set of parameters for its close link with the commodity markets and it will prove to be an important driver of the interdependencies between countries and commodity markets.

System wide dependencies have also been analyzed for financial institutions. Hence, a growing body of the current literature on contagion and network theory tries to understand the main characteristics of the evolution of the connectivity inside the financial market during the crisis of 2008 with contribution such as Billio et al. (2012); Diebold and Yilmaz (2014); Minoiu and Reyes (2013); Tonzer (2015). Indeed, the financial crisis highlighted the propensity of financial institutions risk to spill over to their environment and across border. This high level of interdependency may be explained partly by the contractual linkages existing between this type of firms which increase the counterparty risk. In addition to these spillover effects, the interplay between fiscally strained sovereigns and stressed banks worsen the condition of the economy. Indeed, strained public finances have limited the ability of some countries to support their financial system through bailout. Fragile banking systems provided then less support to the economic activity, which in turn further strains public finances. These inter-dependencies between banks and sovereigns demonstrate the importance of considering the impact of external factors on a country's banking system. We consider therefore the evolution of the banking system as a relevant variable to be considered in our multichannel analysis.

The European sovereign debt crisis of 2011 is another good illustration of contagion between countries as demonstrated by Longstaff et al. (2011) which conclude that sovereign debt returns are even more correlated than equity returns. This period has been the subject of many studies about risk transmission between European sovereigns

(Kalbaska and Gatkowski (2012); Metiu (2012); Alter and Beyer (2013) to quote a few) and could therefore be a privileged channel during this period of time.

Eventually, the volatility index has been chosen to proxy the risk/stress level of a country in order to assess possible stress spillovers between countries. Overall, the selected financial variables have already proven their ability to describe interdependencies between countries. Nevertheless, the proposed multi-channel approach will challenge this ability by considering together all parameters and provide additional information about the evolution of these parameters relevance.

4.2.2 Data

In order to have an overview of the information transmission inside the global economy, we have selected 9 of the most important economies in the world with the United States and Canada for the American continent, France, the United Kingdom, Italy and Germany for Europe and China, Australia and Japan for the Asia-Pacific region. All the financial variables are expressed in terms of daily variations. Table 1 reports the stock market and the volatility indices selected for the empirical application but also the financial institutions used to estimate the financial indices when no aggregate indices were available. In this later case, the reconstructed financial indices were computed as the weighted average of the daily returns of the country's main financial institutions using the market capitalizations as weights. The volatility indices are based on the implied volatility of the selected stock market indices. As for the exchange rate, having included the U.S. in our analysis, we express all currencies in terms of the IMF Special drawing rights (SDR) instead of the usual USD.

Regarding the data-set used to represent the commodity markets, we have selected 8 commodities covering 4 major sectors, precious metals with gold and silver, metal with aluminum and copper, energy with crude oil and food with soybean, wheat and corn. All the data-sets have been taken from Bloomberg and range from January 2001 to June 2016 representing 4030 daily observations, including the end of the early 2000s recession due to the dot-com bubble, the Global Financial Crisis of 2008 and the European Sovereign Crisis of 2011.

4.2.3 Analysis of Cross-country information transmissions

We now turn to the first step of our empirical investigation by looking at cross-country information transmissions. Each financial variables presented in the previous section represents a possible transmitter or receiver of information which implies 25 possible channels between two countries. In order to reduce the dimension of the model to

	Stock Index	Volatility index	Financial institutions
U.S.	SP500	VIX	S&P500 Bank
Canada	S&P/TSX60	VIXC	Royal Bank of Canada
			Toronto-Dominion Bank
			Bank of Nova Scotia
			Bank of Montreal
U.K.	FTSE100	VFTSE	HSBC
			Standard Chatered
			Lloyd
			Barclays
			Royal Bank of Scotland
Germany	DAX	VDAX	Commerzbank
			Deutsche Bank
France	CAC40	VCAC	BNP Paribas
			Société générale
			Credit Agricole
Italy	MIB30	IVMIB30	Intesa SanPaolo
			Unicredit
			Generali bank
China	SSE50	VXFXI	Bank of China
			China Construction bank
			Industrial and commercial bank of China
			Agricultural Bank of China
Japan	Nikkei	VXJ	Nikkei Bank
Australia	S&P/ASX 200	AS51VIX	S&P/ASX 200 Bank

Table 4.1: This table reports the list of the Stock indices, volatility indices selected for the different countries and the list of banks used for the estimation of the financial indices.

be estimated, we start the estimation by applying a filter which eliminates the most distinct non-causal relationships. This filter consists in a simple Granger causality test which is applied for each pair of countries on the 25 possible channels. The significance level of the causality test is set to 0.01 and if no connections are observed we increase this level to 0.05. The Granger Causality test is applied on the complete time series. Once the active channels have been determined, we apply on it our multichannel causality measure which provides for each channel three different information. The first one is the p-value which gives information about the existence of a causal link for a particular channel. Then it provides, for each time step, Ξ_t which includes the matrix of $\xi_{q,t}$ for every possible channels, i.e. the probability of channels' activity for each time step. Eventually it gives the $p_{r,q}$ matrix representing the probability of all possible regime switches.

Once all these information have been computed for each pair of countries, we first look at the evolution of the global connectivity by regrouping the obtained results for the 9 countries. Fig.4.1 and 4.2 report the evolution of the probability of activity for the 25

channels that are considered in this analysis. The activity for each channel has been averaged over the 9 countries and then normalized to properly compare the results. As can be seen from these figures, the probability of channels' activity seems relatively stable in time. However, we observe for several channels a regime switch occurring mainly during the last financial crisis. The relative importance of the stock markets and financial institutions as transmitter stay stable for the entire data-set while the sovereign bond yield and the currencies become more influential. Surprisingly the volatility indices seem to lose their significance during the crisis period of 2008-2010. Those results highlight the interest of the proposed multi-channel approach which is able to provide new insights regarding the information transmission inside the world economy.

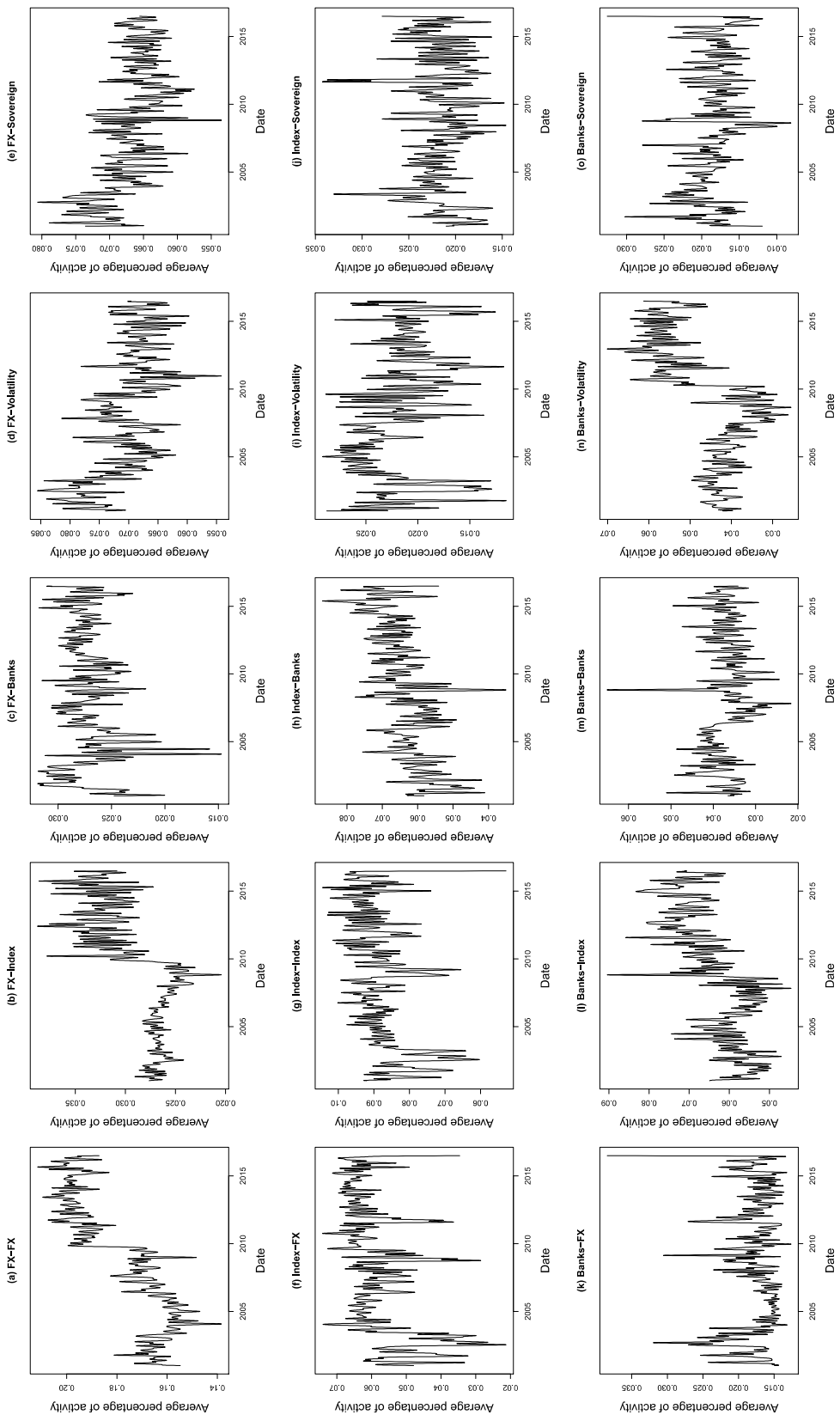


Figure 4.1: These figures report the evolution of the probability of activity for every considered channel. The relative activity has been averaged over the 9 countries of our data-set and then normalized.

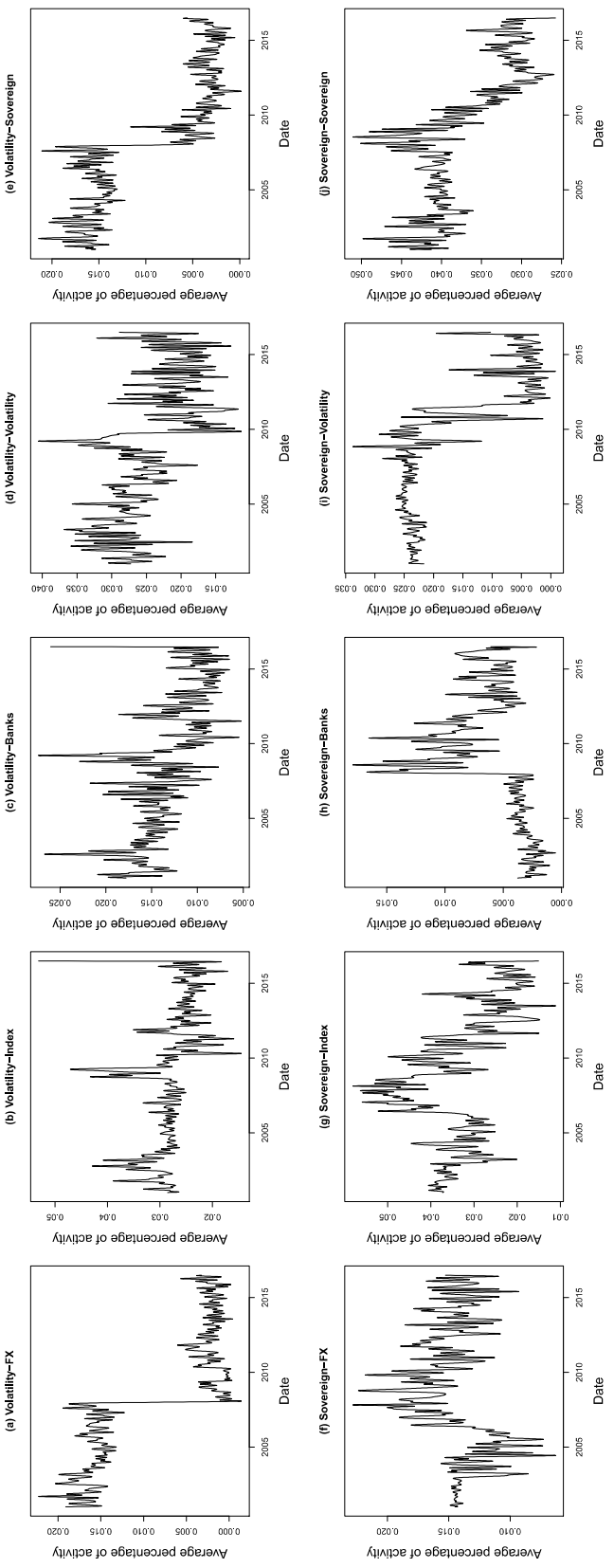


Figure 4.2: These figures report the evolution of the probability of activity for every considered channel. The relative activity has been averaged over the 9 countries of our data-set and then normalized.

Our analysis relying on a large number of channels, we proposed to aggregate the previous results by considering the four main periods of the recent economic history. We therefore divide the data-set into four periods: the pre-crisis period which goes until the start of the Global Financial Crisis on the 9th of August 2007, the Global Financial Crisis period which goes until the 15th of January 2010, the European Sovereign Debt Crisis which goes until the 31st of October 2013 and finally the post crisis period. This will ease the analysis while still allowing us to look at possible changes in the channels dynamics during these different episodes. To get a more comprehensive and global view, we have again summed and normalized the matrices of probability Ξ_t for the 9 countries and averaged them for the considered periods of time.

The results are reported in Table 4.2 and show that overall the main channel of cross-country communication seems to be the currencies, with on average 18 percent of the information traveling through this channel. The transfer of information between stock markets represents the second largest channel with around 9 percent overall. The other main channels are the Volatility to Currency, Sovereign bond yield to Currency and Currency to Stock market, Financial institutions to Stock market and Stock market to Financial institutions with on average 6 percent. We see, moreover, that in contrast with the current literature which focuses its attention mainly on the importance of financial institutions on stress transmission, the main carriers of information are rather the currency and the stock market which are both the main receiver and transmitter of information.

As for the evolution in time of the channels' weight, the results confirms the observation made in Fig.4.1 and 4.2 with an increase of the importance of the currencies and the stock markets, both as transmitters and receivers of information, as from the beginning of the financial crisis. For the other channels, we do not see a clear evolution of the repartition, apart from the volatility, which tends to prove that these channels are relatively stable in time.

After having documented the global repartition of the channels and its evolution in time, we may now look in more details at the repartition for each country of our data-set. We use again the information provided by our multichannel causality measure to compute, for each country, the average probability of activity of every channels in a 5 by 5 matrix by averaging the Ξ_t for the whole period. We then consider each country from two perspectives, first as a receiver of information and then as a transmitter. We select, in both cases, the channels for which the observed p-value is under 0.01. Once normalized the resulting two matrices provide, for every countries in our data-set, an overview of the channels used for both information outflows and inflows.

Entire Period

Global	FX	Index	Banks	Volatility	Sovereign
FX	0,18	0,03	0,03	0,06	0,07
Index	0,06	0,09	0,06	0,02	0,02
Banks	0,02	0,06	0,03	0,05	0,02
Volatility	0,01	0,03	0,01	0,03	0,01
Sovereign	0,01	0,03	0,01	0,02	0,03

Pre-Crisis Period

Pre-C	FX	Index	Banks	Volatility	Sovereign
FX	0,16	0,03	0,03	0,06	0,07
Index	0,06	0,09	0,05	0,02	0,02
Banks	0,01	0,06	0,03	0,04	0,02
Volatility	0,02	0,04	0,02	0,04	0,02
Sovereign	0,01	0,03	-	0,02	0,04

Global Financial Crisis Period

Fin-C	FX	Index	Banks	Volatility	Sovereign
FX	0,18	0,03	0,04	0,06	0,07
Index	0,06	0,09	0,06	0,02	0,02
Banks	0,02	0,06	0,03	0,03	0,02
Volatility	-	0,04	0,01	0,03	0,01
Sovereign	0,01	0,04	0,01	0,02	0,04

European Sovereign Debt Crisis Period

Sov-C	FX	Index	Banks	Volatility	Sovereign
FX	0,19	0,04	0,03	0,06	0,07
Index	0,07	0,10	0,06	0,02	0,02
Banks	0,02	0,07	0,03	0,05	0,02
Volatility	-	0,03	0,01	0,02	-
Sovereign	0,01	0,03	0,01	0,01	0,03

Post-Crisis Period

Post-C	FX	Index	Banks	Volatility	Sovereign
FX	0,19	0,03	0,03	0,06	0,07
Index	0,07	0,10	0,06	0,02	0,02
Banks	0,02	0,07	0,04	0,05	0,02
Volatility	-	0,03	0,01	0,02	-
Sovereign	0,01	0,02	0,01	-	0,03

Table 4.2: This table reports, for 5 different periods of the recent economic history, the probability of activity of every channels of information considered in this application, averaged on all countries.The vertical axis shows the variables of the country receiving the information and the horizontal axis, the variables of the other country from which the information has been transmitting.

Table 4.3 provides the resulting channel's activity index per country, both in terms of inflows and outflows of information. The left part of the table gives the channels through which the information reaches the considered country, with the vertical axis giving the variable of this country, and the horizontal axis, the variable of the aggregated transmitting countries. On the right side of the table, we get the reverse information with the channels through which the information leaves the country, i.e. the variables that impact the most the other countries. The horizontal axis provides now the variables of the country of interest and the vertical axis, the variables of the aggregated receiving countries. As can be seen from theses different tables, the high percentage for the channel Currency-Currency comes mainly from the U.S., Canada, Italy and China for the inflows and from Germany, France, Italy and Japan for the outflows. In contrast with the common idea that the USD is the currency influencing the most the global economy, the results shows on the contrary that the EUR and

the JPY lead the information transmission process to the other currencies such as the USD, the GBP and the CNY.

Looking now at the Stock market indices, we see that the influence of this variable is especially high for the European countries for both information inflows and outflows but also for Japan and Australia for the information inflows. In contrast with the results obtained by Dungey and Gajurel (2014) which demonstrates the importance of the U.S. equity market as a stress transmitter to other equity markets, when looking at the results of our model, the U.S. equity market does not seem to play an important role in the transmission of information as this channel represents only 2 percent.

With regards the financial institutions, our results confirm the importance of the U.K. and U.S. financial sectors as transmitters of risk. The financial sector of the countries from the Asia-Pacific region seems to be the most vulnerable to information coming from other countries especially the Australian one. Australia is also, with the U.S., the country which influenced the most the others with respect to their sovereign bond market. Germany is also a country for which the bond market has a relatively high importance in terms of incoming information flows, especially from the foreign equity market with both the volatility and stock market indices influencing its sovereign bond market. We observed eventually that, except for Italy, countries transmitting a lot of information with their volatility index, are countries for which the volatility index is the least important parameters for information inflows.

After having considered aggregated results, we now turn to a more detailed description by looking at the main channel of communication for each pair of connected countries. As can be seen from Table 4.4, the results confirm the major role of the currencies in the transmission of information to North America while the information transmission inside the European Union is mainly driven by the financial and stock markets both in terms of returns and volatilities. The transmissions from the European Union to Asia involve two main channels, a Currency-Currency channel and a Financial market-Stock market channel, highlighting the importance of the Euro and the European financial industry in Asia. The Asian financial markets and currencies, in turn, tends to influence the European countries. As regards the links between North America and Europe, we see that the European markets are very attentive to North American indicators.

U.S.

IN	FX	Index	Banks	Volatility	Sovereign	OUT	FX	Index	Banks	Volatility	Sovereign
FX	0,48	-	-	0,32	0,11	FX	0,05	0,01	0,10	-	0,11
Index	-	-	-	-	0,02	Index	-	0,02	0,12	0,16	0,12
Banks	-	-	0,02	-	0,05	Banks	-	0,02	0,03	0,02	0,10
Volatility	-	-	-	-	-	Volatility	-	0,01	-	0,03	-
Sovereign	-	-	-	-	-	Sovereign	-	0,08	-	0,02	-

Canada

IN	FX	Index	Banks	Volatility	Sovereign	OUT	FX	Index	Banks	Volatility	Sovereign
FX	0,48	0,07	-	-	0,24	FX	-	-	-	-	0,14
Index	0,08	-	-	-	-	Index	0,20	0,25	-	0,01	0,08
Banks	-	0,02	-	-	-	Banks	0,14	0,01	0,03	0,09	0,02
Volatility	-	-	-	0,10	-	Volatility	-	0,01	0,01	0,01	-
Sovereign	-	-	-	-	-	Sovereign	-	-	-	-	-

U.K.

IN	FX	Index	Banks	Volatility	Sovereign	OUT	FX	Index	Banks	Volatility	Sovereign
FX	0,17	0,20	-	0,14	0,01	FX	0,06	-	-	-	-
Index	0,03	0,17	-	-	-	Index	0,04	-	0,14	-	-
Banks	-	0,14	0,06	-	-	Banks	-	-	0,20	0,08	-
Volatility	-	0,01	0,01	-	-	Volatility	0,09	-	-	0,03	-
Sovereign	-	-	-	-	0,06	Sovereign	-	0,13	0,05	0,18	-

Germany

IN	FX	Index	Banks	Volatility	Sovereign	OUT	FX	Index	Banks	Volatility	Sovereign
FX	-	-	0,06	-	0,27	FX	0,28	0,04	-	-	-
Index	0,14	0,13	-	0,01	0,02	Index	-	0,20	0,10	-	-
Banks	0,03	-	-	-	-	Banks	-	0,07	-	0,08	-
Volatility	-	0,02	0,02	0,10	-	Volatility	-	0,15	0,04	-	0,02
Sovereign	-	0,09	-	0,12	-	Sovereign	-	0,02	0,01	-	-

France

IN	FX	Index	Banks	Volatility	Sovereign	OUT	FX	Index	Banks	Volatility	Sovereign
FX	-	-	-	-	-	FX	0,40	-	-	-	-
Index	0,22	0,26	-	0,14	0,04	Index	-	0,17	0,07	0,02	-
Banks	0,02	0,09	0,07	0,04	0,04	Banks	-	0,03	0,02	0,07	-
Volatility	-	0,02	0,06	0,01	-	Volatility	-	0,02	0,02	0,07	0,02
Sovereign	-	-	-	-	-	Sovereign	-	0,03	-	-	0,06

Italy

IN	FX	Index	Banks	Volatility	Sovereign	OUT	FX	Index	Banks	Volatility	Sovereign
FX	0,18	0,01	0,23	-	-	FX	0,16	0,19	-	0,19	-
Index	-	0,09	-	-	0,08	Index	-	0,11	0,05	-	-
Banks	-	0,05	-	0,08	0,01	Banks	0,01	0,12	0,02	-	-
Volatility	0,01	0,23	-	0,02	-	Volatility	-	0,02	0,04	0,05	0,03
Sovereign	-	-	-	-	-	Sovereign	-	-	-	-	-

China

IN	FX	Index	Banks	Volatility	Sovereign	OUT	FX	Index	Banks	Volatility	Sovereign
FX	0,25	-	-	-	-	FX	-	-	0,19	0,13	-
Index	-	0,03	0,02	-	-	Index	-	-	-	0,04	-
Banks	0,09	0,03	0,08	0,07	0,01	Banks	-	0,29	0,02	0,09	-
Volatility	-	-	-	-	-	Volatility	-	-	-	0,04	-
Sovereign	0,10	0,12	-	0,01	0,16	Sovereign	-	0,11	-	-	0,08

Japan

IN	FX	Index	Banks	Volatility	Sovereign	OUT	FX	Index	Banks	Volatility	Sovereign
FX	0,07	-	-	0,10	-	FX	0,53	-	-	-	-
Index	-	0,06	0,38	0,05	0,05	Index	0,25	-	-	-	-
Banks	-	0,07	0,03	0,10	0,06	Banks	0,01	0,06	-	-	0,04
Volatility	-	-	0,01	0,01	0,01	Volatility	-	-	-	-	-
Sovereign	-	-	-	-	-	Sovereign	0,11	-	-	-	-

Australia

IN	FX	Index	Banks	Volatility	Sovereign	OUT	FX	Index	Banks	Volatility	Sovereign
FX	-	-	-	-	-	FX	-	-	-	0,27	0,32
Index	0,12	0,09	0,12	0,01	-	Index	-	-	0,12	-	0,02
Banks	-	0,19	0,06	0,11	-	Banks	-	-	0,06	-	0,02
Volatility	0,06	-	0,02	-	0,05	Volatility	-	-	-	-	-
Sovereign	0,01	0,05	0,05	-	0,06	Sovereign	-	-	-	-	0,18

Table 4.3: This table reports for the entire period, on the left, the main channels used for incoming information with the vertical axis reporting the variables of the country receiving the information and the horizontal axis reporting the variables of the other countries from which the information has been transmitted. On the right, we have the main channels used for outgoing information with the horizontal axis reporting the variables of the country transmitting the information and the vertical axis reporting the variables of the other countries to which the information has been transmitted.

4.2.4 Commodity markets and Cross-country information transfers

We investigate in this section the role of commodity markets in the transmission of information between countries. The past decade has witnessed a strong modification of

Rec/Trans	US	CA	UK	DE	FR	IT	CN	JP	AU
US	0	FX-Sov	0	FX-FX	FX-FX	FX-Vol	FI-FI	FX-FX	FX-Vol
CA	FX-FX	0	FX-FX	FX-Ind	Vol-Vol	FX-FX	0	FX-FX	FX-Sov
UK	0	FX-Sov	0	Ind-Ind	Sov-Sov	FX-Ind	FI-Ind	FX-FX	FX-Vol
DE	FX-Sov	Ind-FX	Sov-Vol	0	Ind-Ind	Vol-Ind	Sov-Ind	0	FX-Sov
FR	Ind-Vol	Ind-Ind	0	Ind-Ind	0	Vol-FI	FI-Vol	Ind-FX	FI-FI
IT	Ind-Ind	Ind-Sov	FI-Vol	Vol-Ind	Vol-Vol	0	FX-FI	FX-FX	0
CN	Sov-Ind	FI-FX	FI-FI	FX-FX	FX-FX	Ind-Ind	0	Sov-FX	Sov-Sov
JP	Ind-FI	0	Ind-FI	Ind-FI	Ind-FI	FX-FX	FX-Vol	0	Ind-FI
AU	0	Ind-Ind	Vol-FX	Ind-FI	Sov-Ind	Ind-FI	FI-Vol	Ind-FX	0

Table 4.4: This table reports for each pair of connected countries, the main channel driven the information transmission, with the horizontal axis reporting the countries transmitting the information and the vertical axis reporting the countries receiving the information (with FX for the currencies, FI for the financial indices, Ind for the stock market indices, Vol for the volatility indices and Sov for the sovereign bonds).

commodities' connections to the financial market, as they have been gradually included in many portfolios of financial market players such as hedge funds and day traders for a risk diversification purpose. This financialization of commodities Natanelov et al. (2011); Creti et al. (2013) explains the large fluctuation observed in the market during the last global financial crisis. These fluctuations affected many countries leading to a capital outflow for exporting countries when the prices decreased due to the slowdown of emerging market economy; and reducing their competitiveness when the prices increased via currency appreciation. Commodities fluctuations have also impacted financial system as illustrated by the recent drop in crude oil price which jeopardizes many investments of financial institutions in new exploration projects, affecting the stock market. These effects illustrates clearly the influence of commodities in the evolution of the economy of both importing and exporting countries. Commodities could therefore represent an additional channel of communication between importing and exporting countries.

The topics addressed by the literature treating about commodities can be grouped into three major categories, the papers that focus on the interdependencies between commodity markets Escribano and Granger (1998); Ciner (2001), the ones that examine the spillovers between commodities and financial variables such as exchange rates, interest rates or stock prices Bhar and Hammoudeh (2011) and the ones that model the volatility of these commodities Beck (2001); Dahl and Iglesias (2009). In the second type of approach, exchange rates are treated as the most important macroeconomic variable and could therefore be, as mentioned earlier, the preferred channel of transmission between countries and the commodity markets in our analysis. This section proposes therefore to contribute to this second strand of the literature about commodities.

4.2.5 Analysis of commodity markets and Cross-country information transfers

We rely, in this section, on a similar procedure as the one developed for cross-country interdependencies detection, in order to assess the information flows from the commodity markets to the different countries and from the countries to the commodity markets. We create therefore a commodity system represented by the 8 variables presented earlier, gold, silver, aluminum, copper, crude oil, soybean, wheat and corn. We start by estimating the active channel to be considered in our multichannel causality measure by applying once again a bivariate Granger causality test. We then apply our model and average the probability of activity for each channel for the whole sample and for the different periods mentioned before.

We then start by aggregating the information for all the countries to get a global overview of the inflows and outflows for the commodities. As can be seen from Table 4.5, the commodities seems to impact mostly the currencies with for the entire period, 52 percent of the total information flow imputable to the channel Commodity-Currency. This clearly confirms the importance of the relationship between these two markets and reinforce the conclusions of the abundant literature treating this subject Bhar and Hammoudeh (2011); Lee and Chen (2006). The same is true for the stock market indices with 28 percent overall. With regards the distinctive features of each commodity, gold, metals and oil have clearly a much higher influence than food on advanced markets. Gold has the higher impact with 43 percent of the total information flows, influencing mainly the evolution of exchange rates and stock market indices. This high percentage and the impact on stock market indices is mainly due to its role of safe haven asset, especially in time of crisis. As for the other metals, their impact on currencies is mainly due to the importance of theses raw materials for both importing and exporting countries' economy. They seem to have a more global impact than oil which influence only the stock market indices. Similarly to the results obtained for the information flows between countries, we observed a great stability in time for the channels repartition with only small variations when considering the different periods.

Table 4.6 provides information about the impact of countries fundamentals on the commodity markets. In contrast with the major role of gold for the currencies and the equities, we found no information transmitted from these different markets to gold. Instead, it seems that copper and oil are highly influenced by the selected financial variables. Nevertheless, this does not mean that these variables have no influence on gold but just that this influence is less pronounced than for copper or oil. Overall the commodities are mainly influenced by the exchange rate and by the equity market. As for the evolution in time, we observed the same stability as before.

Entire Period

OUT	Gold	Silver	Aluminium	Copper	Oil	Soybean	Wheat	Corn
FX	0,24	-	0,12	0,16	-	0,01	-	-
Index	0,15	-	0,03	-	0,07	0,03	-	-
Banks	0,05	-	-	0,04	-	-	-	-
Volatility	-	0,01	0,02	-	0,03	-	-	0,01
Sovereign	-	-	0,03	-	-	-	-	-

Pre-Crisis Period

OUT	Gold	Silver	Aluminium	Copper	Oil	Soybean	Wheat	Corn
FX	0,25	-	0,12	0,17	-	0,01	-	-
Index	0,12	-	0,03	-	0,05	0,02	-	-
Banks	0,03	-	-	0,06	-	-	-	-
Volatility	-	0,01	0,05	-	0,05	-	-	0,01
Sovereign	-	-	0,02	-	-	-	-	-

Global Financial Crisis Period

OUT	Gold	Silver	Aluminium	Copper	Oil	Soybean	Wheat	Corn
FX	0,25	0,01	0,13	0,16	-	0,01	-	-
Index	0,14	-	0,03	-	0,07	0,03	-	-
Banks	0,02	0,01	-	0,03	-	-	-	0,01
Volatility	-	0,01	0,01	-	0,03	-	-	0,01
Sovereign	-	-	0,04	-	-	-	-	-

European Sovereign Debt Crisis Period

OUT	Gold	Silver	Aluminium	Copper	Oil	Soybean	Wheat	Corn
FX	0,22	-	0,11	0,15	-	0,01	-	-
Index	0,18	-	0,03	-	0,08	0,04	-	-
Banks	0,07	-	-	0,03	-	-	-	0,01
Volatility	-	0,01	-	-	0,01	-	-	-
Sovereign	-	-	0,03	-	-	-	-	-

Post-Crisis Period

OUT	Gold	Silver	Aluminium	Copper	Oil	Soybean	Wheat	Corn
FX	0,22	-	0,10	0,15	-	0,01	-	-
Index	0,17	-	0,04	-	0,08	0,04	-	-
Banks	0,09	-	-	0,04	-	-	-	-
Volatility	-	0,01	-	-	0,01	-	-	-
Sovereign	-	-	0,03	-	-	-	-	-

Table 4.5: This table reports for the different periods, the main channels used for the transfer of information from the commodity market to the countries of our data-set with the horizontal axis reporting the variables of the commodity market transmitting the information and the vertical axis reporting the variables of the countries to which the information is transmitted.

In order to assess the possible use of commodity markets as an additional indirect channel for information transmission between the countries of our data-set, we now consider, in more details, the relationships existing between the selected financial variables and the commodities for each country. To be able to compare properly the strength of these relationships, we create five new data-sets, one for each type of financial variables. These data-sets regroups for each financial variable the time series of every countries. We then estimate using the same methodology as before the links existing between the commodities and each of these data-set, looking at relationships from the commodities to the financial variables and from the financial variables to the commodities.

The resulting matrices are given in Table 4.7 for the transmission from the selected

Pre-Crisis

IN	FX	Index	Banks	Volatility	Sovereign
Gold	-	-	-	-	-
Silver	-	0,02	0,03	-	-
Aluminium	-	0,09	-	-	-
Copper	0,20	0,22	0,05	-	-
Oil	0,20	-	-	-	-
Soybean	-	-	-	-	-
Wheat	-	-	0,04	-	-
Corn	-	0,09	-	0,05	-

Post-Crisis

IN	FX	Index	Banks	Volatility	Sovereign
Gold	-	-	-	-	-
Silver	-	0,02	0,03	-	-
Aluminium	-	0,10	-	-	-
Copper	0,20	0,22	0,05	-	-
Oil	0,20	-	-	-	-
Soybean	-	-	-	-	-
Wheat	-	-	0,04	-	-
Corn	-	0,09	-	0,05	-

Post-Crisis

IN	FX	Index	Banks	Volatility	Sovereign
Gold	-	-	-	-	-
Silver	-	0,02	0,03	-	-
Aluminium	-	0,08	-	-	-
Copper	0,20	0,21	0,05	-	-
Oil	0,20	-	-	-	-
Soybean	-	-	-	-	-
Wheat	-	-	0,04	-	-
Corn	-	0,10	-	0,05	-

Post-Crisis

IN	FX	Index	Banks	Volatility	Sovereign
Gold	-	-	-	-	-
Silver	-	0,02	0,03	-	-
Aluminium	-	0,08	-	-	-
Copper	0,20	0,23	0,05	-	-
Oil	0,20	-	-	-	-
Soybean	-	-	-	-	-
Wheat	-	-	0,05	-	-
Corn	-	0,09	-	0,05	-

Post-Crisis

IN	FX	Index	Banks	Volatility	Sovereign
Gold	-	-	-	-	-
Silver	-	0,02	0,03	-	-
Aluminium	-	0,10	-	-	-
Copper	0,20	0,23	0,05	-	-
Oil	0,20	-	-	-	-
Soybean	-	-	-	-	-
Wheat	-	-	0,04	-	-
Corn	-	0,09	-	0,05	-

Table 4.6: This table reports for the different periods, the main channels used for the transfer of information from the countries of our data-set to the commodity market with the horizontal axis reporting the variables of the countries transmitting the information and the vertical axis reporting the variables of the commodity market to which the information is transmitted.

financial variables to the commodities and in Table 4.8 for the transmission in the reverse direction. As can be seen for Table 4.7 , apart from the exchange rate, we observe a transmission of information from the countries' variables to the commodities for North America and Europe, with food commodities mainly impacted by Europe

and metal and oil by North America. This is in line with the fact that the U.S. and Canada are large producers of oil and metals and that Europe contributes to the commodity market mainly via food production. When looking now at Table 4.8, we observe a clear transmission from the commodity markets to the Asian financial variables, with metals, oil and food impacting their fundamentals. This tends to prove that, in addition to the cross-country channels highlighted in section 4.2.3, there seems to exist a channel from North America to Asia using metals and oil as transmitter and from Europe to Asia using food supply as transmitter. As for the exchange rate, we see clearly that the commodities influence mainly the USD which may be explained by the fact that most commodities, especially metals, are traded in USD and therefore their prices influence the demand for this currency and in turn its value Bhar and Hammoudeh (2011).

As we considered, in our analysis, several different times series in a single measure, co-movements in the set of selected fundamentals may alter the results of our analysis. Therefore, we repeated the analysis presented in the previous sections, by filtering the raw data from common factor using for each country the first principal component of the set of fundamentals characterizing the country. The results are qualitatively the same and available upon request.

Currencies

FX	USD	CAD	GBP	EUR	CNY	JPY	AUD
Gold	-	-	-	-	-	-	-
Silver	-	-	-	-	-	-	-
Aluminium	-	-	0,95	-	-	-	-
Copper	-	0,02	-	0,02	-	-	-
Oil	-	-	-	-	-	-	-
Soybean	-	-	-	-	-	-	-
Wheat	-	-	-	-	-	-	-
Corn	-	-	-	-	-	-	-

Stock market indices

Equi	U.S.	CA	U.K.	DE	FR	IT	CN	JP	AU
Gold	0,31	-	-	-	-	-	-	-	-
Silver	-	-	-	-	-	-	-	-	-
Aluminium	0,42	-	-	-	-	-	-	-	-
Copper	-	-	-	-	-	-	-	-	-
Oil	0,27	-	-	-	-	-	-	-	-
Soybean	-	-	-	-	-	-	-	-	-
Wheat	-	-	-	-	-	-	-	-	-
Corn	-	-	-	-	-	-	-	-	-

Financial indices

FI	U.S.	CA	U.K.	DE	FR	IT	CN	JP	AU
Gold	-	-	-	-	-	-	-	-	-
Silver	-	-	-	-	-	-	-	-	-
Aluminium	-	-	-	-	-	-	-	-	-
Copper	-	-	-	-	-	-	-	-	-
Oil	-	-	-	-	-	-	-	-	-
Soybean	-	-	-	-	-	-	-	-	-
Wheat	-	-	-	0,51	-	-	-	-	-
Corn	-	-	0,04	-	-	0,45	-	-	-

Volatility indices

Vol	U.S.	CA	U.K.	DE	FR	IT	CN	JP	AU
Gold	-	-	-	-	-	-	-	-	-
Silver	0,29	-	-	-	-	-	-	-	-
Aluminium	0,38	-	-	-	-	-	-	-	-
Copper	0,33	-	-	-	-	-	-	-	-
Oil	-	-	-	-	-	-	-	-	-
Soybean	-	-	-	-	-	-	-	-	-
Wheat	-	-	-	-	-	-	-	-	-
Corn	-	-	-	-	-	-	-	-	-

Sovereign bond yield

Sov	U.S.	CA	U.K.	DE	FR	IT	CN	JP	AU
Gold	-	-	-	-	-	-	-	-	-
Silver	-	-	-	-	-	-	-	-	-
Aluminium	-	-	-	-	-	-	-	-	-
Copper	-	-	-	-	-	-	-	-	-
Oil	-	1,00	-	-	-	-	-	-	-
Soybean	-	-	-	-	-	-	-	-	-
Wheat	-	-	-	-	-	-	-	-	-
Corn	-	-	-	-	-	-	-	-	-

Table 4.7: This table reports for the entire period, the main channels through which the financial variables of the selected countries influence commodities.

4.3 Conclusion

In this chapter, we explored the possibility of improving our understanding of the information flows inside the economy taking into account the greater interdependence of markets due to the financial globalization. We proposed a new approach derived from the Markov Switching model to infer the information transmissions between different

Currencies

FX	Gold	Silver	Aluminium	Copper	Oil	Soybean	Wheat	Corn
USD	0,82	-	-	0,17	-	-	-	-
CAD	-	-	-	-	-	-	-	-
GBP	0,01	-	-	-	-	-	-	-
EUR	-	-	-	-	-	-	-	-
CNY	-	-	-	-	-	-	-	-
JPY	-	-	-	-	-	-	-	-
AUD	-	-	-	-	-	-	-	-

Stock market indices

Equi	Gold	Silver	Aluminium	Copper	Oil	Soybean	Wheat	Corn
U.S.	-	-	-	-	-	-	-	-
CA	-	-	-	-	-	-	-	-
U.K.	-	-	-	-	-	-	-	-
DE	-	-	-	-	-	-	-	-
FR	-	-	-	-	-	-	-	-
IT	-	-	-	-	-	-	-	-
CN	-	-	-	-	-	-	-	-
JP	-	-	0,11	-	-	-	0,21	-
AU	-	-	0,31	0,05	0,32	-	-	-

Financial indices

FI	Gold	Silver	Aluminium	Copper	Oil	Soybean	Wheat	Corn
U.S.	-	-	-	-	-	-	-	-
CA	-	-	-	-	-	-	-	-
U.K.	-	-	-	-	-	-	-	-
DE	-	-	-	-	-	-	-	-
FR	-	-	-	-	-	-	-	-
IT	-	-	-	-	-	-	-	-
CN	-	-	0,31	-	-	-	0,65	-
JP	-	-	-	-	-	-	-	-
AU	-	-	-	-	-	-	-	0,03

Volatility indices

Vol	Gold	Silver	Aluminium	Copper	Oil	Soybean	Wheat	Corn
U.S.	-	-	-	-	-	-	-	-
CA	-	-	-	-	-	-	-	-
U.K.	-	-	-	-	-	-	-	-
DE	-	-	-	-	-	-	-	-
FR	-	-	-	-	-	-	-	-
IT	-	-	-	-	-	-	-	-
CN	-	-	-	-	-	-	-	-
JP	-	0,02	0,03	0,01	0,08	0,11	0,30	-
AU	-	0,30	-	-	-	-	-	0,15

Sovereign bond yield

Sov	Gold	Silver	Aluminium	Copper	Oil	Soybean	Wheat	Corn
U.S.	-	-	-	-	-	-	-	-
CA	-	-	-	-	-	-	-	-
U.K.	-	-	-	-	-	-	-	-
DE	-	-	-	-	-	-	-	-
FR	-	-	-	-	-	-	-	-
IT	-	-	-	-	-	-	-	-
CN	-	-	-	-	0,44	-	-	-
JP	-	-	-	-	-	-	-	-
AU	0,26	-	0,28	-	0,01	0,02	-	-

Table 4.8: This table reports for the entire period, the main channels through which the financial variables of the selected countries are influenced by commodities.

multivariate systems. Beyond the global measure proposed by Barrett and Barnett (2010) and the Granger causality test developed by Psaradakis et al. (2005b), our approach provides a detailed view of the channels' dynamics describing the information transmission process between two complex systems, providing information on the

channels' activity at each time step.

We investigated using our new methodology the causal relationships existing between 9 countries representing the world major economies. Instead of looking at simple pairwise relationships, our new methodology enables us to look at the same time at several channels of transmission by considering for each country 5 different financial variables that have been selected to represent the state of each country.

The aggregated results demonstrated the importance of currencies and equities as both transmitters and receivers of information in contrast with the common view that since the Global Financial Crisis, the main vectors of information were the financial institutions. Nevertheless, the U.S. and U.K. financial sector have proven to be relatively important sources of information. As for the currency channel, the Euro and the Japanese Yen seem to be the main spreader of information influencing mainly the U.S. dollar and the Canadian dollar. Eventually, the channels involving the stock market indices were especially important inside the European Union.

The second part of our empirical investigation consisted in examining the relationships existing between countries and the commodity markets and look at the possibility that the commodity markets could represent an additional indirect channel of information transmission between countries. Our results confirmed the major relationships existing between commodities and exchange rates but also between commodities and equities. These close ties concern primarily metals and oil. Eventually we have determined the existence of a link between North America and Asia through oil and metals and from Europe to Asia through agricultural and food commodities.

Conclusion

Each chapter of this thesis has contributed to the literature both methodologically and empirically. The methodological part of the thesis had the objective of understanding causality as a whole considering the many aspect of causality such as the type of causality, the importance of the environment around a causal link or the different channel through which information can travel. Each chapter brought a new piece of this complex causality puzzle.

The objective of the first chapter was to introduce the question of causality by describing the functioning of two pairwise methods, transfer entropy and Granger causality and the main concepts that led to their development. We explore the possibility to better identify causal relations between financial series using transfer entropy, comparing the results with Granger causality which is commonly used in economics and finance. Based on simulations of possible data generating processes (DGPs) for returns with causal relationship both in mean and variance, we studyied the properties of those different measures based on rigorous testing beyond the Gaussian equivalence case put into lights by Barnett et al. (2009).The DGPs were selected in order to replicate some features of financial variables such non stationarity in variance, non linearity, or kurtosis in the probability distribution.

Results show that for highly nonlinear and/or non Gaussian DGPs incorporating extreme and rare events and for causal relationships in variance, transfer entropy leads to better causality detection than standard Granger causality, the statistical tests associated to transfer entropy presented better results in terms of power and size but also in terms of ROC and PR curves.

In the second chapter, we tried to go beyond the simple pairwise approach and look at the impact of the environment around the variables transmitting and receiving the information. Indeed the pairwise approach tends to overestimate the number of causal relationships in a specific system by considering both the direct and indirect links. Drawing on recent contributions proposing strategies to deal with this problem such as the so-called "global silencing" approach of Barzel and Barabasi (2013) or "network

deconvolution" of Feizi et al. (2013), we proposed a novel methodology to infer an effective network structure from multivariate conditional information transfers. The aim of the second chapter was to improve network inference from temporal data acting on two different levels. The proposed approach takes into account the effects of both increasing dimensionality on information transfer estimation and redundancies caused by both inter- and intra-lag relationships. Our parsimonious approach relies on a pre-search step where every connection is inferred using a simple pairwise approach. In a second step, the algorithm uses conditional measure to prune the network from its non-relevant causal relationships.

We relied on a Monte Carlo simulation exercise to demonstrate the effectiveness of our novel approach compared to four other state of the art methodologies. We have shown through these simulations that, regardless of the underlying causality measure selected, the proposed algorithm generally performs better than the other methodologies when examining both intra- and inter-lag relationships. On top of those good results in terms of area under the ROC and PR curve, the other methodology necessitate a good calibration of their parameters in contrast with our approach which necessitate only the definition of a significance level. We also showed that for small samples, conditional Granger causality is the adequate measure for information transfer estimation, transfer entropy being as effective for larger samples as already demonstrated in the first chapter.

Drawing away from the subject of network theory, we could consider in another way the environment around a causal link by introducing in the causality estimation framework more than one possible sources and destinations. The last two chapters are therefore devoted to the development of two multichannel causality measures. These approach allow to consider systems for which more than one channel are used to transfer information; either because the system is represented by several variables or that we can decompose each variable into spectrum using frequency or quantile decomposition. Inspired by Neuroscience where multivariate connectivity is a common problematic and by papers such as Barrett and Barnett (2010), we have tried to address this complex question of multi-channel transmissions, getting away from the usual problematic of multivariate explanatory variables to account for systems described by both multivariate explanatory and dependent variables. The aim of these last chapters was therefore to look at the evolution in time of the relative importance of every channel in one global measure.

We proposed therefore two different frameworks: the first one presented in chapter 3, relied on the simplifying assumption that, at each time step, a specific channel dominates the information transfer between the two considered systems. Based on this new

framework, we developed two different multi-channel causality measures derived from the usual Granger causality to account for linear interactions and from the concept of transfer entropy for nonlinear contribution. Our measures provide different information about the inferred causal links: the strength of the global interaction between the two sub-systems, the average frequency of the channel switches and the channel contributing the most to the information transfer process for each time step.

The first step was to create a measure of the relative importance of each channel at every time steps. Once defined, the algorithm determined, for the considered step, the channel that maximizes the information transfer. These two steps were performed on every observation and the selected channels form a map of the time evolution of the channels' activity. To avoid instability in the channel selection, a softening process is applied.

As in the previous chapters, a Monte Carlo simulation exercise has been performed in order to assess how the proposed models react to different configurations of coupled systems, looking at the effect of the switching frequency, the strength of the relationships, the type of relationships and the complexity of the channels dynamic involved in the information transfer. The results of this simulation exercise demonstrated the ability of our models to infer properly the channel dynamic, especially for low switching frequencies. Both models showed equivalent results when the underlying dynamic is linear, while the transfer entropy based measure showed clearly better results for the non linear dynamic.

In the last chapter we treated again the question of multichannel relationships but from a different point of view. Beyond the global measure proposed by Barrett and Barnett (2010) and the Granger causality test developed by Psaradakis et al. (2005b), our approach provided a detailed view of the channels' dynamics describing the information transmission process between two complex systems; providing information on the channels' activity at each time step. Rather than considering the relative importance of every channel at each time step as in the models proposed in chapter 3, our methodology relied here on the regime switching Granger causality test which enabled us to look at the entire time window in one step. We proposed to modify the regime switching model to take into account multiple dependent variables. Instead of considering each explanatory variable as a possible source of information for one specific dependent variable, we considered every possible channel, represented by a specific pair of dependent and explanatory variable, as a possible state of the world. The Markov regime switching framework provided comprehensive information about the probability of activity of every channel for each time step. A comparison between this regime switching approach and the approaches developed in the chapter 3 is proposed

in the appendix and demonstrate the higher effectiveness of the regime switching approach, although both methods are relatively close.

As mentioned earlier, the aim of this thesis was twofold, first the development of new methodologies related to causality inference, then the investigation of several key questions raised by the literature in finance and economics. The empirical applications had therefore two main objectives, first the demonstration of the empirical interest of the different methodologies developed so far, then to address some crucial questions in finance and economics. The first three chapters have been devoted to the improvement of our understanding of the western financial system and the relationships between both its topology and systemic risk; whereas the last chapter consider some of the most advanced economics from another point of view, relying on a more macroeconomic approach.

In the first chapter, we proposed to use daily CDS observations of European and U.S. financial institutions to achieve two main objectives. The first one was to assess the ability of of Granger causality and transfer entropy to identify the relationships between the considered financial institutions by comparing the information content of the networks estimated with both causality measures. The information content was estimated by looking at the ability of the topological characteristics of the inferred networks to properly describe the evolution of the risk associated with the financial institutions via linear regressions. The second objective was to consider the influence of the surrounding environment of a network in the definition of its connectivity. In order to achieve this second objective we relied on the algorithm developed in chapter 2, as the treatment of the redundancies in the information transfer process requires the conditioning on the environment of the node receiving or transmitting the information. We considered a sub-sample of 24 systematically important financial institutions as the main network of interest and considered additionally 80 other financial institutions divided in three datasets which have been added iteratively to the set of condition to be considered in the network inference procedure.

The results of this first application shows that the information content of the networks estimated with each causality measure explained well the evolution of the risk, considering the high percentage of variance explained by our models. We also showed that both causality measures were complementary. We demonstrated then the positive impact of the surrounding environment on our network, the percentage of explained variance of the topological characteristics of the inferred networks increasing as the number of peripheral financial institutions included in the inference process built up. Such results have been observed for Granger causality when considering all type of financial institutions but only for banks in the case of transfer entropy. These results

seemed to confirm that the ability of a network to benefit from the information content of its surrounding environment depends directly on the ability of the underlying measure to treat high dimensional estimation. In the case of transfer entropy such high dimensional estimation requires longer time series. Nevertheless our empirical investigation was not able to provide a clear limit on the network sample size for the considered application.

In the second chapter, we proposed a similar approach by considering again the impact of topological characteristics of the inferred networks on U.S. financial institutions risk, looking here at the influence of the frequency of the underlying data on the information content of the networks. The results led to similar conclusions compared to the first chapter with the position of an institution inside the network representing a key element to describe the evolution of its risk. The results for the different frequencies provided useful insights about the frequency dependencies of time series information content, showing that it exists a clear trade-off between information and noise leading to the existence of an optimal frequency to retrieve financial networks. Finally, we compared these results with those obtained using a simple pairwise approach instead of our network inference algorithm and confirmed that our algorithm gives a more precise view of the interactions inside networks and therefore gives more information on financial institution fragility.

In contrast with chapter 1 and 2, the application developed in chapter 3 does not consider the entire system in terms of network but look rather at the way financial institutions communicate to each other at a pairwise level using our multichannel approach. At this individual level, the considered channels between financial institutions are expressed both in terms of spectral representation using wavelet transform and probability distribution using quantile regressions. The obtained spectra have been used to estimate a multi-channel information transfer between every connected pair of U.S. financial institutions using the two measures developed in the methodological part. Indeed, every financial institution was represented by a set of time dependent variables representing the different frequencies or the different quantiles. The results of the multi-channel Granger causality test in the frequency domain shows that the transfer of information occurs mainly at high frequency especially during the financial crisis of 2008 which is in line with the common assumption that financial market are more reactive during turmoil period. As for the quantile and cross quantile analysis, it provided a clear insight on the importance of the extreme variations during the financial crisis confirming the importance of the tails in periods of turmoil. The reverse is true for period of stability where the entire probability distribution of the time series outcome is taken into account.

Eventually we examined, in the last chapter, the interrelationships among 9 advanced economies from a macroeconomic point of view. We went, as in the chapter 3, beyond the usual causal relationships, considering the time-varying dimension of the channels conveying causal relationships. Looking at multichannel information transfer, we had to describe each country as a multivariate system. Drawing from the spectral approach of chapter 3, we defined for these countries five different fundamental variables reflecting its state: the main stock market index, an index regrouping the main financial institutions of the country, a volatility index, the exchange rate and the 5 years USD denominated sovereign bond yield. We applied then our multichannel measure on these sets of time series to determine, over time, the main channels through which the information is transmitted between the different countries.

The aggregated results demonstrated the relevance of currencies and equities both as transmitters and receivers of information, contrasting with the common view developed during the last Global Financial Crisis that the main vectors of information were the financial institutions. Nevertheless, the U.S. and U.K. financial sector have proven to be a relatively important source of information. As for the currency channel, the Euro and the Japanese Yen seem to be the main spreader of information influencing mainly the U.S. dollar and the Canadian dollar. Eventually, the channels involving the stock market indices were especially important inside the European Union.

In a second step, we consider the relationships existing between these countries and the commodities market and look at the possible use of the commodities market as an indirect channel of information transmission between countries. Our results confirmed the existence of a clear link between commodities and exchange rates but also between commodities and equities, affecting mainly metals and oil. Eventually we have determined the existence of a link between North America and Asia through oil and metals and from Europe to Asia through agricultural and food commodities.

The aim of this thesis has been to highlight the fact that performing a causality analysis was not an easy task and require a deep understanding of the notions behind the different measures. Indeed, in order to be effective, such analysis has to take into account the many limitations of the existing measures. The definition of these limitations has been one of the backbones of this research considering the many simulation exercises that have been performed in order to assess the usefulness of the different considered causality measures. Without such a rigorous testing procedure, any empirical investigation could not be undertaken without limiting severely the validity of their conclusion; especially in social sciences where no theoretical law can describe the underlying behavior linked to the observations. As a conclusion, I would like to stress again the importance of this positivist research process in all field of science

where every assertion should be assessed before being embedded in a larger empirical or theoretical research.

Bibliography

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2015): "Systemic Risk and Stability in Financial Networks," *American Economic Review*, 105, 564–608.

Acharya, V. (2009): "A Theory of Systemic Risk and Design of Prudential Bank Regulation," *Journal of Financial Stability*, 5, 224–255.

Acharya, V., R. Engle, and M. Richardson (2012): "Capital Shortfall: A New Approach to Ranking and Regulating Systemic Risks," *American Economic Review*, 102, 59–64.

Adrian, T. and M. Brunnermeier (2009): "CoVaR," *Federal Reserve Bank of New York Staff Report*, 348.

Aliferis, C., I. Tsamardinos, and A. Statnikov (2003): "HITON: A Novel Markov Blanket Algorithm for Optimal Variable Selection," in *AMIA Annu Symp Proc.*, 21–25.

Alter, A. and A. Beyer (2013): "The Dynamics of Spillover Effects during the European Sovereign Debt Turmoil," *ECB Working Paper Series*.

Alzahrani, M., M. Masih, and O. Al-Titi (2014): "Linear and non-linear Granger causality between oil spot and futures prices: A wavelet based test," *Journal of International Money and Finance*, 48, 175–201.

Anand, K., P. Gai, and M. Marsili (2012): "Rollover Risk, Network Structure and Systemic Financial Crises," *Journal of Economic Dynamics and Control*, 36, 1088–1100.

Andersen, T., T. Bollerslev, and F. Diebold (2007): "Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility," *Review of Economics and Statistics*, 89, 701–720.

Andersen, T., T. Bollerslev, and F. X. Diebold (2010): *Handbook of Financial Econometrics: Parametric and Nonparametric Volatility Measurement.*, Chapter 2, Elsevier Science B.V.

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2005): "Roughing it Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility," Working Paper 11775, National Bureau of Economic Research.

Ang, A. and J. Chen (2002): "Asymmetric correlations of equity portfolios," *Journal of Financial Economics*, 63, 443–494.

Bae, K.-H., A. Karolyi, and R. M. Stulz (2003): "A New Approach to Measuring Financial Contagion," *Review of Financial Studies*, 16, 717–763.

Bahadori, M. T. and Y. Liu (2012): "Granger Causality Analysis in Irregular Time Series," in *SIAM Conference on Data Mining Proceedings*.

Ballaa, E., I. Ergena, and M. Migueis (2014): "Tail dependence and indicators of systemic risk for largeUS depositories," *Journal of Financial Stability,*, 15, 195–209.

Barabasi, A. and R. Albert (1999): "Emergence of Scaling in Random Networks," *Science*, 286, 509–512.

Bardoscia, M., S. Battiston, F. Caccioli, and G. Caldarelli (2015): "DebtRank: A Microscopic Foundation for Shock Propagation," *PLoS ONE*, 10, e0130406.

Barigozzi, M. and C. Brownlees (2013): "NETS: Network Estimation of Time Series," *Barcelona GSE Working Paper Series, Working Paper n^o 723*.

Barndorff-Nielsen, O. and N. Shephard (2003): "Power and Bipower Variation with Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, 2, 1–37.

Barnett, L., A. B. Barrett, and A. K. Seth (2009): "Granger Causality and Transfer Entropy Are Equivalent for Gaussian Variables," *Phys. Rev. Lett.*, 103.

Barrett, A. B. and A. Barnett, L.Barnett (2010): "Multivariate Granger causality and generalized variance," *Phys. Rev. E*, 81.

Barzel, B. and A. Barabasi (2013): "Network link prediction by global silencing of indirect correlations," *Nature Biotechnology*, 31, 720–725.

Battiston, S., J. Glattfelder, D. Garlaschelli, F. Lillo, and G. Caldarelli (2010): *Network Science*, vol. 131-163, Springer.

Battiston, S., M. Puliga, R. Kaushik, P. Tasca, and G. Caldarelli (2012): "DebtRank: Too Central to Fail? Financial Networks, the FED and Systemic Risk," *Scientific Reports*, 2, 541.

Beck, R. (2001): "Do Country Fundamentals Explain Emerging Market Bond Spreads?" *CFS*.

Bekaert, G., M. Ehrmann, M. Fratzscher, and A. Mehl (2014): "The Global Crisis and Equity Market Contagion," *The Journal of Finance*, 69, 2597–2649.

Bell, J. S. (1964): "On the Einstein-Podolsky-Rosen paradox," *Physics 1*, 195–200.

Berzuini, C., P. Dawid, and L. Bernardinell (2012): *Causality: Statistical Perspectives and Applications*, Wiley.

Bhar, R. and S. Hammoudeh (2011): "Commodities and financial variables: Analyzing relationships in a changing regime environment," *International Review of Economics and Finance*, 20, 469–484.

Bianchi, D., M. Billio, C. R.Casarin, and M. Guidolin (2015): "Modeling Contagion and Systemic Risk," *Syrto Working Paper*.

Billio, M. and S. DiSanzo (2006): "Granger-causality in Markov Switching Models," Working Papers 2006-20, Department of Economics, University of Venice.

Billio, M., M. Getmansky, A. W. Lo, and L. Pelizzon (2012): "Econometric measures of connectedness and systemic risk in the finance and insurance sectors," *Journal of Financial Economics*, 104, 535–559.

Bodart, V. and C. B. (2009): "Evidence of interdependence and contagion using a frequency domain framework," *Emerging Markets Review*, 10, 140–150.

Boss, M., H. Elsinger, M. Summer, and S. Thurner (2004): "Network topology of the interbank market," *Quantitative Finance*, 4, 677–684.

Béreau, S. and C. Dahlqvist (2014): "Testing causal relationships in financial time series: A comparison between Granger and information theoretic based approaches," in *European Conference on Complex Systems*.

Breitung, J. and B. Candelon (2006): "Testing for short and long-run causality: a frequency domain approach." *Journal of Econometrics*, 132.

Bressler, S. and A. K. Seth (2011): "Wiener-Granger Causality: A well established methodology," *NeuroImage*, 58, 323–329.

Brockmann, D., L. Hufnagel, and T. Geisel (2006): "The scaling laws of human travel," *Nature*, 462–465.

Brovelli, A., M. Ding, A. Ledberg, Y. Chen, R. Nakamura, and S. Bressler (2004): "Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality," *Proceedings of the National Academy of Sciences of the United States of America*, 101, 9849–9854.

Brownlees, C. and R. Engle (2017): "SRISK: A Conditional Capital Shortfall Measure of Systemic Risk," *Review of Financial Studies*, 30, 48–79.

Brukner, C. (2014): "Quantum causality," *Nature Physics Insight, Foundations of Quantum Mechanics*, 10, 259–263.

Brummitt, C. D. and T. Kobayashi (2015): "Cascades in multiplex financial networks with debts of different seniority," *Phys. Rev. E*, 91.

Brunnermeier, M., A. Crockett, C. Goodhart, A. Persaud, and H. S. Shin (2009): "The fundamental principles of financial regulation," *Centre for Economic Policy Research*, 11.

Bunge, M. (1963): *The Place of the Causal Principle in Modern Science*, Meridian Books.

Butte, A. and I. Kohane (2000): "Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements," *Pac. Symp. Biocomput.*, 418–429.

Campbell, R., K. Koedijk, and P. Kofman (2002): "Increased Correlation in Bear Markets: A Downside Risk Perspective," *Financial Analysts Journal*, 58, 87–94.

Candelon, B., M. Joets, and S. Tokpavi (2013): "Testing for Granger causality in distribution tails: An application to oil markets integration," *Economic Modelling*, 31, 276–285.

Celik, S. (2012): "The more contagion effect on emerging markets: The evidence of DCC-GARCH model," *Economic Modelling*, 29, 1946–1959.

Chen, C., R. Gerlach, and D. Wei (2009): "Bayesian causal effects in quantiles: Accounting for heteroscedasticity," *Computational Statistics and Data Analysis*, 53, 1993–2007.

Christopoulos, D. K. and M. A. León-Ledesma (2008): "Testing for Granger (non-)causality in a time-varying coefficient VAR model," *Journal of Forecasting*, 27, 293–303.

Chuang, C., C. Kuan, and H. Lin (2009): "Causality in quantiles and dynamic stock return-volume relations," *Journal of Banking & Finance*, 33, 1351–1360.

Ciner, C. (2001): "On the long run relationship between gold and silver prices A note," *Global Finance Journal*, 12, 299–303.

——— (2011): "Commodity prices and inflation: Testing in the frequency domain," *Research in International Business and Finance*, 25, 229–237.

Cohen-Cole, E., E. Patacchinin, and Y. Zenou (2011): "Systemic risk and network formation in the interbank market," *CEPR Discussion Paper*, 8332.

Cont, R. (2001): "Empirical properties of asset returns:stylized facts and statistical issues," *Quantitative Finance*, 1, 223–236.

Corbi, J., J. Prades, and Minds (2000): *Causes and mechanisms: a case against physicalism.*, Oxford (UK): Blackwell Publishers Ltd.

Cosslett, S. R. and L.-F. Lee (1985): "Serial Correlation in Discrete Variable Models," *Journal of Econometrics*, 27, 79–97.

Creti, A., M. Joëts, and V. Mignon (2013): "On the links between stock and commodity markets' volatility," *Energy Economics*, 37, 16–28.

Croux, C. and P. Reusens (2013): "Do stock prices contain predictive power for the future economic activity? A Granger causality analysis in the frequency domain," *Journal of Macroeconomics*, 35, 93–103.

Dahl, C. M. and E. M. Iglesias (2009): "Volatility spill-overs in commodity spot prices: New empirical results," *Economic Modelling*, 26, 601–607.

Dahlqvist, C. and J. Gnabo (2018): "Effective network inference through multivariate information transfer estimation," *Physica A*, 499, 376–394.

Darbellay, G. and I. Vajda (1999): "Estimation of the information by an adaptive partitioning of the observation space," *IEEE Trans. Inform. Theory*, 45, 1315–1321.

Daub, C., R. Steuer, J. Selbig, and S. Kloska (2004): "Estimating mutual information using B-spline functions an improved similarity measure for analysing gene expression data," *BMC Bioinformatics*, 5, 1–12.

Davidson, J. E. H., D. F. Hendry, F. Srba, and S. Yeo (1978): "Econometric Modelling of the Aggregate Time-Series Relationship Between Consumers' Expenditure and Income in the United Kingdom," *The Economic Journal*, 88, 661–692.

Diebold, F. and K. Yilmaz (2009): "Measuring Financial Asset Return and Volatility Spillovers, With Application to Global Equity Markets," *Economic Journal*, 119, 158–171.

——— (2014): "On the network topology of variance decompositions: Measuring the connectedness of financial firms," *Journal of Econometrics*, 182, 119–134.

Diks, C. and H. Fang (2017): "Transfer Entropy for Nonparametric Granger Causality Detection: An Evaluation of Different Resampling Methods," *Entropy*, 19.

Dimpfl, T. and F.-J. Peter (2013): "Using transfer entropy to measure information flows between financial markets," *Studies in Nonlinear Dynamics & Econometrics*, 1081–1826.

Dumitrescu, E. and C. Hurlin (2012): "Testing for Granger non-causality in heterogeneous panels," *Economic Modelling*, 29, 1450–1460.

Dungey, M. and D. Gajurel (2014): "Equity market contagion during the global financial crisis: Evidence from the world's eight largest economies," *Economic Systems*, 38, 161–177.

Dungey, M., M. Luciani, and D. Veredas (2012): "Ranking Systemically Important Financial Institutions," CAMA Working Papers 2012-47, Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University.

ECB (Januray 2011): "Recent advances in modelling systemic risk using network analysis," .

Eccles, J. and K. Popper (1977): *The Self and Its Brain*, Springer.

Edgeworth, F. Y. (1885): "Methods of Statistics," *Jubilee Volume of the Statistical Society, Royal Statistical Society of Britain*, 181–217.

Escribano, A. and C. W. J. Granger (1998): "Investigating the relationship between gold and silver prices," *Journal of Forecasting*, 17, 81–107.

Faber, R. (1986): *Clockwork Garden*, University of Massachusetts Press.

Falcon, A. (2011): *Aristotle on causality*, vol. 71 b 9-11, Stanford Encyclopedia of Philosophy.

Feigl, H. (1953): *Notes on Causality, Readings in the Philosophy of Science*, 408-418, Appleton-Century-Crofts.

Feizi, S., D. Marbach, M. Médard, and M. Kellis (2013): "Network deconvolution as a general method to distinguish direct dependencies in networks," *Nature Biotechnology*, 31.

Feng, X., W. Jo, and B. Kim (2014): "International transmission of shocks and fragility of a bank network," *Physica A*, 403, 120–129.

Fernández-Rodríguеza, F., M. Gómez-Puigb, and S. Sosvilla-Riveroc (2016): "Using connectedness analysis to assess financial stress transmission in EMU sovereign bond market volatility," *Journal of International Financial Markets, Institutions & Money*.

Florens, J. P. and M. Mouchart (1982): "A Note on Noncausality," *Econometrica*, 50, 583–591.

Foucault, M. (1982): *The Subject and Power*, The University of Chicago Press.

Fox, J. and S. Weisberg (2011): *An R Companion to Applied Regression*, Thousand Oaks CA: Sage, second ed.

Frankfurter, G. (2007): *Theory and Reality in Financial Economics: Essays towards a new political finance*, London: World Scientific Publishing.

Frankfurter, G. and E. McGoun (1999): "Ideology and the Theory of Financial Economics," *Journal of Economic Behavior and Organization*, 39, 159–177.

Friedman, M. (1953): *Essays in Positive Economics*, University of Chicago Press.

Friston, K. (1994): "Functional and effective connectivity in neuroimaging: A synthesis," *Human Brain Mapping 2, 56-78*.

Friston, K., R. Moran, and A. Seth (2013): "Analysing connectivity with Granger causality and dynamic causal modelling," *Curr Opin Neurobiol*, 23, 172–178.

Friston, K. J., A. M. Bastos, A. Oswal, B. van Wijk, C. Richter, and V. Litvak (2014): "Granger causality revisited," *NeuroImage*, 101, 796–808.

Galton, F. (1888): "Co-relations and their measurement, chiefly from anthropometric data," *Proceedings of the Royal Society of London*, 45, 135–145.

Geweke, J. (1982): "Measurement of linear dependence and feedback between multiple time series," *Journal of the American Statistical Association*, 77, 304–313.

Geweke, J. F., R. Meese, and W. Dent (1983): "Comparing alternative tests of causality in temporal systems : Analytic results and experimental evidence," *Journal of Econometrics*, 21, 161–194.

Ghysels, E., J. B. Hillb, and K. Motegic (2016): "Testing for Granger causality with mixed frequency data," *Journal of Econometrics*, 192, 207–230.

Goldfeld, S. M. and R. E. Quandt (1973): "A Markov Model for Switching Regressions," *Journal of Econometrics*, 1, 3–16.

Gorea, D. and D. Radev (2014): "The euro area sovereign debt crisis: Can contagion spread from the periphery to the core?" *International Review of Economics and Finance*, 30, 78–100.

Granger, C. W. J. (1969): "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, 37, 424–38.

——— (1980): "Testing for causality : A personal viewpoint," *Journal of Economic Dynamics and Control*, 2, 329–352.

——— (2003): "Some aspects of causal relationships," *Journal of Econometrics*, 112, 69–71.

Guo, S., C. Ladroue, and J. Feng (2010): "Granger Causality: Theory and Applications," in *Frontiers in Computational and Systems Biology*, Springer, chap. 5, 73–98.

Hacker, R. S., H. Karlsson, and K. Månsson (2014): "An investigation of the causal relations between exchange rates and interest rate differentials using wavelets," *International Review of Economics & Finance*, 29, 321–329.

Hafner, C. M. and H. Herwartz (2006): "A Lagrange multiplier test for causality in variance," *Economics Letters*, 93, 137–141.

Haldane, A. (2009): "Rethinking the financial network," in *Speech at the Financial Student Association, Amsterdam, 28 April 2009.*

Haldane, A. and R. May (2011): "Systemic risk in banking ecosystems," *Nature*, 469, 351–355.

Hamilton, J. D. (1988): "Rational-Expectations Econometric Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates," *Journal of Economic Dynamics and Control*, 12, 385–423.

——— (1994): *Time Series Analysis.*

Hardy, L. (2005): "Probability theories with dynamic causal structure: a new framework for quantum gravity," *Preprint at http://arxiv.org/abs/gr-qc/0509120*.

Hassani, H., X. Huang, R. Guptac, and M. Ghodsi (2016): "Does sunspot numbers cause global temperatures? A reconsideration using non-parametric causality tests," *Physica A*, 460, 54–65.

Hausman, D. (1992): *The Inexact and Separate Science of Economics*, Cambridge University Press.

Hegerty, S. W. (2016): "Commodity-price volatility and macroeconomic spillovers: Evidence from nine emerging markets," *North American Journal of Economics and Finance*, 35, 23–37.

Hendry, D. and H.-M. . Krolzig (2001): *Automatic Econometric Model Selection Using PcGets*, Timberlake consultant.

——— (2003): *Econometrics and the Philosophy of Economics: Theory-Data Confrontations in Economics, New developments in automatic general-to-specific modelling*, Chapter 16, Princeton University Press.

Hendry, D. and H.-M. Krolzig (1999): "Improving on Data mining reconsidered," *Econometrics Journal*, 2, 202–219.

Hendry, D. and G. Mizon (1978): "Serial correlation as a convenient simplification, not a nuisance: A comment on a study of the demand for money," *Bank of England. Economic Journal*, 88, 549–563.

Hlavackova-Schindler, K. (2011): "Equivalence of Granger Causality and Transfer Entropy: A Generalization," *Applied Mathematical Sciences*, 5, 3637–3648.

Hlavackova-Schindlera, K., M. Palusb, M. Vejmelkab, and J. Bhattacharyaa (2007): "Causality detection based on information-theoretic approaches in time series analysis," *Physics report*, 441, 1–46.

Hoefer, C. (2016): *Causal determinism*, Stanford Encyclopedia of Philosophy.

Holtz-Eakin, D., W. Newey, and H. S. Rosen (1988): "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56, 1371–95.

Hong, Y., Y. Liu, and S. Wang (2009a): "Granger causality in risk and detection of extreme risk spillover between financial markets," *Journal of Econometrics*, 150, 271–287.

Hong, Y., Y. Liuc, and S. Wang (2009b): "Granger causality in risk and detection of extreme risk spillover between financial markets," *Journal of Econometrics*, 150, 271–287.

Hood, W. and T. Koopmans (1953): *Studies in Econometric Method*, Yale University Press.

Hoover, K. D. (1990): "The Logic of Causal Inference: Econometrics and the Conditional Analysis of Causality," *Economics and Philosophy*, 6, 207–234.

——— (2001): *Causality in Macroeconomics*, Cambridge University Press.

——— (2004): "Lost Causes," *Journal of the History of Economic Thought*, 26, 149–164.

——— (2008): *Causality in Economics and Econometrics, The New Palgrave Dictionary of Economics*, Palgrave Macmillan.

Hoover, K. D. and S. J. Perez (1999): "Data mining reconsidered: encompassing and the general-to-specific approach to specification search," *Econometrics Journal, Royal Economic Society*, 2, 167–191.

Hume, D. (1738): *A Treatise of Human Nature*, vol. Part III, Courier Corporation.

——— (1742): *On Interest*, 304, Liberty Fund.

——— (1748): *An Enquiry Concerning Human Understanding*, Filiquarian Publishing, 2007.

Jammazi, R., R. Ferrer, F. Jareño, and S. M. Hammoudeh (2017): "Main driving factors of the interest rate-stock market Granger causality," *International Review of Financial Analysis*, 52, 260–280.

Jeong, K., W. K. Härdle, and S. Song (2012): "A Consistent Nonparametric Test For Causality In Quantile," *Econometric Theory*, 28, 861–887.

Jondeau, E., S.-H. Poon, and M. Rockinger (2007): *Financial Modelling under Non-Gaussian Distributions*, Springer Finance.

Kafatos, M. and R. Nadeau (1990): *The conscious universe: part and whole in modern physical theory.*, Springer.

Kalbaska, A. and M. Gatkowski (2012): "Eurozone sovereign contagion: Evidence from the CDS market (2005-2010)," *Journal of Economic Behavior & Organization*, 83, 657–673.

Kim, J., G. Kim, S. An, o.-K. Kwon, and S. Yoon (2012): "Entropy-Based Analysis and Bioinformatics: Inspired Integration of Global Economic Information Transfer," *PLOS One*, 8.

Kochen, S. and E. P. Specker (1967): "The problem of hidden variables in quantum mechanics," *Math. Mech*, 17.

Koenker, R. and G. Bassett (1978): "Regression quantile," *Econometrica*, 46.

Kolaczyk, E. (2017): *Topics at the frontier of Statistics and Network Analysis.*, Cambridge University Press.

Koopmans, T. (1950): *Statistical Inference in Dynamic Economic Models*, Wiley.

Krause, A. and S. Giansante (2012): "Interbank lending and the spread of bank failures: A network model of systemic risk," *Journal of Economic Behavior & Organization*, 83.

Krivitsky, P. N. and E. D. Kolaczyk (2015): "On the Question of Effective Sample Size in Network Modeling: An Asymptotic Inquiry," *Statist. Sci.*, 30, 184–198.

Krolzig, H.-M. (1997): *Markov-Switching Vector Autoregressions : Modelling, Statistical Inference, and Application to Business Cycle Analysis.*

Kwon, O. and J. Yang (2008): "Information flow between stock indices," *EPL (Europhysics Letters)*, 82, 68003.

Laplace (1820): *Essai Philosophique sur les Probabilités, Théorie Analytique des Probabilités,*, Paris Bachelier.

Lee, H. Y. and S. L. Chen (2006): "Why use Markov-switching models in exchange rate prediction?" *Economic Modeling*, 23, 662–668.

Lemmens, A., C. Croux, and M. Dekimpe (2008): "Measuring and Testing Granger Causality over the Spectrum: An Application to European Production Expectation Surveys," *International Journal of Forecasting*, 24, 414–431.

Liao, W., J. Ding, D. Marinazzo, Q. Xu, Z. Wang, C. Yuan, Z. Zhang, G. Lu, and H. Chen (2011): "Small-world directed networks in the human brain: Multivariate Granger causality analysis of resting-state fMRI," *NeuroImage*, 54, 2683–2694.

Lipsitch, M., T. Cohen, B. Cooper, J. Robins, S. Ma, L. James, G. Gopalakrishna, S. Chew, C. Tan, M. Samore, D. Fisman, and M. Murray (2003): "Transmission dynamics and control of severe acute respiratory syndrome," *Science*, 300, 1966–70.

Liu, Y. and M. Bahadori (2012): "A survey on Granger Causality: A Computational View," Mimeo.

Lizier, J. and M. Rubinov (2013): "Inferring effective computational connectivity using incrementally conditioned multivariate transfer entropy," *BMC Neuroscience, BioMed Central*.

Longin, F. and B. Solnik (1995): "Is the correlation in international equity returns constant: 1960-1990?" *Journal of International Money and Finance*, 14, 3–26.

Longstaff, F. A., J. Pan, L. H. Pedersen, and K. J. Singleton (2011): "How Sovereign Is Sovereign Credit Risk?" *American Economic Journal: Macroeconomics*, 3, 75–103.

Lundbergh, S. and T. Terasvirta (2002): "Evaluating GARCH models," *Journal of Econometrics*, 110, 417–435.

Marbach, D., J. Costello, R. Küffner, N. Vega, R. Prill, D. Camacho, K. Allison, T. D. Consortium, M. Kellis, J. Collins, and G. Stolovitzky (2012): "Wisdom of crowds for robust gene network inference," *Nature Methods*, 9, 796–804.

Marschinski, R. and H. Kantz (2002): "Analysing the information flow between financial time series," *The European Physical Journal B - Condensed Matter and Complex Systems*, 30, 275–281.

Marshall, A. (1930): *Principles of Economics: An Introductory Volume, 8th ed.*, Macmillan.

Masi, G. D. and M. Gallegati (2012): "Bank-firms topology in Italy," *Empirical Economics*, 43, 851–866.

Matthew, E., M. Jackson, and B. Golub (2014): "Financial Networks and Contagion," *American Economic Review*, 10, 3115–3153.

May, W. E. (1970): *Knowledge of Causality in Hume and Aquinas*, vol. 34.

Mayr, E. (1961): "Cause andEffet in Biology," *Sciences*, 134, 1501–1506.

McGinn, C. (2004): *Consciousness and its objects*, Oxford University Press.

Metiu, N. (2012): "Sovereign risk contagion in the Eurozone," *Economics Letters*, 117, 35–38.

Mezard, M. and A. Montanari (2009): *Information, Physics and Computation*, Oxford University press.

Mill, J. S. (1848): *Political Economy*, Longmans, Green and Co.

Minoiu, C. and J. A. Reyes (2013): "A network analysis of global banking: 1978-2010," *Journal of Financial Stability*, 9, 168–184.

Mäki, U. (Cambridge University Press): *The Economic World View : Studies in the Ontology of Economics.*, 2001.

Moon, Y., B. Rajagopalan, and U. Lall (1995): "Estimation of mutual information using kernel density estimators," *Phys. Rev. E*, 52, 2318–2321.

Morgan, M. S. (1991): *The Stamping Out of Process Analysis in Econometrics, Appraising Economic Theories: Studies in the Methodology of Research Programs*, 237-265, Aldershot.: Elgar.

Mykland, P. and L. Zhang (2009): *Statistical Methods for Stochastic Differential Equations: The econometrics of high-frequency data*, eds. Chapman & Hall/CRC Press.

Nair-Reichert, U. and D. Weinhold (2001): "Causality Tests for Cross-Country Panels: A New Look at FDI and Economic Growth in Developing Countries," *Oxford Bulletin of Economics and Statistics*, 63, 153–71.

Natanelov, V., M. Alam, A. Mc Kenzie, and G. Van Huylenbroeck (2011): "Is thereco-movement of agricultural commodities futures prices and crude oil?" *Energy Policy*, 39, 4971–4984.

Newey, W. and K. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica, 55, 703-708.*

Noble, D. (2008): "Genes and causation," *Phil. Trans. R. Soc. A*, 366, 3001–3015.

Norton, J. D. (2003): "Causation as Folk Science," *Philosophers' Imprint*, 3, 4.

Paninski, L. (2003): "Estimation of entropy and mutual information," *Neural Comput.*, 15, 1191–1253.

Pasquariello, P. (2006): "Imperfect Competition, Information Heterogeneity, and Financial Contagion," *Review of Financial Studies*, 20, 391–426.

Patterson, R. and A. K. Barbey (2012): *A Cognitive Neuroscience Framework for Causal Reasoning, The Neural Representation of Belief Systems*, vol. 77-120, Psychology Press.

Patton, A. J. (2004): "Asymmetric Dependence for Asset Allocation," *Journal of Financial Econometrics*, 2, 130–168.

Pearl, J. (2000a): *Causality: Models, Reasoning, and Inference*, Cambridge University Press.

——— (2000b): *Causality: Models, Reasoning and Inference; 2nd edition*, Cambridge University Press.

——— (2009): *Causality: Models, reasoning and inference*, CAmbridge University Press, 2nd ed.

Pearson, K. (1900): "On the Criterion that a given System of Deviations from the Probable in the Case of a Correlated System of Variables is such that it can be reasonably supposed to have arisen from Random Sampling," *Philosophical Magazine Series 5*, 50, 157–175.

Pesaran, H. and Y. Shin (1998): "Generalized impulse response analysis in linear multivariate models," *Econom. Lett.*, 58, 17–29.

Pradhan, R. P., M. B. Arvin, J. H. Hall, and S. Bahmani (2014): "Causal nexus between economic growth, banking sector development, stock market development, and other macroeconomic variables: The case of ASEAN countries," *Review of Financial Economics*, 23, 155–173.

Pratt, D. (2003): "Consciousness, Causality, and Quantum Physics," *NeuroQuantology*, 1, 58–67.

Psaradakis, Z., M. Ravn, and M. Sola (2005a): "Markov Switching and the money output relationship," *Journal of Applied Econometrics*, 20.

Psaradakis, Z., M. O. Ravn, and M. Sola (2005b): "Markov switching causality and the money output relationship," *Journal of Applied Econometrics*, 20, 665–683.

Ratanapakorn, O. and S. C. Sharma (2007): "Dynamic analysis between the US stock returns and the macroeconomic variables," *Journal Applied Financial Economics*, 17, 369–377.

Renyi, A. (1960): "On measures of information and entropy," in *In Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, 547–561.

Rodriguez, G., M. Leskovec, D. Balduzzi, and B. Scholkopf (2014): "Uncovering the structure and temporal dynamics of information propagation," *Network Science*, 1, 1–40.

Runge, J. (92): "Quantifying information transfer and mediation along causal pathways in complex systems," *Phys. Rev. E*.

Runge, J., J. Heitzig, V. Petoukhov, and J. Kurths (2012): "Escaping the Curse of Dimensionality in Estimating Multivariate Transfer Entropy," *Phys. Rev. Lett.*, 108.

Sandoval, J. L. (2014): "Structure of a Global Network of Financial Companies Based on Transfer Entropy," *Entropy*, 16, 4443–4482.

Sargan, D. (1964): *Wages and Prices in the United Kingdom: A Study in Econometric Methodology, Econometric Analysis for National Economic Planning*, vol. 16, London Butterworths.

Sato, J., E. Amaro, D. Takahashi, M. de Maria Felix, M. Brammer, and P. Morettina (2006): "A method to produce evolving functional connectivity maps during the course of an fMRI experiment using wavelet-based time-varying Granger causality," *NeuroImage*, 31, 187–196.

Savage, L. J. (1972): *The Foundations of Statistics, 2nd ed.*, Dover.

Schraudolph, N. (2004): "Gradient-based manipulation of non-parametric entropy estimates," *IEEE Trans. Neural Networks*, 14, 828–837.

Schreiber, T. (2000): "Measuring Information Transfer," *Phys. Rev. Lett.*, 85, 461–464.

Selva, D. and K. D. Hoover (2003): "Searching for the Causal Structure of a Vector Autoregression," *Oxford Bulletin of Economics and Statistics*, 65, 745–767.

Seth, A. K., A. B. Barrett, and L. Barnett (2015): "Granger Causality Analysis in Neuroscience and Neuroimaging," *The Journal of Neuroscience*, 35, 3293–3297.

Sewell, M. (2011): "Characterization of Financial Time Series," *UCL Research Note*, 11.

Shannon, C. E. (1948): "A Mathematical Theory of Communication," *The Bell System Technical Journal*, 27, 379–423, 623–656.

Simon, H. A. (1953): *Causal Order and Identifiability*, 49-74.

Sims, C. A. (1982): "Policy Analysis with Econometric Models," *Brookings Papers on Economic Activity*, 107–152.

——— (1986): "Are Forecasting Models Usable for Policy Analysis?" *Federal Reserve Bank of Minneapolis Quarterly Review*, 10, 2–15.

Smiech, S., M. Papiez, and M. A. Dabrowski (2015): "Does the euro area macroeconomy affect global commodity prices? Evidence from a SVAR approach," *International Review of Economics and Finance*, 39, 485–503.

Smith, A. (1776): *Wealth of Nation.*

Spirtes, P., C. Glymour, and R. Scheines (2000): *Causation, Prediction, and Search.*, MIT Press.

Staniek, M. and K. Lehnertz (2008): "Symbolic Transfer Entropy," *Phys. Rev. Lett.*, 100.

Steuer, R., J. Kurths, C. Daub, J. Weise, and J. Selbig (2002): "The mutual information: detecting and evaluating dependencies between variables," *Bioinformatics*, 18, 231–240.

Stigler, S. M. (1989): "Francis Galton's Account of the Invention of Correlation," *Statistical Science*, 4, 73–79.

Stramaglia, S., G. Wu, M. Pellicoro, and D. . Marinazzo (2012): "Expanding the transfer entropy to identify information circuits in complex systems," *Phys. Rev. E*, 86.

Suppes, P. (1970): *A Probabilistic Theory of Causality*, North-Holland Publishing.

——— (1993): "The Transcendental Character of Determinism," *Midwest Studies in Philosophy*, 18, 242–257.

Taamouti, A., T. Bouezmarni, and A. E. Ghouch (2012): "Nonparametric estimation and inference for Granger causality measures," Economics Working Papers we1217, Universidad Carlos III, Departamento de Economía.

Taskin, L. and M. de Nanteuil (2001): *Pespectives critiques en management. Pour une gestion citoyenne.*

Thompson, I. (1990): "Quantum Mechanics and Consciousness: A Causal Correspondence Theory," .

Thurner, S. and S. Poledna (2013): "DebtRank-transparency: controlling systemic risk in financial networks," *Scientific Reports*, 3.

Tibshirani, R. (1996): "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

Toben, B. (1974): *Space-Time and Beyond*, Dutton.

Tonzer, L. (2015): "Cross-border interbank networks, banking risk and contagion," *Journal of Financial Stability*, 18, 19–32.

Tsallis, C. (1988): "Possible generalization of Boltzmann-Gibbs statistics," *Journal of Statistical Physics*, 52, 479–487.

Tsamardinos, I., C. Aliferis, and A. Statnikov (2003): "Algorithms for Large Scale Markov Blanket Discovery," in *at the 16th International FLAIRS Conference*.

Tung, T., T. Ryu, D. Lee, and D. Lee (2007): "Inferring Gene Regulatory Networks from Microarray Time Series Data Using Transfer Entropy," in *Twentieth IEEE International Symposium on Computer-Based Medical Systems*.

van Nes, E. H., M. Scheffer, V. Brovkin, T. M. Lenton, H. Ye, E. Deyle, and G. Sugihara4 (2015): "Causal feedbacks in climate change," *Nature Climate Change*, 5, 445–448.

Vicente, R., M. Wibral, M. Lindner, and G. Pipa (2011): "Transfer entropy: a model-free measure of effective connectivity for the neurosciences," *J Comput Neurosci*, 30, 45–67.

Wada, T. and H. Suyari (2013): "The k-generalizations of stirling approximation and multinominal coefficients," *Entropy*, 15, 5144–5143.

Walker, H. M. (1975): *Studies in the history of statistical method*, Arno Press.

Wallinga, J. and P. Teunis (2004): "Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures," *American Journal of Epidemiology*, 160, 509–16.

Walras, L. (1954): *Elements of Pure Economics*.

Wang, J. and Z. F. Yu (2012): "Symbolic transfer entropy-based premature signal analysis," *Chin. Phys. B*, 21, 018702.

Weinhold, D. (1996): "Investment, Growth and CausalityTesting in Panels," *Economie et Prevision*.

White, G. (2013): *Medieval Theories of Causation*.

Wiener, N. (1956): "The theory of prediction," in *Modern mathematics for engineers*, McGraw-Hill, 1, 125–139.

Wigner, E. (1962): *Remarks on the Mind-Body Question, The Scientist Speculates*, vol. 284-302, Basic Books.

Winnie, J. A. (1977): *Foundations of Space-Time Theories*, University of Minnesota Press.

Wurzman, R. and J. Giordano (2009): "Mind, matter, neuroscience, and physics," *NeuroQuantology*, 7, 368–381.

Yang, J. and Y. Zhou (2013): "Credit Risk Spillovers among Financial Institutions around the Global Credit Crisis: Firm-Level Evidence," *Management Science*, 59, 2343–2359.

Yu, J. and R. Meyer (2006): "Multivariate Stochastic Volatility Models: Bayesian Estimation and Model Comparison," *Econometric Reviews*, 25, 361–384.

Yuan, K. (2005): "Asymmetric Price Movements and Borrowing Constraints: A Rational Expectations Equilibrium Model of Crises, Contagion, and Confusion," *Journal of Finance*, 60, 379–411.

Zaremba, A. and T. Aste (2014): "Measures of Causality in Complex Datasets with Application to Financial Data," *Entroy*, 16, 2309–2349.

Zeileis, A. (2004): "Econometric Computing with HC and HAC Covariance Matrix Estimators," *Journal of Statistical Software*, 11, 1–17.

Zeileis, A. and T. Hothorn (2002): "Diagnostic Checking in Regression Relationships," *R News*, 2, 7–10.

Zellner, A. A. (1979): *Causality and Econometrics ,Three Aspects of Policy Making: Knowledge, Data and Institutions*, vol. 10, Carnegie-Rochester Conference Series on Public Policy.

Zhang, H. J., J.-M. Dufour, and J. W. Galbraith (2016): "Exchange rates and commodity prices: Measuring causality at multiple horizons," *Journal of Empirical Finance*, 36, 100–120.

Zhang, X., X. Zhao, K. He, L. Lu, Y. Cao, J. Liu, J. Hao, Z. Liu, and L. Chen (2012): "Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information." *Bioinformatics*, 28, 98–104.

Zhao, W., E. Serpedin, and E. R. Dougherty (2006): "Inferring gene regulatory networks from time series data using the minimum description length principle," *Bioinformatics*, 22, 2129–2135.

Zinkernagel, H. (2010): *Causal fundamentalism in physics, EPSA Philosophical Issues in the Sciences*, Springer.

Appendix A.Testing causality in financial time series

A.1. Selected financial institutions

Name	Ticker	Type	Name	Ticker	Type
ING Bank NV	ING	SIFI	MGIC Investment Corp	MGIC	FS
BNP Paribas SA	BNP	SIFI	Wachovia Corp	WB	FS
Banco Santander SA	BME	SIFI	Western Union Co/The	WU	FS
HSBC Bank PLC	HSBA	SIFI	American Express Co	AXP	FS
Credit Agricole SA	ACA	SIFI	AXA SA	CS	I
Societe Generale SA	SG	SIFI	Allianz SE	ALV	I
Deutsche Bank AG	DBK	SIFI	Assicurazioni Generali SpA	G	I
Barclays Bank PLC	BARC	SIFI	Muenchener Ru.-Ges. AG	MUV	I
Royal Bank of Scotland PLC	RBS	SIFI	Aviva PLC	AV	I
UniCredit SpA	UCG	SIFI	Zurich Insurance Co Ltd	ZURN	I
Banco Bilbao Vizcaya Argentaria	BBVA	SIFI	Aegon NV	AGN	I
Credit Suisse Group AG	CSGN	SIFI	Legal & General Group PLC	LGEN	I
UBS AG	UBSG	SIFI	Swiss Reinsurance Co Ltd	SREN	I
Dexia Credit Local SA	DEXB	SIFI	Old Mutual PLC	OML	I
Nordea Bank AB	NDA	SIFI	Hannover Rueck SE	HNR1	I
Lloyds Bank PLC	LLOY	SIFI	SCOR SE	SCR	I
Standard Chartered Bank	STAN	SIFI	XLIT Ltd	XLGLF	I
Commerzbank AG	CBK	SIFI	Aon PLC	AON	I
JPMorgan Chase & Co	JPM	SIFI	Genworth Holdings Inc	GNW	I
Morgan Stanley	MS	SIFI	Hartford Financial Services Group	HIG	I
Goldman Sachs Group Inc	GS	SIFI	Lincoln National Corp	LNC	I
Bank of America Corp	BAC	SIFI	Prudential Financial Inc	PRU	I
Citigroup Inc	C	SIFI	Allstate Corp/The	ALL	I
Wells Fargo & Co	WFC	SIFI	Travelers Cos Inc/The	TRV	I
Banque Populaire	KN	B	Unum Group	UNM	I
Intesa Sanpaolo SpA	ISB	B	CNA Financial Corp	CNA	I
Cooperatieve Rabobank UA	RBK	B	MBIA Insurance Corp	MBIA	I
Banca Monte dei Paschi di Siena	BMPS	B	Safeco Corp	SAF	I
Erste Group Bank AG	EBS	B	Washington Mutual Inc	WAMUQ	I
Allied Irish Banks PLC	ALBK	B	American International Group	AIG	I
Banco Popular Espanol SA	POP	B	Chubb Corp/The	CB	I
Banco Popolare SC	BP	B	Manor Care Inc	HCR	MI
Banco Comercial Portugues SA	BCP	B	MetLife Inc	MET	MI
Banco de Sabadell SA	SAB	B	Cigna Corp	CI	MI
Banca Popolare di Milano Scarl	PMI	B	Aetna Inc	AET	MI
Danske Bank A/S	DANSKE	B	Humana Inc	HUM	MI
Svenska Handelsbanken AB	SHBA	B	UnitedHealth Group Inc	UNH	MI
DNB Bank ASA	DNB	B	Land Securities Group PLC	LSGOF	RE
Skandinaviska Enskilda Banken AB	SEB	B	Unibail-Rodamco SE	UL	RE
Mediobanca SpA	MB	FS	Hammerson PLC	HMSO	RE
Bear Stearns Cos LLC	BS	B	Klepierre	LI	RE
Charles Schwab Corp	SCHW	B	Gecina SA	GFC	RE
Federal Home Loan Mortgage Corp	FMCC	B	Compass Group PLC	CPG	RE
Federal National Mortgage Association	FNMA	B	ERP Operating LP	ERPOP	RE
Lehman Brothers Holdings Inc	LEH	B	HCP Inc	HCP	RE
Merrill Lynch & Co Inc	MER	B	Kimco Realty Corp	KIM	RE
H&R Block Inc	HRB	FS	Prologis	PLD	RE
Invesco Ltd (EUR)	IVZ	FS	Simon Property Group LP	SPG	RE
Loews Corp	L	FS	DDR Corp	DDR	RE
Marsh & McLennan Cos Inc	MMC	FS	AvalonBay Communities Inc	AVB	RE
Torchmark Corp	TMK	FS	Boston Properties LP	BXP	RE
Janus Capital Group Inc	JNS	FS	Capital One Financial Corp	COF	RE

Table A.1. This table reports the name, the ticker, the type of institution (SIFI for systematically important financial institutions, B for banks, FS for financial services which have been included in the set of banks, I for insurance, MI for medical insurance and RE for real estate) and the region of the selected financial institutions for the application part.

A.2. Topological measures

In this appendix, we define mathematically the different topological measures used in the application of chapter 1.

The transitivity also called clustering coefficient provides a measure of the degree to which nodes tend to cluster together or more precisely the extend to which the neighborhood of a node is connected. Considering a graph $\boldsymbol{G} = \{\boldsymbol{V}, \boldsymbol{E}\}$ composed of a set of nodes $\nu \in \boldsymbol{E}$ linked by a set of edges $e \in \boldsymbol{E}$, the coefficient for a node i is given by:

$$C_i = \frac{2 \mid \{e_{jk} : \nu_j, \nu_k \in N_i, e_{jk} \in \boldsymbol{E}\} \mid}{k_i(k_i - 1)}$$

with e_{jk}, the edge between the node ν_j and ν_k; N_i, the neighbors of the node ν_i; and k_i the number of neighbors of ν_i.

The betweeness centrality assesses the centrality of a node in a graph based on shortest paths. A high betweeness centrality indicates that a given node interacts with other nodes on relatively short paths. Its estimation is done by computing the number of shortest paths from all the other nodes of the network that pass through the considered node. For a given node ν_i, the measure is defined as follows:

$$BC_i = \sum_{\nu_i \neq \nu_j \neq \nu_k} \frac{\sigma_{st}(\nu_i)}{\sigma_{st}}$$

with σ_{st}, the total number of shortest paths from node ν_j to node ν_k; and $\sigma_{st}(\nu_i)$, the number of shortest paths from node ν_j to node ν_k that pass through ν_i.

The eigenvector centrality measures the influence of each node on the rest of the network based on the idea that connections to highly connected nodes contribute more. For a given graph $\boldsymbol{G}$ described by an adjacency matrix $\boldsymbol{A} = \{a_{jk} = 1 \forall e_{jk} \in \boldsymbol{E} \mid a_{jk} = 0\}$, the eigenvector centrality for a node ν_i is given by:

$$EC_i = \frac{1}{\lambda} \sum_{j \in \boldsymbol{G}} a_{i,j} x_j$$

with x_j and λ given by $\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$.

Appendix B. Effective Inference Algorithm

B.1. Information transfer measures

In this appendix, we describe two information measures that we used with our algorithm, that is, Granger causality and transfer entropy.

We first introduce the concept of Granger causality. Formally, a process Y is said to Granger cause another process X if the future values of X can be better predicted using the past values of X and Y rather than only the past values of X. The formal test developed by Granger (1969) relies on the linear regression model described by:

$$x_t = \beta_0 + \sum_{k=1}^{K} \alpha_l x_{t-k} + \sum_{k=1}^{K} \beta_l y_{t-k} + \varepsilon_t,$$

where the null hypothesis of Y that does not Granger cause X corresponds to the joint nullity of $\beta_k,\ \forall k \in \{1, \ldots, K\}$, leading to:

$$x_t = \beta_0 + \sum_{k=1}^{K} \alpha_k x_{t-k} + \eta_t$$

The extension to conditional Granger causality requires the introduction of conditions in the model. This leads to a similar null hypothesis corresponding to the joint nullity of β_l.

$$\begin{aligned} x_t = \quad & \beta_0 + \textstyle\sum_{k=1}^{K} \alpha_l x_{t-k} + \sum_{k=1}^{K} \beta_l y_{t-l} \\ & + \textstyle\sum_{i=1}^{n} \sum_{l=1}^{L} \delta_l^i w_{t-l} + \varepsilon_t \\ x_t = \quad & \beta_0 + \textstyle\sum_{k=1}^{K} \alpha_l x_{t-k} \\ & + \textstyle\sum_{i=1}^{n} \sum_{k=1}^{K} \delta_k^i w_{t-k} + \eta_t \end{aligned}$$

$\forall i \in \{1, \ldots, c\}$ with c the number of conditions.

We propose to rely on the usual Granger–Sargent statistic to test for the null hypothesis of joint nullity of β_k for the conditional and unconditional form of Granger causality.

$$\mathcal{F}_{Y\to X} = \frac{(SSR_2 - SSR_1)/K}{SSR_1/(T-(2+n)K)} \sim_{H_0} \mathcal{F}_{K,T-(2+n)K}$$

As for transfer entropy, this measure was first proposed by Schreiber (2000) and folows the main idea behind Granger causality. Indeed, its definition of causality or information transfer can be seen as the information gained on the future of a process X due to the past of another process Y, already knowing the past of process X. Transfer entropy defines this gain as the difference between two conditional Shannon entropies, an unrestricted entropy including X and Y, and a restricted entropy including only X. Assuming X and Y, two stationary Markov processes of order n and m, the transfer entropy determines to what extent the knowledge of process Y should reduce the uncertainty of process X, knowing the history of X itself.

$$\begin{aligned} T_{Y\to X} &= H(x \mid x_{t-n}) - H(x \mid x_{t-n}, y_{t-m}) \\ &= \sum_{x_t, x_{t-n}, y_{t-m}} p(x_t, x_{t-n}, y_{t-m}) \log_\alpha(p(x_t \mid x_{t-n}, y_{t-m})) \\ &\quad - \sum_{x_t, x_{t-n}, y_{t-m}} p(x_t, x_{t-n}, y_{t-m}) \log_\alpha(p(x_t \mid x_{t-n})) \\ &= \sum_{x_t, x_{t-n}, y_{t-m}} p(x_t, x_{t-n}, y_{t-m}) \log_\alpha(\tfrac{p(x_t \mid x_{t-n}, y_{t-m})}{p(x_t \mid x_{t-n})}) \end{aligned}$$

Taking the conditional form by introducing Z the set of conditions, the conditional transfer entropy is as follows:

$$\begin{aligned} T_{Y\to X} = H(x \mid x_{t-n}, z_{t-n}) \\ -H(X \mid x_{t-n}, \quad y_{t-m}, z_{t-n}) \end{aligned}$$

We estimate the four joint probabilities using quantile symbolization. To determine the true transfer entropy level, we need processes of infinite length so as to compute the right transition probabilities. To avoid the finite sample bias, we use the bootstrapped methodology proposed by Marschinski and Kantz (2002) to estimate the effective transfer entropy. The effective transfer entropy can simply be defined as follows:

$$ET_{Y\to X} \quad = T_{Y\to X} - T_{Y_{boot}\to X}$$

Variable $T_{Y_{boot}\to X}$ represents the transfer entropy with the series Y bootstrapped. By this bootstrapping of Y, all the statistical dependencies between the two series are

removed and therefore all the non-zero values of $T_{Y_{boot} \to X}$ demonstrate the presence of bias due to the small-sample effect and should therefore be removed from the transfer entropy estimation. This bias is determined at a specific level using a Kernel estimation based on multiple bootstrapped transfer entropy estimations and then removed from the original transfer entropy. This follows a two-step approach: first, estimate the n transfer entropy with the series Y bootstrapped to build a probability distribution function using a Gaussian kernel, and then obtain from the Kernel density the probability distribution of transfer entropy levels in the case of no information transfer, from which we derive the value of the bias at a specific level. The number of bootstraps used in the Kernel estimation for the effective transfer entropy has been set to 15 in this chapter. As for quantile symbolization, the data has been divided into three regions, two regions positioned at 46 percent on either side of the 50 percent quantile with a bandwidth of 4 percent, and the rest forming the third region.

B.2. Algorithm code description

The code starts with the pre-analysis step which removes all the non causal links via a for loop considering the $n(n-1)$ potential causal links from $lag = 1$ to $lag = lag_{max}$ (2-4). It then select the lag for which the information transfer is maximal (5). The pruning step follows, going from one node to another. Once a node X is selected, the algorithm starts by defining all the parents of X, called Y (8). Then one specific parent Y is selected and all its parents and children also connected to X is included in a set on which the transfer of information between Y and X is conditioned (9-13). If the causal link survives to this step (14), the algorithm moves to the greedy part which deals with the elimination of indirect links due to instantaneous relationships. The algorithm select all the parents of X sharing the same lag as Y (15). The selected nodes are then individually added to the set of conditions applied on the information transfer between Y and X (17-19). The nodes leading to the maximum reduction in the amount of information transferred is definitely added to the set of conditions (20). The remaining nodes are tested iteratively until no more reduction is observed. The resulting information transfer is kept in the network matrix Net (if > 0) and the algorithm goes to the next parent of X (23). When all parents of X have been treated, the algorithm moves to the next nodes until all nodes have been treated.

1:Input: time series data set $\mathbf{X} = (X^1, X^2, \cdots, X^n)$
2: **For** $k = 1$ to lag_{max} **do**
3: $\forall (i,j) \in \{1, \cdots, n\}^2, i \neq j, Net_{i,j,k} = IT(x^i_t, x^j_{t-k} \mid x^i_{t-\boldsymbol{k_x}})$
4: **end for**
5: $Net = \max(Net_{i,j,})$
6:**For** $i = 1$ to n **do**
7: **For** $k_p = 1$ to lag_{max} **do**
8: $\mathbf{Y}^i_{k_p} = \mathcal{P}_{k_p}(X^i)$
9: **For** $j = 1$ to $length(\mathbf{Y}^i_{t-k})$ **do**
10: $\boldsymbol{V}^{i,j}_{\boldsymbol{k^x_p}} = \left\{\mathcal{P}_{\boldsymbol{k^y_p}}(Y^{i,j}_{k_p}) \cap \mathcal{P}_{\boldsymbol{k^x_p}}(X^i)\right\}$
11: $\boldsymbol{W}^{i,j}_{\boldsymbol{k^{x\prime}_p}} = \left\{\mathcal{C}_{\boldsymbol{k^y_c}}(Y^{i,j}_{k_p}) \cap \mathcal{P}_{\boldsymbol{k^{x\prime}_p}}(X^i)\right\}$
12: $\boldsymbol{Z}^{i,j}_{\boldsymbol{k_{pc}}} = \left\{\boldsymbol{V}^{i,j}_{\boldsymbol{k^x_p}} \cup \boldsymbol{W}^{i,j}_{\boldsymbol{k^{x\prime}_p}}\right\}$
13: $Net(i, y^i_j, k_p) = IT(x^i_t, y^{i,j}_{t-k_p} \mid x^i_{t-\boldsymbol{k_x}}, \mathbf{z}^{i,j}_{t-\boldsymbol{k_{pc}}})$
14: **If** $Net(i, y^i_j, k_p) > 0$ **do**
15: $\mathbf{Y}^{i,greedy}_{k_p} = \mathbf{Y}^i_{k_p} \setminus Y^{i,j}_{k_p}$
16: **While** $\mathbf{Y}^{i,greedy}_{k_p} > 0$ **do**
17: **For** $h = 1$ to $length(\mathbf{Y}^{i,greedy}_{k_p})$ **do**
18: $IT^{i,j}_{greedy,h} = IT(x^i_t, y^{i,j}_{t-k_p} \mid x^i_{t-\boldsymbol{k_x}}, y^{i,h}_{t-k_p}, \mathbf{z}^{i,j}_{t-\boldsymbol{k_{pci}}})$
19: **End for**
20: $\mathbf{Z}^{i,j}_{\boldsymbol{k_{pci}}} = \mathbf{Z}^{i,j}_{\boldsymbol{k_{pci}}} \cup Which(\mathbf{Y}^{i,greedy}_{k_p} = \max(IT^{i,j}_{greedy}))$
21: $\mathbf{Y}^{i,greedy}_{k_p} = \mathbf{Y}^{i,greedy}_{k_p} \setminus$
$\left\{Which(\mathbf{Y}^{i,greedy}_{k_p} = \max(IT^{i,j}_{greedy})), Which(\mathbf{Y}^{i,greedy}_{k_p} = 0)\right\}$
22: **End While**
23: $Net(i, y^i_j, k_p) = IT(x^i_t, y^{i,j}_{t-k_p} \mid x^i_{t-\boldsymbol{k_x}}, \mathbf{z}^{i,j}_{t-\boldsymbol{k_{pci}}})$
24: **End if**
25: **End for**
26: **End for**
27:**End for**

B.3. Additional step treating Boolean relationships

C. For each node $S^{i,j}_{k_p} = N \setminus \mathcal{P}_{k_p}(X^i)$ with lag k_p not connected to X^i, estimate the following conditional information transfer:

* Condition first on the set of connected nodes $\mathbf{Y}^i_{k_p} = \mathcal{P}_{k_p}(X^i)$ with lag k_p and on their parents and children $\mathbf{Z}^i_{\boldsymbol{k_{pc}}}$

$$IT(x^i_t, s^{i,j}_{t-k_p} \mid x^i_{t-\boldsymbol{k_x}}, \boldsymbol{y}^i_{t-k_p}, \boldsymbol{z}^i_{\boldsymbol{k_{pc}}}) \geq 0$$

For each positive conditional information transfer, connect $S^j_{k_p}$ to X^i

* Condition then on each $S^{i,l}_{k_p}$ $\forall j \neq l$, on the set of connected nodes $\mathbf{Y}^i_{k_p} = \mathcal{P}_{k_p}(X^i)$ with lag k_p and on their parents and children $\mathbf{Z}^i_{\boldsymbol{k_{pc}}}$

$$IT(x^i_t, s^{i,j}_{j,t-k_p} \mid x^i_{t-\boldsymbol{k_x}}, s^{i,l}_{t-k_p}, \boldsymbol{y}^i_{t-k_p}, \boldsymbol{z}^i_{\boldsymbol{k_{pc}}}) \geq 0 \quad \forall j \neq l$$

For each positive conditional information transfer, connect $S_{k_p}^{i,j}$ and $S_{k_p}^{i,l}$ to X^i

B.4. Characterization of simulated network topology

We propose, in this appendix to analyze the connectivity distribution obtained with our network simulator. As stated by Battiston et al. (2010), the study of the topology of financial networks shows clearly the features of scale-free networks. Indeed, following the definition of a scale-free network, they verify that the in-degree and out-degree probability density functions display a power law for financial network. The objective of this chapter is the proper inference of effective financial network. Therefore, in order to be able to properly define the ability of the algorithm to infer the topology of a financial network, the simulated networks should display the main features of a true financial network. We verify this assumption by looking at the distribution of the in-degree inside the simulated network. We differentiate the in-degree for the causal links with a lag of one and two periods and for the instantaneous relationship. The three probability density functions are then estimated based on simulated networks composed of 20 nodes with up to 6 lagged connections per node and one instantaneous connection per node. The results of Fig.B.1 confirm the powerlaw behavior of the distribution function for the three types of connections considered in the simulation. It implies that the results of the Monte-Carlo simulations may be extended to real financial network.

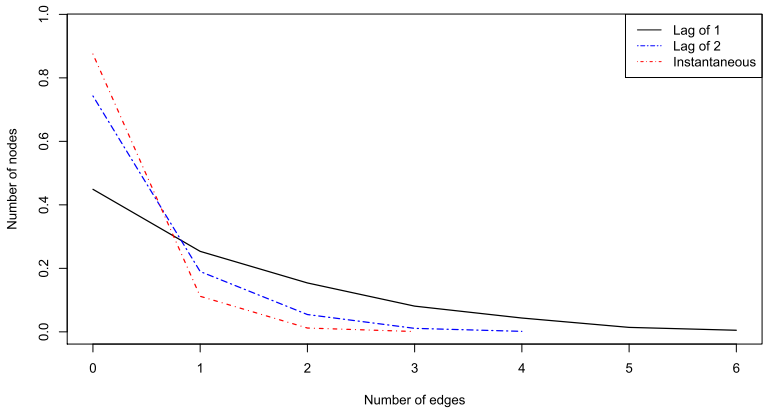


Figure B.1. The figure provides the distribution of the nodes connectivity for the instantaneous links, and lag 1 and lag 2 links.

B.5. Characterization of simulated network topology

In this appendix, we demonstrate that the good results obtained by our algorithm can't be replicated by a simple ordering procedure based on the pre-search step. Indeed, we may wonder if the proposed algorithm does not simply select the links presenting the highest information transfer level through some sort of ordering procedure on the results of the pairwise causality measures. We show here that it is not the case by considering three possible approaches involving an ordering procedure to select the direct links which differs only by the threshold applied to define the true direct links. The three approached are based on the ordering of the adjacency matrix obtained at the end of the pre-search step. For the first approach, there is no explicit threshold but we select in the ordered adjacency matrix the connections with the lowest p-value, the number of selected connections equals the number of direct links inferred by our algorithm. We then look at the number of connections in the adjacency matrix obtained by our algorithm, with a p-value higher than the highest p-value of the selected connections in the ordered adjacency matrix from the pre-search step. The obtained number of connections represents the connections inferred by our algorithm not present in the selection based on the ordering procedure. This number is a good proxy of the distance existing between the network inferred by our algorithm and the one obtained by a simple ordering procedure. Considering the good results of our algorithm, it shows also the ability of this ordering procedure to infer the true underlying network.

The second approach is based on the definition of a threshold. The threshold is defined using either the network inferred with our algorithm or the true underlying network. Indeed, we take respectively the highest p-value in the adjacency matrix obtained with our algorithm or the highest p-value in the pre-search step adjacency matrix considering only the connections shared by this adjacency matrix and the one of the true underlying network. Every connections from the ordered adjacency matrix with a p-value lower than the threshold are selected. We estimate then the difference between the number of selected connections and the number of connections obtained with our algorithm in the case of the first threshold and the number of connections in the true underlying network in the other case. This number is again considered as a proxy of the distance between the network obtained with the ordering procedure and the one obtained with our algorithm or the true one. Nevertheless, in contrast with the number obtained in the previous paragraph, here we do not consider the number of connections not included in the network inferred with our algorithm, but rather the number of additional connections compared to the network inferred with our algorithm, i.e. the number of additional indirect links not treated.

Table B.1. presents the results for the three different approaches based on the sim-

ulations of 100 networks with 20 nodes and up to 6 incoming edges represented by times series of 300 observations. The results are expressed in percent by dividing the numbers of true links detected by the number of connections inside either the network inferred with our algorithm (first two) and the true underlying algorithm (last two). As can be seen, the results obtained with our algorithm can't be replicated by using a simple ordering procedure even when considering the true underlying network to define the threshold use for the direct links selection. Indeed, the first approach provides the right number of connections but not the correct ones, with only 15.9 percent of the inferred link corresponding to true relationships for the Granger based estimation. The second and third approaches have better results but still much lower than our algorithm in terms of true positive rate with respectively 53.8 and 38.1 percent of the true underlying relationships detected for the Granger based estimation.

True positive rate	Number threshold	Algorithm based Threshold	True network based threshold	ENIA Algorithm
TE ENIA	30,2%	63,1%	30,1%	83,5%
GC ENIA	15,9%	53,8%	38,1%	99,6%

Table B.1. This table reports for the different ordering methodologies the true positive rate which has to be compared to the true positive rate of effecitve network inference algorithm (ENIA) provided in the last column

Appendix C. Cross-country Information Transmissions

C.1. Simulation exercise

This appendix consists in a Monte Carlo simulation exercise allowing the comparison between the model developed in chapter 4, i.e. Multichannel Markov Switching Granger causality and the one proposed in chapter 3 called Multi-Channel Granger Causality. For both models, we consider the results obtained with the algorithm including or not a pre-search step. This additional step is used to reduce the complexity of the estimation by reducing the number of possible channels to be taken into account. The step consists in the estimation of a bivariate Granger Causality for each channel in order to include in the further steps only the channels for which an average causal link is observed for the whole sample. The effect of this pre-search step should therefore be greater on the Multi-Channel Markov Switching Granger Causality (MCMSGC) as the number of parameters to be estimated increases sharply with the number of channels to be considered.

The data generating process used here, simulate the interaction between two systems composed of three different variables. The system $\boldsymbol{X}$ receives information from the system $\boldsymbol{Y}$ through three possible and exclusive channels, meaning that at each time step only one channel is used for the transfer of information. The times series of the system $\boldsymbol{X}$ are defined as follow for the active channels:

$$\begin{aligned} x_t^1 &= \alpha x_{t-1}^1 + \beta y_{t-1}^1 + (1-\alpha-\beta)\varepsilon_t^1 \qquad \text{when } s_t = 0 \\ x_t^2 &= \alpha x_{t-1}^2 + \beta y_{t-1}^2 + (1-\alpha-\beta)\varepsilon_t^2 \qquad \text{when } s_t = 1 \\ x_t^3 &= \alpha x_{t-1}^3 + \beta y_{t-1}^1 + (1-\alpha-\beta)\varepsilon_t^3 \qquad \text{when } s_t = 2 \end{aligned}$$

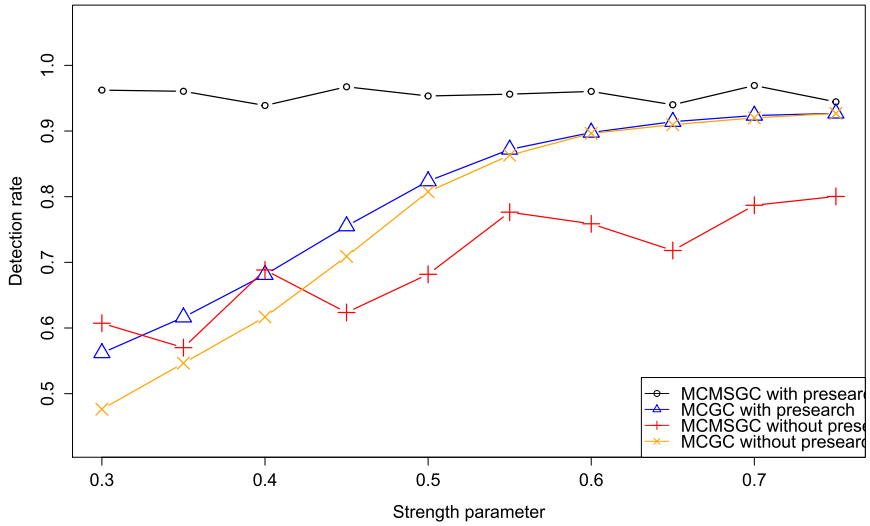
with ε_t^i a Gaussian white noise with zero mean and unit standard deviation.

The state of nature s_t is defined using a Poisson process which determines the switching frequency. The non connected series X^i and all the time series Y^j follow the usual

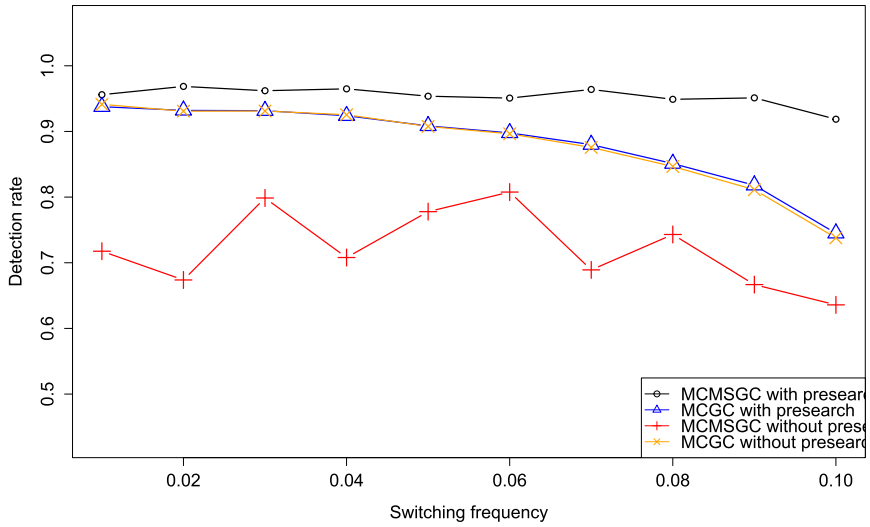
autoregressive process (AR):

$$x_t^i = \alpha x_{t-1}^i + (1-\alpha)\varepsilon_t^i$$
$$y_t^j = \alpha y_{t-1}^j + (1-\alpha)\eta_t^j$$

where $i \in \{1,2,3\}$ depending on the period and $j \in \{1,2,3\}$.



(a)Detection rates in terms of relationship strength



(b)Detection rates in terms of switching frequency

Figure C.1. The figure (a) reports the detection rates for both the MCGC and MCMSGC with and without a pre-search step as a function of the strength of the causal relationship between the variables of the system Y and X. The figure (a) reports the detection rates for the same models but here in terms of switching frequency with the intensity of the Poisson process reported in the X-axis.

We tested our DGP for different parametrization by changing the switching frequency, i.e. the intensity of the Poisson process, and the strength of the causal link represented by the β parameters. In the first case, the intensity varies from 1 regime switch each 10 observations to 1 regime switch each 100 observations with a fixed β of 0.6. In the second case, the switching frequency is fixed at 1 regime switch per 50 observations and the strength parameter β ranges from 0.25 to 0.7. Fig.C.1 (a) and (b) give the rate of detection of the true underlying channels for both models, with and without pre search step, and for both parametrization. These results regroup for each set of parameters, 100 simulated data-sets of 1000 observations. At each time step, the most active channel for the MCMSGC is determined by selecting the channel providing the highest inferred probability of information transfer for the considered time step (see Sect.4.1). For the MCGC, this information is directly provided by the algorithm. We then compute the detection rate by comparing the inferred active channels with the true underlying active channel. As can be seen from Fig.C.1 (a), the MCMSGC including a pre-search step shows a higher detection rate compared to the other models when the strength of the causal link reduces with a detection rate stable for all the value of the β parameters considered in the simulation. We have a similar trend when testing different switching frequencies, with a stable detection rate for the MCMSGC including a pre-search step and a decreasing one when the switching frequency increases for the MCGC. We see nevertheless, with both parametrization, that the MCMSGC gives lower detection rates compared to the other methods when considering all the possible channels. This arises from the fact that the number of parameters to be estimated increases quickly with the number of channels and therefore reduces the ability of the maximization algorithm to find an optimal solution. It highlights the importance of the filtering procedure when using the MCMSGC as it improves both the computing speed and the detection rates.

Appendix D. Summary of main contributions and publication status

Chapter 1: Testing causality in financial time series

The first chapter explores the possibility to better identify causal relations between financial series using empirical tools derived from information theory. Indeed, a large literature in physics and information theory shows that close links exist between the informational entropy transfer measures (Shannon, 1948; Tsallis, 1988) and that of causality developed by Granger (1969) which is commonly used in economics and finance. Based on simulations of possible data generating processes (DGPs) for returns with causal relationship both in mean and variance, we study the properties of those different measures based on rigorous testing beyond the Gaussian equivalence case put into lights by Barnett et al. (2009). Results show that for nonlinear and/or non Gaussian DGPs, transfer entropy measures perform better in terms of power and size and thus better capture causal relations, when they exist, between two series. Having documented the added value of informational measures to the identification of causal relationships in specific cases, we explore, in the context of network inference, the ability of both measures to identify the relationships existing between European and U.S. Systematically Important Financial Institutions (SIFIs). We propose to analyze how network characteristics affect the systemic fragility of the financial system and also determine the impact of the surrounding environment of the considered network on the information content of its topology, addressing empirically the question of optimal network sampling.

"Testing causality in financial time series: A comparison between Granger and information theoretical based approaches" submitted to the Journal of Financial Econometrics.

Chapter 2: Effective Network Inference Algorithm

Network representation has steadily gained in popularity over the past decades. In many disciplines such as finance, genetics, neuroscience or human travel to cite a few, the network may not directly be observable and needs to be inferred from time-series data, leading to the issue of separating direct interactions between two entities forming the network from indirect interactions coming through its remaining part. Drawing on recent contributions proposing strategies to deal with this problem such as the so-called "global silencing" approach of Barzel and Barabasi or "network deconvolution" of Feizi et Al.(Nature Biotechnology, 31 (2013)), we propose a novel methodology to infer an effective network structure from multivariate conditional information transfers. Its core principal is to test the information transfer between two nodes through a step-wise approach by conditioning the transfer for each pair on a specific set of relevant nodes as identified by our algorithm from the rest of the network.The methodology is model free and can be applied to high-dimensional networks with both inter-lag and intra-lag relationships. It outperforms state-of-the-art approaches for eliminating the redundancies and more generally retrieving simulated artificial networks in our Monte-Carlo experiments. We apply the method to stock market data at different frequencies (15 minutes, 1 hour, 1 day) to retrieve the network of US largest financial institutions and then document how bank's centrality measurements relate to bank's systemic vulnerability.

"Effective network inference through multivariate information transfer estimation" published in Physica A, 499 (2018), pp. 376-394, 10.1016/j.physa.2018.02.053

R package for the Effective Network Inference Algorithm available at:
https://github.com/chdahlqvist/enia/releases

Chapter 3: Multichannel Information Transfer Estimation

The past decade has seen the development of new methods to infer causal relationships in biological and socio-economic complex systems, following the expansion of network theory. Nevertheless, the standard estimation of causality still involves a single pair of time dependent variables which could be conditioned, in some instance, on its close environment. However, interactions may appear at a higher level between parts of the considered systems represented by more than one variable. We propose to study these types of relationships and develop a multi-channel framework, in the vein of Barrett and Barnett (Phys. Rev. E, 81 (2010)), allowing the inference of causal relationships between two sets of variables. Each channel represents the possible in-

teraction between a variable of each sub-system. Based on this new framework, we develop two different multi-channel causality measures derived from the usual Granger causality to account for linear interactions and from the concept of transfer entropy for nonlinear contribution. Our measures provide different information about the inferred causal links: the strength of the global interaction between the two sub-systems, the average frequency of the channel switches and the channel contributing the most to the information transfer process for each time step. After having demonstrated the ability of our measures to infer linear as well as nonlinear interactions, we propose an application looking at the U.S. financial sector in order to better understand the interactions between individual financial institutions, as well as parts of the financial system. At the individual level, the considered channels between financial institutions are expressed both in terms of spectral representation using wavelet transform and probability distribution using quantile regressions. Beyond the application presented in the chapter, this new multi-channel framework should be easy to implement in other fields of complex systems science such as neuroscience, biology or physics.

Chapter 4: Cross-country information transmissions

This chapter examines the interrelationships among 9 advanced economies using a novel multichannel approach to investigate, beyond the usual causal relationships, the time-varying dimension of the channels conveying causal relationships. The model is derived from the well-known Markov Switching setting and account for systems described by multiple variables. A Markov switching causality measure is adapted to account for information transmissions between distinct multivariate systems. Each country is described by 5 different fundamental variables reflecting its state. Our multichannel causality measure is then applied on these sets of time series to determine, over time, the main channels through which the information is transmitted between the different countries. In a second step, we investigate the relationships existing between these countries and the commodity markets and look at the possible use of the commodity markets as an indirect channel of information transmission between countries.

"Cross-country information transmissions and the role of commodity markets: A multichannel Markov switching approach" accepted for publication in Plos One on August 3, 2018.

.