Energy derivative markets and systemic risk

Delphine Lautier

Franck Raynaud

Report 2009-2010

List of Figures

Glossary

Mnemonic

Energy Futures Contracts
NCL : American Crude Oil
LLC : European Crude Oil
NHO : Heating Oil
LLE : European Gas Oil
NNG : American Natural Gas
LNG : European Natural Gas

Financial Futures Contracts
IED : Eurodollar, interest rate futures contract
NGC : Gold
CEU : Exchange rate, Dollar-Euro
ISM : Mini S&P500

Agricultural Futures Contracts
CC : Corn
CW : Wheat
CS : Soy Bean
CBO : Soy Oil

Methodology

$\boldsymbol{MST}$: Minimum Spanning Tree
$\boldsymbol{\rho_{ij}}$: Correlation Coefficient
$\boldsymbol{C^T}$: Mean Correlation
$\boldsymbol{\sigma_C^2}$: Variance of the Mean Correlation
$\boldsymbol{d_{ij}}$: Distance
$\boldsymbol{A_i}$, $\boldsymbol{B_i}$: Allometric Coefficients
$\boldsymbol{\eta}$: Allometric Exponent
$\boldsymbol{S_i}$: Node's Strength
$\boldsymbol{\mathcal{L}}$: tree's Length
$\boldsymbol{l(v_i)}$: Level of the Vertex v_i
$\boldsymbol{L}$: Mean Occupation Layer
$\boldsymbol{S_R}$: Survival Ratio

Executive Summary

The objective of this research project is to study systemic risk in energy derivatives markets. Concerns about systemic risk have recently grown in financial markets, notably in energy markets. The latter become more and more integrated, both as regards each other and as regards other markets. For some years now, price increases in energy commodities have often been invoked to explain that of soft commodities like corn, wheat, rape or sugar cane. Moreover, since commodities are considered as a new class of assets, they are intensively used by portfolio managers for diversification purposes. Consequently, part of the price movements recently recorded on commodity markets might be explained by external events like the fluctuations recorded in stock prices or dollar exchange rates.
The financial literature has investigated the question of systemic risks on commodity markets in different ways: herding behavior, co-integration techniques, spatial and temporal integration, etc. These studies of the way shocks appear in financial markets, and the way they disseminate among other markets generally take into account one or two of the dimensions of integration, which can be examined according to three different points of view: space, observation time, and maturity. The analysis of the relationships linking different markets, for a single commodity being simultaneously negotiated in several places, has to do with the spatial dimension of integration. When the focus is placed on how the relationships between several commodities evolve over time, it is the temporal dimension of integration which is examined. Lastly, it is possible to consider a third dimension, related to the term structure of commodity prices, i.e. the relationships linking, at a specific date, several futures contracts with different maturities.
So, while it is highly likely that integration and systematic risk are progressing in energy derivatives markets, the previous studies always gave preference to one or two dimensions of the integration. Such a conclusion naturally leads to the following question: "why not try to study the three simultaneously?" Part of the answer lies probably in the fact that its is not that easy. Taking into account simultaneously the three dimensions of integration implies the possibility of, firstly being able to collect a huge amount of data, secondly being able to analyze the data, and thirdly taking into account the possible complexity of the system described by the data.

This report presents the results of our investigations on that subject, one year after we started our research. It contains seven chapters. The first chapter is dedicated to a general introduction, the presentation of our main objectives and the scientific relevance of our project, first for the French Energy Council, second from an academic point of view. The aim of the second chapter is the presentation of the database and its main characteristics. In the third chapter we provide the methodology and tools used to measure the integration of the markets. In the fourth chapter we expose the results of the empirical analysis of the energy markets. In the fifth chapter we provide for a systemic approach for all markets in the spatial, maturity and three dimensions. In the sixth chapter we present a short review of the so-called Ising model which is a major model of collective behavior. These pages are taken as an opportunity to expose important features of statistical physics as well as the concept of the minimal model. We then present our conclusions and explain the policy implications of our study.

In the first chapter of this report, we underline the objective of our research project, which is the three-D investigation of systemic risk in energy derivative markets. This research aims at enhancing the understanding of market mechanisms. To do so, we intend to transfer technical skills from physics to economics in order to extract relevant information from the three dimensional space (time, space and maturity). We then explain why systemic risk seems an interesting investigation field for statistical physics. Systemic risk can be briefly defined as the sudden manifestation of a dysfunction occurring on a large scale and resulting from a strong integration of the markets. At a microscopic scale, the interactions between the individual actors operating in such a market draw a complex network which is partly responsible for the integration and may lead to non predictable movements spreading throughout a whole economic sector. In other words, such a network might give rise to the emergence of a global self organized behavior resulting from local interactions. Such a phenomenon appears in physics when a system changes from one state to another one, for example when water (liquid phase) turns into ice (solid phase). According to the authors of this report, such an analogy between finance and physics has a high potential, because physicists developed many theoretical or numerical tools allowing them to investigate the behavior of complex systems.
We also found it important to underline the relevance of such a subject for an organization like the French Energy Council. This point was underlined at the beginning of this executive summary, through the presentation of the objectives of this report. So let us just underline here that it is crucial, today, to know how strongly, in the spatial as well as in the maturity dimension, energetic markets are self-integrated or integrated with other derivatives markets, and whether or not they could be affected by systemic risk. If it were the case, this risk would need to be characterized and quantified.
Finally, we give our point of view on the scientific and academic pertinence of the project. From an academic point of view, the aim of this project is to connect two different fields of investigation: finance and physics. For twenty years, the economy was of interest to an increasing number of physicists. More precisely, minimal models, that is to say, models

relying on very simple assumptions, have recently offered fruitful insights into the understanding of the role of simple mechanisms in the emergence of complex patterns or information transfers. Thus we hope that, as was the case for biology, dynamical systems and also economy, we will be able to construct a new model capturing the main features of co-movements in commodity derivative markets. Before reaching such a long-term objective, we needed to explore the empirical relationships linking energy derivative markets. This is the aim of this report.

Over the past year, we first realized an important effort with the data. Extracting the data and analyzing the database represented a huge amount of work and time. During this period, we collected nearly two million daily data and analyzed more than six hundred fifty thousand prices. In the second chapter, firstly we present the markets selected for the empirical study, namely energy, agriculture and financial assets. On the basis of the Futures Industry Association's monthly volume reports, we retained those contracts characterized by the largest transaction volumes. The choice of these three sectors is motivated by recent observations of changes in financial markets. Commodities derivatives are more and more integrated: within the sector of commodities and also with other markets. Furthermore, commodities are considered as full fledged assets used for diversification purposes by portfolio managers. Consequently, price increases in the given commodities has in part been associated with price fluctuations of energy commodities or could be explained by a priori foreign events like the drop in equities or exchange rates.
On these markets, we collected settlement futures prices as well as opening prices, transactions volumes and open interest. We however only used futures prices for this part of the research. We leave the information provided by transaction volumes and open interest for further investigations, as well as, if necessary, the extension of the database. Then we expose the main characteristics of our database and present a brief overview of the behavior of futures prices over our study period. Finally, we propose a discussion on the seasonal behavior of the commodities under examination. We tried to identify a seasonal pattern in the futures prices of petroleum products, through - among others - the Discrete Fourier Transform method. However, so far, the results have not proved convincing.

In the third chapter, we suggest a method which enables us to measure the integration of the markets empirically. Among the different tools available in physics, one seems naturally relevant to study the three dimensional integration of derivative markets: the graph theory. A graph is a mathematical representation of pair wise relations within a collection of discrete entities. A financial market is composed of a large number of assets, such as equities, bonds or derivative products, which are linked together with different intensities. Thus a representation of the financial markets through the prism of graphs could be interesting, as economic information will emerge from the topology of the graphs. Among the different graphs available, we chose to use the minimum spanning tree because of its uniqueness and simplicity. The latter indeed provides the shortest path linking the nodes of the graph to each other. Thus, it reveals the geometric aspects of correlations between the different entities under examination. As the markets are intrinsically time dependant, it is necessary to study the dynamical properties of both the correlations and the minimum spanning trees. We examine the time dependant properties of the correlations, as well as the markets' strength, which gives information on how much a market is correlated to the others. A further characterization is obtained by quantifying the degree of randomness of the minimum spanning trees. The former information is given by the so-called allometric exponent, which indicates what kind of path a shock can follow to spread through to other markets. We also study the robustness of the tree topology in respect of market events using the survival ratio. This gives the fraction of survival links and shows the importance of rearrangements between two consecutive minimum spanning trees.

In the fourth chapter, we apply this method to our data and carry out empirical tests in the energy markets. The maturity dimension is explored for two markets: American crude oil and heating oil. Before proceeding with these tests, we assumed that they would reflect the presence of the Samuelson effect on the data. In an ideal case, the graph representing the maturities would be perfectly organized, ranging regularly from the first to the last delivery date. In other words, the topology of the minimum spanning trees would be linear. We were nevertheless surprised to observe our results on heating oil. The maturities between one and 36 months are perfectly ordered. As far as crude oil is concerned, the results are less perfect, but still very interesting. We suppose that what we observe on crude oil is partially the results of the maturation process of the market over time, but we intend to make further investigations on this market before coming to a more definite conclusion. As far as spatial integration is concerned, the results are also very interesting. In order to give greater insight on the empirical relationships linking five markets, we carried out several series of tests, on different maturities: one, two, three, six and twelve months. It so happens that the topologies of the graph change with the maturity under consideration: they are the same for the one, two and three months' maturities, but not for the six and twelve months' maturities. In each case however, the links between markets, via the representation of the minimum spanning trees, have an economical interpretation that satisfies intuition. We interpret this result as a positive test for the relevance of our method and its application to derivative markets. Comparing the results obtained with the different maturities, we found that the strength of the integration increases with the maturity. The latter result is original and has not as yet been mentioned in other studies. In particular, the authors of [19] identified the spatial links between oil markets but omitted the information provided by the maturity dimension.

The fifth chapter provides a systemic analysis of all markets selected for the study and performs empirical tests in the spatial dimension, maturity dimension as well as on the three dimensions. To the best of our knowledge the maturity dimension and the three

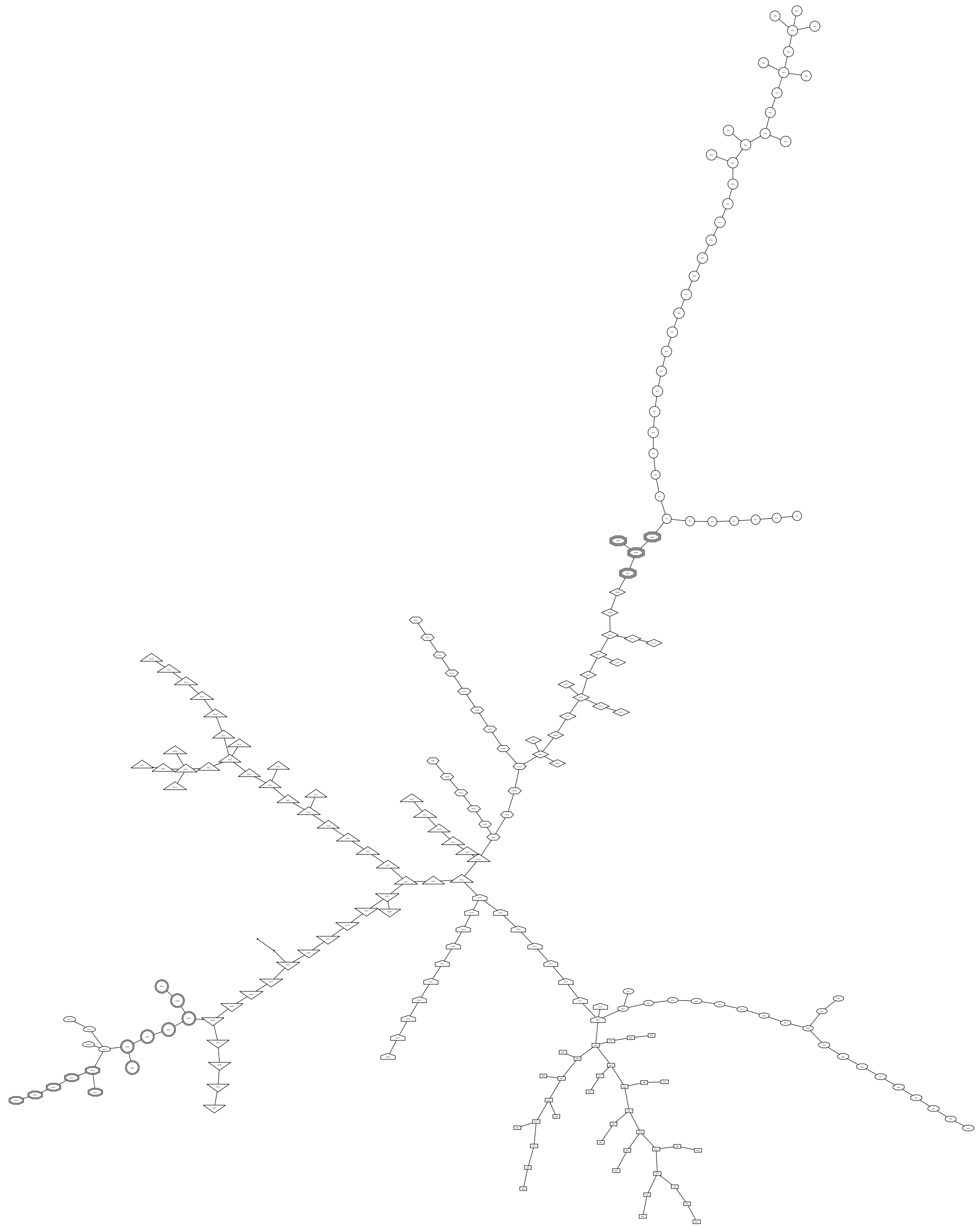

Figure 1: MST for the three-dimensional analysis, 06/27/2000-08/12/2009. The different futures contracts are represented by the following symbols: empty circle: *IED*, point: *ISM*, octagon: *LNG*, ellipse: *LLE*, box: *NNG*, hexagon: *LLC*, triangle: *NCL*, house: *NHO*, diamond: *NGC*, inverted triangle: *CBO*, triple octagon: *CEU*, double circle: *CS*, double octagon: *CW*, egg: *CC*. For a given futures contract, all maturities are represented with the same symbol. The distance between the nodes is set to unity.

dimensions have not been studied previously.
The first part of the study is devoted to the visualization of the MST of the three sectors simultaneously. The visualization of the MST firstly gives evidence of a star-like organization of the trees in the spatial dimension, whereas the maturity dimension is characterized by chain-like trees. These two topologies merge in the three-dimensional analysis. A typical three-dimensional MST is presented in Figure 5.2. In order to help the visualization of the tree all maturities are given with the same symbol and distance between the nodes is set to unity. The star-like organization that appears on this Figure reproduces the three different sectors under examination: energy, agriculture and finance. We emphasize that the linear shape of the trees observed in the energy sub-group is also valid for the agriculture as the financial sectors.
American and European crude oils simultaneously occupy the center of the graph and ensure the links with agricultural products and financial assets. Thus our first important conclusion is that crude oil is the best candidate for transmitting prices shocks. If such a shock appears at the periphery of the graph, unless it is absorbed quickly, it will necessarily pass through crude oil before spreading to the other energy products and sectors. Moreover, a shock will have an impact on the whole system that will be all the greater the closer it is to the heart of the system.
In the third section of the chapter we explore the dynamical properties of the sys- tem. We have seen that the level of integration is greater in the maturity dimension than in the spatial one. This result reflects the fact that arbitrage operations are far easier with standardized futures contracts written on the same underlying asset than with products of different natures. The analysis of how this level evolves over time shows that integration increases significantly on both the spatial and maturity dimensions. Such an increase can be observed on the whole prices system. It is even more evident in the energy sector (with the exception of the American and European natural gas markets) as well as in the agricultural sector. The latter is highly integrated in the end of our period. Lastly, as far as the financial sector is concerned, no remarkable trend can be highlighted. Thus, as time goes on, the heart of the price system becomes stronger whereas the peripheral assets do not change place significantly.

The sixth chapter of this report is devoted to a theoretical study focusing on the way we could build a minimal model. This chapter should naturally be read as an initial attempt to think about such a model rather than to build it.
We suggest the use of recent methods originated from statistical physics, hoping that the tools and ideas previously developed for complex systems will also be relevant for financial markets. If we indeed reformulate from a physicists point of view the question of systemic risks and the spreading of shocks among markets, the following question arises: “How can global collective behavior occur from local inter- actions?”. We have the intuition that concepts originated from the physics of phase transition and critical phenomena, such as collective behavior, scale invariance and renormalization may be useful to con-

sider that question theoretically. Even if financial markets are not ruled by natural laws, they cannot escape the fascinating ubiquity of collective motion. Like statistical physics, economics aims to describe the equilibrium and dynamics of a large number of entities, such as economic agents, which interact with each other. In both cases these interactions lead to sudden collective phenomena.

We first point out the spirit of a minimal model. Such a model indeed results from a choice to be made between two contradictory objectives: first, the formulation of equations which remain simply enough and secondly the need to capture all the main features of the phenomena under examination.

We then present what a physicist means by the concept of collective behavior and what kind of tools or measures exist in order to represent and to detect such behavior. The seminal model on that subject is the so-called "Ising model", which describes the spontaneous magnetization in a magnetic material. A magnetic field results from the collective behavior of microscopic entities. On a microscopic scale, each electron carries a physical quantity, the spin. The latter is more or less the same for all electrons and creates a macroscopic magnetization. The Ising model allows the interactions between the spins to be determined. We firstly present it in a one dimensional space. Naturally, higher dimensions are needed. Unfortunately, to deal with a high dimensional system - or a more complicated model - some approximations are necessary. A very important and useful approximation method is the mean-field approximation. The mean field method replaces, within a population, the influence that an individual's neighbors might have on that specific individual, by their average impact. This chapter of the report must be considered as the reflection of our state of advancement so far. Indeed, the core of our next project is the development of such a model inspired by statistical physics.

At the end of this report, we intend to carry on our investigations in the following directions: We will first expand our empirical analysis in the maturity dimension. We observed some regular and recurrent correlation patterns in the maturity dimension that need deeper investigation and might reflect some universal mechanisms of price's curve segmentation. The latter result would be of interest to both the financial and physics communities, whilst up until now the literature has greatly omitted this important feature.

We also aim to enrich our results with an analysis of the transaction volumes and the open interests. Firstly we could use the same graph theory formalism in order to analyze trees of correlated transactions and open interests. We could then try to consider returns fluctuations weighted by volumes and/or open interests. Thus, such questions as the robustness of the centrality of crude oil with respect to interest rates will be addressed.

Another field of empirical investigation will be the study of shocks affecting the markets. In particular we could determine the topological properties of trees during strong events, as the nature of the affected links or the time required to go back to initial configuration.

The main part of our further studies will however be devoted to modeling the collective behavior of derivatives energy markets and systemic risk. We aim to use theoretical concepts inspired by statistical physics, especially the use of the minimum model. Our former results will lead us to establish fundamental hypotheses and will act as a guideline in developing the model. In particular we want to determine, within a single framework, the mechanisms of price's term structure (which lead to linear tree), as the interactions between markets (which lead to star-like tree). Once the two typical shapes have been achieved, we will be able to use the model in order to understand the complex process of branching that appeared while the three dimensions of integration, namely where, in their price's curves, two different derivatives markets are most correlated. A major contribution of this part of the modeling will be to understand how (and where) links appear between markets. Secondly, we will proceed to a shock analysis and consider such questions as the existence of the tree's shape that encourage or prevent strong shocks, the required number of markets involved in an event to propagate it, or the amplitude of shocks that can involved in systemic risk.

Contents

Chapter 5
Analysis of the systemic risk in the spatial, maturity and spatio-maturity dimensions

Chapter 6
Building a minimal model

1

Issues of this research project

Contents

In this chapter we present the objective of our project, that is to say the three dimensional investigation on systemic risk in energy derivative markets. Then, we expose a physicist point of view of systemic risk and explain why the latter seems an interesting investigation field for statistical physics. Finally, we give our point of view on the scientific and academic pertinence of the project.

1.1 Objectives

The objective of this research project is a statistical physics approach of systemic risk occurring in commodities derivative markets. In a usual framework, the study of shock prices arising in a market and propagating in one or more markets is usually considered according three different dimensions. There is one spatial dimension and two temporal. The spatial aspect deals with commodities traded simultaneously in different geographical places. As we are working with future contracts, we need to consider two temporal dimensions. The first one is the prices changing over time and the second one is the prices term structure, *ie* how prices change with maturity.

The recent works in the area of energy commodities show a higher and higher market integration, involving a most probable systemic risk. Until now, most of the economical and financial investigations focus on one of the integration [14, 4] at the expense of a global vision of the three dimensions.

In the last decade physicists paid a lot of attention to economics. They used a wide spectrum of methods, models and tools of modern physics as non-linear physics [13], stochastic process [3], critical phenomena or networks [18, 31, 2]. Due to the large quantities of data available on commodities, we will consider the problem of systemic risk within the framework of statistical physics and complex systems. This innovative approach will need to resort to a novel vision of markets mechanisms. Different technical skills will be transferred from physics to economics in order to extract relevant informations from the three dimensional space (space, time and maturity).

1.2 Interest of statistical physics for the question of systemic risk

Considering that previous works focus only on one or two dimensions of integration, we will challenge to consider from a global point of view the propagation of systemic risk.

Physics extends its knowledge to a wide variety of subjects far from the laboratories. An exhaustive list of the applications of physical models applied far from classical physics area is not in the scope of this report. As examples, fire forests are studied with percolation methods [9], traffic flows are understood in terms of hydrodynamical and shock wave equations, elastic properties of biological membranes or blood cells are well described by statistical fields theory [7, 10], birds flocks can be described with simple models and statistical physics [26].

A key question is: *Why is there an interest in using statistical physics to approach the question of systemic risk?*. Systemic risk is the sudden manifestation of a dysfunction occurring at large, eventually global, scale due to a strong markets integration. At a microscopic scale, the interactions between agents make up a complex network responsible in part of the strong integration and may lead to non predictable drawup (drawdown) spreading over a whole economic sector. In other words, one can assists to the emergence of global cooperative behavior, self organized, resulting from local interactions.

This kind of behaviors appear in phase transitions when a system changes from one to another state of matter, like the water (liquid phase) turns to ice (solid phase) or vapor (gaseous phase) as the temperature of the environment is tuned, or when a material displays suddenly a spontaneous magnetization. As one can observe large collective behavior in financial markets, the analogy with critical phenomena appears to be a field of investigations with a very high potential, because statistical physicists have many theoretical or numerical tools, to investigate the behavior of complex systems and understand the mechanisms at stake.

If there is an interest for economists in the results of statistical physics, finance appear to be a very challenging and fascinating field of investigation for physicists. Derivative markets, and financial markets in general, are open systems, it means that there are some quantities as money, number of agents or volume of contracts that change in time. These kind of systems are often called non-equilibrium systems and are at the top of the recent investigations in statistical physics. Prices are not pure uncorrelated random processes, they are not normally distributed and at contrary and fat tails and power law distribution. In conclusion, there are common interests between economy and physics that merge in studying and understanding through the scope of the science of complex systems.

1.3 Scientific relevance of the project for the French Energy Council

The last evolutions observed on financial markets raise fears about the associated systemic risk. Commodities derivative markets are more and more integrated: within the sector of commodities and also with other markets. Indeed, the price increase of given commodities (corn, rapeseed, wheat, sugar cane) has been in part associated to the change of energetic commodities prices. Furthermore, commodities are considered as full fledged assets used in a diversification purpose by portfolio managers. Consequently, price movements recorded on some commodities markets could be explained by *a priori* foreign events like the decrease of equities or exchange rates. Then, the aggressive behavior of speculators is invoked to explain some price behavior and could be at the origin of shocks rising from the merging of markets and propagating to the physical market.
It is crucial to know if energetic markets can be affected by systemic risk and, in such case, be able to quantify this risk and its characteristics. It is undoubtedly interesting to focus on systematic risk. If it turns out that a significant part of price fluctuations is caused by noise, the efficiency of hedging strategies on derivative markets is likely to be affected. Moreover, from the point of view of regulation, it is important to consider the quality of the services offered by derivative markets, and to ask how effective they are in transferring risk among operators and in providing, through futures prices, informative signals.

Figure 1.1: Left figure: real starling flock. Right figure: three-dimensional numerical flock [26]. Natural complex pattern and typical density fluctuations can be observed in a minimal model of interacting agents.

1.4 Academic relevance of this research project

From an academic point of view, the aim of our project is to connect two different fields of investigation: finance and physics.

Since twenty years, economy raised the interest of an increasing number of physicists. From the pioneer works based on mechanical equilibrium approach to recent works on random matrix theory. Such an interest in financial markets can be explained by the fact that the events, on financial markets, result from the interactions of heterogenous agents in a non equilibrium environment.

Systemic risk can lead to important defaults due to markets integration. Considering this situation from a physicist point of view, one can address the question of systemic in the following way: How a can global collective consensus can emerge from local interactions, correlations or decisions? Concepts from the phase transitions and critical phenomena, but also from chaos or non-linear physics, can be very helpful to answer questions about systemic risks.

Recently, Chaté and his co-workers developed stochastic algorithms aiming at characterizing the nature and the properties of a spontaneous macroscopic motion raising in an assembly of identical agents interacting locally [5, 12]. Despite their minimality, these models with a reduce number of parameters are full of informations about the motion of bacteria or complex patterns arising in liquid crystal [6]. Furthermore, still in the spirit of minimal models done by Chaté and coworkers, a slightly modified model of collective motion has been successfully applied into the description of starlings flocks, that are known to display a very high degree of complexity (figure 1.1) [26].

If these minimal models are fruitful to understand the role of simple mechanisms in the emergence of complex patterns or information transfer, like in biology, dynamical systems and also in economy, we hope to build a new model capturing the main features of co-movements in commodities derivative markets.

2

Data

Contents

In this chapter, we present the characteristics of our data. The first section is devoted to a description of the selected markets and the economic intuition which has motivated the choice of fourteen markets corresponding to three different sectors: energy, agriculture and financial assets. In the second section we present the database and technical skills related to the building of prices' curves. The third section gives a brief overview ot the time series'behavior. The last section concerns the seasonality of petroleum products and the attempt to identify periodic patterns in the datas.

2.1 Derivatives markets selected

In this section we present the markets selected for the empirical study, namely energy, agriculture and financial assets.

On the basis of the Futures Industry Association's monthly volume reports, we retained those contracts characterized by the largest transaction volumes. The choice of these three sectors is motivated by the last observations of financial markets evolutions. Commodities derivative are more and more integrated: within the sector of commodities and also with other markets. Furthermore, commodities are considered as full fledged assets used in a diversification purpose by portfolio managers. Consequently, the price increase of given commodities has been in part associated to the change of energy commodities prices or could be explained by a priori foreign events like the decrease of equities or exchange rates. The choice of the different markets within the three sectors allow us to study different kinds of relationships, that is to say the upstream/downstream phases of the industrial process in the petroleum field as between the soy oil and soy bean futures contracts.

2.1.1 Energy markets

For this report, we collected data on seven [1] energy markets: more precisely, we choose three futures contracts on crude oil and four futures contracts on petroleum products.

The three futures contracts on crude oil correspond to:

- the American light sweet crude oil negotiated in the United States, on the Chicago Mercantile Exchange Group (formerly the New York Mercantile Exchange), which

[1] data on seven markets have been collected, but due to a too short historical record or unusable recontructed term structure three markets (*LTC, LHO* and *RBOB*) have not been considered for the empirical investigation.

is referred to as the *NCL* in this report. As illustrated by figure 2.1, this futures contract is, by far, the most widely traded commodity contract in the world since several years.

- the European light sweet crude oil negotiated in Europe, on the InterContinental Exchange, referred to as the *LLC*. This futures contract is usually the second one, worldwide, as far as its transaction volumes are concerned.

- the American light sweet crude oil negotiated in Europe, on the InterContinental Exchange, which is referred to as the *LTC* [2].

As far as the qualities are concerned, we have data on the American crude, usually called the WTI (West Texas Intermediate) and the European one, usually called the Brent. These two qualities are negotiated in two different geographic places. Consequently, the price differential between these two products should reflect both quality differential and transportation costs [3].

These crude oil markets being the most important - in the commodity field - worldwide, they are characterized by the presence of long term expiration dates: up to 9 years in the American market.

We also retain the main futures contracts on petroleum products, namely:

- the American heating oil negotiated in the United States, on the CME Group, which is called: *NHO* in this report

- the American gasoline in the United States, on the CME Group, which is called: *RBOB* in this report

[2] This contract, which was recently launched in Europe, immediately encountered a huge success. It soon became the third commodity futures contract exchanged worldwide. We thus have two different qualities of crude oil and two transactions places.

[3] The third contract has the American crude oil as its underlying asset. It is however negotiated in Europe. Thus only difference between the *NCL* and the *LTC* futures contracts lies in the transaction place and we expect that the price differentials between these two contracts will be very small (in other words, the links between these contracts should be strong.)

- the European heating oil negotiated in Europe, on the ICE, namely the *LHO*
- the European gas oil negotiated in Europe,on the ICE, namely the *LLE*.

Finally, we have completed the energy sector with two natural gas:

- the European natural gas negociated in Europe on the ICE, namely the *LNG*.
- the American natural gas negociated on the CME group, namely the *NNG*.

Thus this database gives us the possibility to study several kinds of relationships, that is to say:

- the upstream and downstream phases of the industrial process in the petroleum field, in Europe and in the United States
- futures contracts corresponding to different qualities and negotiated in different geographical places

2.1.2 Agricultural markets

The recent develoment of biofuel rise our interest as it could introduce strong connections between soft commodities and petroleum products. We selected four significative agricultural markets:

- the Corn futures contract negociated on the CBOT, namely the *CC*
- the Wheat futures contract negociated on the CBOT, namely the *CW*
- the Soy Bean futures contract negociated on the CBOT, namely the *CS*
- the Soy Oil futures contracts negociated on the CBOT, namely the *CBO*

The two former futures contracts, namely *CS* and *CBO*, give us the ability to study the upstream/downstream phases of the industrial process as the link between biofuel and other agricultural markets.

2.1.3 Financial assets

We have completed our databse with three financial assets:

- the Eurodollar interest rate futures contract *IED*
- the Gold, considered as a safety asset *NGC*
- the Exange rate, Dollar/Euro *CEU*
- the Mini S&P500 *ISM*

2.2 Presentation of the database

We selected the futures markets characterized by the most important transaction volumes in three different sectors[4], namely energy, agriculture and financial assets. We used two databases, *Datastream* and *Reuters*, in order to collect on a daily basis, settlement prices, opening prices, open interests and transaction volumes for each market. In order to concentrate on the methodology of the empirical study we limited our empirical work on settlement prices in this report. The original time series gave us the data through the life of a specific futures contract. For example we obtain, for a contract having a nine years maturity, a time series beginning at the birth date of the contracts and finishing at its death. As one of the aims of our study is the analysis of prices behavior according to the maturity of the futures contracts, we had to arrange these futures prices in order to reconstitute, for each market, daily term structures of futures prices. The term structure or prices curve indeed represents the relationship, at a specific date, between futures prices having different delivery dates. So we reconstructed the delivery calendars of all the futures contracts, on the seven markets. Then we determined, month by month, when a specific contract has, for example, a one- or a two-month maturity, and we identified the

[4]Source: Futures Industry Association, Monthly volumes reports

Top 20 Energy Futures and Options Worldwide

Ranked by Number of Contracts Traded and/or cleared in 2007

Rank	Contract	2007	2006	% Change
1	Light Sweet Crude Oil Futures, Nymex	121,525,967	71,053,203	71.04%
2	Brent Crude Oil Futures, ICE Futures Europe	59,728,941	44,345,927	34.69%
3	WTI Crude Oil Futures, ICE Futures Europe	51,388,362	28,672,639	79.22%
4	European Style Natural Gas Options, Nymex Clearport *	29,921,068	19,515,968	53.32%
5	Natural Gas Futures, Nymex	29,786,318	23,029,988	29.34%
6	Light Sweet Crude Oil Options on Futures, Nymex	28,398,793	21,016,562	35.13%
7	Gas Oil Futures, ICE Futures Europe	24,509,884	18,289,877	34.01%
8	NY Harbor RBOB Gasoline Futures, Nymex	19,791,439	3,883,261	409.66%
9	No. 2 Heating Oil Futures, Nymex	18,078,976	13,990,589	29.22%
10	Henry Hub Swap Futures, Nymex Clearport *	16,207,044	24,157,726	-32.91%
11	Crude Oil Futures, MCX	13,938,813	4,466,538	212.07%
12	Fuel Oil Futures, SHFE	12,005,094	12,734,045	-5.72%
13	Henry Hub Penultimate Swap Futures, Nymex Clearport *	10,117,889	7,973,290	26.90%
14	Gasoline Futures, Tocom	7,529,706	12,932,848	-41.78%
15	miNY Crude Oil Futures, Nymex	5,185,214	9,323,467	-44.39%
16	Natural Gas Options on Futures, Nymex	5,051,879	9,581,663	-47.28%
17	Gasoline Futures, C-Com	3,635,329	4,953,168	-26.61%
18	Kerosene Futures, C-Com	2,685,345	4,027,192	-33.32%
19	Kerosene Futures, Tocom	2,350,819	4,492,904	-47.68%
20	European Style Crude Oil Options, Nymex Clearport *	1,879,999	379,250	395.71%

* *Traded primarily off-exchange.*

Source : FIA

Figure 2.1: Energy Futures and Options Worldwide

day when the contracts falls from the two- to the one-month maturities, from the three- to the two-month maturities, etc. This allowed us to obtain daily prices curves for all the markets.

With such a database, one of the difficulties comes from the fact that for one underlying asset, the beginning of the time series frequently contains less information than the end. In other words, the prices curve is shorter at the beginning of the time period. Indeed, as time goes on, the maturities of the futures contracts usually rise on a derivative market. The growth in the transaction volumes of existing contracts induces the introduction of new delivery dates. Thus, in order to keep sufficiently long time periods for our analyses, and in order to have continuous time series, we had to withdraw some maturities from the database. Once this selection has been done, our database still contains more than 655000 prices.

Figure 2.2 summarizes the main characteristics of our database, and the data available, once the term structures reconstituted. First remark, the periods are quite different for each market. Datastream actually does not give the possibility to reconstitute very long time series, and the length of the available time series changes with the market. So the longer time series displayed in table 2.2 (for the American crude oil, for example) rely on databases previously collected by the authors of this report.

Last remark, the American gasoline negotiated in the United States, namely the *RBOB* does not appear in this table. We indeed find out that once the term structures were reconstructed, the gasoline data where not utilizable, as illustrated by figure 2.3. In this figure, the black line represents the one-month maturity (F_{1M}), whereas the dotted line stands for the twelve months' maturity (F_{12M}). The abscissa represents the time period, between 02/2007 and 09/2009. A quick glance at this figure shows that they are sudden drops and spikes in the two series of futures prices. These variations also appear for the other maturities and can not be explained by economic events. We thus decide to neglect these data.

2.3 A brief overview of the time series

Figure 2.4 gives an overview of the behavior of the crude oil futures prices for the *LLC*, that is to say, for the Brent, between 2000 and 2009. Two maturities were selected : one month (black line) and eighteen months. The figure exhibits first of all a huge change in the prices level, in $2004-2009$: they indeed range from a lower level of 40$/b before 2004 to the highest level of 150$/b. It also gives evidence of quite a dramatic change in the prices' volatility, which clearly increases since 2004. Finally, it shows that whereas the one-month maturity was usually higher, before 2004, than the eighteen months maturity,

Energy Futures Contracts					
Mnemonic	**Future contracts**	**Place**	**Period**	**Maturities**	**Trading days**
NCL	Crude	US	1988 – 2009	up to 84	2965
LLC	Crude	EU	2000 – 2009	up to 18	2523
NHO	Heating	US	1998 – 2009	up to 18	2835
LGO	Gas	EU	2000 – 2007	up to 12	2546
NNG	Natural	US	1998 – 2009	up to 36	3140
LNG	Natural	EU	1997 – 2009	up to 9	3055

Financial Futures Contracts					
Mnemonic	**Future contracts**	**Place**	**Period**	**Maturities**	**Trading days**
IED	Eurodollar 3M	US	1997 – 2009	up to 120	3056
NGC	Gold (100 oz)	US	1998 – 2009	up to 60	2877
CEU	Exchange rate	US	1999 – 2009	up to 12	2864
ISM	Mini SP 500	US	1997 – 2009	up to 6	3011

Agricultural Futures Contracts					
Mnemonic	**Future contracts**	**Place**	**Period**	**Maturities**	**Trading days**
CC	Corn	US	1998 – 2009	up to 25	2569
CW	Wheat	US	1998 – 2009	up to 15	3026
CS	Soy Bean	US	1998 – 2009	up to 14	2977
CBO	Soy Oil	US	1997 – 2009	up 15	3056
Total = 655406 daily settlement prices					

Figure 2.2: Main characteristics of the collected datas. The column "mnemonic” refers to the label of the futures contract in *Datastream*, and the column entitled "Place" indicates the geographic localization of transactions. The column "maturities" indicates the last maturity available.

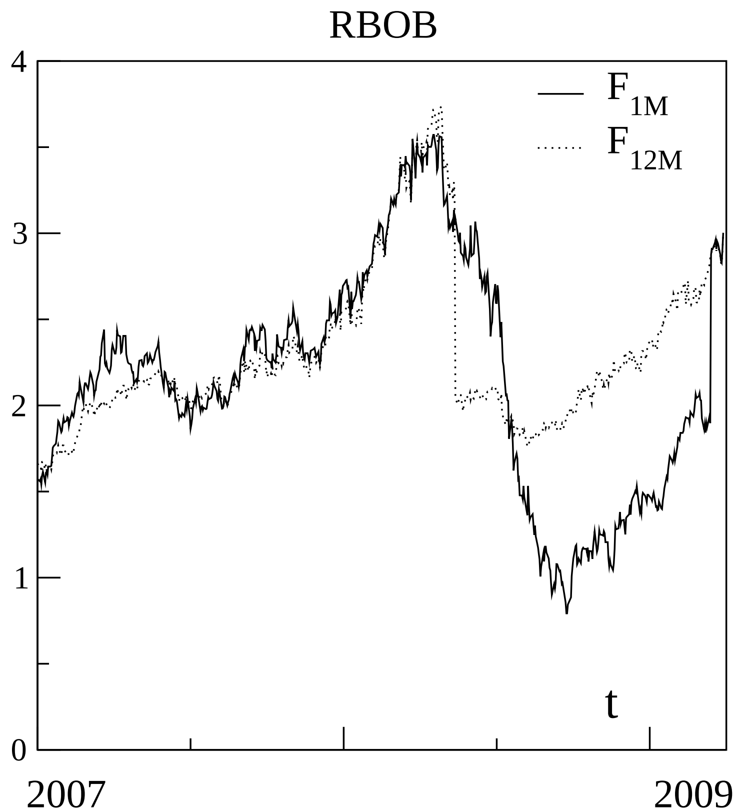

Figure 2.3: Temporal evolution of the maturities 1 and 12 months

this hierarchy seems to change in the end of the period.

We further explored this question trough the calculation of futures prices bases [5] as illustrated by figure 2.5. The figure 2.5 (b) represents, for example, two temporal bases for the American crude oil negotiated in the United States, the so-called WTI, between 1988 and 2009. The dotted line stands for the relationship between the one- and the two-month maturities, that is to say:

[5]The basis considered in this report is a temporal basis: it represents the difference between the spot and futures prices. More precisely, we expressed the temporal bases in percentage, in order to avoid size order effects. When the basis is positive, the market is in backwardation; otherwise, it is in contango. Lastly, as we did not have time series for spot prices, we choose to approximate this variable with the one-month futures price. Such an approximation is very frequent in empirical studies on commodity markets.

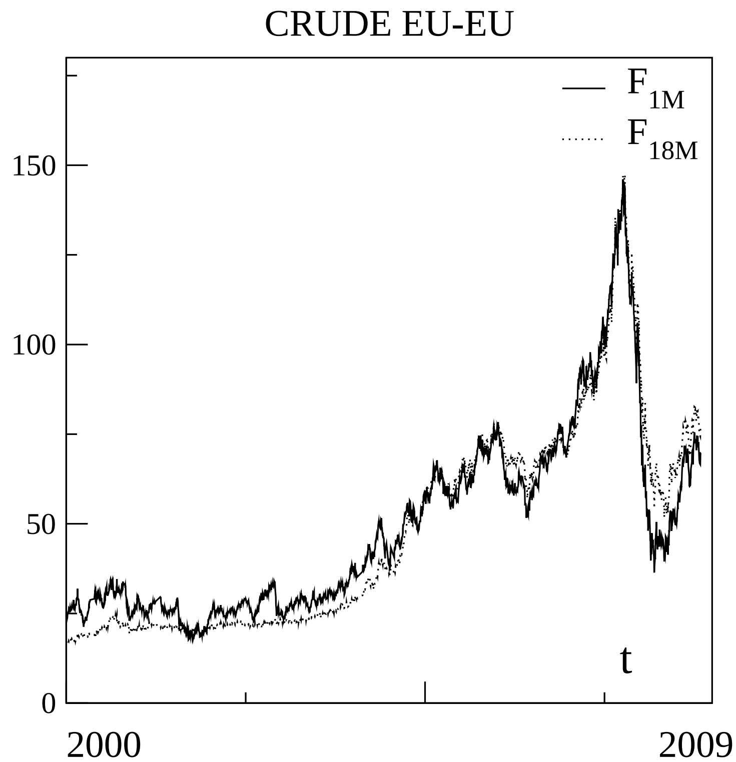

Figure 2.4: Temporal evolution of the maturities 1 and 18 months

$$\frac{F_{2M} - F_{1M}}{F_{1M}},$$

whereas the black line represent the basis between the two- and the three-months maturities, namely:

$$\frac{F_{3M} - F_{2M}}{F_{2M}},$$

As the differences between the two lines are not really easy to see on figure 2.5 (b), let us just underline that, during this very long period (almost 21 years), the basis was most of the time positive. In other words, the market was in backwardation. This is a well-known characteristic of the crude oil market. However, since a few months, it seems that backwardation (i.e. positive bases) might not be the rule anymore. Moreover, a very high contango (i.e. negative basis) distinguishes the very end of the period. This spike is all the more surprising that, in derivative markets, arbitrage operations between the physical and paper markets should impose a limit on contango situation, this limit corresponding to the storage costs of the commodity. Thus, either there was a storage difficulty in the physical market at that date, or there is a shock that must be explained by another phenomenon.

Figure 2.6 presents the time evolution of the prices of the first (black lines) and the late maturity (dotted lines) for three different futures contracts, namely the crude oil negotiated on the New York Mercantile Exchange (*NCL*), the corn (*CC*) and the mini contract on the S&P500 (*ISM*), both of them traded on the CME Group. The illustration shows that our time period covers one crisis for the commodities, highlighted by a significant price's rise in the crude oil (*NCL*) and in the gold (*CC*) which does not appear for the financial asset (*ISM*). The similarity between the commodities price's behavior is at the core of the debate related to the possible impact of institutional investors on the prices of commodities.

The black and dotted lines also give an illustration on one of the most important features of the commodity prices curve's dynamic: the differences in the behavior of first nearby contracts and deferred contracts. The movements in the price of the prompt contracts are larger than the other ones. This results in a decreasing pattern of volatilities along the prices curve. Indeed, the variance of futures prices and the correlation between the nearest and subsequent futures prices decline with maturity. This phenomenon is called the *Samuelson effect*. For the financial asset (figure 2.6 (c) and its inset), the two lines are almost mixed up, while we can clearly observe a backwardation for the energy commodity (figure 2.6 (a)) and a contango for the agricultural commodity (figure 2.6 (b)).

A more precise insight on the prices behavior at the end of the period is illustrated by the figure 2.5 (c). This figure indeed represents the same bases, on a shorter period, from 2000 to 2009. Now the differences between the dark and the dotted lines appear clearly. Thus the figure gives an illustration of one of the most important features of the commodity

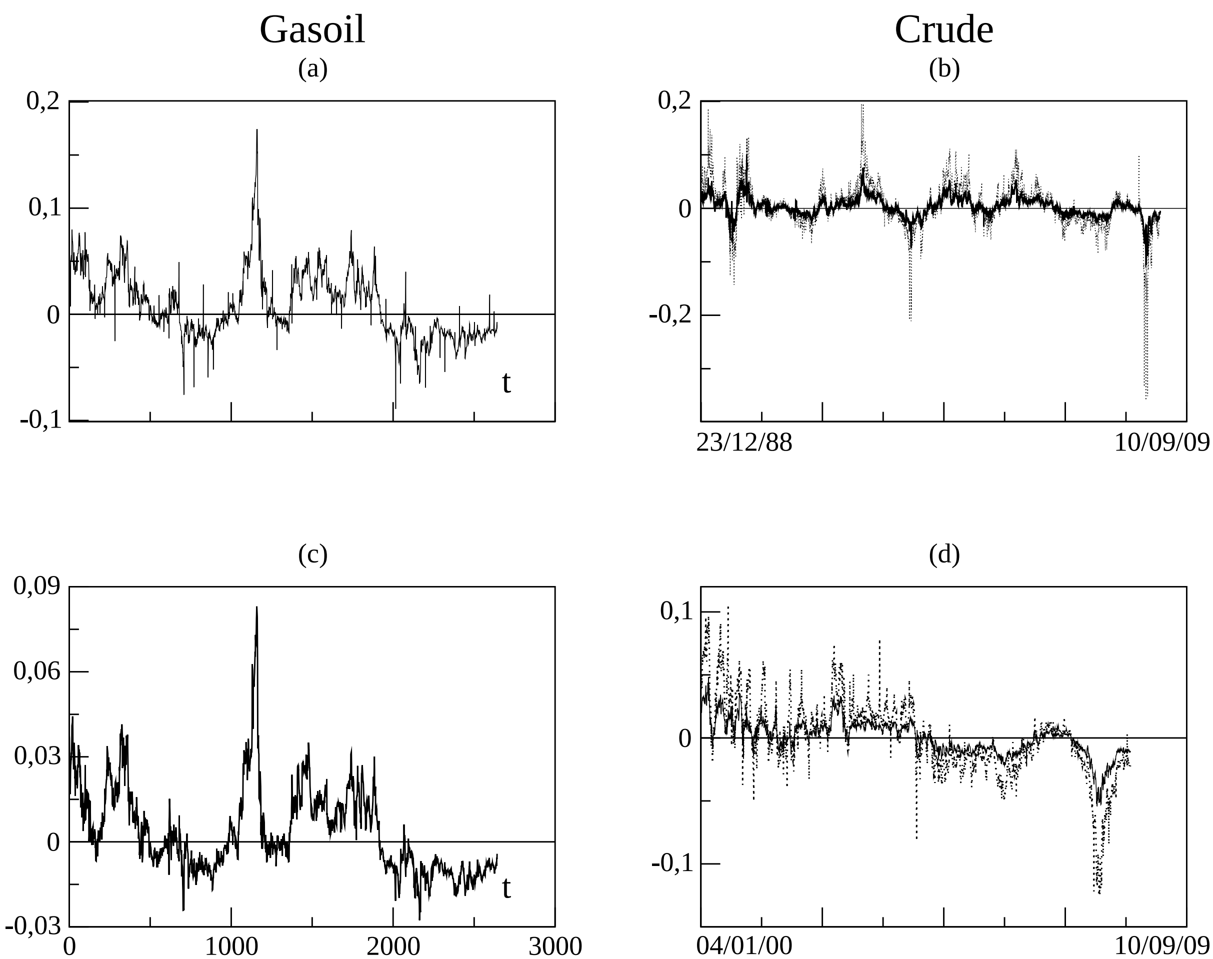

Figure 2.5: Base behavior of the LLE (left panel) and the LLC (right panel).

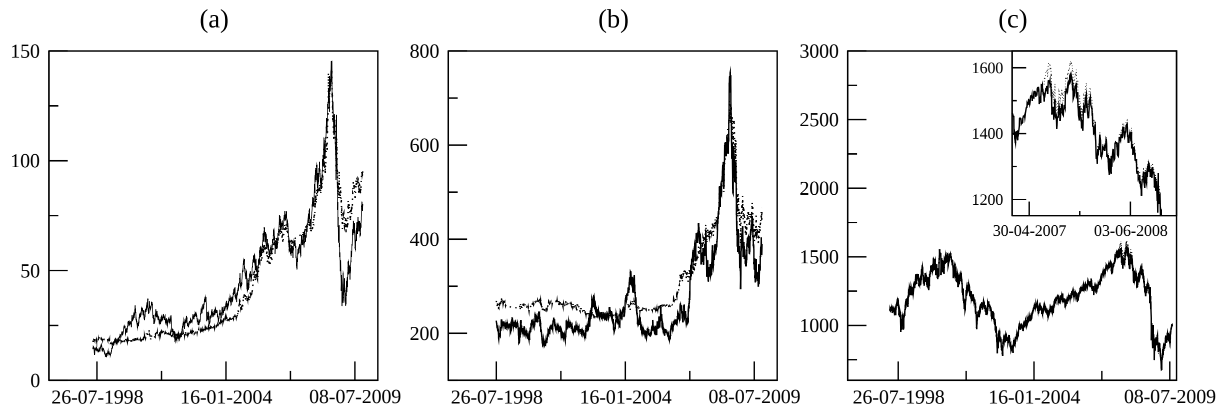

Figure 2.6: Time evolution of the prices of the first (black lines) and last maturities (dotted lines) for futures contracts representative of each sector between 1998 and 2009. Figure (a) represents the evolution of crude oil prices (*NCL*), Figure (b) exhibits that of corn prices (*CC*), whereasFigure (c) is devoted to the S&P index (*ISM*). The inset in Figure (c) represents a smaller time window where the two lines are easily distinguishable.

prices curve's dynamic: the difference between the price behavior of first nearby contracts and deferred contracts [6]. The movements in the prices of the prompt contracts are larger than the other ones. This results in a decreasing pattern of volatilities along the prices curve. Indeed, the variance of futures prices and the correlation between the nearest and subsequent futures prices decline with maturity. This phenomenon is usually called *the Samuelson effect*. Intuitively, it happens because a shock affecting the nearby contract price has an impact on succeeding prices that decreases as maturity increases [28]. As futures contracts reach their expiration date, they react much stronger to information shocks, due to the ultimate convergence of futures prices to spot prices upon maturity. These price disturbances influencing mostly the short-term part of the curve are due to the physical market, and to demand and supply shocks. The Samuelson effect seems however particularly important in the end of our observation period, suggesting that there are quite a lot of noises (or shocks) affecting the one-month time series.

Lastly, figures 2.5 (a) and 2.5 (c) display two temporal bases (respectively three-month

[6]We obtained almost exactly the same figure for the european crude oil market.

minus two-month over two-month and two-month minus one-month over one-month) for the European gas oil between 2000 and 2007. The same general comments than those already proposed for the crude oil can be made for this petroleum product: backwardation is more frequent than contango, except for the end of the period. Moreover, short term prices are more volatile.
This brief overview of the futures prices behavior lead us to think, first that it could be interesting to make a separation between the period before and after 2004-2005 and second, that a separate analysis of backwardation and contango situations could be fruitful.

2.4 The seasonality of petroleum products

Before proceeding with the empirical tests, we first took the time to study the question of the seasonality of petroleum products. The presence of such a phenomenon indeed might influence the futures priceś behavior in two ways: first, it can create autocorrelations in the time series; second, it is frequently associated with the sign of the temporal basis, namely the difference between the spot and the futures prices.
In this paragraph, we will first recall what literature says about the seasonality of energy commodities, and second, we will present our attempt to identify seasonal patterns in the extracted time series.

2.4.1 What literature says about seasonality

In the energy field, some commodities are known for showing seasonal fluctuations: electricity, natural gas, heating oil and gasoline / gas oil (as we did not collect, for this report, data on natural gas and electricity, we only mention them here for the record). This seasonality is due to changing consumption as a result of weather patterns. Usually, it is not supposed to influence the crude oil markets.

As far as seasonality [7] is concerned, heating oil and gas oil (and/or gasoline) markets usually move in opposite ways. In both cases, there is a low and a high season. However, the high season for heating oil corresponds to the low season for gas oil and vice and versa. In the case of gasoline, the high consumption period corresponds to the holidays and, more specifically to the summer, whereas in the case of heating oil, prices spikes essentially take place in winter. Since the turn of the century however, a second period of high consumption for heating oil takes places in the summer, for air-conditioning purposes. This is especially true for the United States. Thus a lot of authors recognize the presence of a seasonality effect in petroleum products, with one or two periods of high consumption, according to the observation period and to the place of consumption.

Another interest aspect of the seasonality is that it creates a specific behavior of their temporal bases. The heating oil market is frequently in backwardation from December to March, whereas the gasoline market in characterized by the presence of inverse carrying charges from June to November. Moreover the two products are supposed to have opposite behavior: when one of them is in contango, the other one is in backwardation. This is due to the fact that the refining process is a joint production process [8].

[7] For more information on the seasonality of petroleum products, see for example D. Pilipovic, 2007, Energy risk : valuing and managing energy derivatives, 512 p, Mac Graw [24].

[8] For more details on these points, see for example D. Lautier, 2000, La structure par terme des prix des commodités : analyse théorique et applications au marché pétrolier, Thèse, Université Paris Dauphine [16], or Edwards F.R., Canters M .S., 1995, The Collapse of Metallgesellschaft : unhedgeable risks, poor hedging strategy, or just bad luck ?, The journal of futures markets, 15(3), 211-264 [8].

2.4.2 An attempt to identify the seasonal patterns

Our first attempt to identify the seasonal patterns [9] in the data consisted in a graphical analysis. We examined the heating oil and gas oil data, year by year, in order to find evidence of a seasonal patterns. We first found that it was probably possible to identify, for each market, a one-period seasonality, with a high and a low season. However these high and low seasons slightly changed each year. We also try to compute the temporal basis for this petroleum products, in order to make sure that the high seasons corresponded to backwardation whereas the low season exhibited contangos. The results however were not convincing. We thus proceeded with a more formal method: the frequency domain analysis.

In this paragraph, we first briefly expose this method. We then present its applications to our data. Lastly, we conclude.

Frequency domain analysis

The frequency domain representation, or Discrete Fourier Transform (DTF) is a mathematical procedure that transforms a discrete function into another, which is called the frequency domain representation. The latter provides a decomposition in frequencies, and their associated strength, rather than a decomposition in time. While the original signal expresses the value of the function f(t) at the time t, the frequency domain gives the strength of each frequency which is present in the original time series. When a specific pattern occurs fairly regularly in the signal, the DTF gives the value as well as the

[9] Two main methods are usually used to take account of seasonal fluctuations. The first is to *deseasonalise* the data by, for example, a moving average method before estimation. The second is to use particular dummy variables during estimation. The first method can be criticized on several counts. Firstly, the moving averages are based on an overlapping process that creates additional autocorrelation. Secondly, cumulating averaging is a smoothing device and may tend to obscure some of the finer movements in the series we are considering. The second method amounts to enhance the information associated to a specific variable and to specify, for each price, its season. For more details on these methods, see for example Introductory econometrics M.B. Stewart and K.F.Wallis, Basil Blackwell, second edition, 1990, 337 p [30].

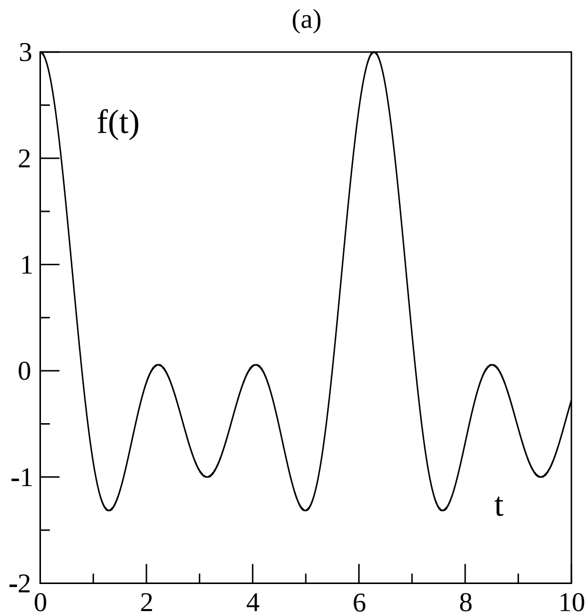

Figure 2.7: Frequency domain representation. Figure (a) plot of the function $f(t) = \cos(\omega_1 t) + \cos(\omega_2 t) + \cos(\omega_3 t)$. Figure (b) discrete fourier transform of f(t).

strength of this repeated pattern.

The figure 2.7 gives an illustration of this method. On 2.7 (a), we plot the function:

$$f(t) = \cos(\omega_1 t) + \cos(\omega_2 t) + \cos(\omega_3 t)\,, \tag{2.1}$$

where the ω_i are the frequencies and the periods T_i are given by:

$$T_i = \frac{\omega_i}{2\pi} \tag{2.2}$$

In this example, the signal comes from a mathematical periodic function. The graphic representation of this signal shows that there are three periods: a long one, of duration 6, a shorter one, of duration 3, and the shortest, of duration 2. The figure 2.7 (b) gives the frequency representation of this signal. We can observe three peaks localized at a characteristic period of the original signal. These three peaks confirm what can be seen on figure 2.7 (a): there are three distinct patters repeated with periods T_1, T_2 and T_3.

In the case of empirical data, however, the interpretation of the original signal is far from obvious. The data are affected by noises and finding periodic patterns is not simple.

Discrete Fourier Transform of petroleum prices

In our attempt to identify the seasonal pattern of petroleum prices, we decided to use a discrete fourier transform for two different markets: crude oil and heating oil. The first of these two markets was used as a reference precisely because it is not supposed to exhibit a cyclical behavior. We decided to use three months futures prices in order to avoid the presence of potential noises in the nearest contracts. The figure 2.8 presents the results obtained on these two markets. The left side of the figure pictures the behavior of the futures prices on the observation period. The right side gives the result of the transformation.

The interpretation of the results is by far not straightforward. The two series exhibit quite similar patterns, that is to say, we do not find evidence of a specific cyclical behavior in the heating oil data, especially if we compare them with the crude oil data. Several reasons can explain such a result. First, our observation period on heating oil starts in 1998, namely roughly at the date when the consumption of this petroleum product for air-conditioning purposes really began. Maybe the presence of this second period of high consumption smoothes the prices behavior, and makes it more difficult to identify the cyclical behavior. Another possible reason is that for these markets, even if we have 4000 (for the heating oil) and 8000 (for the crude oil) daily observations, the data are not sufficiently abundant to give evidence of a seasonal pattern. These results leaded us to conclude that we should not try to take into account the seasonality of our data.

We briefly remind the main features of this chapter. Firstly we built a database which includes a crisis affecting ,at least, the energy and agricultural markets. Then, we detected a similar prices's behavior of the commodities, which is not observed for the financial assets. Finally, we were not able to give an evidence of the seasonality of the petroleum

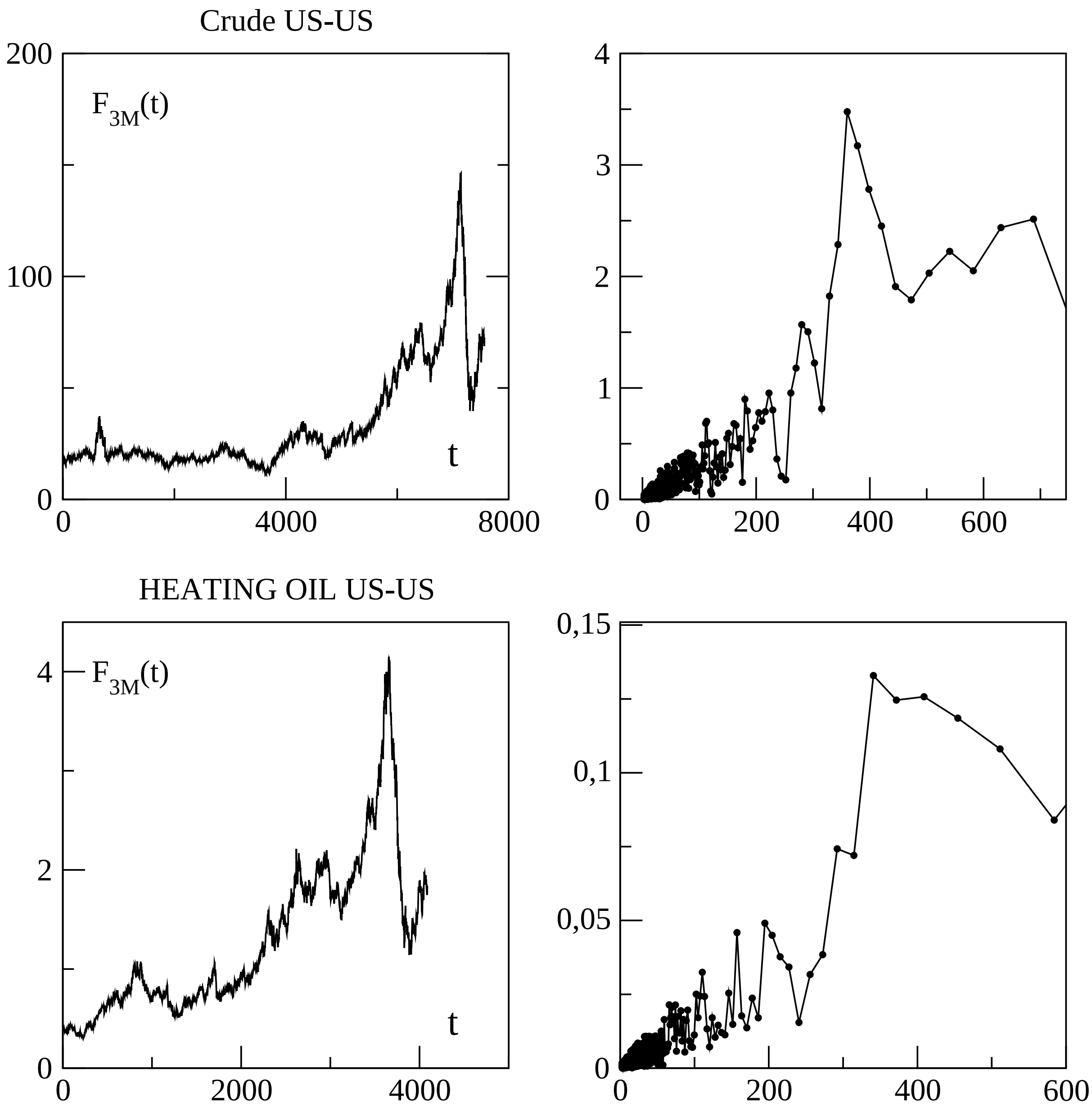

Figure 2.8: Temporal series of the 3 month maturity price $F_{3M}(t)$ (left panel) and discrete fourier transform (right panel).

products. Consequently, the question of the seasonality will not been taken into account in our analysis.

3

Methodology

Contents

In this chapter we present the tools used to perform the analysis of the integration derivative markets. In the first section we expose a correlation-based method which allows to transform a correlation matrix into a distance matrix. The former is then used to build the so-called Minimum Spanning Tree which allows to study the links between the most correlated assets. The second section is devoted to the description of the topological properties of the trees. We present how to extract economical information from the visualization as from allometric properties of the trees. In the third section we expose the dynamical measures used to study the temporal dimension.

From a physicist point of view, financial markets are very challenging complex systems. Among the different tools which were developed and used in this field, one seems naturally relevant to perform a three-dimensional analysis of the integration of derivative markets: the graph theory, also referred to as networks analysis. A graph can be defined as a mathematical representation of pairwise relationships within a collection of discrete entities. Representing a financial market as a graph is appealing because such a system is composed of a large number of assets, such as equities, bonds, stocks or derivative products. The graph gives a way to describe all the links (edges) connecting these entities (nodes). Moreover, the graph can be weighted in order to represent the different intensities of these links and/or nodes.

The graph theory has received a lot of attention from the physicist community during the last decade. Today, it is used in order to describe various complex systems such as biological cells, biochemical reactions, the Internet, and financial markets. Among the different tools and measures recently developed we selected, for our study, those allowing us to analyze market integration in a three-dimensional approach.

We first decided to represent our prices system through the study of the correlations of returns. Having transformed these correlations into distances, we were able to draw a full connected graph of our system, where the nodes of the graph represent the different time series of futures prices. The dimension of our correlation matrix being high, in order to filter the information contained in the graph, we rely on specific graphs, namely minimum spanning trees. The method used to build minimum spanning trees is presented in the second section of this chapter. In order to understand the organizing principles and the dynamic behavior of these trees, we employed a method which is presented in the third section of this chapter.

3.1 Preliminary studies on integration

The financial literature has investigated the question of integration through different ways. As early as 1990, [25] began to study the herding behavior of investors on commodity derivative markets. Their seminal work shows that the persistent tendency of commodity prices to move together can not be totally explained by the common effects of inflation, exchange rates, interest rates and other macro-economic variables. It has inspired several other researches on co-movement. Yet, in this kind of work the identification of the relevant economic variables is tricky. This could explain why empirical tests do not really succeed in concluding that there is herding behavior in commodity markets.

Focusing on spatial integration, [14] initiated another approach to the systemic risk in commodity markets. Such a study is centered on the relationships between the prices of raw materials negotiated in different places. The authors initiated several works on spatial integration, based on the methodology of the co-integration. The empirical tests show that commodity markets are more and more spatially integrated. In the same vein, [4] examined the links between stock and commodity markets. They were however not able to conclude that the former have an influence on the latter.

Integration has also a temporal dimension, in the sense of the preferred habitat theory ([21]). In [17], the author studied the segmentation of the term structure of commodity prices and examined the propagation of shocks along the prices curve, on the crude oil

symbol	name	market	symbol	name	market
AA.F	aluminum alloy	LME	MW.F	wheat spring	MGEX
AL.F	aluminum	LME	NG.F	natural gas	NYMEX
BO.F	soybean oil	CBOT	NI F	nickel	LME
C.F	corn	CBOT	OJ.F	orange juice	NYBOT
CC.F	cocoa	NYBOT	PA.F	palladium	NYMEX
CL.F	crude oil	NYMEX	PB.F	pork bellies	CME
CO.F	copper	LME	PL.F	platinum	NYMEX
CT.F	cotton	NYBOT	RR.F	rough rice	CBOT
FC.F	feeder cattle	CME	RS.F	canola	WCE
GC.F	gold	NYMEX	S.F	soybean	CBOT
HG.F	copper	NYMEX	SB.F	sugar	NYBOT
HO.F	heating oil	NYMEX	SC.F	brent oil	ICE
KC.F	coffee	NYBOT	SI.F	silver	NYMEX
KW.F	wheat	KCBT	SM.F	soybean meal	CBOT
LB.F	lumber	CME	TI F	tin	LME
LC.F	live cattle	CME	W.F	wheat	CBOT
LE.F	lead	LME	ZI.F	zinc	LME
LH.F	lean hogs	CME			

Figure 3.1: List of commodities investigated by the authors of [29].

petroleum markets. She showed that temporal integration progresses trough time.

In statistical physics, the minimum spanning tree is the tool that is the most heavily used in order to understand the evolution of complex systems, especially when these systems are financial assets. Other filtering procedures have been used by different authors, [19], and provide different aspects of the information stored in the investigated sets.

In the pionneer work of [18], relying on minimum spanning trees, the author sinvestigates cross correlations of asset returns and identifies a clustering of the companies under investigation. In [2], the authors use this correlation based method in order to examine stocks portfolios and financial indexes at different time horizons. They also apply this method in order to falsify widespread markets models, on the basis of a comparison between the topological properties of networks related to real and artificial markets. The filtering approach based on the minimum spanning tree can also be used to construct a correlation based classification of relevant economic entities such as banks or hedge funds, [20]. Last but not least, the robustness over time of the minimum spanning tree's characteristics has also been examined in a series of studies, like for example [15] and [23].

As far as commodities are concerned, [29] recently proposed a study of commodities clustering using minimum spanning trees and statistical physics tools. Whereas they found evidence of a market synchronization, which is a crucial point when aiming at understanding systemic risk, some criticisms can be made. First, the database used in [29] contains commodities characterized by a low transaction volume, which can introduce noise in the correlation matrix. Second, they do not examine the maturity dimension of the futures contracts.

3.2 Minimum Spanning Trees: a correlation-based method

In order to study the links between assets and /or maturities, we first of all compute the synchronous correlation coefficients of the prices returns. This coefficient matrix is the starting point of our analysis. In order to use the graph theory, there was however a need to quantify a distance between the elements under examination. We thus extracted a metric distance from the correlation matrix. We then had the possibility to build some graph. Lastly, we used a filtering procedure in order to identify the minimum spanning tree ([18]). Such a tree can be briefly defined as the one providing the best arrangement of the different points of the network, that is to say, the one that identify the most relevant connections between points.

3.2.1 The correlation matrix

In order to measure the degree of similarity between the synchronous time evolution of futures contracts, we built a matrix of correlation coefficients. The latter are defined as follows:

$$\rho_{ij}(t) = \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\sqrt{\left(\langle r_i^2 \rangle - \langle r_i \rangle^2\right)\left(\langle r_j^2 \rangle - \langle r_j \rangle^2\right)}}, \tag{3.1}$$

where i and j are pairs of assets (that is to say, i and j stand for the nearby futures prices of pairs of assets like commodities, interest rates, stocks, and currencies) when spatial integration is under scrutiny, pairs of delivery dates when we focus on maturity integration, or a mix of the two when a three-dimensional analysis is performed. The daily logarithm price differential stands for prices returns, with $r_i = (\ln F_i(t) - \ln F_i(t - \Delta t)) / \Delta t$, where $F_i(t)$ is the settlement price of the futures contract at the instant t. Δt is the time horizon, and $\langle . \rangle$ denotes the statistical average performed other time, on the trading days of the investigated time period.

By definition the correlation coefficient $\rho_{ij}(t)$ can vary from -1 (complete anti-correlation) to 1 (complete correlation). Lastly, when $\rho_{ij}(t) = 0$, there is no correlation.

For a given time period and a given set of data, we thus computed the $N \times N$ matrix of correlation coefficients C, for all the pairs ij. C is symmetric with $\rho_{ij} = 1$ in the main diagonal where $i = j$. Consequently, such a matrix is characterized by $N(N-1)/2$ coefficients.

In order to study the correlations of returns and their statistical properties, we mainly follow [29]. More precisely, in order to examine the time evolution of our system and its sensibility to specific market events, we also investigate the mean correlations of the returns and their variances.

The mean correlation $C^T(t)$ for the correlation coefficient ρ_{ij}^T in a time window $[t - \Delta T, t]$ can be defined as follows:

$$C^T(t) = \frac{2}{N(N-1)} \sum_{i<j} \rho_{ij}^T(t), \tag{3.2}$$

The variance σ_C^2 of the mean correlation is given by:

$$\sigma_C^2 = \frac{2}{N(N-1)} \sum_{i<j} \left(\rho_{ij}^T(t) - C^T(t)\right)^2. \tag{3.3}$$

While computing the mean correlations and their variances, we examined the window size dependance of these quantities (needless to say, this problem also has an influence on other investigated quantities, such as node strength, tree length, main occupation layers and survival ratios). The choice of the size of the time window ΔT is a compromise between the noise's level and a good statistical averaging. A large window decreases the noise's level but also gives averages over a too long time period. The selection of different time windows (i.e. $\Delta T = 20, 120, 240, 480$ and 960 trading days) and the study of the corresponding size effects lead us to choose a window's length of 480 days for all the quantities which are used in this study. This value is smaller than those usually encountered in physics. It however allows us to grasp finer market's evolutions.

3.2.2 From the correlation matrix to the distance matrix

In the search for a description of futures prices relying on the graph theory, there is a need to introduce a metric. The correlation coefficient ρ_{ij} indeed cannot be used as a distance d_{ij} between i and j because it does not fulfill the three axioms that define a metric [11]:

- $d_{ij} = 0$ if and only if $i = j$,
- $d_{ij} = d_{ji}$
- $d_{ij} \leq d_{ik} + d_{kj}$

A metric d_{ij} can however be extracted from the correlation coefficients through a non linear transformation. Such a metric is defined as follows:

$$d_{ij} = \sqrt{(2(1-\rho_{ij}))}. \tag{3.4}$$

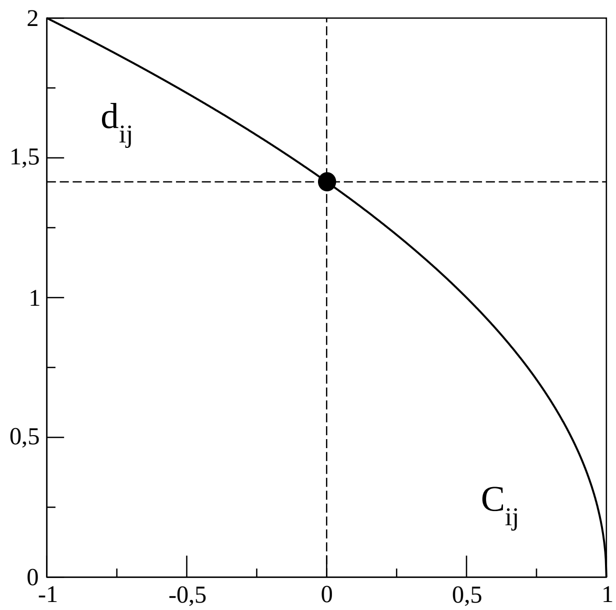

Figure 3.2: Distance d_{ij} between two commodities or delivery dates as a function of the correlation coefficient ρ_{ij}.

Thus the distance matrix D is extracted from the correlation matrix C according to the equation (5.2). C and D are both $N \times N$ dimensional. As illustrated by figure (3.2), while the coefficients ρ_{ij} can be positive for correlated returns or negative for anti-correlated returns, the distance d_{ij} representing the distance between prices returns is always positive.

With such a metric, the first axiom defining a metric is valid because $d_{ij} = 0$ if and only if the correlation is total (namely, only if the two futures prices follow the same stochastic process). The second axiom is valid because the correlation coefficient matrix C is symmetric. Hence, the distance matrix D is symmetric by definition. The third axiom is valid because equation (5.2) is equivalent to the Euclidian distance between two vectors .

The distance matrix D can then be employed in order to build the graph that connect all the elements of the system.

3.2.3 From full connected graphs to Minimum Spanning Trees (MST)

A graph gives a representation of pairwise relationships within a collection of discrete entities. A simple connected graph represents all the possible connections between the N points under examination with $N-1$ edges. Each point of the graph constitutes a node or a vertex.
A weighted graph gives more information than a simple one, which just describe the existing relationships between the elements of the system being described. The weights indeed give some information about the intensity of the relationships. Such weights can for example represent the distance separating the nodes.
Previous studies of complex networks have lead to the conclusion that such systems cannot be described by simple or even weighted graphs. In order to fully understand the dynamics of the system and its organizing principles, there is a need to span the graph, *i.e.* to traverse all its nodes. However, starting from one node and going to the next one until all the graph has been spanned, there are a lot of different paths. In other words, there are a lot of spanning trees. The aim of the analysis relying on graph theory is to retain all the important information while having a simple representation of this information. Minimum spanning trees have proven to give a simple and efficient answer to this problem. In a weighted graph, the minimum spanning tree (MST) is the tree spanning of the nodes of the graph, without loops, having less weight than any other. It thus gives the shortest path linking all nodes between them. The links of the MST are a subset of the links of the initial graph. Figure (3.3) gives an example of such a MST, for a graph having weighted links between its different nodes. It also shows that a single connected graph has many spanning trees.
The MST associated with a distance matrix can be built as follows ([18]). The MST is progressively built up by linking all the elements of the set under examination in a tree characterized by a minimal distance between nodes. The first step consists in ordering the

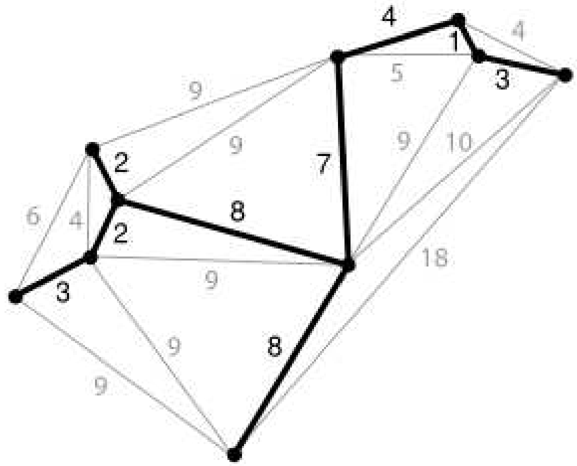

Figure 3.3: Example of a minimum spanning tree constructed from a graph with weighted edges. The minimum spanning tree (MST) is a particular subgraph of the original one. The MST covers all the points with the less weighted path (black line) and without forming any loop.

non-diagonal elements of the distance matrix D in increasing order. The starting point is then the pair of elements with the shortest distance. At this stage, the MST is just composed by these two elements. Starting from one or the other of these two elements, the next smallest distance is determined, adding thus a third element in the MST. By continuing, the tree includes a fourth element, a fifth one, and so on. If the next smallest distance concerns two elements which are already in the MST, this distance is ignored, in order to avoid loops. In our study, we used Prim's algorithm ([27]) in order to built the minimum spanning trees. As described above, starting from a single node, this algorithm continuously increases the size of a tree until it spans all the vertices of a connected graph. The MST is attractive because through a filtering procedure[10], it provides for an arrangement of the different points of the graph which reveals the most relevant connections of each elements of the system. In the context of our study, it gives us a synthetic way to observe the connections between different assets and maturities. As the minimum spanning

[10]The MST can be considered as a filter as during its construction, we are reducing the information space from $N(N-1)/2$ separate correlation coefficients to $N-1$ tree edges.

tree is a path between nodes with a minimal distance, it is also, according to equation (5.2) the path between the most correlated nodes. Thus, such a method can be seen as a way to reveal the underlying mechanisms of systemic risk: the minimal spanning tree can be interpreted as the easiest path for a shock to propagate in three dimensions: space, maturity and observation time.

3.3 Topology of the Minimum Spanning Trees

The first information given by a minimum spanning tree is the kind of arrangement found between the vertices. So a first step of the study of minimum spanning trees lies in the visualization of the trees. After a simple graphic representation of the MST, we use the method of the allometric coefficients in order to quantify wether the filtered network is totally organized, totally random, or is situated somewhere between these two extreme kinds of organization.

3.3.1 Visualization and description of the MST

The visualization of the trees is the first step of the analysis of a complex system through the method of the MST. It is a very important step, as the meaningfulness of the taxonomy that will emerge of the system through the representation of the trees will be one of the main justifications for the use of the method. In our study, the analysis of the groups formed by the different underlying assets in the space and the examination of the organization of the different delivery dates will be very interesting.

After visualizing the MST, there is a need to describe and interpret the graphs. In our three-dimensional analysis of the integration of derivative markets, we propose the use of a distinct terminology according to the dimension under examination. In order to describe the grouping of underlying assets in the space, we will use the term sector, whereas in order to describe the grouping of delivery dates in the maturity dimension, we will retain the word cluster. In both cases, the term branch will refers to a subset of the tree. In addition, in order to describe the graph, there is a need for a reference point. In our case, the reference is the central node. We will come back to this concept while presenting the notion of the mean occupation layer.

3.3.2 Allometric behavior of the MST

One step further in the interpretation of the information given by the MST is the analysis of its randomness. Star-like trees are symptomatic of a random organization of the elements of the system, whereas chain-like trees reveal a very strong organization. In order to determine wether our filtered networks are totally organized, totally random, or where there are located between these two extreme kinds of organization, we decided to study the allometric behavior of the MST.
The first model of the allometric scaling on a spanning tree was developed by [1]. The first step of the procedure consists in initializing each node of the tree with the value 1. Then the root or central vertex of the spanning tree must be identified. In what follows, the root is defined as the node having the highest number of links attached to him[11]. Starting from this root, the method consists in assigning two coefficients A_i and B_i to each node i of the tree. Such coefficients are defined as follows:

$$A_i = \sum_j A_j + 1 \text{ and } B_i = \sum_j B_j + A_i, \tag{3.5}$$

where j stands for all nodes connected to i in the MST. The allometric scaling relation is defined as the relation between the two allometric coefficients A_i and B_i:

$$B \sim A^{\eta}, \tag{3.6}$$

[11] There are actually several definitions for the central vertex, as will be explained a bit later in this section, in the paragraph devoted to the mean occupation layer

where the exponent η is called the allometric exponent. The latter represents the degree or randomness of the tree and stands between two extreme values: 1^+ for star-like trees and 2^- for chain-like trees.

3.4 Dynamic analysis of the Minimum Spanning Trees

Minimum Spanning Trees are appealing because of the information revealed by their topology. However, such a correlation based method is intrinsically time dependent. Thus, there is a need to study the time dependent properties of the MST. In our case, as the MST reflects the temporal state of the markets under consideration, we will particularly focus on the possible consequences of markets events on the structure of the system.
In order to study the robustness of the trees' topology, we use several measures. We first calculate the nodes strength, which gives an information on how much a node is correlated to the others in the MST. The graph lengths reveals the state of the system at a specific time. We next use the concepts of central vertex and mean occupation layer in order to appreciate the compactness of the trees. Lastly, survival ratios indicate how the topology of the trees evolves with time.

3.4.1 Node's strength

The node's strength S_i is defined as follows:

$$S_i = \sum_{i \neq j} \frac{1}{d_{ij}}. \tag{3.7}$$

This quantity, calculated for each node i, indicates the closeness of one node i to the others and in our case, give thus an information on the intensity of the correlation between this node and the others. When S_i is high, the node is close to the others whereas when it is low, the node is far from the others. Lastly, the node strength gives the possibility to undertake static and dynamic analysis. It indeed can be computed over the entire period under examination or it can be measured on the basis of rolling windows having a size ΔT.

3.4.2 Tree's length

Another interesting quantity is the normalized tree's length, which can be defined as the sum of the lengths of the edges belonging to the MST:

$$\mathcal{L}(t) = \frac{1}{N-1} \sum_{(i,j) \in MST} d_{ij}, \tag{3.8}$$

where t denotes the time at which the tree is constructed, and $N-1$ is the number of edges present in the MST.

The length of a tree is all the more important that the distances are high, that is to say, in our case, that the correlations are low. Thus, the more the length of the tree diminishes, the more integrated the system is.

3.4.3 Central vertex and mean occupation layer

The central vertex and the mean occupation layer allow to appreciate the degree of the compactness of a graph. Understanding the concept of the central vertex, or root of the tree, is a prerequisite for the use of the mean occupation layer. Such a concept is very important for the analysis of the topology and dynamic behavior of the networks, especially when studying financial markets integration.

The central vertex can be defined as the parent of all other nodes in the tree. It is thus a reference point in the three, against which the localization of all other vertices is set. This concept is very important in our case, as if a shock emerges at this specific node, it will have a more important impact than anywhere else in the tree. Moreover, such a node will be the preferred one for the transmission of a shock.

There are several ways to define the central vertex of a tree. [23] propose three alternative definitions. According to the first one, the central vertex is the node with the highest vertex degree, namely the highest number of edges which are incident with the vertex (this definition is the one retained for the computation of the allometric coefficients). The second definition corresponds to the weighted vertex degree criterion and defines the central vertex as the one with the highest sum of those correlation coefficients that are associated with the incident edges of the vertex. In such a case, more weight is given to short links, whereas in the first definition, each departing node was weighted in the same way. In order to present the third definition, let us first introduce the mean occupation layer proposed by the authors.

This quantity L can be computed in the following way:

$$L = \frac{1}{N} \sum_{i} l\left(v_i\right), \tag{3.9}$$

where $l\left(v_i\right)$ is the level (layer) of the vertex i. This measure must not be confused with the distance d_{ij} between nodes. The level says, indeed, how far the node i is to the central node, whose level is set to zero.

The mean occupation layer L indicates the layer where the mass of the tree is located. The node minimizing the mean occupation layer is the center of the mass, given that all nodes are assigned an equal weight and consecutive layers are at equidistance from one another.

This quantity, displaying information about the topology of the tree, can be computed on the whole period, or dynamically. In a dynamic analysis, a low value of the mean occupation layer reflects an homogeneous behavior of the different elements of the system under investigation.

The existence of several definitions naturally induces to discuss the choice of a specific definition and the identification of the central vertex. Following [23], we choose the mean occupation layer because the first and second definitions of the central vertex are local in nature, whereas the mean occupation layer gives a global appreciation of the topology of the tree. Even if this choice can be considered as arbitrary, it is not crucial, as the authors showed that the three definitions yield to similar and even, in most cases, identical conclusions.

3.4.4 Survival ratios

Finally we examined the robustness of the MST over time by analyzing the single step survival ratio S_R of links. This quantity refers to as the fraction of common links between two consecutive MST, at times t and $t-1$ ([23]):

$$S_R\left(t\right) = \frac{1}{N-1} \left|E\left(t\right) \cap E\left(t-1\right)\right|. \tag{3.10}$$

In this equation, E(t) refers to the set of edges of the tree at time t, $\cap$ is the intersection operator, and |..| gives the number of elements in the set. Under normal circumstances, the topology of the trees for two consecutive steps should be very stable, at least for small values of the windows length's parameter ΔT. While some fluctuations of the survival ratios might be due to real changes in the behavior of the system, it is worth noting that others may simply be due to noise. In our study, we will focus on strong fluctuations of survival ratios and examine whether or not strong trees reconfigurations do coincide with specific market events. We will also naturally examine the eventual presence of trends in the evolution of these ratios.

4

Networks analysis of the Energy sector

Contents

In this chapter we apply the methodology of minimum spanning trees on the Energy subset and mainly focus on the topological properties of the correlations. We first investigate the correlations between maturities for two different markets, namely the NCL Crude and the NHO Heating Oil. Then we examine the spatial relations between crude oil and petroleum products at different maturities.

4.1 Maturity integration of two energy markets

The test of temporal integration should reflect the presence of the Samuelson effect on the data. In an ideal case, the maturities would be perfectly organized, ranging regularly from the first to the last delivery date. Consequently, the topology of the minimum spanning trees should be linear.

4.1.1 Temporal integration of the Heating Oil (*NHO*)

In order to examine correlation between maturities for the NHO Heating Oil, we have considered two series of eighteen and thirty-six maturities. In both case, the minimum spanning tree is extracted from the distance matrix and the correlation coefficients are computed for all pairs of maturities between 07/07/98 and 10/09/09 for eighteen maturities and between 16/04/07 and 10/09/09 for the thirty-six maturities. Within these two periods we have only averaged over the days when all the contracts were traded. The figure 4.1 shows the links between one month to eighteen months' maturities 4.1 (a) and one to thirty-six months' maturities 4.1 (b). In both cases, the topology of the minimum spanning trees are identical. The trees are linear, without branching point, and the maturities are perfectly ordered. This results act as a test for our methodology as a linear structure is expected from the Samuelson effect.

4.1.2 Temporal integration of the Crude Oil (*NCL*)

We have then studied three series of maturities for the NCL. We have observed one to fifteen months' maturities between 06/21/89 and 07/29/09, one to eighteen maturities between 07/16/90 and 07/29/09 and one to twenty-eight months' maturities (plus thirty-six, forty-eight and sixty months' maturities) between 03/20/97 and 07/29/09.
For the three series, the topology are similar to those obtained for the NHO Heating Oil, and are roughly linear with ordered maturities 4.2 (a, b) and 4.3. A comparison of the

HEATING OIL US-US

(a)

1M-2M-3M-4M-5M-6M-7M-8M-9M-10M-11M-12M-13M-14M-15M-16M-17M-18M

(b)

1M-2M-3M-4M-5M-6M-7M-8M-9M-10M-11M-12M-13M-14M-15M-16M-17M-18M

36M-35M-34M-33M-32M-31M-30M-29M-28M-27M-26M-25M-24M-23M-22M-21M-20M-19M

Figure 4.1: Links between maturities for NHO Heating Oil. Figure (a) eighteen maturities between 07/07/98 and 09/10/09. Figure (b) thirty-six maturities between 16/04/07 and 09/10/09.

three minimum spanning trees indicates that when longer maturities are added to the graph, the shortest maturities are more ordered. As the longer maturities appear at the end of our database, this ordering process of the shortest maturities could be interpreted as the result of the maturation processus of the market.

We have already noticed that the different maturities of the NHO Heating Oil are perfectly ordered. It could be very interesting to test over a longer period for the NHO Heating Oil if both the maturation of the market and the add of longer maturities tends to stabilize the shortest ones.

The different behavior at long maturity (more than fifteen months) is surprising. It is unexpected that the maturities of the NCL Crude are less ordered than the NHO Heating Oil, while the NCL Crude has a more important activity. Different arguments can be given to explain this empirical observation:

- The is a problem with the temporal series extracted from DATASTREAM

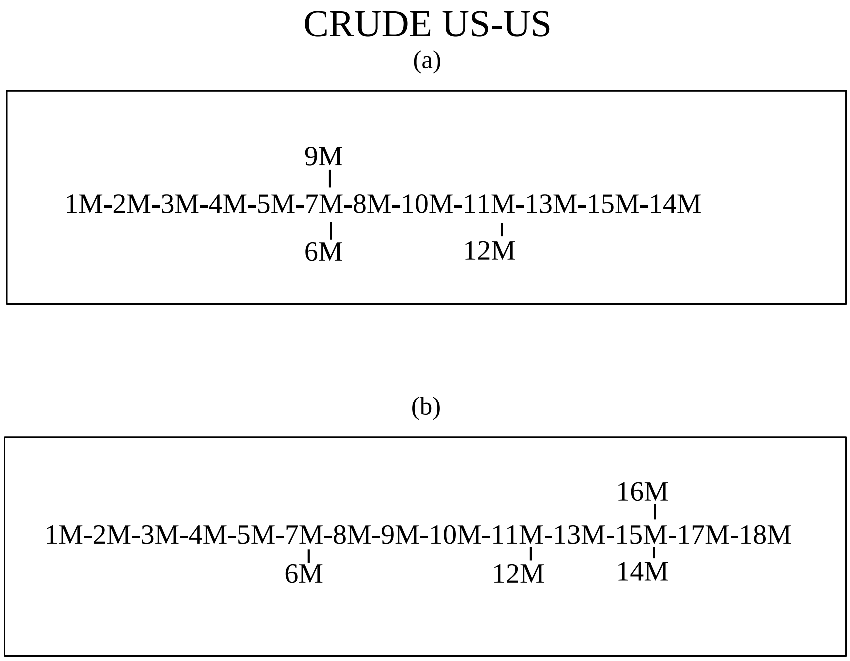

Figure 4.2: Links between maturities up to fifteen maturities (a) and eighteen maturities (b).

- the long maturities are less reliable for the crude oil rather than the heating oil (which would be a surprise). We will have to check in our next investigations the volume of contracts traded for the long maturities.

- the maturities form blocks that act in cooperation. There is a first block between one to fifteen months (moreover the fifteen months' maturity is important for the market). There is a second block of maturities between sixteen and twenty (including the key maturity eighteen month). Finally, there is a last group made up of long maturities with the particular twenty-four months' maturity astonishing localized at the periphery of the tree.

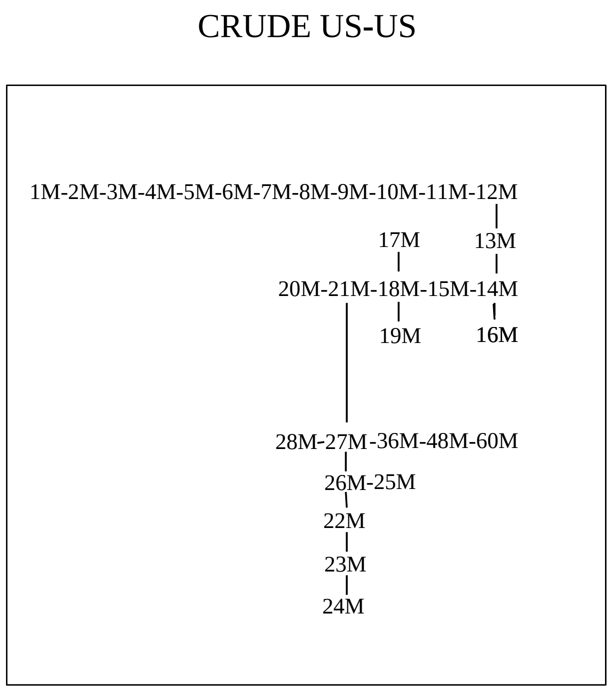

Figure 4.3: Links between maturities up to sixty months.

4.2 Spatial integration of energy markets

We study in this section the spatial aspect of the integration of the three futures contracts on crude oil and the two futures contracts on petroleum products between 21/04/06 and 07/29/09.

In order to give more insight on the empirical relationship linking the market, we realized several tests on different maturities, from one to three months, six months and twelve months.

4.2.1 Maturities 1, 2 and 3 months

The figure 4.4 represents the link between the markets in the minimal spanning tree. For the maturity one to three months, the topology of the tree is the unchanged and the

NCL − LTC − LLC − LHO − NHO

Figure 4.4: Links between the five energy markets NCL, LLC, LTC, NHO, LHO at maturity one month.

different qualities are well separated.

The two commodities NCL and LTC are naturally connected because it is the same quality but traded in different place. This first link is a pure geographical connection.

Then, there is a connection between LTC and LLC based on the quality. The two commodities are traded at the same place (London), but there is an American crude oil and an European crude oil.

The next link is an upstream-downstream relation, in term of industrial process, with the crude oil traded in London as input and the heating oil traded in London as output.

The last edge in the graph is again a purely spatial link between the two heating oil LHO and NHO. The latter connection is surprising while one could imagine a preferred link between the American crude oil NCL and the American heating oil NHO. A possible explanation is the presence of noise for the short maturities prices.

Let us notice that while the topology of the trees does not change between one and three months but the length of the tree decreases with the maturity.

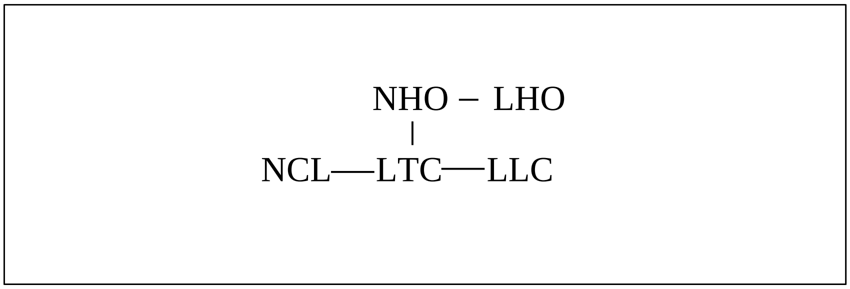

Figure 4.5: Links between the five markets NCL, LLC, LTC, NHO, LHO at maturity six months.

4.2.2 Maturity 6 months

The links between the five markets at a maturity six months are given on the figure 4.5. While the topology remained unchanged for the three first maturities, at six months the tree is no more linear and a node with a connectivity equal to three appears. The most connected market is the LTC Crude. The latter acts as pivot in the system and is *between* NCL and LLC in term of quality and trading place. Moreover, the LTC connects the two heating oil, first the LHO traded at the same place and then the NHO.

4.2.3 Maturity 12 months

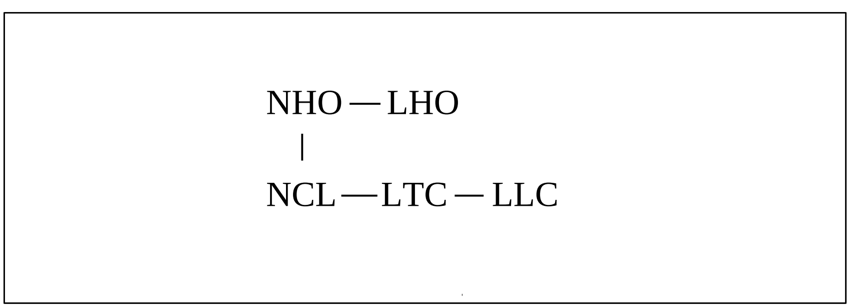

Figure 4.6: Links between the five markets NCL, LLC, LTC, NHO, LHO at maturity twelve months.

The relations between the prices at the maturity twelve months are represented on the figure 4.6. The topology of the networks is again modified. The pivot, which allows to move from the crude oil to the heating oil is no more the American crude oil negotiated in London but the American crude oil traded in United States. Furthermore, the NHO is before the LHO. As the NCL is the most important market, it is relevant that the latter plays the role of the pivot when long maturities are considered. On the other hand, it is also interesting to notice that the distance between the NHO and the LHO is smaller than the distance between the LLC and LHO. We can have a intuition, further tests are necessary to conclude, that the links between financial markets have the upper hand over the prices behavior of the same commodity but traded in different places. The financial markets, easily arbitrable could have stronger links rather than the upstream/downstream industrial link.

We can summarize all the results by plotting the path length of each minimum spanning tree as a function of the maturity. The figure 4.7 shows that the total length of the minimum spanning tree is decreasing function the maturity. The latter result implies that the edges of the graph become shorter. We have also compute the average correlation $\langle C_{ij} \rangle_{i \sim j}$, where $\langle ... \rangle_{i \sim j}$ denotes the average over the edges of the graph. $\langle C_{ij} \rangle_{i \sim j}$ increases with the maturity and tends to a value close to 1. We can highlight that the value of the average correlation is high because we are considered edges of the minimum spanning tree, and then the most correlated markets. But the relevant point is that the closest markets become more and more correlated at long maturity.

4.3 Conclusion

In this chapter, we have estimated and quantified, through the scope of the minimum spanning tree, the spatial integration of the five most important energy markets. Our results are full of promise and we are confident to extract relevant economic information

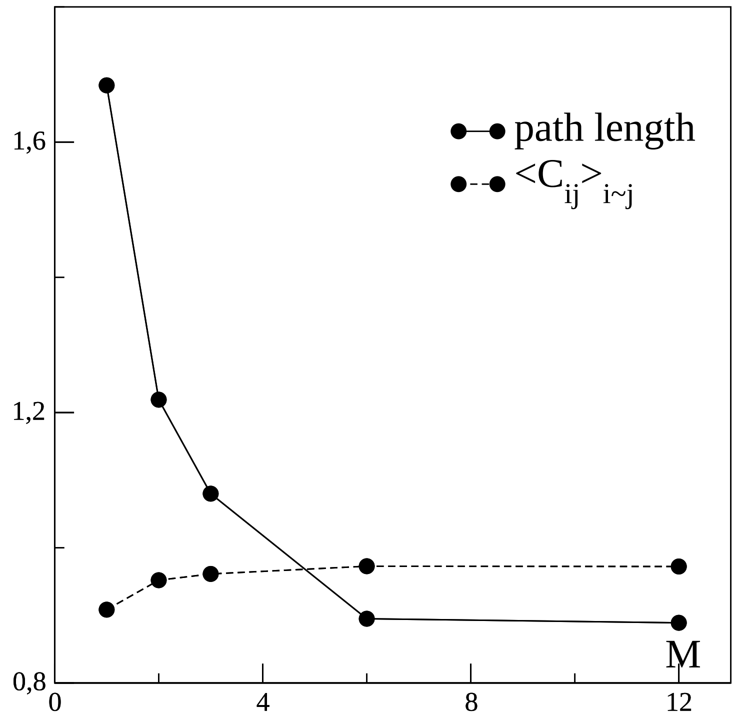

Figure 4.7: Path length (black line) and average correlation (dotted line) as a function of the maturity M.

from the complex topology of commodities networks.

Our main result is that the links between markets, the edges of the minimum spanning tree, have an economical interpretation that satisfies the intuition. We interpret this result as a positive test for the relevance of our method and the application to derivative markets.

We also observed that the strength of the integration increases with the maturity. The latter result is original and has not been yet mentioned in other works. In particular, the authors of [29] saw the spatial links between oil markets but miss the information into the maturity dimension. The increase of the correlation with the maturity is coherent with the Samuelson effect. As we are considering long maturity, the noise vanishes, and the markets follow fundamental rules behavior.

5

Analysis of the systemic risk in the spatial, maturity and spatio-maturity dimensions

Contents

In this chapter we present the network analysis of fourteen derivatives markets in different dimensions. In the first section we briefly remind the principal tools used in the networks analysis. In order to study the integration of derivative markets, we rely on the graph-theory. Among the different tools provided by this method, we selected those allowing us to analyze market integration using a three-dimensional approach. We first decided to represent our prices system by studying the correlation of price returns. Having transformed these correlations into distances, we were able to draw a fully connected graph of the prices system, where the nodes of the graph represent the time series of futures prices. In order to filter the information contained in the graph, we rely on specific graphs: minimum spanning trees (MST). The method used for the identification of the MST is presented in the first part of this chapter. We then study the topology of the trees (second section) and their dynamic behavior (third section).

5.1 Reminder on minimum spanning trees

The first step towards the analysis of market integration was in our case the computation of the synchronous correlation coefficients of price returns. In order to use the graph theory, we needed to quantify the distance between the elements under examination. We thus extracted a metric distance from the correlation matrix. We were then able to construct graphs. Lastly, we used a filtering procedure in order to identify the MST [18]. Such a tree can be defined as the one providing the best arrangement of the network's different points.

5.1.1 The correlation matrix

In order to measure the similarities in the synchronous time evolution of the futures contracts, we built a matrix of correlation coefficients. The latter are defined as follows:

$$\rho_{ij}(t) = \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\sqrt{\left(\langle r_i^2 \rangle - \langle r_i \rangle^2\right)\left(\langle r_j^2 \rangle - \langle r_j \rangle^2\right)}}, \tag{5.1}$$

When focusing on the spatial dimension, i and j stand for the nearby futures prices of pairs of assets, like crude oil or corn. In the absence of reliable spot data, we approximate the spot prices with the nearest futures prices. When focusing on the maturity dimension, they stand for pairs of delivery dates. They are a mix of the two when a three-dimensional analysis is performed. The daily logarithm price differential stands for price returns, with $r_i = (\ln F_i(t) - \ln F_i(t - \Delta t)) / \Delta t$, where $F_i(t)$ is the settlement price of the futures contract at t. Δt is the time window, and $\langle . \rangle$ denotes the statistical average performed other time, on the trading days of the study period.

For a given time period and a given set of data, we thus computed the matrix of $N \times N$ correlation coefficients C, for all the pairs ij. C is symmetric with ρ_{ij} when $i = j$. Thus, is characterized by $N(N-1)/2$ coefficients.

5.1.2 From correlations to distances

In order to use the graph-theory, we needed to introduce a metric. The correlation coefficient ρ_{ij} indeed cannot be used as a distance d_{ij} between i and j because it does not fulfill the three axioms that define a metric [11]:

- $d_{ij} = 0$ if and only if $i = j$,
- $d_{ij} = d_{ji}$
- $d_{ij} \leq d_{ik} + d_{kj}$

A metric d_{ij} can however be extracted from the correlation coefficients through a non linear transformation. Such a metric is defined as follows:

$$d_{ij} = \sqrt{(2\,(1-\rho_{ij}))}. \tag{5.2}$$

A distance matrix D was thus extracted from the correlation matrix C according to Equation (5.2). C and D are both $N \times N$ dimensional. Whereas the coefficients ρ_{ij} can be positive for correlated returns or negative for anti-correlated returns, the distance d_{ij} representing the distance between price returns is always positive.

5.1.3 From full connected graphs to Minimum Spanning Trees (MST)

A graph gives a representation of pairwise relationships within a collection of discrete entities. A simple connected graph represents all the possible connections between N points under examination with $N-1$ links (edges). Each point of the graph constitutes a node (vertex). The graph can be weighted in order to represent the different intensities of the links and / or nodes. Such weights can represent the distances between the nodes. In order to understand the organizing principles of a system through its representation as a graph, it needs to be spanned, *i.e.* all its nodes need to be traversed. However, there are a lot of paths spanning a graph. For a weighted graph, the minimum spanning tree (MST) is the one spanning all the nodes of the graph, without loops. It has less weight than any other tree. Its links are a subset of those of the initial graph.

Through a filtering procedure (the information space is reduced from $N(N-1)/2$ to $N-1$), the MST reveals the most relevant connections of each element of the system. In our study, they provide for the shortest path linking all nodes. Thus, they can be seen as a way of revealing the underlying mechanisms of systemic risk: the minimal spanning tree is indeed the easiest path for the transmission of a prices shock.

5.2 Topology of the Minimum Spanning Trees: empirical results

The first information given by a minimum spanning tree is the kind of arrangement found between the vertices. Therefore, the first step in studying MST lies in their visualization. We then use the allometric coefficients method in order to determine whether a MST is totally organized, totally random, or is situated somewhere between these two extreme topologies. In this part of the study, we consider the whole time period as a single window and thus perform a static analysis.

5.2.1 Visualization and description of the MST

The visualization of the trees is a very important step, as it addresses the meaningfulness of the taxonomy that emerges from the system. Before going further, let us make two remarks: first, we are considering links between markets and/or delivery dates belonging to the MST. Thus, if a relationship between two markets or maturities does not appear in the tree, this does not mean that this relation does not exist. It just does not correspond to a minimal distance. Second, our results naturally depend on the nature and number of markets chosen for the study.

In what follows, we will use the term "sector" in order to describe the grouping of underlying assets, whereas we will retain the term "cluster" in order to describe, for a specific market, the grouping of delivery dates in the maturity dimension.

Figure (5.1) presents the MST obtained for the spatial and for the maturity dimensions. As far as the spatial dimension is concerned, the MST looks like a star. In Figure (5.1)-a the three sectors can be identified: energy is at the bottom. It gathers American as well as European markets and is situated between agricultural (on the left) and financial assets (mainly on the right). Moreover, the most connected node in the graph is European crude oil (LLC), which makes it the best candidate for the transmission of price fluctuations in

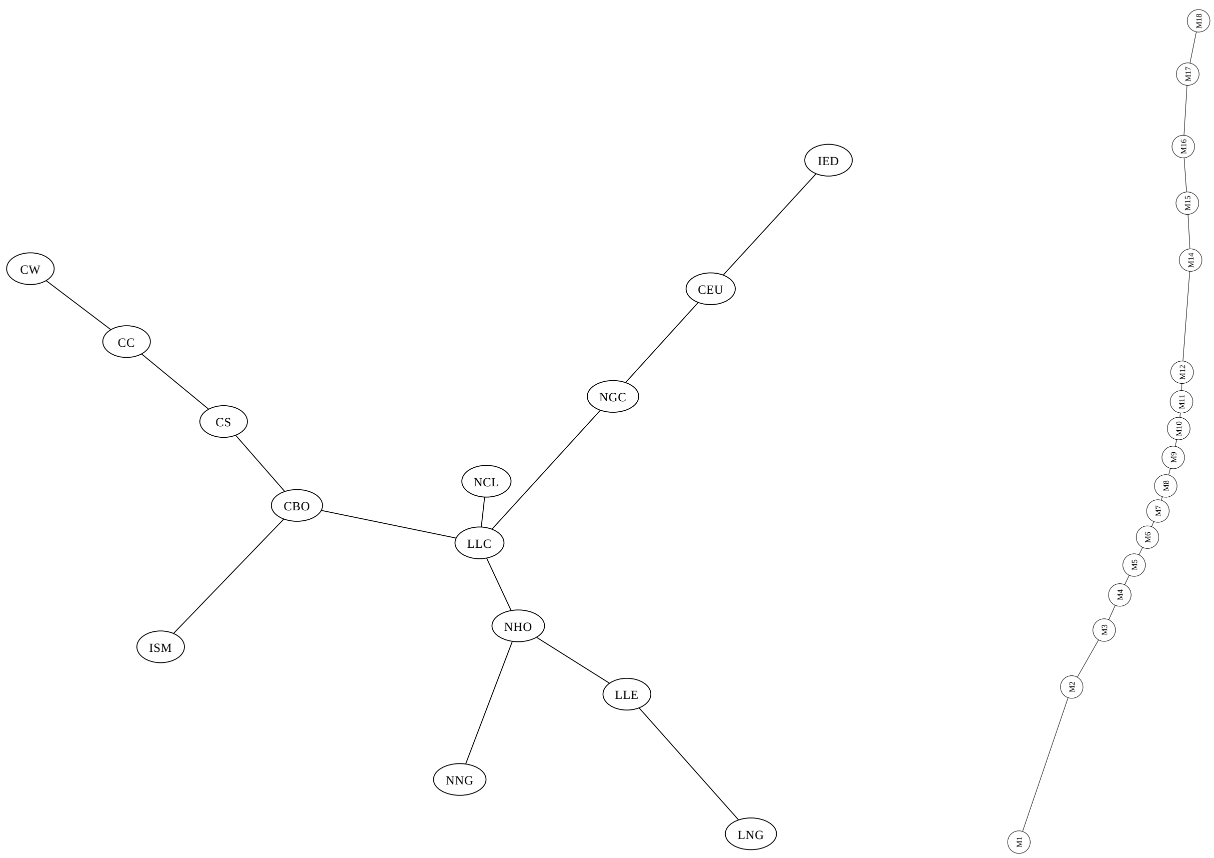

Figure 5.1: Static minimum spanning trees. Left panel: MST for the spatial dimension, built from the correlation coefficients of prices returns, 30/04/01-01/08/09. Right panel: MST on the maturity dimension, built from the correlation coefficients of the Brent crude oil LLC, 01/04/2000-06/11/09.

the tree (actually, the same could have been said for American crude oil (NLC), as the distance between these products is very short). Last but not least, the energy sector seems the most integrated, as the distances between the nodes are short. The link between the energy and agricultural products passes through soy oil (CBO). This is interesting, as the latter can be used for fuel. The link between commodities and financial assets passes through gold (NGC), which can be seen as a commodity but also as a reserve of value. The only surprise comes from the S&P500 (ISM), which is more correlated to soy oil (CBO) than to other financial assets.

Such an organization leads to specific conclusions regarding systemic risk. Let us suppose that a prices shock reaches interest rates (IED). The star-like organization of the tree does not enable us to determine whether this shock comes from the energy or the agricultural sectors. Things are totally different in the maturity dimension.

In this case, it was not possible to give an illustration for each tree, as the database gathers 14 futures contracts. We thus retained a representative tree, that of Brent crude (LLC). The latter is illustrated by Figure (5.1)-b. The MST is linear and the maturities are regularly ordered from the first to the last delivery dates.

The analysis on the maturity dimension gives rise to three remarks. Firstly, this linear topology reflects the presence of the *Samuelson effect*. In derivative markets, the movements in the prices of the prompt contracts are larger than the other ones. This results in a decreasing pattern of volatilities along the prices curve. Secondly, this type of organization impacts the possible transmission of prices shocks. The most likely path for a shock is indeed unique and passes through each maturity, one after the other. Thirdly, the short part of the curves are generally less correlated with the other parts. This phenomenon can result from prices shocks emerging in the physical market with the most nearby price being the most affected; it could also reflect noises introduced on the first maturity by investors in the derivative market.

Let us now turn to the three-dimensional analysis. Figure (5.2) represents the 3-D static MST. Its shape brings to mind that observed in the spatial dimension. However, it is enhanced by the presence of the different maturities available for each market. The latter are clearly linearly organized. As previously, the tree shows a clear separation between the sectors. Three energy contracts, American crude oil (NCL), European crude oil (LLC) and American heating oil (NHO) are found at the center of the graph. They are the three closest nodes of the graph. Moreover, the agricultural sector is no longer linked to gold. It is now directly linked to American crude oil (NCL).

It would have been interesting to know which maturities connect two markets or sec-

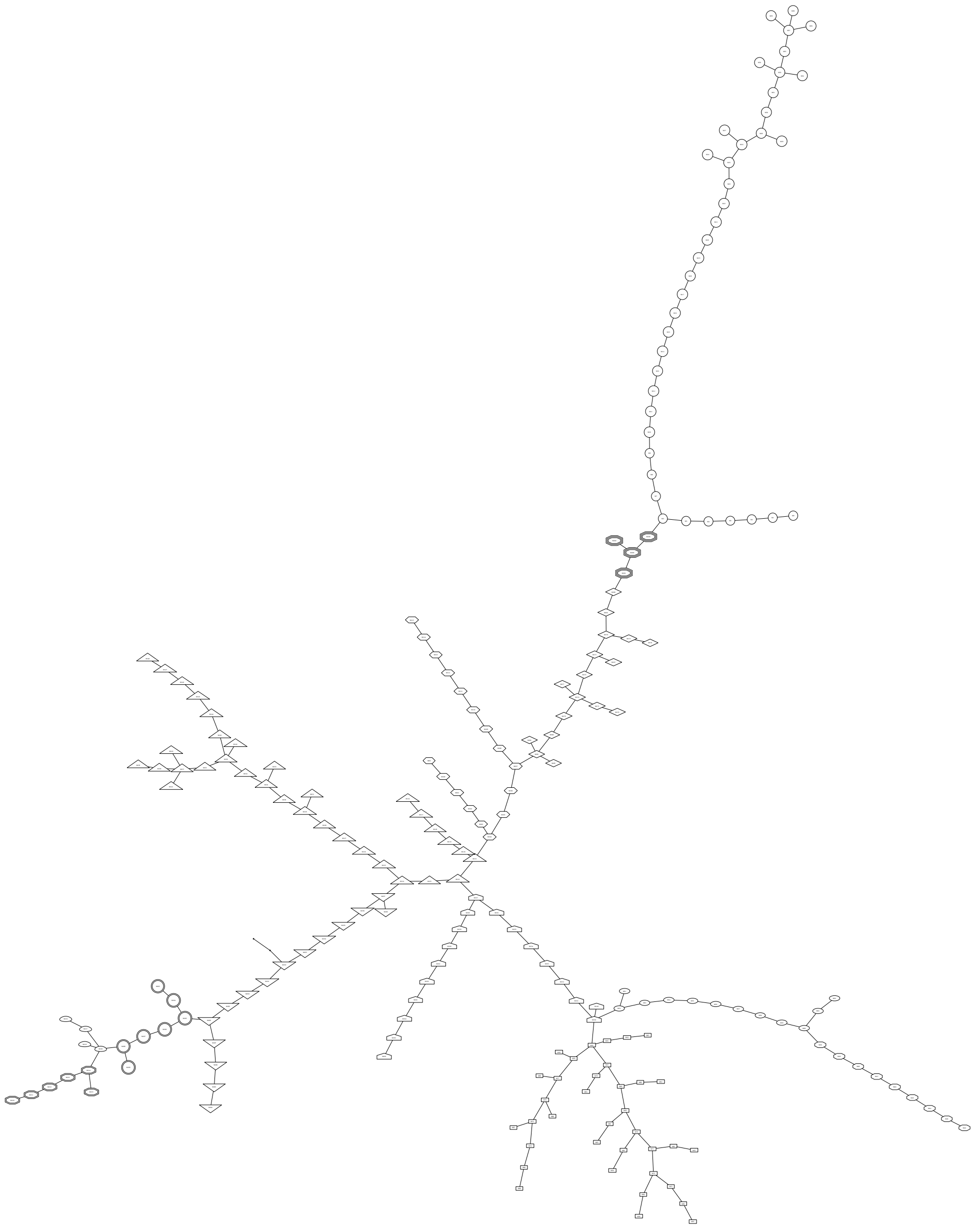

Figure 5.2: Static minimum spanning tree for the three-dimensional analysis, 27/06/2000-12/08/2009. The different futures contracts are represented by the following symbols: empty circle: *IED*, point: *ISM*, octagon: *LNG*, ellipse: *LLE*, box: *NNG*, hexagon: *LLC*, triangle: *NCL*, house: *NHO*, diamond: *NGC*, inverted triangle: *CBO*, triple octagon: *CEU*, double circle: *CS*, double octagon: *CW*, egg: *CC*. For a given futures contract, all maturities are represented with the same symbol. The distance between the nodes is set to unity.

Allometric coefficient		
Name	**Static coefficient**	**Dynamical coefficient**
IED	$1{,}927 \pm 0{,}056$	$1{,}913 \pm 0{,}011$
LNG	$1{,}874 \pm 0{,}002$	$1{,}886 \pm 0{,}059$
LLE	$1{,}88 \pm 0{,}003$	$1{,}943 \pm 0{,}02$
NNG	$1{,}75 \pm 0{,}037$	$1{,}774 \pm 0{,}018$
LLC	$1{,}889 \pm 0{,}003$	$1{,}904 \pm 0{,}095$
NCL	$1{,}994 \pm 0{,}045$	$1{,}906 \pm 0{,}013$
NGC	$1{,}732 \pm 0{,}092$	$1{,}908 \pm 0{,}013$
CBO	$1{,}889 \pm 0{,}003$	$1{,}886 \pm 0{,}032$
CS	$1{,}848 \pm 0{,}095$	$1{,}822 \pm 0{,}095$
CW	$1{,}864 \pm 0{,}13$	$1{,}761 \pm 0{,}125$
CC	$1{,}88 \pm 0{,}003$	$1{,}834 \pm 0{,}024$
Spatial	$1{,}493 \pm 0{,}056$	$1{,}621 \pm 0{,}024$
3-D	$1{,}757 \pm 0{,}023$	$1{,}85 \pm 0{,}009$

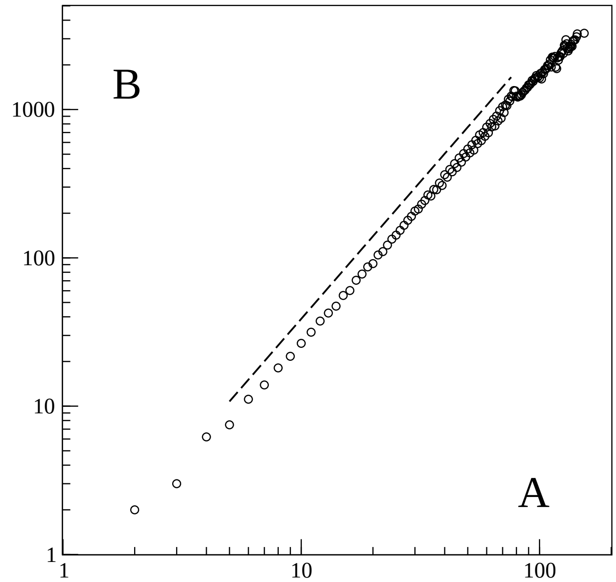

Figure 5.3: Allometric properties of the trees. Left panel: static and dynamical exponents for each futures contract (maturity dimension), as well as for the spatial and 3-D analyses. Right panel: 3-D dynamical allometric coefficients in log-log scale. The dashed line corresponds to the best fit with an exponent equal to 185.

tors. Economic intuition suggests two kinds of connections: they could appear on the shortest or on the longest part of the curves. In the first case, the price's system would be essentially driven by underlying assets; in the second one, it would be dominated by derivative markets. However, a closer analysis of the 3-D trees does not provide evidence of either kind of expected organization. Moreover, the analysis of the tree at different periods does not lead to the conclusion that there is something like a pattern in the way connections occur. Further investigations are thus necessary in order to study the links between markets and sectors more precisely. We offer an initial response to this problem at the end of this section.
1

5.2.2 Allometric properties of the MST

Star-like trees are symptomatic of a random organization, whereas chain-like trees reveal a strong structure. The computation of the allometric coefficients of the MST provides a means of quantifying the degree of randomness in the tree.

The first model of the allometric scaling on a spanning tree was developed by [1]. The first step of the procedure consists in initializing each node of the tree with the value 1. Then the root or central vertex of the tree must be identified. In what follows, the root is defined as the node having the highest number of links attached to it. Starting from this root, the method consists in assigning two coefficients A_i and B_i to each node i of the tree, where:

$$A_i = \sum_j A_j + 1 \text{ and } B_i = \sum_j B_j + A_i, \tag{5.3}$$

j stands for all the nodes connected to i in the MST. The allometric scaling relation is defined as the relationship between A_i and B_i:

$$B \sim A^{\eta}, \tag{5.4}$$

η is the allometric exponent. It represents the degree or randomness of the tree and stands between two extreme values: 1^+ for star-like trees and 2^- for chain-like trees.

Figure (5.3) summarizes the allometric properties of the MST for each dimension. The left panel reproduces the different exponents and gives the error resulting from a non-linear regression. Figure (5.3) gives an illustration of the allometric coefficients in 3-D. The dashed line corresponds to the best fit with an exponent equal to 185. The figure shows that the coefficients are well described by the power law with an exponent.

As far as the spatial dimension is concerned, the exponents indicate that even if Figure (5.1) seems to show a star-like organization, the shape of the MST is rather complex and stands exactly between the two asymptotic topologies. There is an ordering of the tree, which is well illustrated by the agricultural sector, which forms a regular branch.

Within the maturity dimension, the coefficients tend towards their asymptotic value $\eta = 2^-$. They are however a bit smaller than 2, due to finite size effects (there is a finite number of maturities). Such a result is rather intuitive but nevertheless interesting:

arbitrage operations on the futures contracts related to the same underlying asset are easy and rapidly undertaken, resulting in a perfect ordering of the maturity dates.
Even if the topology of the spatial and 3-D trees seems similar, they are quantitatively different. The allometric exponent for the three-dimensional is higher: the best fit from our data gives an exponent close to 1.757, which must be compared to the value of 1.493 for the spatial case. Thus, the topology of our system, in 3-D, is rather complex. It is the result of two driving forces: the star-like organization induced by the spatial dimension and the chain-like organization arising from the maturity dimension.

5.3 Dynamical studies of the systems

Because they are based on correlation coefficients, our Minimum Spanning Trees are intrinsically time dependent. Therefore, its is necessary to study the time dependent properties of the graphs. On the basis of the entire graph, firstly we examined the dynamical properties of the correlation coefficients, as well as the node's strength, which provides information on how far a given node is correlated to the other nodes. In order to study the robustness of the topology of the MST, we then computed the graph's length, which reveals the state of the system at a specific time. Lastly, survival ratios indicate how the topology of the trees evolves over time. In what follows, we retained a rolling time window with a size of $\Delta T = 480$ consecutive trading days.

5.3.1 Correlation coefficients

In order to examine the time evolution of our system, we investigated the mean correlations of the returns and their variances ([29]). The mean correlation $C^T(t)$ for the correlation coefficient ρ_{ij}^T in a time window $[t-\Delta T, t]$ can be defined as follows:

$$C^T(t) = \frac{2}{N(N-1)} \sum_{i<j} \rho_{ij}^T(t), \tag{5.5}$$

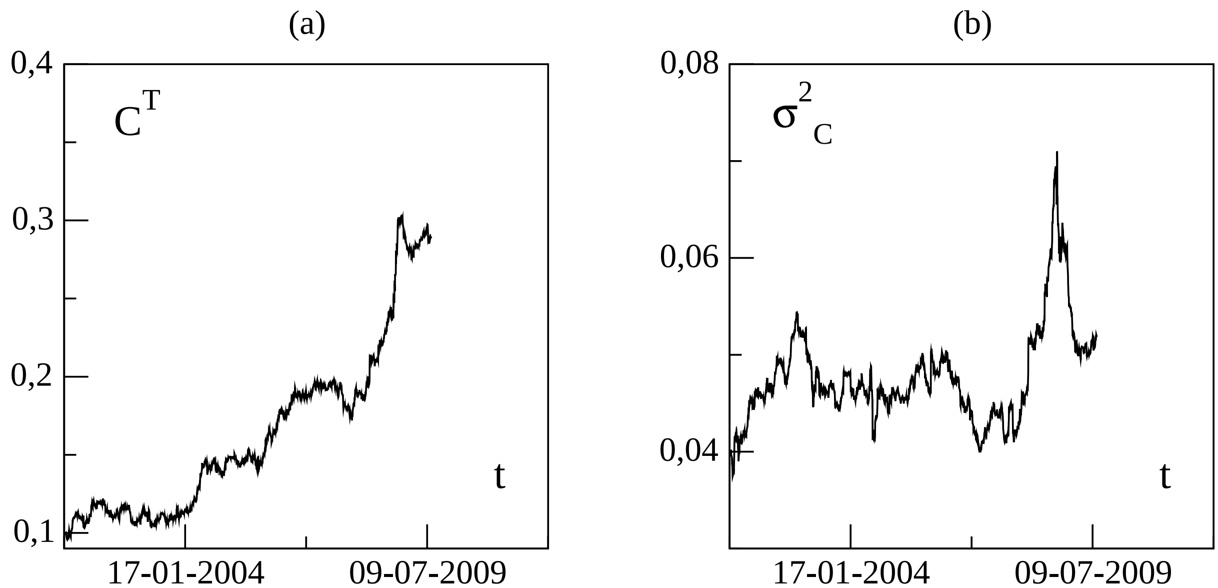

Figure 5.4: Correlation coefficients in the spatial dimension. Figure (a): Mean of the correlation coefficients; Figure (b): Variance of the correlation coefficients.

The variance $\sigma_C^2(t)$ of the mean correlation is given by:

$$\sigma_C^2(t) = \frac{2}{N(N-1)} \sum_{i<j} \left(\rho_{ij}^T(t) - C^T(t)\right)^2 . \tag{5.6}$$

Figure (5.4) represents the mean correlation and its variance on the spatial dimension. It shows that the mean correlation of the prices system increases over time, especially after 2007. The variance exhibits a similar trend. Moreover, it reaches its maximum on the 09/19/2008, four days after the Lehman Brothers' bankruptcy.

We then examine the maturity dimension. Firstly, we focus on the statistical properties of the correlation coefficients of two futures contracts, represented by Figure (5.5). They are very different for these contracts. The maturities of Brent crude oil (LLC) are more and more integrated over time: at the end of the period, the mean correlation is close to 1. Such a trend does not appear for the eurodollar contract (IED). This is consistent with the peripheral position of the interest rate market in the correlation landscape. As far as crude

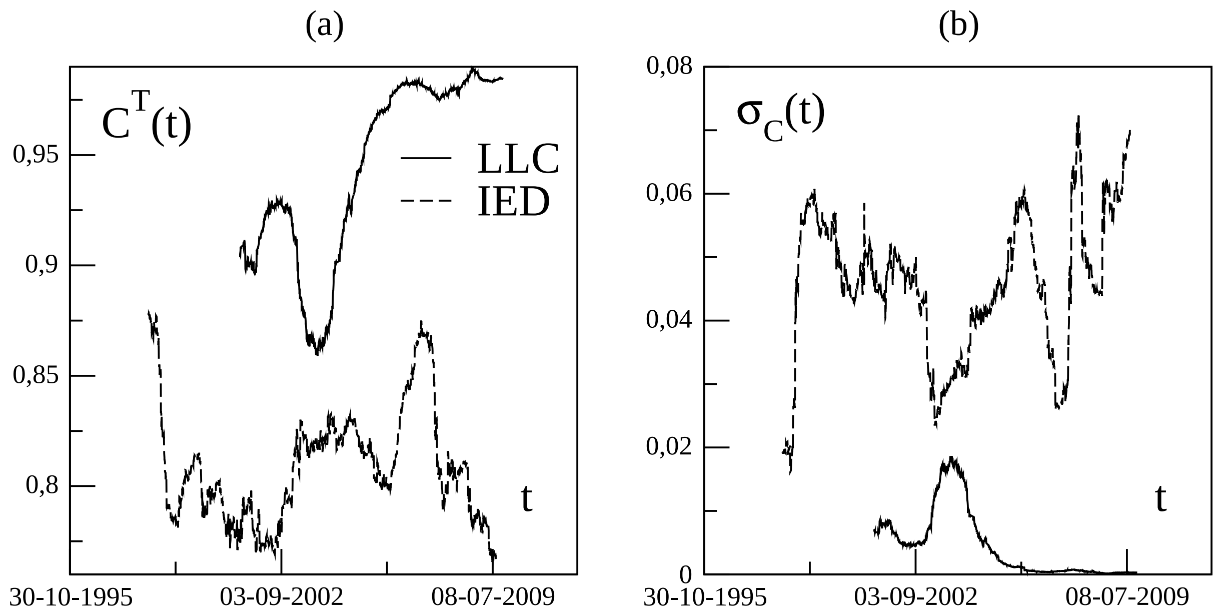

Figure 5.5: Correlation coefficients in the maturity dimension for the eurodollar *IED* (dashed lines) and the Brent crude oil *LLC* (black lines). Figure (a): Mean of the correlation coefficients; Figure (b): Variance of the correlation coefficients.

oil is concerned, the level of integration becomes so strong that the variance decreases and exhibits an anti correlation with the mean correlation. The result was totally different in the spatial case: the mean correlation and its variance where correlated ([22]) also observe such a positive correlation during prices growth and financial crises).

Figure (5.7) summarizes the statistical properties of the mean correlations and variances for the 14 markets, on the maturity dimension. It confirms that, for almost every contract, the mean correlation is very high and anti correlated with the mean variance. The two natural gases however exhibit more specific figures. Their correlation level is quite low, when compared with other markets, especially for London Natural Gas. Meanwhile, their mean variance is high.

Merging space and maturity, in three dimensions, we also observe an important rise in the mean correlation and variance, as shown in Figures (5.6)-a and (5.6)-b. Moreover, these values are correlated.

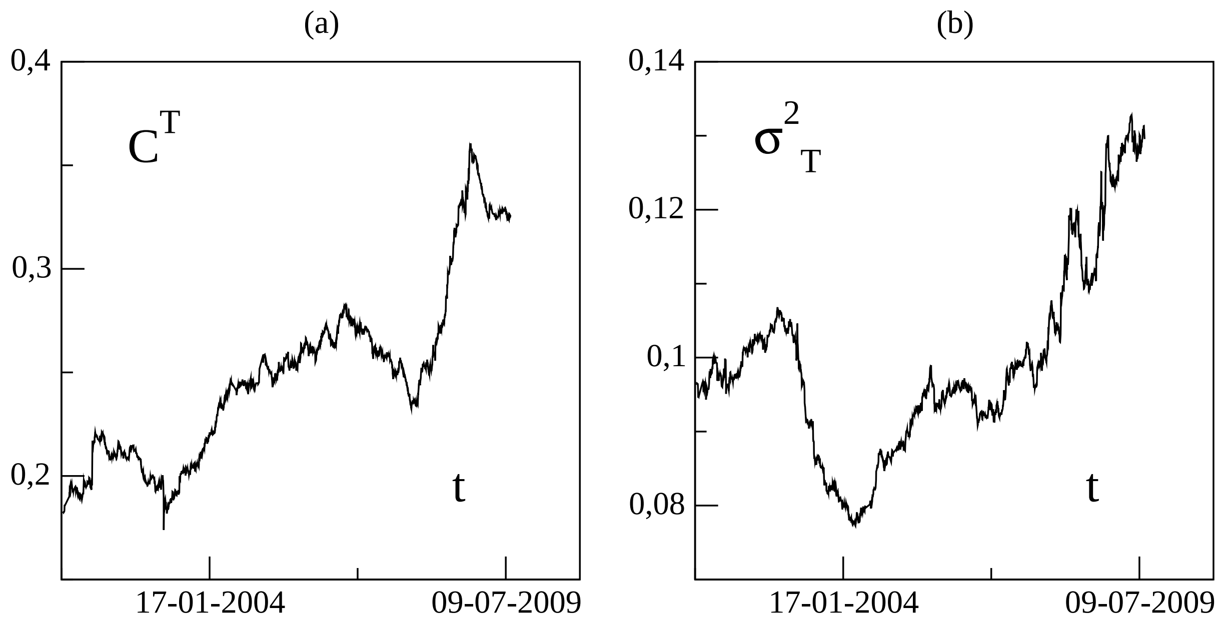

Figure 5.6: Correlation coefficients in three dimensions. Figure (a): Mean of the correlation coefficients. Figure (b): Variance of the correlation coefficients.

5.3.2 Node's strength

The node's strength, calculated for each node i, indicates the closeness of one node i to the others. It is defined as follows:

$$S_i = \sum_{i \neq j} \frac{1}{d_{ij}}. \tag{5.7}$$

In our case, the node's strength provides information on the intensity of the correlations linking a given node to the others. When S_i is high, the node is close to the others.
Figure (5.8) represents the time evolution of the node's strength for each node within the fully connected graph, in the spatial dimension. The figure has been separated into four panels: the energy sector is at the top, with American products on the left and European products on the right, the agricultural sector is at the bottom left and financial assets are

Energy Futures Contracts					
Mnemonic	**Mean**	**Variance**	**Correlation**	**Min**	**Max**
NCL	0,936	$0{,}984\ 10^{-3}$	$-0{,}981$	0,863	0,973
LLC	0,945	$0{,}167\ 10^{-2}$	$-0{,}966$	0,859	0,99
NNG	0,617	0,0226	$-0{,}962$	0,393	0,855
LNG	0,254	0,0275	$-0{,}887$	$-0{,}169\ 10^{-3}$	0,601
LLE	0,950	$0{,}553\ 10^{-3}$	$-0{,}921$	0,891	0,981

Financial Futures Contracts					
Mnemonic	**Mean**	**Variance**	**Correlation**	**Min**	**Max**
NGC	0,983	$0{,}140\ 10^{-3}$	$-0{,}882$	0,939	0,996
IED	0,809	$0{,}690\ 10^{-3}$	$-0{,}709$	0,765	0,878

Agricultural Futures Contracts					
Mnemonic	**Mean**	**Variance**	**Correlation**	**Min**	**Max**
CBO	0,940	$0{,}199\ 10^{-2}$	$-0{,}972$	0,826	0,997
CC	0,84	$0{,}330\ 10^{-2}$	$-0{,}962$	0,709	0?902
CS	0,912	$0{,}279\ 10^{-2}$	$-0{,}797$	0,769	0,974
CW	0,918	$0{,}232\ 10^{-2}$	$-0{,}943$	0,814	0,993

Figure 5.7: Correlation coefficients in the maturity dimension. Mean correlation coefficients and their variances for all markets.

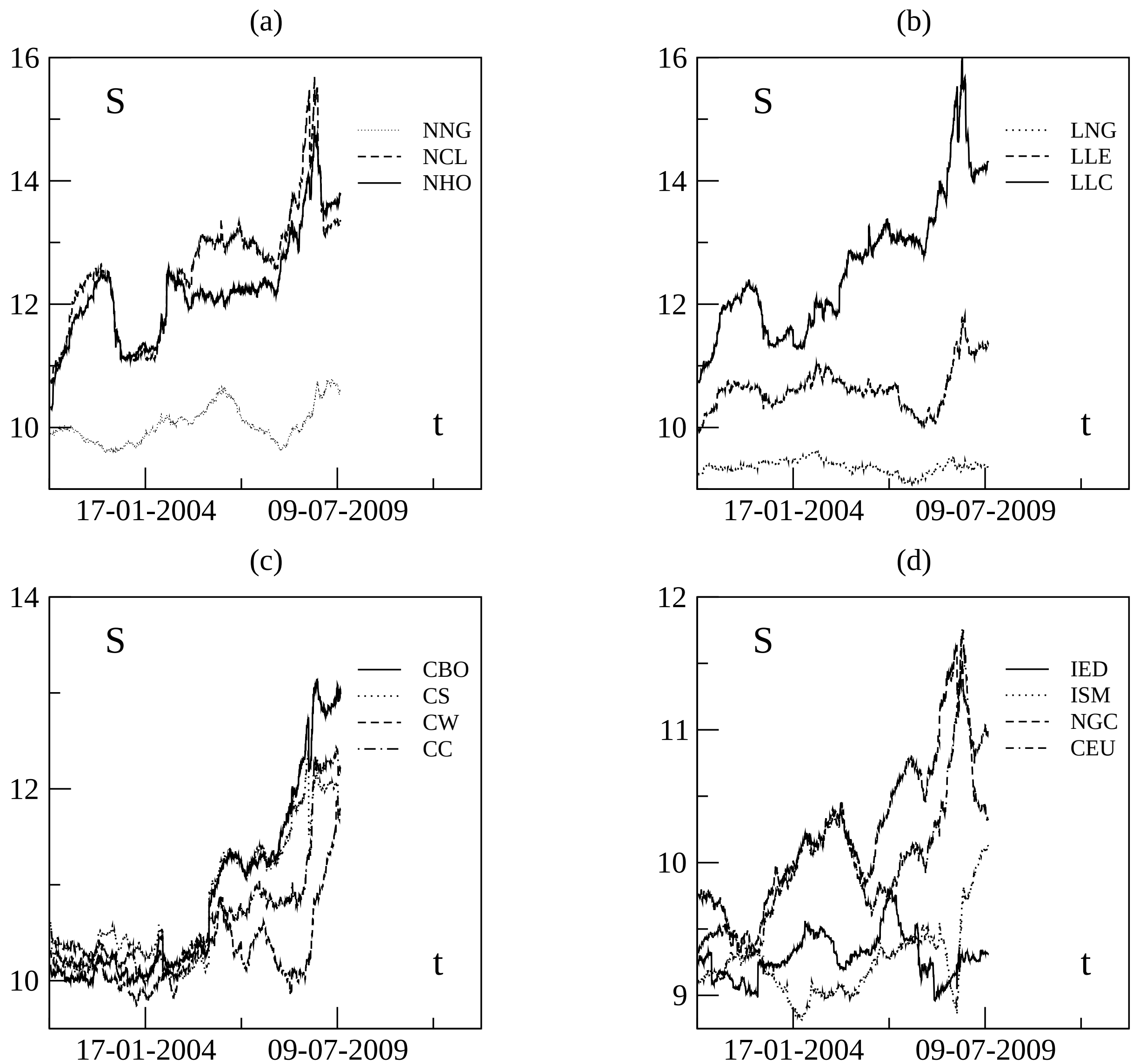

Figure 5.8: Nodes strength of the markets in the spatial dimension. Figure (a): American energy products. Figure (b): European energy products. Figure (c): Agricultural products. Figure (d): Financial assets.

at the bottom right.

Figure 5.8) prompts the following remarks: at the end of the period, out of all the assets studied, the two crude oils and American heating oil show the greatest node's strength. These are followed by soy oil (CBO), other agricultural assets, the S&P500 contract

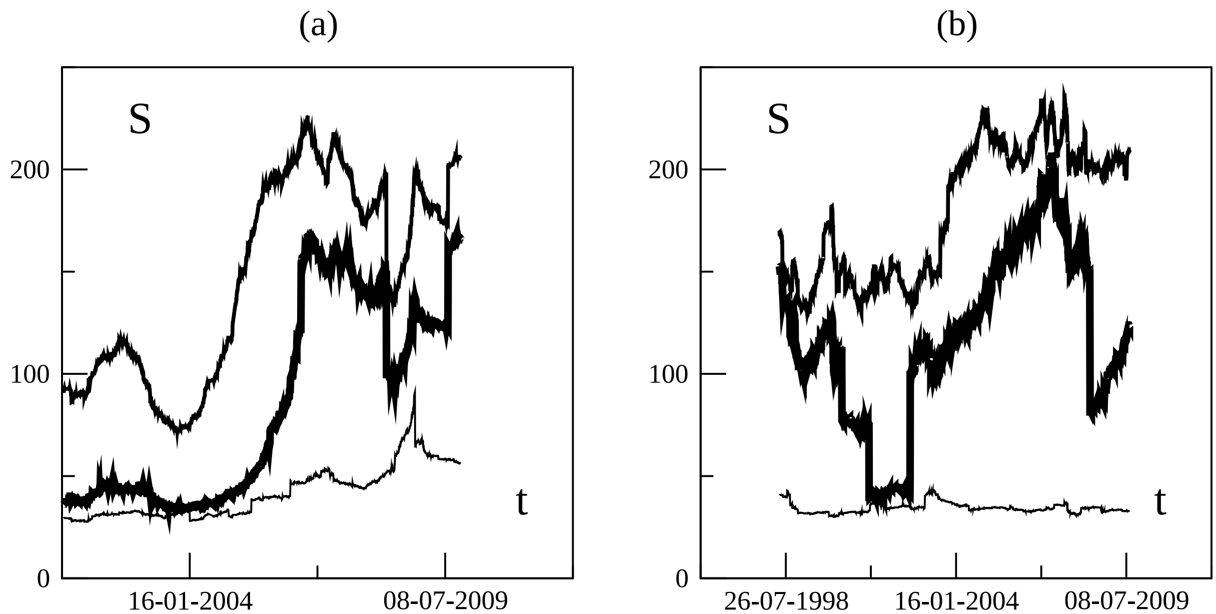

Figure 5.9: Node strength in the maturity dimension, for three maturities. Figure (a): Eurodollar *IED*. Figure (b): Brent *LLC* (b).

(*ISM*), gold (*NGC*), the euro dollar exchange rate (*CEU*) and European gas oil (*LLE*). The more distant nodes are those representing the eurodollar (*IED*) and natural gases (*NNG* and *LNG*).

When the time evolution of this measure is concerned, the sector shows different patterns: the integration movement, characterized by an increase in the node's strength, emerges earlier for the energy sector than for the agricultural sector. However, it decreases for energy at the end of the period, which is not the case for agricultural products. Moreover, the agricultural sector behaves very homogeneously, with a high increase after October 2005. Last but not least, most of the products exhibit a strong increase, except for natural gases and interest rate contracts. Thus, whereas the core of the tree becomes more and more integrated, the peripheral assets do not follow this movement.

As far as the maturity dimension is concerned, it was not possible to represent the node's strength for all futures contracts. Moreover, the computation of mean node's strength, on all maturities for each contract, would lead to the same kind of results as those provided by

Figure (5.7). Therefore, we again retained, the Brent crude oil (LLC) and the eurodollar contract (IED) examples. We then chose three delivery dates for these contracts, as shown in Figures (5.9)-a and (5.9)-b. The first maturity is drawn with a fine line, the last maturity with a wide line and the intermediary maturity with a medium width line. All the observed nodes' strength grow over time, except for the first eurodollar (IED) maturity. Moreover, in each case, the strongest node is the one which corresponds to the intermediary maturity, whereas the weakest one represents the first maturity.

5.3.3 Normalized tree's length

Let us now examine some of the properties of the filtered information. The normalized tree's length can be defined as the sum of the lengths of the edges belonging to the MST:

$$\mathcal{L}(t) = \frac{1}{N-1} \sum_{(i,j)\in MST} d_{ij}, \tag{5.8}$$

where t denotes the date of the construction of the tree and $N-1$ is the number of edges in the MST. The length of a tree is longer as the distances increase, and consequently when correlations are low. Thus, the more the length shortens, the more integrated the system is.

Figure (5.10)-a represents the dynamic behavior of the normalized length of the MST in its spatial dimension. The general pattern is that the length decreases, which reflects the integration of the system. This information confirms what was observed on the basis of the node's strengths. However we must remember that we are now analyzing a filtered network. Thus, what we see on Figure (5.10)-a is that the most efficient transmission path for price fluctuations becomes shorter as times goes on. From a systemic point of view, this means that a prices shock will be less and less absorbed as it passes through the tree. A more indepth examination of the graph also shows a very important decrease between October 2006 and October 2008, as well as significant fluctuations in September

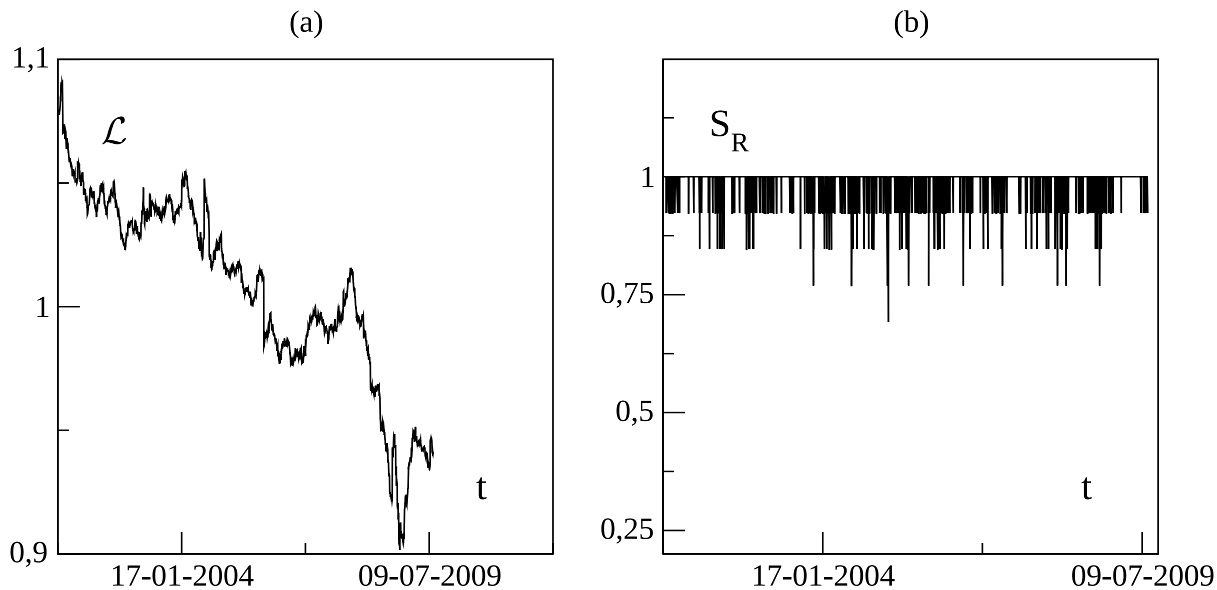

Figure 5.10: Spatial dimension. Figure (a): Normalized tree's length. Figure (b): Survival ratios.

and October 2008. We leave the analysis of such events for future studies.

In the maturity dimension, as integration increases, the normalized tree's length also diminishes. This phenomenon is illustrated by Figures (5.11)-a and -b, which represent the evolutions recorded for the eurodollar contract (*IED*) and for Brent crude (*LLC*). As far as the interest rate contract is concerned, the tree's length first increases, then in mid-2001 it drops sharply and remains fairly stable after that date. For crude oil, the decrease is constant and steady , except for a few surges.

Figure (5.12) summarizes the main results concerning the tree's length for each futures contract. However, it is not easy to compare the tree's lengths of futures contracts when the latter have a different number of delivery dates.

5.3.4 Survival ratios

The robustness of the MST over time is examined by computing the single step survival ratio of the links, S_R. This quantity refers to the fraction of edges in the MST, that survives between two consecutive trading days ([23]):

$$S_R(t) = \frac{1}{N-1} \left| E(t) \cap E(t-1) \right| . \tag{5.9}$$

In this equation, $E(t)$ refers to the set of the tree's edges at date t, $\cap$ is the intersection operator, and $| \, . \, |$ gives the number of elements contained in the set. Under normal circumstances, the topology of the trees, between two dates, should be very stable, at least when of the window lengths parameter ΔT presents small values. While some fluctuations of the survival ratios might be due to real changes in the behavior of the system, it is worth noting that others may simply be due to noise. In this study, we mostly examine the presence of trends in the way these ratios evolve.

Figure (5.10)-b represents their evolution in the spatial dimension. Most of the time, this measure remains constant, with a value greater than 0.9. Thus, the topology of the tree, in the spatial dimension, is very stable. The shape of the most efficient path for the transmission of prices shocks does not change much over time. However, it is possible to identify four events where 1/4 of the edges has been shuffled. Such a result also calls for further investigation, as a reorganization of the system can be interpreted as the result of a prices shock.

In the maturity dimension, Figures (5.11)-a and -b exhibit different patterns for crude oil (LLC) and interest rates (IED). As far as crude oil is concerned, while the trees shrink in the metric sense, the organization of the MST is very stable. Few events seem to destabilize the edges of the trees, except for the very end of the period, *i.e.* from the end of 2008. Again, what happens on the eurodollar is totally different. In mid-2001, around the time of the internet crisis, when the length of the tree increases, the tree also becomes more spaced out. This sparseness comes with an important amount of reorganizations, and fluctuations in the survival ratio are greater as the length increases.

A more complete view of what happens in the maturity dimension is offered by Figure (5.12). It exhibits the high level of stability of the trees in the way delivery dates are

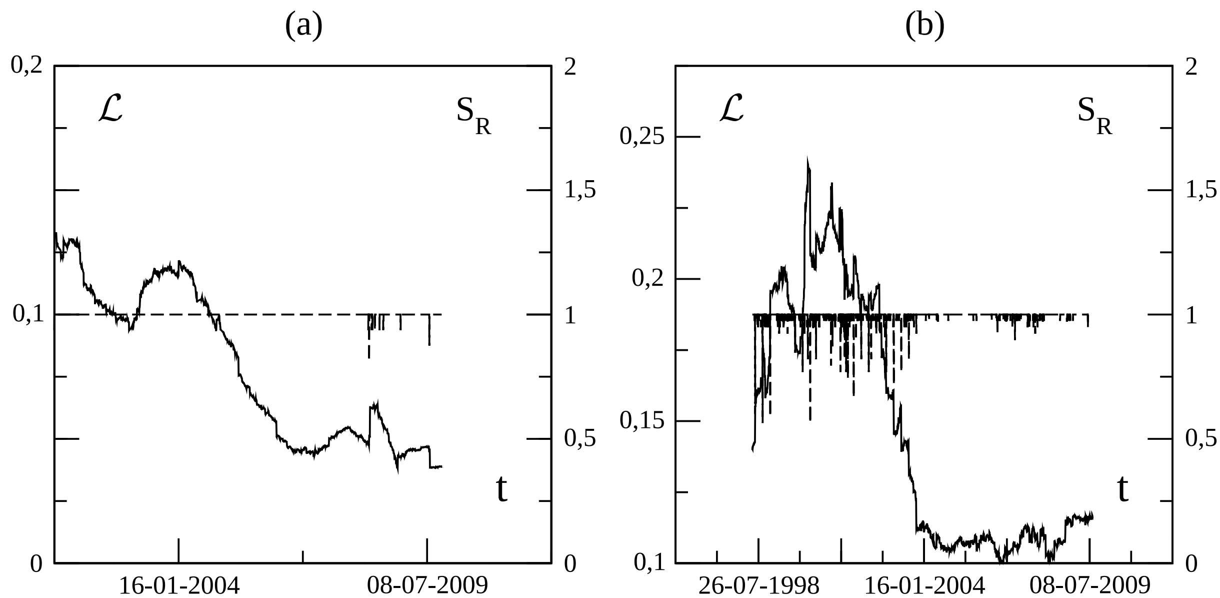

Figure 5.11: Maturity dimension, normalized tree's length and survival ratios for the eurodollar *IED* (a) and the Brent *LLC* (b)

organized.

Lastly, as far as the 3-D trees are concerned, the survival ratios do not give any further information than in the spatial and maturity dimension. However, a more specific analysis of these trees, based on a pruning method, provides some interesting results.

5.3.5 Mean occupation layer

We then present the time evolution of the mean occupation layer. This measure characterizes the topological compactness of the tree which usually shrinks, topologically, during crashes. This effect could be measured as a decrease of $L(t)$. The time evolution for three integration are presented on Figure (5.13). For the spatial integration (dashed line of Figure (5.13)-a, and the spatio-maturity integration (black line of Figure (5.13)-a. As observed by the authors of [29], there are fluctuations but no constant trends for the value. For the maturity integration, represented by Figure (5.13)-b, we can observe that the eurodollar*IED* becomes sparser with a sudden jump, while the Brent crude oil *LLC*

Energy Futures Contracts		
Mnemonic	**Mean MST length**	**Mean Survival Ratio**
NCL	0,0611	0,996
LLC	0,0773	0,999
NNG	0,376	0,995
LNG	0,754	0,999
LLE	0,119	0,999

Financial Futures Contracts		
Mnemonic	**Mean MST length**	**Mean Survival Ratio**
NGC	0,0543	0,997
IED	0,145	0,994

Agricultural Futures Contracts		
Mnemonic	**Mean MST length**	**Mean Survival Ratio**
CBO	0,169	0,996
CC	0,289	0,979
CS	0,270	0,997
CW	0,268	0,999

Figure 5.12: Table summarizing the mean normalized tree's length and the mean survival ratio of the minimum spanning trees in the maturity dimension, for all markets.

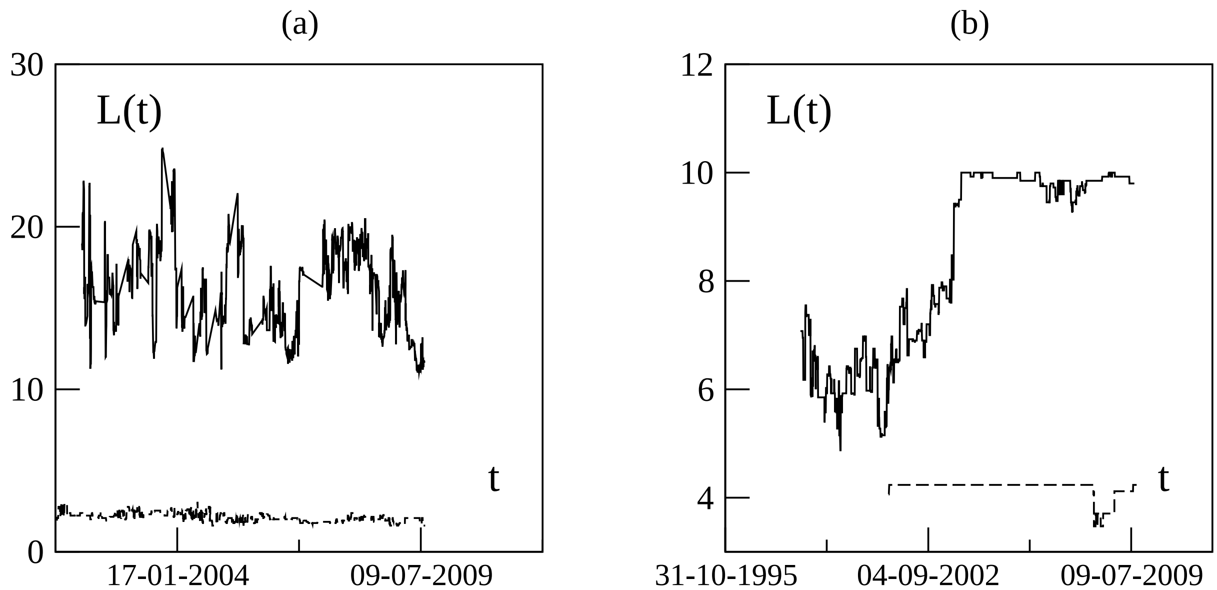

Figure 5.13: Mean occupation layer. Figure (a): Three-dimensional integration (black line) and spatial integration (dashed line); Figure (b): Maturity integration for the eurodollar *IED* (black line) and the Brent crude oil *LLC* (dashed line).

is constant in time with an important fluctuation in the end of the period.

5.3.6 Pruning the trees

As far as the stability of the trees is concerned, especially in 3-D, when focusing on the whole system, it is interesting to distinguish between reorganizations occurring in a specific market, between different delivery dates of the same contract, and reorganization that changes the nature of the links between two markets or even between two sectors. Equation (5.9) however gives the same weight to every kind of reorganization, whatever its nature. The trouble is, a change in intra-maturity links does not have the same meaning, from an economic point of view, as a movement affecting the relationship between two markets or sectors. As we are interested, at least initially, in strong events affecting the markets, inter markets and inter sectors reorganizations seem more relevant. Thus, in order to distinguish between these categories of displacements, we decided to prune the 3-D

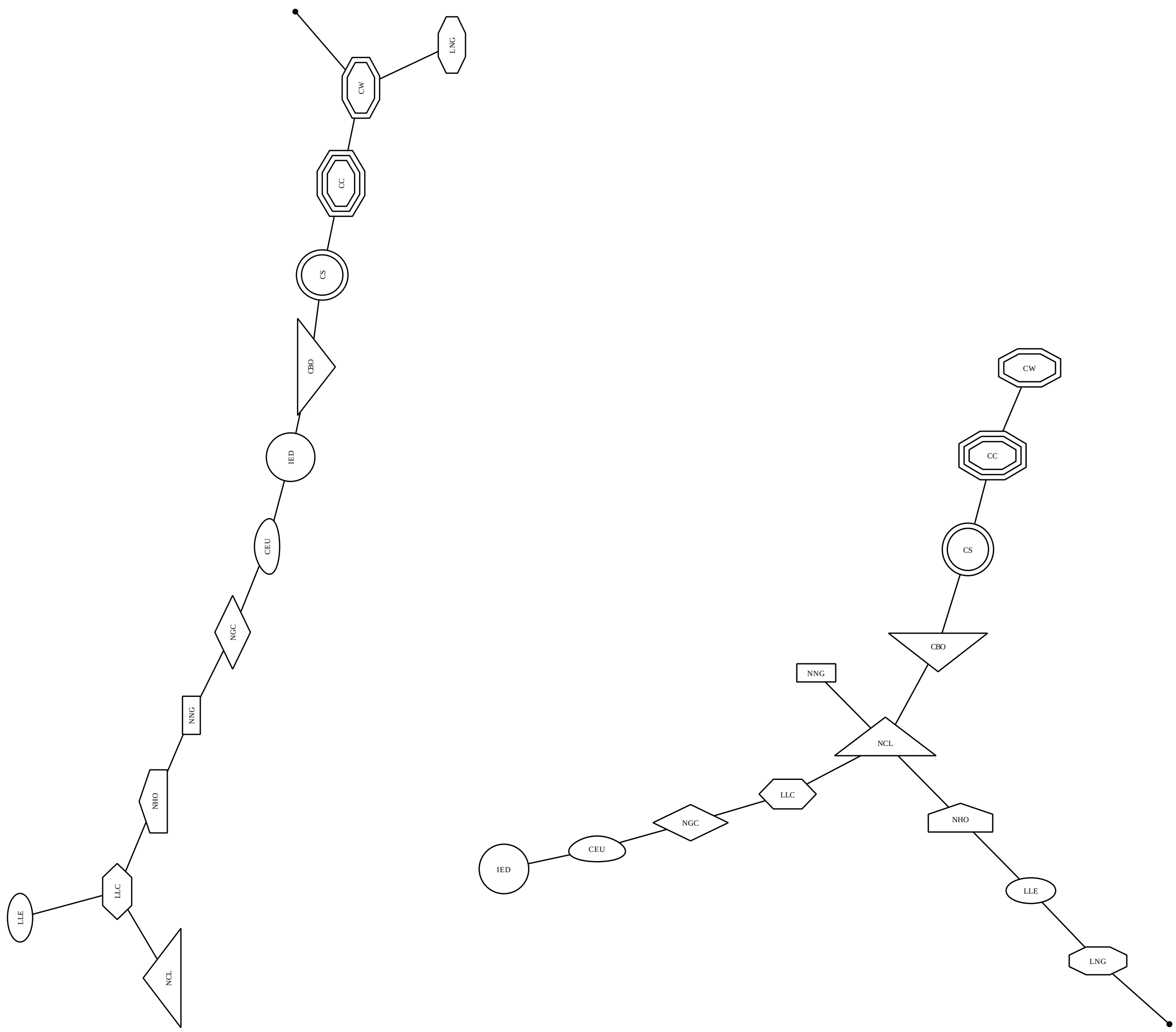

Figure 5.14: Pruned minimum spanning trees of the events 09/02/2004 (left panel) and 16/09/2008 (right panel).

trees, *i.e.* to only consider the links between markets, whatever the maturity considered. This does not mean, however, that maturity is removed from the analysis. It signifies that with pruned trees, the information on the specific maturity that is responsible for the connexion between markets is no longer relevant. Such trees enable us to compute the survival ratios on the sole basis of market links.

Figure (5.15)-a displays the survival ratio of the reduced trees. As observed previously, the ratio is fairly stable. However, several events cause a significant rearrangement of the tree. This is the case, for example, for two specific dates, namely 02/09/04 and 09/16/08. A brief focus on these two dates shows that the tree is totally rearranged. In 2004, the trees become highly linear, the financial assets sector is at the center of the graph, and commodities appear mainly at the periphery of the system. Conversely, in 2008, the tree has a typical star-like shape showing an organization based on the different sectors studied.

Another interesting characteristic of the pruned survival ratios is that they provide information on the length of periods of market stability. Over the entire period of our study, we measured the length of time τ corresponding to a stability period, and we computed the occurrences $N(\tau)$ of such periods. Figure (5.15)-b displays our results. It shows that $N(\tau)$ decreases strongly with τ, with a possible power law behavior, as shown in the log-log scale inset of Figure (5.15)-b. There are few stable periods that last a long time, and much more stable periods that last a short time. We need to refine the former result, but if such a power law is confirmed, it will mean that the markets can have stable periods of any length.

Finally, another interesting result lies in the analysis of those links which are most frequently responsible for the reorganization of the trees. With fourteen markets, there are ninety one links in our system. Some of them - twenty six - never appear. Among the remaining sixty- five trees, some appear very frequently and, on the contrary, others display very few occurrences. Figure (5.16) reproduces these two categories of links and the frequency in which they appear in the MST. The most robust links have a frequency equal to one, which means that the links are always present. They mainly correspond to the agricultural sector, with the following pairs: wheat and corn (CW-CC), soy beans and corn (CS-CC), soy oil and soy beans (CBO-CS). The link between gold and the euro-dollar exchange rate (NGC-CEU) is also always present. As expected, the relationships between the two crude oils (NCL-LLC) are very stable, with a frequency greater

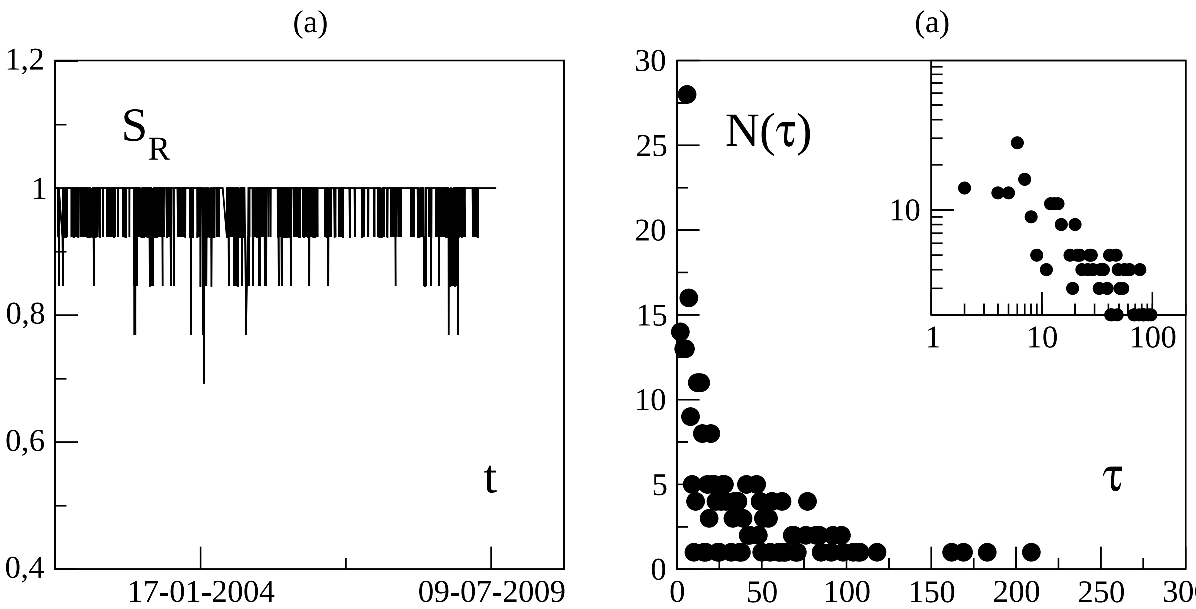

Figure 5.15: Properties of pruned trees. Figure (a): Survival ratio. Figure (b): Number of occurrences of stable periods of length τ. Inset: same as Figure (b), but in log-log scale.

than 0.9. The same is true for the links between the interest rates and the exchange rate (*IED-CEU*), which is also rather intuitive, from an economic point of view, as interest rates are embedded in forward exchange rates. The other tail of the curve contains ten links characterized by a frequency lower than 0.01. The lowest values correspond to the association of interest rates and gas oil and that of interest rates and gold.

5.4 Conclusion

In this chapter, we study the question of systemic risk in energy derivative markets based on two choices. First we focus on market integration, as it can be seen as a necessary condition for the propagation of a prices shock. More specifically, we focus on the simultaneous correlations of price returns. Secondly, based on the fact that previous studies mainly focused solely on the spatio-temporal dimension of integration, we introduce a

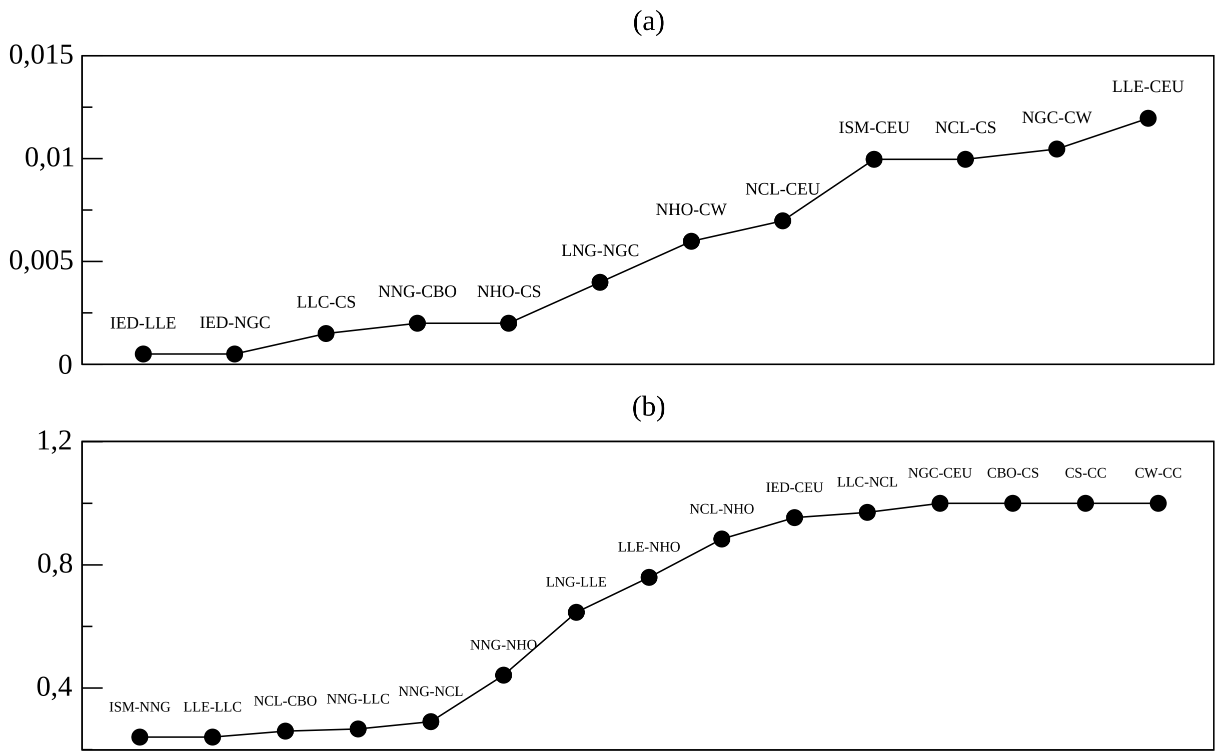

Figure 5.16: Frequency of links apparition in the pruned minimum spanning trees. Figure (a): frequency lower than 0,012. Figure (b): frequency greater than 0,25.

maturity dimension analysis and we perform a three-dimensional analysis.

The visualization of the MST first shows a star-like organization of the trees in the spatial dimension, whereas the maturity dimension is characterized by chain-like trees. These two topologies merge in the three-dimensional analysis, but the star-like organization still dominates. The star-like organization reproduces the three different sectors studied: energy, agriculture and finance, and the chain-like structure reflects the presence of a Samuelson effect. These intuitive results are very important, as they are a key justification for the use of our methodology.

The American and European crude oils are both found at the center of the graph and ensure the links with agricultural products and financial assets. Thus the first conclusion of importance that we come to is that crude oil is the best candidate for the transmission of prices shocks. If such a shock appears at the periphery of the graph, unless it is absorbed quickly, it will necessarily pass through crude oil before spreading to other energy products and sectors. Moreover, a shock will have an impact on the whole system that will be all the greater the closer it is to the heart of the system.

Another important conclusion is that the level of integration is more important in the maturity dimension than in the spatial one. Once again, this result is intuitive: arbitrage operations are far easier with standardized futures contracts written on the same underlying asset than with products of different natures such as corn bushels and interest rates. The analysis of how this level evolves over time shows that integration increases significantly on both the spatial and maturity dimensions. Such an increase can be observed on the whole prices system. It is even more evident in the energy sector (with the exception of the American and European natural gas markets) as well as in the agricultural sector. The latter is highly integrated at the end of our period. Lastly, as far as the financial sector is concerned, no remarkable trend can be highlighted. Thus, as time goes on, the heart of the price system becomes stronger whereas where the peripheral assets are found does not change significantly.

Last but not least, the dynamic analysis also reveals, by using survival ratios, that the system is fairly stable. This is true, except for specific events leading to important reconfigurations of the trees and requiring a specific analysis. We leave these studies for future analyses.

Such results have very important consequences, for regulatory as well as for hedging and diversification purposes. The move towards integration started some time ago and there is probably no way to stop or refrain it. However, knowledge of its characteristics is important, as regulation authorities may act in order to prevent prices shocks from occurring, especially in places where their impact may be important. As far as diversification is concerned, portfolio managers should probably focus on the less stable parts of the graph. The links in the trees which change the most should be the best candidates for diversification opportunities. Lastly, one important concern for hedging is the information conveyed by futures prices and its meaning. The increasing integration of derivative markets is probably not a problem for hedging purposes, unless a prices shock appears somewhere in the system. In such a case, the information related to the transmission path of the shock is important, as prices might temporarily become irrelevant.

6

Building a minimal model

Contents

This chapter is devoted to a theoretical work focusing on the way we could build a minimal model. We first remind of the spirit of a minimal model and finally attempt to describe the behavior of linked commodity derivatives, on the basis of physical tools.

6.1 The spirit if a minimal model

Following a common strategy for physicists, we aim to build a model of commodity prices behavior which gives a simplified representation of their real dynamic. Such a construction relies on a balance between:

- the formulation of equations which are simple enough to be solved, either analytically or at least numerically
- remaining faithful to all the main features of the phenomena we want to study

Of course these two points are the core of any modeling process. As far as we know, our investigations however are pretty original, and physicists who attempt ti build a new model in a new field of investigation tend to make very rough simplifications. Most of the time they propose models with a very small number of parameters. The key of this modeling is to retrieve most of the information through the reduced number of parameters and keeping in mind that some details are lost and not in the scope of the build minimal model. We must emphasize that knowing what kind of information is lost make richer the understanding of the model and is helpful for further investigations and refinements.

6.2 A minimal model for collective behavior and statistical physics toolbox

In the preceding sections , we briefly introduced certain words which sound familiar for physicists. These words are probably less obvious for scientists belonging to other communities. More precisely, we talked about phase transition, order parameter, collective behavior, universality, noise or temperature... All this terminology can be explained with rather simple concepts or pictures. This is the aim of this section. We will explain these words, either formally or, when possible, with simple illustrations. In the next paragraph, all these concepts will be joined in a well defined context, namely a simple model exhibiting collective behavior and phase transition.

Let us start this pedagogical section by introducing what a physicists means with the expression *collective behavior* and what kind of tools or measures exist to detect them.

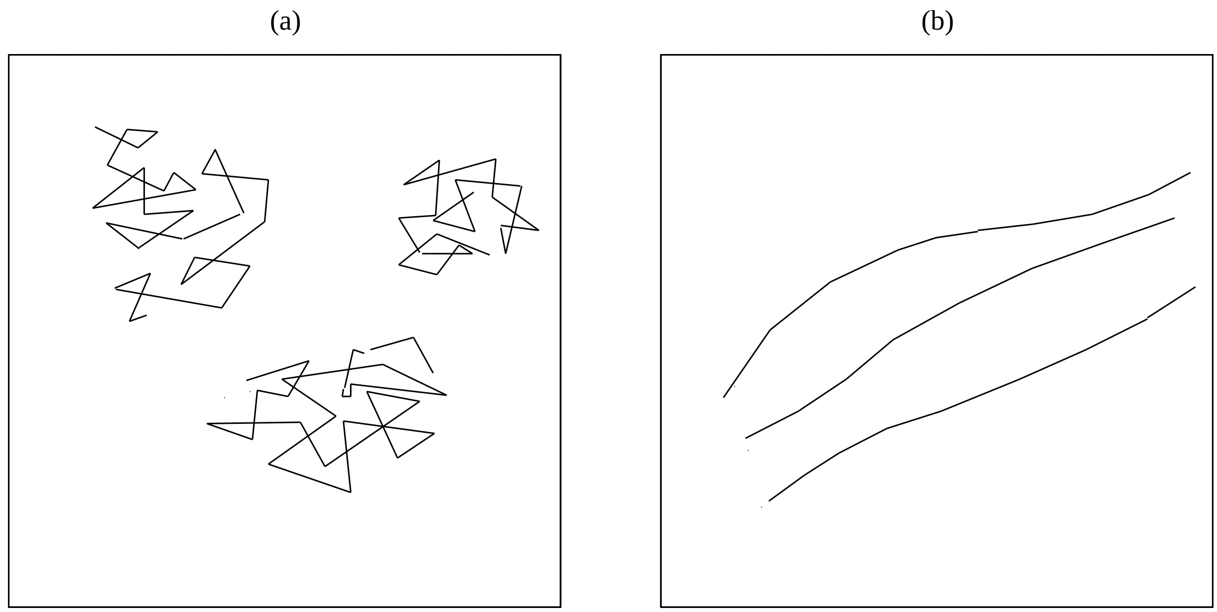

Figure 6.1: Sketch of three trajectories in a swarming crowd (figure (a)) in a moving crowd (figure (b)).

We define a collective behavior as a spontaneous consensus reached by all (or a signifiant part) of an assembly of homogeneous or heterogeneous individuals. For instance, one can imagine a crowd made up of individuals (having different ages, sex, social classes or political opinions) walking around with no specific direction, which suddenly moves in one direction. How can this kind of movement emerge in the absence of a leader and/or an external stimulus? Such a collective motion is all the more surprising if individuals inside the crowd only have access to local information. This a good picture of a collective motion for a physicist. There is a large number of individuals, the available information is local and a global movement is spontaneous. Naturally, a global motion resulting from an external signal is also of interest but the questions addressed will be different. Then we need a value to discriminate if the crowd is moving or not. This is the role of the so-called order parameter. The latter is a value that gives an information about the macroscopic state of the system. For instance if the crowd is moving, the order parameter is equal to one. Otherwise it is equal to zero. All the value between one and zero indicate the degree of order in the crowd. Let us consider the velocity $\vec{u}_i(t)$ of the individual (labelled by i) at the time t. To make it more simple we take the modulus $|u_i(t)| = 1$. The value of the order parameter φ is defined by:

$$\varphi(t) = \langle u_i(t) \rangle_i \, , \tag{6.1}$$

where $<>_i$ denotes the average over all the population. φ belongs to the interval $[0, 1]$. It takes the value 0 if the crowd does not move. conversely, if the assembly is moving in the same direction, it is equal to 1.

Now we have introduced some concepts of phase transitions, we can collect them within a simple model. The so-called Ising model, which describes the spontaneous magnetization in magnetic material.

6.2.1 Ising Model

In this paragraph, we will present with all possible details a simple model describing how a spontaneous magnetic field can appear in a material. A magnetic field results from a collective behavior of microscopic entities. Questions about magnetization is out of the scope of this report but the Ising model presents a great advantage: it provides a good explanation of what is a collective behavior. Moreover, it gives some tool that proved to be useful in the analysis of phase transition.

The ferromagnetism is a very complex phenomenon. Some materials, such as iron, cobalt, nickel, for instance, are naturally magnetized. The existence of magnetization signifies that, at a microscopic scale each electron carries a physical quantity, the *spin*. The latter is more or less the same for all the electrons and create a macroscopic magnetization. At a high temperature, the spin of each electron is different. The magnetization vanishes and the material is said to be paramagnetic. Again, the transition from a ferromagnetic state to a paramagnetic state is a very complex phenomenon and we will see how we can face such a complexity. We first need simple enough equations to be solved analytically or at least numerically. Secondly, we have to conserve the main features we aim to study. While the equations must remain very simple, the phase transition must appear in the development of the model. We will not think any more about a specific material, nor about electrons. We will replace it with a network of N nodes. The latter is characterized by the presence of a spin $\vec{S}_i$ at each node i. Then we consider the interactions between spins. Only spins that are neighbors on the network can interact together. The latter lust be aligned to create magnetization, the simplest interaction describing the alignment is given by the hamitltonian $\mathcal{H}$:

$$\mathcal{H} = -J \sum_{\langle i,j \rangle} \vec{S}_i \vec{S}_j, \tag{6.2}$$

where $J > 0$ is a coupling constant, $\sum_{\langle i,j \rangle}$ is the sum over the nearest neighbors and the modulus $\left| \vec{S}_i \right|$ is set to one 1.
Despite the simplicity of the Ising model, the latter is still too much complicated and cannot be solved. There is a need for an additional approximation in order to obtain a analytical solution. Thus all the vectors $\vec{S}_i$ are replaced by a number S_i, which can take two values: 1 and -1.
Equation (6.2) thus becomes:

$$\mathcal{H} = -J \sum_{\langle i,j \rangle} S_i S_j, \quad S_i = \pm \pm 1 \tag{6.3}$$

and the partition function $\mathcal{Z}$ is :

$$\mathcal{Z} = \sum_{[S_i]} e^{\frac{J}{kT} \sum_{\langle i,j \rangle} S_i S_j}, \tag{6.4}$$

where the sum is over all the possible configurations:

$$\sum_{[S_i]} = \sum_{S_1=\pm 1} \sum_{S_2=\pm 1} \sum_{S_N=\pm 1} . \tag{6.5}$$

Let us consider a possible configuration state α of the system. One can prove that the probability of having the state α is :

$$P_\alpha = \frac{1}{\mathcal{Z}} e^{\frac{1}{kT} E_\alpha}, \tag{6.6}$$

where k is the Boltzmann factor, T the temperature of the system and E_α the energy of the state α. The partition function is a normalization constant which ensures the sum of probabilities to be equal to one. The partition function contains also how the probabilities are distributed between all the configurations.

In a one dimensional space, the N spins are on a line, the partition is $\mathcal{Z} = 2^N \left(\cosh{(K)}^{N-1}\right)$. The partition function covers the statistical properties of the system and allows for checking wether a phase transition can occur. Here the transition corresponds to a transition from a zero magnetization state to a non-zero magnetization state.

Once $\mathcal{Z}$ has been calculated, we have access to the two point correlation function $\langle S_i S_j \rangle$. The correlation function gives the possibility to appreciate the influence of the value of one spin S_i on the other spins S_j. The influence of S_i on S_j depends on the distance r_{ij} between i and j. The value $\langle S_i S_j \rangle$ is high when r_{ij} is small and decreases as r_{ij} increases. The correlation function is given by:

$$\langle S_i S_j \rangle = e^{-r_{ij}/\xi}, \tag{6.7}$$

where $\xi = a/\left|ln\left(\tanh J/kT\right)\right|$ is the correlation length, that gives the typical distance above which the information is lost.

In this paragraph, we have presented the Ising model, a minimal model describing magnetization in ferromagnetic material. More precisely, we have seen how to model the interaction between spins. We have also extracted the partition function, that gives access to all the available information. For instance, the correlation function which gives the scale at which a spin influences its neighbors. One can easily notice that a large value of ξ implies large correlations and thus is the signal of a collective behavior.

We have exposed the model in a one dimensional space. However, for practical reasons there is a need for higher dimensions. Unfortunately, despite all the simplifications made to obtain the Ising model, the latter remains not trivial. Ernst Ising solved the problem in 1925. It was not until 1936 that Rudolf Peierls proved that the two-dimensional Ising model has a phase transition. The critical temperature for the phase transition has been obtained by Kramers and Wannier in 1941 and finally a general solution for the two-dimensional case has been found by Onsager in 1944. Eighty years later, no one has found an analytical solution in three dimensions. To deal with a high dimensional system, or a more complicated model, additional approximations are necessary. A very important and useful method is the so called mean field approximation.

6.2.2 Mean field approximation

Weiss proposed the mean field method in 1907. The latter consists in replacing the influence of the neighbors by their average impact. Let us consider one spin S_i. If we want to calculate its energy, it is possible to approximate the effect of the other spins S_j by introducing their average $\langle S_j \rangle$. The problem becomes a classical paramagnetism calculation and the interaction is given by:

$$\mathcal{H} = -J \sum_{\langle i,j \rangle} S_i S_j - \mu B \sum_i S_i, \tag{6.8}$$

where B is an external magnetic field and μ the magnetic moment.
The energy for one spin S_i is:

$$E_i = -JS_i \sum_j \langle S_j \rangle - \mu B S_i. \tag{6.9}$$

In equations (6.8) and (6.9), there a supplementary term $\mu B \sum_i S_i$ that reflects an exogenous effect and could be later set at 0 later. The mean field approximation is obvious in (6.9) where the sum over the neighbors has been transformed into a sum over the average value $\langle S_j \rangle$. Within the mean field approximation, the magnetization M is:

$$\tanh^{-1} M = \frac{qJ}{kT} M + \frac{\mu B}{kT}, \tag{6.10}$$

where q is the number of neighbors, which depends of the dimension oh the space. When B tends towards zero, the mean field approximation predicts a spontaneous magnetization:

- $M = 0$ if $T > qJ/k$
- $M \neq 0$ if $T < qJ/k$

As the parameter T is tuned, the magnetization M changes from zero to a non-zero value. Thus T is thus called the *control parameter* because a variation in T changes the macroscopic property of the system. The value M, which indicates if there is magnetization, is the *order parameter*.
The mean field approach predicts a spontaneous magnetization as soon as the temperature T is below qJ/k. This magnetization is the sign of a collective behavior. Indeed, at a microscopic scale, due to the local interactions described by (6.3), the spins take collectively the same value and create a macroscopic magnetization. There is a special value of T. This value is named the critical temperature, T_c, and is equal to qJ/k. The mean field theory provides important results for a temperature T close to T_c.

In particular, the magnetization M behaves as:

$$M \sim (T_c - T)^{\frac{1}{2}}\,. \tag{6.11}$$

The magnetization response of a material to an exogenous magnetic field is called the susceptibility and is defined as:

$$\chi = \left.\frac{\partial M}{\partial B}\right|_{B=0}. \tag{6.12}$$

On both sides of T_c, the susceptibility χ behaves like:

$$\chi \sim (T_c - T)^{-1}\,, \tag{6.13}$$

with a different coefficient according to the sign of $(T_c - T)$. The behavior of the magnetization M in function the external applied magnetic field B is:

$$B \sim M^3. \tag{6.14}$$

The equations (6.11), (6.13), (6.14), give the behavior of some physical quantities close to the critical temperature T_c. These behaviors are described power laws. A power law function $f(x) = x^\alpha$ is invariant under the rescaling $x \to \lambda x$ because:

$$f(x) = \lambda^{-\alpha} f(\lambda x). \tag{6.15}$$

Thus, the observed phenomena close to the critical temperature T_c are scale invariant and remain identical whatever the size at which the system is observed. The exponents of the power law functions do not depend on details, it is the reason why details are removed to reach a minimal model, and are named universal exponents.

6.3 Some conclusions and general ideas about the Ising model and the mean field approximation

As a conclusion, let us briefly summarize the two preceding paragraphs in order to insist, first on the philosophy underlying the construction of a minimal model, second on the advantages and drawbacks of such a model.

A minimal model focuses on some main characteristic(s) and do not pay intention to the details. For instance, the Ising model predicts that a transition will occur at a given temperature T_c. However T_c is not universal while the transition is. We cannot know which spin is in charge when the magnetization appears, but the behaviors of the correlation length or some physical quantities described by power laws are the same for all materials. Despite the simplicity of minimal models, most of the time they have analytical solutions and one need numerical investigations or approximations like the mean field theory. It is a very useful method, still used and in general is a first step to study a new model, but must be considered carefully. The reason is that fluctuations are neglected and if the phenomenology is driven by fluctuations the mean field approach will not capture it.

7

Conclusion

In this report we have presented our first results about the study of systemic risk in energy derivatives markets. In the financial literature, the studies of the way shocks appear in financial markets and the way they disseminate among other markets generally take into account one or two of the dimensions of the integration. While the integration can be examined according three dimensions, space, observation time and maturity, we have been naturally lead to the following question: “Why not trying to study the three dimensions simultaneously?”. On that purpose, we apply recent methods from statistical physics.

We first present the energy markets selected for our empirical study: three crude oil, two heating oil and natural gas markets. We then expose the main characteristics of the database and present some futures prices behavior on our study period. We finally discuss the seasonal behavior of the commodities under consideration.

We then propose a method allowing us to empirically measure the integration of the markets. This method is a filtering procedure which transforms a correlation matrix into a distance matrix in order to compute a particular graph, the minimum spanning tree. The latter provides the shortest path linking together all the nodes of the graph.

We apply this method to our data and realize two series of empirical tests, on the maturity as well as on the spatial dimensions. These results reflect the presence of the

Samuelson effect associated with a linear organization of the graph, from the first to the last maturities. We performe an analysis of the spatial integration for five markets on different maturities. It happens that the topology of the resulting graphs changes with the maturity under consideration. In each case, the links between markets, through the representation of the minimum spanning tree, have an economical interpretation that satisfies the intuition. Comparing the results for different maturities, we have found that the strength of the integration increases with the maturity. The latter result is original and has not been yet mentioned in other works.

We then study the question of systemic risk in energy derivative markets based on two choices. First we focus on market integration, as it can be seen as a necessary condition for the propagation of a prices shock. More specifically, we focus on the simultaneous correlations of price returns. Secondly, based on the fact that previous studies mainly focused solely on the spatio-temporal dimension of integration, we introduce a maturity dimension analysis and we perform a three-dimensional analysis.

The visualization of the MST first shows a star-like organization of the trees in the spatial dimension, whereas the maturity dimension is characterized by chain-like trees. These two topologies merge in the three-dimensional analysis, but the star-like organization still dominates. The star-like organization reproduces the three different sectors studied: energy, agriculture and finance, and the chain-like structure reflects the presence of a Samuelson effect. These intuitive results are very important, as they are a key justification for the use of our methodology.

The American and European crude oils are both found at the center of the graph and ensure the links with agricultural products and financial assets. Thus the first conclusion of importance that we come to is that crude oil is the best candidate for the transmission of prices shocks. If such a shock appears at the periphery of the graph, unless it is absorbed quickly, it will necessarily pass through crude oil before spreading to other energy products and sectors. Moreover, a shock will have an impact on the whole system that will be all the greater the closer it is to the heart of the system.

Another important conclusion is that the level of integration is more important in the maturity dimension than in the spatial one. Once again, this result is intuitive: arbitrage operations are far easier with standardized futures contracts written on the same underlying asset than with products of different natures such as corn bushels and interest rates. The analysis of how this level evolves over time shows that integration increases significantly on both the spatial and maturity dimensions. Such an increase can be observed on the whole prices system. It is even more evident in the energy sector (with the exception of the American and European natural gas markets) as well as in the agricultural sector. The latter is highly integrated at the end of our period. Lastly, as far as the financial sector is concerned, no remarkable trend can be highlighted. Thus, as time goes on, the heart of the price system becomes stronger whereas where the peripheral assets are found does not change significantly.

Last but not least, the dynamic analysis also reveals, by using survival ratios, that the system is fairly stable. This is true, except for specific events leading to important reconfigurations of the trees and requiring a specific analysis. We leave these studies for future analyses.

Such results have very important consequences, for regulatory as well as for hedging and diversification purposes. The move towards integration started some time ago and there is probably no way to stop or refrain it. However, knowledge of its characteristics is important, as regulation authorities may act in order to prevent prices shocks from occurring, especially in places where their impact may be important. As far as diversification is concerned, portfolio managers should probably focus on the less stable parts of the graph. The links in the trees which change the most should be the best candidates for diversification opportunities. Lastly, one important concern for hedging is the information conveyed by futures prices and its meaning. The increasing integration of derivative markets is probably not a problem for hedging purposes, unless a prices shock appears somewhere in the system. In such a case, the information related to the transmission path of the shock is important, as prices might temporarily become irrelevant.

The last part of the report is devoted to a theoretical work focusing on the way a minimal model could be built. We finally intend to carry on our investigations in the following directions:

We will first expand our empirical analysis in the maturity dimension. We observed some regular and recurent correlation patterns in the maturity dimension that need deeper investigations and might reflect some universal mechanisisms of price's curve segmentation. The latter result would rise the interest of both communities, finance and physics, whilst up to now the litterature mainly missed this important feature.

We also aim to enrich our results with an analysis of the transaction volumes and the open interests. First we could use the same graph theory formalism in order to analyse trees of correlated transactions and open interests. We could then try to consider returns fluctuations weighted by volumes and/or open interests. Thus, such questions as the robustness of the centrality of crude oil with respect to interest rates will be adressed.

Another field of empirical investigation will be the study of shocks affecting the markets. In particular we could determine the topological properties of trees during strong events, as the nature of the affected links or the time required to go back to initial configuration.

The main part of our further studies will be devoted to modeling the collective behavior of derivatives energy markets and systemic risk. We aim to use theoritical concepts inspired by statistical physics, espacially the use of minimum model. Our former results will lead us to establish fundamental hypothesis and play the role of guideline in the development of the model. In particular we want to determine in a single framework the mechanisms of price's term structure (which lead to linear tree), as the interactions between markets (which lead to star-like tree). Once the two typical shapes will be achieved, we will be able to use the model in order to understand the complex process of branching that appeared while the three dimensions of integration, namely where in their price's curves two different derivatives markets are most correlated. A major contribution of this part of the modelling will be to understand how (and where) links appear between markets. Secondly, we will proceed to a shock analysis and consider such questions as the existence of tree's shape that help or prevent to strong shocks, the required number of markets involved in an event to propagate it, or the amplitude of shocks that can involved in systemic risk.

Bibliography

[1] BANAVAR, J., MARITAN, A., AND RINALDO, A. *Nature 399* (1999), 130.

[2] BONNANO, G., CALDARELLI, G., LILLO, F., MICCICHÈ, S., VANDEWALLE, N., AND MANTEGNA, R. Networks of equities in financial markets. *Eur. Phys. J. B 38* (2004).

[3] BOUCHAUD, J., GEFEN, Y., POTTERS, M., AND WYART, M. Fluctuations and response in financial markets: the subtle nature or random price changes. *Quantitative Finance 4* (2004).

[4] BUYUKSAHIN, B., HAIGH, M., AND ROBE, M. Commodities and equities: a market of one? *Working paper* (2008).

[5] CHATÉ, H., GINELLI, F., GRÉGOIRE, G., AND RAYNAUD, F. Collective motion of self-propelled particles interacting without cohesion. *Phys. Rev. E 77* (2008).

[6] CHATÉ, H., GINELLI, F., AND MONTAGNE, R. Minimal model for active nematics:quasi-long-range-order and giant fluctuations. *Phys. Rev. Lett 96* (2006).

[7] DUBUS, C., AND FOURNIER, J. A gaussian model for the membrane of red blood cells with cytoskeletal defects. *Europhys. Lett 75* (2006).

[8] Edwards, F., and Canters, M. The collapse of metallgesellschaft : unhedgeable risks, poor hedging strategy, or just bad luck? *The journal of futures markets 15(3)* (1995), 211–264.

[9] Favier, C. Percolation model of fire dynamic. *Physics Letters A 330* (2004).

[10] Fournier, J., and Barbetta, C. Direct calculation from the stress tensor of the lateral surface tension of fluctuating fluid membranes. *Phys. Rev. Lett. 100* (2008).

[11] Gower, J. Some distance properties of latent roor and vector methods used in multivariate analysis. *Biometrika 53*, 3/4 (1966).

[12] Grégoire, G., Chaté, H., and Tu, Y. Moving and staying together without a leader. *Physica D 181* (2004).

[13] Ide, K., and Sornette, D. Oscillatory finite-time singularities in finance, population and rupture. *Physica A 63* (2002).

[14] Jumah, A., and Karbuz, S. Interest rate differentials, market integration, and the efficiency of commodity futures markets. *Applied Financial Economics 9* (1999).

[15] Kullmann, L., Kertész, J., and Kaski, K. Time-dependent cross-correlations between different stock returns: A directed network of influence. *Phys. Rev. E 66*, 2 (Aug 2002), 026125.

[16] Lautier, D. *La structure par terme des prix des commodités: analyse théorique et applications au marché pétrolier*. PhD thesis, Université Paris Dauphine, 2000.

[17] Lautier, D. Segmentation in the crude oil term structure. *Quarterly Journal of Finance IX*, 4 (2005), 1003–1020.

[18] Mantegna, R. Hierarchical structure in financial markets. *Eur. Phys. J. B 11* (1999).

[19] Marsili, M. *Quantitative Finance 2* (2002).

[20] Miceli, M., and Susinno, G. *Risk 16* (2003).

[21] Modigliani, F., and Stutch, R. Innovation in interest rate policy. *American Economic Review 56* (1966), 178–197.

[22] Onnela, J.-P., Chakraborti, A., Kaski, K., Kertész, J., and Kanto, A. Dynamics of market correlations: Taxonomy and portfolio analysis. *Phys. Rev. E 68*, 5 (Nov 2003), 056110.

[23] Onnela, J.-P., Chakraborti, A., Kaski, K., and KertÃ©sz, J. Dynamic asset trees and black monday. *Physica A 324* (2003).

[24] Pilipovic, D. *Energy risk : valuing and managing energy derivatives*. Mac Graw, 2007.

[25] Pindyck, R., and Rotenberg, J. The excess co-movement of commodity prices. *Economic Journal 100* (1990).

[26] Raynaud, F. *Modèles de comportements collectifs tri-dimensionnels*. PhD thesis, Université Denis Diderot, 2009.

[27] R.C, P. *Bell System Technical Journal 36* (1957).

[28] Samuelson, P. Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review 6* (1965).

[29] Sieczka, P., and Holyst, J. A. Correlations in commodity markets. *Physica A 388* (2009).

[30] Stewart, M., and Wallis, K. *Introductory Econometrics*. Basil Blackwell, second edition, 1990.

[31] VANDEWALLE, N., BRISBOIS, F., AND TORDOIR, X. Non-random topology of stock markets. *Quantitative Finance 1* (2001).