



HAL
open science

An empirical analysis of systemic risk in commodity futures markets

Julien Ling

► **To cite this version:**

Julien Ling. An empirical analysis of systemic risk in commodity futures markets. Business administration. Université Paris sciences et lettres, 2018. English. NNT : 2018PSLED022 . tel-01916619

HAL Id: tel-01916619

<https://theses.hal.science/tel-01916619>

Submitted on 8 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à PSL Université Paris-Dauphine

An empirical analysis of systemic risk in commodity futures markets

École doctorale n°543

ECOLE DOCTORALE DE DAUPHINE

Spécialité SCIENCES DE GESTION

Soutenue par Julien LING
le 26 septembre 2018

Dirigée par **Delphine LAUTIER**

COMPOSITION DU JURY :

Pr. Frédéric ABERGEL
CentraleSupélec, Membre du jury

Pr. René AÏD
PSL Université Paris-Dauphine
Président du jury

Pr. Olivier BRANDOUY
Université de Bordeaux, Rapporteur

Pr. Delphine LAUTIER
PSL Université Paris-Dauphine
Directrice de thèse

Dr. Yannick LE PEN
PSL Université Paris-Dauphine
Membre du jury

Pr. Benoît SÉVI
Université de Nantes, Rapporteur

An empirical analysis of systemic risk in commodity futures markets

DOCTORAL DISSERTATION

submitted by

Julien LING

May 22, 2018

Thesis supervisor : **Delphine LAUTIER**

Thesis jury

Frédéric ABERGEL	Professor	Member of the jury
René AID	Professor	President of the jury
Olivier BRANDOUY	Professor	Examiner
Delphine LAUTIER	Professor	Thesis supervisor
Yannick LE PEN	Associate Professor	Member of the jury
Benoît SEVI	Professor	Examiner

L'Université Paris-Dauphine n'entend donner aucune approbation, ni improbation aux opinions émises dans cette thèse ; elles doivent être considérées comme propres à leurs auteurs.

Remerciements

Ces quelques années de thèse ont indéniablement été les plus riches de ma vie. Elles ont été remplies de nouvelles expériences et découvertes. Qu'elles aient été bonnes ou mauvaises, elles m'ont fait grandir, m'épanouir et devenir la personne que je suis maintenant. Je me rends compte du long chemin parcouru pour en arriver ici et je ne saurais remercier assez ma directrice de thèse, Delphine Lautier, sans qui je n'aurais pas pu évoluer autant.

Notre rencontre, lors de mon année de Master 1 Gestion – Finance à Dauphine, ne nous destinait pas vraiment à faire ce chemin ensemble, n'ayant pas toujours été le plus sérieux des élèves dans son cours. Nous retrouver en Master 2 Recherche ne présageait pas non plus d'une collaboration aussi approfondie, n'ayant suivi son cours qu'en auditeur libre. Heureusement, Delphine a su me "voler" à un autre professeur au moment où je devais choisir le sujet de mon mémoire de recherche ; c'est à ce moment que toute cette aventure a vraiment commencé. Elle m'a rapidement plongée dans son univers et fait vivre très tôt les expériences qui font un chercheur : les présentations en conférences, l'écriture d'articles de recherche et l'organisation de conférence.

Delphine a aussi été d'un soutien et d'une patience sans faille, aussi bien dans les périodes les plus difficiles que dans les meilleures. Travailler ensemble a été un véritable plaisir, bien que nous ayons eu quelques petits différends sur mon travail. Je ne regrette absolument pas de m'être fait réprimander, car ses conseils m'ont permis de grandement m'améliorer et de combattre ma timidité. Encore une fois – et ce ne sera pas la dernière, merci.

Je remercie également Olivier Brandouy de l'Université de Bordeaux et

Benoît Sévi de l'Université de Nantes d'avoir accepté de rapporter cette thèse qui, je l'espère, leur plaira. Mes remerciements vont aussi à Frédéric Abergel de CentraleSupélec pour avoir accepté de participer au jury de ma thèse. L'enthousiasme de René Aïd lors de nos discussions, notamment pendant la pré-soutenance, et son aide précieuse auront largement contribué à ces quelques années. Je remercie chaleureusement René. Je suis également reconnaissant envers Yannick Le Pen, qui m'a prodigué de nombreux conseils pour améliorer mon travail et qui a aussi accepté de faire partie de ce jury.

Je tiens aussi à remercier les professeurs, enseignants, chercheurs et l'équipe de DRM Finance : ceux dont j'ai pu suivre les cours et qui m'ont permis d'apprendre et d'élargir mes champs d'intérêt dans la recherche ; ceux qui m'ont guidé depuis que j'ai décidé de faire de la recherche, je pense particulièrement à Gaëlle Le Fol, Fabrice Riva, Edith Ginglinger, Carole Gresse, Yves Simon, Milo Bianchi et bien sûr Delphine ; ceux qui m'ont permis de débiter cette thèse grâce au contrat doctoral ; ceux qui m'ont accordé leur confiance en me permettant d'enseigner une partie de leurs cours ; et ceux qui ont contribué à ma réflexion, au bon fonctionnement et à la vie du laboratoire et à cet environnement stimulant de recherche. Je pense aussi à l'équipe de Tunis, qui a fait de mon court séjour là-bas une expérience très plaisante.

Je souhaite particulièrement exprimer toute ma gratitude, mon respect et mon amitié à Edouard Jaeck, mon "binôme de thèse", avec qui j'ai pu partager nombre de ces expériences durant ces années, mais aussi Mario Milone, Yasmine Skalli et Marion Sierra. Merci aussi à tous les doctorants et anciens doctorants du laboratoire pour les discussions animées au CROUS et pour la bonne ambiance dans la salle et leur souhaite bon courage pour la suite. J'aimerais tout particulièrement remercier le Laboratoire de Finance des Marchés de l'Energie (FiME) et l'initiative de recherche Modélisation de l'Intégration des Marchés des Oléagineux (MIMO) sans qui toute cette aventure n'aurait pas été possible. Leurs membres ont été très présents au cours de ma thèse à tous les plans, particulièrement René, Damien, Clémence,

Bertrand, Ivar, Marc, Emmanuel et Imen. La recherche qui s'y fait, ainsi que les événements organisés, dont une école d'été à Florence et les Journées ateliers FiME à l'Abbaye des Vaux de Cernay, ont aussi largement stimulé ma réflexion. Je leur en serai éternellement reconnaissant.

Je remercie aussi les membres des autres laboratoires que j'ai pu côtoyer et qui m'ont donc aussi beaucoup apporté, notamment ceux d'économie, de mathématiques et de sciences de gestion de Dauphine.

Mes séjours à l'étranger ont aussi été l'occasion d'approcher des membres de laboratoires différents mais aussi des mondes différents. Je remercie donc les équipes du Pacific Institute for the Mathematical Sciences de la University of British Columbia à Vancouver, qui m'ont permis d'assister à leur école d'été et workshop sur le risque systémique. L'organisation et l'encadrement ont été exemplaires et la possibilité de discuter longuement de recherche avec des professeurs de grandes universités américaines a grandement contribué à ma réflexion. Vivre pour la première fois sur un magnifique campus nord-américain, découvrir cette ville chaleureuse et y faire des rencontres intéressantes m'a fait grandir à plusieurs niveaux. Je pense notamment à Charles et Chaojun.

Je remercie sincèrement aussi les équipes de l'Institute for Pure and Applied Mathematics de la University of California, Los Angeles, qui m'ont offert la chance de participer à un de leurs beaux "long program". Leurs compétences, leur accueil chaleureux et leur sympathie ont fait de ce programme un événement d'exception, regroupant des chercheurs de divers horizons sur une longue période, animée par de très bons moments de détente et de partage. J'aimerais en particulier remercier Roland, toujours prêt à aider, et sans qui je n'aurais pas eu la chance d'assister à un match de NBA en loge VIP, ni de séjourner dans une magnifique maison près de l'océan (merci Diane pour nous avoir loué cette maison!). J'ai aussi pu faire la rencontre de personnes très sympathiques, avec qui j'ai partagé de très bons moments durant ce séjour, Pierre et Ioane particulièrement.

Mon passage au département Operations Research and Financial Engineering de la Princeton University a également été l'occasion de rencontres

intéressantes, autant sur le plan personnel que professionnel. Je tiens à remercier profondément l'équipe du département pour avoir fait de ce séjour un véritable plaisir, sans problème aucun. Merci aussi à Sheldon et Tatiana pour m'avoir accueilli chez eux, les rencontrer a été particulièrement surprenant et intéressant, ne serait-ce qu'en ayant partagé avec moi leurs expériences et histoires. Mes remerciements vont aussi à Brandon et Rikki, avec qui j'ai partagé de nombreux bons moments et expériences et qui m'auront beaucoup appris sur moi-même.

Je n'oublierai jamais tous ces moments.

Je remercie du fond du cœur tous mes amis, qui ont su me motiver et me remonter le moral dans les moments difficiles, qui m'ont montré l'exemple et avec qui j'ai pu partager toutes mes pensées, tous ces moments heureux et toutes ces expériences. Tout particulièrement Ziyed, Théo, Patrick, François, Yixin et Marc, merci infiniment.

Enfin, je ne peux remercier assez ma famille, qui a toujours été là pour moi, qui m'a compris et soutenu jusqu'au bout et sans qui je n'aurais pas pu aller aussi loin. J'aimerais en particulier remercier du fond du cœur ma belle-mère, Isabelle, qui nous a portés vers nos objectifs. Je n'oublie pas tous mes cousins, oncles, tantes, grands-parents, frères et sœur, qui ont tant participé à la personne que je suis, qui m'ont tant donné et avec qui j'ai tant partagé. Je ne suis pas la plus expressive des personnes, mais sachez que je n'en pense pas moins et que je vous aime. Je clos ces remerciements en pensant à mes parents, qui ne sont malheureusement plus là pour voir cet aboutissement, mais qui ont toujours voulu le meilleur pour moi, qui ont fait tout leur possible pour que j'aille toujours plus loin et que j'espère rendre fiers. Je leur dédie donc cette thèse.

Table des matières

Remerciements	i
Table des figures	vii
Liste des tableaux	ix
Introduction	1
0.1 Les sources de risque systémique	2
0.2 La financiarisation des matières premières, nouvelle source de risque systémique?	12
Références	24
1 Are commodity markets still different from other asset classes?	29
Introduction	31
1.1 Related literature	37
1.2 Data description	41
1.3 Comparing the different sectors in terms of unconditional moments and shocks	43
1.4 Incorporating the information provided by the basis	53
1.5 The potential for diversification and systematic risk	63
Conclusion	69
References	71
Appendices	73
1.A Statistical measures	73

1.B	Conditioning returns on the sign of the basis	75
1.C	Correlation matrices by subsample	88
1.D	Additional results on PCA	91
2	Integration of Commodity Derivative Markets: Has It Gone Too Far?	93
	Introduction	95
2.1	The price system	97
2.2	The efficiency of the shock transmission	101
2.3	The organization of the tree	108
2.4	Examining the centrality of the graphs	112
	Conclusion	118
	References	119
	Appendices	121
2.A	Timelines around the events	121
2.B	Additional results	124
3	Disentangling systemic risk in financialized commodity markets	131
	Introduction	133
3.1	Data	138
3.2	Methodology	139
3.3	The reference graphs (static, full sample)	144
3.4	Shocks and propagation (dynamic, rolling window)	155
	Conclusion	166
	References	167
	Appendices	179
3.A	Correlations vs. partial correlations	179
3.B	Explanation of LASSO penalization	182
3.C	Measures in the dynamic analysis over the whole period	185
	Conclusion et développements futurs	189

Table des figures

0.1	Schéma synthétique des interactions entre le système financier et l'économie réelle	12
1.1	Time series of the spot prices/rates for all our markets over 2000-2014.	32
1.2	Examples of time series of spot returns	33
1.3	Heatmap of pairwise Kolmogorov-Smirnov two-sample test p-values	48
1.4	Heatmap of pairwise Hellinger distances between Gaussian distributions	49
1.5	Examples of time series of spot returns	55
1.6	Examples of conditional return distributions	56
1.7	Heatmap of the unconditional correlation matrix over the whole sample	64
1.8	Results of a PCA in the spatial dimension	68
1.9	Time series of spot returns	76
1.10	Conditional spot return distributions	80
1.11	Heatmap of the unconditional correlation matrix per subsample	88
1.12	Weights of the first 4 principal components	91
2.1	Scaled MST in the spatial dimension, 2000-2009.	103
2.2	Scaled MST in 3D, 2000-2009.	105
2.3	Normalized tree's length in the spatial dimension, 2000-2009.	105
2.4	Normalized tree's length in the spatial dimension and in 3D for each event.	107
2.5	Scaled MST in the maturity dimension, Eurodollar market.	109

2.6	Survival ratios in the spatial dimension, 2000-2009.	110
2.7	Extreme types of structures	111
2.8	Allometric coefficients, in the spatial dimension and in 3D, for each event.	113
2.9	Scaled MST in the spatial dimension at Lehman Brothers bankruptcy (September 15, 2008).	114
2.10	Connectivity (A, B, or C) versus centrality (S).	115
2.11	Normalized length in the spatial dimension and in 3D, around the events	124
2.12	Survival ratios in the spatial dimension and in 3D, around the events	125
2.13	Allometric coefficients in the spatial dimension and in 3D, around the events	125
2.14	Evolution of the ranks in the spatial dimension around the Subprime crisis.	127
2.15	Evolution of the ranks in the spatial dimension, around the Lehman Brothers bankruptcy.	129
3.1	Granger causality directed graph from VAR(1) on metals spot markets	140
3.2	Partial correlation matrix graph in the spatial dimension	146
3.3	Partial correlation graph in 3D	149
3.4	Autoregressive matrix graph (Granger causality graph) in spatial dimension	152
3.5	Autoregressive matrix graph (Granger causality graph) in 3D . .	154
3.6	Granger causality (above) and partial correlation (below) graphs on the day of default of Lehman Brothers	158
3.7	Granger causality (above) and partial correlation (below) graphs on the day of lowest return of the S&P500	161
3.8	Granger causality (above) and partial correlation (below) graphs on the day of the Flash Crash	165
3.9	Geometry point of view, difference between LASSO (left) and Ridge (right) prenalizations	184

3.10	Measures related to the links of the propagation (Granger causality)	186
3.11	Measures related to the integration of markets (PC)	187
3.12	Measures related to the structure of the partial correlation graphs (3)	188

Liste des tableaux

1.1	Futures contracts selected, data information	42
1.2	Moments of the spot return distributions	45
1.3	Spot returns distribution quantiles	51
1.4	Moments of the empirical returns distributions conditional on the sign of the 3m-basis	57
1.5	Probabilities of spot returns shocks (at 1% and 5%), conditional on contango or backwardation	62
1.6	p-values for the KS tests of pairwise distribution differences	73
1.7	Moments of the empirical returns distributions conditional on the Last-basis sign	85
1.8	Probabilities (in %) of shocks (negative: '1%' and '5%'), conditional on 3m-contango and 3m-backwardation	86
1.9	Probabilities (in %) of shocks (negative: '1%' and '5%'), conditional on Last-contango and Last-backwardation	87
2.1	Characteristics of the collected data	99
2.2	Bonacich's centrality measure in the spatial dimension, 2000-2009.	117
2.3	Important events around the subprime crisis (S denotes the date of the trigger of the crisis, on August 9 th , 2008)	122

2.4	Important events around Lehman Brothers bankruptcy (L denotes the date of Lehman Brothers default, on September 15, 2008)	123
3.1	Data summary	138
3.2	Calibration results of a VAR(1) model on metals markets in the spatial dimension.	140
3.3	Communicability centrality of markets (from high to low) in the reference graph of partial correlations in the spatial dimension	147
3.4	Communicability centrality of the 10 most (left) and least (right) central markets (from high to low) in the reference graph of partial correlations in 3D	151

Introduction

La crise économique de 2007-2008 a débuté par une crise financière. L'origine de cette dernière peut être imputée à une recherche excessive de profits par les banques qui auraient octroyé toujours plus de crédits, d'une qualité de moins en moins bonne (subprimes), dans un contexte de croissance économique soutenue. Pour pouvoir offrir ces importants crédits, les banques ont usé de divers moyens, de plus en plus sophistiqués et complexes, tels que la titrisation de titres de crédits déjà titrisés, afin de contourner les contraintes réglementaires. L'innovation financière, soutenue par cette recherche de profits, est allée toujours plus loin et a engendré cette crise financière, d'abord par le défaut de particuliers sur leurs prêts hypothécaires, puis par le défaut d'institutions financières exposées à ces marchés. Ces derniers défauts ont généré une grande défiance sur les marchés interbancaires, gelant alors le marché des crédits. Le fait que cette crise ait alors affecté l'économie réelle par ce biais (les crédits) l'a rendue "systémique".

Dans une première section, nous allons prendre comme référentiel la régulation post-2008 pour détailler les différents aspects et sources de risque systémique : les agents présents sur les marchés (banques, fonds d'investissement, infrastructures, etc.) et les marchés eux-mêmes. L'étude de l'importance systémique d'un marché est plus difficile et nécessite d'en évaluer les interconnexions, en temps normal et en temps de "crise".

Dans une seconde section, nous verrons en quoi les marchés *futures* de matières premières sont un potentiel moyen de transmission à l'économie réelle, du fait d'interactions fortes – et parfois nouvelles – entre différents secteurs et marchés et avec leurs marchés physiques. Etudier le risque systémique

du point de vue des marchés et instruments nécessite donc des approches en grande dimension, à cause de la diversité des marchés et des sources d'information. Cette thèse cherche à répondre à ce besoin d'analyse de données de marché en grande dimension et à évaluer dans quelle mesure les marchés de matières premières peuvent être un vecteur de transmission de chocs financiers à la sphère économique et inversement.

0.1 Les sources de risque systémique

A la faveur de la crise de 2008, la notion de risque systémique est devenue primordiale, car les conséquences de cette crise ont été telles qu'une mobilisation internationale a été nécessaire. Les différentes institutions nationales ont dû coopérer afin de mettre en place un cadre réglementaire plus à même d'assurer la stabilité financière au niveau mondial et homogène afin d'éviter des arbitrages légaux (mouvements de capitaux vers les marchés plus souples). Le Conseil de Stabilité Financière (CSF, ou FSB pour *Financial Stability Board*) est né de cette nécessité. En coopération avec le Fonds Monétaire International (FMI, ou IMF pour *International Monetary Fund*) et la Banque des Règlements Internationaux (BRI, ou BIS pour *Bank for International Settlements*), il a rapidement redéfini le risque systémique comme :

"le risque d'une interruption des services financiers qui est (i) causée par une perturbation de tout ou d'une partie du système financier et qui (ii) peut potentiellement affecter l'économie réelle de manière significative et négative."¹ dans BIS, FSB et IMF (2009).

Cette définition, largement utilisée par les régulateurs, met en évidence deux fondements de ce risque : il provient du système financier et il doit avoir de

¹En anglais : "*The paper defines systemic risk as a risk of disruption to financial services that is (i) caused by an impairment of all or parts of the financial system and (ii) has the potential to have serious negative consequences for the real economy*"

grandes chances d'affecter l'économie réelle, qui en est donc séparée. Notons aussi que dans ce rapport initial, il y eut une volonté d'avoir une définition très générale, afin de pouvoir englober ce risque dans son entièreté et dans toute sa complexité. Il peut en effet se manifester au niveau "d'une institution, d'un marché ou d'un instrument".

Pour répondre à cette distinction, plusieurs régulateurs ont été créés pour l'Union Européenne (UE). Les institutions financières ont été subdivisées en deux, les banques et les assurances (et assimilées), respectivement supervisées par l'Autorité Bancaire Européenne (ABE, ou EBA pour *European Banking Authority*) et l'Autorité Européenne des Assurances et des Pensions Professionnelles (AEAPP, ou EIOPA pour *European Insurance and Occupational Pensions Authority*). Les marchés et instruments sont gérés par l'Autorité Européenne des Marchés Financiers (AEMF, ou ESMA pour *European Securities Markets Authority*). Enfin, le Conseil Européen du Risque Systémique (CERS, ou ESRB pour *European Systemic Risk Board*) est "en charge de la surveillance macroprudentielle du système financier de l'UE et de la prévention et de l'atténuation du risque systémique"². Il doit donc veiller à la bonne articulation des différents régulateurs et a autorité sur tout le système financier européen : "les banques, les assureurs, les gestionnaires d'actifs, le secteur bancaire parallèle, les infrastructures des marchés financiers et autres établissements et marchés financiers". Un comité a donc été créé spécifiquement pour le risque systémique, car celui-ci relève d'une compétence de niveau plus élevé qu'une institution ou un marché. Les premiers travaux ont cependant commencé par la micro-régulation, en s'intéressant aux institutions.

0.1.1 Le risque systémique comme propagation de défauts d'institutions

Suite aux événements de 2007-2008, où de nombreuses institutions ont fait défaut (et faillite), les interconnexions entre institutions et leur place dans

²<https://www.esrb.europa.eu/about/background/html/index.fr.html>

le système ont dû être revues. Comme indiqué encore dans le rapport de BIS, FSB et IMF (2009), la notion d'externalité négative est au cœur de ce risque. En effet, nombre d'événements ayant eu lieu lors de la crise de 2007-2008 étaient dus à une recherche de profit au niveau individuel (une institution), sans prendre en compte les potentielles conséquences de cette recherche sur leurs contreparties, les marchés et le système financier. Comme traumatisés par cette crise, les régulateurs, mais aussi les chercheurs, se sont donc principalement focalisés sur ces défauts d'institutions financières dites "systémiques" (à définir, plus loin) et leurs conséquences : les cascades de défauts. La régulation qui s'ensuivit s'est concentrée sur les institutions, comme en témoigne cette définition du risque systémique par le régulateur américain des marchés dérivés, la *Commodity Futures Trading Commission* (CFTC) : 'Le risque systémique est "le risque que le défaut d'un acteur du marché se répercute sur les autres du fait de la nature intriquée des marchés financiers".³'

Les régulateurs leur impose de nouvelles contraintes, telles que des contraintes de liquidité, avec le ratio de liquidité à court terme (*Liquidity Coverage Ratio*, LCR) et le ratio structurel de liquidité à long terme (*Net Stable Funding Ratio*, NSFR). Une grande part de cette nouvelle régulation a aussi cherché à faire internaliser ces externalités aux acteurs financiers. Une contrainte supplémentaire est donc à appliquer aux institutions systémiques pour prévenir leur défaut. Mais comment identifier ces institutions ?

Pour ne citer que quelques uns des travaux répondant à cette question⁴, ACHARYA et al. (2017) ont développé leur "*Systemic Expected Shortfall*" (SES) comme une mesure de la propension d'une institution à être sous-capitalisée lors que le système l'est. Le fait que ce dernier soit en crise affecterait donc plus fortement les institutions financières et permettrait alors

³En anglais : "The risk that a default by one market participant will have repercussions on other participants due to the interlocking nature of financial markets." https://www.cftc.gov/ConsumerProtection/EducationCenter/CFTCGlossary/glossary_s.html

⁴Veillez vous référer à l'excellente revue de littérature de BENOÎT, COLLIARD et al. (2017) pour plus de références.

d'identifier les plus vulnérables. On peut noter que, dans cette approche, le système est sous-capitalisé avant que l'institution le soit. La SES ne mesure certes pas complètement la contribution d'une institution au risque systémique, mais permet néanmoins d'estimer dans quelle mesure elle serait affectée, donc si elle ferait alors défaut et propagerait davantage les chocs. Elle semble donc répondre au besoin de connaître les conséquences d'un événement pour en évaluer la portée systémique ou non, donc de prévenir son occurrence ou non en imposant une contrainte en capital adéquate.

Une autre méthode très appréciée des régulateurs est celle de ADRIAN et BRUNNERMEIER (2016). Leur CoVaR mesure, quant à elle, la variation de *Value-at-Risk* (VaR) du système, conditionnellement au fait qu'une institution se trouve dans une mauvaise situation (par rapport à son état médian). Ici, contrairement à ACHARYA et al. (2017), l'institution affecte le système et non l'inverse. Il est donc possible dans ce cas, de connaître la contribution de chaque institution au "risque" du système, donc de lui faire internaliser ses externalités.

D'autres approches vont plus en détail, comme celle de HAUTSCH, SCHAUMBURG et SCHIENLE (2015), qui cherchent les contributions de chaque institution à la VaR des autres. Ces contributions se basent sur diverses informations disponibles, telles que les bilans des institutions (notamment leur levier, leur taille, etc.) ou le contexte macroéconomique et peuvent ensuite s'agréger pour obtenir la VaR du système. Plus détaillé encore, le modèle de CAPPONI et LARSSON (2015) cherche à représenter les cascades de défauts en prenant en compte les interactions entre institutions par la totalité de leurs bilans. Toutes les prises de participation d'une institution dans une autre, toutes les expositions aux mêmes actifs sont considérées, ce qui permet alors d'observer tous les mécanismes en jeu lors d'un choc.

Sur la base de ces travaux et de consultations et discussions, plusieurs critères d'évaluation du caractère systémique d'une institution ont émergé, tels que la taille, le caractère multi-juridictionnel et la complexité de l'activité, la (non-)substituabilité de l'institution et enfin son interconnexion. Le Comité de Bâle (ou BCBS pour *Basel Committee on Banking Supervision*),

hébergé à la BRI, a en effet proposé une méthodologie⁵ permettant d'évaluer l'importance systémique des banques de son échantillon avec ces critères. La méthodologie donne ensuite la contrainte de capitaux propres supplémentaires correspondante à appliquer, appelée "higher loss absorbency requirement". Le Comité soumet enfin sa liste de banques d'importance systémique globale (G-SIB pour *Global Systemically Important Banks*) au CSF chaque année, qui ensuite la valide ou non, puis la publie.

Les risques auxquels font face les banques proviennent cependant de leur activité, à savoir les actifs (dont hors-bilan) qu'ils détiennent, leurs sources de financement et capitaux propres, etc. Il convient donc de ne pas perdre de vue que les instruments et marchés peuvent aussi être responsables de crises systémiques, car un choc les affectant se répercutera alors sur tous leurs participants, les rendant vulnérables aux mouvements de prix adverses.

0.1.2 Le risque systémique comme risque de chocs sur les marchés

Les instruments et marchés financiers sont aussi l'objet de mesures visant à réduire le risque systémique qu'ils génèrent. Par exemple, certaines bourses, qui ne les avaient pas encore adoptés, ont mis en place des "circuit breakers", qui interrompent les échanges lorsque le prix varie trop brusquement. Cela fait suite aux événements du 6 mai 2010, lors du Flash Crash ayant eu lieu sur les marchés *futures* de l'indice *Standard and Poor's 500* (S&P500) et sur les actions le constituant. En effet, en raison d'arbitrages entre ces *futures*, les fonds indiciels et les actions composant l'indice, certaines actions ont connu des baisses de prix allant jusqu'à -99% pendant que d'autres ont connu des hausses importantes. Les régulateurs ont alors dû annuler de nombreuses transactions ayant eu lieu durant cette journée.

⁵La méthodologie a été finalisée en novembre 2011 dans BCBS (2011), mise à jour dans BCBS (2013) et devrait bientôt recevoir une nouvelle mise à jour, car elle doit faire l'objet d'une réévaluation tous les 3 ans.

Un choc sur un marché affecte donc tous les participants de ce marché et pourrait déjà se révéler systémique. Ce choc pourrait en outre se répercuter sur leurs contreparties, si les premiers font défaut sur leurs contrats, d'où la volonté de forcer les participants à avoir une capitalisation suffisante – au vu de leur activité. Il est cependant difficile de connaître les positions de chaque agent, sauf à leur imposer de les déclarer, par exemple auprès des régulateurs. Cela soulève le problème de la disponibilité des données, leur fréquence et leur latence. C'est pourquoi BENOÎT, HURLIN et PÉRIGNON (2015) tentent d'inférer leurs positions à partir des informations que ces agents ont obligation de divulguer, comme leurs VaR (par secteur), mais étant donnée la vitesse à laquelle il est possible d'échanger, il est probable qu'un tel travail soit toujours en retard. Agir sur les sources de leurs risques, les marchés, pourrait donc se révéler plus efficace.

Plus loin dans le processus de traitement des transactions, la technologie "blockchain" génère elle aussi un engouement certain et pourrait profondément modifier le fonctionnement des marchés, notamment le système de règlement-livraison. Par exemple, les blockchains auraient le potentiel de remplacer les chambres de compensation – mais ce n'est pas le sujet ici. Une chambre de compensation joue en général le rôle de contrepartie centrale lors d'une transaction sur un produit qu'elle compense : elle est l'acheteur du produit pour l'agent qui le vend et le vendeur du produit pour l'agent qui l'achète. Elle peut ainsi internaliser une transaction et compenser certains risques sur les deux positions (risque de contrepartie notamment). Jusqu'à présent, ces chambres de compensation ont généralement fait preuve d'une excellente gestion des risques, même en cas de défaut d'une des contreparties ; c'est la raison pour laquelle les régulateurs tentent d'inciter les agents à faire appel à leurs services. Par exemple, l'AEMF oblige la compensation de certains produits dérivés, même de gré-à-gré (ou OTC pour *Over-The-Counter*), dont la liste est publique⁶. Dans cette liste figurent aussi les chambres autorisées à, et acceptant de, remplir cette tâche. On y voit notamment que le *Chicago Mercantile Exchange* (CME) a décidé de réduire

⁶La liste est disponible à l'adresse : https://www.esma.europa.eu/system/files_force/library/public_register_for_the_clearing_obligation_under_emir.pdf

l'activité de sa branche européenne au profit de l'entité américaine (des chambres hors Union Européenne peuvent donc opérer sur ces contrats). Pour réduire au plus ces risques, il pourrait être optimal de regrouper toute la compensation dans une seule chambre. Mais faire compenser plus de produits pourrait concentrer les risques sur ces infrastructures essentielles ; leur défaut entraînerait donc un choc sur tous les marchés concernés donc tous leurs participants, pouvant aisément devenir systémique. Si on revient à une définition plus ancienne du risque systémique,

"Une crise systémique est une perturbation qui nuit au fonctionnement du système financier et, dans le pire des cas, en cause un arrêt complet. Les risques systémiques sont les risques qui causeraient potentiellement de telles crises. Les crises systémiques peuvent avoir diverses origines, mais affecteront inévitablement au moins une de ces trois fonctions clés du système financier : l'allocation de crédits, le système de paiements, et l'évaluation des actifs financiers."⁷ dans BIS (1992).

En cas de défaut d'une chambre de compensation, ce seraient donc le système de paiements et potentiellement l'évaluation des actifs compensés (et ceux qui leur sont liés) qui seraient touchés. Les acteurs des marchés concernés le seraient à leur tour, ce qui pourrait aisément devenir une crise systémique. Ces chambres, ainsi que les autres infrastructures de marché, font donc l'objet de nouvelles recommandations en termes de pratique de gestion des risques – dont le risque systémique – et de gouvernance dans CPSS et IOSCO (2012). En particulier, les instruments compensés par ces contreparties centrales (très souvent standardisés) doivent faire l'objet de collatéralisation adéquate par des actifs liquides et/ou peu risqués (avec décote en fonction de l'actif), puis d'appels de marge au moins quotidiens en cas

⁷En anglais : "A systemic crisis is a disturbance that severely impairs the working of the financial system and, at the extreme, causes a complete breakdown in it. Systemic risks are those risks that have the potential to cause such a crisis. Systemic crises can originate in a variety of ways, but ultimately they will impair at least one of three key functions of the financial system : credit allocation, payments, and pricing of financial assets."

de pertes. Il est donc nécessaire de préciser les manifestations possibles de ce type de crise du point de vue des marchés.

Un choc de prix sur un marché affecte déjà tous ses participants et peut les mettre en difficulté, voire en situation de défaut, en fonction de leur position sur ce marché et leur levier. Plus important encore, les marchés sont liés entre eux, économiquement ou financièrement. Donc si l'un subit un choc, il est probable que d'autres aussi, ce qui pourrait se répercuter (davantage) sur leurs participants et même déclencher un cercle vicieux où les chocs sont amplifiés par des "ruées", entraînant encore plus de pertes et mouvements. Il ne faut donc pas négliger les interactions entre marchés et agents, qui sortent cependant du cadre de cette thèse.

On trouve une première caractérisation de l'importance systémique d'un marché dans le rapport post-crise BIS, FSB et IMF (2009). Comme pour les institutions, la taille, la (non-)substituabilité et l'interconnexion étaient considérées comme des critères pertinents. La taille est relativement facile à évaluer, avec par exemple le volume, la capitalisation totale ou les positions ouvertes si ces informations sont disponibles. La non-substituabilité et l'interconnexion sont plus difficiles à évaluer. C'est sur ce dernier point que cette thèse va tenter de contribuer.

L'interconnexion est en effet essentielle si l'on souhaite évaluer le caractère systémique. Elle peut exister sous diverses formes : des liens économiques, comme entre produits pétroliers ; des liens financiers, comme entre un actif et ses produits dérivés ; etc. Ces liens peuvent être statiques (ou contemporains), comme c'est le cas par exemple avec la relation de parité *call-put*, qui stipule qu'à une date t , les prix d'un sous-jacent, d'une option d'achat (*call*) et d'une option de vente (*put*) européennes de mêmes caractéristiques sur ce sous-jacent sont liés par :

$$C_t - P_t = S_t - K e^{-r(T-t)} \quad (1)$$

avec C_t le prix du call à la date t , P_t le prix du put de mêmes caractéristiques à la date t aussi, S_t le prix du sous-jacent, K le prix d'exercice des

options, r le taux d'intérêt et T la date d'échéance des options. Lorsque cette relation n'est pas vérifiée, des opérations d'arbitrage permettent de rétablir l'égalité.

D'autres relations contemporaines existent, mais peuvent être variables dans le temps. Il peut être plus difficile de les évaluer du fait de cette variabilité. Le coefficient de corrélation linéaire permet de mesurer la dépendance linéaire (statistique) entre deux variables (à une date donnée). Par exemple, il était coutume de considérer que les rendements des actions et des obligations, ou plus généralement les variations de taux d'intérêt, n'étaient pas corrélés. Cette corrélation fluctue dans le temps et n'est pas toujours nulle.

D'autres connexions encore ne sont pas contemporaines mais peuvent se manifester dans le temps. On peut notamment penser aux influences de certains marchés géographiquement lointains. Le décalage horaire fait que la bourse de Tokyo, ouvrant à 9 heures (heure locale) et fermant à 15 heures (heure locale), n'est pas active en même temps que la bourse de New York, qui ouvre de 21 heures (heure de Tokyo) à 4 heures (heure de Tokyo). Ainsi, si des informations (pertinentes pour Tokyo) arrivent sur la bourse américaine, elles ne seront intégrables que le lendemain pour la bourse japonaise. Il est donc important de prendre en compte cette dimension temporelle dans les connexions entre marchés.

La notion de "causalité de Granger" développée par GRANGER (1969) prend en compte ce type de relations. Si une variable à une date t dépend des observations d'une autre à une date antérieure (elles apportent de l'information), alors cette dernière "cause" la première. Un modèle de séries temporelles de type Vecteur AutoRégressif (VAR) permet d'intégrer cette notion directement. Si on appelle \mathbf{Y}_t un ensemble de variables aléatoires à une date t , un VAR d'ordre 1, VAR(1) s'écrit alors

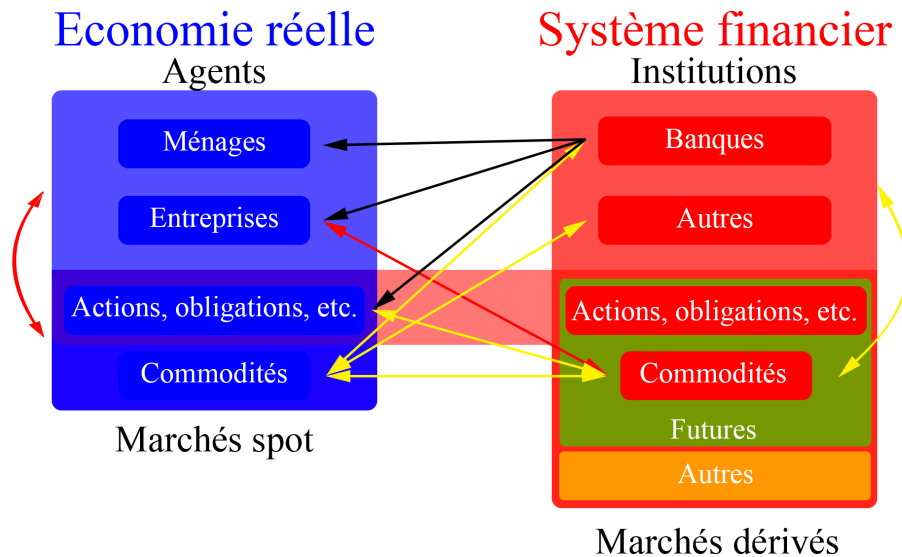
$$\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_{t-1} + \epsilon_t \quad (2)$$

avec \mathbf{A} la matrice qui regroupe les coefficients indiquant dans quelle mesure chaque variable dépend des autres (observées à la date précédente) et ϵ_t un vecteur de bruit pour chacune de ces variables.

Nous proposons donc de décomposer le risque systémique sur les marchés en deux. La première composante représentera le risque systématique – risque que tous les marchés soient affectés par un choc en même temps – par l'étude des corrélations. La deuxième composante représentera le risque de propagation de chocs (individuels) par l'étude de la causalité à la Granger. En particulier, nous nous intéresserons à des variables pouvant directement affecter l'économie réelle : les prix des matières premières. Des agents économiques (par exemple des entreprises) peuvent avoir recours aux différents instruments financiers permettant de gérer les risques (de prix) de leur activité. En effet, certains ont leurs propres équipes intervenant directement sur les marchés, comme EDF notamment. Un choc sur les marchés financiers pourrait, par le biais de ces instruments, affecter l'économie réelle et générer une crise systémique.

La Figure 0.1 présente une vision simplifiée du système financier et de ses liens avec l'économie réelle. Elle montre en rouge les liens dont nous venons de parler, entre les entreprises et les marchés dérivés (ici plus précisément les futures), afin de gérer leurs risques sur les commodités. Les flèches noires indiquent les canaux de transmission "traditionnels" des banques à la sphère réelle et souvent mis en cause : l'octroi de crédits et le financement de l'économie. En outre, depuis 2003-2004, on observe ce qu'on appelle la "financiarisation des matières premières", qui se manifeste de différentes façons – entre autres par une corrélation accrue entre les rendements des marchés de matières premières et des marchés financiers. Ce sont les liens jaunes dans la Figure 0.1, sur lesquels nous reviendrons dans la section suivante. La finance est par ailleurs amenée à prendre une place encore plus importante dans les matières premières et l'environnement car on parle récemment de financer la transition énergétique. Cette relation de plus en plus forte a soulevé de nombreuses inquiétudes, dont celle d'une potentielle influence de la finance sur les commodités, et peut donc affecter l'économie réelle.

FIGURE 0.1 – Schéma synthétique des interactions entre le système financier et l'économie réelle



L'ensemble bleu représente l'économie réelle, tandis que l'ensemble rouge, incluant une partie de l'économie réelle, représente le système financier. Seuls les liens importants ici sont dessinés. En noir, on trouve les liens naturels des banques, avec les ménages et entreprises par l'octroi de crédits, et avec les marchés actions, obligataires, etc. par leur activité traditionnelle d'investissement. Les liens naturels des agents économiques avec les matières premières et leur couverture sont en rouge. Enfin, en jaune, on trouve les canaux de transmission créés ou facilités par la financiarisation des commodités. Le cadre vert indique l'univers d'étude de cette thèse.

0.2 La financiarisation des matières premières, nouvelle source de risque systémique ?

Les matières premières, notamment les sources d'énergie, sont à la base de nombreuses activités économiques. Elles peuvent aussi faire partie des indices de prix pour les calculs d'inflation et ont donc une importance cruciale dans l'économie. Un choc sur leurs prix a donc le potentiel pour générer des

difficultés pour tous les agents et institutions, comme par exemple lors des chocs pétroliers. Bien qu'il soit impossible d'avoir un prix unique pour une même matière première dans le monde (notamment à cause des coûts de transport et de stockage), certains prix servent de référence internationale, comme par exemple les prix du West Texas Intermediate (WTI) ou du Brent pour le pétrole. Un choc de prix sur ces marchés pourrait rapidement se répercuter à tous les marchés auxquels ils sont liés, dans le monde entier, et à tous les agents les utilisant, donc devenir une crise économique. En outre, un phénomène relativement récent a fortement lié les marchés de commodités et les actifs financiers : la financiarisation des matières premières⁸. Elle a commencé par l'utilisation de contrats *futures* portant sur des matières premières en tant qu'actifs en portefeuilles.

Les "*futures*" sont des produits dérivés, des contrats permettant d'effectuer une transaction à une date ultérieure. Un acheteur de *futures* s'engage alors à acheter une certaine quantité de l'actif sous-jacent (dont les caractéristiques sont aussi fixées) à la date de maturité du contrat. En face, le vendeur du *futures* s'engage à vendre cette quantité de sous-jacent à l'échéance du contrat. Les "*futures*" se caractérisent par une standardisation des termes du contrat, définis par la bourse : actif sous-jacent (qualité, catégorie, etc.), quantité, lieu de livraison (si livraison il y a) et la maturité. Cette standardisation, certes contraignante, permet une plus grande liquidité du marché par rapport aux contrats "*forward*" complètement sur-mesure. Les contrats *futures* peuvent être échangés rapidement, voire instantanément, car beaucoup sont cotés en continu, du fait de l'électronisation des marchés. Ainsi, de nombreux spéculateurs (définis ici de manière large comme les agents ne cherchant pas à couvrir un risque physique) se sont intéressés à ces contrats, leur octroyant une exposition à ces actifs sous-jacents sans (devoir) les détenir et récupérant l'éventuelle prime de risque associée.

L'attrait pour les matières premières n'est pas nouveau. Plusieurs indices sur matières premières ont vu le jour avant 2000, comme le Standard and Poor's - Goldman Sachs Commodity Index (S&P-GSCI) ou le Bloomberg Commodity Index (anciennement Dow Jones - UBS Commodity Index).

⁸CHENG et XIONG (2014) fournissent une très bonne revue de cette littérature.

Cependant, leur essor n'a commencé qu'en 2003-2004, en même temps que la hausse des prix des matières premières. Coïncidence ou non, à cette période, plusieurs articles ont étudié les corrélations entre matières premières et entre commodités et actifs financiers. L'étude de GORTON et ROUWENHORST (2006), dont la première version disponible remonte à 2004, a montré qu'un indice constitué de *futures* sur matières premières pouvait constituer un support d'investissement intéressant. Cet indice offrait des rendements similaires au S&P500 sur la période 1959-2004 mais avec une volatilité plus faible et une corrélation faible voire négative avec cet indice actions. ERB et HARVEY (2006) (déjà cité dans la première version de Gorton et Rouwenhorst, donc aussi disponible en 2004) ont montré qu'un indice de *futures* sur matières premières était aussi performant qu'un indice actions, bien qu'en moyenne ces *futures* aient individuellement une rentabilité nulle. Ils montrent que cela est dû à la fois à une réactualisation des poids des *futures* et à un effet de diversification parmi ces contrats sur matières premières. Les futures sur matières premières permettaient donc aux agents financiers de diversifier leur portefeuille, donc en en réduisant le risque, tout en leur offrant la liquidité nécessaire à la vitesse de leur activité. De nombreux fonds échangeables en bourse (ou ETF pour Exchange Traded Fund) ont alors été créés pour offrir une certaine diversité de matières premières dans lesquelles investir.

Au même moment, le prix de nombreuses commodités a commencé à augmenter, tout comme leurs corrélations, comme le notent TANG et XIONG (2012). Plus précisément, ils trouvent que la participation plus importante de ces investisseurs (prenant des positions longues, les Commodity Index Traders, CIT) depuis 2004 a entraîné une augmentation des corrélations entre matières premières, davantage encore si elles font partie d'un des deux plus gros indices (ceux mentionnés ci-dessus). Ils arrivent aussi à la conclusion que les prix des commodités sont alors devenus plus volatils, ces investisseurs apportant alors la volatilité des marchés financiers. De leur côté, BÜYÜKŞAHİN et ROBE (2014) trouvent qu'à partir de 2008, une plus grande participation des "spéculateurs" (définis très généralement)

augmente la corrélation entre les rentabilités des futures sur matières premières et des marchés financiers. De plus, les positions des hedge funds (présents sur les deux secteurs) permettent de prédire la corrélation leurs rentabilités, mais pas celles des CIT de TANG et XIONG (2012). Ils travaillent en effet sur une des bases de données de la CFTC, confidentielle et extrêmement détaillée, ce qui leur permet d'identifier les agents et leur type d'activité. Enfin, ils montrent que cette plus grande corrélation disparaît en temps de crise, du fait du retrait de ces agents.

Ces résultats indiquent qu'un choc sur les marchés financiers pourrait affecter de nombreuses matières premières par le biais de ces contrats futures, largement utilisés par des agents financiers.

Les marchés futures de matières premières représentaient au moins 38% du volume total des marchés futures, avec 5,6 milliards de contrats échangés dans le monde en 2017, d'après le rapport⁹ de la World Federation of Exchanges (WFE). Le volume des futures a crû en tendance depuis 2005, avec une baisse en 2017 due uniquement à une baisse des volumes en Asie pacifique. Les commodités représentent la deuxième classe de futures en volume, derrière les actions et indices actions. Comme pour ces derniers, ces marchés dérivés ont joué un rôle important dans le processus de formation des prix *spot* (prix valable à l'instant présent), au point qu'il est usuel de prendre les prix du futures le plus "proche" comme prix de référence pour des transactions *spot*. Or ces produits étant "dérivés" par nature, leur prix devrait dépendre du prix *spot* et non l'inverse. Nous verrons que ces deux prix ont en fait des interactions fortes et qu'il est donc nécessaire d'étudier les prix futures si l'on souhaite comprendre le fonctionnement des marchés financiers et de matières premières.

Nous nous intéresserons dans un premier temps aux mécanismes reliant les prix *spot* et futures. Dans un second temps, nous verrons les interactions

⁹Le rapport est disponible à l'adresse suivante : <https://www.world-exchanges.org/home/index.php/files/53/IOMA-Derivatives-Market-Survey/500/2017-IOMA-Derivatives-Market-Survey.pdf>. N'ayant pas le détail par sous-jacent ni maturités, il n'est pas possible de trouver la valeur totale de ces contrats échangés.

– anciennes et nouvelles – entre commodités et entre commodités et actifs financiers, créant ainsi un lien entre les marchés financiers traditionnels et les prix spot de matières premières, donc l'économie réelle.

0.2.1 Les relations entre prix spot et prix futures pour les matières premières

Les contrats futures sont des produits dérivés, c'est-à-dire que leur valeur dérive de la valeur de leur sous-jacent. La théorie du stockage par KALDOR (1939), WORKING (1949) et BRENNAN (1958) est ainsi formulée. Elle stipule que le prix futures est une fonction du prix spot, du coût de stockage du sous-jacent, du financement de ces opérations et des revenus associés. En temps continu, on pourrait écrire :

$$F_{t,T} = S_t \exp((r + cs - y) * (T - t)) \quad (3)$$

si on note $F_{t,T}$ le prix futures à la date t pour une échéance T , S_t le prix spot à la date t , r le taux d'intérêt, cs les coûts de stockage en pourcentage du prix spot et y les revenus associés à la détention de l'actif sous-jacent. Pour les matières premières, ces revenus pourraient correspondre à ce qui est généralement appelé le *convenience yield*. Il pourrait être vu comme une option de vente américaine sur le sous-jacent (qu'on détient), permettant de le vendre au "bon" moment ou de l'utiliser lorsqu'on en aura besoin, comme l'étudient HEINKEL, HOWE et HUGHES (1990).

Si cette relation n'est pas vérifiée, il est en théorie possible de la rétablir par des arbitrages. Prenons le cas où le prix futures est supérieur au montant donné par l'expression à droite du signe d'égalité dans l'équation 3. Il serait intéressant de vendre ce futures, d'acheter le sous-jacent et de le stocker. Il est possible de clore ces positions à tout moment (lorsque cela est profitable), mais les conserver jusqu'à la maturité évite les risques de prix. Si la livraison est demandée par l'acheteur du futures, il suffit de livrer le sous-jacent possédé, sinon, on peut le revendre (et clore sa position en rachetant les futures). Ces opérations peuvent être mises en place jus-

qu'à ce que l'égalité revienne et/ou dans la limite des contraintes, telles que la non-négativité des stocks (on ne peut pas vendre à découvert un actif physique) ou le maximum de capacité de stockage. Il faut de plus pouvoir évaluer les différentes composantes du prix futures. Le *convenience yield* est particulièrement difficile à évaluer, car on peut le représenter par une option, donc une fonction non linéaire du prix, et dont les avantages sont difficilement quantifiables. En contrôlant pour le taux d'intérêt et les coûts de stockage, on peut cependant le déduire des prix observés, comme le font FAMA et FRENCH (1988) ou NG et PIRONG (1994) pour les métaux.

La différence entre le prix futures et le prix spot (souvent rapportée au prix spot) est appelée la base. Une base négative correspond à une situation de déport (ou backwardation), où le "convenience yield" est donc élevé, relativement aux autres composantes. Celui-ci est d'autant plus important que l'offre en actif sous-jacent est insuffisante pour combler la demande. Il correspond donc à une situation de rarification de la matière première (soumise à une contrainte de non-négativité), limitant aussi les arbitrages. Celle-ci devrait donc contenir des informations relatives à l'offre, la demande et les inventaires sur le marché. Suivant DEATON et LAROQUE (1992) et GORTON, HAYASHI et ROUWENHORST (2013), cette insuffisance d'inventaires se répercuterait en termes de volatilité des prix spot et des futurs prix spot. En effet, ces inventaires permettent d'atténuer les variations de prix en répondant aux chocs de demande (ou d'offre en augmentant les inventaires).

Comme cette contrainte ne s'applique pas aux actifs financiers, leurs rendements ne devraient pas en subir les conséquences. Nous allons donc, dans le chapitre 1, étudier les statistiques univariées des rendements des différents actifs. Nous cherchons à évaluer dans quelle mesure les prix de futures sur matières premières reflètent encore leurs fondamentaux et leurs contraintes et dans quelle mesure ces marchés se sont financiarisés.

Il peut cependant y avoir d'autres explications à une situation de déport. Initialement appelée théorie du déport normal par KEYNES (1930), la théorie de la pression à la couverture stipule qu'une prime de risque doit ré-

munérer les agents acceptant de prendre les risques à la place des agents économiques. A l'époque, HICKS (1946) considérait que les agents souhaitant vendre à terme (producteurs, etc.) étaient plus nombreux, car avaient de plus fortes raisons de se couvrir, que ceux souhaitant acheter, donc la pression à la couverture était nette vendeuse. Pour pouvoir vendre à terme, ces agents doivent donc accepter de vendre à un prix plus faible que celui que le marché anticipe, afin d'encourager des agents à se porter contrepartie : la différence entre l'anticipation et le prix futures est la prime de risque. Ce dernier est donc fixé en dessous des anticipations de prix, qui peuvent même être inférieures au prix spot (d'où l'envie de se couvrir), entraînant une situation de déport.

Enfin, BLACK (1976) estime que dans le Modèle d'Evaluation Des Actifs Financiers (MEDAF, ou CAPM pour Capital Asset Pricing Model), les contrats sur matières premières devraient être indépendants du portefeuille de marché. En effet, comme ils représentent un jeu à somme nulle, ce portefeuille ne contiendrait pas ces contrats (car la position nette est nulle). Ainsi, pour lui, le prix futures devrait être égal au prix spot anticipé, n'offrant pas de rémunération car il n'y aurait pas de risque systématique à rémunérer dans ces contrats.

Les marchés futures, par leur liquidité, sont cependant devenus des lieux de découverte et formation des prix. Les futures étant plus faciles à échanger que des actifs physiques, les spéculateurs préféreraient a priori intervenir sur ces premiers (certains sont aussi actifs sur les marchés physiques). La présence de nombreux agents permet en effet de confronter leurs informations et anticipations et donc de former un prix "de marché". Pour les futures sur matières premières, bien que le lieu de livraison soit prédéfini, donc d'une portée a priori locale, leur prix sert souvent d'étalon au niveau national, voire international, comme les prix des pétroles américain (West Texas Intermediate) ou britannique (Brent). Comme le dit aussi BLACK (1976), les futures permettent en effet d'avoir un prix de référence pour les agents et pour leur marché spot, relativement proche si on considère le contrat futures de maturité la plus courte.

Ces *front-month contracts* sont généralement les plus liquides et les plus liés aux marchés spot, du fait de leur proximité temporelle. En effet, plus la maturité d'un contrat diminue, plus les opérations d'arbitrage sont aisées : stockage moins long voire inutile (si on prend en compte le temps de livraison, ...), etc. A la maturité du contrat, si les arbitrages sont possibles, on devrait donc observer ce qu'on appelle la convergence de la base : elle devrait devenir nulle à l'échéance.

Plusieurs auteurs se sont donc attachés à montrer ce transfert de fonction, à savoir que les prix futures influencent les prix spot. GARBADE et SILBER (1983) montrent que les marchés futures dominant en termes de flux d'information (mais les prix spot jouent encore un rôle) pour une variété de matières premières : agricoles (blé, maïs, avoine et jus d'orange) et métaux (or, argent et cuivre). FIGUEROLA-FERRETTI et GONZALO (2010) approfondissent le travail de Garbade et Silber et l'appliquent aux métaux non-ferreux. Alors que QUAN (1992) montrait justement que les prix spot sur pétrole remplissaient encore la fonction de découverte des prix, en appliquant la même méthode que Garbade et Silber, SCHWARZ et SZAKMARY (1994) montrent l'inverse. Il est fort probable que les deux soient vrais, que les prix futures et spot soient fortement interconnectés.

Ces flux d'informations des marchés futures vers les marchés spot existeraient même sur les marchés financiers, comme le montrent STOLL et WHALEY (1990), alors que ces actifs ne sont pas aussi lourds à échanger ou stocker. Même en incluant les ETFs, qui prennent une part de plus en plus importante de l'activité, HASBROUCK (2003) trouve que les futures sur S&P500 apportent le plus d'informations.

Dans le contexte de la financiarisation, les ETFs sur matières premières ont aussi connu un essor considérable et participeraient aussi à la découverte des prix. Pour s'exposer, ils peuvent recourir soit à une réplique synthétique, en utilisant des produits financiers (futures, swaps, etc.), soit à une réplique physique en détenant les matières premières concernées. BENOÎT, HURLIN et PÉRIGNON (2015) montrent que même les banques auraient des expositions aux matières premières, certaines en détenant physiquement. Leur influence en termes de prix de commodités serait alors partagée entre

les marchés spot et dérivés. Mais comme mentionné plus haut, la présence de certains agents sur les marchés de matières premières et financiers aurait créé un "lien" entre ces deux secteurs, l'un pouvant alors influencer l'autre. Ainsi, l'activité des agents financiers sur ces marchés forme un pont entre les sphères financière et économique, pouvant potentiellement faciliter la propagation de chocs de la première à la seconde, donc faciliter l'occurrence de crises systémiques.

0.2.2 Les relations entre matières premières et avec d'autres actifs

Certaines matières premières sont naturellement liées, comme par exemple le pétrole et les produits obtenus à partir de celui-ci (gasoil, etc.) ou le soja et ses produits (huile de soja, etc.). Les prix (rendements) des produits devraient donc être fortement corrélés à ceux de leur source qui pourrait donc propager des chocs de prix. On retrouve ainsi une sectorisation des marchés, où les marchés naturellement liés sont fortement corrélés (intégration intra-secteur) et les liens entre différents secteurs (inter-secteurs) sont plus faibles.

Il y a aussi des liens de substituabilité : certaines matières premières agricoles sont substituables. Lorsque l'une, comme le blé, devient trop chère, il est possible de se rabattre sur d'autres, dans le cadre de l'alimentation par exemple. Un effet d'égalisation des prix devrait donc apparaître entre ces matières premières. De tels liens pourraient même se matérialiser par des corrélations négatives s'il y a parfaite substituabilité, si les stocks des différentes commodités le permettent (évitant ainsi les problèmes de facteurs de production). Il semble cependant que tel ne soit pas le cas, comme nous verrons dans le chapitre 1 et comme le montrent *ERB* et *HARVEY* (2006). Au contraire, les corrélations entre matières premières agricoles sont positives. Depuis l'introduction des biocarburants, les prix des commodités agricoles utilisées pour leur production ont aussi commencé à se lier à ceux du secteur énergétique, créant alors de fortes connexions inter-secteurs.

On distingue généralement métaux industriels et métaux précieux. Les premiers étant moins rares et largement utilisés dans divers secteurs économiques, devraient être liés à l'activité économique et donc avec les marchés actions, comme pour FAMA et FRENCH (1988).

Les seconds sont non seulement utilisés comme tels, par exemple pour les bijoux ou pour l'électronique, mais aussi comme actifs défensifs, valeurs refuge lorsque les marchés actions chutent. Leurs futures sont donc parfois considérés comme des actifs financiers, comme le font GORTON, HAYASHI et ROUWENHORST (2013). Leurs corrélations avec les actifs financiers sont donc généralement négatives, comme nous le trouvons pour certaines matières premières dans le chapitre 1. Les prix des métaux précieux (qu'il faut aussi subdiviser) sont en revanche liés aux changements quotidiens dans les taux de change, d'après SARI, HAMMOUDEH et SOYTAS (2010) et peuvent donc être soumis aux variations économiques relatives des différents pays.

Bien que ces derniers auteurs ne trouvent pas de lien entre le pétrole et les taux de change, S.-S. CHEN et H.-C. CHEN (2007) montrent qu'il y a une relation de cointégration de long terme entre les deux (en termes réels et non nominaux). De plus, les prix du pétrole permettraient de prédire "hors échantillon" les taux de change réels à tous horizons, avec de meilleures performances à long terme. Comme le soulignent FERRARO, ROGOFF et ROSSI (2015), la devise d'un pays dépend fortement de ses plus grosses exportations. Ils trouvent ainsi que les prix du pétrole, nominaux cette fois, sont un bon prédicteur hors échantillon des taux de change à un horizon quotidien, moins à plus long terme.

La liste des connexions possibles entre marchés et secteurs peut être longue. Du fait de la mondialisation, les échanges ont été facilités et de nombreux liens peuvent évoluer dans le temps. Avec la financiarisation des matières, certains liens peuvent être nouveaux et éphémères. Il est donc important d'avoir des indicateurs clairs de ces liens au cours du temps, à des fréquences aussi courtes que journalières, voire intra-journalières. Comme

pour le risque systémique, il est possible de différencier deux types de liens : les liens contemporains, mesurant l'intégration (avec les corrélations par exemple) et les liens temporels, généralement qualifiés de *spillovers* (propagation). De plus, il est nécessaire de considérer un large système de marchés afin de ne pas oublier des liens qui pourraient être importants, d'où l'utilisation des statistiques en grande dimension ou de réduction de dimension. Par exemple, LE PEN et SÉVI (2018) expliquent les corrélations entre différents futures sur matières premières en contrôlant pour de nombreuses variables économiques et financières, réduites à un plus petit nombre grâce à un modèle à facteurs. Ils trouvent ainsi que ces corrélations sont dues à la demande de couverture et de spéculation, donc à la financiarisation. SILVENNOINEN et THORP (2013) étudient aussi un large système de futures sur commodités (résumés dans une seule variable par commodité) et actifs financiers et trouvent aussi des preuves de financiarisation.

Cette nécessité a aussi fait émerger l'utilisation des graphes/réseaux pour étudier de tels systèmes. On trouve notamment DIEBOLD, YILMAZ et LIU (2017) qui étudient les *spillovers* de volatilité entre différentes matières premières et peuvent ainsi identifier celles qui en transmettent le plus ou en reçoivent le plus. De même, LAUTIER et RAYNAUD (2012) appliquent des outils de la théorie des graphes à un système de 220 variables de différents secteurs et montrent leur intégration en termes d'intensité. Ils trouvent cependant une sectorisation en termes de structure, mettant l'énergie au centre du système.

Cette thèse s'inscrit dans ce courant, en cherchant à travailler sur un large système de variables, à la fois de commodités (contrats futures ici, de différents secteurs) et d'actifs financiers.

Le chapitre 1 de cette thèse s'attache à étudier les marchés individuellement, car un choc de prix sur un de ces marchés affecte tous ses participants et pourrait se propager à d'autres. Nous tentons de retrouver dans les distributions de rentabilités des faits stylisés et des signes de la pertinence des fondamentaux économiques des matières premières. En particulier, la contrainte de non-négativité des stocks pour les actifs physiques (commodités) devrait se traduire par des distributions de rentabilités particulières.

Nous utilisons donc la base, apportant théoriquement des informations sur les stocks, pour mettre en avant ces signes et voir s'ils se traduisent aussi en termes de chocs. Enfin, nous analysons brièvement les corrélations entre ces variables et confirmons l'apparition d'un facteur commun aux matières premières d'abord à partir de 2004, puis englobant ensuite actifs financiers à partir de 2008.

Le chapitre 2 entre plus en détail dans l'étude des corrélations. Estimant que chaque maturité peut apporter des informations différentes, nous ajoutons à chaque marché plusieurs de ses contrats futures, travaillant alors en grande dimension. Nous fournissons donc ce que nous qualifierons de travail en trois dimensions (3D), car nous intégrons dans notre système : 1) des marchés différents en termes de sous-jacents et de localisation (dimension spatiale) ; 2) les maturités de chaque marché considéré (dimension maturité) ; 3) la dimension temporelle (évolution dynamique). Une approche dynamique suffisamment sensible à l'évolution (journalière) du système et le grand nombre de variables rendent les modèles économétriques classiques difficiles à estimer correctement, d'où l'utilisation de la théorie des graphes. Suivant la méthodologie de LAUTIER et RAYNAUD (2012), nous l'appliquons à l'étude de chocs financiers afin d'en déterminer les conséquences sur les actifs financiers et sur les matières premières. Nous introduisons d'autres outils de la théorie des graphes et montrons que le commencement de la crise des subprimes n'a pas eu d'effet brusque sur ce système. Au contraire, le défaut de Lehman Brothers a profondément mais temporairement modifié la structure du système, sans propagation du choc aux matières premières.

Enfin, le chapitre 3 applique des outils d'apprentissage statistique permettant d'obtenir des résultats interprétables malgré la grande dimension, en sélectionnant les variables pertinentes. Sur la base de ces outils d'apprentissage, nous définissons un ensemble de mesures (graphiques) pertinentes pour évaluer les deux composantes du risque systémique : le risque systématique (issu de l'intégration des marchés) et le risque de propagation. Nous nous appuyons sur un modèle de type Vecteur AutoRégressif (VAR), estimé en grande dimension grâce à une pénalisation LASSO. L'utilisation des cor-

rélations partielles au lieu des corrélations inconditionnelles dans ce modèle permet en effet d'avoir une matrice creuse, propice à cette pénalisation et donc à la sélection des liens pertinents. Cette matrice met aussi en évidence la sectorisation des marchés au début de l'échantillon (année 2001), qui s'efface progressivement. Cela prouve l'intégration croissante entre marchés financiers et de matières premières, d'abord en termes de structure, puis en termes d'intensité des liens. Cette étude montre cependant peu de liens de propagation temporelle, même autour de chocs financiers importants tels que le début de la crise des subprimes, le défaut de Lehman Brothers, les rentabilités minimale et maximale du S&P500 et le Flash Crash du 6 mai 2010.

Références

- ACHARYA, Viral V. et al. (2017). 'Measuring Systemic Risk'. In : *The Review of Financial Studies* 30(1), p. 2–47.
- ADRIAN, Tobias et Markus K. BRUNNERMEIER (2016). 'CoVaR'. In : *American Economic Review* 106(7), p. 1705–41.
- BCBS (2011). *Global systemically important banks : assessment methodology and the additional loss absorbency requirement*. Rapp. tech. Basel Committee on Banking Supervision, Bank for International Settlements.
- BCBS (2013). *Global systemically important banks : updated assessment methodology and the higher loss absorbency requirement*. Rapp. tech. Basel Committee on Banking Supervision, Bank for International Settlements.
- BENOÎT, Sylvain, Jean-Edouard COLLIARD et al. (2017). 'Where the Risks Lie : A Survey on Systemic Risk'. In : *Review of Finance* 21(1), p. 109–152.
- BENOÎT, Sylvain, Christophe HURLIN et Christophe PÉRIGNON (2015). 'Implied Risk Exposures'. In : *Review of Finance* 19(6), p. 2183–2222.
- BIS (1992). *Recent Developments in International Interbank Relations*. Rapp. tech. Bank for International Settlements.

- BIS, FSB et IMF (2009). *Guidance to Assess the Systemic Importance of Financial Institutions, Markets and Instruments : Initial Considerations*. Rapp. tech. Bank for International Settlements, Financial Stability Board et International Monetary Fund.
- BLACK, Fischer (1976). 'The pricing of commodity contracts'. In : *Journal of Financial Economics* 3(1), p. 167–179.
- BRENNAN, Michael J (1958). 'The supply of storage'. In : *The American Economic Review* 48(1), p. 50–72.
- BÜYÜKŞAHİN, Bahattin et Michel A. ROBE (2014). 'Speculators, commodities and cross-market linkages'. In : *Journal of International Money and Finance* 42, p. 38–70.
- CAPPONI, Agostino et Martin LARSSON (2015). 'Price Contagion through Balance Sheet Linkages'. In : *The Review of Asset Pricing Studies* 5(2), p. 227–253.
- CHEN, Shiu-Sheng et Hung-Chyn CHEN (2007). 'Oil prices and real exchange rates'. In : *Energy Economics* 29(3), p. 390–404.
- CHENG, Ing-Haw et Wei XIONG (2014). 'Financialization of commodity markets'. In : *Annu. Rev. Financ. Econ.* 6(1), p. 419–441.
- CPSS et IOSCO (2012). *Principles for financial market infrastructures*. Rapp. tech. Committee on Payment, Settlement Systems, Bank for International Settlements et International Organization of Securities Commissions.
- DEATON, Angus et Guy LAROQUE (1992). 'On the behaviour of commodity prices'. In : *The Review of Economic Studies* 59(1), p. 1–23.
- DIEBOLD, Francis X., Kamil YILMAZ et Laura LIU (2017). 'Commodity Connectedness'. In : *NBER, Working Paper No. 23685*.
- ERB, Claude B. et Campbell R. HARVEY (2006). 'The Strategic and Tactical Value of Commodity Futures'. In : *Financial Analysts Journal* 62(2), p. 69–97.
- FAMA, Eugene F. et Kenneth R. FRENCH (1988). 'Business Cycles and the Behavior of Metals Prices'. In : *The Journal of Finance* 43(5), p. 1075–1093.

- FERRARO, Domenico, Kenneth ROGOFF et Barbara ROSSI (2015). ‘Can oil prices forecast exchange rates? An empirical analysis of the relationship between commodity prices and exchange rates’. In : *Journal of International Money and Finance* 54, p. 116–141.
- FIGUEROLA-FERRETTI, Isabel et Jesús GONZALO (2010). ‘Modelling and measuring price discovery in commodity markets’. In : *Journal of Econometrics* 158(1), p. 95–107.
- GARBADE, Kenneth D. et William L. SILBER (1983). ‘Price Movements and Price Discovery in Futures and Cash Markets’. In : *The Review of Economics and Statistics* 65(2), p. 289–297.
- GORTON, Gary B, Fumio HAYASHI et K Geert ROUWENHORST (2013). ‘The fundamentals of commodity futures returns’. In : *Review of Finance* 17(1), p. 35–105.
- GORTON, Gary B et K Geert ROUWENHORST (2006). ‘Facts and Fantasies about Commodity Futures’. In : *Financial Analysts Journal* 62(2), p. 47–68.
- GRANGER, C. W. J. (1969). ‘Investigating Causal Relations by Econometric Models and Cross-spectral Methods’. In : *Econometrica* 37(3), p. 424–438.
- HASBROUCK, Joel (2003). ‘Intraday Price Formation in U.S. Equity Index Markets’. In : *The Journal of Finance* 58(6), p. 2375–2400.
- HAUTSCH, Nikolaus, Julia SCHAUMBURG et Melanie SCHIENLE (2015). ‘Financial Network Systemic Risk Contributions’. In : *Review of Finance* 19(2), p. 685–738.
- HEINKEL, Robert, Maureen E. HOWE et John S. HUGHES (1990). ‘Commodity convenience yields as an option profit’. In : *Journal of Futures Markets* 10(5), p. 519–533.
- HICKS, J. R. (1946). *Value and Capital : An Inquiry into some Fundamental Principles of Economic Theory*. Oxford University Press.
- KALDOR, Nicholas (1939). ‘Speculation and Economic Stability’. In : *The Review of Economic Studies* 7(1), p. 1–27.
- KEYNES, John Maynard (1930). *A Treatise on Money : In 2 Volumes*. Macmillan & Company.

- LAUTIER, Delphine et Franck RAYNAUD (2012). ‘Systemic Risk in Energy Derivative Markets : A Graph-Theory Analysis’. In : *The Energy Journal* 33(3), p. 215–239.
- LE PEN, Yannick et Benoît SÉVI (2018). ‘Futures Trading and the Excess Co-movement of Commodity Prices’. In : *Review of Finance* 22(1), p. 381–418.
- NG, Victor K. et Stephen Craig PIRRONG (1994). ‘Fundamentals and Volatility : Storage, Spreads, and the Dynamics of Metals Prices’. In : *The Journal of Business* 67(2), p. 203–230.
- QUAN, Jing (1992). ‘Twostep testing procedure for price discovery role of futures prices’. In : *Journal of Futures Markets* 12(2), p. 139–149.
- SARI, Ramazan, Shawkat HAMMOUDEH et Ugur SOYTAS (2010). ‘Dynamics of oil price, precious metal prices, and exchange rate’. In : *Energy Economics* 32(2), p. 351–362.
- SCHWARZ, Thomas V. et Andrew C. SZAKMARY (1994). ‘Price discovery in petroleum markets : Arbitrage, cointegration, and the time interval of analysis’. In : *Journal of Futures Markets* 14(2), p. 147–167.
- SILVENNOINEN, Annastiina et Susan THORP (2013). ‘Financialization, crisis and commodity correlation dynamics’. In : *Journal of International Financial Markets, Institutions and Money* 24, p. 42–65. ISSN : 1042-4431.
- STOLL, Hans R. et Robert E. WHALEY (1990). ‘The Dynamics of Stock Index and Stock Index Futures Returns’. In : *The Journal of Financial and Quantitative Analysis* 25(4), p. 441–468.
- TANG, Ke et Wei XIONG (2012). ‘Index Investing and the Financialization of Commodities’. In : *Financial Analysts Journal* 68(6), p. 54–74.
- WORKING, Holbrook (1949). ‘The Theory of Price of Storage’. In : *The American Economic Review* 39(6), p. 1254–1262.

Chapter 1

Are commodity markets still different from other asset classes?

Preliminary document, not to be distributed.

Abstract

In light of the financialization of commodity markets, we want to assess the extent to which commodity markets behave like financial markets by confronting data to the theories on commodities: the theory of storage, the theory of normal backwardation, but also the theory of expectations, and to the empirical literature. We thus study how spot returns behave, in particular when we account for the basis, which should theoretically provide information on the inventories on the market. This economic fundamental should be of utmost importance for commodities, due to a nonnegativity constraint on physical inventories, but not for financial markets.

Our database consists of 17 assets, divided into 4 sectors. We use daily data from 2000 to 2014 and our results are mostly consistent with the literature in terms of spot average return and standard deviation: the more difficult an asset is to store, the larger the standard

deviation and average return. Markets thus group into sectors accordingly. When we differentiate contango and backwardation, we indeed find larger standard deviation in backwardation (scarce inventories), except for the Brent crude oil and silver. For the former, contango has been a "normal" market condition only recently (after 2004) and may be due to its financialization... When we look into skewness and shocks, results are not always consistent with the theory: returns are not positively skewed in backwardation for all commodities. We also find that some commodity prices experience their minimum return on the day after the S&P500 does, same for the maximum...

Finally, we investigate the diversification potential of these commodities by looking at their correlation matrix and the presence of a potential common factor for all markets (with Principal Component Analysis) over different samples. We find sectorisation again before 2004, but afterwards, there is integration of commodities and finally, after 2008, integration with financial markets.

Contents

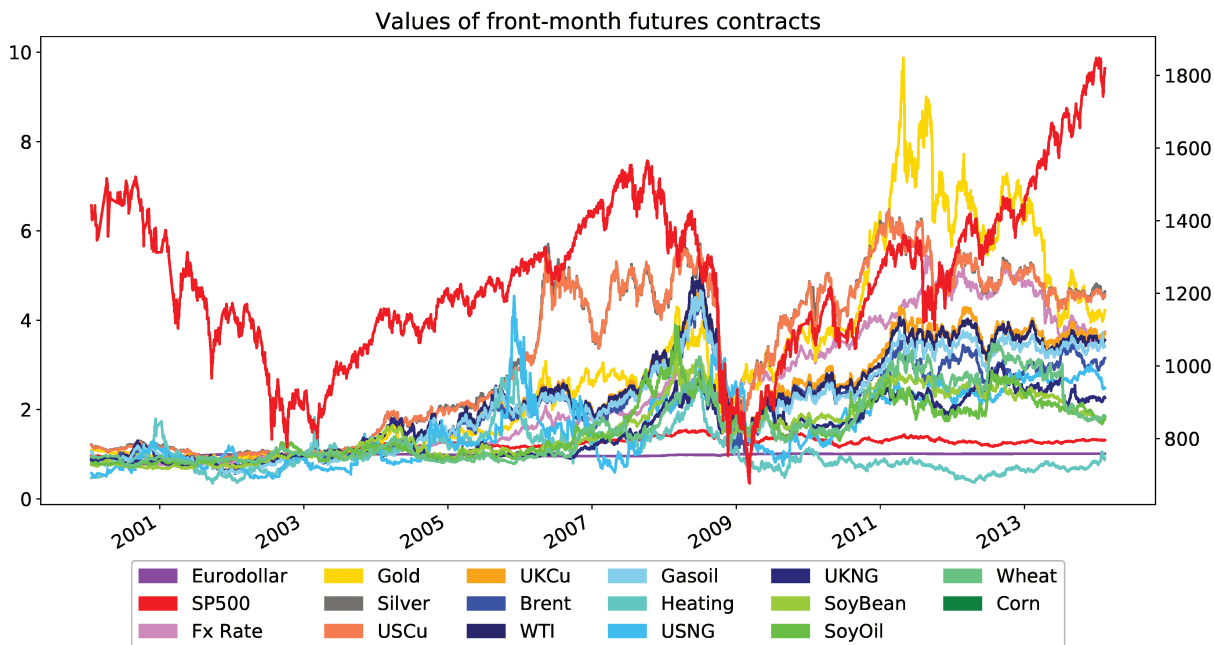
Introduction	31
1.1 Related literature	37
1.2 Data description	41
1.3 Comparing the different sectors in terms of unconditional moments and shocks	43
1.3.1 Investigating stylized facts about spot returns	44
1.3.2 Checking the similarity of the whole distributions	47
1.3.3 Investigating the tails of the unconditional return distributions	50
1.4 Incorporating the information provided by the basis	53
1.4.1 Studying spot returns by differentiating the state of inventories	54
1.4.2 What inventories can tell us about spot return shocks	59

1.5	The potential for diversification and systematic risk . . .	63
1.5.1	Diversification and correlations	63
1.5.2	The presence of common factors	65
	Conclusion	69
	References	71
	Appendices	73
1.A	Statistical measures	73
1.A.1	The Kolmogorov-Smirnov distance and test . . .	73
1.A.2	The Hellinger distance	74
1.B	Conditioning returns on the sign of the basis	75
1.B.1	Time series of spot returns and shocks conditional on the 3m-basis	75
1.B.2	Empirical distributions of returns conditional on the sign of the 3m-basis	79
1.B.3	Computing the moments of the conditional distributions of spot returns	85
1.B.4	Probabilities of shocks conditional on contango and backwardation	86
1.C	Correlation matrices by subsample	88
1.D	Additional results on PCA	91
1.D.1	PCA results for each subsample	91

Introduction

Since around 2004, commodity and financial markets have experienced an upward trend in their prices and rates, followed by a huge drop from late 2008 to early 2009, right after the default of Lehman Brothers. Figure 1.1 shows a plot of the price time series where all series have been normalised to 100 for 2003-01-02 except the S&P500 (in red). We see that almost all

Figure 1.1 – Time series of the spot prices/rates for all our markets over 2000-2014.



Prices and rates have been standardised to 100 for 2003-01-02, except for the S&P500 (right scale).

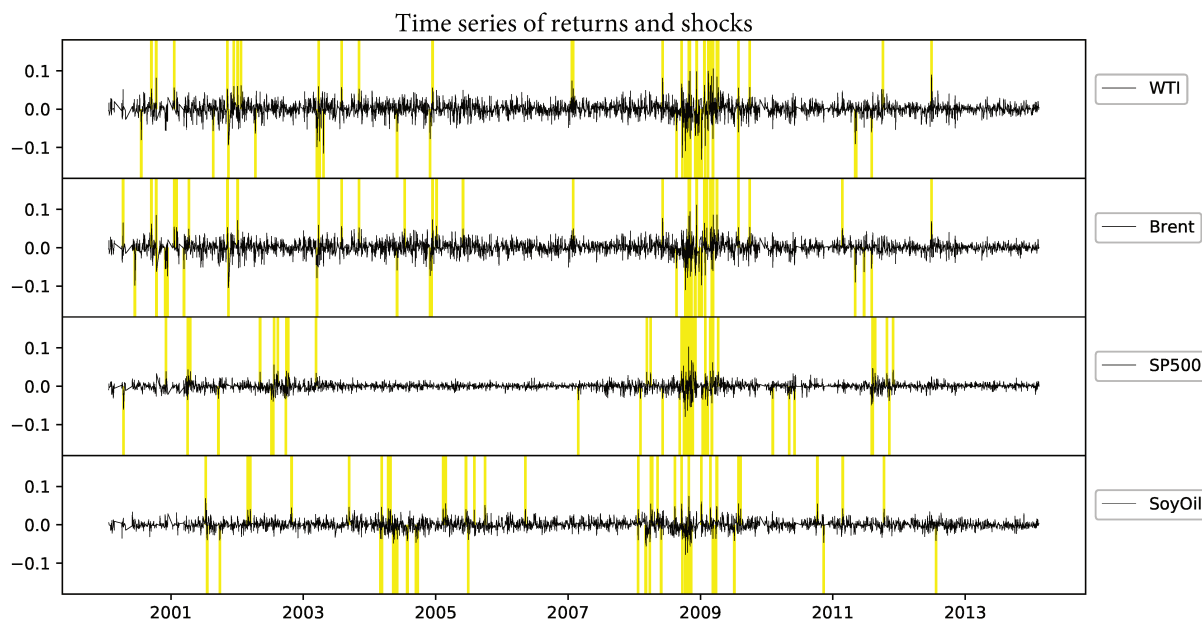
markets exhibit an upward trend around 2003-2007. This common evolution has raised concerns as to whether financial markets would be interfering too much with commodities, with fears that financial crises could affect commodities (and vice versa), hence systemic risk.

Indeed, Figure 1.2 shows the time series of returns for some markets. We highlighted the shocks in yellow: negative shocks (below the 1% quantile) are plotted as yellow bars below the horizontal axis and the positive shocks (above the 99% quantile) are above the horizontal axis. We see that many shocks clustered from late 2008 to early 2009 (we will come back to this figure in the last section, using the basis).

The year 2004 coincides with the creation of many investment vehicles that provide exposure to commodities¹. Large inflows of money followed,

¹One of the first commodity Exchange Traded Fund (ETF) was created in 2003 by Gold Bullion Securities (now ETF Securities). The famous ETF SPDR Gold Shares was

Figure 1.2 – Examples of time series of spot returns



The yellow bars represent "shocks" at 1% (positive if the bar is above 0, negative if below).

because, as Gorton and Rouwenhorst (2006) (GR) show for 1959-2004, their correlations with equities and bonds were historically low or even negative. Commodity futures provide an easy way of getting exposure, since only a deposit is required, instead of buying and storing the actual underlying commodities. In addition, these futures markets have turned to electronic trading, allowing agents to get in and out as fast as they need. Therefore, synthetic ETFs (which use futures contracts) have become more and more popular, creating more and more long positions in futures markets.

But this new participation of financial agents changed that low correlation. For example, Tang and Xiong (2012) find that, since 2004, the correlations between commodities have increased, even more so for those belonging to commodity futures indices. Additionally, Büyüksahin and Robe (2014) used a CFTC proprietary dataset to provide evidence that there is also an increasing correlation with equity markets, but not because of these indices. Rather, the increased participation of speculators (more specifically hedge

 listed on NYSE in November 2004.

funds) in both commodity and equity markets increases their correlation with financial markets in "normal" times. They find that this evidence does not hold when financial markets are in distress though, which raises questions, especially if agents can create an "unnatural connection" between the two sectors.

Before investigating that connection (which we leave to the following chapters), we first want to focus on price returns and shocks on individual markets. Indeed, shock propagation is an important part of systemic risk, which was neglected until the 2008 crisis. One should thus investigate the univariate properties of the markets too.

In particular, we confront data to the theory of storage (Working (1949)), since there is no physical aspect (deterioration, storage, etc.) for financial assets. More specifically, the most important physical feature that we will investigate is the nonnegativity constraint of the inventories. This theory provides testable implications in terms of the behaviour of spot and futures prices. For example, when inventories are scarce, the market should be in backwardation (negative basis, defined as futures price minus spot price, all divided by spot price) because the spot price should have increased (high demand consumes inventories). Low inventories are thus associated with backwardation: Gorton, Hayashi and Rouwenhorst (2013) (GHR) relate inventories to different derived measures, such as spot return volatility, prior returns or the basis. Low inventories also mean a higher volatility of spot prices since inventories are insufficient to be used as buffers for price fluctuation.

Futures markets have thus become a place for price discovery and the prices of those derivative contracts are said to reflect the information on the market, from the fundamentals of this market (as described in the previous paragraph) to expectations of agents about the future evolution of the price of the underlying asset.

This basis may be also related to the primary function of futures markets, the transfer of risk: the difference between the futures price and the expected future spot price should represent the risk premium paid to the agents

bearing the risk. This risk premium may have been a consistent source of return for speculators and financial agents, both in commodity and financial markets, and we will also investigate whether it has been rewarding in the past.

In addition to unconditional univariate return distributions, we will analyse return distributions when we condition them on the sign of the basis. We will thus assess whether the basis still matters for commodity markets or whether they have integrated with financial markets and display similar return distributions. Since commodity prices should exhibit positive skewness in backwardation (low inventories), we will also work on conditional probabilities of shocks and see whether this prediction holds and compare them with financial assets.

When we look at the unconditional spot returns, we find results consistent with financial intuition: commodities provide a larger average return than financial markets, but it comes with larger volatility too (consistent with the theory of storage). This volatility translates into shocks: commodity markets experience more shocks than financial markets. A concerning result is that the most extreme price shocks on the S&P500 futures seems to have triggered the most extreme price shocks (maximum and minimum return) on several commodities on the following day, all in October 2008, during the financial crisis...

When we condition on the inventories of commodities to investigate the spot returns, we find that the amplitude of this average is larger in backwardation for all our markets, though the sign may change. For example, energy markets (except the UK natural gas, which behaves the opposite way) have a negative average spot return in contango and positive in backwardation, just like the S&P500.

Consistent with the theory of storage, we also find that financial markets are less volatile both in contango and backwardation. Nevertheless, the only commodities that are more volatile in backwardation are the agricultural ones, copper and US natural gas. In this case too, volatility translates into

shocks: for these markets, there is generally a higher probability of shocks in backwardation. On the other hand, more shocks on crude oil prices occurred while in contango instead of backwardation, though contango seems to have become a "normal" situation since 2005, after the beginning of the financialization.

Finally, we analyse the potential for diversification of commodity futures by computing the correlations between the returns of all our markets. When we split our sample into three subsamples, we find results consistent with the literature: there is sectorization (thus diversification) before 2004, then commodity markets integrate, and finally, financial markets too, after 2008. We also check this by assessing whether there are common factors driving them all by conducting a Principal Component Analysis (PCA). With the same subsamples, we find that the first principal component emerges, first (before 2004), as an energy component, then (2004-2008) as a commodity component, and finally (after 2008) as a common component for all markets, confirming the progressive integration.

In a first section, we review some previous studies and theories related to the futures markets and the basis in commodity markets. In the second section, we briefly describe our database. In the third section, we compare financial and commodity markets by looking at the unconditional return distributions and shocks. In the fourth section, we do the same but by incorporating the basis in the analysis: we condition the return distributions and shocks on the sign of the basis and assess resemblance or not of the different sectors depending on market conditions (inventories). In the fifth section, we examine the diversification benefits of commodities over time and the presence of potential common factors that would provide evidence of integration and financialization.

1.1 Related literature

Futures markets as they exist now were created in 1864 by the Chicago Board of Trade, starting with commodity futures contracts. Since then, many researchers and practitioners have tried to understand them and several main theories have emerged.

Since the primary function of the futures contracts is to provide hedging against price fluctuation risk, Keynes (1930) developed the theory of normal backwardation, also called theory of hedging pressure. It states that a risk premium must be paid to those who bear the price risk: if the hedging pressure (the difference between long and short positions on a futures market) is net short, the premium should be paid to the long side. At that time, Hicks (1946) explains that producers are more likely to hedge, hence the reasoning of Keynes, that backwardation (the basis, equal to futures price minus spot price, is negative) was the normal situation: producers accept to sell in the future at a lower price to insure against price risk. The futures price should thus contain information on the risk premium. But since the beginning of the financialization, with the large inflows of investments wanting exposure to commodities, the "hedging pressure" may have shifted, making the long side (much) larger than the short side, hence shifting the risk premium. This may explain for example why the crude oil markets were in contango starting 2005.

But with competition and new knowledge, one may expect that this risk premium should decrease (in absolute value), potentially to zero. Indeed, some speculators (banks, funds, etc.), bearing risk or using these futures as an asset class, have also been involved in the physical market (buying storage, monitoring crops, etc.), hence hedging their risk from futures with physical commodities.

Such an evolution should lead to the futures price being equal to the expected future spot price, as in the theory of expectations. For example, Black (1976) considers that if producers hedge against price risk (and if

we assume that the Capital Asset Pricing Model applies for futures), the futures prices should exactly be equal to the expected future spot prices. Therefore, "futures prices provide wealth of valuable information for those who produce, store, and use commodities". Futures markets have thus become a place of price discovery for their underlying assets. Garbade and Silber (1983) indeed find that futures markets account for 75% of the price discovery for metals (copper, gold, silver) and agricultural markets (wheat, corn, oats, frozen orange juice), while Schwarz and Szakmary (1994) studied crude oil markets and found 65%. Even though financial assets are much easier to trade than commodities, Hasbrouck (2003) found that S&P500 E-mini futures contract accounts for about 90% of price discovery for the S&P500.

Building on this, especially on another statement of Black: "participants in this market can decide on the best times to plant, harvest, buy for storage, sell from storage, or process the commodity", futures prices could affect spot prices or at least provide information on the latter.

A simpler explanation of contango than that of shifting hedging pressure is provided by the theory of storage of Kaldor (1939), Working (1949), Brennan (1958) and Telser (1958). It states that the futures price today should reflect the cost of buying now and storing/holding the underlying asset until the maturity of the contract. If not, there could be (reverse) cash-and-carry arbitrage.

It was first designed for commodity markets, but can be applied to other assets too, as in Hasbrouck (2003) for equities. Nevertheless, this should only entail situations of contango for commodities, which does not always prevail on the market.

A way to explain situations where the basis is negative (backwardation) is to consider inventory as an American "option". Someone holding an asset can exercise the option to sell it only when the price is satisfactory (high), i.e. when the demand is consuming too much supply or inventories. This "option" has been given names such as "convenience yield" (Kaldor (1939)), which is thus seen as a return for holding the asset, but is difficult to assess,

since it is unobservable.

Backwardation has thus been associated with situations where inventories are low: the demand being too strong relative to supply and inventories, the spot price should increase (and its volatility too). On the contrary, spot prices should be less volatile in contango, since arbitrage is possible. This result can be tested empirically and should apply to markets where inventories exist and cannot be negative – commodities for example, but not financial markets where we can short-sell equities or borrow money.

These theories have thus been empirically tested, with the availability and quality issues that can arise with the corresponding data. For instance, Gorton, Hayashi and Rouwenhorst (2013) (GHR) provide a thorough study of commodity markets, both theoretically and empirically. To them, low inventory is associated with high basis, high spot and futures momentum. It should also come with high spot prices and high volatility (of the future spot price, due to a risk of stock-out). They also claim that futures prices should be high too but to a lesser extent, because of the expectation that inventories will replenish, thus still allowing backwardation. In the same vein, Fama and French (1988) think that the correlation between spot and futures prices should be perfect in the case of high inventory, while it should be more variable in the case of low inventory (arbitrage is not always possible). In their study, they explain the extreme backwardation of metals markets with business cycles: there is a large demand shock around business cycle peaks, while production adjusts slowly. As a matter of fact, it makes the spot price skyrocket, while the futures price may not change (as much).

In terms of the expectation theory, this would mean that the increase in spot price is expected to be temporary. This is consistent with Gorton, Hayashi and Rouwenhorst (2013) too, since a smaller variation of the futures price compared to that of the spot price would mean an imperfect correlation. On the other hand, this would mean that price changes when the market is in contango are expected to be permanent (perfect correlation of spot and futures prices).

Furthermore, there seems to be a linear relationship between average returns and the basis: the longer a market is in backwardation, the larger the average return (see Nash (2001)) and the larger the average backwardation, the larger the average return (see Till (2006) for reference by Nash & Strayer (2004)), which would come from roll returns.

Gorton, Hayashi and Rouwenhorst (2013) also examine the higher moments of the spot return distributions and find that they are highly volatile, positively skewed and have significant kurtosis, which they also found previously in Gorton and Rouwenhorst (2006) (GR). They thus find that spot prices should not experience downward spikes, but can have infrequent positive ones, like the ones from Fama and French (1988).

In addition, Gorton and Rouwenhorst (2006) build an equally-valued (1\$ for each component) index of commodity futures and find that it performs better than the S&P500: it has about the same (positive) average return but with lower volatility on the 1959-2004 period. This paper may have contributed to the success of commodity-related investments from around 2004, which may have triggered the "financialization" of commodity markets. Subsequently, the investment in commodity indices has increased the correlation between commodities inside but also outside those indices, as found by Tang and Xiong (2012). Finally, Büyüksahin and Robe (2014) find that speculators (hedge funds) create correlation between commodities and financial markets (after 2008) by holding positions on both, and hence make this correlation disappear as they pull out. Cheng and Xiong (2014) provide a great literature review on the subject.

We will instead focus on empirically assessing whether commodity markets and financial markets have similarities in terms of their univariate returns distributions and shocks. Our aim is to investigate whether market integration has been too strong or whether commodities still represent a different (physical) asset class – and thus rely on the basis and the above literature on commodities. In light of the 2008 crisis and the evolution of commodity prices at that time, we indeed fear that the connection between the two sectors may be too strong and without fundamentals justifying it,

which raises concerns about the propagation of shocks and crises, i.e. systemic risk.

1.2 Data description

Our database consists of daily observations of futures prices extracted from Datastream. It goes from 2000/01/21 to 2014/02/12. It (currently) consists of 17 markets, classified into four (sub-)sectors:

- 4 for the agricultural sector: soybean, soybean oil, corn and wheat;
- 6 for the energy sector: Brent, WTI, heating oil, gasoil, US natural gas and UK natural gas;
- 4 for the metals sector: gold, silver, US copper and UK copper;
- 3 for the financial sector: 3-month Eurodollar interest rate, S&P500 and USD/EUR exchange rate.

Table 1.1 presents the relevant information on our database, where we grouped our markets by sub-sectors. In the Exchange column, we display the exchange and region where the contracts are quoted. The Last maturity column provides the longest maturity (in months) contract that we kept for our database. Finally, the "# of maturities" represent the number of contracts that we have for each underlying asset, which may be different from the longest maturity, because not every month is quoted. We need to be careful when analysing the results based on the longest contract, due to the large differences of maturity of those contracts: for example, the eurodollar has 40 contracts (and a longest maturity of 114 months), versus the 3 contracts of the S&P500 (up to 6 months only). We will thus focus on the 3-month contracts in the body of the paper, but also because these contracts and the longest ones may involve different dynamics: the Samuelson effect (the volatility of the futures prices decreases with the maturity)

Table 1.1 – Futures contracts selected, data information

Underlying asset	Exchange	Last maturity (in months)	# of maturities
Brent	ICE-EU	30	16
WTI	CME-US	72	32
Gasoil	ICE-US	12	12
Heating oil (HO)	CME-US	12	12
US Natural Gas (USNG)	CME-US	36	36
UK Natural Gas (UKNG)	ICE-EU	9	9
Eurodollar	CME-US	114	40
S&P500	CME-US	6	3
FX rate (USD/EUR)	CME-US	12	4
Gold	CME-US	54	17
Silver	CME-US	60	19
US Copper (USCu)	CME-US	24	20
UK Copper (UKCu)	LME-UK	24	24
Soybean	CME-US	14	7
Soy oil	CME-US	15	9
Wheat	CME-US	12	5
Corn	CME-US	15	6

CME stands for Chicago Mercantile Exchange, LME for London Metals Exchange and ICE for InterContinental Exchange. US and EU refer to the location where the contracts are listed.

and the difference of anticipations. We do provide the results for both maturities, taking advantage of the term structure for each market.

We faced several difficulties with data in our study. First, we had to merge all those time series in order to provide coherent comparisons between markets/maturities. We thus had to remove the dates where at least one observation was missing, which may lead to irregular time intervals between observations and hence change the results. We did transform them into daily returns (i.e. dividing the returns by the number of days between the observations). In addition, for some markets, due to Datastream reporting or no actual

change (for instance on the Eurodollar market, the short-term rate may not change because the repo rate does not change), we find the same quote for successive observations. We did remove the duplicates (same observations for the whole term structure at two successive lines or when rolling) though to mitigate this issue.

Our initial database had more than 935,000 prices consisting of 3450 observations for each of our 271 time series, but this number decreased to 2949 (2889 when we use both the returns and the 3-month basis) due to this treatment.

We analyse several variables from market data: "spot" returns and the basis. We use the front-month futures returns to approximate spot returns. For the basis, we compute it in two ways: a short-term basis (using the 3-month contract, labelled 3m-basis) and a long-term basis (using the last contract we have for each market, refer to Table 1.1, labelled Last-basis). So when we write "3m-contango" (3m-backwardation), it means that the 3-month basis (3m-basis) is positive (negative) and likewise for Last-contango (Last-backwardation).

1.3 Comparing the different sectors in terms of unconditional moments and shocks

Before using the basis, we want to confront our data to previous studies on commodities and the stylized facts that they uncovered, namely a positive average return, relatively high volatility, positive skewness and high kurtosis. We will thus first look at the moments of the unconditional return distributions and compare commodities to financial markets. We will also compare the overall distributions using Kolmogorov-Smirnov and Hellinger distances. Only after that will we turn to our focus: the analysis of shocks on our futures markets.

1.3.1 Investigating stylized facts about spot returns

In terms of unconditional moments of spot returns, Gorton and Rouwenhorst (2006) compare an equally valued (1\$ for each component) index of commodity futures to the S&P500 and the Ibbotson corporate bond total return index over the 1959-2004 period. They find that their index has about the same average return as the S&P500 (larger than the bonds index), but with lower standard deviation and positive skewness, compared to the negative skewness of the equity index. Overall, their commodity index provides about the same average return but with a lower downward risk. They also provide the moments for individual commodities: the volatility of the index value is lower than all the individual standard deviation; there is diversification among commodities, as evidenced by the average correlations with other commodities that they present.

In their more recent paper, Gorton, Hayashi and Rouwenhorst (2013) examine individual commodities and relate their risk premium to inventories and to other price measures, such as the basis, past returns, etc. They confirm their findings on standard deviation, positive skewness and significant kurtosis.

Table 1.2 presents the moments of the unconditional daily spot return distributions. We did not annualize them as in other studies.

Average spot return If we set apart the Eurodollar (because it moves mostly with central bank decisions), we first see that only the wheat has a negative average spot return, but very close to 0. Also, the way we compute our spot returns (log-return of the front-month futures price) *may* be assimilated to the "excess future return" of Gorton, Hayashi and Rouwen-

Table 1.2 – Moments of the spot return distributions

	Mean	SD	Skew	Kurt
Brent	0.069%	1.840%	-0.23	4.01
WTI	0.049%	1.988%	-0.33	4.51
Gasoil	0.039%	1.748%	-0.46	5.32
Heating oil	0.083%	1.886%	0.01	2.37
USNG	0.032%	2.949%	1.06	7.94
UKNG	0.066%	3.411%	4.16	43.20
Eurodollar	-0.119%	1.908%	1.40	37.40
FX rate	0.009%	0.555%	-0.14	1.76
S&P500	0.006%	1.084%	0.03	8.47
Gold	0.033%	1.048%	-0.35	5.20
Silver	0.041%	1.817%	-1.26	10.87
US Copper	0.042%	1.650%	-0.21	6.59
UK Copper	0.043%	1.601%	-0.04	7.44
Soybean	0.037%	1.516%	-1.24	44.07
Soy oil	0.015%	1.308%	0.08	3.11
Wheat	-0.002%	1.758%	0.08	3.31
Corn	0.014%	1.607%	-0.15	3.90

"SD" stands for "Standard Deviation", "Skew" stands for "Skewness" and "Kurt" stands for "Kurtosis" (computed here as Fisher's: "normal" kurtosis – 3).

horst (2013)², but they work on a monthly basis. They compute it as a proxy for the risk premium on the markets and thus find, as we do, that most futures offer a positive average "excess" return, meaning a positive risk premium.

If we exclude the eurodollar and wheat, which have negative average spot return, we find that financial markets have a lower risk premium than commodity markets. This result is slightly different from that of Gorton and Rouwenhorst (2006), since they find that their equally valued index of commodity futures (\$ 1 for each futures in the index) has about the same average

²Their "excess future return" is computed as follows: $\frac{F_{t+1,T} - F_{t,T}}{F_{t,T}}$, where $F_{t,T}$ is the futures price at the end of month t on the nearest contract whose expiration date T is after the month $t + 1$, and $F_{t+1,T}$ is the price of the same contract at the end of month $t + 1$. In our daily framework, T would be the first maturity and t and $t + 1$ would be the daily observations (before the closest contract expires).

return as the equity index.

Volatility and the theory of storage Consistent with both GR and GHR, we find that commodity spot returns have a relatively high standard deviation. Furthermore, consistent with the theory of storage, the more difficult an asset is to store, the more volatile its price is: financial markets are less volatile than commodities. An exception in the world of commodities is gold, which has low volatility (due to easy storage, no seasonality, etc.), but then silver should behave the same (though to a lesser extent). Both are basically always in contango (no shortage of inventories). This may be the reason why Gorton, Hayashi and Rouwenhorst (2013) considered both of these markets as financial assets and excluded them from their study.

As for the higher moments, all unconditional return distributions exhibit (very) high kurtosis, consistent with both GR and GHR. In terms of skewness, Deaton and Laroque (1992) find positive (and significant) skewness for commodity markets, just like GR and GHR, and in GR, negative skewness for equities. In our case, however, it is not as obvious. Equity markets are known for their negative skewness, but we do not find it here, though it is only slightly positive. For commodity markets, we get different results for each sub-sector both in terms of signs and of amplitude, but crude oils have negative skewness, just like metals, while natural gases have quite positive skewness.

We will investigate these results on the tails (shocks) into more details in a following section.

To summarise, consistent with the theory of storage, we find that overall commodity futures markets provide a positive and larger average return than financial futures (while GR find a similar average). This ordering holds if we look at the standard deviation too, thus the higher average return, if we consider, as financial agents, that it rewards risk. This is also consistent with the theory of hedging pressure, where the risk premium would be

larger for commodities than for financial assets. In terms of skewness, we cannot find general results, though we note that crude oils have negative skewness, just like the S&P500 is supposed to have (it is slightly positive here).

The potential consequences of the physical constraints on prices may be subtly differentiable in the moments. We then compare the distributions with distance measures, to assess how similar they are and account for the whole of the distribution.

1.3.2 Checking the similarity of the whole distributions

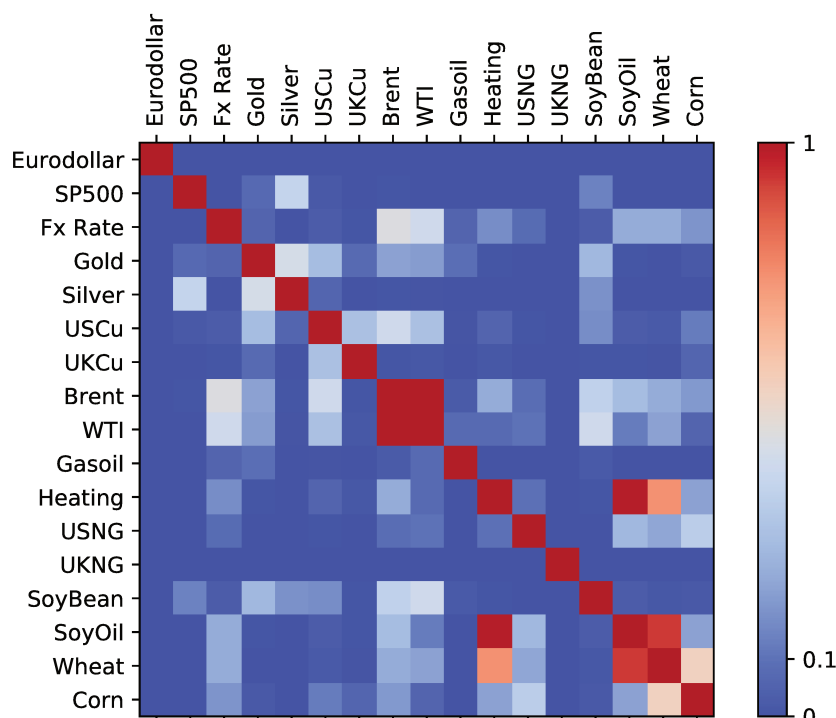
We further study the resemblance of the probability distributions of the returns by computing two statistical distances between pairs of random variables (see Appendix 1.A for some details on the measures). It should account for the shape of the distributions more precisely and better highlight the differences.

The Kolmogorov-Smirnov (KS) statistic is the largest difference between the cumulative distributions of two samples and this difference can then be statistically tested. We thus compute the corresponding p-values and plot them.

Figure 1.3 is a heatmap of the pairwise KS test p-values, with the lowest p-values (close to 0) in blue and the largest (close to 1) in red (Table 1.6 in Appendix gives the values). Low p-values mean that we can reject the null hypothesis that the samples are drawn from the same distribution.

We first note that the Eurodollar and UK natural gas price return distributions are statistically different from all others. Hence, when referring to energy or commodity markets, we will exclude the UK natural gas. On the contrary, the Brent and WTI return distributions seem indiscernible. Some small clusters seem to emerge from this heatmap: an agricultural cluster (except the soybean market), an energy cluster and a metals cluster. A larger cluster of commodities can also be evidenced, especially the markets

Figure 1.3 – Heatmap of pairwise Kolmogorov-Smirnov two-sample test p-values



related to Brent and WTI (agricultural markets, gold, silver, US copper and other energy markets). Though the UK copper is not directly highly related to those crude oil markets, its return distribution is expectedly close to that of the US copper, hence indirectly looks like that of other commodity markets. The USD/EUR exchange rate also seems to be quite close to commodity markets, much more than to other financial markets. This is consistent with the view that imports and exports of commodities affect the exchange rate and vice versa and that crude oil prices are highly related to the exchange rate fluctuations, as shown by Zhang et al. (2008).

Several odd results need to be investigated in the future: the soybean market is more related to crude oil and metals markets (and the S&P500); heating oil and US natural gas are more related to agricultural markets.

Overall, it seems that commodity prices in general do not evolve the same

way as financial markets. The exception is the USD/EUR exchange rate, which is naturally related to imports and exports of commodities.

We then turn to the Hellinger distance. It basically differs from the KS statistic in that it sums the differences (actually the squared differences of square roots of distribution functions) over the space of the random variables (refer to Appendix 1.A for more details). In addition, the returns are not normalised here.

Figure 1.4 – Heatmap of pairwise Hellinger distances between Gaussian distributions

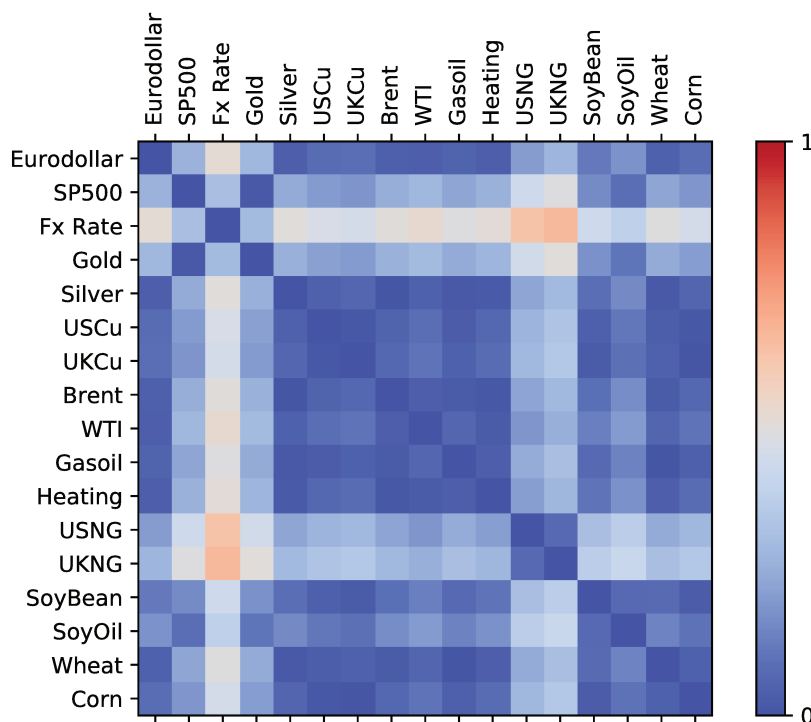


Figure 1.4 shows the heatmap of the pairwise Hellinger distances between our return distributions, with shorter distances (smaller differences) in blue. This measure gives slightly different results. We do find again that commodities in general form a cluster (from the dark blue diagonal blocks and

inter-sector blocks), but this time, the USD/EUR exchange rate does not resemble those commodities. Instead, the eurodollar market does, being very close to energy and metals markets in particular. It means that the largest difference in cumulative density function between eurodollar and commodities is not as important as the other smaller differences, from around the peak to the tails. The exchange rate is instead quite close to only gold and the S&P500. These two markets are particularly close to each other, which is a little puzzling: they should be anti-correlated, so potentially mirror images of each other. In addition, the skewness of one should translate into an opposite skewness for the other, creating some distance between their distributions. Another pair is standing out: the natural gas markets are only related to each other, not so much to the others, and thus have a particular behaviour in terms of price variations. This is consistent with their more difficult and specific storage.

Overall, be it with one or the other distance measure used, we find that some financial markets do have similar return distributions with commodities. Commodity prices do seem to exhibit a specific behaviour. their physical constraints (e.g. nonnegativity constraint) potentially translating into return distributions, consistent with theory.

We will then investigate the tails (high kurtosis and skewness found before) into more details as we want to assess the similarity between sectors in terms of shocks too.

1.3.3 Investigating the tails of the unconditional return distributions

The skewness of the return distributions for our markets is not always large or even positive, so the prediction that commodity spot returns should exhibit infrequent upward spikes but no downward spikes (Gorton, Hayashi and Rouwenhorst (2013)) is not obvious here. We thus take a look at some

quantiles of the spot returns. Table 1.3 provides the 1% and 5% positive and negative returns quantiles (99% and 95% for positive ones), with the minimum and maximum returns and the date when they occurred.

Table 1.3 – Spot returns distribution quantiles

	1%	5%	95%	99%	Min	Date Min	Max	Date Max
Brent	-5.30%	-2.86%	2.99%	4.85%	-10.95%	2008-10-10	11.13%	2008-12-11
WTI	-5.58%	-3.18%	3.10%	5.15%	-13.07%	2009-01-07	10.53%	2009-03-12
Gasoil	-4.77%	-2.85%	2.77%	4.77%	-15.07%	2003-03-13	8.50%	2009-04-02
Heating	-5.07%	-2.95%	3.19%	5.48%	-9.02%	2008-10-10	8.67%	2003-03-27
USNG	-7.57%	-4.49%	4.60%	8.90%	-14.47%	2000-12-12	26.77%	2009-09-29
UKNG	-7.46%	-3.69%	3.82%	11.02%	-26.28%	2006-02-28	47.77%	2006-09-29
Eurodollar	-6.10%	-2.44%	1.89%	6.34%	-14.31%	2010-05-27	30.19%	2010-05-06
SP500	-3.13%	-1.76%	1.54%	3.29%	-7.92%	2008-10-09	10.25%	2008-10-28
Fx Rate	-1.54%	-0.90%	0.92%	1.42%	-2.47%	2008-09-30	2.52%	2008-11-04
Gold	-3.14%	-1.70%	1.61%	2.59%	-7.58%	2006-06-13	7.77%	2000-02-04
Silver	-5.68%	-2.92%	2.76%	4.71%	-19.55%	2011-09-23	10.93%	2008-10-29
USCu	-5.21%	-2.45%	2.57%	4.80%	-11.69%	2008-10-10	11.77%	2008-10-29
UKCu	-4.54%	-2.38%	2.52%	4.54%	-10.32%	2008-10-30	11.92%	2008-10-29
SoyBean	-4.21%	-2.24%	2.21%	3.74%	-26.46%	2004-07-15	20.32%	2008-09-12
SoyOil	-3.57%	-2.13%	2.19%	3.78%	-7.75%	2008-10-10	7.40%	2008-10-29
Wheat	-4.79%	-2.70%	2.83%	4.69%	-9.22%	2011-06-30	8.79%	2008-10-29
Corn	-4.32%	-2.45%	2.60%	4.67%	-10.41%	2011-06-30	8.07%	2006-09-15

The dates of minimum and maximum returns of the S&P500 are displayed in black bold. The dates of minimum and maximum returns for other markets on the following day are in red bold.

The distributions do not seem very asymmetric in terms of "shocks", except for the natural gases at 1%. For a same shock in absolute value, natural gas prices would experience more positive shocks than negative ones, which would explain their high positive skewness and consistency with the theory. The storage of natural gas is indeed relatively difficult, compared to other kinds of commodities (needs a spherical storage otherwise pressure will build up and the storage facility could explode), which seems to translate into larger shocks (and larger volatility).

On the contrary, precious metals and soybean are much easier to store. They experience slightly more negative shocks which may partly explain

their negative skewness. Precious metals should exhibit positive skewness as they are used as a defensive asset and because equities should have negative skewness. Strangely, both these anticipations do not appear when looking at the quantiles (S&P500 does display negative "skewness" for the 5% quantile).

Overall, energy markets and the Eurodollar market seem to experience many "extreme" shocks (more than 5% in absolute value, which is a daily return). And if we exclude the Eurodollar market, we can again find the same ranking as for the standard deviation – financial markets and gold have lower quantiles in absolute value.

If we look more closely at the date at which the minimum and maximum returns occur, we find 19 out of 34 dates in late 2008 or early 2009 (around and after the Lehman bankruptcy and the other important events)³. Out of these 19, only 4 are from financial markets (exchange rate and S&P500). Why would the others come from commodity markets? Fama and French (1988) explain that spot prices increase significantly around business cycles peaks. This could explain why commodities related to the economic activity, such as energy markets and industrial metals (coppers) could experience positive price shocks around crises. We find consistent results overall, with 10 out of the remaining 15 largest shocks for these markets.

Several other results are worth noting. The S&P500 experience its minimum on October 9, 2008. We see that the Brent, heating oil, US copper and soy oil – all sectors – have their minimum the following day, on October 10, 2008. Additionally, the S&P500 has its largest positive price shock on October 28, 2008 and coppers, silver, soy oil and wheat have theirs the following day, on October 29, 2008. These observations raise even more concerns about the financialization of commodity markets, especially since they are not completely intuitive: silver being a precious metal, it should be anti-correlated to equity markets (though here, there is a lag); there is no

³See Chapter 2 for a summary timeline around Lehman Brothers' failure or Brunnermeier (2009) for some details

intuitive reason for agricultural markets to be affected that much by equity markets either.

To summarise, we do find that commodity futures offer a larger "risk premium" than financial markets, but with a larger volatility. In addition, this volatility translates into larger shocks for commodities too. We also note some hints of financialization from these shocks: the S&P500 futures minimum and maximum returns seem to trigger minimum and maximum for some commodities, namely Brent, heating oil, copper, silver, soy oil and wheat markets.

There is no intuitive reason for agricultural markets being led by financial markets. We will thus explore the reasons why it happened, especially by looking at what the basis can tell us about the inventories and whether there is an economic explanation for this result.

1.4 Incorporating the information provided by the basis

As we mentioned earlier, for commodities, the basis contains information from two sources. On the one hand, economic fundamentals (inventories) are reflected as storage costs and provide an arbitrage relationship between the futures price and the spot price. On the other hand, the basis also contains information about the risk premium and expectations. Fama and French (1987) and Gorton, Hayashi and Rouwenhorst (2013) provide evidence for these two components of the basis. For financial markets, though these two components exist too, the first one is not that prominent in terms of volatility of spot returns since there is no nonnegativity constraint here. Instead, it could mean that the agents may expect the value to in-

crease when the market is in contango (e.g. economic growth if we take the S&P500) and to decrease when it is in backwardation (recessions).

We compute the basis as follows:

$$B(t, T) = \frac{1}{T-t} \frac{F_{t,T} - S_t}{S_t} \quad (1.1)$$

where $F_{t,T}$ stands for the futures price at date t for maturity T (which is either 3 months for 3m-basis, or the longest maturity available for Last-basis) and S_t for the "spot" price (that we approximate by the closest futures price). We divide by the residual maturity to allow comparison between markets, which have different last maturities.

We thus continue confronting our data to theories and stylized facts that differentiate the two situations: contango and backwardation.

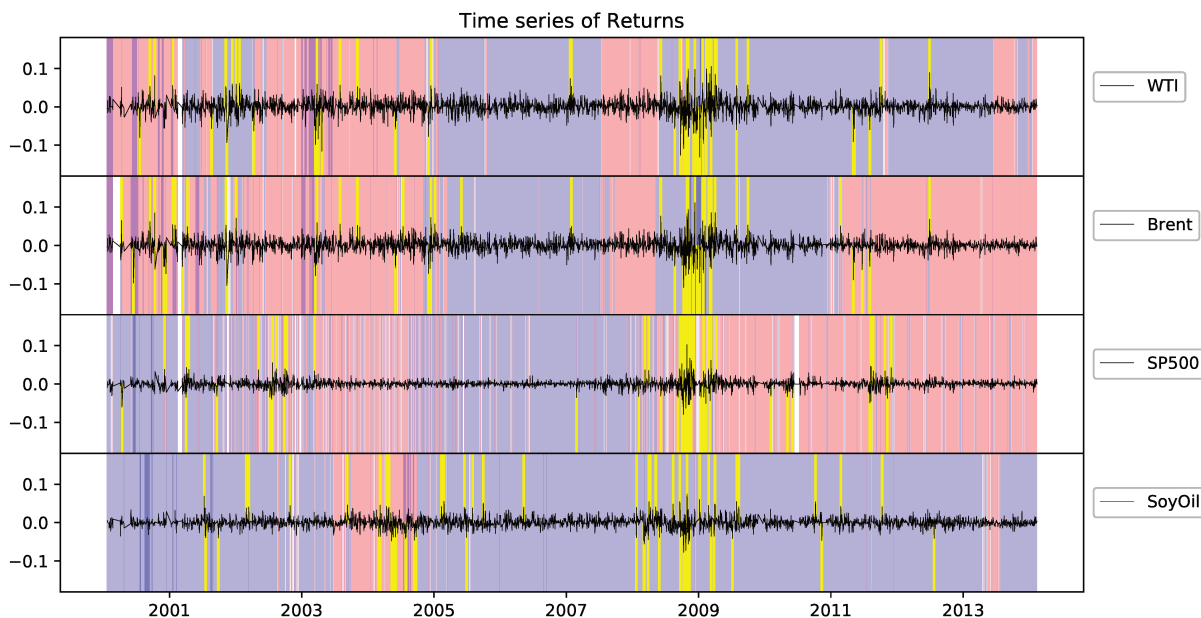
1.4.1 Studying spot returns by differentiating the state of inventories

Spot prices should be more volatile when inventories are not sufficient, since they cannot dampen price fluctuations. This low inventory situation should be associated with backwardation, when the spot price exceeds the futures price. The first step is thus to see how often a market is in contango or backwardation.

We first plot in Figure 1.5 the time series of returns for four chosen markets again (Brent, WTI, S&P500 and soy oil), but this time adding the information on the basis: when the background is blue, it means that the market is in 3m-contango, while when it is red, it is in 3m-backwardation. The others are available in Figure 1.9.

We will go into more details below, but we can first observe that the WTI is more often in contango, while the Brent is more often in backwardation. The clear distinction occurs in 2011, when storage issues arose in

Figure 1.5 – Examples of time series of spot returns

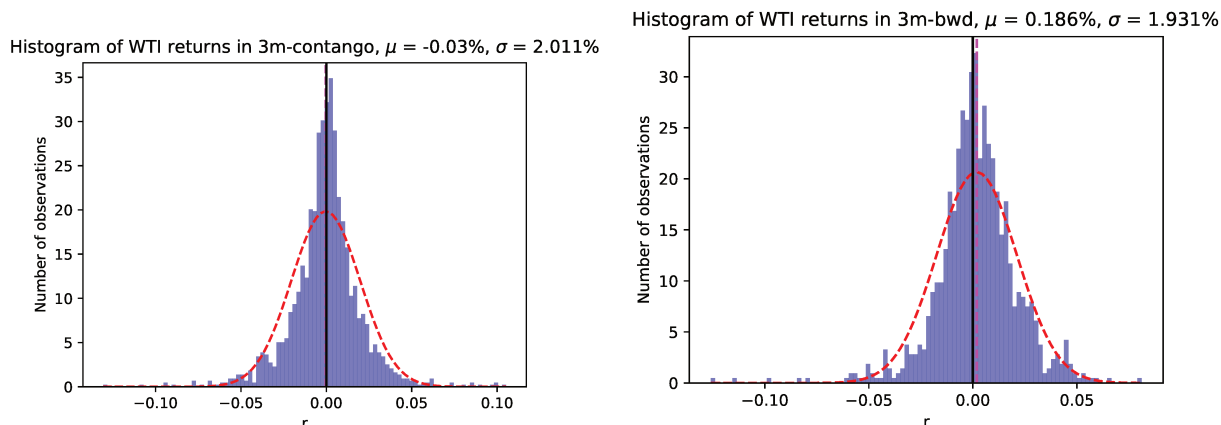


The yellow bars represent "shocks" at 1% (positive if the bar is above 0, negative if below) and the background is red if the market is in 3m-backwardation and blue if it is in 3m-contango. If it is darker, it means that the basis has an amplitude of more than half its maximum over the whole sample.

Cushing, OK. They were both mostly in backwardation (historical "normal" market condition) until 2005, when contango became (temporarily for the Brent) the norm. Nevertheless, their prices have very similar behaviour, especially in terms of shocks, displayed in yellow. In particular, there is a huge cluster of shocks in late 2008 and early 2009 (also for the other two markets), but apparently not really in backwardation. Gorton, Hayashi and Rouwenhorst (2013) and the nonnegativity constraint on commodities predict that commodity prices should experience infrequent upward spikes when inventories are low (backwardation) and no negative spike. We will thus compute the moments of the conditional return distributions and check this prediction quantitatively.

The conditional spot return distributions generally look like the ones in Figure 1.6: they are mostly similar in both contango and backwardation

Figure 1.6 – Examples of conditional return distributions



The pink dashed line represents the mean of the return distributions and the red dashed curve represents a Gaussian density function that has the same mean μ and volatility σ .

(except for markets that are rarely in backwardation). The other distributions are available in Figure 1.10. Table 1.4 presents various statistics of these distributions, market by market. We first computed the probability of being in 3-month contango, $P(C)$, then computed the moments of the return distributions conditional on this contango and on the complementary event, backwardation.

Frequency of contango and backwardation Looking at Table 1.4, we can first see that almost all markets are most of the time (more than 50%) in contango, with the notable exceptions of the Brent and S&P500 (the exchange rate being very close). For the latter, we see in Figure 1.5 that backwardation became the most frequent in 2003-2004 and starting late 2008 (after the events triggered by Lehman Brothers' default). In addition, the variability of the sign is also striking. For the former, backwardation was the "normal regime" on the market, until 2005. The basis was negative again around late 2007-early 2008 and after 2011, which may indicate a return to normal conditions. Since 2011, though, the UK and US markets began diverging: the American market was saturated in crude oil because of the development of shale oil. This huge supply created a storage issue (the

Table 1.4 – Moments of the empirical returns distributions conditional on the sign of the 3m-basis

		Conditional on 3m-contango (= C)				Conditional on 3m-backwardation			
	P(C)	Mean	SD	Skew	Kurt	Mean	SD	Skew	Kurt
Brent	44.26%	-0.01%**	2.05%***	-0.22	3.86	0.13%	1.64%	-0.21	3.70
WTI	64.29%	-0.03%***	2.01%	-0.15	4.61	0.19%	1.93%	-0.71	4.87
Gasoil	54.05%	-0.03%**	1.73%	-0.08	3.37	0.12%	1.77%	-0.91	7.84
Heating	70.97%	0.02%***	1.85%	0.01	2.49	0.22%	1.95%**	-0.03	2.35
USNG	87.82%	-0.06%**	2.83%	1.28	10.00	0.38%	3.45%***	0.02	2.31
UKNG	64.05%	0.20%***	3.65%***	5.34	47.79	-0.15%	2.92%	0.24	16.61
Eurodollar	67.96%	-0.04%***	1.87%	3.12	50.79	-0.26%	2.02%***	-1.47	14.53
SP500	49.27%	-0.06%***	1.07%	-0.66	7.36	0.07%	1.10%	0.65	9.48
Fx Rate	50.03%	0.00%	0.50%	0.05	1.91	0.02%	0.61%***	-0.25	1.51
Gold	97.72%	0.03%	1.05%	-0.33	5.34	0.12%	1.02%	-0.47	2.09
Silver	98.62%	0.04%	1.82%**	-1.27	11.04	0.40%	1.36%	-0.12	-0.05
USCu	68.75%	0.03%	1.56%	0.10	5.78	0.08%	1.79%***	-0.55	7.43
UKCu	59.86%	0.00%*	1.53%	0.28	7.77	0.11%	1.70%***	-0.39	7.22
SoyBean	61.66%	0.02%	1.36%	-0.24	3.72	0.06%	1.74%***	-2.03	65.00
SoyOil	88.96%	0.02%	1.27%	0.11	3.52	-0.05%	1.50%***	-0.04	1.11
Wheat	96.23%	-0.01%	1.75%	0.09	3.13	0.19%	2.01%**	-0.05	6.24
Corn	87.79%	0.01%	1.60%	-0.03	3.88	0.04%	1.71%*	-0.91	4.24

"SD" stands for "Standard Deviation", "Skew" stands for "Skewness" and "Kurt" stands for "Kurtosis" (computed here as Fisher's: "normal" kurtosis – 3).

We tested for the difference in means between the two samples and for the ratio of their variances. Stars indicate the level of statistical significance of the difference at 1% (***), 5% (**), 10% (*). For means, either column is used, for variance (though standard deviation is displayed), stars are on the sample with the larger variance.

Cushing cushion) and changed the behaviour of the front month futures. This explains the difference in frequency of contango for the two crude oils.

Average spot return We do not find any clear distinction between finance and commodities in each conditioning. Nonetheless, we see that for energy markets and financial markets (except exchange rate), the average spot returns are statistically different when conditioning on contango and when conditioning on backwardation. It is positive and much larger in backwardation than in contango for most of them (all except UK natural gas

and eurodollar). This result is consistent with the theory of hedging pressure: backwardation is associated with a situation where the risk premium is positive and contango with a negative risk premium. Moreover, the positive average in backwardation is also in line with the Nash & Strayer (2004) result that backwardation provides return on average and with the theory of storage: backwardation corresponds to scarce inventories, hence a demand that makes the spot price increase on average. On the contrary, it seems to contradict the theory of expectations, since backwardation should correspond to an anticipation of decreasing price and contango with an increasing price. We have to keep in mind that here, the average returns are not those until the maturity of the contract (which would be the expectation), only the daily return on the day of observation of the basis.

Volatility and the theory of storage According to the theory of storage, when the market is in backwardation, inventories should be insufficient, hence, the spot price should be more volatile. It should apply to all commodity markets, since storage is a concern, but four energy markets (both crude oils, UK natural gas and gasoil) are not consistent with this prediction. When we look at Figure 1.5, it seems that most of the "shocks" for the crude oils occurred during contango (the blue background) – which became the "normal" condition after the beginning of the financialization–, which may explain the higher standard deviation in contango.

The prediction applies to markets from all other sectors, but equity and precious metals. Why would it work for interest and exchange rates markets? Backwardation for the eurodollar interest rate means that money can be invested or borrowed at a lower rate in 3 months than now. It means that markets may have expected the rates to drop, for example if there was a central bank announcement. An interest rate decrease is generally meant to boost the economy, e.g. in times of crisis, which is consistent with more variability. As for the exchange rate, backwardation means that the market anticipates that the dollar will become stronger, relative to the euro, just reflecting the difference in growth or recession.

Gold and silver being precious metals, they are rarely in backwardation (respectively 2.28% and 1.38% of times); contango is their "normal regime". When they are though, it should be a concern as we would expect the same results as Fama and French (1988): there will be a huge demand for these whenever a crisis occurs, leading to backwardation. Their study looks at business cycles and how their peaks (potentially right before crises) generate "extreme backwardation", due to a peak in demand and a slowly adapting supply, hence a skyrocketing spot price and a potentially stable futures price (expecting inventories to replenish). We will thus examine these potential spikes and check whether this result holds.

1.4.2 What inventories can tell us about spot return shocks

More particularly, we want to assess whether the basis can provide information on the occurrence of shocks on spot returns, for instance if a specific configuration of the market could help anticipate and prevent shocks.

Skewness and kurtosis Looking at the kurtosis of the conditional spot returns, all distributions are indeed leptokurtic in both conditionings. These fat tails are overall quite different across all our markets. We do find that they are almost all negatively skewed in backwardation, though with the exception of S&P500 and natural gases. This contradicts theories and previous studies, since they predict that prices should experience positive shocks in backwardation, due to low inventories and their nonnegativity constraint. But they do not resemble equity markets either. Furthermore, this similarity between equities and natural gases disappears in contango, since we find the stylized fact that the S&P500 has negative skewness, while natural gases still have a positive one (even higher).

Overall, it seems like our commodity markets are different from our equity

futures in terms of shocks, which is reassuring. Nevertheless, interest and exchange rate markets seems to have similarities with some commodities: their skewness is positive in contango and negative in backwardation, just like the heating oil, coppers, soy oil and wheat markets.

We still need to investigate these tails into more details.

Probabilities of shocks To this aim, we compute the probabilities of shocks conditional on the sign of the basis (see Table 1.8 in appendix for the detailed results). We then compare the conditional probabilities and classify the markets into three categories for each kind of shocks (positive or negative) in Table 1.5a:

- 3m-contango (or Last-contango): for the case where conditioning on contango gives a higher probability of shock.
- 3m-backwardation (or Last-backwardation): for the case where conditioning on backwardation gives a higher probability of shock.
- Unclear: when we condition on the sign of the basis, we do not find consistent results.

The predictions about shocks from literature should not apply to financial markets, since they are not subject to a nonnegative inventory constraint. But if we instead consider that the futures prices (or interest and exchange rates) represent the market anticipations about the future spot price (or rate), then backwardation would mean that the spot price (or rate) is expected to decrease. For equity markets, this translates as an expectation of "recession", which would be a more turbulent period, hence, potentially many shocks too.

What we find is that both crude oil markets and silver have a higher probability of shocks (positive *and* negative) when they are in contango (both short- and long-term). Since arbitraging should be more difficult for energy markets, shocks occur more easily than for financial markets. But

the fact that they occurred in contango is concerning, since it has been a recent (and temporary for Brent) situation (please refer to Figure 1.5).

On the contrary, USD/EUR exchange rate and soy markets experience all their shocks consistently when they are in backwardation (with both definitions).

As for other similarities between commodity and financial markets, we find many common categorizations between crude oils and eurodollar.

If we now split between positive and negative shocks, we observe that the S&P500 experiences more positive shocks when it is in backwardation (both short- and long-term), just like wheat, in addition to exchange rate and soy markets (which experience all their shocks in backwardation).

If we split between short- and long-term basis, we additionally find that silver experiences more shocks when in short-term contango and in long-term backwardation. The case of gold is strange, since it should behave like silver (both are rarely in backwardation and are precious metals), but instead, it experiences price shocks when it is in long-term contango (its behaviour conditioning on the short-term basis is unclear). Just like silver, copper markets also experience shocks when in short-term contango, though they have different behaviours when looking at the long-term basis sign.

Table 1.5 – Probabilities of spot returns shocks (at 1% and 5%), conditional on contango or backwardation

(a) Probabilities of spot return shocks, conditional on 3m-basis.

	3m-contango	3m-backwardation	Unclear
Higher probability of negative shocks	<u>Crude oils</u> , silver and <i>wheat</i>	<u>Natural gases</u> , Eurodollar , exchange rate , coppers, <i>soy oil</i> , <i>soybean and corn</i>	<u>Heating oil, gasoil</u> , S&P500 and gold
Higher probability of positive shocks	<u>Brent</u> , <u>UK natural gas</u> , Eurodollar , silver and <i>corn</i>	<u>US natural gas</u> , exchange rate , S&P500 , coppers, <i>soy oil</i> , <i>soybean and wheat</i>	<u>Light crude, heating oil, gasoil</u> and gold

(b) Probabilities of spot return shocks, conditional on Last-basis.

	Last-contango	Last-backwardation	Unclear
Higher probability of negative shocks	<u>Crude oils, heating oil, UK natural gas, gasoil</u> , Eurodollar , gold (both never in bwd)	Exchange rate , silver, <i>soybean, soy oil, and corn</i>	<u>US natural gas</u> , S&P500 , coppers and <i>wheat</i>
Higher probability of positive shocks	<u>Crude oils</u> , Eurodollar , gold (both never in bwd) and <i>corn</i>	<u>Natural gases</u> , S&P500 , exchange rate , silver, <i>soy oil</i> , <i>soybean and wheat</i>	<u>Heating oil, gasoil</u> and coppers

We coloured the names of the markets by sectors (blue for energy, red for financial markets, green for agricultural markets) and use signs (underlined for energy, bold for financial markets, italic for agricultural markets) to easily see if we can clearly classify a sector into one category. "EW" means "Equally-Weighted", for the sector indices.

"3m-contango/backwardation" (resp. "Last-contango/backwardation") means that we condition the spot returns on 3m-contango/backwardation (resp. Last-contango/backwardation).

Ex: if a market is in the "Higher probability of negative shocks" "3m-contango" cell, it means that when we condition on 3m-contango, it has a higher probability of experiencing a shock both at 1% and 5%.

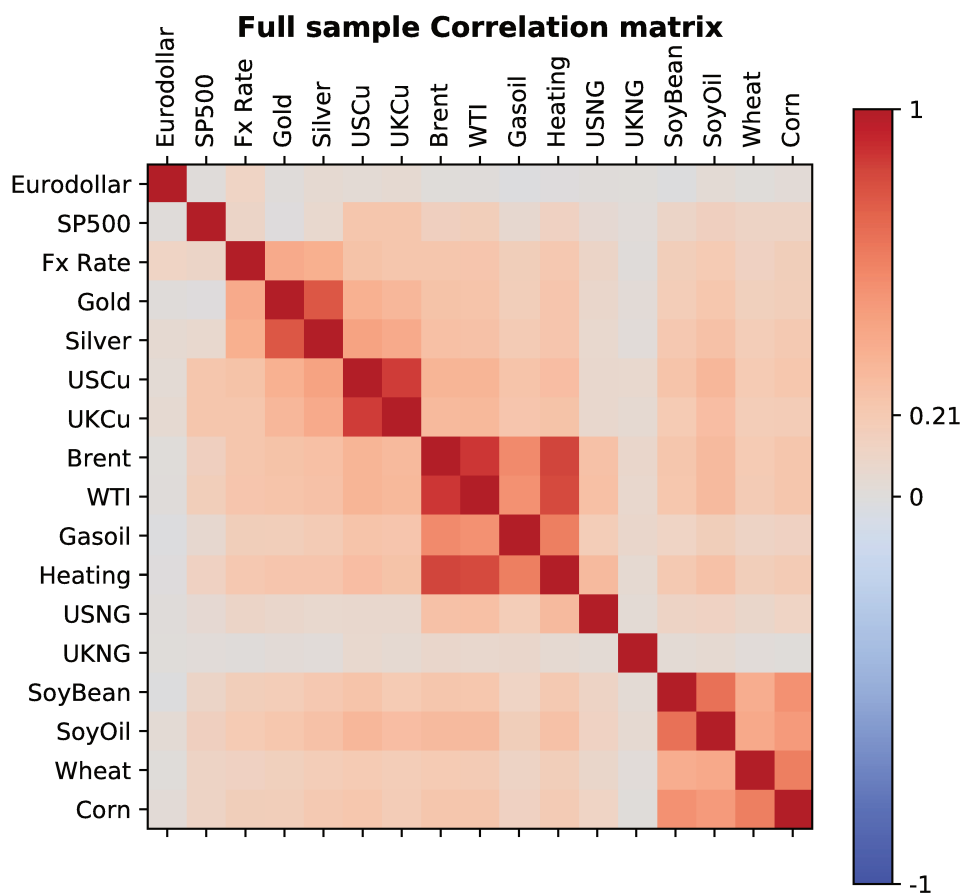
1.5 The potential for diversification and systematic risk

Since many commodities are related to the business cycle, e.g. crude oils and industrial metals, in case of downturn, they may exhibit similar price variations. This similarity can be captured by correlation coefficients. We will first look at the correlation matrix and then assess whether there could be diversification or whether there is integration (and potentially systematic risk) between the sectors. We also perform a Principal Component Analysis (PCA) on this correlation matrix and check whether one component can be composed of markets from each sector. We normalise the returns to 0 mean and 1 variance for each univariate series, in order to control for the potential initial difference in contribution to the variance of the system.

1.5.1 Diversification and correlations

Figure 1.7 shows the pairwise correlation coefficients (in the correlation matrix), with perfect anti-correlation (-1) in blue and perfect correlation (1) in red. Similar to the pairwise Kolmogorov-Smirnov tests and the Hellinger distances, we find that the UK natural gas prices and eurodollar interest rates evolve differently of the other prices (here, close to 0 correlation). Moreover, sectors are also clustered (dark red blocks around the diagonal), with an even sharper difference between them. The energy sector seems split into at least two: crude oils plus heating oil markets vs. the others. The metals sector is even subdivided into precious metals and industrial metals. The exchange rate is also more correlated to precious metals than other financial or commodity markets. The S&P500, as could be expected, correlates most with pro-cyclical commodities, i.e. coppers and crude oils here (which are also slightly more correlated than other inter-sector markets). The off diagonal blocks are lighter, but still indicate inter-sector correlations.

Figure 1.7 – Heatmap of the unconditional correlation matrix over the whole sample



We also compute the correlation matrix for sub-samples, splitting it into three: 2000-2003, 2004-2008 and 2009-2014 (see Appendix 1.C for the heat-maps).

At the beginning of the sample, the correlation matrix displayed much clearer sectorization (high intra-sector correlations) with an average correlation of only 0.1026. The exchange rate was correlated with precious metals and the S&P500 only with industrial markets, while it was anti-correlated to those precious metals and exchange rate. This is consistent with what Gorton and Rouwenhorst (2006) find: zero or even negative cor-

relation between an index of commodity futures and traditional financial assets (equities and bonds indices).

After the beginning of the financialization (2003-2004), the average correlation increased to 0.23. This is in line with Tang and Xiong (2012), who find that the correlation between commodities increases after 2004. This increase in average comes from both intra- and inter-sector correlations, but the intra-sector correlations were already significant before 2004, thus increased only slightly. Almost all commodity inter-sector correlations, not just those with oil markets, contrary to Tang and Xiong (2012). In addition, the eurodollar and S&P500 markets are not really correlated with commodities, as one would expect from their financialization. This process thus seems to be a long-term evolution of markets.

The correlation increase again after 2008, with an average of 0.2734 and this time, as Büyükşahin and Robe (2014) find, large correlations of S&P500 with the exchange rate, industrial metals and energy markets appear... Moreover, commodities in general also become more correlated.

Though simple, the correlation matrix allows to recover many results from the literature that use more sophisticated methodologies. We will further investigate the drivers of the variance in the system by conducting a Principal Component Analysis (PCA).

1.5.2 The presence of common factors

Integration and financialization could result in the existence of a common factor for both commodities and financial markets. We thus investigate this by conducting a Principal Component Analysis (PCA) on our markets: it will tell whether there will be a single component related to all sectors with the same sign (i.e. there is financialization) or if components are weighting markets differently.

Figure 1.8 presents the weights of each market into each of the first 4 components (numbered from 1 to 4, subfigure 1.8a) and the explained variance

ratios (subfigure 1.8b) for all components, for the whole sample ("Full") and for each subsample.

From the latter, we see that the first component explains almost 30% of the variance of the system. If we look at the markets involved in this component (the blue line in the top figure), we find that almost all markets participate in it, with weights having the same sign (could be all positive, for easier reading). Lower weights are put on natural gas and financial markets (eurodollar close to 0), showing little integration of all sectors, but still existent.

The second component (orange line), explaining around 12% of the variance of the system, is also composed of commodities, but confronts energy markets (negative weights) to agricultural markets and, to a lesser extent, metals and finance (all positive weights). This component and the following ones are less easy to interpret, but some make sense. For example, the fourth one (red line) contrasts the S&P500 and industrial metals (pro-cyclical) with precious metals (contra-cyclical) and exchange rate (expressed as dollars per euro).

Conducting PCA on financial markets only, we find that the first factor is driven almost entirely by the S&P500 and explains almost 80% of the variance of the system. The second one, explaining around 20% of it, is for the exchange rate. The last one, for the eurodollar market, doesn't explain much. There is thus no commonality among traditional financial assets.

On agricultural markets, the first component is common to all with weights between 35% and 60% and explains more than 60% of the variance of the sector, confirming a sectorial factor. The second one confronts soy markets to wheat and corn and explains 20% of the variance.

Metals markets also have a sectorial factor, as the first component weights all of them positively, between 20% and 60% and explains more than 65% of their variance. Further down, copper markets face precious metals on the second component, which explains around 25% of the variance. Finally, the last two components put gold and silver facing each other and UK copper and US copper facing each other.

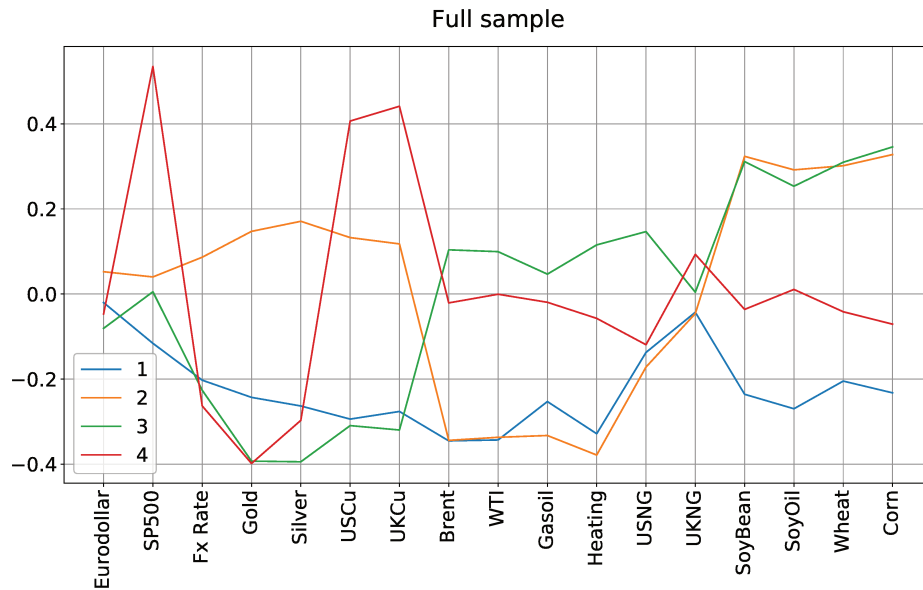
Looking at energy markets, we also find a common factor, but it explains only around 40% of their variance. The second component is highly weighting UK natural gas, opposed to all other energy markets and explains about 33% of the variability in this sector. As we saw before, its spot return distribution has huge kurtosis and the largest volatility, followed by the its US counterpart, who is contributing a lot to the third component, explaining around 20% of the variance. The others do not explain much of it.

Finally, we run a PCA on all the commodities together. We do find a common factor for all of them, but a little weak as it explains only 34% of the variance. The second component explains around 14% of it and confronts energy markets to the others, while the third one puts metals and the others facing each other, explaining around 12% of the variability. These factors are consistent with those found when including financial markets, which seem more related to metals markets, considering their weights have the same sign in the similar components.

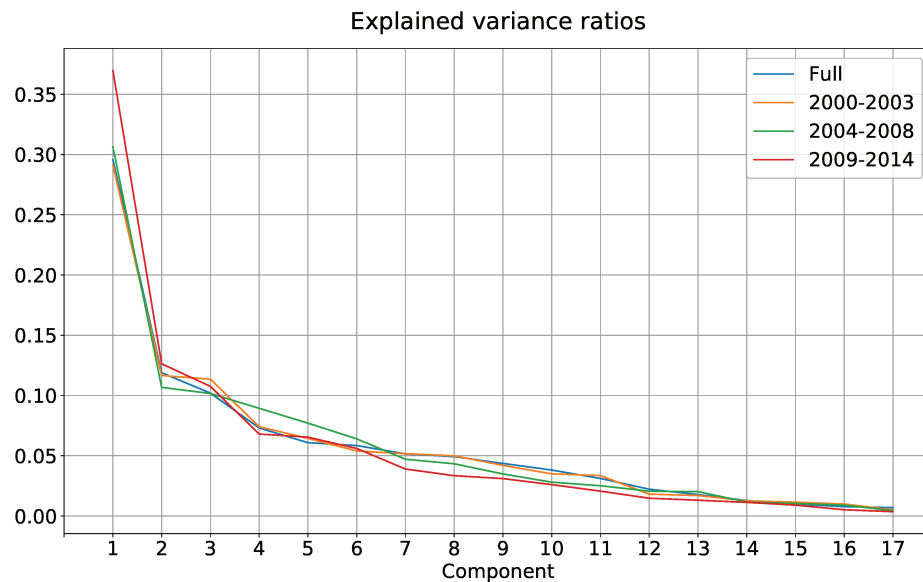
We also conduct a PCA for each subsample, as for the correlation matrix (2000-2003, 2004-2008, 2009-2014). The first 6 components and explained variance ratios are given in Appendix 1.D. The previous results on the correlation matrix per subsample are confirmed here. The beginning of the sample sees the energy sector separated from the others (first component). Then, in 2004-2008, commodities in general participate in the first component (which explain a little more than 30% of the variance), as evidenced by Tang and Xiong (2012), with a little contribution from financial markets, mostly the exchange rate. Finally, after 2008, the S&P500 increases its contribution to the first component, which has about the same shape as in the whole sample results, confirming the results of Büyükşahin and Robe (2014) again.

Figure 1.8 – Results of a PCA in the spatial dimension

(a) Weights of the first 4 components in the spatial dimension over the whole sample (2000-2014)



(b) Explained variance ratios per principal component in the spatial dimension



"Full" denotes the explained variance ratios for the PCA performed over the whole sample (2000-2014), others are for the PCA performed on each subsample.

Conclusion

In light of the early literature on commodity futures, we wanted to reassess the previous results, considering the evolution of these prices around the 2008 crisis. Indeed, Gorton and Rouwenhorst (2006) found that until 2004, commodity futures returns were not or even negatively correlated to equities and bonds, while providing the same average return with lower volatility. On the contrary, the theory of storage predicts that commodities prices should be more volatile, though they use a commodity index (which may benefit from commodity diversification). Individually, they are more difficult to store (and trade, physically) and are subject to a physical non-negativity constraint compared to financial assets. This should translate into positive skewness (because the demand can exceed the supply), which they indeed find at that time, while financial markets exhibit negative skewness in general.

We thus investigated the moments of the univariate empirical distributions to 1) check whether their results still held (and commodity futures could still be used as an asset class on their own) and 2) if the physical aspects of commodities do translate into different return distributions. We compared the moments of the distributions, their quantiles and extrema and finally the whole distributions by computing their pairwise Kolmogorov-Smirnov and Hellinger distances. We used several markets for the energy, metals and agricultural markets and for financial assets, and equity index (S&P500), a short-term interest rate (eurodollar) and the USD/EUR exchange rate to compare individual commodities to other commodities and to traditional financial assets.

We first find that, consistent with the theory of storage, there is sectorisation: commodities from the same sector have similar distributions. Moreover, the more difficult a commodity is to store, the larger the standard deviation and thus the average return. We do not find obvious results in terms of skewness. Looking at the unconditional extrema, we find a hint of financialization: several commodity prices experience their extrema on the day after the S&P500 experiences its extrema. We need to investigate

this further and postpone this to Chapter 3.

But to account for different market conditions, we differentiate two states: when the market seems to have enough inventories to match the demand (contango) and when inventories are scarce (backwardation). We thus condition the returns for one market on its contango and on backwardation and compare the results between the two conditionings and between markets and sectors. We find that indeed, backwardation is associated with larger standard deviation (since inventories cannot dampen price variations) for most commodities, but it does not always translate into larger average return (only for most energy markets). This conditioning also allowed to evidence that most price shocks on crude oils occurred while it was in contango, which was the "normal" market condition only recently (2005-2007 and after mid-2008), maybe due to its financialization.

To further investigate the potential changes in return distributions for commodities, we look at the correlation matrix of all the markets retained here and try to find common factors using Principal Component Analysis. We do find sectorisation, but it fades away over time: before 2004, sectors were quite clearly separated, then commodities integrated, as found by Tang and Xiong (2012), then, after 2008, financial markets also correlate with commodities, as evidenced by Büyüksahin and Robe (2014), with the first principal component involving all markets.

We leave more advanced analyses of the correlation matrix to the following chapters. In Chapter 2, we will filter this correlation matrix estimated using a rolling window to analyse what happens around the beginning of the subprime crisis and around the default of Lehman Brothers, two shocks that could thus affect commodity markets. In Chapter 3, we will further improve this analysis by adding a propagation component to systemic risk and using a new high dimensional methodology to investigate the observations about the maximum and minimum return of the S&P500 and other commodities.

References

- Black, Fischer (1976). ‘The pricing of commodity contracts’. In: *Journal of Financial Economics* 3(1), pp. 167–179.
- Brennan, Michael J (1958). ‘The supply of storage’. In: *The American Economic Review* 48(1), pp. 50–72.
- Brunnermeier, Markus K. (2009). ‘Deciphering the Liquidity and Credit Crunch 2007–2008’. In: *Journal of Economic Perspectives* 23(1), pp. 77–100.
- Büyüksahin, Bahattin and Michel A. Robe (2014). ‘Speculators, commodities and cross-market linkages’. In: *Journal of International Money and Finance* 42, pp. 38–70.
- Cheng, Ing-Haw and Wei Xiong (2014). ‘Financialization of commodity markets’. In: *Annu. Rev. Financ. Econ.* 6(1), pp. 419–441.
- Deaton, Angus and Guy Laroque (1992). ‘On the behaviour of commodity prices’. In: *The Review of Economic Studies* 59(1), pp. 1–23.
- Fama, Eugene F. and Kenneth R. French (1987). ‘Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage’. In: *The Journal of Business* 60(1), pp. 55–73.
- Fama, Eugene F. and Kenneth R. French (1988). ‘Business Cycles and the Behavior of Metals Prices’. In: *The Journal of Finance* 43(5), pp. 1075–1093.
- Garbade, Kenneth D. and William L. Silber (1983). ‘Price Movements and Price Discovery in Futures and Cash Markets’. In: *The Review of Economics and Statistics* 65(2), pp. 289–297.
- Gorton, Gary B, Fumio Hayashi and K Geert Rouwenhorst (2013). ‘The fundamentals of commodity futures returns’. In: *Review of Finance* 17(1), pp. 35–105.
- Gorton, Gary B and K Geert Rouwenhorst (2006). ‘Facts and Fantasies about Commodity Futures’. In: *Financial Analysts Journal* 62(2), pp. 47–68.
- Hasbrouck, Joel (2003). ‘Intraday Price Formation in U.S. Equity Index Markets’. In: *The Journal of Finance* 58(6), pp. 2375–2400.

- Hicks, J. R. (1946). *Value and Capital: An Inquiry into some Fundamental Principles of Economic Theory*. Oxford University Press.
- Kaldor, Nicholas (1939). 'Speculation and Economic Stability'. In: *The Review of Economic Studies* 7(1), pp. 1–27.
- Keynes, John Maynard (1930). *A Treatise on Money: In 2 Volumes*. Macmillan & Company.
- Nash, Daniel J. (2001). 'Long-term Investing in Commodities'. In: *Global Pensions Quarterly, Morgan Stanley Global Pensions* 4(1), pp. 25–31.
- Schwarz, Thomas V. and Andrew C. Szakmary (1994). 'Price discovery in petroleum markets: Arbitrage, cointegration, and the time interval of analysis'. In: *Journal of Futures Markets* 14(2), pp. 147–167.
- Tang, Ke and Wei Xiong (2012). 'Index Investing and the Financialization of Commodities'. In: *Financial Analysts Journal* 68(6), pp. 54–74.
- Telser, Lester G (1958). 'Futures trading and the storage of cotton and wheat'. In: *Journal of Political Economy* 66(3), pp. 233–255.
- Till, Hilary (2006). 'A Review of the Historical Literature on Commodity Futures Contract Returns'. In: *Working paper*.
- Working, Holbrook (1949). 'The Theory of Price of Storage'. In: *The American Economic Review* 39(6), pp. 1254–1262.
- Zhang, Yue-Jun et al. (2008). 'Spillover effect of US dollar exchange rate on oil prices'. In: *Journal of Policy Modeling* 30(6), pp. 973–991.

Appendices

1.A Statistical measures

1.A.1 The Kolmogorov-Smirnov distance and test

The Kolmogorov-Smirnov distance is defined as:

$$D_{1,2} = \sup_x |F_1(x) - F_2(x)| \quad (1.2)$$

where F_i is the cumulative distribution function (CDF) of variable i . It represents the largest distance between the two CDFs. Thus, if it is small, it means that the two CDFs are relatively close to each other, i.e. the samples may be drawn from same distribution. Statistical tests exist for assessing this null hypothesis and thus allows to reject it or not depending on the value of the distance.

Table 1.6 – p-values for the KS tests of pairwise distribution differences

	Eurodollar	SP500	Fx Rate	Gold	Silver	USCu	UKCu	Brent	WTI	Gasoil	HO	USNG	UKNG	SoyBean	SoyOil	Wheat	Corn
Eurodollar	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SP500	0.00	1.00	0.00	0.07	0.41	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.14	0.00	0.00	0.00
Fx Rate	0.00	0.00	1.00	0.06	0.00	0.03	0.01	0.49	0.45	0.06	0.17	0.08	0.00	0.03	0.25	0.25	0.19
Gold	0.00	0.07	0.06	1.00	0.47	0.31	0.08	0.22	0.21	0.09	0.01	0.00	0.00	0.29	0.01	0.00	0.02
Silver	0.00	0.41	0.00	0.47	1.00	0.06	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.18	0.00	0.00	0.00
USCu	0.00	0.02	0.03	0.31	0.06	1.00	0.32	0.45	0.32	0.00	0.06	0.01	0.00	0.17	0.03	0.02	0.13
UKCu	0.00	0.00	0.01	0.08	0.00	0.32	1.00	0.01	0.01	0.00	0.01	0.00	0.00	0.01	0.01	0.00	0.06
Brent	0.00	0.01	0.49	0.22	0.01	0.45	0.01	1.00	1.00	0.03	0.25	0.09	0.00	0.39	0.31	0.25	0.20
WTI	0.00	0.00	0.45	0.21	0.01	0.32	0.01	1.00	1.00	0.08	0.08	0.10	0.00	0.45	0.13	0.22	0.06
Gasoil	0.00	0.00	0.06	0.09	0.00	0.00	0.00	0.03	0.08	1.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
HO	0.00	0.00	0.17	0.01	0.00	0.06	0.01	0.25	0.08	0.00	1.00	0.10	0.00	0.01	0.99	0.77	0.22
USNG	0.00	0.00	0.08	0.00	0.00	0.01	0.00	0.09	0.10	0.00	0.10	1.00	0.00	0.00	0.29	0.24	0.37
UKNG	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
SoyBean	0.00	0.14	0.03	0.29	0.18	0.17	0.01	0.39	0.45	0.02	0.01	0.00	0.00	1.00	0.03	0.01	0.02
SoyOil	0.00	0.00	0.25	0.01	0.00	0.03	0.01	0.31	0.13	0.00	0.99	0.29	0.00	0.03	1.00	0.94	0.22
Wheat	0.00	0.00	0.25	0.00	0.00	0.02	0.00	0.25	0.22	0.00	0.77	0.24	0.00	0.01	0.94	1.00	0.57
Corn	0.00	0.00	0.19	0.02	0.00	0.13	0.06	0.20	0.06	0.00	0.22	0.37	0.00	0.02	0.22	0.57	1.00

Table 1.6 provides the p-values for all the pairwise KS tests, that has been represented as a heatmap in Figure 1.3.

1.A.2 The Hellinger distance

The Hellinger distance quantifies how different two probability distributions are different over their domain. More formally, it can be written as:

$$H(F_1, F_2) = \sqrt{\frac{1}{2} \int \left(\sqrt{f_1(x)} - \sqrt{f_2(x)} \right)^2 dx} \quad (1.3)$$

where f_i is the density function of variable i . This measure thus sums all the differences between the two density functions and should provide a more accurate distance measure than the Kolmogorov-Smirnov distance. In the case of two normal distributions, it can even be expressed in closed form:

$$H^2(F_1, F_2) = 1 - \sqrt{\frac{2\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2}} \exp\left(-\frac{1}{4} \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}\right) \quad (1.4)$$

with μ_i and σ_i respectively the mean and standard deviation of the normal distribution of variable i .

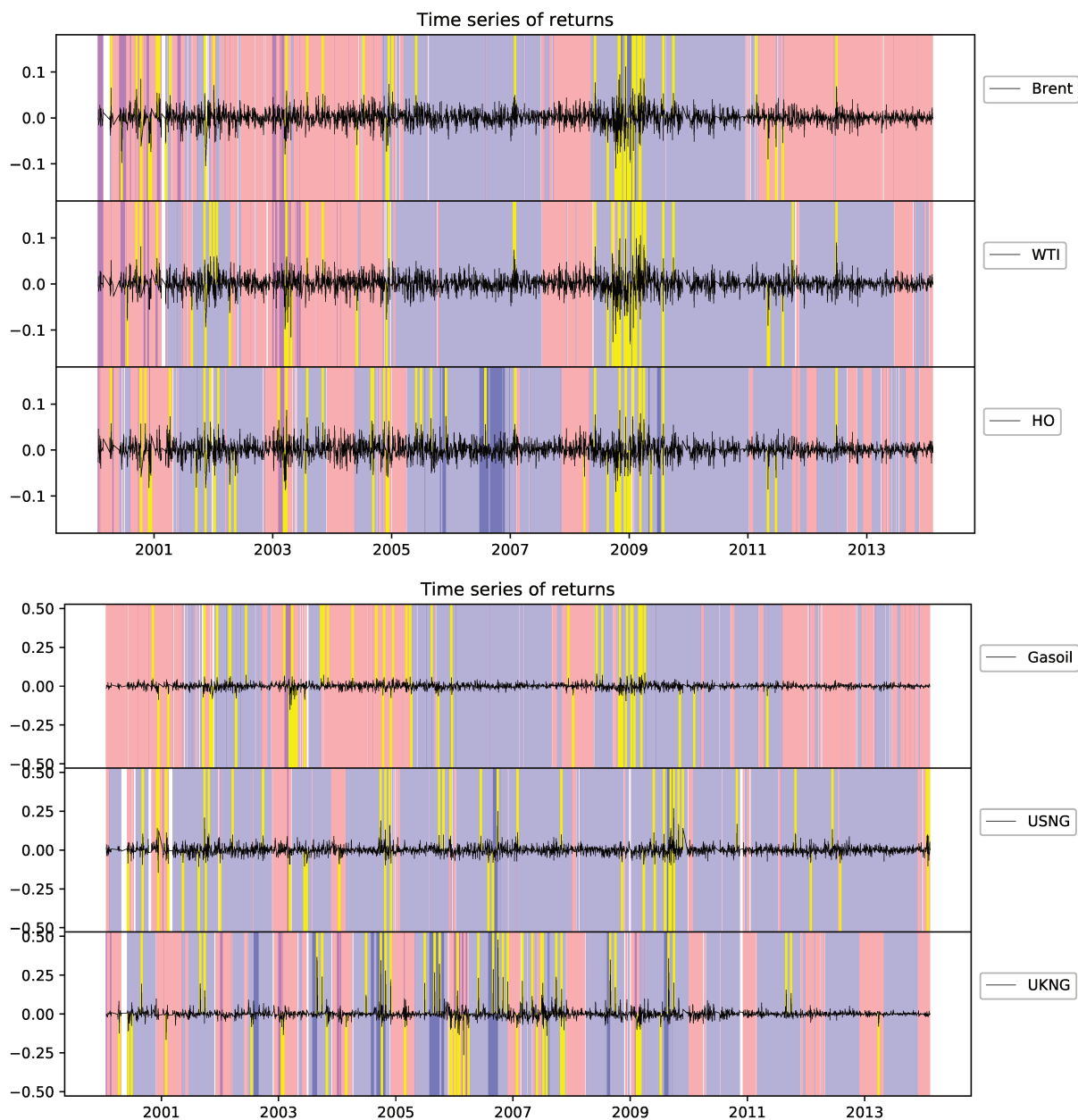
This measure requires that the normal distributions are thus not normalised and we will use this last formula to compute the distances, approximating the empirical distributions with univariate Gaussian distributions.

1.B Conditioning returns on the sign of the basis

1.B.1 Time series of spot returns and shocks conditional on the 3m-basis

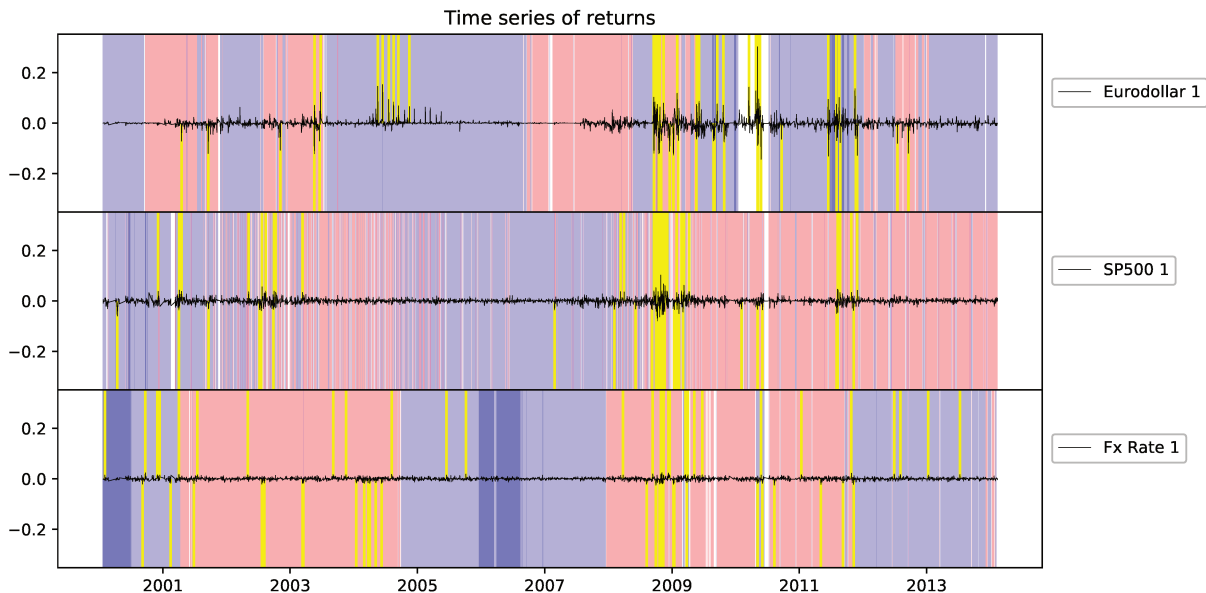
Figure 1.9 – Time series of spot returns

(a) For energy markets

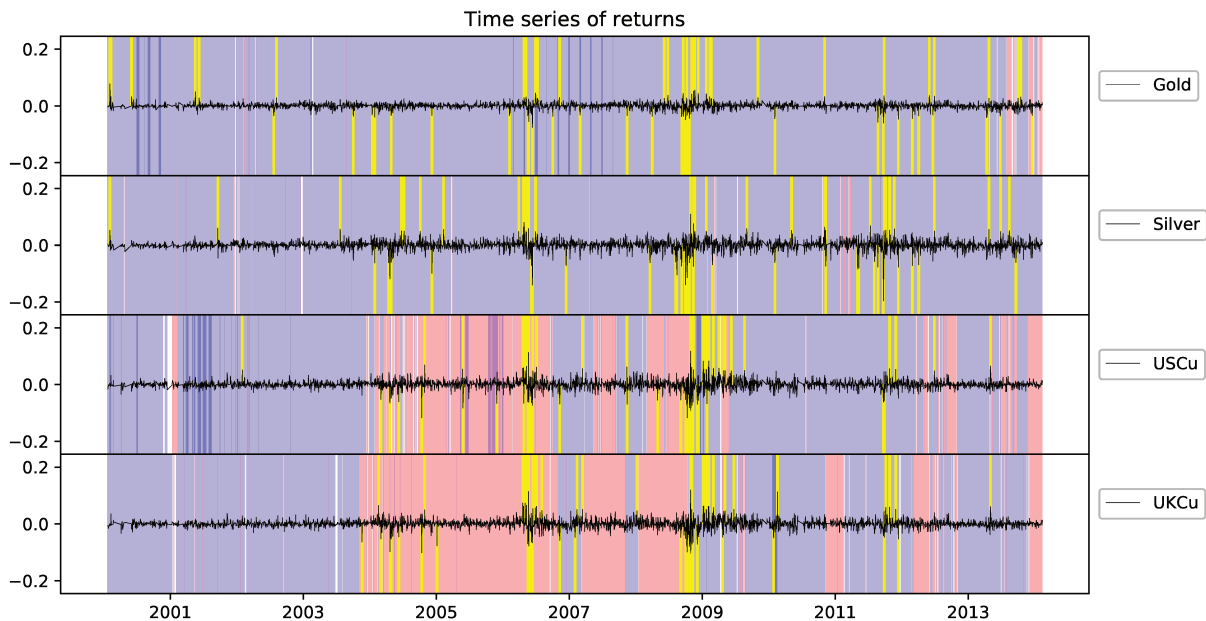


The yellow bars represent "shocks" at 1% (positive if the bar is above 0, negative if below) and the background is red if the market is in 3m-backwardation and blue if it is in 3m-contango. If it is darker, it means that the basis has an amplitude of more than half its maximum over the whole sample.

(b) For financial markets

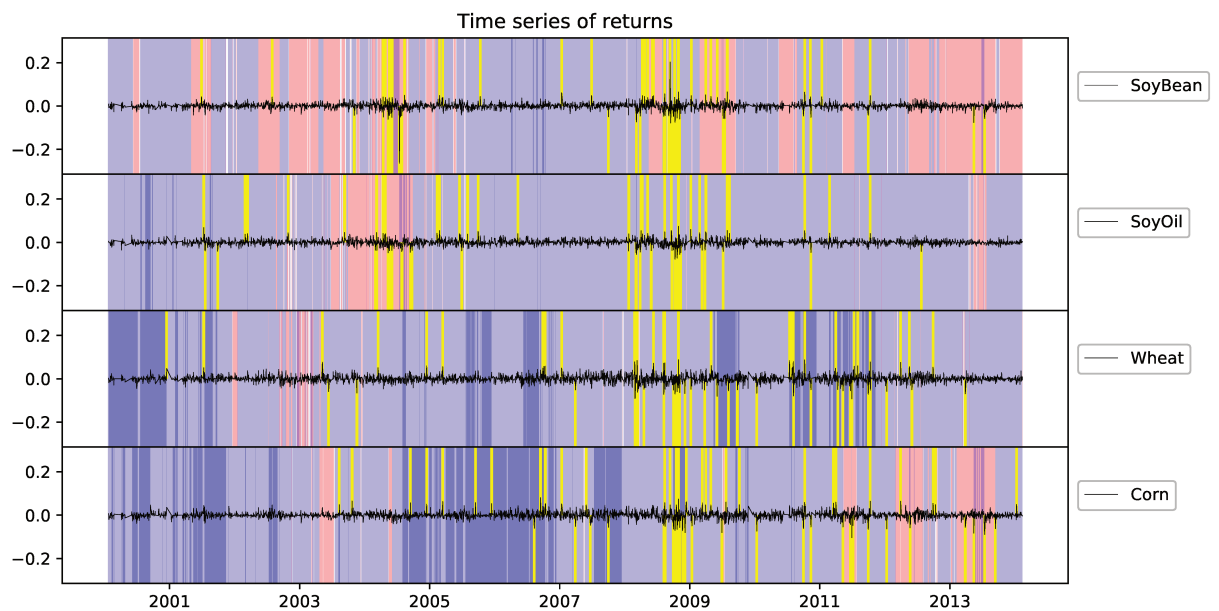


(c) For metals markets



The yellow bars represent "shocks" at 1% (positive if the bar is above 0, negative if below) and the background is red if the market is in 3m-backwardation and blue if it is in 3m-contango. If it is darker, it means that the basis has an amplitude of more than half its maximum over the whole sample.

(d) For agricultural markets

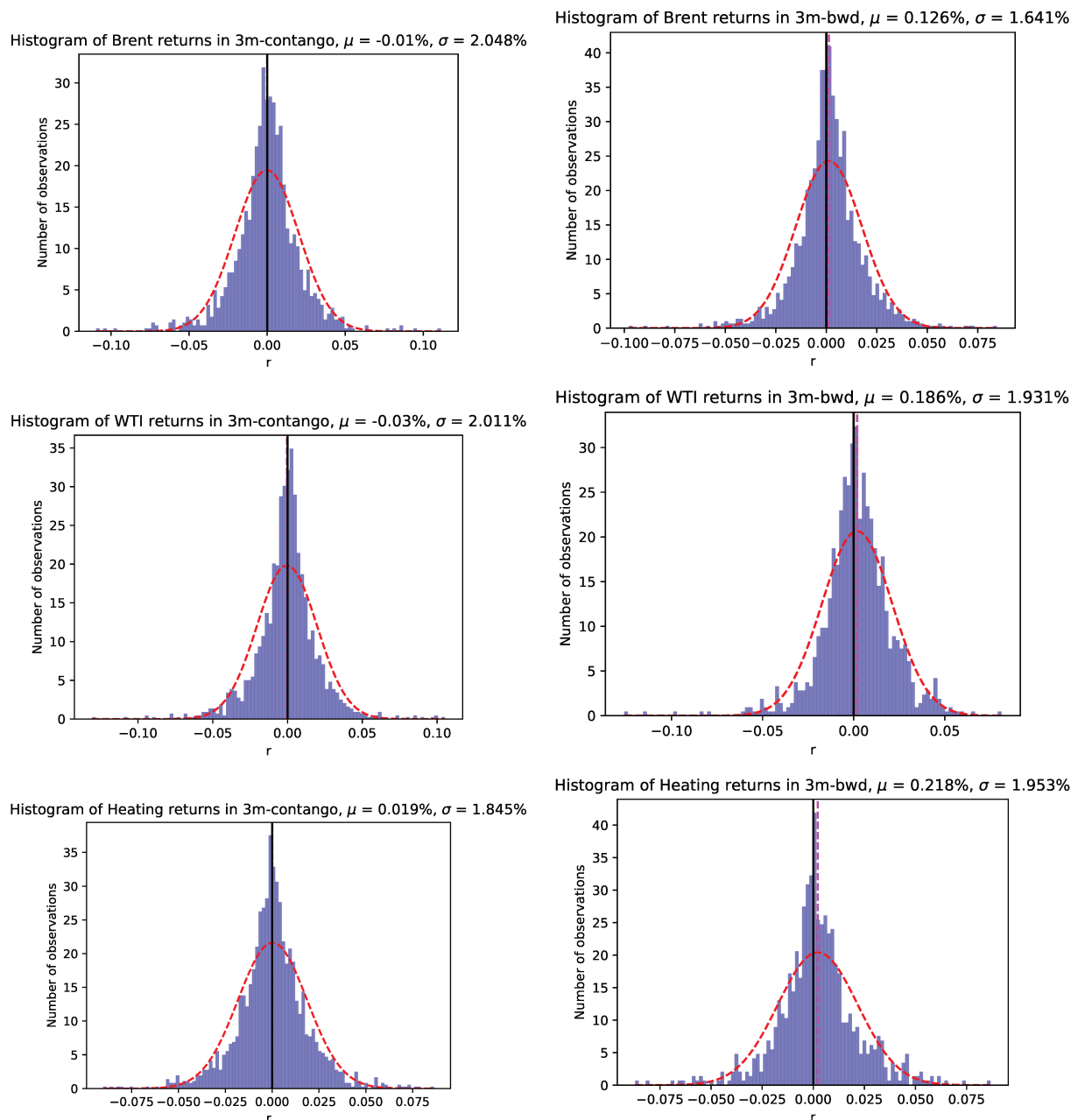


The yellow bars represent "shocks" at 1% (positive if the bar is above 0, negative if below) and the background is red if the market is in 3m-backwardation and blue if it is in 3m-contango. If it is darker, it means that the basis has an amplitude of more than half its maximum over the whole sample.

**1.B.2 Empirical distributions of returns conditional
on the sign of the 3m-basis**

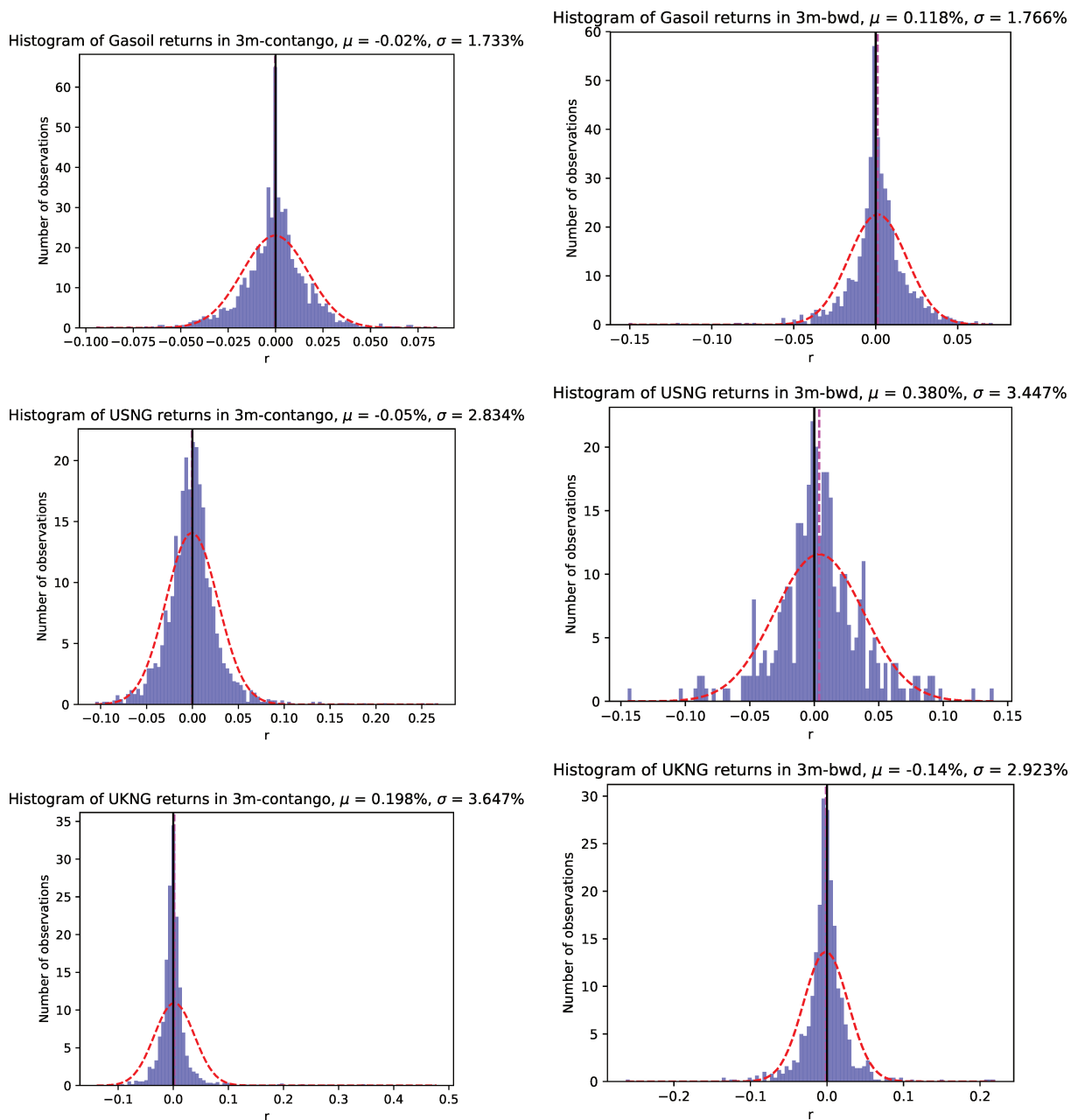
Figure 1.10 – Conditional spot return distributions

(a) For energy markets (1)



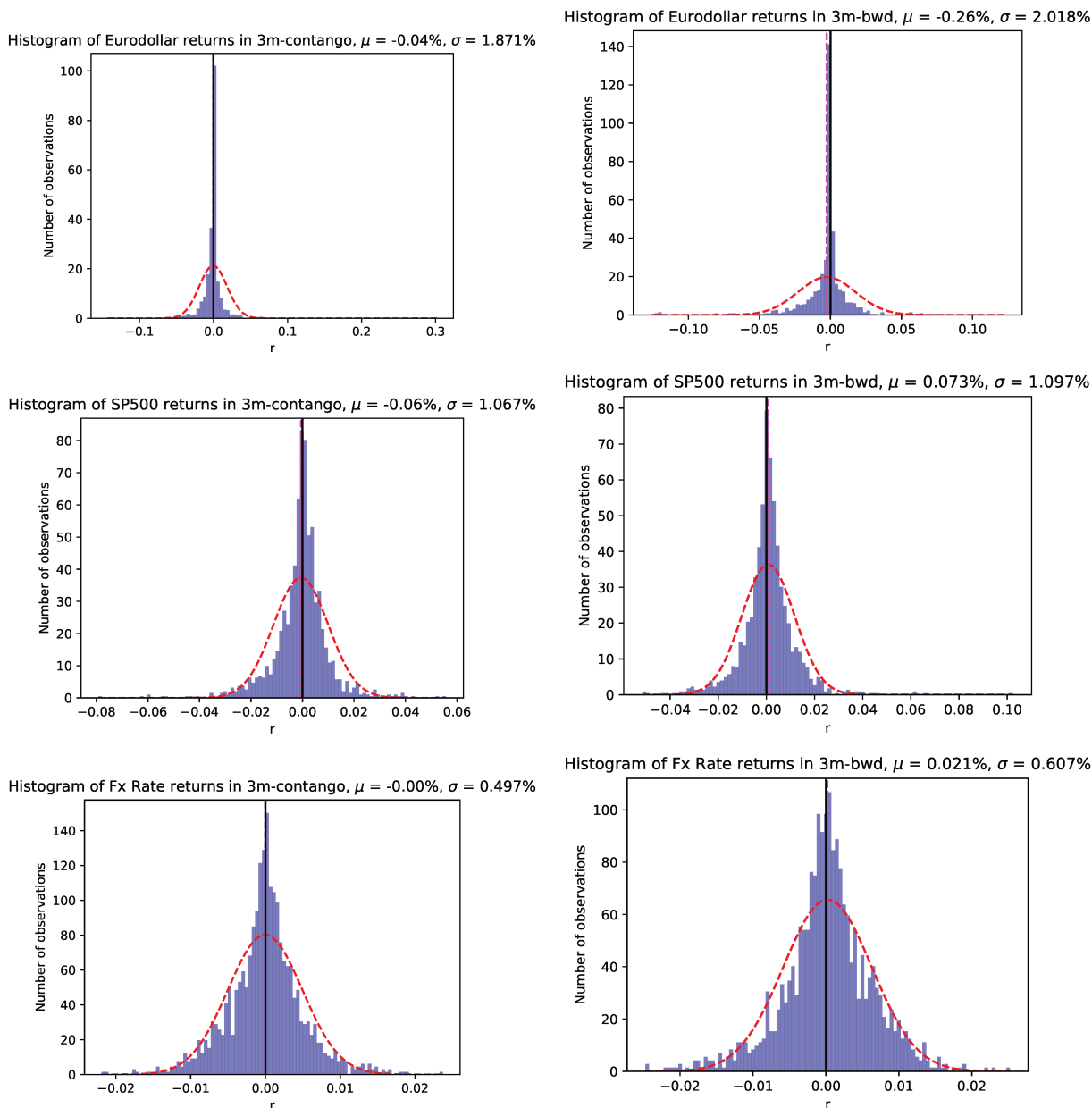
The pink dashed line represents the mean of the return distributions and the red dashed curve represents a Gaussian density function that has the same mean μ and volatility σ .

(b) For energy markets (2)



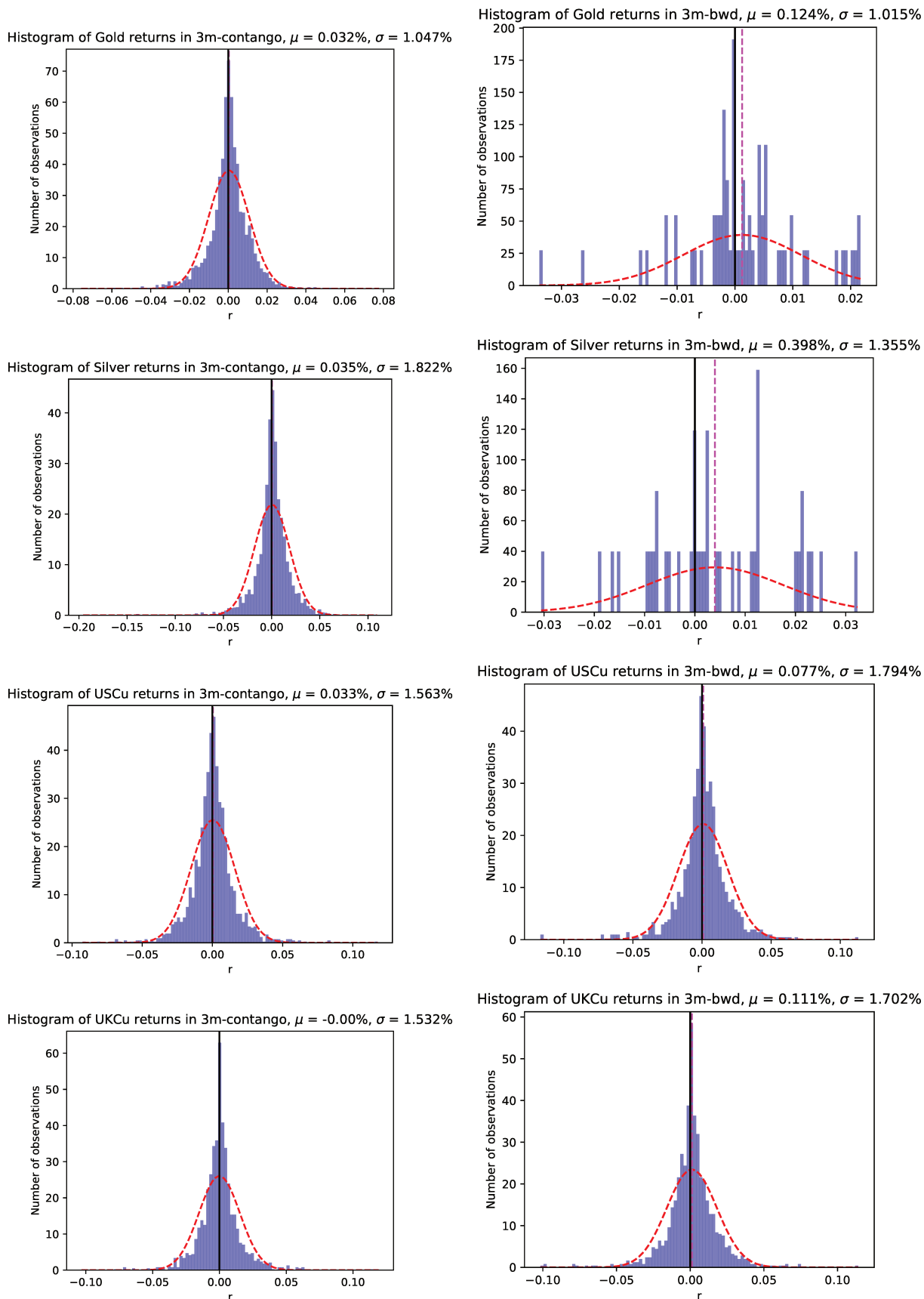
The pink dashed line represents the mean of the return distributions and the red dashed curve represents a Gaussian density function that has the same mean μ and volatility σ .

(c) For financial markets



The pink dashed line represents the mean of the return distributions and the red dashed curve represents a Gaussian density function that has the same mean μ and volatility σ .

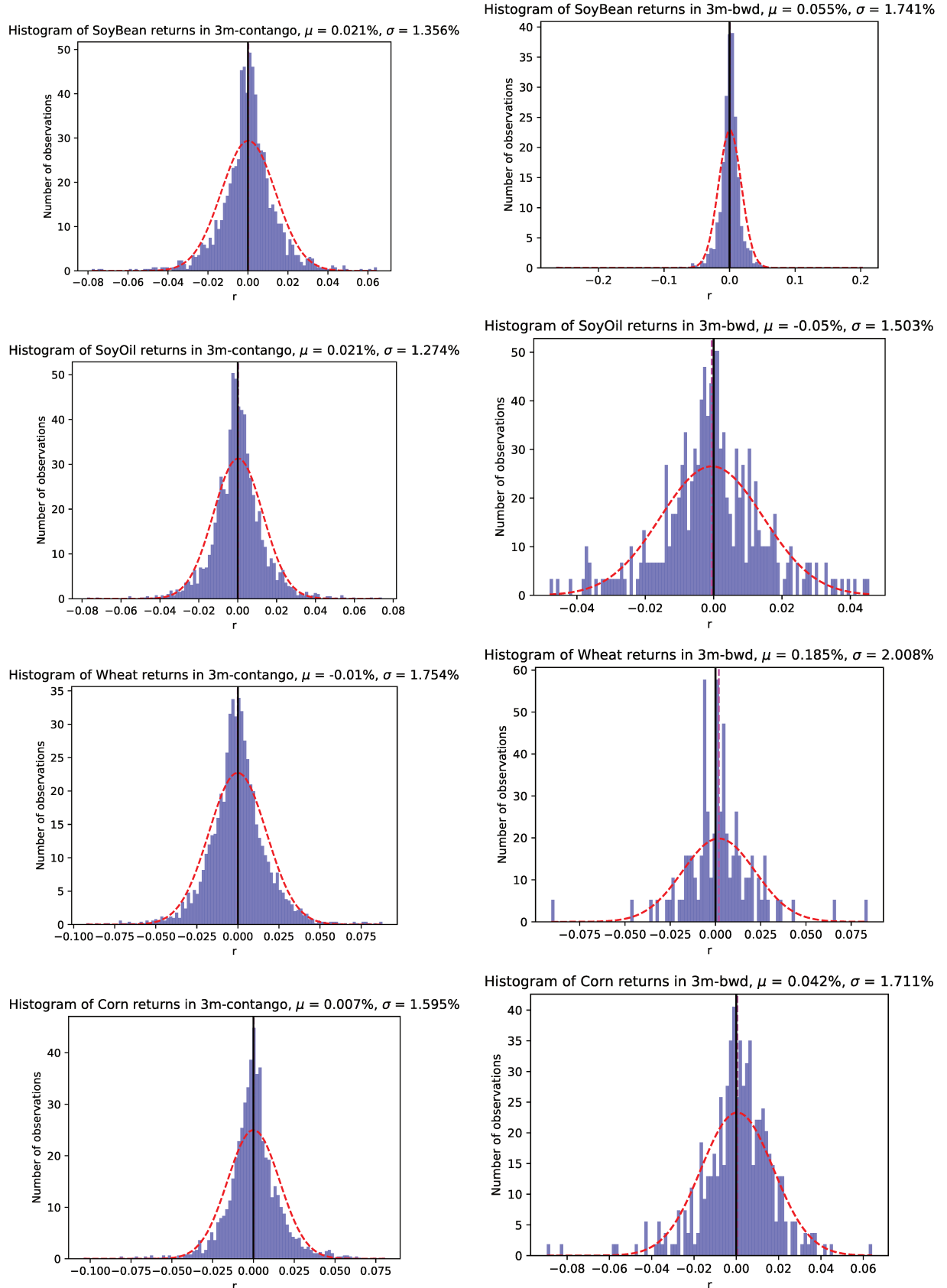
(d) For metals markets



The pink dashed line represents the mean of the return distributions and the red dashed curve represents a Gaussian density function that has the same mean μ and volatility σ .

CHAPTER 1. ARE COMMODITY MARKETS STILL DIFFERENT FROM OTHER ASSET CLASSES?

(e) For agricultural markets



The pink dashed line represents the mean of the return distributions and the red dashed curve represents a Gaussian density function that has the same mean μ and volatility σ .

1.B.3 Computing the moments of the conditional distributions of spot returns

Table 1.7 – Moments of the empirical returns distributions conditional on the Last-basis sign

	P(C)	Conditional on Last-contango				Conditional on Last-backwardation			
		Mean	SD	Skew	Kurt	Mean	SD	Skew	Kurt
Brent	29.83%	-0.037%*	2.111%***	-0.25	4.45	0.108%	1.710%	-0.18	3.07
WTI	30.07%	-0.075%*	2.399%***	-0.20	4.41	0.096%	1.782%	-0.39	3.25
Gasoil	41.74%	-0.048%**	1.793%	-0.12	2.91	0.103%	1.717%	-0.73	7.51
Heating	44.71%	-0.037%***	1.897%	-0.04	2.42	0.173%	1.873%	0.05	2.35
USNG	65.07%	-0.100%***	2.903%	1.37	10.40	0.252%	3.005%	0.53	4.61
UKNG	67.84%	-0.086%***	2.967%	2.94	37.99	0.371%	4.201%***	4.70	37.64
Eurodollar	100%	-0.114%	1.914%	1.40	37.31				
SP500	48.17%	-0.066%***	1.051%	-0.67	7.16	0.072%	1.115%**	0.55	9.20
Fx Rate	53.23%	0.002%	0.507%	0.06	1.71	0.015%	0.607%***	-0.26	1.57
Gold	100%	0.032%	1.051%	-0.35	5.18				
Silver	84.02%	0.037%	1.697%	-1.05	8.82	0.039%	2.376%***	-1.56	10.91
USCu	48.41%	-0.030%**	1.580%	0.19	6.93	0.105%	1.719%***	-0.52	6.36
UKCu	38.80%	-0.068%***	1.507%	0.04	9.23	0.110%	1.661%***	-0.10	6.61
SoyBean	43.00%	-0.044%**	1.317%	-0.20	3.71	0.091%	1.648%***	-1.71	54.79
SoyOil	77.59%	0.003%	1.300%	0.09	3.58	0.030%	1.328%	0.00	1.54
Wheat	85.36%	-0.036%**	1.704%	0.11	3.28	0.209%	2.060%***	-0.11	2.99
Corn	74.58%	-0.003%	1.593%	-0.06	3.84	0.052%	1.657%	-0.42	4.09

"C" stands for "contango" and "|" stands for conditioning. "SD" stands for "Standard Deviation", "Skew" stands for "Skewness" and "Kurt" stands for "Kurtosis" (computed here as Fisher's: "normal" kurtosis - 3).

Stars indicate the level of statistical difference at 1% (***), 5% (**), 10% (*). For means, either column is used, for variance (though standard deviation is displayed), stars are on the sample with the larger variance.

Note: Eurodollar and Gold are never in backwardation, so we cannot compare the moments of the two conditionings.

1.B.4 Probabilities of shocks conditional on contango and backwardation

Table 1.8 – Probabilities (in %) of shocks (negative: '1%' and '5%'), conditional on 3m-contango and 3m-backwardation

	P(B)	Neg. shocks				Pos. shocks			
		P(1% B)	P(1% \bar{B})	P(5% B)	P(5% \bar{B})	P(5% B)	P(5% \bar{B})	P(1% B)	P(1% \bar{B})
Brent	44.27	1.564	0.559	6.489	3.849	6.020	4.221	1.329	0.745
WTI	64.31	1.076	0.872	5.651	3.876	4.736	5.523	1.130	0.775
Gasoil	54.07	0.960	1.054	5.442	4.518	4.353	5.798	1.024	0.979
Heating	70.99	0.878	1.311	5.266	4.410	4.291	6.794	1.024	0.954
USNG	87.82	0.788	2.557	4.728	7.102	4.413	9.375	0.827	2.273
UKNG	64.07	0.702	1.540	4.430	6.064	5.024	5.005	1.351	0.385
Eurodollar	67.95	0.611	1.836	4.226	6.695	5.092	4.860	1.069	0.864
SP500	49.26	0.913	1.091	5.548	4.502	4.213	5.798	0.843	1.160
Fx Rate	50.02	0.415	1.593	3.804	6.233	3.596	6.440	0.968	1.039
Gold	97.72	0.992	1.515	5.064	3.030	4.887	10.606	1.027	0.000
Silver	98.62	1.018	0.000	5.053	2.500	5.053	2.500	1.018	0.000
UKCu	59.85	0.867	1.207	4.913	5.172	4.451	5.862	0.983	1.034
USCu	68.74	0.805	1.440	4.731	5.648	4.529	6.091	0.956	1.107
SoyBean	61.65	0.842	1.264	4.770	5.415	4.714	5.505	0.898	1.173
SoyOil	88.96	0.778	2.821	4.706	7.524	4.823	6.583	0.972	1.254
Wheat	96.23	1.007	0.917	5.034	4.587	4.998	5.505	0.971	1.835
Corn	87.78	0.946	1.416	4.848	6.232	5.085	4.533	1.104	0.283

"B" stands for 3m-contango and " \bar{B} " stands for 3m-backwardation. 1% is actually 1.0038% (due to 2889 observations) and 5% is actually 5.019%

Table 1.9 – Probabilities (in %) of shocks (negative: '1%' and '5%'), conditional on Last-contango and Last-backwardation

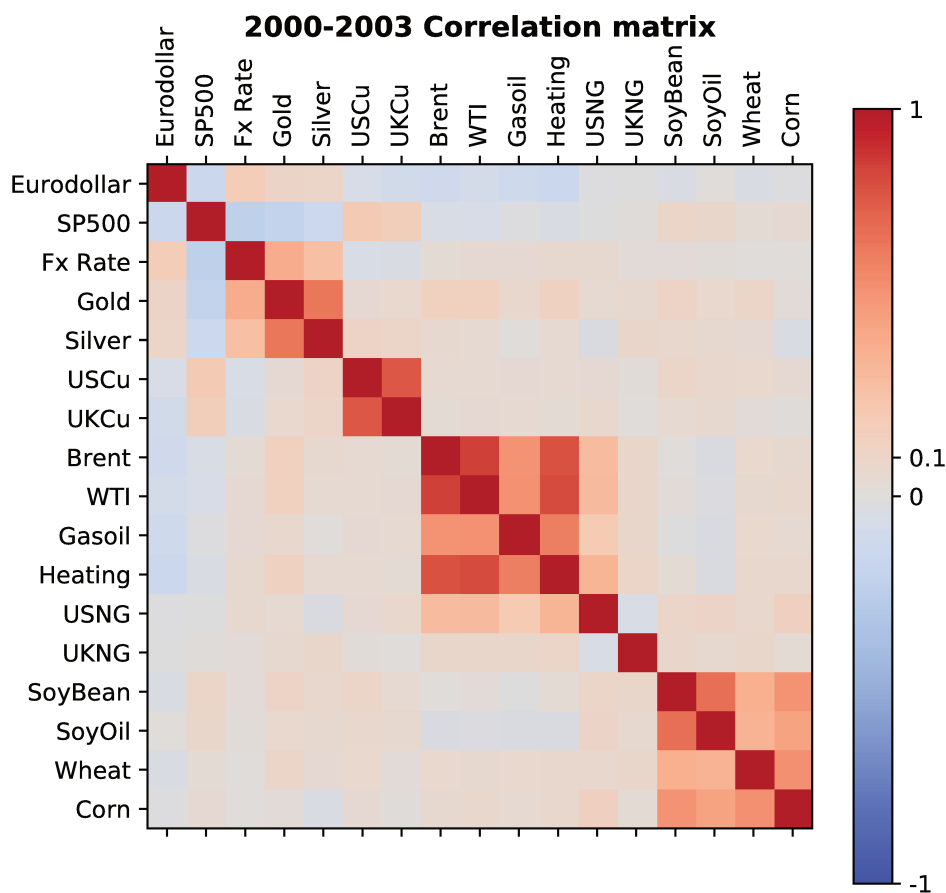
	Last-contango (= B)								
	P(B)	Neg. shocks				Pos. shocks			
		P(1% B)	P(1% \bar{B})	P(5% B)	P(5% \bar{B})	P(5% B)	P(5% \bar{B})	P(1% B)	P(1% \bar{B})
Brent	29.84	1.950	0.634	6.651	4.339	5.963	4.632	1.376	0.878
WTI	30.08	2.048	0.587	7.395	4.012	5.916	4.648	2.162	0.538
Gasoil	41.75	1.230	0.881	6.230	4.169	4.672	5.285	1.066	0.998
Heating	44.73	1.148	0.928	6.197	4.084	4.285	5.631	1.071	0.990
USNG	65.06	0.789	1.469	5.152	4.799	4.048	6.856	0.841	1.371
UKNG	67.86	1.059	0.957	5.245	4.574	4.085	7.021	0.605	1.915
Eurodollar	100	1.026		5.029		5.029		1.026	
SP500	48.15	0.852	1.188	5.540	4.554	4.048	5.941	0.710	1.320
Fx Rate	53.22	0.450	1.683	3.920	6.291	4.113	6.072	0.964	1.097
Gold	100	1.026		5.029		5.029		1.026	
Silver	84.02	0.814	2.141	4.520	7.709	4.112	9.850	0.774	2.355
USCu	48.39	0.848	1.194	5.088	4.973	4.028	5.968	1.201	0.862
UKCu	38.77	1.058	1.006	4.762	5.198	3.527	5.981	1.058	1.006
SoyBean	42.98	0.955	1.080	4.694	5.282	3.660	6.062	0.955	1.080
SoyOil	77.58	0.926	1.374	4.806	5.802	4.850	5.649	0.970	1.221
Wheat	85.35	1.042	0.935	4.689	7.009	4.369	8.879	0.962	1.402
Corn	74.57	0.872	1.480	4.862	5.518	5.046	4.980	1.055	0.942

"B" stands for Last-contango and " \bar{B} " stands for Last-backwardation. 1% is actually 1.0267% (due to 2922 observations) and 5% is actually 5.0308%.

1.C Correlation matrices by subsample

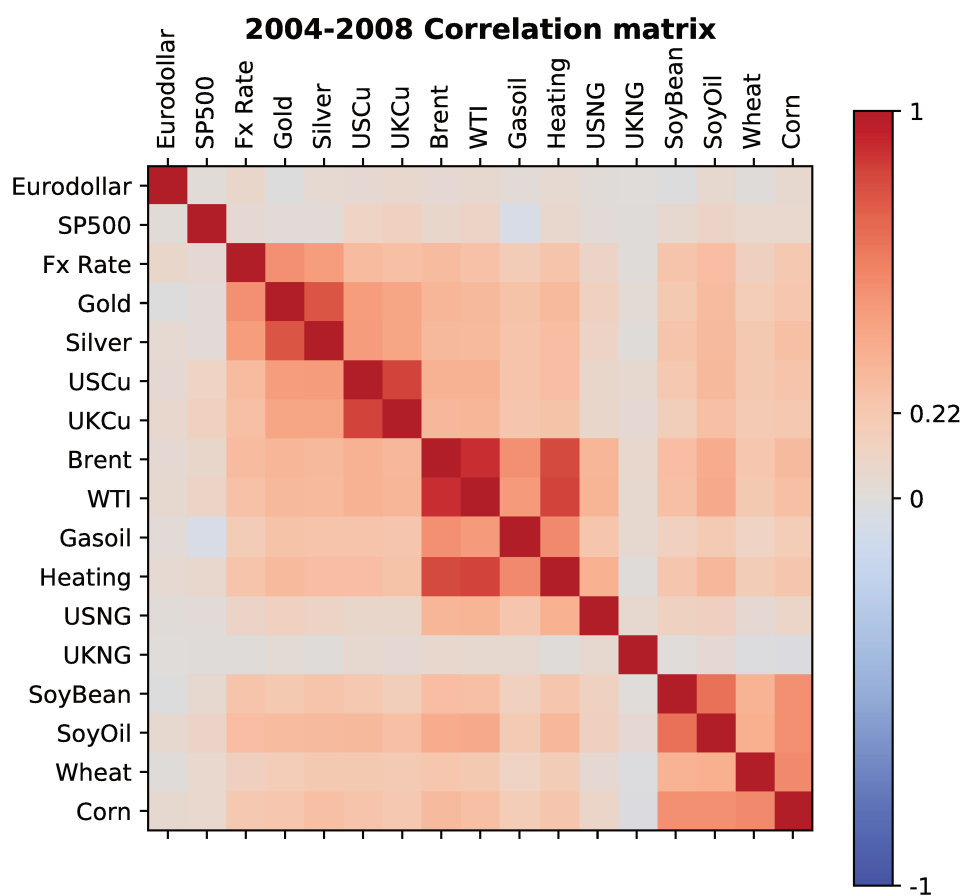
Figure 1.11 – Heatmap of the unconditional correlation matrix per subsample

(a) Heatmap of the unconditional correlation matrix over 2000-2003



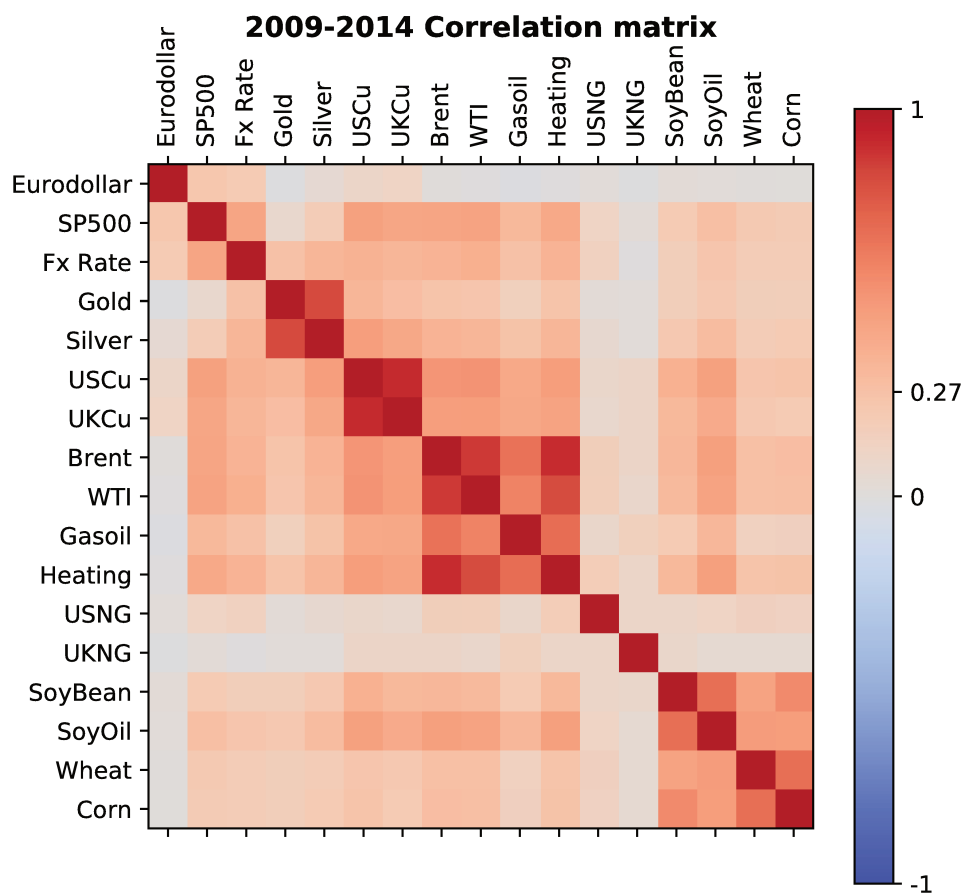
The warm colours (red shades) represent positive correlations, the cold colours (blue shades) represent negative correlations. The average of the off-diagonal correlation coefficients is provided on the scale, 0.1 here.

(b) Heatmap of the unconditional correlation matrix over 2004-2008



The warm colours (red shades) represent positive correlations, the cold colours (blue shades) represent negative correlations. The average of the off-diagonal correlation coefficients is provided on the scale, 0.22 here.

(c) Heatmap of the unconditional correlation matrix over 2009-2014



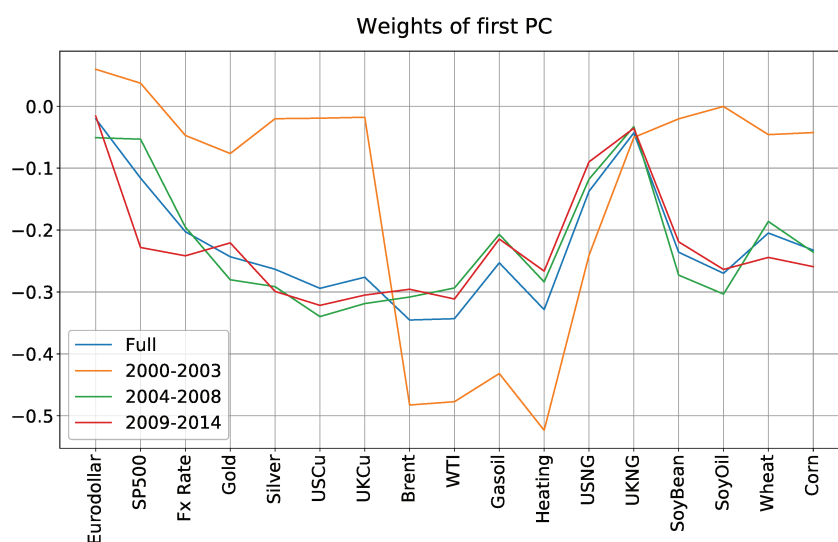
The warm colours (red shades) represent positive correlations, the cold colours (blue shades) represent negative correlations. The average of the off-diagonal correlation coefficients is provided on the scale, 0.27 here.

1.D Additional results on PCA

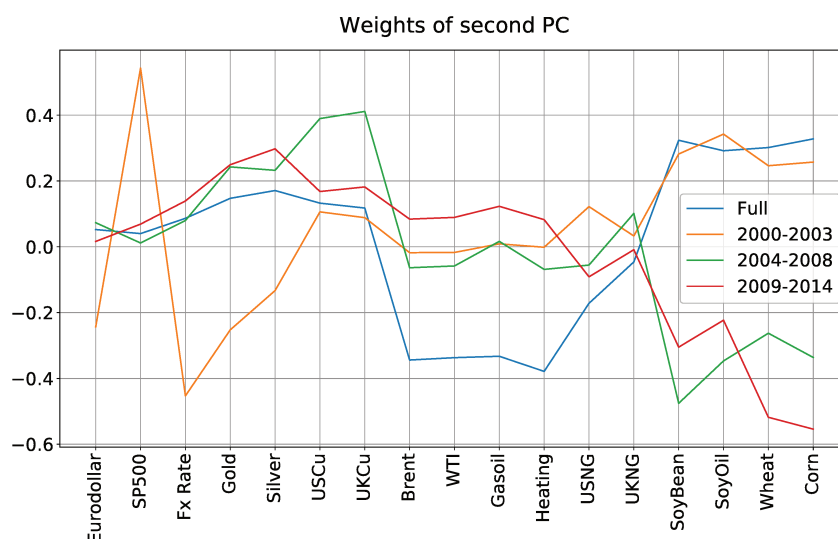
1.D.1 PCA results for each subsample

Figure 1.12 – Weights of the first 4 principal components

(a) Weights for the first principal component

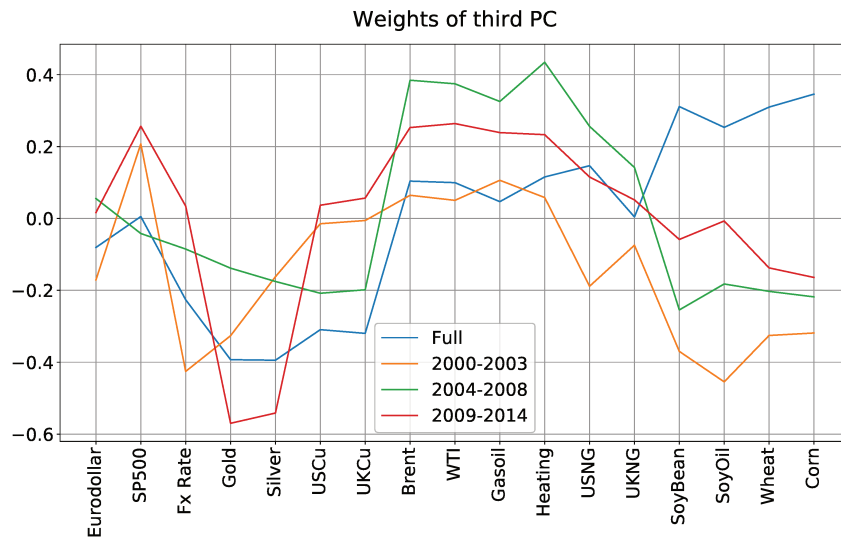


(b) Weights of the first 4 principal components

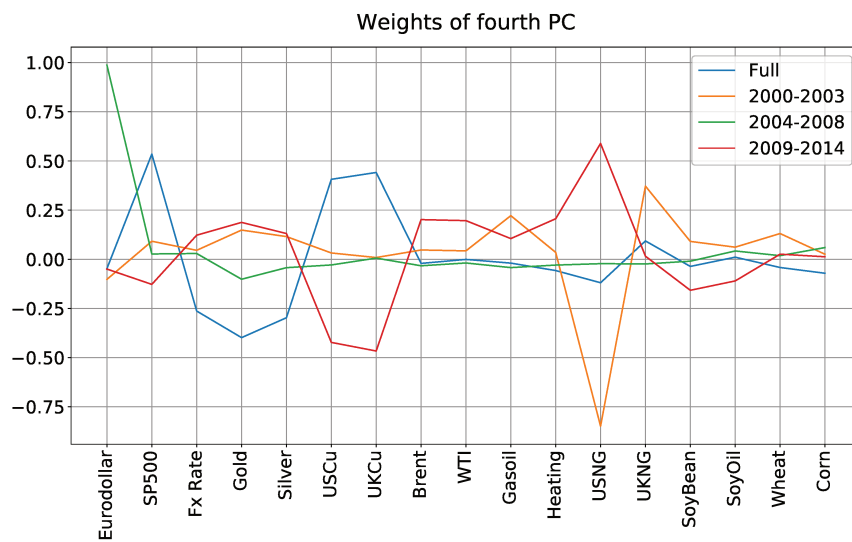


"Full" denotes the weights of the principal components for the PCA performed over the whole sample (2000-2014), others are for the PCA performed on each subsample.

(c) Weights of the third principal component



(d) Weights of the fourth principal component



"Full" denotes the weights of the principal components for the PCA performed over the whole sample (2000-2014), others are for the PCA performed on each subsample.

Chapter 2

Integration of Commodity Derivative Markets: Has It Gone Too Far?

Abstract

We examine the impact of two financial crises on commodity derivative markets: the subprime crisis and the bankruptcy of Lehman Brothers. These crises are "external" to the commodity markets because they occurred in the financial sphere. Still, because commodity markets are now highly integrated with each other and with other financial markets, such events could have had an impact. In order to fully comprehend this possible impact, we rely on tools inspired by the graph theory that allow for the study of large databases. We examine the daily price fluctuations recorded in 14 derivative markets from 2000 to 2009 in three dimensions: the observation time, the space dimension — the same underlying asset can be traded simultaneously in two different places — and the maturity of the transactions. We perform an event study in which we first focus on the efficiency of the price shock's transmission to the commodity markets during the crises. Then we concentrate on whether the paths of shock transmission are modified. Finally, relying on the measure proposed by Bonacich (1987) for social networks, we focus

on whether the centrality of the price system changes.¹

JEL Codes: E44, F15, G01, Q02, Q40

Keywords: Commodity markets, derivatives, market integration, financial crises, graph theory, minimum spanning tree, centrality.

Contents

Introduction	95
2.1 The price system	97
2.1.1 Data	98
2.1.2 Building the graphs	98
2.2 The efficiency of the shock transmission	101
2.2.1 The minimum spanning tree	101
2.2.2 How does the length of the MST behave?	104
2.3 The organization of the tree	108
2.3.1 The survival ratios	108
2.3.2 The allometric coefficients	110
2.4 Examining the centrality of the graphs	112
2.4.1 The degree of the nodes	114
2.4.2 The Katz-Bonacich centrality measure	114
Conclusion	118
References	119
Appendices	121
2.A Timelines around the events	121

¹This paper is a joint work with Franck Raynaud, EPFL and Delphine Lautier, PSL Université Paris-Dauphine. It has been published as a chapter of the book "Commodities, Energy and Environmental Finance", Springer (Fields Institute). It is based on work supported by the Chair Finance and Sustainable Development and the FIME Research Initiative. Comments by the referees and the audience at the 7th Financial Risks International Forum and at the 31st AFFI conference are gratefully acknowledged. The same is true for fruitful remarks from the reviewer of this special volume and from Michel Robe.

2.A.1	Some important events around the subprime crisis	121
2.A.2	Some important events around Lehman Brothers bankruptcy	122
2.B	Additional results	124
2.B.1	Robustness checks	124
2.B.2	Evolution of the markets rankings by centrality, around the events and sector by sector	126

Introduction

In this paper, we examine the impact of two financial crises on the commodity derivative markets: the subprime crisis and the bankruptcy of Lehman Brothers. These crises are exogenous to the commodity markets because they occurred in the financial sphere. Still, such events could have propagated to the commodity markets because these markets are highly integrated with each other and with other financial markets (see Fattouh, Killian and Mahadeva (2013), Büyüksahin, Robe and Bruno (2014), Büyüksahin and Robe (2011), Büyüksahin and Robe (2014), Tang and Xiong (2012), Irwin and Sanders (2011)).

Specifically, in this paper, we analyze the shock transmission through the dynamic behavior of the correlations between price returns. Following Forbes and Rigobon (2002), we consider that there is transmission if market co-movements increase significantly after a shock.

In order to fully comprehend the potential impact of such crises on the commodity derivative markets, we perform an event study in which we examine price fluctuations in three dimensions: the observation time, the space dimension — the same underlying asset can be traded in two exchanges simultaneously — and the maturity of the transactions. We focus on a time window of one month (i.e., ten trading days before and after the beginning of the crises). We situate the triggering event on August 9, 2007 for the subprime crisis and on September 15, 2008 for the Lehman Brothers

bankruptcy (see the Appendices 2.A.1 and 2.A.2 for more details on the chronology of the crises).

Such an analysis requires the use of high dimensional data. In this context, the tools of the graph theory have already proved to be very interesting in various fields of finance. First, they provide a way to synthesize the information contained in the data and to obtain meaningful visual representations, second because they allow for the quantification of high dimensional information (see for instance Onnela et al. (2003), Cohen-Cole et al. (2012), Lautier and Raynaud (2012)). In what follows, we rely mainly on the methodology proposed by Lautier and Raynaud (2012). These authors provide a *long-term* analysis of the connections between 14 derivative markets between 2000 and 2009. They give evidence of an increasing integration along the time period under scrutiny, and they show that it is a condition for systemic risk to appear. Taking advantage of the fact that between 2000 and 2009 two main financial crises occurred, we perform an *event study* on the same markets. This study gives us the possibility to concretely assess the potential consequences of market integration. Moreover, we introduce a new method that was initially proposed by Bonacich (1987) for social networks. This method allows us to better evaluate the organization of the graph. It gives insights into the localization of the center of the graph that, as far as systemic risk is concerned, is crucial.

Following Lautier and Raynaud (2012), the nodes of the graphs correspond to price returns: there is one node per futures contract and per maturity. The link between each pair of nodes depends on the correlations between their returns. Relying on several measures, we provide a dynamic analysis of these graphs and their behavior around the crises. We also empirically compute how exceptional these events are compared to what can be observed in the whole period.

First, in order to filter the information contained in the graphs, we use Minimum Spanning Trees —MST— (Mantegna (1999)). Because they capture the most important links between the markets, they are the most probable and the most efficient paths of price shock transmission. Taking into

account the length of the MST, we can ask a first question: does the efficiency of the price shock transmission improve during crises? We then concentrate on the organization of the graph, namely the topology of the MST and ask a second question: do the paths of shock transmission change during crises and how? In order to answer these questions, several tools are used. First, we use survival ratios that indicate the number of links that change from one day to the other and give indications about large reorganizations of the graphs. Second, the allometric coefficients measure how far a tree stands from a linear or, on the contrary, a star-like organization. These two extreme configurations have radically opposite consequences from the systemic point of view: with a chain-like tree, a shock appearing at one extremity of the tree must spread through all nodes before reaching the other extremity. On the contrary, with a star-like tree, a shock arising at the center of the graph might rapidly affect all other nodes. Finally, we focus on the centrality of the price system: does it change? Does it increase? In a first approach, we simply identify the center of the price system as the most connected node. We then improve this analysis with the measure developed by Bonacich (1987): in a nutshell, instead of focusing on one single node, we take into account the whole organization of the network, that is, the number and proximity of *the direct as well as the indirect neighbors* of a node.

This paper is organized as follows. We first explain how to build a graph on the basis of our data. We then examine the efficiency of the shock transmission, the organization of the price system and its centrality. At each step, we compare the behavior of the price system in the whole period with what happened during the crises.

2.1 The price system

After a short description of the data used for the study, we explain the way we build price graphs.

2.1.1 Data

For the empirical study, we examine 14 futures markets corresponding to three different sectors of activity: 6 energy markets that comprise 2 markets each of crude oil, natural gas and petroleum products; 4 agricultural markets (wheat, corn, soy oil and soy bean) and 4 financial assets (Mini S&P500 index, gold, USD/EUR exchange rate, and 3-month Eurodollar interest rate). We selected the contracts that were characterized by the largest transaction volumes over a long time period, thanks to the Futures Industry Association's monthly volume reports. We used Datastream in order to collect settlement prices on a daily basis.

In the absence of reliable spot data for most commodity markets, we approximated all spot prices with the nearest futures prices. Such an approximation is very common in finance. We also rearranged the futures prices in order to reconstitute the daily term structures, i.e., the relationships linking, at a specific date, several futures contracts with different delivery dates. We removed some maturities from the database because the price curves were shorter at the beginning of the period. The number of contract maturities indeed usually rises on a derivative market; the growth in the transaction volumes of existing contracts results in the introduction of new delivery dates. Finally, when performing spatial and 3D analyses, we used the longest common time period for all of the underlying assets, from 2000/01/04 to 2009/08/12. Once these selections have been carried out, our database still contains more than 655,000 prices, that comprise 220 time series in the 3D analysis and a subset of 14 in the spatial one.

Table 2.1 summarizes the main characteristics of our database.

2.1.2 Building the graphs

Our graphs are built on the basis of the correlations between the price returns. We use this measure in order to capture the synchronous price movements in the system. To obtain a graph, these correlations are transformed into distances.

Table 2.1 – Characteristics of the collected data

Underlying asset	Exchange-Zone	Maturities	# contracts
Light crude oil	CME-US	up to 84	33
Brent crude	ICE-EU	up to 18	17
Heating oil	CME-US	up to 18	18
Gasoil	ICE-EU	up to 12	12
Natural gas (US)	CME-US	up to 36	36
Natural gas (Eu)	ICE-EU	up to 9	9
Wheat	CME-US	up to 15	6
Soy bean	CME-US	up to 14	7
Soy oil	CME-US	up to 15	15
Corn	CME-US	up to 25	4
Eurodollar	CME-US	up to 120	40
Gold	CME-US	up to 60	17
USD/EUR Exchange rate	CME-US	up to 12	4
Mini S&P500	CME-US	up to 6	2

nature of the underlying asset, trading location (CME stands for Chicago Mercantile Exchange, ICE for Inter Continental Exchange, US for United States and EU for Europe), longest maturity traded (in months), number of contracts (this number is added just after the name of the underlying asset in the figures).

Correlations of price returns The first step towards the analysis of market integration is the computation of the synchronous correlation coefficients $\rho_{ij}(t)$ of the price returns, defined as follows:

$$\rho_{ij}(t) = \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\sqrt{(\langle r_i^2 \rangle - \langle r_i \rangle^2) (\langle r_j^2 \rangle - \langle r_j \rangle^2)}}, \quad (2.1)$$

In the spatial dimension, i and j stand for the nearby futures contracts of a pair of assets (crude oil or corn for example), whereas they stand for pairs of delivery dates in the maturity dimension. Both are present in the 3D analysis with the 220 time series. The daily logarithm price differential stands for the price returns r_i , with $r_i = (\ln F_i(t) - \ln F_i(t - \Delta t)) / \Delta t$, where $F_i(t)$ is the price of the futures contract i at date t . The time interval is Δt and $\langle \cdot \rangle$ denotes the statistical average performed over time, for the trading days of the study period.

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

Performing dynamic studies on the basis of rolling windows requires the choice of a proper window length. On the one hand, we want it to represent typical economic periods (one semester, one year, five years...) and to be as short as possible in order to give evidence of sudden changes. On the other hand, we are confronted with a technical constraint: in order to ensure representative results, the number of observations has to be larger than the number of nodes. Having to deal with 220 series of price returns (i.e., 220 nodes), we thus use a rolling window of one year (252 trading days). We do the same in the spatial dimension for comparison purposes. As robustness checks, we also perform computations with 2-year windows, as illustrated in the appendix 2.B.1. Further, we use rolling windows situated before the observation date. So when we look at what happens on August 9, 2007, the information used is situated one year before that event. Fortunately, because the two crises are separated by more than one year, there is no overlap between them.

From correlations to distances In order to use the tools of the graph theory, we need to introduce a metric. The correlation coefficient ρ_{ij} cannot be used as a distance d_{ij} between i and j because it does not fulfill the three axioms that define a metric (Fréchet (1906), p30):

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

However, a metric d_{ij} can be extracted from the correlation coefficients through a nonlinear transformation. This Euclidean distance is defined as

follows²:

$$d_{ij}(t) = \sqrt{2(1 - \rho_{ij}(t))}. \quad (2.2)$$

A distance matrix \mathbf{D} is thus extracted from each correlation matrix \mathbf{C} (at each date t) according to the Eq. (2.2). The matrices \mathbf{C} and \mathbf{D} are both $N \times N$ dimensional. While the coefficients $\rho_{ij}(t)$ can be positive for the correlated returns or negative for the anti-correlated returns, the distance $d_{ij}(t)$ is always positive. The distance matrix corresponds to a fully connected graph; it represents all the possible connections in the price system.

2.2 The efficiency of the shock transmission

Considering the dimensionality of our price system and the number of nodes in our graph, it is very difficult to visualize. We thus resort to a filtering technique which is especially suited to our context: the Minimum Spanning Tree (MST).

2.2.1 The minimum spanning tree

In order to understand the organizing principles of a system through its representation as a graph, the latter needs to be spanned. However, there are a lot of paths that span a graph. For a weighted graph like ours, the MST divulges the most relevant connections of each element of the system and it reduces the information space from $N(N - 1)/2$ to $N - 1$.

The MST is the path spanning all the nodes of the graph without any loop. It has less weight than any other tree and is unique. The distance $d_{ij}(t)$ is more than just an Euclidean metric; it is an ultrametric that satisfies the triangular inequality: $d_{ij}(t) \leq \max \{d_{ik}(t); d_{kj}(t)\}$.

When the graph is weighted with distances, the latter corresponding to the correlations between the price returns, the MST is especially useful for the study of systemic risk. In an analogy with signal transmission, the ultrametric provides the shortest path between all of the nodes, that is,

²Taking the square of $\rho_{ij}(t)$ has no impact on the results (computations are available on request)

the path where the signal suffers the least losses and travels the fastest. We interpret this feature as the efficiency (in speed and in accuracy) in the transmission of the signal. Furthermore, if a price shock is assimilated to a signal and if transmission is appreciated through the analysis of the dynamic behavior of the correlations between the price returns, then the MST "can be assimilated into the shortest and most probable path for the propagation of price shocks" (Lautier and Raynaud (2012)).

The visualization of the trees (which are plotted with the software Graphviz) addresses the meaningfulness of the taxonomy that emerges from the system. Because we are considering the links between markets and/or delivery dates belonging to the MST, if a link between two markets or maturities does not appear in the tree, it only means that this link does not correspond to a minimal distance. Note also that, in such an analysis, the results depend on the nature and the number of markets chosen for the study.

Figure 2.1 presents the MST obtained on the basis of our price system for the spatial dimension and over the whole period. It is scaled: the closest nodes correspond to the most correlated price series. Three sectors can be identified: energy is in the top left-hand. It gathers American as well as European markets and is situated between agriculture (on the right) and financial assets (at the bottom).

The link between the energy and agricultural products passes through soy oil. This is interesting because soy oil can be used for fuel. The link between commodities and financial assets passes through gold, which is also meaningful, because gold can be seen as a commodity as well as a reserve of value. The only surprise comes from the Mini S&P500 that is more correlated to soy oil than to financial assets. This connection between the Mini S&P500 and agricultural markets could be interpreted as evidence of the financialization of the commodity markets. However, in a dynamic analysis, this connection is very unstable. At least two reasons could explain such a result: first, Büyükaşahin, Robe and Bruno (2014) find that the correlations

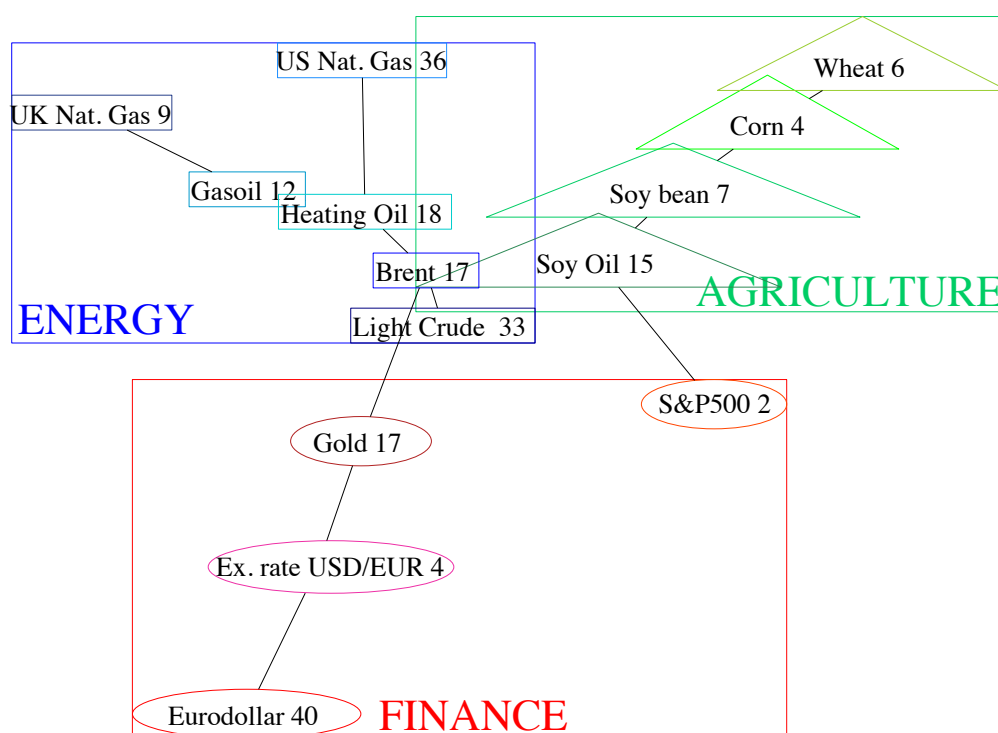


Figure 2.1 – Scaled MST in the spatial dimension, 2000-2009.

between grains and equities fluctuate a lot; and second, compared to all other contracts taken into account, the Mini S&P500 is the least actively traded.

At first glance (if we accept that counting the number of links allows for the identification of the center of the graph) the most connected node is the one corresponding to Brent crude oil, which makes it — a priori — the best candidate for the transmission of price fluctuations in the tree (actually, the same could be said for American crude oil — Light Crude — because 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.

Such a star-like organization leads to specific conclusions regarding systemic risk. A price movement appearing in the energy markets, situated at the heart of the price system, will have more impact than a fluctuation affecting

the peripheral markets such as interest rates or wheat. This configuration explains why we consider the subprime and the Lehman Brothers crises as exogenous events in this study.

The 3D MST comprises 220 time series (nodes). Depicted by Fig. 2.2, it is less easy to read (this is why we removed the captions in the nodes), but it can be interpreted through the prism of the spatial tree. The same topology prevails, except that adding the maturities introduces linear branches in each market (with the noticeable exception of American natural gas). Moreover, this scaled representation shows that some markets are more integrated than others: clusters of maturities can be seen, at the center of the graph, for the energy sector (except for the two natural gas markets). Strong integration can also be observed in the financial branch; this is especially true for the Eurodollar contract after the eighth maturity.

Because these topologies are very stable over time (Lautier and Raynaud (2012)), we use them as references in the remainder of this study.

2.2.2 How does the length of the MST behave?

We first explain how this measure can be obtained and how it behaves on the whole sample. We then study it around crises.

2.2.2.1 The measure

The normalized length of the tree can be defined as the average of the lengths of the edges belonging to the MST:

$$\mathcal{L}(t) = \frac{1}{N-1} \sum_{(i,j) \in MST} d_{ij}(t), \quad (2.3)$$

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 higher when the distances increase and consequently when correlations are low. Thus, the more the length diminishes, the more integrated the system is.

Figure 2.3 represents the dynamic behavior of the normalized length of the MST in the spatial dimension over the whole period under consideration.

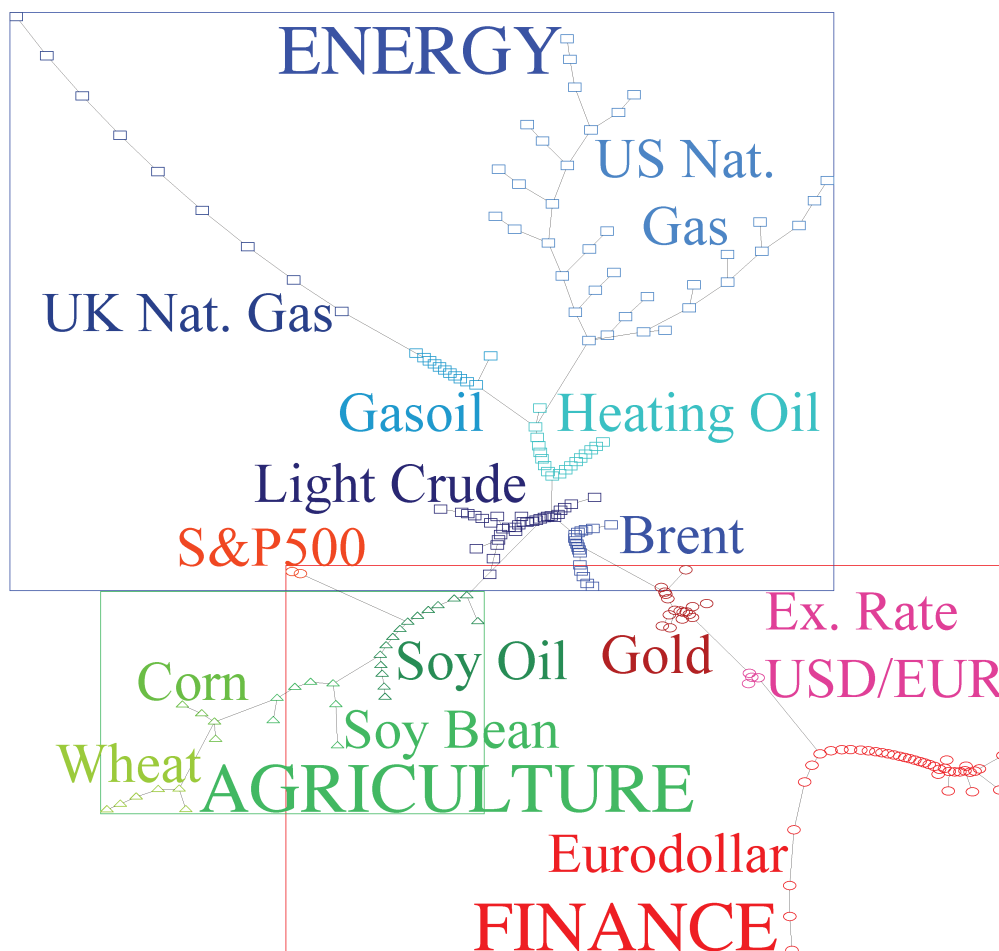


Figure 2.2 – Scaled MST in 3D, 2000-2009.

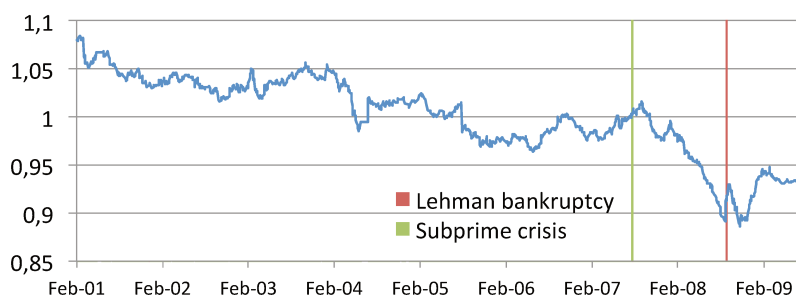


Figure 2.3 – Normalized tree's length in the spatial dimension, 2000-2009.

The general pattern is that the length decreases, which reflects the increasing integration of the system. Thus the most efficient transmission path for price fluctuations becomes shorter as time goes by. This finding is consistent with e.g., **SilvennoinenThorp2013** and Tang and Xiong (2012). A more in-depth examination of the graph also shows some very important moves at specific dates, one of them being around the Lehman Brothers bankruptcy.

2.2.2.2 The length of the trees around the crises

A first appraisal of the importance of the crises consists in measuring whether the changes in the length of the MST that occurred around the events were tail events or not.

We compute the empirical distribution of the length variations over the whole sample and examine the probability of the occurrence of fluctuations situated above (for increases) or below (for decreases) those observed around the crises. At 5%, the changes recorded on August 16, 2007 (five trading days after the beginning of the subprime crisis) and on September 12, 2008 (one trading day before the bankruptcy of Lehman Brothers) are in the tail of the distribution, both in the spatial dimension and in 3D. In the spatial dimension only, we can add August 14, 2007, and in 3D only September 17, 2008. These last two events and the one recorded on September 12, 2008 have a probability of occurrence that is close to 1%. Consequently, compared to what was observed between 2000 and 2009, the two crises have generated exceptional changes in the length of the MST.

A recurrent result in finance is the observation of an increase in the price correlations just after a crisis (see, e.g., Chakraborti et al. (2003) for an analysis of the equity market around Black Monday on October 19, 1987, Büyüksahin and Robe (2014) and Tang and Xiong (2012) for commodity-equity markets, or Pesaran and Pick (2007) for a review of several studies on these topics). Figures 2.4 (a) to 2.4 (d), which represent the evolution of the length of the trees on a 1-month time window around the crises under

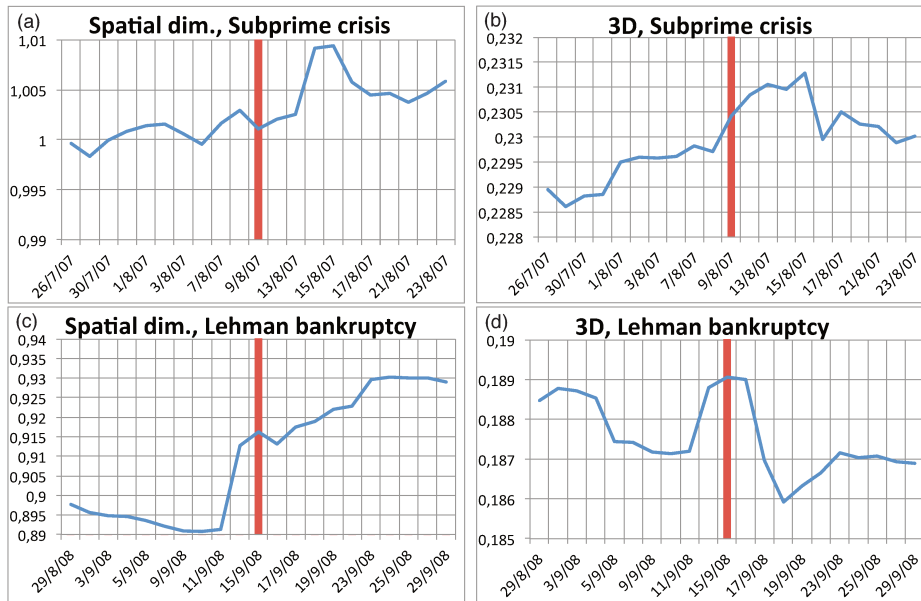


Figure 2.4 – Normalized tree’s length in the spatial dimension and in 3D for each event.

consideration both in the spatial dimension and in 3D, do not exhibit such behavior. On the contrary, in three cases (subsets (a), (b) and (c)) out of four, we find an increase in the length of the MST. For the subprime crisis, the peak appears on August 15, 2008, four trading days after the beginning of the crisis. For the Lehman Brothers bankruptcy, the change in the behavior of the tree arrives before the event, between September 11 and 12 of 2008. These dates correspond to the period when the difficulties encountered by the bank became public knowledge (see the Appendices 2.A.1 and 2.A.2).

However, this increase in the global length of the MST comes with a decrease in certain subsets of the trees. This is especially the case for the Eurodollar market around the Lehman Brothers bankruptcy as shown by the scaled MST on Fig. 2.5, where there clearly is a shrink in the trees around the two crises. Such a result is reasonable: first because the 3-month interest rate is a pure financial asset and second because what we observe here is a branch of the tree where only the maturity dimension is taken into account. As mentioned by Lautier and Raynaud (2012), under

the pressure of arbitrage operations, the markets are more integrated in the maturity dimension than in the spatial one.

The analysis of the length of the trees shows that, even if our price system becomes more and more integrated between 2000 and 2009, these two crises, born in the financial sphere, did not harm the commodity markets as a whole. This conclusion is consistent with the findings of Büyüksahin and Robe (2011) who observed that the links between the equity index and the energy futures is weaker in times of crises or of Corsetti, Pericoli and Sbracia (2005) who find that correlations decreases in some episodes of crisis. As expected, these crises had an impact on the financial sphere: there is a local increase in the integration of the futures contracts written on the financial assets. However, as far as commodity markets are concerned, they became temporarily less connected with the financial assets.

2.3 The organization of the tree

Measuring the length of the MST does not give the possibility to ask whether or not the paths for shock transmission change during the crises. In order to answer this question, the graph theory provides several tools: first the survival ratios and second the allometric coefficients.

2.3.1 The survival ratios

This measure (S_R) indicates the fraction of links that survives, in the MST, between two consecutive trading days (Chakraborti et al. (2003)):

$$S_R(t) = \frac{1}{N-1} |E(t) \cap E(t-1)|. \quad (2.4)$$

In this equation, $E(t)$ refers to the set of the tree edges at date t , \cap is the intersection operator and $|\cdot|$ gives the number of elements contained in the set. Due to the finite number of links, the ratios take discrete values. The use of this measure naturally raises the same question as before: how exceptional are the values of the survival ratios observed around the crises?

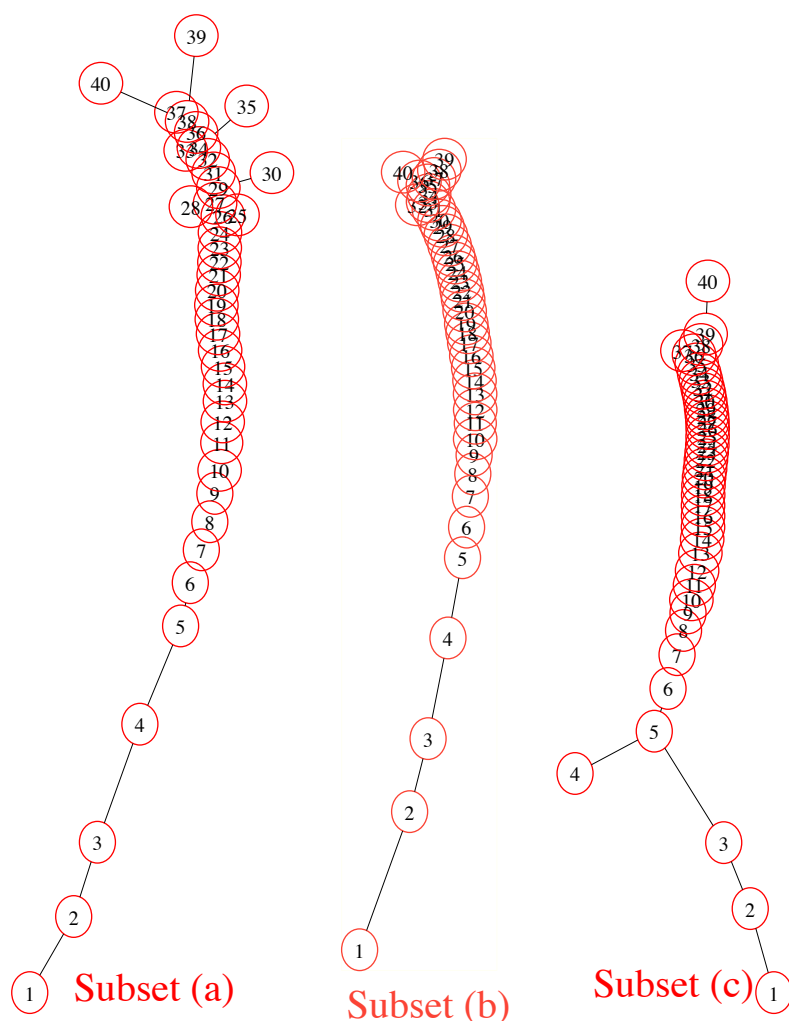


Figure 2.5 – Scaled MST in the maturity dimension, Eurodollar market.

Subset (a): 2000-2009 ; subset (b): 1-month time window including the subprime crisis; subset (c): 1-month time window including the Lehman Brothers bankruptcy.

As before, we evaluate the probability of the occurrence of high reconfigurations in the graph. We find that only the changes recorded on September 18 and 19 of 2008 (the 17th is close) are below the 5% probability of occurrence in the spatial dimension. In 3D, only September 17 and 24 of 2008 appear below the 5% threshold. According to these figures, the subprime crisis shows nothing specific: even if, as shown by the length of the MST,

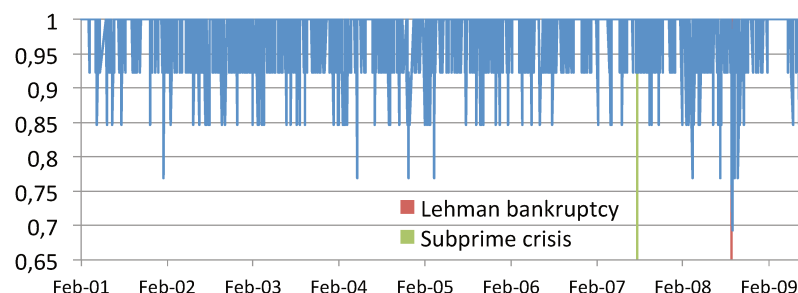


Figure 2.6 – Survival ratios in the spatial dimension, 2000-2009.

the trees locally shrink in the financial sphere on this occasion, the path of the price shock transmission remains the same.

This result is confirmed by Fig. 2.6. The figure shows first that under normal circumstances, the topology of the trees is very stable between two dates, in the spatial dimension as well as in 3D: most of the time, between 2000 and 2009, more than 85% of the links remain unchanged from one day to the next. Second, nothing special happens around the subprime crisis. This is far less obvious for the Lehman Brothers bankruptcy. In this case, the most important reorganizations appear in the spatial dimension, where more than 30% of the graph is reorganized.

Finally, 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 might simply be due to noise. This is why a deeper analysis is needed. We will perform it through the use of allometric coefficients.

2.3.2 The allometric coefficients

The computation of the allometric coefficients of a MST quantifies where this tree stands between two asymptotic topologies: star-like trees and chain-like trees. These two topologies have very different implications for systemic risk.

The first model of the allometric scaling on a spanning tree was developed by Banavar, Maritan and Rinaldo (1999). In their method, the first step consists in assigning a value A_i equal to 1 to each node i . Then the root

(also called the central node) of the graph must be identified. In what follows, the root is determined with Bonacich's measure defined in section 2.4. As a robustness check, we perform the same tests with a root identified as the node with the highest number of links. The results remain qualitatively the same and are available on request

Starting from the root, the second step of the method consists in updating the coefficients A_i and in assigning the coefficients B_i of each node i as follows:

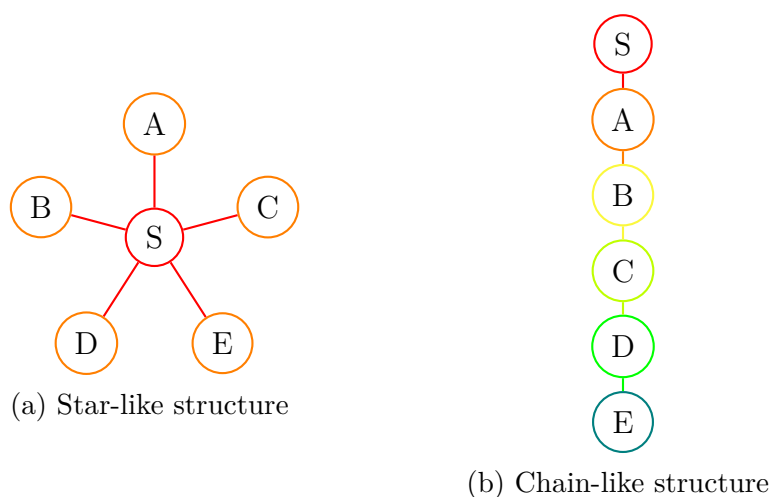
$$A_i = \sum_j A_j + 1 \text{ and } B_i = \sum_j B_j + A_i, \quad (2.5)$$

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

$$B \sim A^\eta, \quad (2.6)$$

where η is the allometric exponent. It represents the degree or complexity of the tree and stands between two extreme values: 1^+ for star-like trees (Fig. 2.7a) and 2^- for chain-like trees (Fig. 2.7b).

Figure 2.7 – Extreme types of structures



A MST belonging to the first or to the second structure will not have the same implications in terms of shock transmission. One way to explain such

an interpretation is to rely once again on the analogy with the transmission of a signal in a network. Let us assume that a signal is transmitted in each network represented by Figs. 2.7a and 2.7b. In each case, the signal is transmitted from node S at time t and there is some latency in the transmission. In the star-like tree, all of the others nodes (A, B, C, D and E) will receive the transmission simultaneously at time $t+1$. Comparatively, in the chain-like tree, the first receiver is node A, the second is node B, etc. In such a topology with N nodes, it takes $N - 1$ time periods (i.e., five in the Fig. 2.7b) before reaching the end of the network. Meanwhile, if there is noise in the transmission channel, the signal will suffer some losses. In our case, where the distances in the networks stand for correlations between price returns, a price shock emerging at node S will spread more efficiently if the structure of the tree is star-like, because it will more quickly reach all of the other nodes. It is thus crucial to correctly identify the center of the graph.

Relying on the allometric coefficients, Lautier and Raynaud (2012) show that: i) the MST are almost linear in the maturity dimension of most markets, ii) they stand right in the middle of the two extreme configurations in the spatial dimension at 1.5, and iii) the allometric coefficients are around 1.75 in the 3D case. Around the crises, as shown by Fig. 2.8, the levels of the allometric coefficients remain the same. Moreover, their variations are *not* exceptional at 5% except those recorded in 3D on September 2, 2008 and on September 29, 2008, around the bankruptcy of Lehman Brothers.

2.4 Examining the centrality of the graphs

When studying systemic risk, it is important to correctly detect the center of the trees. For regulatory authorities, such nodes can be assimilated to regions of higher fragility. Even though we examine exogenous events in this study, the question of centrality remains crucial. What if these events create shocks that reach the center of the graph? They would then spread rapidly to all of the other markets, as noted in the above subsection.

The most common way to identify the center of a graph is to assess the de-

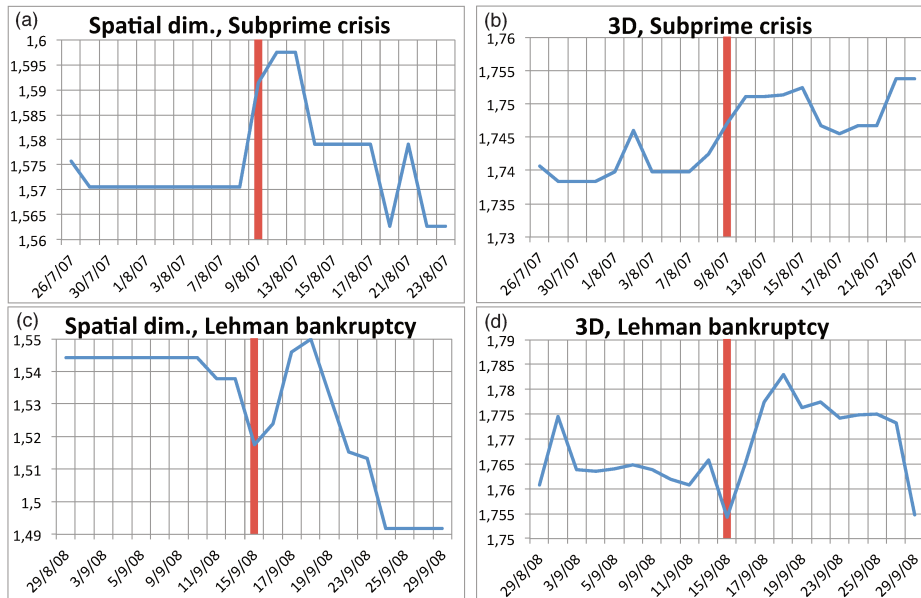


Figure 2.8 – Allometric coefficients, in the spatial dimension and in 3D, for each event.

gree (i.e., the number of links) of each node in the trees. However, such an analysis might be insufficient: first because it does not take into consideration the distances between the nodes, and second because it only accounts for the direct neighbors of a node. It could be interesting, on the contrary, to be mindful of the overall configuration of the graph.

In what follows, we first give an example of an analysis based on degree only: we focus on the evolution of the trees in the spatial dimension around the Lehman Brothers bankruptcy. As noted on the basis of the survival ratios, there is indeed an important reconfiguration of the graph on this occasion. Then, we propose the use of a new measure of centrality that was introduced by Bonacich in 1987 for social networks and recently used by Bloch and Qu erou (2013). To the best of our knowledge, this is the first time that such a method has been employed in finance.

2.4.1 The degree of the nodes

The scaled MST in the spatial dimension at the Lehman Brothers bankruptcy is depicted by Fig. 2.9. If we compare this tree with the one computed for the whole period as illustrated by Fig. 2.1 (as shown in subsection 2.3.1 the MST is very stable; the tree computed on the whole period can thus be taken as a reference) then we can see some changes: the Mini S&P500 is not linked to soy oil anymore, but now to wheat; the UK natural gas is not directly connected to the energy sector anymore; and, more importantly, gold now stands at the center of the graph. From an economic point of view, such a result is very reasonable. In a situation where high uncertainty affects the whole financial system, we indeed expect investors to consider gold as a reserve of value. Yet the story is not so simple.

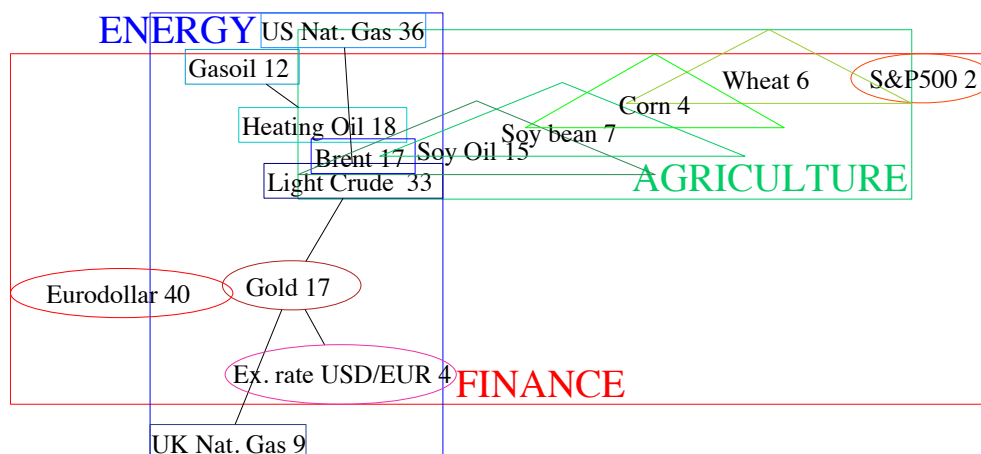


Figure 2.9 – Scaled MST in the spatial dimension at Lehman Brothers bankruptcy (September 15, 2008).

2.4.2 The Katz-Bonacich centrality measure

We first present the method and its advantages. Then we use it for the event study.

2.4.2.1 The method

The Katz-Bonacich centrality measure aims at taking into consideration the whole configuration of a graph, that is, the direct as well as the indirect neighbors of a node. Looking only at the direct neighbors, as done when one relies on the degree, might be insufficient as illustrated by Fig. 2.10: the node labelled A (or B or C) exhibits the highest degree (four in this case). However, the S node is obviously the most central one.

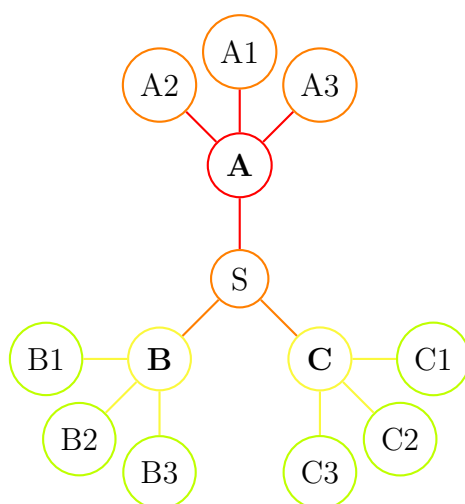


Figure 2.10 – Connectivity (A, B, or C) versus centrality (S).

The measure proposed by Bonacich (1987) is an extension of the one developed by Katz (1953). This author was the first to pay attention to the indirect neighbors of a node. In addition, the measure developed by Bonacich (1987) gives the possibility of taking into account the "negative" relations, *i.e.* the fact that, if the value of a node increases, then its neighbors' value decreases.

The centrality vector, which gives one value per node, is computed as follows:

$$c(\alpha, \beta) = \alpha (\mathbf{I} - \beta \mathbf{R})^{-1} \mathbf{R} \mathbf{1}$$

where \mathbf{I} is the identity square matrix, \mathbf{R} is the matrix of the weights of the graph, and $\mathbf{1}$ a vector of 1s. The coefficient α is a scale factor. According to Bonacich, the coefficient β can be interpreted in different ways: "the

degree to which an individual's status is a function of the statuses of those to whom he or she is connected" or "a radius within which the researcher wishes to assess centrality". Note also that the centrality values are sensitive to both the weights of the graph and its topology. Since these values take into account infinitely far neighbors, a small change in the topology of the graph can result in large changes in the centrality values.

The use of this relationship matrix requires first a measure of similarity: the quantities in \mathbf{R} must be such that, the higher the β , the easier the transmission. A second requirement is that all R_{ij} are positive. Third the R_{ii} must be equal to zero. To fulfill the first requirement, we use the correlation matrix for \mathbf{R} . More precisely, because we are interested in the central node of the MST, we consider the prices correlations in the MST, and we compute \mathbf{R} as follows:

$$R_{ij}(t) = C_{ij}(t) * E_{ij}^{MST}(t),$$

where $\mathbf{C}(t)$ is the correlation matrix and $\mathbf{E}^{MST}(t)$ is the edge matrix of the tree; $E_{ij}^{MST}(t)$ equals to one if there is a link between i and j in the MST and zero otherwise. This matrix is symmetric, with $N - 1$ ones.

The use of the filtered correlation matrix for \mathbf{R} simplifies the application of the method developed by Bonacich. This matrix can be directly identified to \mathbf{R} , because it fits all of the requirements. Moreover, such a choice leads to more precise results, because it allows for taking into account the specific value of each link instead of averaging them into a β coefficient (which we thus drop).

2.4.2.2 Empirical results

For comparison purposes, it is interesting to go back to the scaled MST in the spatial dimension commented on in section 2.2 and represented by Fig. 2.1. When relying on the degrees of the nodes, the root of the tree corresponds to crude oil. However, taking into account the overall organization of the tree leads to a conclusion that is more nuanced. Table 2.2 presents the results of the method when it is applied in the spatial dimension between 2000 and 2009. Relying on the centrality measures leads to

putting more emphasis on both heating oil and crude oil; the heating oil is ranked first. Moreover, a dynamic analysis shows that, especially after August 17, 2005, the agricultural markets play a more important role. This result calls for further analysis, but it is probably due to the introduction of the rules regarding bioethanol in the United States in 2005. Second, half of the markets under consideration in the spatial analysis never reach a centrality value above 1: this is true for the 3-month eurodollar, the USD/EUR exchange rate, the Mini S&P500 index, gold, gasoil and for the US and UK natural gases. These markets thus have a centrality that is unusually low and are hence less important.

Table 2.2 – Bonacich’s centrality measure in the spatial dimension, 2000-2009.

Market	Centrality measure	Rank
Heating Oil	1.148228	1
Brent	1.108484	2
Light Crude	0.856703	3
Gasoil	0.591487	4
US Natural Gas	0.364067	5
Gold	0.231502	6
USD/EUR Exchange Rate	0.036973	7
UK Natural gas	0.034241	8
Eurodollar	-0.00875	9
Mini S&P500	-0.189855	10
Wheat	-1.144788	11
Soy Oil	-1.159204	12
Corn	-1.890017	13
Soy Bean	-1.979338	14

The results associated with the centrality measures around the crises are depicted, for the spatial dimension, in Appendices 2.B.2.1 and 2.B.2.2. Once again, the subprime crisis does not affect the organization of the trees, whereas the Lehman Brothers bankruptcy has an impact (mostly temporary, though). Around this event, the ranking of the nodes puts light crude oil first, gold second and heating oil third.

In 3D, the most central nodes are about the same as in the spatial dimension. Due to the large number of nodes (220), we cannot display the tables in this case but the results are available on request. As before, we do not find many changes around the subprime crisis and many more around the Lehman Brothers bankruptcy.

Finally, the most interesting phenomena appear in the maturity dimension around the Lehman Brothers bankruptcy. There are some changes in the direction of certain propagation paths. The most illustrative example of such behavior is that of light crude on September 10, 2008: before that date, many short-term maturities of light crude oil are among the most central nodes of the tree (they are situated above the rank of 20 according to the centrality measure), while most of the long-term maturities are among the least central (below the rank 200). From one day to the next, however, there is an inversion: the least central nodes become the most central ones (they even reach the rank of one) while the previously most central ones go as low as rank 220. Finally, things revert back to the initial state.

Conclusion

For a decade, commodity derivative markets have been experiencing a process of financialization due to managers seeking the diversification of their portfolios and to the arrival of new actors. This phenomenon has raised questions and worries about the eventuality of meaningless links, from an economic point of view, between commodities and more traditional financial markets like bonds and stocks. These fears have been largely confirmed by the acknowledgment of a growing integration between commodity markets as well as between commodities and other financial assets. One could wonder to what extent a shock originating from financial markets could propagate to commodities and strongly impact them. Investigating such a question is the purpose of this paper.

To this aim we examine the impact, on commodity markets of two recent

financial crises: the Subprime crisis and the Lehman Brothers bankruptcy. Using the insightful tools of the graph theory, on the basis of several measures, we show that those shocks did not affect the commodity markets as hard as one might have expected.

References

- Banavar, Jayanth R., Amos Maritan and Andrea Rinaldo (1999). ‘Size and form in efficient transportation networks’. In: *Nature* 399(6732), pp. 130–132.
- Bloch, Francis and Nicolas Qu  rou (2013). ‘Pricing in social networks’. In: *Games and Economic Behavior*.
- Bonacich, Phillip (1987). ‘Power and Centrality: A Family of Measures’. In: *American Journal of Sociology* 92(5), pp. 1170–1182.
- Brunnermeier, Markus K. (2009). ‘Deciphering the Liquidity and Credit Crunch 2007–2008’. In: *Journal of Economic Perspectives* 23(1), pp. 77–100.
- B  y  k  ahin, Bahattin and Michel A. Robe (2011). ‘Does Paper Oil Matter? Energy Markets’ Financialization and Equity-Commodity Co-Movements’. In: *Working Paper, American University, (Revised) July*.
- B  y  k  ahin, Bahattin and Michel A. Robe (2014). ‘Speculators, commodities and cross-market linkages’. In: *Journal of International Money and Finance* 42, pp. 38–70.
- B  y  k  ahin, Bahattin, Michel A. Robe and Valentina G. Bruno (2014). ‘The Financialization of Food?’ In: *Working Paper*.
- Chakraborti, Anirban et al. (2003). ‘Dynamic asset trees and Black Monday’. In: *Physica A* 324(1-2), pp. 247–252.
- Cohen-Cole, Ethan et al. (2012). ‘Strategic Interactions on Financial Networks for the Analysis of Systemic Risk’. In: *Handbook on Systemic Risk*, p. 306.
- Corsetti, Giancarlo, Marcello Pericoli and Massimo Sbracia (2005). ‘Some contagion, some interdependence’: More pitfalls in tests of financial con-

- tagion'. In: *Journal of International Money and Finance* 24(8), pp. 1177–1199.
- Fattouh, Bassam, Lutz Killian and Lavan Mahadeva (2013). 'The Role of Speculation in Oil Markets: What Have We Learned So Far?' In: *The Energy Journal* 34(3).
- Forbes, Kristin J. and Roberto Rigobon (2002). 'No Contagion, Only Interdependence: Measuring Stock Market Comovements'. In: *The Journal of Finance* 57(5), pp. 2223–2261.
- Fréchet, Maurice (1906). 'Sur quelques points du calcul fonctionnel'. In: *Rendiconti del Circolo Matematico di Palermo* 22, pp. 1–74.
- Irwin, Scott H and Dwight R Sanders (2011). 'Index funds, financialization, and commodity futures markets'. In: *Applied Economic Perspectives and Policy* 33(1), pp. 1–31.
- Katz, Leo (1953). 'A New Status Index Derived From Sociometric Analysis'. In: *Psychometrika* 18(1), pp. 39–43.
- Lautier, Delphine and Franck Raynaud (2012). 'Systemic Risk in Energy Derivative Markets: A Graph-Theory Analysis'. In: *The Energy Journal* 33(3), pp. 215–239.
- Mantegna, R. N. (1999). 'Hierarchical structure in financial markets'. In: *The European Physical Journal B* 11, pp. 193–197.
- Onnela, J-P et al. (2003). 'Dynamics of market correlations: Taxonomy and portfolio analysis'. In: *Physical Review E* 68(5), p. 056110.
- Pesaran, M Hashem and Andreas Pick (2007). 'Econometric issues in the analysis of contagion'. In: *Journal of Economic Dynamics and Control* 31(4), pp. 1245–1277.
- Tang, Ke and Wei Xiong (2012). 'Index Investing and the Financialization of Commodities'. In: *Financial Analysts Journal* 68(6), pp. 54–74.

Appendices

2.A Timelines around the events

2.A.1 Some important events around the subprime crisis

Based on Brunnermeier (2009), News feeds, Wikipedia

Table 2.3 – Important events around the subprime crisis (S denotes the date of the trigger of the crisis, on August 9th, 2008)

Trading date	Calendar date	Events
S-10	2007-07-26	Home sales declined and largest home builder reported loss
S-7	2007-07-31	American Home Mortgage Investment Corp. faces difficulties
S-6 to S	2007-08-01 - 2007-08-09	Quantitative hedge funds suffered losses that trigger margin calls, fire sales, and correlation across strategies
S-6	2007-08-01	US Crude oil prices reach a new high due to declining stocks and decreased output
S-4	2007-08-03	Officials state that the housing crisis should not spread
S-3	2007-08-06	America Home Mortgage Investment Corp. goes bankrupt
S	2007-08-09	BNP Paribas froze redemption of 3 of its investment funds due to inability to value structured products Triggered the first illiquidity wave on the interbank market and support from Central Banks
S+1	2007-08-10	Decreases propagate to Asian markets, triggering support from Central Banks
S+2 S+8	to 2007-08-13 - 2007-08-21	Central Banks increase their support and lower rates

2.A.2 Some important events around Lehman Brothers bankruptcy

Based on Brunnermeier (2009), News feeds, Wikipedia

Table 2.4 – Important events around Lehman Brothers bankruptcy (L denotes the date of Lehman Brothers default, on September 15, 2008)

Trading date	Calendar date	Events
L-6	2008-09-05	US Government's plan to bail out Fannie Mae and Freddie Mac leaks
L-3	2008-09-10	OPEC will cut oil production by 500,000 barrels a day Announcement of the worst losses of Lehman
L-1	2008-09-12	The Federal Reserve tries to find buyers for Lehman and warns CME of a potential default
L	2008-09-15	Lehman files for bankruptcy in the morning, because of lack of buyers and of bail out Merrill Lynch is sold to Bank of America
L+1	2008-09-16	AIG is bailed out
L+2	2008-09-17	Russia helps its biggest banks
L+3	2008-09-18	Russia extends help Lloyds TSB purchases HBOS, largely exposed to subprime mortgages
L+4	2008-09-19	The Troubled Asset Relief Program leaks US Treasury guarantees money market mutual funds up to \$50 billion Nigerian oil production is cut by 280,000 barrels per day and a pipeline of Royal Dutch Shell was destroyed
L+5	2008-09-22	G7 commits to protect the financial system
L+9	2008-09-26	The Federal Deposit Insurance Corporation seizes Washington Mutual to sell it to JP-Morgan Chase

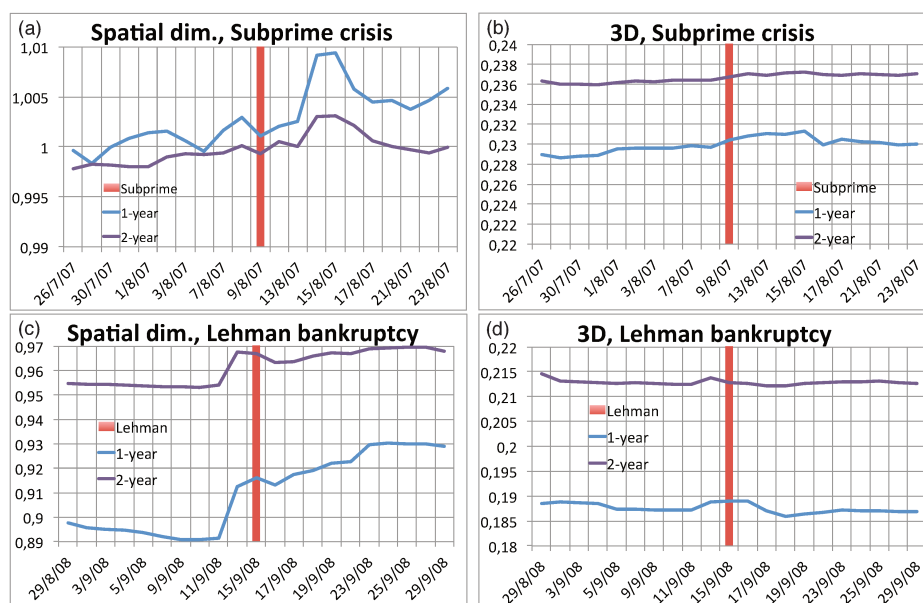


Figure 2.11 – Normalized length in the spatial dimension and in 3D, around the events

2.B Additional results

2.B.1 Robustness checks

This section of the appendix is devoted to a sensitivity analysis. It provides the results obtained around the two events, with the different measures used in the analysis (length of the MST, survival ratios and allometric coefficients) when the rolling window is extended to two years instead of one year. The comparison shows that overall, the behavior remains qualitatively the same. As expected, compared with the 1-year rolling window, the two-year window has a smoothing effect.

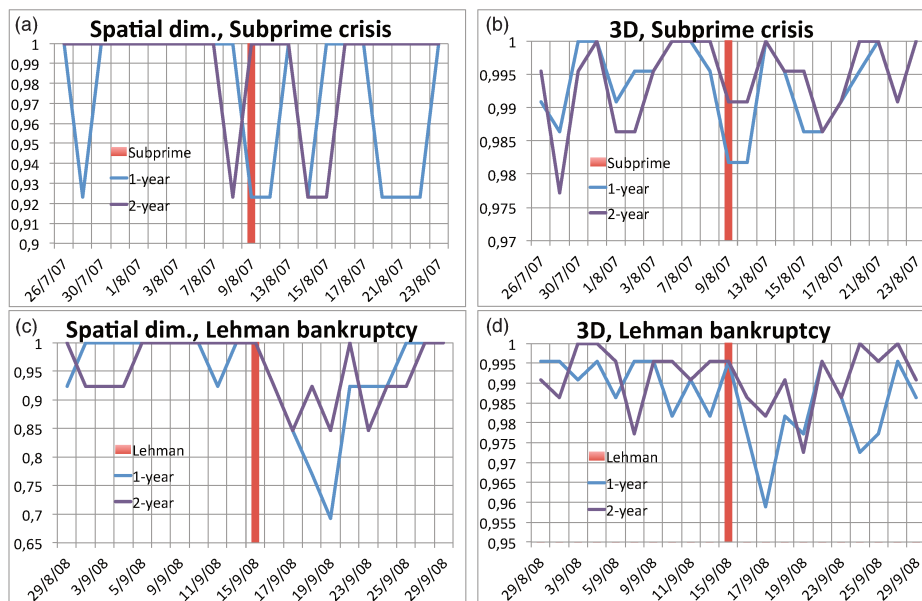


Figure 2.12 – Survival ratios in the spatial dimension and in 3D, around the events

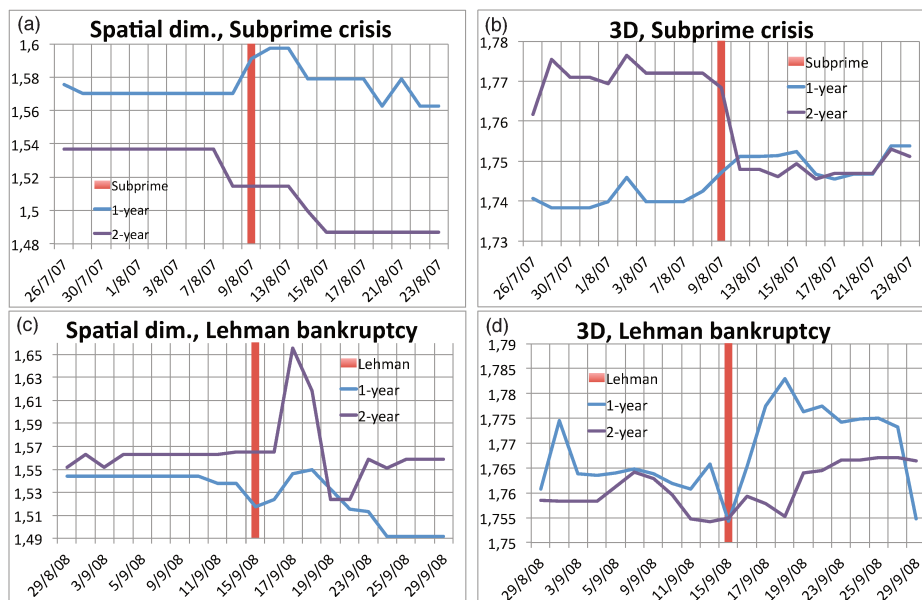


Figure 2.13 – Allometric coefficients in the spatial dimension and in 3D, around the events

**2.B.2 Evolution of the markets rankings by
centrality, around the events and sector by
sector**

**2.B.2.1 Ranking by centrality measure in the spatial dimension,
around the Subprime crisis (August 9, 2007)**

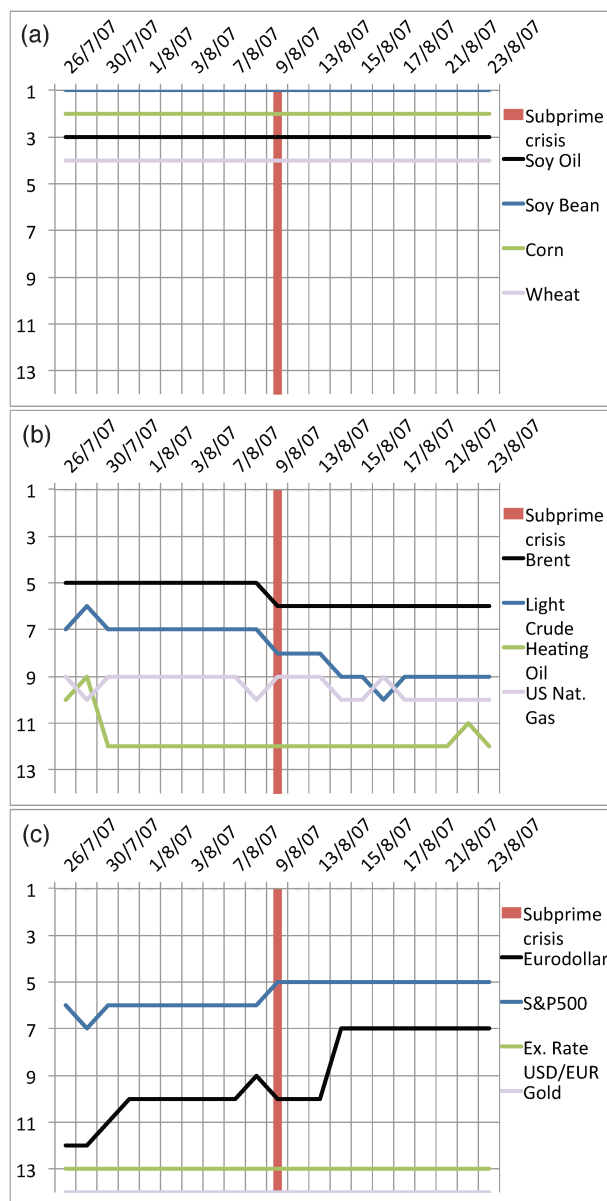


Figure 2.14 – Evolution of the ranks in the spatial dimension around the Subprime crisis.

Subset (a) is for agricultural markets, subset (b) is for only 4 energy markets (for readability) and subset (c) is for financial markets.

2.B.2.2 Rankings by centrality measure in the spatial
dimension, around the Lehman Brothers bankruptcy
(September 15, 2008)

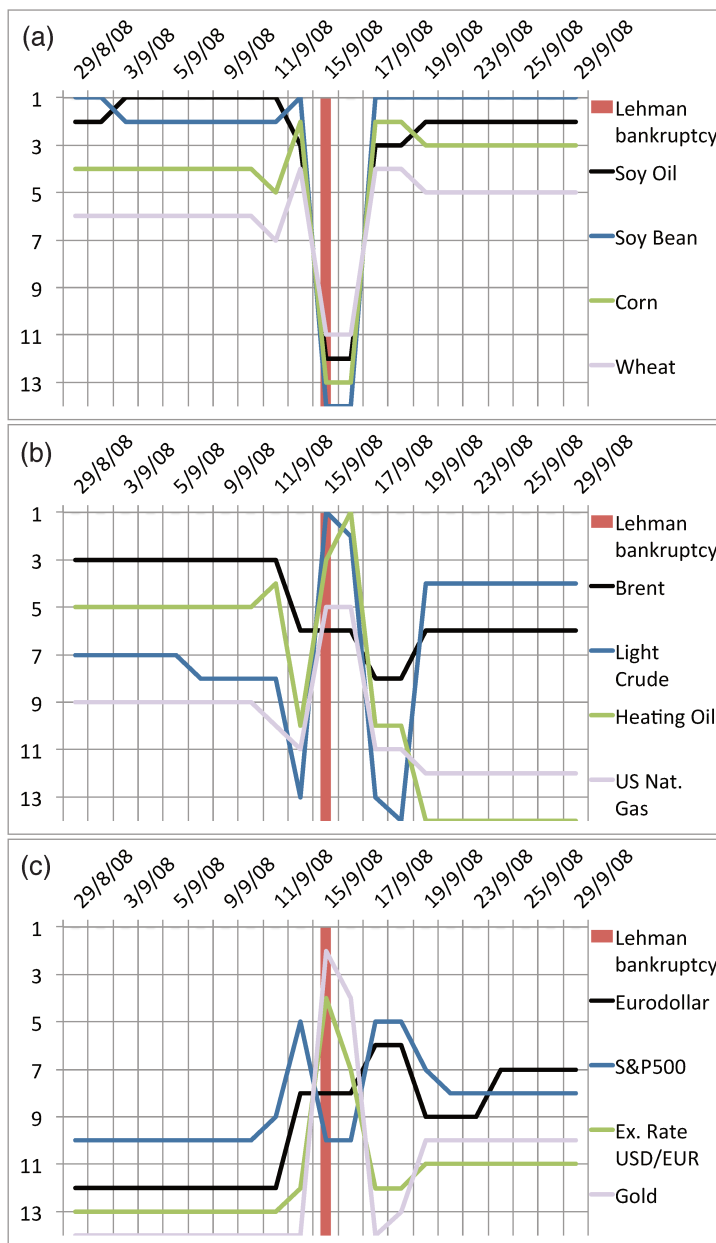


Figure 2.15 – Evolution of the ranks in the spatial dimension, around the Lehman Brothers bankruptcy.

Subset (a) is for agricultural markets, subset (b) is for only 4 energy markets (for readability) and subset (c) is for financial markets.

Chapter 3

Disentangling systemic risk in financialized commodity markets

Preliminary document

Abstract

To analyse whether the financialization of commodity markets contributes to systemic risk, we study a system consisting of both commodity and financial futures markets in a sparse Vector Autoregression (VAR) framework. It allows us to distinguish two components of systemic risk: we can assess systematic risk (integration) and propagation risk. This work aims at providing (non exhaustive) tools to help regulators analyse and monitor markets, with a focus on systemic risk. In particular, we can identify which markets are influential in any component of systemic risk and thus conduct a more in-depth investigation if necessary. We rely on an algorithm that gives sparsity in both the autoregression and partial correlation matrices. In a static analysis, in the spatial dimension, we find that sectors are separated, except for metals and finance. We also show that including the maturity dimensions is necessary, since they connect all the sectors and thus cause the integration of the whole

system. In our dynamic analysis, we focus on major financial events (the beginning of the subprime crisis, the default of Lehman Brothers, the maximum and minimum returns of the S&P500 futures and the 2010-05-06 S&P500 e-mini Flash Crash). We find that integration is very high between commodities and financial assets and among commodities, making systematic shocks a realistic possibility around each of these events, but that there is little daily propagation (slightly more around the Flash Crash).

Contents

Introduction	133
3.1 Data	138
3.2 Methodology	139
3.3 The reference graphs (static, full sample)	144
3.3.1 The partial correlation graph	144
3.3.2 The Granger causality graph	151
3.4 Shocks and propagation (dynamic, rolling window) . . .	155
3.4.1 The beginning of the subprime crisis (2007-08-09)	155
3.4.2 The default of Lehman Brothers (2008-09-15) .	157
3.4.3 The lowest return on S&P500 (2008-10-09) (overlapping with Lehman)	160
3.4.4 The largest return on S&P500 (2008-10-28) (overlapping with minimum S&P500)	162
3.4.5 The Flash Crash (2010-05-06)	164
Conclusion	166
References	167
Appendices	179
3.A Correlations vs. partial correlations	179
3.B Explanation of LASSO penalization	182
3.C Measures in the dynamic analysis over the whole period	185

3.C.1	Measures related to the propagation (Granger causality)	185
3.C.2	Measures related to the integration (partial correlation, PC)	187

Introduction

Since the 2007-2008 financial and economic crises, regulators have tried to improve their understanding and monitoring of financial markets. The notion of systemic risk has gained interest and became their focus, considering how difficult it is to assess. The BIS (2001) defines systemic risk as "the risk that an event will trigger a loss of economic value or confidence in, and attendant increases in uncertainty about, a substantial portion of the financial system that is serious enough to quite probably have significant adverse effects on the real economy." Since the main events that are considered as the start of the crises were the default of financial institutions, the majority of the works study the contribution of financial institutions to systemic risk and the propagation of shocks to help regulators in maintaining healthy markets. For example, Acharya et al. (2017) use their Systemic Expected Shortfall (SES), which measures the propensity of an institution "to be undercapitalized when the system as a whole is undercapitalized". They thus consider that the system being undercapitalized (in crisis) generates more risk for a financial institution, that the causality goes from the system to the institution. On the contrary, Adrian and Brunnermeier (2016)'s CoVaR measures "the change in the value at risk of the financial system conditional on an institution being under distress relative to its median state", so the institution contributes to the risk of the financial system. There have been other measures created, but these two are the most famous ones, even used by regulators; refer to Benoît, Colliard et al. (2017) for a review on this topic.

Nevertheless, the presence of financial institutions on risky markets could make them more vulnerable to default and hence generate cascade effects,

thus proper monitoring of financial institutions requires knowing the composition of their assets. Regulators are more and more pushing through transparency and require disclosure of their exposures. Benoît, Hurlin and Pérignon (2015) for example try to infer a bank's exposures through its more general disclosures like value-at-risk and find that commodity markets are part of their portfolios. Gorton and Rouwenhorst (2006) indeed show that before 2004, commodity futures presented interesting diversifying and investment benefits, with returns about the same as equity returns on average and zero or even negative correlations with them. Since then, there has been an increasing similarity between the evolution of prices of financial assets and commodities (undistinctively spot or futures, since they interact), which is due to the financialization of commodity markets (see Cheng and Xiong (2014) for a review of this strand of literature). They actually interact in several ways. Commodities are thus influenced by financial institutions using them: Büyükşahin and Robe (2014) show that the presence of speculators on both equity and commodity markets generate correlation between their prices, and some funds are even dedicated to the latter. Since commodities appear in portfolios for the reasons mentioned above, they may in turn influence many financial assets, from bank stock prices to those of companies using or producing commodities.

In this paper, we revisit the integration of financial and commodity markets in addition to potential propagation of information and of shocks between them and among commodities. Many papers have focused on particular links. For example, Y.-J. Zhang et al. (2008) looked at the spillover effects between US dollar exchange rate and oil prices. Hammoudeh and Yuan (2008) considered spillover between metals by taking into interest rates and oil price shocks, Park and Ratti (2008) assess how oil prices affect equity indices in European countries and the US. Ewing, Malik and Ozfidan (2002) examined volatility spillovers between crude oil and natural gas markets. Silvennoinen and Thorp (2016) study the correlation between crude oil and agricultural markets. There are many other articles on the relations between different markets, but we cite only a few here. All in all,

it is important to incorporate all those markets (and even others) into a single system to avoid missing relevant variables in the information transmission mechanism.

This is what we do here: our system consists of 51 time series, which represent 3 futures contracts (of different maturities) for 17 assets from 4 sectors (see Section 3.1 for details on data). We use a Vector AutoRegression (VAR) framework on daily data to untangle two components of systemic risk: we assess the integration of these markets, at the link level (pairwise) but also at a more global level (sectors, system), and whether there is propagation of returns between markets. Integration is measured in the partial correlation (PC) matrix, which gives the correlation coefficients of each pair of variables, conditional on all the other variables in the system. It thus filters out their influence and provides a better measure of the dependences, hence of integration. The use of partial correlations here improves over Chapter 2, which relied on unconditional correlations (see Appendix 3.A for an explanation and example). Propagation is measured in the Granger causality structure of the system, given by the AutoRegression (AR) matrix. Chapter 2 indeed lacked this temporal dependence, which is an important feature in determining whether one market influences another one (with causality). We found in Chapter 2 that the Minimum Spanning Trees (MST) were sparse and structured in an economically meaningful way (nodes organised by markets and by sector). In addition, since every contract of every market is not dependent on every other ones (contemporaneously or not), many partial correlation coefficients should be 0, hence the partial correlation matrix should be sparse. The novelty here thus relies on the use of LASSO penalization in an algorithm newly developed by Barigozzi and C. Brownlees (2017). We choose LASSO for several reasons (see Appendix 3.B for more details), but the most important ones are that it allows to work in high dimension (so we can increase the size of our system as much as we want or need to) and to give us sparse matrices. Their algorithm calibrates a VAR model (both matrices) in one step, which has several advantages, e.g. in terms of convergence properties of the parameters. More particularly in our case, it allows to put more or less weight to propagation or to

integration depending on the data, instead of first estimating one component and then let the rest for the other to explain.

The sparsity of the matrices enables their representation as graphs: the partial correlation graph and the Granger causality graph. The nodes of the graph will represent the time series of our system and the edge between a pair of nodes will represent, in the former, the partial correlation (undirected edge) and in the latter, their Granger causality (directed edge). Contrary to Chapter 2, we do not constrain the correlation graph to be a tree (which has no cycle) or even to be connected, since we want to be able to identify common factors and since cycles do increase integration. Indeed, if some nodes are connected (i.e. form a component of the graph), they may be subject to a common factor, driving all their partial correlations. Even though Barigozzi and C. Brownlees (2017) recommend controlling for those common factors to get sparser matrices, we want to have the possibility to see whether financial and commodity markets are influenced by a common factor or not.

Based on these graphs, we derive graph theoretic measures that can help monitor potential propagation and integration and help identify which markets should be investigated in more depth. More particularly, the sparsity of the matrices allows to assess the importance of propagation and integration in terms of number of connections involved (degree centrality). In addition, we use the total communicability centrality measure by Benzi and Klymko (2013) to identify important contracts and also to assess how easy information can flow in the network (which is also a measure of integration when to the PC matrix). It improves on the Katz-Bonacich measure used in Chapter 2 as the rankings are more stable and there is no sign issue. We adapt this measure to our application by using a weighted matrix instead of an adjacency matrix, which is done for the first time, to the best of our knowledge. It thus accounts for the individual weights (partial correlations) instead of using a single weight parameter, as we will explain below. This work should thus provide means for regulators to monitor markets and eventually prevent the occurrence of crises by taking action on the identified

markets if propagation risk or integration are too high.

This work also differs from Diebold, Yilmaz and L. Liu (2017), who also use a sparse VAR framework and, as in Barigozzi and C. Brownlees (2017), encompass both the contemporaneous and lagged influences into a single matrix. They derive variance decompositions and then aggregate them into node-level and system-level directional connectedness to provide a global measure of systemic risk. Here, as stated above, we work on partial correlations (from the concentration matrix), which gives us the conditional (in)dependence structure of the system, compared to the variance decompositions. We also keep the contemporaneous and the lagged influences separated in order to untangle system integration for the former – the systematic component of systemic risk – and the propagation component. It thus allows to identify which markets are actually involved in each component and monitor and intervene on them if necessary.

Other works have also used sparse frameworks, but have only focused on one matrix: a sparse VAR with only the AR matrix being sparse or only the concentration/covariance matrix being sparse (see Barigozzi and C. Brownlees (2017) for some examples).

In Section 3.1 we briefly introduce our database. In Section 3.2, we present the methodology we use and how it is relevant to study propagation and integration. In Section 3.3, we study our whole database (2000 to 2014) to provide a long-term reference graphs to compare with and explain the measures and results we derive. In Section 3.4, we then conduct a dynamic analysis (with a rolling window), in which the problem becomes highly dimensional, thus the relevance of using LASSO. More specifically, we focus on periods around major financial events that may have affected commodity markets. In chronological order, first, we look at the day BNP Paribas froze the redemption of some of its investment funds because it couldn't value some of its structured products (August 9, 2007, one of the dates that are defined as the beginning of the subprime crisis). Second, we examine integration and propagation around the default of Lehman Broth-

ers (September 15, 2008), which is often assumed to be the beginning of the economic crisis. Third and fourth, we assess whether the minimum and maximum returns on the S&P500 index in our sample, respectively October 9, 2008 and October 28, 2008, propagated to commodity markets, since we observed in Chapter 1 that many of them experienced their extremum return on the following day. Fifth, we look into what may have happened around the day of the May 6, 2010 Flash Crash, since it has affected some contracts in our system, namely the S&P500 e-mini futures.

3.1 Data

We collected futures prices for 17 different underlying assets from 4 different sectors (energy, finance, metals and agriculture) from Datastream, constructed continuous time series with constant maturity and computed daily returns.

We have 208 time series, with many maturities for some markets, but will only keep 3 for each market (short-, medium- and long-term contracts) for several reasons. First, not all maturities of each market are relevant. Second, we want to have an overall balanced representation for each market. Third, working with too many time series, even in a high dimension framework with variable selection and filtering, can still lead to results that are difficult to interpret.

Table 3.1 details the underlying assets (Market column) we retained, the exchange on which they are traded (Exchange column) and the maturity of the contracts (Node labels and maturity) we kept. We thus chose three maturities for each market: the front-month contract, representing the spot value (short term); the 3-month maturity (or closest higher than 3), representing the medium term; and the 12-month maturity (or largest available if less than 12), representing the long term.

We thus end up with 51 nodes/variables, with a total of 2889 daily observations of return for each, from 2000-01-24 to 2014-02-14.

Table 3.1 – Data summary

Market	Exchange	Node labels (maturity in months)
WTI	CME-US	WTI1 (1), WTI2 (3), WTI3 (12)
Brent	ICE-EU	Brent1 (1), Brent2 (3), Brent3 (12)
Heating Oil	CME-US	H.O.1 (1), H.O.2 (3), H.O.3 (12)
Gasoil	ICE-US	Gasoil1 (1), Gasoil2 (3), Gasoil3 (12)
US Nat. Gas	CME-US	USNat.Gas1 (1), USNat.Gas2 (3), USNat.Gas3 (12)
UK Nat. Gas	ICE-EU	UKNat.Gas1 (1), UKNat.Gas2 (3), UKNat.Gas3 (9)
Wheat	CME-US	Wheat1 (3), Wheat2 (5), Wheat3 (12)
Soy Bean	CME-US	Soybean1 (2), Soybean2 (4), Soybean3 (12)
Soy Oil	CME-US	Soyoil1 (1), Soyoil2 (3), Soyoil3 (12)
Corn	CME-US	Corn1 (3), Corn2 (5), Corn3 (12)
Eurodollar	CME-US	IR1 (1), IR2 (3), IR3 (12)
USD/EUR Ex Rate	CME-US	FX1 (3), FX2 (6), FX3 (12)
S&P500	CME-US	SP5001 (Spot), SP5002 (3), SP5003 (6)
Gold	CME-US	Gold1 (1), Gold2 (4), Gold3 (12)
Silver	CME-US	Silver1 (1), Silver2 (3), Silver3 (12)
US Copper	CME-US	USCu 1 (1), USCu 2 (3), USCu 3 (12)
UK Copper	LME-EU	UKCu 1 (1), UKCu 2 (3), UKCu 3 (12)

3.2 Methodology

An obvious way to view propagation is to consider the effect of one (or several) market(s) on others. This is exactly what a Vector Autoregression (VAR) is doing: it assumes that previous observations of a vector of random variables influences the contemporaneous observation. More formally, if \mathbf{Y}_t

is our vector of random variables $Y_i, t, i = 1, \dots, N$, we have that:

$$\mathbf{Y}_t = \sum_{k=1}^p \mathbf{A}_k \mathbf{Y}_{t-k} + \mathbf{u}_t \quad (3.1)$$

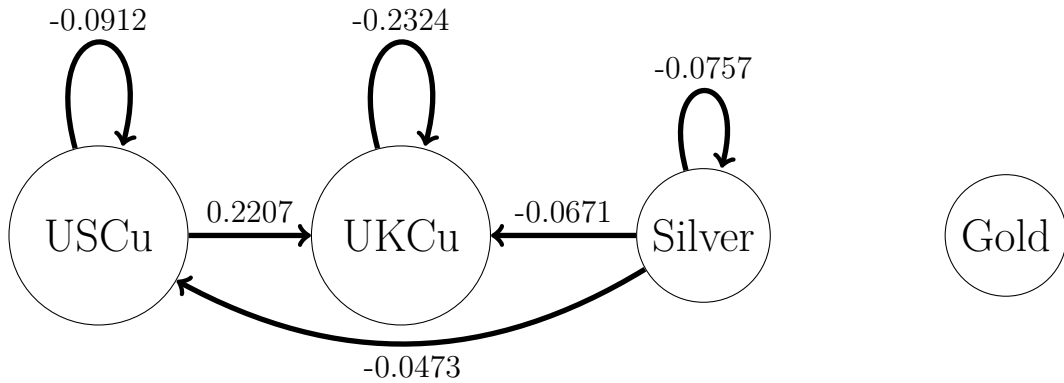
with p being the order of the VAR and $\mathbf{u}_t \sim \mathcal{N}(0, \Sigma_u)$.

Table 3.2 – Calibration results of a VAR(1) model on metals markets in the spatial dimension.

	UKCu 1	USCu 1	Silver 1	Gold 1
const	0.0004	0.0005	0.0004	0.0004**
L1.UKCu 1	-0.2324***	0.0407	0.0333	0.0141
L1.USCu 1	0.2207***	-0.0912**	-0.0358	-0.0343
L1.Silver 1	-0.0671**	-0.0473*	-0.0757**	-0.0117
L1.Gold 1	0.0627	0.0454	0.0803	0.0167

* = statistically significant at 10%, ** = statistically significant at 5%, *** = statistically significant at 1%. L1 means lagged value of 1 period.

Figure 3.1 – Granger causality directed graph from VAR(1) on metals spot markets



The nodes represent the time series for each spot market. The directed links (arrows) represent the lagged Granger causality between the different markets: for example, the arrow from USCu to UKCu means that the return on the copper traded in the US will Granger cause (influence) the return on the copper traded in London on the following day. The numbers on each link represent the autoregression coefficients from Table 3.2.

Applying this framework on small systems gives interesting results. For example, if we consider only metals spot markets (here, two coppers, silver

and gold) and calibrate a VAR(1), we find the parameter estimates of Table 3.2. Keeping only the statistically significant coefficients, which correspond to Granger causality relationships, we can summarise these parameters as links in a graph, as in Figure 3.1. The nodes represent time series of returns of each futures contract. The directed links represent the Granger causality relationships. For example, the link going from USCu to UKCu means that the return on USCu at a date t will affect the return on UKCu at date $t + 1$ (which is given by the corresponding entry in the VAR(1)).

The directionality of these links give interesting insights. There is some autocorrelation for both coppers and for silver and it is negative for all of them. This is consistent with the common view that commodity markets exhibit a mean-reverting behaviour (see Lutz (2010) for a review of explanations and tests). The link from USCu to UKCu could reflect the time difference between the markets: the information from the closing of the Chicago market would be incorporated the following day for the London market, for about 22%. Silver (Chicago) is influencing the two copper markets but is contributing only little to these markets. What is also interesting is that the gold market is not affected or affecting the others (at a statistical significance of 1%), but only with a lag 1. There is thus a clear separation between the reserve of value of the gold and industrial metals. This may be explained by some cross-correlation at longer lags or maybe even in the contemporaneous observations.

We thus also study the contemporaneous correlation matrix. More precisely, we turn to the partial correlation matrix \mathbf{C} , because it encodes the conditional dependence structure of the time series. Indeed, if $c_{ij} = 0$, then Y_i and Y_j are independent conditional on the other variables ($Y_k, k \neq i, j$). It thus filters out the influence of the other variables, which could result in exaggerated correlations if the two variables are both correlated with another one for example (see Appendix 3.A for more details). But if there exists a path between two variables in the partial correlation graph, then their unconditional correlation will be nonzero (while their partial correla-

tion may be 0). By considering partial correlations, we can suppose that there may be sparsity in the dependence structure (every one of our markets should be independent of many of the others), hence the LASSO penalization applied to estimating partial correlations instead of correlations. As per the results from Chapter 2 in particular, the partial correlation matrix (more specifically the concentration matrix) could potentially be sparse, which comforts us in using this approach.

Instead of successively estimating the Autoregressive matrix and then the partial correlation matrix of the residuals, we will estimate them simultaneously. Indeed, having a two-step estimation procedure is challenging in terms of convergence and of properties of the estimators (see Barigozzi and C. Brownlees (2017)). Nevertheless, having to estimate both the Autoregression matrices ($N \times N$ for each lag) and the contemporaneous partial correlation matrix ($\frac{N \times (N-1)}{2}$) of the residuals makes this problem high-dimensional even for relatively small systems. For example, if we take our 17 different assets and keep only 3 contracts for each, we have a system of $N = 51$ nodes/variables. In a VAR(1) model, this would mean $51 * 51 = 2,601$ autoregressive coefficients, plus $51 * 50/2 = 1,275$ contemporaneous partial correlation coefficients for a total of 3,876 parameters. Recent machine learning techniques can be applied here to work with high dimensional systems and provide consistent estimates without overfitting and even with a number of observations lower than the number of parameters to estimate.

In addition, in this plethora of parameters, not all of them are relevant. First, as noted in the introduction, we find a kind of sectorization in Chapters 1 and 2: markets are generally "clustered" into sectors, financial markets being separated from commodity markets. Moreover, the maturity dimensions for each asset has a linear structure in general. Thus, especially based on Chapter 2, we may have sparsity in the matrices of parameters: many parameters may be set to 0, but we want to find which ones (not imposing our assumptions). The Least Absolute Shrinkage and Selection Operator (LASSO) regularization is particularly suited for this task. In the

end, we will calibrate a high-dimension VAR model on our data, with sparse AR matrix and partial correlation matrix. We use an algorithm developed by Barigozzi and C. Brownlees (2017), called "nets algorithm", in order to simultaneously estimate a sparse VAR and a sparse contemporaneous partial correlation matrix of the innovations thanks to a LASSO regularization. These two sparse elements are then represented as two graphs. A directed graph for the Granger causality links from the AR matrix will inform us about potential propagation of information and shocks, while an undirected graph (as in Chapter 2) for the contemporaneous partial correlations will rather inform us about the integration of the system, and hence, the potential for systematic shocks.

This framework and the sparsity will allow us to assess systemic risk in different ways. We can, for instance, derive some graph theoretic measures such as centrality in the graphs, to untangle which one(s) is (are) the most important market(s).

More particularly, we will briefly consider degree centrality (the number of links/neighbours of a node) and compare the results with total communicability centrality, developed by Benzi and Klymko (2013). This measure allows to take into account not only direct neighbours, but also indirect neighbours (even infinitely far ones). More formally, if \mathbf{A} is the adjacency matrix of the graph, its powers (\mathbf{A}^k) provides the number of paths of length k between each pair of nodes. Hence, summing the powers of this matrix gives the total number of paths between each pair of nodes. But to dampen the effect of longer paths, it is common to add weights to the powers of the adjacency matrix. Here, the weights will be the $\beta^k/k!$, which allows the sum (power series) to converge to the exponential of the weighted adjacency matrix as follows:

$$\sum_{k=0}^{\infty} \frac{\beta^k}{k!} \mathbf{A}^k = e^{\beta \mathbf{A}} \quad (3.2)$$

Each coefficient $[e^{\beta \mathbf{A}}]_{i,j}$ thus gives the communicability between nodes i and j . To find the total communicability centrality of each node, we just sum each row.

In this paper, we will slightly change it: instead of taking $\beta^k \mathbf{A}^k$, we will directly take the sparse partial correlation matrix (with its diagonal set to 0 to avoid self-loops): it allows us to take into account each partial correlation of the different links instead of a general β coefficient. To the best of our knowledge, this is the first time this measure is used this way in finance.

In addition, taking the sum of the centralities gives the total communicability of the network, allowing to have an overall measure of the ability for information to flow in the system (and see how it evolves over time for example) and even compare different network structures.

3.3 The reference graphs (static, full sample)

We will first study the case where we calibrate a sparse VAR(1) using all our observations, which will provide us a reference case for future comparisons. We first calibrate on a subset of our variables (spatial dimension: only the front month contract for each of the 17 assets) and then calibrate on our 51 variables (3D). We can thus analyse the two graphs that emerge from this calibration: the graph of pairwise partial correlations of the residuals and the graph of Granger causality.

3.3.1 The partial correlation graph

Since the partial correlation between two variables measures their dependence conditional on the observation of all the other variables, it filters out their influence. If despite this filtering, some variables are still dependent, it means that they are directly and actually dependent on each other (instead of being correlated because they would both be correlated with another variable).

We can assume that most of our markets are not conditionally dependent of many others, that there should be some clustering (into sectors for ex-

ample). If we instead observe that markets that should be unrelated are connected in the graph, it would mean that they may be influenced by a common phenomenon. Hence, the number of components observed in the estimated graphs will tell us whether the markets under study are integrated (if different markets cluster) or if they are still subject to different fundamentals. In particular, if all the nodes belong to a single component, it would mean that the system is completely integrated, prone to systematic shocks.

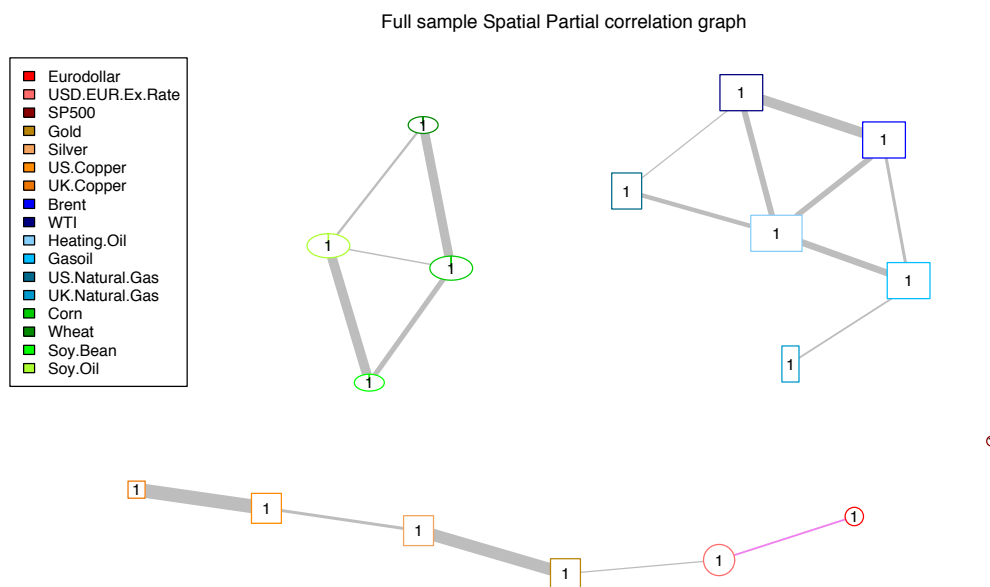
We can refine this analysis by looking at the values of the partial correlations. The average partial correlation will tell whether this integration is strong or not. We can also check the minimum and maximum values of these partial correlations, in order to check their amplitude in the system. We will also assess which markets are the most central in terms of dependence with the others by looking at the communicability centrality of the nodes.

3.3.1.1 In the spatial dimension

Figure 3.2 represents the calibrated partial correlation graph in the spatial dimension only. Nodes are coloured according to their sector (red for finance, blue for energy, green for agriculture and orange for metals), edges are coloured according to the sign of the partial correlation and their width represents the absolute value of the coefficient. The partial correlations range from -0.04 to 0.84, with an average of 0.28 (only one link has negative partial correlation). Correlations range from -0.04 to 0.85, but are not as sparse (74% sparse vs. around 87% for partial correlation matrix, 18 links among the 136 possible ones). We remind that if there exist a path between two nodes (for example if they belong to the same component), their (unconditional) correlation coefficient will be nonzero.

We can notice several interesting insights from this graph. The first one is that it consists of three clusters, while we have theoretically 4 sectors: one

Figure 3.2 – Partial correlation matrix graph in the spatial dimension



The shape of the nodes represents their sector: circles for Finance, rectangles for Energy, ellipses for Agriculture and squares for Metals, with details of each market on the figure (different colours). The colour of the edges represents the sign of the weight/dependence: grey for positive, violet for negative. The size of the nodes represent their degree: the bigger the node, the higher its degree. The width of the edges represent the absolute value of their weight/dependence: the wider the edge, the higher the dependence (in absolute value).

cluster for energy markets, one for agricultural markets and one for metals and financial markets. Lautier and Raynaud (2012) find similar results based on the filtering of the correlation matrix using Minimum Spanning Trees (MSTs), which constrains the graph to be connected (single component) but did not have metals markets. Here, allowing several components helps us assess whether the integration is systemwide or still "sectorwide". In addition, we allow for cycles here (while the MSTs do not), which tell us precisely which markets are connected with which others. The financial and metals sectors are connected through USD/EUR exchange rate and gold market, which can be considered a reserve of value, directly followed by silver (their link is strong, as represented by the width of the link). This

component is moreover organised linearly (whereas the other two components are quite connected), meaning that there is no "direct" interdependence between all these markets but the (unconditional) correlation coefficients between the markets of this component will still be nonzero (they will be dependent through their neighbours). Finally, we can note that the S&P500 is not connected to others in this graph, meaning that it did not have relevant links with the other markets and that its return innovations are independent of the others.

Table 3.3 – Communicability centrality of markets (from high to low) in the reference graph of partial correlations in the spatial dimension

	Centrality
Heating Oil 1	2.75711552
Brent 1	2.64689923
Light Crude 1	2.55987776
US Copper 1	2.52091147
UK Copper 1	2.39170158
Silver 1	2.27630636
SoyBean 1	2.18569361
Corn 1	2.14969241
Gold 1	2.13325709
SoyOil 1	1.97115158
Wheat 1	1.90001059
Gasoil 1	1.88942044
US Natural Gas 1	1.3720924
UK Natural Gas 1	1.04974776
SP500 1	1
USD EUR Ex Rate 1	0.9630333
Eurodollar 1	0.96096818

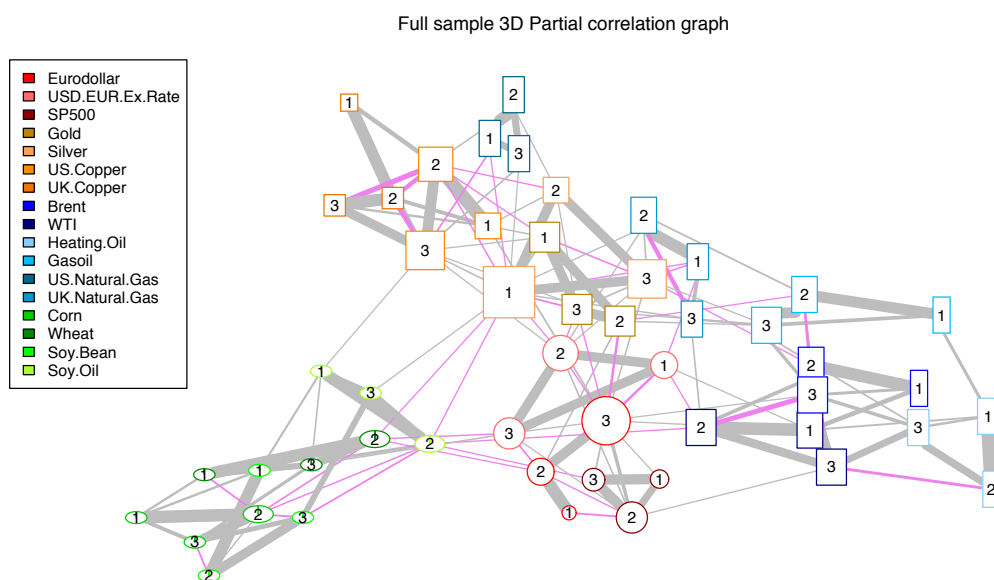
The degree of the nodes counts the number of direct neighbours that they have. But this is not enough, as indirect connections also matter, as a path in the partial correlation graph means there is an unconditional correlation between the two extremities of the path. Hence, the communicability centrality measure accounts for (infinitely) further "neighbours", in

the sense that it accounts for every possible path from each node, but also weights them (to dampen the influence of further neighbours). Therefore, if the graph consists of three components, the centrality of the nodes will represent their centrality only in their component. Table 3.3 displays the communicability centrality measure computed on this graph. We note that the heating oil has the largest degree (4) and is also the most central market, followed by the brent and WTI, which have a strong connection. Hence, these markets are the most susceptible to influence the others, but since the sectors are separated, the influence would be limited to energy markets. In the metals/financial component, the copper markets have the most potential for information propagation, which can be explained by their strong link and their distance to the link of negative partial correlation between eurodollar and USD/EUR exchange rate. Finally, in the agricultural sector, soybean and corn are the most central, as their links are among the strongest.

3.3.1.2 In 3D

Figure 3.3 presents the estimated partial correlation graph on the whole sample, in three dimensions. When we include different maturities for each market, we find many more links than in the spatial dimension, forming a single component (all nodes are present) instead of several ones as in the spatial dimension only. This means that all those markets are integrated (prone to systematic shocks)... But this integration is quite diverse: partial correlations range from -0.24 to 0.75 with an average of 0.16 (and most of them, 72%, are positive). The nets algorithm filtered 88% of the partial correlation coefficients. This partial correlation matrix translates into a non-sparse correlation matrix, with coefficients ranging from -0.82 to 0.99 with an average of 0.10 (only 60% positive ones). We see here the effect of other variables on the correlation coefficients, which have a much larger amplitude than the partial correlations.

Figure 3.3 – Partial correlation graph in 3D



The shape of the nodes represents their sector: circles for Finance, rectangles for Energy, ellipses for Agriculture and squares for Metals, with details of each market on the figure (different colours). The colour of the edges represents the sign of the weight/dependence: grey for positive, violet for negative. The size of the nodes represent their degree: the bigger the node, the higher its degree. The width of the edges represent the absolute value of their weight/dependence: the wider the edge, the higher the dependence (in absolute value).

Markets are again clustered into sectors and here, finance and metals are in the middle of energy and agriculture, which may generate systematic risk if these former two sectors are affected. An exception is the US natural gas markets, which seems to connect with metals instead of other energy markets.

Table 3.4 presents the 10 most and least central markets according to the communicability centrality measure. The most striking result is that the 12-month S&P futures contract is the most central one. This result naturally raises concerns of potential shock lead by financial markets in this system; we can thus look into it more particularly. Its maximum partial

correlation is around 0.28, so its influence is not among its neighbours but further down the paths. Nevertheless, the differences in centrality values are not very large until we reach the bottom of the ranking (so the system is highly integrated), so the influence of the S&P500 futures may not be that much larger than others. In particular, we see that all 4 sectors are represented at the top, with four ranks taken by energy markets, three by metals, two by agricultural markets and only one for financial markets. We note that there is only one front-month contract, therefore futures seem to be conveying and receiving the most information contemporaneously, compared to spot markets. In conjunction with degree centrality, the influence of that contract (silver) is not only direct, it can easily reach infinitely further nodes, unlike the 12-month eurodollar contract, which is among the least central nodes. In addition, silver also has another contract in the top, its 12-month one, making it very influential: shocks from either its spot or futures markets could lead to systematic shocks. Then, investigating their maximum partial correlation, we find that they are close to 0.5, larger than those of the S&P500 futures, but still not the largest ones. This result emphasises that we should not just focus on the largest partial correlations to find the most influential markets.

If we now look at the least central markets, we find many energy markets, more particularly, most of the natural gas contracts. Though the UK natural gas contracts appear around the "center" of the graph of Figure 3.3, they are the least central ones. The same holds for the 12-month eurodollar contract, which even has a quite high degree (13, second after the front-month silver, which has 15) and its partial correlations range from -0.08 to 0.52 (with 4 negative partial correlations).

If the focus is only on the transmission of information, it could be more relevant to conduct this analysis with the absolute value of the coefficients to assess the amplitude of transmission instead of allowing for negative coefficients mitigating the influence of positive ones.

Table 3.4 – Communicability centrality of the 10 most (left) and least (right) central markets (from high to low) in the reference graph of partial correlations in 3D

	Centrality		Centrality
SP500 3	2.84667941	Wheat 3	2.4987086
Gasoil 2	2.81147241	Soy Bean 1	2.46805167
Soy Oil 2	2.8034704	US Natural Gas 1	2.46415903
WTI 2	2.77946501	Gasoil 1	2.46101264
Silver 3	2.77499497	US Natural Gas 3	2.16668475
US Copper 2	2.77100391	Eurodollar 1	2.07323956
Brent 2	2.76630783	Eurodollar 3	2.01416571
Soy Bean 2	2.76122455	UK Natural Gas 1	1.98864143
Heating Oil 2	2.76078463	UK Natural Gas 2	1.86925835
Silver 1	2.7540656	UK Natural Gas 3	1.13756682

3.3.2 The Granger causality graph

Since the AR coefficients are directly related to the notion of Granger causality, they first tell us whether some returns are Granger-caused by others (nonzero AR coefficients). They also inform us on the amplitude of this influence (autocorrelation is also allowed), we can see this as a kind of daily propagation.

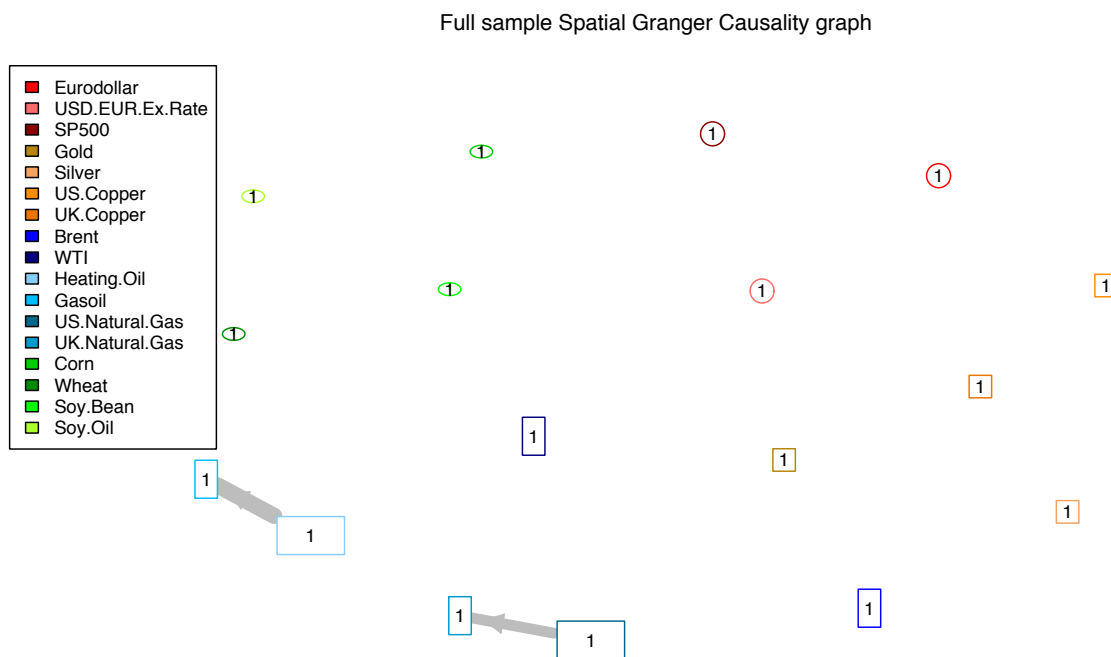
Several measures derived from the graph can be useful in assessing systemic risk (propagation risk). The number of clusters, of nodes and of links will tell us the range of the propagation, whether it is widespread or if it is contained within a subset of variables.

The average AR, minimum AR and max AR will tell us the possible amplitude of this propagation, whether it is positive or negative.

In addition, we can study the centrality of the nodes, to quantitatively assess which ones will be propagating information the most (which may need monitoring), and which ones the least, but considering that there are few links, it may not be necessary.

3.3.2.1 In the spatial dimension

Figure 3.4 – Autoregressive matrix graph (Granger causality graph) in spatial dimension



The shape of the nodes represents their sector: circles for Finance, rectangles for Energy, ellipses for Agriculture and squares for Metals, with details of each market on the figure (different colours). The colour of the edges represents the sign of the weight/dependence: grey for positive, violet for negative. The size of the nodes represent their degree: the bigger the node, the higher its degree. The width of the edges represent the absolute value of their weight/dependence: the wider the edge, the higher the dependence (in absolute value).

It seems like there is not much Granger causality in the spatial dimension. Figure 3.4 shows the estimated Granger causality graph (from the AR matrix). Only two links, both positive, seem relevant when using our whole database: one from heating oil to gasoil (for 0.18) and one from US natural gas to UK natural gas (0.11). The latter link is natural, as it would reflect the difference of trading hours (the US market closing after the UK market,

information of the day in the US market is incorporated the following day in the UK market). The inverse relationship is not seen, as it should appear in the contemporaneous correlations instead of here. The former link is probably reflecting the same phenomenon (gasoil being quoted in EU). Naturally, as we look at the centrality of this graph, the nodes influencing the others are the most central ones, i.e. heating oil and US natural gas here, though the influence is limited.

Thus, there is not much propagation risk in the spatial dimension at a daily frequency.

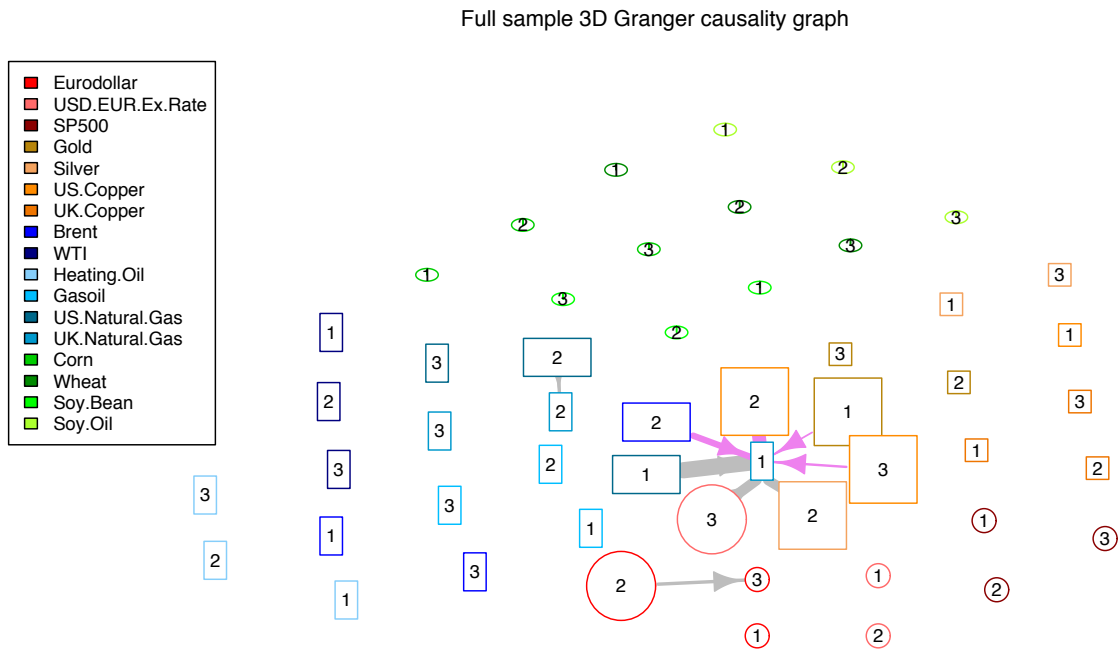
3.3.2.2 In 3D

When including maturities in the system, the Granger causality graph in Figure 3.5 remains quite simple but adds interesting features to the spatial dimension. From that dimension, only the link from US natural gas to UK natural gas remains. Many other contracts, mostly metals, join in and influence this UK natural gas front month contract... But their influence is not homogeneous; some have a negative influence, others have a positive one, but the most important one in absolute value is still from the US natural gas (0.12). So here again, the propagation risk is still limited. We also find a link from the 3-month US natural gas to 3-month UK natural gas and a link from 3-month eurodollar to 12-month eurodollar, but both influences are very limited (0.015).

Also, in terms of centrality, as in the spatial dimension, the nodes influencing the others are the most central ones and here the UK natural gas is obviously the one receiving the most influence.

We will study these different measures around several events of interest, but can also look at the overall picture (their evolution over our sample).

Figure 3.5 – Autoregressive matrix graph (Granger causality graph) in 3D



The shape of the nodes represents their sector: circles for Finance, rectangles for Energy, ellipses for Agriculture and squares for Metals, with details of each market on the figure (different colours). The colour of the edges represents the sign of the weight/dependence: grey for positive, violet for negative. The size of the nodes represent their degree: the bigger the node, the higher its degree. The width of the edges represent the absolute value of their weight/dependence: the wider the edge, the higher the dependence (in absolute value).

3.4 Shocks and propagation (dynamic, rolling window)

We use a rolling window of 252 observations, corresponding to approximately 1 trading year. We computed all measures for our whole sample and provide the most interesting ones in Appendix 3.C, providing the bigger picture for analysing where the events lie. In addition to the previous measures, in a dynamic setting, we can also assess the stability of the graphs by looking at the survival ratio of the graph, compared to the reference graphs or the graph of the previous day for example (see Figure 3.12 in Appendix).

We retain several dates to analyse: from Chapter 1, we retain the minimum and maximum observed return for the spot S&P500 index, respectively 2008-10-09 and 2008-10-28 (because the following days, many commodities experienced extremum return). We will also study what happens around the beginning of the subprime crisis (2007-08-09), the default of Lehman Brothers (2008-09-15) and the Flash Crash (2010-05-06) since they are financial events that may have affected the system.

Please note that for the animation to play, you need to use Adobe Acrobat Reader or similar software (does not work with Preview on Mac OS).

3.4.1 The beginning of the subprime crisis (2007-08-09)

One of the dates that have been identified as the beginning of the subprime crisis is August 9, 2007. On that day, BNP Paribas stopped the valuation (and subscriptions and redemptions) of three of its funds due to "the complete evaporation of liquidity in certain market segments of the U.S. securitization market" (NYT 2007-08-09)¹. We will thus look at how our graphs (Figures are available upon request, due to their size) and measures

¹<http://www.nytimes.com/2007/08/09/business/worldbusiness/09iht-09bnp.7054054.html>

(Appendix 3.C) behave around that date.

As in the reference graphs, UK natural gas receives information from metals markets (which is strange) and from the US natural gas, though there are only 4 of the 7 (reference) relevant links to UK natural gas. The front-month soybean contract joins on 2007-08-13 and the silver contract goes away on 2007-08-21, after central banks have increased their support and lowered their rates.

In the partial correlation graph, markets are also clustered into sectors overall, with metals markets in the middle and with the exception of the gasoil, US natural gas and eurodollar (the S&P500 also has an unstable connection). The US natural gas is a peculiar market, as evidenced by Lautier and Raynaud (2012), so adding metals markets, compared to them, seems to separate this market from other energy markets. One particularity is that the front-month eurodollar contract is not part of this graph around this event and other eurodollar contracts were even separated from all other contract on 2007-07-25, to then stick on metals markets. Moreover, the gasoil contracts also separated from this single component on 2007-08-20, to go back to normal afterwards, sticking back to the energy sector on 2007-08-28.

The range of partial correlations fluctuates quite a lot until one week before this event. There is a large drop in the minimum partial correlation from 2007-08-20 to 2007-08-22, while the gasoil market is living on its own and central banks were supporting financial markets. But since this drop was still limited (from -0.2 to -0.3).

Around the beginning of the subprime crisis, however, the gold contracts have taken the role of most connected nodes (instead of silver). Since they play a role of reserve of value, we naturally find that 60-75% of their links are negative partial correlations, but are in general very low, except for those with other metals. Silver markets are not as important in terms of reserve of value, but still have many negative partial correlations with other markets (81% and 83% of their links for the 3-month and 12-month respectively). Nevertheless, looking at the rankings provided by the total

communicability centrality, the 12-month S&P500, which seems peripheral, was the second most important contract on 2007-07-25 and became first the following day. Other important contracts are naturally mainly metals. In terms of total communicability, the trend is negative around the beginning of the subprime crisis, but looking at the bigger picture in Appendix 3.11, it explodes very early after.

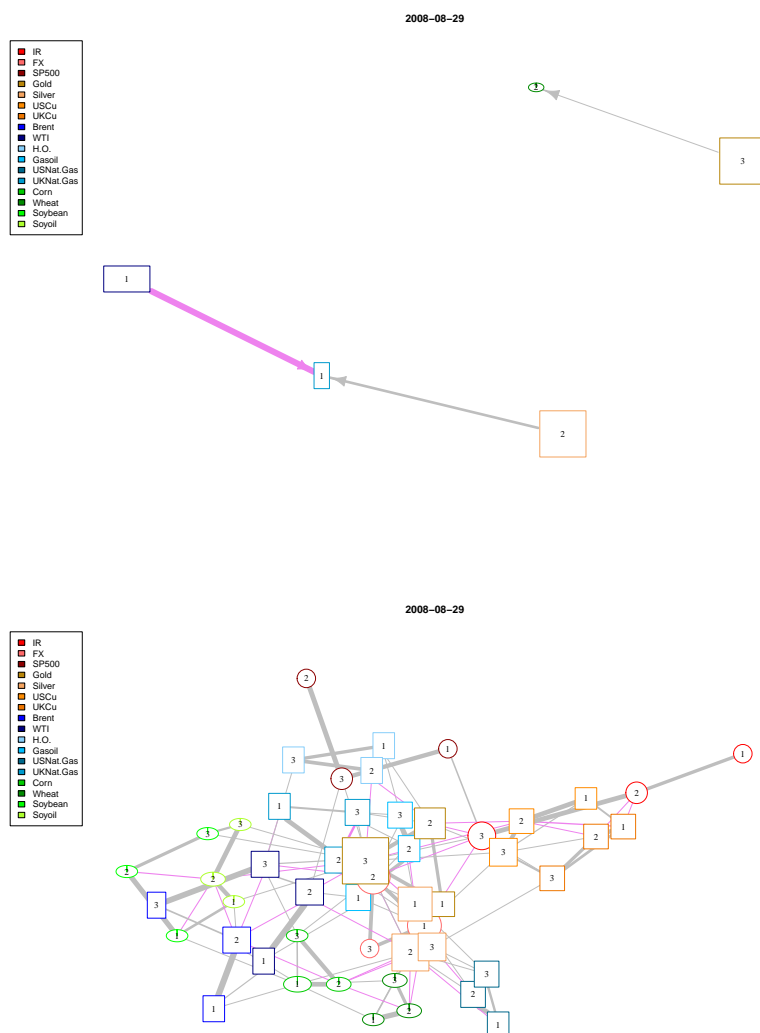
Overall, we do not find evidence of substantial systemic risk. Nevertheless, some indicators point to towards potential vulnerability to systematic shocks coming from stock markets: the presence of a large component in the partial correlation graph, meaning that the system is integrated, partial correlations ranging from -0.3 to more than 0.9 and the S&P500 being the most central contract in the system...

3.4.2 The default of Lehman Brothers (2008-09-15)

Lehman Brothers, one of the most important derivative dealers, faced tremendous difficulties, until it had to default on 2008-09-15 (see Appendix in Chapter 2 for a short timeline). This event has triggered the global financial crisis, that has rippled through the whole financial system, affecting economies worldwide and thus, being a systemic event, coming from the financial sector. Many interconnections were neglected, leading to largely unexpected losses for many entities in different sectors, plus huge commodity price drops. We thus want to analyse what happened around that date and other subsequent events.

We find overall two components in the Granger causality graph: one revolving around the UK natural gas, as usual, and one around an agricultural market (wheat or soybean). The latter component is dominated by the 12-month gold contract, which positively influences the 12-month wheat until Lehman Brothers defaults on 2008-09-15, to then turn to the front-month soybean contract (negative influence). Starting on 2008-09-19, day on which the Troubled Asset Relief Program (TARP) was announced,

Figure 3.6 – Granger causality (above) and partial correlation (below) graphs on the day of default of Lehman Brothers



The shape of the nodes represents their sector: circles for Finance, rectangles for Energy, ellipses for Agriculture and squares for Metals, with details of each market on the figure (different colours). The colour of the edges represents the sign of the weight/dependence: grey for positive, violet for negative. The size of the nodes represent their degree: the bigger the node, the higher its degree. The width of the edges represent the absolute value of their dependence: the wider the edge, the higher the dependence (in absolute value).

the 12-month eurodollar becomes positively influenced by this same gold contract... Financial agents may have been looking for safety in the longer contracts at that time. The former component sees the WTI, Brent and/or silver (various maturities) influence the UK natural gas.

The partial correlation graph still has a single huge component, but sectors are all scattered and with a smaller range of partial correlations (from -0.3 to less than 0.9). The maximum partial correlation seemed to decrease steadily, but skyrockets back to its pre-decrease level on 2008-09-23, just one day after the G7 commits to protect the financial system, while the proportion of negative partial correlations also increases.

In terms of centrality, gold, silver, WTI and USD/EUR exchange rate have the highest degrees and are at the center of the graph. The total communicability centrality gives overall the same results, but also allows to see the substantial change that occurred on 2008-09-19. On that day (day the TARP was revealed), the 3-month maturity eurodollar contract went from 26th to 2nd most important contract (after the 12-month S&P500) and total communicability of the network spiked, so information could flow more easily. By acting on this contract, regulators may have been able to mitigate the effect of the default of Lehman Brothers and its subsequent cascade effects. In addition, the front-month USD/EUR exchange rate contract also became quite important in the system (from 28th on 2008-09-11 to 9th on 2008-09-19), showing that exchanges between countries (capital reallocations) may have driven the returns in the system and may even have propagated the crisis. Nonetheless, regulators could also have intervened on these currency markets to try to avoid too much reallocation. Other contracts got to lower ranks accordingly, but the two shortest maturities of gold contracts in particular became quite less important (even if the second one is the most connected one). This may not be that counterintuitive, as gold would be anticorrelated to procyclical assets and relevant as an alternative mostly for financial investors.

All in all, markets are integrated (they form a single component in the partial correlation graph) and there may have been capital flows between

countries or assets at play around the demise of Lehman Brothers. Nevertheless, there is no significant sign of potential for daily propagation, since the number of Granger causality links and their coefficients are low.

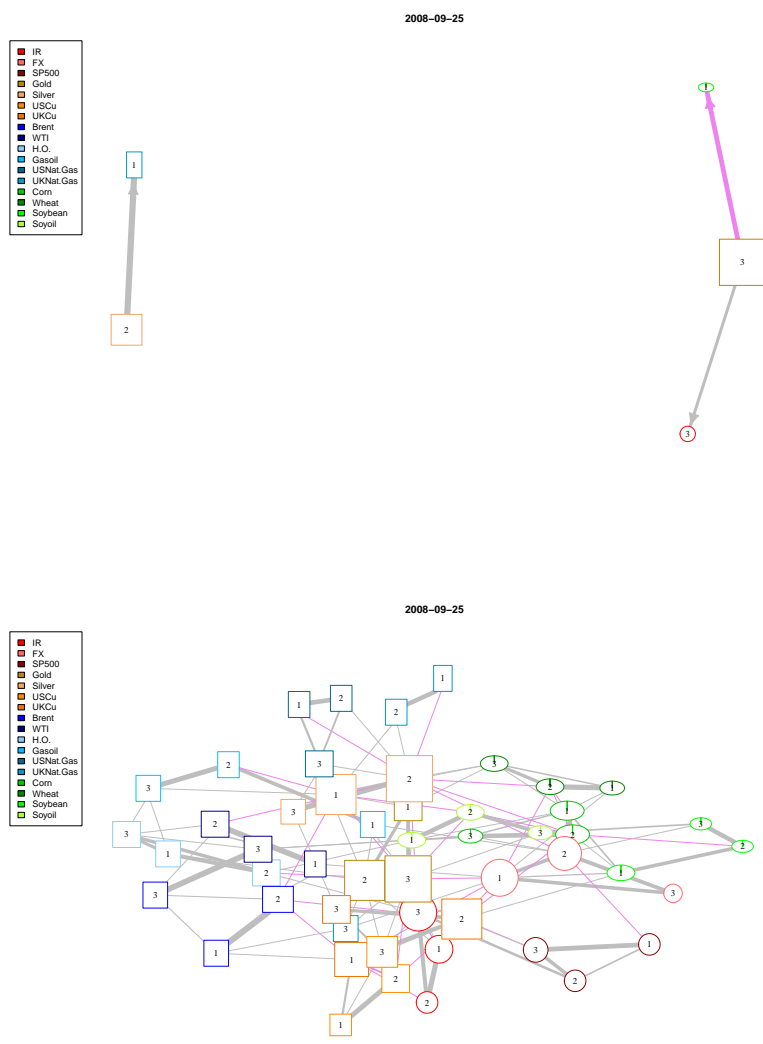
3.4.3 The lowest return on S&P500 (2008-10-09) (overlapping with Lehman)

We chose to study this event, because we noticed in Chapter 1 that all the S&P500 contracts experienced their minimum return (going down by at least -7.23%) on that day, some commodities experienced their minimum (or very infrequent) return the following day, on 2008-10-10. 33 of our 51 contracts had 1% negative shocks (observed), among which 10 are their minimum: USC_u1, USC_u3, UKC_u2, UKC_u3, H.O.1, Brent1, Brent2, WTI2, Soyoil1 and Soybean3. We thus wanted to check whether this was due to propagation from the S&P500 or not.

Unfortunately, the S&P500 does not appear in the Granger causality graphs around that day, so it does not explain what was observed. Nevertheless, there are more propagation links on 2008-10-09. The link from 12-month gold to 12-month eurodollar is still here and its coefficient increased to 0.20 (4 times larger than the maximum on the day before...). In addition, the 3-month UK copper joins to influence the 12-month eurodollar contract negatively (-0.11) and the 3-month eurodollar one positively (0.02). These extremum coefficients slowly decrease afterwards. And on 2008-10-17, agricultural markets disappear from the Granger causality graph. The system still forms a single component in the partial correlation graph and sectors are still scattered, but the USD/EUR exchange rate has made its way to the middle of the graph (and the front-month made it to the top of the centrality ranking), with silver and gold. As seen for Lehman Brothers, this may mean that capital flows have driven most of the returns, with these two precious metals. The 12-month gold contract temporarily dropped in centrality ranking on 2008-10-09 for a few days while other

3.4. SHOCKS AND PROPAGATION (DYNAMIC, ROLLING WINDOW)

Figure 3.7 – Granger causality (above) and partial correlation (below) graphs on the day of lowest return of the S&P500



The shape of the nodes represents their sector: circles for Finance, rectangles for Energy, ellipses for Agriculture and squares for Metals, with details of each market on the figure (different colours). The colour of the edges represents the sign of the weight/dependence: grey for positive, violet for negative. The size of the nodes represent their degree: the bigger the node, the higher its degree. The width of the edges represent the absolute value of their dependence: the wider the edge, the higher the dependence (in absolute value).

metals and longer maturities of brent and 3-month soy oil contracts became among the most central ones. Again, the 12-month S&P500 was also among the most important contracts, but not the spot. They had a 0.6076 PC on that day though, so they still could have been the source of the returns on that day (the following day, it was slightly lower, at 0.6010). Coefficients still range from about -0.2 to 0.9, but with a temporary drop in the minimum starting on the day of the event (2008-10-09). So in terms of total communicability, it dropped on that day and recovered slightly on the following day (day of minimum for several commodities, mentioned above).

In the end, we find that in those troubled times, markets have been largely changing (as seen with the centrality measure), but the lowest return observed on the S&P500 on 2008-10-09 did not affect those of commodities the following day. Other than that, the 3-month contract on S&P500 seems to be an important source of information in the system and could thus have triggered a systematic shock...

3.4.4 The largest return on S&P500 (2008-10-28) (overlapping with minimum S&P500)

The S&P500 maximum returns occurred (on 2008-10-28, up by more than 10.24%) in expectation of imminent rate cuts by central banks. On the following day (2008-10-29), 35 contracts experienced 1% positive shocks (observed), among which USCu1, USCu2, UKCu1, UKCu2, UKCu3, Silver1, Silver2, Silver3, Soyoil1, Soyoil2, Soyoil3, Wheat1, Soybean2 and Corn3 experienced their maximum return. Almost all markets experienced important returns (positive or negative), except for the UK natural gas. In particular, the returns on S&P500 were this time in the 10% bottom quantile, those of the eurodollar and gold contracts in the 5% quantile, respectively bottom and top. We thus wanted to assess whether financial equities shock propagated to these commodities and other assets.

Figures are available upon request, due to the large size of this file. Again, in terms of propagation, the S&P500 does not appear in the Granger causality graph. And around this event, the graph is rather stable, with a negative link from front-month silver to 12-month eurodollar coming in and out. The average coefficient was slightly positive but became negative after 2008-10-23 (not more than 10% in absolute value), with very little change. As for the partial correlation graphs, markets are still integrated (single component) and seem to come and go in the graph, with a widely spread energy sector (crude oils stand around the middle and meddling with others). For example, on the day of the event (2008-10-28), all sectors but energy seem well separated. While the S&P500 nodes seem at the periphery of the graph, its 12-month contract is still among the most central (though it is only connected to S&P500 contracts and with PC between 0.5 and 0.56), with silver, gold and the exchange rate. But on the following day, when some commodities experience their maximum return, all sectors are intertwined, with USD/EUR exchange rate, precious metals and Brent markets highly connected (high degree). Some agricultural markets have been very central on that day and experienced large returns. Maybe because of this centrality, other commodities like metals did too. The partial correlations ranged from more than -0.3 (though they increase and stabilise to about -0.1) to less than 0.9 (it dropped to 0.85 on 2008-11-11), just like in the previous 2008 events. Though the range has become smaller because of the decreasing maximum, the increasing minimum may present some risks as it means potential diversification effects have reduced.

To summarise, the highest return observed in our sample for the S&P500 does not seem to contribute through propagation to the highest return observed for several commodities on the following day. Though they are at the periphery of the partial correlation graph, the 12-month one belongs to the most central nodes in the network, joining precious metal markets. This may explain why many commodity markets experienced high returns on 2008-10-28.

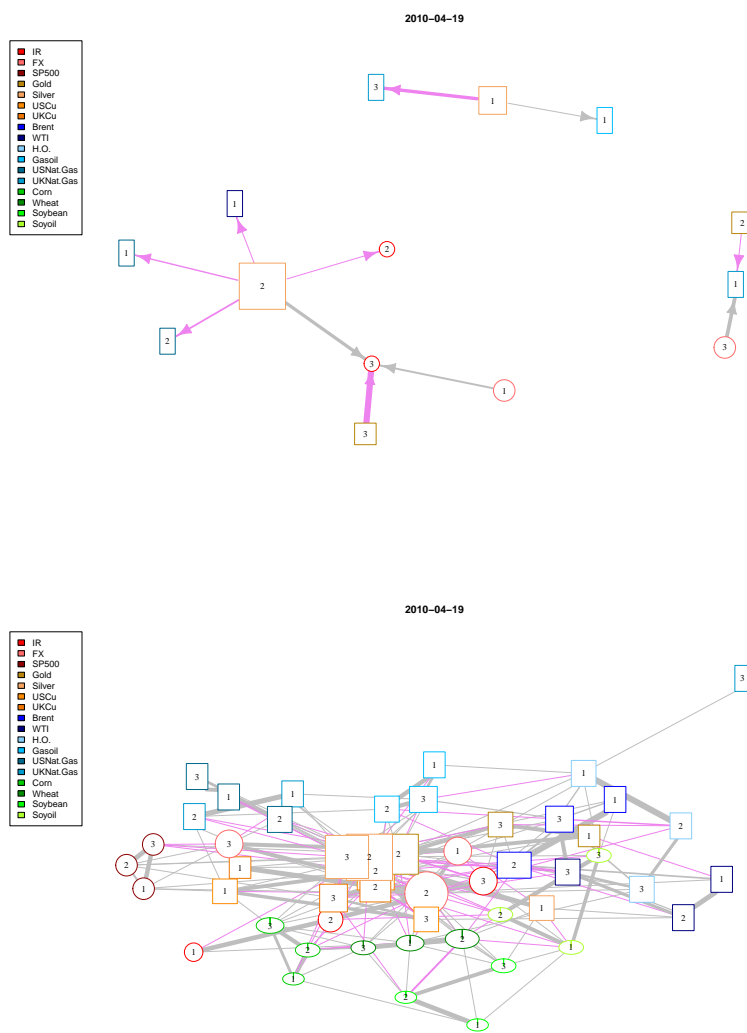
3.4.5 The Flash Crash (2010-05-06)

The Flash Crash corresponds to a financial event occurring on the e-mini S&P500 futures (that we have in our sample), at that time for maturity June 2010 (our SP500 2). There has been a sell order for 75,000 contracts, corresponding to around 4 billion dollars, which has triggered many reactions from other trading algorithms. Most of the losses were recovered quickly, but some equities were still impacted. We thus examine whether this few-hour financial "shock" could have had some impact on the real economy, through commodities.

On the day of the Flash Crash, the Granger causality graph has many more links than during our other events of interest and looks quite different from our reference graph. The S&P500 is still not present on the following day, so it may not have triggered propagation to commodities. Though this event occurred on the front-month E-mini S&P500 futures contract, its daily return on that day was -3.63%, while those of the front-month and 3-month eurodollar contract were more than 30% (the two interest rates increased from 5.25 bp to 7.10 bp). These returns do not seem to come from propagation, since other markets contributed only +0.009% and +0.039%. The 12-month eurodollar (which receives the most from others) seems to behave differently as we observe a 1.81% return, while the observations of the previous day contributed -0.86%. Here, silver contracts seem to propagate information to several energy markets and to the 3-month (new link) and 12-month eurodollar. In addition, the 12-month gold and some USD/EUR exchange rate contracts influence eurodollar ones too. This 12-month eurodollar contract gets the most influence from the others, with coefficients of -0.20 from the 12-month gold and 0.18 from the front-month exchange rate on the eve of the Flash Crash, turning to -0.34 and 0.33 respectively on the day of the event (then slightly taming down to -0.27 and 0.25 respectively)... Propagation risk has thus increased substantially on

3.4. SHOCKS AND PROPAGATION (DYNAMIC, ROLLING WINDOW)

Figure 3.8 – Granger causality (above) and partial correlation (below) graphs on the day of the Flash Crash



The shape of the nodes represents their sector: circles for Finance, rectangles for Energy, ellipses for Agriculture and squares for Metals, with details of each market on the figure (different colours). The colour of the edges represents the sign of the weight/dependence: grey for positive, violet for negative. The size of the nodes represent their degree: the bigger the node, the higher its degree. The width of the edges represent the absolute value of their dependence: the wider the edge, the higher the dependence (in absolute value).

the day of the Flash Crash, though towards a long-term eurodollar contract. In terms of partial correlations, the corresponding graph still consists of a single component, but there are many more links than the other events, as can be seen in Figure 3.11. The metals and financial sectors are at the center of the graph and intertwined, with gold, silver and exchange rate being the most connected markets. Energy markets are also scattered and among the most central nodes (still with the 12-month S&P500 at the top of the ranking), while the agricultural sector is still clustered. Before the day of the event, the number of links in the graph increases, being mostly negative partial correlation links. It hence makes the average partial correlation decrease, but we also find a large increase in the maximum partial correlation on the eve of the Flash Crash (from 0.81 for Brent1-Brent2 to 0.93 for WTI1-WTI2), which could raise concerns about systematic risk again since they are among the most central nodes.

Around the Flash Crash, markets seem to have integrated a lot as sectors are not as clearly separated as before in the partial correlation graph, with a large increase of the range of dependencies, in turn increasing the risk of systematic shock. In addition, relatively many propagation links exist at that time, also largely increasing their influence in absolute value, which means that propagation risk is also larger, but did not seem to represent a substantial risk.

Conclusion

The evolution of commodity prices since the early 2000s, raised concerns, since they seem closely related to equity prices. In addition, we observe that some commodity markets experienced their extremum return on the day after the S&P500 experienced its extremum over our sample (which occurred in October 2008, after the default of Lehman Brothers). This influence could lead to systemic risk, as commodities are part of the real economy. In order to properly assess systemic risk, it is important to dis-

entangle two components we identify as its sources: propagation risk and systematic risk (from market integration). A VAR framework is thus intuitive and suited for this task. We thus rely on the algorithm of Barigozzi and C. Brownlees (2017), which estimates the sparse AutoRegression (AR, Granger causality) and Partial Correlation (PC) matrices in a single step, allowing for an accurate balance between propagation and integration respectively.

We claim that the maturity dimension (using futures of different maturities) is necessary for assessing systemic risk as futures markets are a place for price discovery, according to the literature. We indeed find in our spatial reference graph that when the maturity dimension is excluded, the different sectors are separated, i.e. they do not exhibit common factors (in particular, financial markets and all commodity markets). On the contrary, once we include the 3-month and 12-month contracts in our 3D reference graph, all the sectors connect: they are highly integrated.

To study propagation, in particular the aforementioned observations, we conduct a dynamic analysis using a one-trading-year rolling window. We find that there was little propagation at play and that the S&P500 contracts do not even appear in the Granger causality graph. Instead, integration was the largest from mid-2008 in terms of number of factors and between 2008-09-15 (the default of Lehman Brothers) and 2009, many mode links were created (sharp increase). If we search for the most central markets in this PC graph, we find that gold, silver, S&P500 and USD/EUR exchange rate contracts are the ones driving the contemporaneous return innovations.

The next steps are to provide simulations and scenarios testing for a better assessment of the consequences of such scenarios on this system of markets.

References

Acharya, Viral V. et al. (2017). 'Measuring Systemic Risk'. In: *The Review of Financial Studies* 30(1), pp. 2–47.

- Adamic, Lada et al. (2009). ‘On the Informational Properties of Trading Networks’. In: *Working Paper*.
- Adrian, Tobias and Markus K. Brunnermeier (2016). ‘CoVaR’. In: *American Economic Review* 106(7), pp. 1705–41.
- Aït-Sahalia, Yacine, Julio Cacho-Diaz and Roger JA Laeven (2015). ‘Modeling financial contagion using mutually exciting jump processes’. In: *Journal of Financial Economics*.
- Allen, Franklin and Ana Babus (2009). ‘The Network Challenge: Strategy, Profit, and Risk in an Interlinked World’. In: ed. by Paul R. Kleindorfer and Yoram Wind. Upper Saddle River, N.J: Pearson Education. Chap. 21, pp. 367–382.
- Babus, Ana and Péter Kondor (2013). ‘Trading and information diffusion in over-the-counter markets’. In: *Working paper*.
- Banavar, Jayanth R., Amos Maritan and Andrea Rinaldo (1999). ‘Size and form in efficient transportation networks’. In: *Nature* 399(6732), pp. 130–132.
- Barigozzi, Matteo and Christian Brownlees (2017). ‘Nets: Network Estimation for Time Series’. In: *Working Paper*.
- BCBS (2011). *Global systemically important banks: assessment methodology and the additional loss absorbency requirement*. Tech. rep. Basel Committee on Banking Supervision, Bank for International Settlements.
- BCBS (2013). *Global systemically important banks: updated assessment methodology and the higher loss absorbency requirement*. Tech. rep. Basel Committee on Banking Supervision, Bank for International Settlements.
- Bekaert, Geert, Campbell R Harvey and Angela Ng (2005). ‘Market Integration and Contagion’. In: *Journal of Business* 78(1).
- Benoît, Sylvain, Jean-Edouard Colliard et al. (2017). ‘Where the Risks Lie: A Survey on Systemic Risk’. In: *Review of Finance* 21(1), pp. 109–152.
- Benoît, Sylvain, Christophe Hurlin and Christophe Pérignon (2015). ‘Implied Risk Exposures’. In: *Review of Finance* 19(6), pp. 2183–2222.
- Benzi, Michele and Christine Klymko (2013). ‘Total communicability as a centrality measure’. In: *Journal of Complex Networks* 1(2), pp. 124–149.

- Bessembinder, Hendrik et al. (1996). 'Is there a term structure of futures volatilities? Reevaluating the Samuelson hypothesis'. In: *Journal of Derivatives* 4(2), pp. 45–58.
- Billio, Monica and Massimiliano Caporin (2010). 'Market linkages, variance spillovers, and correlation stability: Empirical evidence of financial contagion'. In: *Computational Statistics and Data Analysis* 54(11), pp. 2443–2458.
- Billio, Monica, Mila Getmansky et al. (2012). 'Econometric measures of connectedness and systemic risk in the finance and insurance sectors'. In: *Journal of Financial Economics* 104(3), pp. 535–559.
- BIS (1992). *Recent Developments in International Interbank Relations*. Tech. rep. Bank for International Settlements.
- BIS (2001). *Report on Consolidation in the Financial Sector*. Tech. rep. Bank for International Settlements.
- BIS, FSB and IMF (2009). *Guidance to Assess the Systemic Importance of Financial Institutions, Markets and Instruments: Initial Considerations*. Tech. rep. Bank for International Settlements, Financial Stability Board and International Monetary Fund.
- Black, Fischer (1976). 'The pricing of commodity contracts'. In: *Journal of Financial Economics* 3(1), pp. 167–179.
- Bloch, Francis and Nicolas Qu erou (2013). 'Pricing in social networks'. In: *Games and Economic Behavior*.
- Bonacich, Phillip (1987). 'Power and Centrality: A Family of Measures'. In: *American Journal of Sociology* 92(5), pp. 1170–1182.
- Bonanno, G. et al. (2004). 'Networks of equities in financial markets'. In: *Eur. Phys. J. B* 38(2), pp. 363–371.
- Brennan, Michael J (1958). 'The supply of storage'. In: *The American Economic Review* 48(1), pp. 50–72.
- Brownlees, Christian T and Robert F Engle (2012). 'Volatility, correlation and tails for systemic risk measurement'. In: *Working paper*.
- Brunnermeier, Markus K. (2009). 'Deciphering the Liquidity and Credit Crunch 2007–2008'. In: *Journal of Economic Perspectives* 23(1), pp. 77–100.

- Büyüksahin, Bahattin, Michael S. Haigh and Michel A. Robe (2010). ‘Commodities and equities: ever a "market of one"?’ In: *Journal of Alternative Investment*, pp. 76–95.
- Büyüksahin, Bahattin and Michel A. Robe (2011). ‘Does Paper Oil Matter? Energy Markets’ Financialization and Equity-Commodity Co-Movements’. In: *Working Paper, American University, (Revised) July*.
- Büyüksahin, Bahattin and Michel A. Robe (2014). ‘Speculators, commodities and cross-market linkages’. In: *Journal of International Money and Finance* 42, pp. 38–70.
- Büyüksahin, Bahattin, Michel A. Robe and Valentina G. Bruno (2014). ‘The Financialization of Food?’ In: *Working Paper*.
- Caporale, Guglielmo Maria, Andrea Cipollini and Nicola Spagnolo (2005). ‘Testing for contagion: a conditional correlation analysis’. In: *Journal of Empirical Finance* 12(3), pp. 476–489.
- Capponi, Agostino and Martin Larsson (2015). ‘Price Contagion through Balance Sheet Linkages’. In: *The Review of Asset Pricing Studies* 5(2), pp. 227–253.
- Castiglionesi, Fabio, Fabio Feriozzi and Guido Lorenzoni (2009). ‘Financial integration, liquidity, and the depth of systemic crisis’. In: *Unpublished working paper. University of Tilburg*.
- Chakraborti, Anirban et al. (2003). ‘Dynamic asset trees and Black Monday’. In: *Physica A* 324(1-2), pp. 247–252.
- Chen, Shiu-Sheng and Hung-Chyn Chen (2007). ‘Oil prices and real exchange rates’. In: *Energy Economics* 29(3), pp. 390–404.
- Cheng, Ing-Haw and Wei Xiong (2014). ‘Financialization of commodity markets’. In: *Annu. Rev. Financ. Econ.* 6(1), pp. 419–441.
- Chinazzi, Matteo and Giorgio Fagiolo (2013). ‘Systemic risk, contagion, and financial networks: A survey’. In: *LEM Working Paper Series*.
- Cohen-Cole, Ethan et al. (2012). ‘Strategic Interactions on Financial Networks for the Analysis of Systemic Risk’. In: *Handbook on Systemic Risk*, p. 306.
- Corsetti, Giancarlo, Marcello Pericoli and Massimo Sbracia (2005). ‘Some contagion, some interdependence’: More pitfalls in tests of financial con-

- tagion'. In: *Journal of International Money and Finance* 24(8), pp. 1177–1199.
- CPSS and IOSCO (2012). *Principles for financial market infrastructures*. Tech. rep. Committee on Payment, Settlement Systems, Bank for International Settlements and International Organization of Securities Commissions.
- Deaton, Angus and Guy Laroque (1992). 'On the behaviour of commodity prices'. In: *The Review of Economic Studies* 59(1), pp. 1–23.
- Diebold, Francis X. and Kamil Yilmaz (2014). 'On the network topology of variance decompositions: Measuring the connectedness of financial firms'. In: *Journal of Econometrics* 182(1), pp. 119–134.
- Diebold, Francis X., Kamil Yilmaz and Laura Liu (2017). 'Commodity Connectedness'. In: *NBER, Working Paper No. 23685*.
- Dumontaux, Nicolas and Adrian Pop (2013). 'Contagion Effects in the Aftermath of Lehman's Collapse: Evidence from the US Financial Services Industry'. In:
- ECB (2012). *Financial Stability Review*. Tech. rep. European Central Bank.
- Eichengreen, Barry, Andrew K. Rose and Charles Wyplosz (1996). 'Contagious currency crises: first tests'. In: *Scandinavian Journal of Economics* 98, pp. 463–484.
- Erb, Claude B. and Campbell R. Harvey (2006). 'The Strategic and Tactical Value of Commodity Futures'. In: *Financial Analysts Journal* 62(2), pp. 69–97.
- Ewing, Bradley T, Farooq Malik and Ozkan Ozfidan (2002). 'Volatility transmission in the oil and natural gas markets'. In: *Energy Economics* 24(6), pp. 525–538.
- Fama, Eugene F. and Kenneth R. French (1987). 'Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage'. In: *The Journal of Business* 60(1), pp. 55–73.
- Fama, Eugene F. and Kenneth R. French (1988). 'Business Cycles and the Behavior of Metals Prices'. In: *The Journal of Finance* 43(5), pp. 1075–1093.

- Fattouh, Bassam, Lutz Killian and Lavan Mahadeva (2013). ‘The Role of Speculation in Oil Markets: What Have We Learned So Far?’ In: *The Energy Journal* 34(3).
- Fernando, Chitru S, Anthony D May and William L Megginson (2012). ‘The value of investment banking relationships: evidence from the collapse of Lehman Brothers’. In: *The Journal of Finance* 67(1), pp. 235–270.
- Ferraro, Domenico, Kenneth Rogoff and Barbara Rossi (2015). ‘Can oil prices forecast exchange rates? An empirical analysis of the relationship between commodity prices and exchange rates’. In: *Journal of International Money and Finance* 54, pp. 116–141.
- Figuerola-Ferretti, Isabel and Jesús Gonzalo (2010). ‘Modelling and measuring price discovery in commodity markets’. In: *Journal of Econometrics* 158(1), pp. 95–107.
- Forbes, Kristin J. and Roberto Rigobon (2002). ‘No Contagion, Only Interdependence: Measuring Stock Market Comovements’. In: *The Journal of Finance* 57(5), pp. 2223–2261.
- Fréchet, Maurice (1906). ‘Sur quelques points du calcul fonctionnel’. In: *Rendiconti del Circolo Mathematico di Palermo* 22, pp. 1–74.
- Fund, International Monetary (2011). ‘Global Financial Stability Report’. In: *World Economic and Financial Surveys*.
- Gai, Prasanna and Sujit Kapadia (2010). ‘Contagion in financial networks’. In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. The Royal Society.
- Garbade, Kenneth D. and William L. Silber (1983). ‘Price Movements and Price Discovery in Futures and Cash Markets’. In: *The Review of Economics and Statistics* 65(2), pp. 289–297.
- Giesecke, Kay and Baeho Kim (2011). ‘Systemic Risk: What Defaults Are Telling Us’. In: *Management Science* 57(8), pp. 1387–1405.
- González-Hermosillo, Brenda and Heiko Hesse (2011). ‘Global market conditions and systemic risk’. In: *Journal of Emerging Market Finance* 10(2), pp. 227–252.

- Gorton, Gary B, Fumio Hayashi and K Geert Rouwenhorst (2013). ‘The fundamentals of commodity futures returns’. In: *Review of Finance* 17(1), pp. 35–105.
- Gorton, Gary B and K Geert Rouwenhorst (2006). ‘Facts and Fantasies about Commodity Futures’. In: *Financial Analysts Journal* 62(2), pp. 47–68.
- Gower, John C. (1966). ‘Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis’. In: *Biometrika* 53(3/4), pp. 325–338.
- Granger, C. W. J. (1969). ‘Investigating Causal Relations by Econometric Models and Cross-spectral Methods’. In: *Econometrica* 37(3), pp. 424–438.
- Haldane, Andrew G and Robert M May (2011). ‘Systemic risk in banking ecosystems’. In: *Nature* 469(7330), pp. 351–355.
- Hamilton, James D. (2009). ‘Causes and Consequences of the Oil Shock of 2007-08.’ In: *Brookings Papers on Economic Activity*, pp. 215–261.
- Hammoudeh, Shawkat and Yuan Yuan (2008). ‘Metal volatility in presence of oil and interest rate shocks’. In: *Energy Economics* 30(2), pp. 606–620.
- Hartmann, Philipp, Stefan Straetmans and Casper G De Vries (2004). ‘Asset market linkages in crisis periods’. In: *Review of Economics and Statistics* 86(1), pp. 313–326.
- Hasbrouck, Joel (2003). ‘Intraday Price Formation in U.S. Equity Index Markets’. In: *The Journal of Finance* 58(6), pp. 2375–2400.
- Hautsch, Nikolaus, Julia Schaumburg and Melanie Schienle (2015). ‘Financial Network Systemic Risk Contributions’. In: *Review of Finance* 19(2), pp. 685–738.
- Heinkel, Robert, Maureen E. Howe and John S. Hughes (1990). ‘Commodity convenience yields as an option profit’. In: *Journal of Futures Markets* 10(5), pp. 519–533.
- Hicks, J. R. (1946). *Value and Capital: An Inquiry into some Fundamental Principles of Economic Theory*. Oxford University Press.

- Illing, Mark and Ying Liu (2006). ‘Measuring financial stress in a developed country: An application to Canada’. In: *Journal of Financial Stability* 2(3), pp. 243–265.
- IMF (2008). *Global Financial Stability Report*. Tech. rep. International Monetary Fund.
- IMF (2010). *Global Financial Stability Report*. Tech. rep. IMF.
- IMF (2011a). *Global Financial Stability Report*. Tech. rep. International Monetary Fund.
- IMF (2011b). *Global Financial Stability Report*. Tech. rep. International Monetary Fund.
- Irwin, Scott H and Dwight R Sanders (2011). ‘Index funds, financialization, and commodity futures markets’. In: *Applied Economic Perspectives and Policy* 33(1), pp. 1–31.
- Ishihara, Yoichiro (2005). ‘Quantitative analysis of crisis: Crisis identification and causality’. In: *World Bank Policy Research Working Paper* (3598).
- Joëts, Marc, Valérie Mignon and Tovonony Razafindrabe (2017). ‘Does the volatility of commodity prices reflect macroeconomic uncertainty?’ In: *Energy Economics*.
- Jorion, Philippe and Gaiyan Zhang (2007). ‘Good and bad credit contagion: Evidence from credit default swaps’. In: *Journal of Financial Economics* 84(3), pp. 860–883.
- Jorion, Philippe and Gaiyan Zhang (2009). ‘Credit Contagion from Counterparty Risk’. In: *The Journal of Finance* 64(5), pp. 2053–2087.
- Kaldor, Nicholas (1939). ‘Speculation and Economic Stability’. In: *The Review of Economic Studies* 7(1), pp. 1–27.
- Katz, Leo (1953). ‘A New Status Index Derived From Sociometric Analysis’. In: *Psychometrika* 18(1), pp. 39–43.
- Kaufman, George G. (1995). *Research in Financial Services*. Vol. 7. London: JAI Press.
- Keynes, John Maynard (1930). *A Treatise on Money: In 2 Volumes*. Macmillan & Company.

- Kilian, Lutz (2009). ‘Causes and Consequences of the Oil Shock of 2007-08: Comment’. In: *Brookings Papers on Economic Activity*, pp. 267–278.
- Korniotis, George M. (2009). ‘Does Speculation Affect Spot Prices Levels? The Case of Metals with and without Futures Markets’. In: *Finance and Economics Discussion Series Working Paper 2009-29, Board of Governors of the Federal Reserve*.
- More on oil and speculation* (2008) May 13.
- Lautier, Delphine, Julien Ling and Franck Raynaud (2015). ‘Integration of Commodity Derivative Markets: Has It Gone Too Far?’ In: *Commodities, Energy and Environmental Finance*. Ed. by René Aïd, Michael Ludkovski and Ronnie Sircar. Vol. 74. Fields Institute Communications. Springer New York, pp. 65–90.
- Lautier, Delphine and Franck Raynaud (2012). ‘Systemic Risk in Energy Derivative Markets: A Graph-Theory Analysis’. In: *The Energy Journal* 33(3), pp. 215–239.
- Lautier, Delphine and Franck Raynaud (2014). ‘Information flows in the term structure of commodity prices’. In: *Working Paper*.
- Le Pen, Yannick and Benoît Sévi (2018). ‘Futures Trading and the Excess Co-movement of Commodity Prices’. In: *Review of Finance* 22(1), pp. 381–418.
- Longin, Francois and Bruno Solnik (2001). ‘Extreme correlation of international equity markets’. In: *The Journal of Finance* 56(2), pp. 649–676.
- Longstaff, Francis A (2010). ‘The subprime credit crisis and contagion in financial markets’. In: *Journal of Financial Economics* 97(3), pp. 436–450.
- Lutz, Björn (2010). *Pricing of Derivatives on Mean-Reverting Assets*. Vol. 630. Lecture Notes in Economics and Mathematical Systems. Springer-Verlag Berlin Heidelberg.
- Mantegna, R. N. (1999). ‘Hierarchical structure in financial markets’. In: *The European Physical Journal B* 11, pp. 193–197.
- Mantegna, R.N. and H.E. Stanley (1999). *Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press.

ISBN: 9781139431224. URL: <https://books.google.fr/books?id=SzgXWCS7Nr8C>.

- Markose, Sheri M. (2012). ‘Systemic Risk from Global Financial Derivatives: A Network Analysis of Contagion and Its Mitigation with Super-Spreader Tax’. In: *IMF Working paper WP12282*.
- Markowitz, Harry (1952). ‘Portfolio Selection’. In: *The Journal of Finance* 7(1), pp. 77–91.
- Martinez-Jaramillo, Serafin et al. (2014). ‘An empirical study of the Mexican banking system’s network and its implications for systemic risk’. In: *Journal of Economic Dynamics and Control* 40, pp. 242–265.
- Martínez-Jaramillo, Serafin et al. (2010). ‘Systemic risk, financial contagion and financial fragility’. In: *Journal of Economic Dynamics and Control* 34(11), pp. 2358–2374.
- Masters, Michael W. (2008). *before the Committee on Homeland Security and Governmental Affairs*. Speech.
- Miller, J. Isaac and Ronald A. Ratti (2009). ‘Crude oil and stock markets: Stability, instability, and bubbles’. In: *Energy Economics* 31(4), pp. 559–568.
- Moshirian, Fariborz and Qiongbing Wu (2009). ‘Banking industry volatility and banking crises’. In: *Journal of International Financial Markets, Institutions and Money* 19(2), pp. 351–370.
- Nash, Daniel J. (2001). ‘Long-term Investing in Commodities’. In: *Global Pensions Quarterly, Morgan Stanley Global Pensions* 4(1), pp. 25–31.
- Newbery, David M. (1987). ‘When do Futures Destabilize Spot Prices?’ In: *International Economic Review* 28 (2), pp. 291–297.
- Ng, Victor K. and Stephen Craig Pirrong (1994). ‘Fundamentals and Volatility: Storage, Spreads, and the Dynamics of Metals Prices’. In: *The Journal of Business* 67(2), pp. 203–230.
- Ng, Victor K. and Stephen Craig Pirrong (1996). ‘Price dynamics in refined petroleum spot and futures markets’. In: *Journal of Empirical Finance* 2(4), pp. 359–388.
- Nier, Erlend et al. (2007). ‘Network models and financial stability’. In: *Journal of Economic Dynamics and Control* 31(6), pp. 2033–2060.

- Oet, Mikhail V et al. (2013). 'SAFE: An early warning system for systemic banking risk'. In: *Journal of Banking & Finance* 37(11), pp. 4510–4533.
- Onnela, J-P, Anirban Chakraborti et al. (2003). 'Dynamics of market correlations: Taxonomy and portfolio analysis'. In: *Physical Review E* 68(5), p. 056110.
- Onnela, J-P, A Chakraborti et al. (2002). 'Dynamic asset trees and portfolio analysis'. In: *The European Physical Journal B-Condensed Matter and Complex Systems* 30(3), pp. 285–288.
- Park, Jungwook and Ronald A. Ratti (2008). 'Oil price shocks and stock markets in the U.S. and 13 European countries'. In: *Energy Economics* 30(5), pp. 2587–2608.
- Pesaran, M Hashem and Andreas Pick (2007). 'Econometric issues in the analysis of contagion'. In: *Journal of Economic Dynamics and Control* 31(4), pp. 1245–1277.
- Pindyck, Robert and Julio Rotemberg (1990). 'The excess co-movement of commodity prices'. In: *Economic Journal* 100.
- Quan, Jing (1992). 'Twostep testing procedure for price discovery role of futures prices'. In: *Journal of Futures Markets* 12(2), pp. 139–149.
- Sari, Ramazan, Shawkat Hammoudeh and Ugur Soytas (2010). 'Dynamics of oil price, precious metal prices, and exchange rate'. In: *Energy Economics* 32(2), pp. 351–362.
- Schwarcz, Steven L. (2008). 'Systemic Risk'. In: *Georgetown Law Journal* 97, p. 193.
- Schwartz, Eduardo S. (1997). 'The Stochastic Behavior of Commodity Prices: Implications for Valuation and Hedging'. In: *The Journal of Finance* 52(3), pp. 923–973.
- Schwarz, Thomas V. and Andrew C. Szakmary (1994). 'Price discovery in petroleum markets: Arbitrage, cointegration, and the time interval of analysis'. In: *Journal of Futures Markets* 14(2), pp. 147–167.
- Silvennoinen, Annastiina and Susan Thorp (2013). 'Financialization, crisis and commodity correlation dynamics'. In: *Journal of International Financial Markets, Institutions and Money* 24, pp. 42–65. ISSN: 1042-4431.

- Silvennoinen, Annastiina and Susan Thorp (2016). ‘Crude Oil and Agricultural Futures: An Analysis of Correlation Dynamics’. In: *Journal of Futures Markets* 36(6), pp. 522–544.
- Stoll, Hans R. and Robert E. Whaley (1990). ‘The Dynamics of Stock Index and Stock Index Futures Returns’. In: *The Journal of Financial and Quantitative Analysis* 25(4), pp. 441–468.
- Tang, Ke and Wei Xiong (2012). ‘Index Investing and the Financialization of Commodities’. In: *Financial Analysts Journal* 68(6), pp. 54–74.
- Telser, Lester G (1958). ‘Futures trading and the storage of cotton and wheat’. In: *Journal of Political Economy* 66(3), pp. 233–255.
- Till, Hilary (2006). ‘A Review of the Historical Literature on Commodity Futures Contract Returns’. In: *Working paper*.
- West, Douglas B. (2001). *Introduction to graph theory*. Prentice Hall.
- Working, Holbrook (1949). ‘The Theory of Price of Storage’. In: *The American Economic Review* 39(6), pp. 1254–1262.
- Zhang, Yue-Jun et al. (2008). ‘Spillover effect of US dollar exchange rate on oil prices’. In: *Journal of Policy Modeling* 30(6), pp. 973–991.

Appendices

3.A Correlations vs. partial correlations

When we want to assess the dependence between two random variables, the linear correlation coefficient is commonly used, even though it has flaws. One of them is that it may be inflated if we do not take into account variables that may have an effect on both variables under consideration. More formally, if we take a set of three random variables X, Y, Z , we can compute their unconditional correlation coefficients, in their unconditional correlation matrix ρ :

$$\rho = \begin{pmatrix} 1 & \rho_{X,Y} & \rho_{X,Z} \\ \rho_{X,Y} & 1 & \rho_{Y,Z} \\ \rho_{X,Z} & \rho_{Y,Z} & 1 \end{pmatrix} \quad (3.3)$$

If we want to assess the dependence between X and Y and just look at $\rho_{X,Y}$, we may get a biased result. Indeed, what if Z is a common factor for these two variables or an intermediary variable between the two? It would create a "artificial" unconditional correlation $\rho_{X,Y}$. Hence, we need to filter out the influence of the variable Z from the dependence between X and Y . The partial correlation $C_{X,Y}$ between the variables X and Y precisely serves this purpose: it corresponds to the linear correlation between these two variables, conditional on the other variables, here Z . We would thus get the following partial correlation matrix \mathbf{C} :

$$\mathbf{C} = \begin{pmatrix} 1 & C_{X,Y} & C_{X,Z} \\ C_{X,Y} & 1 & C_{Y,Z} \\ C_{X,Z} & C_{Y,Z} & 1 \end{pmatrix} = \begin{pmatrix} 1 & \rho_{(X,Y)|Z} & \rho_{(X,Z)|Y} \\ \rho_{(X,Y)|Z} & 1 & \rho_{(Y,Z)|X} \\ \rho_{(X,Z)|Y} & \rho_{(Y,Z)|X} & 1 \end{pmatrix} \quad (3.4)$$

where $|A$ denotes the conditionality on variable $A \in \{X, Y, Z\}$. This partial correlation matrix can be easily obtained from the unconditional correlation matrix: we take the inverse of the unconditional correlation matrix ρ (or covariance matrix), then normalise it (by dividing each term by the corresponding diagonal terms) and finally multiply by -1 as follows:

$$\mathbf{C} = - \begin{pmatrix} \frac{D_{X,X}}{D_{X,X}^{1/2} D_{X,X}^{1/2}} & \frac{D_{X,Y}}{D_{X,X}^{1/2} D_{Y,Y}^{1/2}} & \frac{D_{X,Z}}{D_{X,X}^{1/2} D_{Z,Z}^{1/2}} \\ \frac{D_{X,Y}}{D_{X,X}^{1/2} D_{Y,Y}^{1/2}} & \frac{D_{Y,Y}}{D_{Y,Y}^{1/2} D_{Y,Y}^{1/2}} & \frac{D_{Y,Z}}{D_{Y,Y}^{1/2} D_{Z,Z}^{1/2}} \\ \frac{D_{X,Z}}{D_{X,X}^{1/2} D_{Z,Z}^{1/2}} & \frac{D_{Y,Z}}{D_{Y,Y}^{1/2} D_{Z,Z}^{1/2}} & \frac{D_{Z,Z}}{D_{Z,Z}^{1/2} D_{Z,Z}^{1/2}} \end{pmatrix} \quad (3.5)$$

$$= \begin{pmatrix} -1 & -\frac{D_{X,Y}}{D_{X,X}^{1/2} D_{Y,Y}^{1/2}} & -\frac{D_{X,Z}}{D_{X,X}^{1/2} D_{Z,Z}^{1/2}} \\ -\frac{D_{X,Y}}{D_{X,X}^{1/2} D_{Y,Y}^{1/2}} & -1 & -\frac{D_{Y,Z}}{D_{Y,Y}^{1/2} D_{Z,Z}^{1/2}} \\ -\frac{D_{X,Z}}{D_{X,X}^{1/2} D_{Z,Z}^{1/2}} & -\frac{D_{Y,Z}}{D_{Y,Y}^{1/2} D_{Z,Z}^{1/2}} & -1 \end{pmatrix} \quad (3.6)$$

where $\mathbf{D} = \rho^{-1}$ (or the inverse of covariance matrix). The partial correlation matrix is to the concentration matrix (inverse of the covariance matrix) what the correlation matrix is to the covariance matrix. We can also obtain it by regressing each variable on the others and after some computations on the regression coefficients.

Let us take the three eurodollar contracts from our 3D reference graph (Figure 3.3) as an example. If we compute the correlation matrix ρ for these three time series (IR1, IR2, IR3), we get very high correlation coefficients:

$$\rho = \begin{pmatrix} 1 & 0.9994 & 0.9424 \\ 0.9994 & 1 & 0.9533 \\ 0.9424 & 0.9533 & 1 \end{pmatrix} \quad \mathbf{\Gamma} = \begin{pmatrix} 0.0003615 & 0.0003557 & 0.0002504 \\ 0.0003557 & 0.0005651 & 0.0004958 \\ 0.0002504 & 0.0004958 & 0.0011149 \end{pmatrix} \quad (3.7)$$

We note that the highest coefficients are, in descending order, the one between IR1 and IR2 (0.9994), the one between IR2 and IR3 (0.9533) and finally the one between IR1 and IR3 (0.9424). These values may be inflated by the influence of the variables not involved in each pair. Let us denote \mathbf{F} the concentration matrix (inverse of the covariance matrix $\mathbf{\Gamma}$, also called information or precision matrix). If we look at the partial correlations by normalising the concentration matrix, we find the following:

$$\mathbf{C} = \begin{pmatrix} -1 & 0.7403 & -0.1716 \\ 0.7403 & -1 & 0.5172 \\ -0.1716 & 0.5172 & -1 \end{pmatrix} \quad \mathbf{F} = \mathbf{\Gamma}^{-1} = \begin{pmatrix} 7838.30 & -5507.77 & 632.03 \\ -5507.77 & 7061.22 & -1807.80 \\ 632.03 & -1807.80 & 1730.46 \end{pmatrix} \quad (3.8)$$

We see that after conditioning on all the other variables in our system (of 51 variables), the linear dependence between IR1 and IR2 is only $C_{IR1,IR2} = 0.7403$ now instead of $\rho_{IR1,IR2} = 0.9994$. Hence, the difference of 0.2591 was due to the influence of other variables. The same applies to the other coefficients, but we interestingly find that the partial correlation between IR1 and IR3 is actually negative, $C_{IR1,IR3} = -0.1716$, while the unconditional one was $\rho_{IR1,IR3} = 0.9424$. Therefore, if we do not control for the influence of other variables that may alter the unconditional correlation matrix, we may get different results.

In addition, the concentration matrix \mathbf{F} encodes the conditional dependence between pairs of variables in a multivariate normal distribution. Hence, an entry $C_{i,j}$ of the partial correlation matrix (which has the same 0s as the information matrix) is equal to 0 if and only if X_i and X_j are independent, conditional on $\{X_k\}_{k \neq i,j}$. Though the multivariate Gaussian distribution is a strong assumption to make in empirical studies, it is a common one.

We can thus build a graph based on those coefficients, with a link when the partial correlation is nonzero as we do in this paper. Nevertheless, even if some variables are conditionally independent, they can still be unconditionally dependent. Let us consider three variables X, Y, Z again. If their partial correlation graph is

$$X - -Y - -Z$$

then the partial correlation between X and Z is 0, but their unconditional correlation is not (see Eq. 3.12 and 3.13 for a numerical example).

This feature of paths of partial correlations can highlight the presence of potential common factors when some nodes in the graph form a separate component, as in the reference partial correlation graph in the spatial dimension, in Figure 3.2, which we use as a way to visualise and measure the integration in our system (if different components/sectors start connecting, they become integrated).

All in all, we choose to work on partial correlations instead of unconditional correlations for their many benefits, but still come and go between the two, since unconditional correlations are also useful.

3.B Explanation of LASSO penalization

Let us consider a simple linear model of the form:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \epsilon \quad (3.9)$$

where Y , the dependent variable, is explained by a constant and p explanatory variables $\{X_1, \dots, X_p\}$. If we have a set of data of length n (we have n observations of Y , (y_1, \dots, y_n) , and the corresponding n observations of $\{X_1, \dots, X_p\}$, $(x_{i,1}, \dots, x_{i,p})_{i=1..n}$), we will try to find the parameters $(\beta_0, \dots, \beta_p)$ that will allow the estimated values \hat{Y} (on the regression line) to be the closest to the observed Y . Formally, in the case of an Ordinary Least Squares (where we minimise the sum of squared errors) estimation, the objective function would be:

$$\min_{\beta_0, \beta_1, \dots, \beta_p} \frac{1}{2} \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 \quad (3.10)$$

where j stands for variable X_j and i stands for observation i (of Y or X_j). In traditional Econometrics works, we do not work in high dimension and have parameter estimates that can be tested. We rely on these statistical tests to identify the important explanatory variables of the regressions and analyse their parameters more particularly.

Penalizing with LASSO would slightly change that objective function by adding a term (which represents a constraint):

$$\min_{\beta_0, \beta_1, \dots, \beta_p} \frac{1}{2} \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (3.11)$$

with $\lambda \geq 0$ the penalizing parameter. This parameter basically defines how much sparsity we want in the parameters: the greater this λ , the more

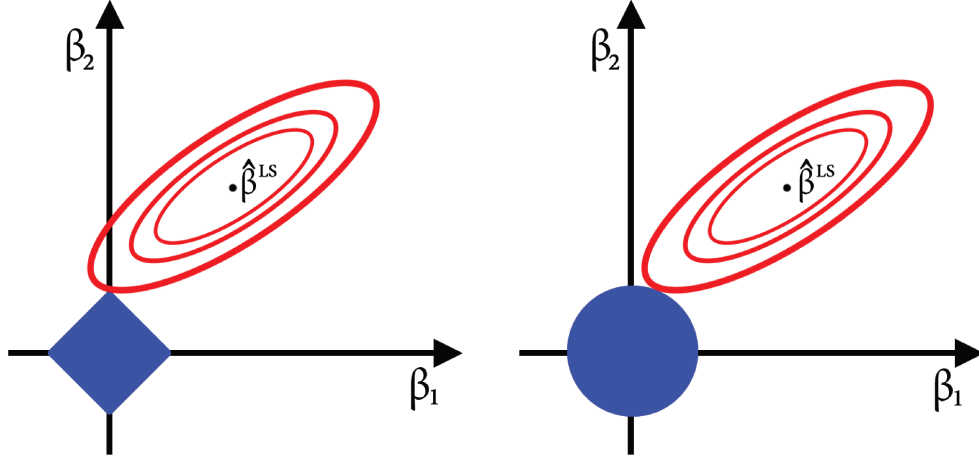
coefficients will be set to 0. Indeed, if coefficients are very small, a large λ will increase the value of the objective function, so in order to minimise this function, we should set them to 0, so that they don't contribute to the minimisation problem anymore. Meanwhile, the nonzero parameters are calibrated at the same time, taking into account the 0s. We should thus obtain more realistic estimates of those nonzero parameters than if we calibrate the parameters first (if it is even possible, since we are in high dimension) and only keep the statistically significant ones. As we will see below, the estimates for the nonzero parameters will compensate for setting some parameters to 0.

Other regularizations are possible, but they may not set the small ones to 0. Another famous one is the Ridge regularization, which will provide many nonzero coefficients, due to the form of the constraint. Figure 3.9 shows how these two types of constraints translate into geometry. LASSO penalises the objective function with an L1-norm on the parameters, $|\beta|$, while ridge penalises it with an L2-norm, β^2 , hence their shape in the figure (square vs. circle respectively, in blue). If we consider these regularizations as constraints in the optimization problem, we see in this figure that, on the one hand, the LASSO constraint will generally give an optimum in a corner of the square, setting a parameter to 0 (β_1 here). On the other hand, with the Ridge constraint, an optimum may be found with both parameters being different from 0.

As per the results from Chapter 2 in particular, the partial correlation matrix (or the concentration matrix) could potentially be sparse, which comforts us in using this approach. It should at least be sparser than the unconditional correlation matrix, since it is "polluted" by the influences of all the variables. Applying this penalization to the calibration of the the concentration matrix \mathbf{F} , we would get many 0s, meaning conditional independence between many variables, after filtering out the influence of the others.

Let us come back to our example with the three eurodollar contracts (IR1, IR2 and IR3) from Appendix 3.A. We may wonder whether the coeffi-

Figure 3.9 – Geometry point of view, difference between LASSO (left) and Ridge (right) prenalizations



The figure was taken from Wikipedia. The red ellipses represent different level ellipses of the objective function of a least square error minimisation problem. The blue areas represent the constraints added: square-shaped for LASSO (L1-norm, in the form of $|\beta|$, on the left) and circle-shaped for ridge regularization (L2-norm, in the form of β^2 , on the right).

cient $F_{IR1,IR3} = 632.0388$ (and maybe even $F_{IR2,IR3} = -1807.8044$ and $F_{IR3,IR3} = 1730.4653$, considering other values are very large) is "normal" or if it is small enough compared to the others to be set to 0. If we look at their partial correlations from the whole 51×51 matrix, we get the following PC matrix \mathbf{C}^{LASSO} and concentration matrix \mathbf{F}^{LASSO} (which was calibrated with LASSO and then the PC was derived from it):

$$\mathbf{C}^{LASSO} = \begin{pmatrix} -1 & 0.6548 & 0 \\ 0.6548 & -1 & 0.5211 \\ 0 & 0.5211 & -1 \end{pmatrix} \quad \mathbf{F}^{LASSO} = \begin{pmatrix} 6999.69 & -4313.20 & 0.000 \\ -4313.20 & 6199.35 & -1632.13 \\ 0.000 & -1632.13 & 1582.37 \end{pmatrix} \quad (3.12)$$

Their Looking at the reference graph in Figure 3.3, we indeed see that there is no link between these two nodes. Comparing with the partial correlation matrix (and the corresponding concentration matrix) found in Eq. 3.8, the coefficient between IR2 and IR3 is slightly larger when using LASSO (0.5211 vs. 0.5172). The coefficient between IR1 and IR2 is lower when penalizing with LASSO (0.6548 vs. 0.7403). This may be explained by the 0 set for

$C_{IR1,IR3}^{LASSO}$: while it was negative when not using LASSO, to compensate for this mitigating effect (of the negative coefficient), the dependence between IR1 and IR2 may have been decreased (or it could have affected other partial correlations too).

Looking at the corresponding unconditional correlations, we find the following matrix (extracted from the 51×51 matrix):

$$\rho = \begin{pmatrix} 1 & 0.9994 & 0.9424 \\ 0.9994 & 1 & 0.9533 \\ 0.9424 & 0.9533 & 1 \end{pmatrix} \quad \rho^{LASSO} = \begin{pmatrix} 1 & 0.9169 & 0.8370 \\ 0.9169 & 1 & 0.9230 \\ 0.8370 & 0.9230 & 1 \end{pmatrix} \quad (3.13)$$

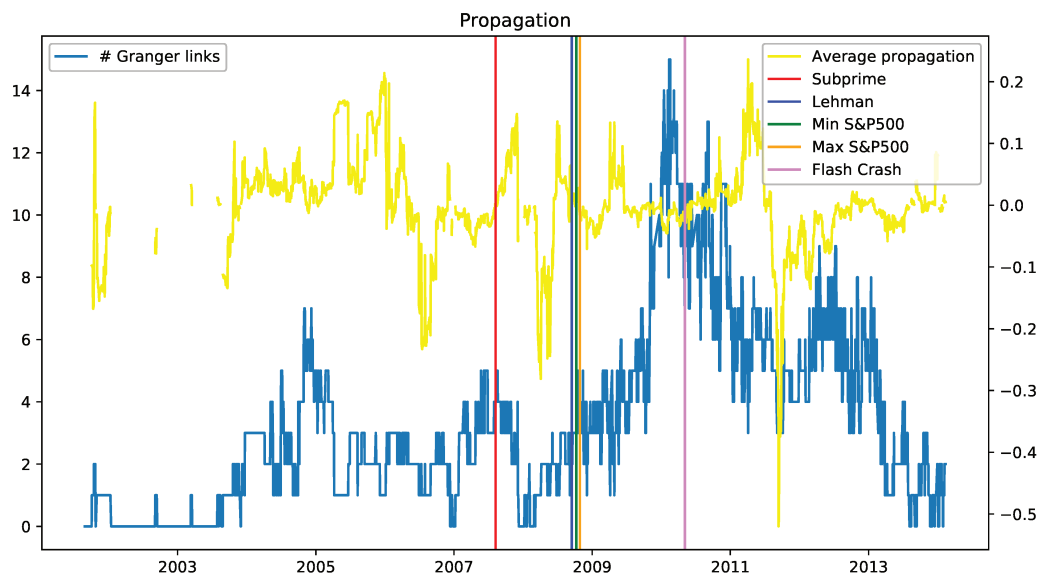
The coefficients are not exactly equal to the ones in the previous section, without using LASSO (and the ordering is not the same), but are still quite close. The difference may also be due to some noise in the unconstrained case.

Moreover, applying the *nets* algorithm of Barigozzi and C. Brownlees (2017) applies LASSO penalization to the estimation of both the AutoRegression and partial correlation matrices in one step. This allows the optimization to have better convergence properties and allows it to select the parameters among both AR coefficients and partial correlation coefficients, instead of selecting the former first and then the latter.

3.C Measures in the dynamic analysis over the whole period

3.C.1 Measures related to the propagation (Granger causality)

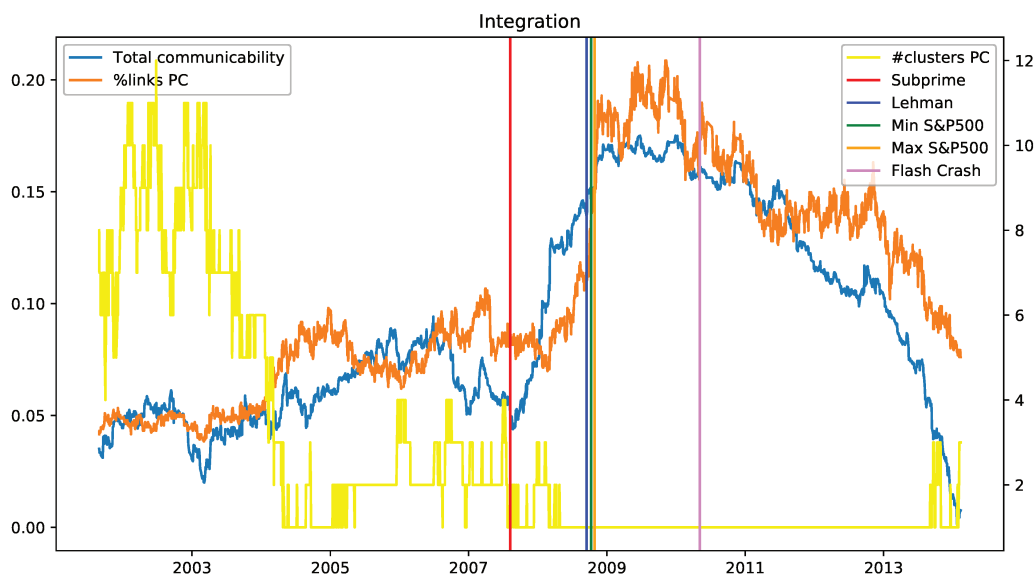
Figure 3.10 – Measures related to the links of the propagation (Granger causality)



The vertical bars represent the events under consideration: red for the beginning of the subprime crisis, blue for the default of Lehman Brothers, green for the day of minimum return on the S&P500, orange for the day of its maximum return and violet for the Flash Crash. The blue line (#Granger links) represents the number of links of propagation. The yellow line (Average propagation) represents the average value of the propagation coefficients. Note that when there is no link, the average is NA, hence the discontinuity.

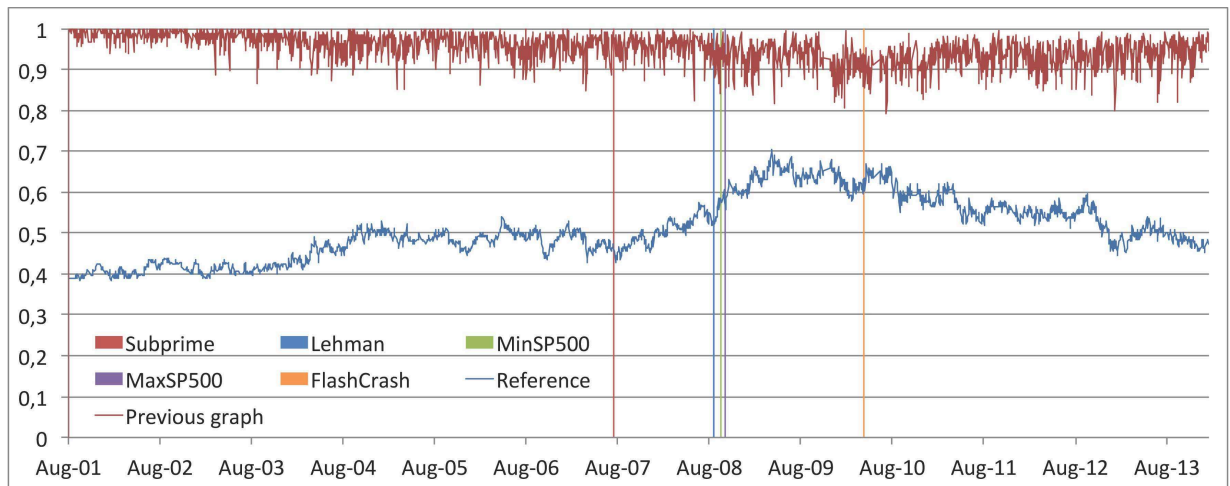
3.C.2 Measures related to the integration (partial correlation, PC)

Figure 3.11 – Measures related to the integration of markets (PC)



The vertical bars represent the events under consideration: red for the beginning of the subprime crisis, blue for the default of Lehman Brothers, green for the day of minimum return on the S&P500, orange for the day of its maximum return and violet for the Flash Crash. The blue line (Total communicability) represents the total of the easiness of information flow in the system; the values have been normalized by the minimum value to fit on the same scale. The orange line (%links PC) represents the inverse of the sparsity in the graph; it gives the proportion of nonzero PC in terms of total links possible (full graph). The yellow line (#clusters PC) represents the number of factors (components, clusters) in the PC graph.

Figure 3.12 – Measures related to the structure of the partial correlation graphs (3)



The vertical lines represent the events under consideration: red for the beginning of the subprime crisis, blue for the default of Lehman Brothers, green for the day of minimum return on the S&P500, purple for the day of its maximum return and orange for the Flash Crash. The blue line (Reference) represents the proportion of links that are common to the daily graphs and the reference graph. The red line (Previous graph) represents the proportion of links that are common to the graph on that day and the graph on the previous day.

Conclusion et développements futurs

La financiarisation des matières premières a entraîné une profonde modification de ces marchés. La présence de nombreux agents financiers a créé une forte intégration des commodités, plus encore lorsqu'elles apparaissent dans la composition des grands indices tels que le S&P500-GSCI ou le Bloomberg Index (anciennement DJ-UBSCI). Les positions longues de nombre d'investisseurs institutionnels sur ces indices et de spéculateurs, notamment les *hedge funds*, a en outre provoqué une forte intégration de ces matières premières avec les actifs financiers traditionnels.

Cette financiarisation s'est d'abord faite à l'aide de contrats *futures* sur matières premières, beaucoup plus faciles à échanger que leurs sous-jacents physiques. Or ces contrats ont progressivement hérité de la majeure partie du processus de formation des prix, l'information provenant ainsi principalement de ces futures. Ces derniers interagissent fortement avec leurs marchés physiques, directement impliqués dans l'activité économique, la consommation, etc. La financiarisation a créé un potentiel canal de propagation facile de chocs financiers à l'économie réelle, entrant ainsi dans le domaine du risque systémique. Les positions prises par ces agents financiers sur ces marchés, futures et/ou physiques, les soumettent aussi au risque que des chocs de prix sur matières premières fragilisent leurs bilans et méritent donc une attention accrue.

L'objectif de cette thèse est d'étudier l'intégration d'un grand nombre de matières premières et marchés financiers, prenant en compte de nombreux liens, économiques ou financiers. Elle tente de contribuer à la compréhension

des marchés, notamment de commodités et de fournir des outils appropriés à une analyse de l'intégration et de la propagation en grande dimension.

Dans le chapitre 1, nous cherchons les conséquences de la financiarisation dans les distributions de rentabilité des prix de matières premières. La contrainte de non-négativité des stocks pour les actifs physiques devrait se traduire en termes de dynamique de prix. Par exemple, la théorie du stockage prédit que la difficulté à stocker un actif peut rendre ses prix plus volatils. Nous avons donc étudié les moments des distributions et plus particulièrement l'occurrence de chocs sur les prix des matières premières et des actifs financiers, en vue d'en étudier ensuite leur potentielle propagation ou leur simultanéité. Conformément à la théorie du stockage, nous trouvons une sectorisation (en termes de volatilité des rendements) des marchés considérés, reflétant la difficulté ou non à stocker. De plus, en comparant les distributions, nous trouvons que les commodités se ressemblent ainsi que certains marchés financiers, en fonction de la mesure choisie. Il n'y a cependant pas de ressemblance avec le S&P500. En s'intéressant aux extrêmes, nous observons que les prix de plusieurs matières premières connaissent leur variation minimale le jour suivant la variation minimale de la valeur du S&P500 et d'autres leur variation maximale après le S&P500.

Des stocks faibles, en plus de cette difficulté à stocker, sont aussi susceptibles de provoquer des chocs de prix à la hausse, mais pas à la baisse. La littérature identifie dans la base – prix futures moins prix spot – des informations sur les stocks. Nous avons donc inclus cette information en conditionnant les rentabilités au signe de la base et trouvons que les rendements sont en effet plus volatils lorsque le marché est en déport (stocks faibles).

Enfin, une étude préliminaire des corrélations et des facteurs communs (à l'aide d'une ACP) confirme les résultats de la littérature sur le sujet. Nous retrouvons une intégration des matières premières à partir de 2004 puis avec les marchés financiers à partir de 2008.

Le chapitre 2 mène une étude plus approfondie de cette intégration en

ajoutant à l'analyse une plus grande granularité de la dimension maturité pour chaque marché, donnant alors 208 variables. Les liens d'intégration (les corrélations) étant très nombreux (21 528), ils sont difficiles à analyser, mais indispensables pour l'étude du risque systémique. Nous avons eu recours à la théorie des graphes afin de filtrer l'information et ne garder que les liens représentant l'intégration la plus forte (207), donc les canaux de transmission les plus efficaces. Le graphe obtenu permet ensuite de calculer différentes mesures, telles que la longueur, représentant l'intégration, le ratio de survie, indiquant la stabilité du graphe dans le temps ou une mesure de la structure du graphe. Il est aussi possible d'identifier les contrats les plus importants à l'aide de mesures de centralité.

Dans une première approche, les variations de ces corrélations les plus fortes donnent des informations sur la potentielle propagation de chocs. Nous appliquons cette méthodologie à l'étude de deux événements financiers majeurs : le début de la crise des subprimes et le défaut de Lehman Brothers. Nous trouvons que le premier n'a pas eu d'influence sur les matières premières à très court terme, mais a déclenché une phase d'intégration rapide. En revanche, le second a provoqué des modifications temporaires de la structure du graphe et une désintégration entre secteurs. Il n'y a donc pas eu de propagation vers les matières premières, malgré l'évolution de leurs prix autour de ces événements.

Le chapitre 3 intègre de manière plus explicite la composante propagation du risque systémique à l'aide d'un modèle de type VAR. La dimension maturité a été gardée ici, mais dans une moindre mesure, par souci d'égalité en termes de nombre de contrats par marché. Grâce à un algorithme développé par BARIGOZZI et C. BROWNLEES (2017), il est possible d'estimer un tel modèle en grande dimension et en une seule étape, permettant ainsi aux données de dévoiler leurs relations, de pondérer les deux composantes du risque systémique en conséquence. La paramétrisation par la matrice de concentration au lieu de la matrice de corrélation permet en effet d'avoir une matrice de paramètres creuse, propice à la pénalisation par LASSO. De plus, nous pouvons obtenir la matrice de corrélations partielles à partir de

cette matrice de concentration. Celles-ci représentent les "vraies" corrélations entre paires de variables, après avoir retiré l'influence des autres. Elles fournissent donc une information plus claire sur l'intégration des marchés. Cette nouvelle matrice peut aussi être représentée par un graphe et diverses mesures peuvent encore en être déduites.

Nous trouvons ainsi que dans le début des années 2000, les marchés sont désintégrés, segmentés. Nous observons aussi une intégration croissante en tendance à partir de 2003, mais celle-ci disparaît avant le début de la crise des subprimes avant d'exploser ensuite. Le défaut de Lehman Brothers a créé de nombreux liens d'intégration, plus nombreux encore suite aux variations minimale et maximale du S&P500. En revanche, la composante propagation est peu représentée dans les données, sauf dans l'année autour du Flash Crash du 6 mai 2010.

Ces outils permettent donc de mettre en évidence les véritables liens de propagation et d'intégration, en apprenant des données les composantes du risque systémique et en calculant des mesures appropriées.

Ces travaux ouvrent plusieurs pistes de recherche possibles. Il est important de pouvoir anticiper les potentiels chocs et crises (point de vue prospectif). Les *stress tests* sont en effet un outil indispensable aux régulateurs afin d'estimer les conséquences de certains scénarii potentiellement dangereux, comme des chocs macroéconomiques (potentiellement systématiques). Sur la base des outils présentés dans cette thèse, un premier développement possible serait de faire des simulations de l'évolution du système et des analyses de scénarii, tels qu'un choc sur un marché central ou périphérique, un choc sur tel secteur, etc.

La multitude de liens économiques ou financiers possibles demande d'inclure le plus d'actifs possibles dans l'étude du risque systémique. On parle notamment de financiarisation de nombreux autres "actifs", comme l'immobilier, la nourriture ou la nature. Il apparaît alors nécessaire d'étendre l'univers de marchés et variables à inclure dans le système. Dans la continuité de ces travaux, nous envisageons par exemple d'ajouter d'autres taux de change, d'autres matières premières et de s'intéresser aux marchés d'autres pays.

Il faut toutefois garder à l'esprit les critères dévaluation soulevés par les régulateurs : la taille, la non-substituabilité et l'interconnexion. Tous les marchés ne sont donc pas indispensables. De plus, l'historique des données doit être suffisant et celles-ci doivent être fiables.

Une autre extension possible serait dans la fréquence d'observation. Etant donnée la vitesse à laquelle l'information circule aujourd'hui, le moindre choc (ou même une erreur ou fausse information) peut se propager presque instantanément. Afin de mieux appréhender le risque systémique et éviter par exemple la propagation de *flash crashes*, il devrait être capital de travailler avec des données haute fréquence, voire en temps réel. Analyser ces marchés à cette fréquence constitue donc un prolongement naturel de ces travaux, mais pose le problème de la vitesse de calcul et de la mémoire utilisée. Il faudrait donc adapter ces méthodes à ce type d'analyse.

Enfin, pour boucler l'étude des marchés financiers, il est nécessaire d'introduire les interactions entre les agents et les marchés. Les "impacts de marché" générés par un ou plusieurs agents peuvent en effet déclencher des cercles vicieux, par exemple de vente, ou de ruées, qui peuvent rapidement consommer la liquidité d'un marché et en perturber le fonctionnement. De plus, les chocs de prix sur un marché peuvent déclencher des modifications d'allocations dans leurs portefeuilles, affectant alors les autres actifs les constituant. Les stratégies des agents financiers étant souvent les mêmes, ils pourraient alors propager et amplifier ces chocs. Pour évaluer correctement le risque systémique, il est donc essentiel de comprendre les interactions entre les niveaux microéconomiques (les agents) et "macroéconomiques" (les marchés).

Nous envisageons de poursuivre ces travaux et de développer ces extensions ultérieurement, en espérant apporter encore des connaissances sur les marchés et des outils de régulation, afin d'éviter les crises systémiques.

Bibliographie

- ACHARYA, Viral V. et al. (2017). ‘Measuring Systemic Risk’. In : *The Review of Financial Studies* 30(1), p. 2–47.
- ADAMIC, Lada et al. (2009). ‘On the Informational Properties of Trading Networks’. In : *Working Paper*.
- ADRIAN, Tobias et Markus K. BRUNNERMEIER (2016). ‘CoVaR’. In : *American Economic Review* 106(7), p. 1705–41.
- AÏT-SAHALIA, Yacine, Julio CACHO-DIAZ et Roger JA LAEVEN (2015). ‘Modeling financial contagion using mutually exciting jump processes’. In : *Journal of Financial Economics*.
- ALLEN, Franklin et Ana BABUS (2009). ‘The Network Challenge : Strategy, Profit, and Risk in an Interlinked World’. In : sous la dir. de Paul R. KLEINDORFER et Yoram WIND. Upper Saddle River, N.J : Pearson Education. Chap. 21, p. 367–382.
- BABUS, Ana et Péter KONDOR (2013). ‘Trading and information diffusion in over-the-counter markets’. In : *Working paper*.
- BANAVAR, Jayanth R., Amos MARITAN et Andrea RINALDO (1999). ‘Size and form in efficient transportation networks’. In : *Nature* 399(6732), p. 130–132.
- BARIGOZZI, Matteo et Christian BROWNLEES (2017). ‘Nets : Network Estimation for Time Series’. In : *Working Paper*.
- BCBS (2011). *Global systemically important banks : assessment methodology and the additional loss absorbency requirement*. Rapp. tech. Basel Committee on Banking Supervision, Bank for International Settlements.
- BCBS (2013). *Global systemically important banks : updated assessment methodology and the higher loss absorbency requirement*. Rapp. tech.

- Basel Committee on Banking Supervision, Bank for International Settlements.
- BEKAERT, Geert, Campbell R HARVEY et Angela NG (2005). 'Market Integration and Contagion'. In : *Journal of Business* 78(1).
- BENOÎT, Sylvain, Jean-Edouard COLLIARD et al. (2017). 'Where the Risks Lie : A Survey on Systemic Risk'. In : *Review of Finance* 21(1), p. 109–152.
- BENOÎT, Sylvain, Christophe HURLIN et Christophe PÉRIGNON (2015). 'Implied Risk Exposures'. In : *Review of Finance* 19(6), p. 2183–2222.
- BENZI, Michele et Christine KLYMKO (2013). 'Total communicability as a centrality measure'. In : *Journal of Complex Networks* 1(2), p. 124–149.
- BESSEMBINDER, Hendrik et al. (1996). 'Is there a term structure of futures volatilities? Reevaluating the Samuelson hypothesis'. In : *Journal of Derivatives* 4(2), p. 45–58.
- BILLIO, Monica et Massimiliano CAPORIN (2010). 'Market linkages, variance spillovers, and correlation stability : Empirical evidence of financial contagion'. In : *Computational Statistics and Data Analysis* 54(11), p. 2443–2458.
- BILLIO, Monica, Mila GETMANSKY et al. (2012). 'Econometric measures of connectedness and systemic risk in the finance and insurance sectors'. In : *Journal of Financial Economics* 104(3), p. 535–559.
- BIS (1992). *Recent Developments in International Interbank Relations*. Rapp. tech. Bank for International Settlements.
- BIS (2001). *Report on Consolidation in the Financial Sector*. Rapp. tech. Bank for International Settlements.
- BIS, FSB et IMF (2009). *Guidance to Assess the Systemic Importance of Financial Institutions, Markets and Instruments : Initial Considerations*. Rapp. tech. Bank for International Settlements, Financial Stability Board et International Monetary Fund.
- BLACK, Fischer (1976). 'The pricing of commodity contracts'. In : *Journal of Financial Economics* 3(1), p. 167–179.
- BLOCH, Francis et Nicolas QUÉROU (2013). 'Pricing in social networks'. In : *Games and Economic Behavior*.

- BONACICH, Phillip (1987). 'Power and Centrality : A Family of Measures'.
In : *American Journal of Sociology* 92(5), p. 1170–1182.
- BONANNO, G. et al. (2004). 'Networks of equities in financial markets'. In :
Eur. Phys. J. B 38(2), p. 363–371.
- BRENNAN, Michael J (1958). 'The supply of storage'. In : *The American
Economic Review* 48(1), p. 50–72.
- BROWNLEES, Christian T et Robert F ENGLE (2012). 'Volatility, correlation
and tails for systemic risk measurement'. In : *Working paper*.
- BRUNNERMEIER, Markus K. (2009). 'Deciphering the Liquidity and Credit
Crunch 2007–2008'. In : *Journal of Economic Perspectives* 23(1), p. 77–
100.
- BÜYÜKŞAHİN, Bahattin, Michael S. HAIGH et Michel A. ROBE (2010).
'Commodities and equities : ever a "market of one"?' In : *Journal of
Alternative Investment*, p. 76–95.
- BÜYÜKŞAHİN, Bahattin et Michel A. ROBE (2011). 'Does Paper Oil Mat-
ter? Energy Markets' Financialization and Equity-Commodity Co-Movements'.
In : *Working Paper, American University, (Revised) July*.
- BÜYÜKŞAHİN, Bahattin et Michel A. ROBE (2014). 'Speculators, commo-
dities and cross-market linkages'. In : *Journal of International Money
and Finance* 42, p. 38–70.
- BÜYÜKŞAHİN, Bahattin, Michel A. ROBE et Valentina G. BRUNO (2014).
'The Financialization of Food?' In : *Working Paper*.
- CAPORALE, Guglielmo Maria, Andrea CIPOLLINI et Nicola SPAGNOLO
(2005). 'Testing for contagion : a conditional correlation analysis'. In :
Journal of Empirical Finance 12(3), p. 476–489.
- CAPPONI, Agostino et Martin LARSSON (2015). 'Price Contagion through
Balance Sheet Linkages'. In : *The Review of Asset Pricing Studies* 5(2),
p. 227–253.
- CASTIGLIONESI, Fabio, Fabio FERIOZZI et Guido LORENZONI (2009). 'Fi-
nancial integration, liquidity, and the depth of systemic crisis'. In : *Un-
published working paper. University of Tilburg*.
- CHAKRABORTI, Anirban et al. (2003). 'Dynamic asset trees and Black Mon-
day'. In : *Physica A* 324(1-2), p. 247–252.

- CHEN, Shiu-Sheng et Hung-Chyn CHEN (2007). ‘Oil prices and real exchange rates’. In : *Energy Economics* 29(3), p. 390–404.
- CHENG, Ing-Haw et Wei XIONG (2014). ‘Financialization of commodity markets’. In : *Annu. Rev. Financ. Econ.* 6(1), p. 419–441.
- CHINAZZI, Matteo et Giorgio FAGIOLO (2013). ‘Systemic risk, contagion, and financial networks : A survey’. In : *LEM Working Paper Series*.
- COHEN-COLE, Ethan et al. (2012). ‘Strategic Interactions on Financial Networks for the Analysis of Systemic Risk’. In : *Handbook on Systemic Risk*, p. 306.
- CORSETTI, Giancarlo, Marcello PERICOLI et Massimo SBRACIA (2005). ‘Some contagion, some interdependence’ : More pitfalls in tests of financial contagion’. In : *Journal of International Money and Finance* 24(8), p. 1177–1199.
- CPSS et IOSCO (2012). *Principles for financial market infrastructures*. Rapp. tech. Committee on Payment, Settlement Systems, Bank for International Settlements et International Organization of Securities Commissions.
- DEATON, Angus et Guy LAROQUE (1992). ‘On the behaviour of commodity prices’. In : *The Review of Economic Studies* 59(1), p. 1–23.
- DIEBOLD, Francis X. et Kamil YILMAZ (2014). ‘On the network topology of variance decompositions : Measuring the connectedness of financial firms’. In : *Journal of Econometrics* 182(1), p. 119–134.
- DIEBOLD, Francis X., Kamil YILMAZ et Laura LIU (2017). ‘Commodity Connectedness’. In : *NBER, Working Paper No. 23685*.
- DUMONTAUX, Nicolas et Adrian POP (2013). ‘Contagion Effects in the Aftermath of Lehman’s Collapse : Evidence from the US Financial Services Industry’. In :
- ECB (2012). *Financial Stability Review*. Rapp. tech. European Central Bank.
- EICHENGREEN, Barry, Andrew K. ROSE et Charles WYPLOSZ (1996). ‘Contagious currency crises : first tests’. In : *Scandinavian Journal of Economics* 98, p. 463–484.

- ERB, Claude B. et Campbell R. HARVEY (2006). 'The Strategic and Tactical Value of Commodity Futures'. In : *Financial Analysts Journal* 62(2), p. 69–97.
- EWING, Bradley T, Farooq MALIK et Ozkan OZFIDAN (2002). 'Volatility transmission in the oil and natural gas markets'. In : *Energy Economics* 24(6), p. 525–538.
- FAMA, Eugene F. et Kenneth R. FRENCH (1987). 'Commodity Futures Prices : Some Evidence on Forecast Power, Premiums, and the Theory of Storage'. In : *The Journal of Business* 60(1), p. 55–73.
- FAMA, Eugene F. et Kenneth R. FRENCH (1988). 'Business Cycles and the Behavior of Metals Prices'. In : *The Journal of Finance* 43(5), p. 1075–1093.
- FATTOUH, Bassam, Lutz KILLIAN et Lavan MAHADEVA (2013). 'The Role of Speculation in Oil Markets : What Have We Learned So Far?' In : *The Energy Journal* 34(3).
- FERNANDO, Chitru S, Anthony D MAY et William L MEGGINSON (2012). 'The value of investment banking relationships : evidence from the collapse of Lehman Brothers'. In : *The Journal of Finance* 67(1), p. 235–270.
- FERRARO, Domenico, Kenneth ROGOFF et Barbara ROSSI (2015). 'Can oil prices forecast exchange rates? An empirical analysis of the relationship between commodity prices and exchange rates'. In : *Journal of International Money and Finance* 54, p. 116–141.
- FIGUEROLA-FERRETTI, Isabel et Jesús GONZALO (2010). 'Modelling and measuring price discovery in commodity markets'. In : *Journal of Econometrics* 158(1), p. 95–107.
- FORBES, Kristin J. et Roberto RIGOBON (2002). 'No Contagion, Only Interdependence : Measuring Stock Market Comovements'. In : *The Journal of Finance* 57(5), p. 2223–2261.
- FRÉCHET, Maurice (1906). 'Sur quelques points du calcul fonctionnel'. In : *Rendiconti del Circolo Matematico di Palermo* 22, p. 1–74.
- FUND, International Monetary (2011). 'Global Financial Stability Report'. In : *World Economic and Financial Surveys*.

- GAI, Prasanna et Sujit KAPADIA (2010). 'Contagion in financial networks'. In : *Proceedings of the Royal Society of London A : Mathematical, Physical and Engineering Sciences*. The Royal Society.
- GARBADE, Kenneth D. et William L. SILBER (1983). 'Price Movements and Price Discovery in Futures and Cash Markets'. In : *The Review of Economics and Statistics* 65(2), p. 289–297.
- GIESECKE, Kay et Baeho KIM (2011). 'Systemic Risk : What Defaults Are Telling Us'. In : *Management Science* 57(8), p. 1387–1405.
- GONZÀLEZ-HERMOSILLO, Brenda et Heiko HESSE (2011). 'Global market conditions and systemic risk'. In : *Journal of Emerging Market Finance* 10(2), p. 227–252.
- GORTON, Gary B, Fumio HAYASHI et K Geert ROUWENHORST (2013). 'The fundamentals of commodity futures returns'. In : *Review of Finance* 17(1), p. 35–105.
- GORTON, Gary B et K Geert ROUWENHORST (2006). 'Facts and Fantasies about Commodity Futures'. In : *Financial Analysts Journal* 62(2), p. 47–68.
- GOWER, John C. (1966). 'Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis'. In : *Biometrika* 53(3/4), p. 325–338.
- GRANGER, C. W. J. (1969). 'Investigating Causal Relations by Econometric Models and Cross-spectral Methods'. In : *Econometrica* 37(3), p. 424–438.
- HALDANE, Andrew G et Robert M MAY (2011). 'Systemic risk in banking ecosystems'. In : *Nature* 469(7330), p. 351–355.
- HAMILTON, James D. (2009). 'Causes and Consequences of the Oil Shock of 2007-08.' In : *Brookings Papers on Economic Activity*, p. 215–261.
- HAMMOUDEH, Shawkat et Yuan YUAN (2008). 'Metal volatility in presence of oil and interest rate shocks'. In : *Energy Economics* 30(2), p. 606–620.
- HARTMANN, Philipp, Stefan STRAETMANS et Casper G DE VRIES (2004). 'Asset market linkages in crisis periods'. In : *Review of Economics and Statistics* 86(1), p. 313–326.

- HASBROUCK, Joel (2003). 'Intraday Price Formation in U.S. Equity Index Markets'. In : *The Journal of Finance* 58(6), p. 2375–2400.
- HAUTSCH, Nikolaus, Julia SCHAUMBURG et Melanie SCHIENLE (2015). 'Financial Network Systemic Risk Contributions'. In : *Review of Finance* 19(2), p. 685–738.
- HEINKEL, Robert, Maureen E. HOWE et John S. HUGHES (1990). 'Commodity convenience yields as an option profit'. In : *Journal of Futures Markets* 10(5), p. 519–533.
- HICKS, J. R. (1946). *Value and Capital : An Inquiry into some Fundamental Principles of Economic Theory*. Oxford University Press.
- ILLING, Mark et Ying LIU (2006). 'Measuring financial stress in a developed country : An application to Canada'. In : *Journal of Financial Stability* 2(3), p. 243–265.
- IMF (2008). *Global Financial Stability Report*. Rapp. tech. International Monetary Fund.
- IMF (2010). *Global Financial Stability Report*. Rapp. tech. IMF.
- IMF (2011a). *Global Financial Stability Report*. Rapp. tech. International Monetary Fund.
- IMF (2011b). *Global Financial Stability Report*. Rapp. tech. International Monetary Fund.
- IRWIN, Scott H et Dwight R SANDERS (2011). 'Index funds, financialization, and commodity futures markets'. In : *Applied Economic Perspectives and Policy* 33(1), p. 1–31.
- ISHIHARA, Yoichiro (2005). 'Quantitative analysis of crisis : Crisis identification and causality'. In : *World Bank Policy Research Working Paper* (3598).
- JOËTS, Marc, Valérie MIGNON et Tovonony RAZAFINDRABE (2017). 'Does the volatility of commodity prices reflect macroeconomic uncertainty?' In : *Energy Economics*.
- JORION, Philippe et Gaiyan ZHANG (2007). 'Good and bad credit contagion : Evidence from credit default swaps'. In : *Journal of Financial Economics* 84(3), p. 860–883.

- JORION, Philippe et Gaiyan ZHANG (2009). 'Credit Contagion from Counterparty Risk'. In : *The Journal of Finance* 64(5), p. 2053–2087.
- KALDOR, Nicholas (1939). 'Speculation and Economic Stability'. In : *The Review of Economic Studies* 7(1), p. 1–27.
- KATZ, Leo (1953). 'A New Status Index Derived From Sociometric Analysis'. In : *Psychometrika* 18(1), p. 39–43.
- KAUFMAN, George G. (1995). *Research in Financial Services*. T. 7. London : JAI Press.
- KEYNES, John Maynard (1930). *A Treatise on Money : In 2 Volumes*. Macmillan & Company.
- KILIAN, Lutz (2009). 'Causes and Consequences of the Oil Shock of 2007–08 : Comment'. In : *Brookings Papers on Economic Activity*, p. 267–278.
- KORNIOTIS, George M. (2009). 'Does Speculation Affect Spot Prices Levels? The Case of Metals with and without Futures Markets'. In : *Finance and Economics Discussion Series Working Paper 2009-29, Board of Governors of the Federal Reserve*.
- More on oil and speculation* (2008) May 13.
- LAUTIER, Delphine, Julien LING et Franck RAYNAUD (2015). 'Integration of Commodity Derivative Markets : Has It Gone Too Far?' In : *Commodities, Energy and Environmental Finance*. Sous la dir. de René AÏD, Michael LUDKOVSKI et Ronnie SIRCAR. T. 74. Fields Institute Communications. Springer New York, p. 65–90.
- LAUTIER, Delphine et Franck RAYNAUD (2012). 'Systemic Risk in Energy Derivative Markets : A Graph-Theory Analysis'. In : *The Energy Journal* 33(3), p. 215–239.
- LAUTIER, Delphine et Franck RAYNAUD (2014). 'Information flows in the term structure of commodity prices'. In : *Working Paper*.
- LE PEN, Yannick et Benoît SÉVI (2018). 'Futures Trading and the Excess Co-movement of Commodity Prices'. In : *Review of Finance* 22(1), p. 381–418.

- LONGIN, Francois et Bruno SOLNIK (2001). 'Extreme correlation of international equity markets'. In : *The Journal of Finance* 56(2), p. 649–676.
- LONGSTAFF, Francis A (2010). 'The subprime credit crisis and contagion in financial markets'. In : *Journal of Financial Economics* 97(3), p. 436–450.
- LUTZ, Björn (2010). *Pricing of Derivatives on Mean-Reverting Assets*. T. 630. Lecture Notes in Economics and Mathematical Systems. Springer-Verlag Berlin Heidelberg.
- MANTEGNA, R. N. (1999). 'Hierarchical structure in financial markets'. In : *The European Physical Journal B* 11, p. 193–197.
- MANTEGNA, R.N. et H.E. STANLEY (1999). *Introduction to Econophysics : Correlations and Complexity in Finance*. Cambridge University Press. ISBN : 9781139431224. URL : <https://books.google.fr/books?id=SzgXWCS7Nr8C>.
- MARKOSE, Sheri M. (2012). 'Systemic Risk from Global Financial Derivatives : A Network Analysis of Contagion and Its Mitigation with Super-Spreader Tax'. In : *IMF Working paper WP12282*.
- MARKOWITZ, Harry (1952). 'Portfolio Selection'. In : *The Journal of Finance* 7(1), p. 77–91.
- MARTINEZ-JARAMILLO, Serafin et al. (2014). 'An empirical study of the Mexican banking system's network and its implications for systemic risk'. In : *Journal of Economic Dynamics and Control* 40, p. 242–265.
- MARTÍNEZ-JARAMILLO, Serafín et al. (2010). 'Systemic risk, financial contagion and financial fragility'. In : *Journal of Economic Dynamics and Control* 34(11), p. 2358–2374.
- MASTERS, Michael W. (2008). *before the Committee on Homeland Security and Governmental Affairs*. Speech.
- MILLER, J. Isaac et Ronald A. RATTI (2009). 'Crude oil and stock markets : Stability, instability, and bubbles'. In : *Energy Economics* 31(4), p. 559–568.

- MOSHIRIAN, Fariborz et Qiongbing WU (2009). 'Banking industry volatility and banking crises'. In : *Journal of International Financial Markets, Institutions and Money* 19(2), p. 351–370.
- NASH, Daniel J. (2001). 'Long-term Investing in Commodities'. In : *Global Pensions Quarterly, Morgan Stanley Global Pensions* 4(1), p. 25–31.
- NEWBERY, David M. (1987). 'When do Futures Destabilize Spot Prices?' In : *International Economic Review* 28 (2), p. 291–297.
- NG, Victor K. et Stephen Craig PIRRONG (1994). 'Fundamentals and Volatility : Storage, Spreads, and the Dynamics of Metals Prices'. In : *The Journal of Business* 67(2), p. 203–230.
- NG, Victor K. et Stephen Craig PIRRONG (1996). 'Price dynamics in refined petroleum spot and futures markets'. In : *Journal of Empirical Finance* 2(4), p. 359–388.
- NIER, Erlend et al. (2007). 'Network models and financial stability'. In : *Journal of Economic Dynamics and Control* 31(6), p. 2033–2060.
- OET, Mikhail V et al. (2013). 'SAFE : An early warning system for systemic banking risk'. In : *Journal of Banking & Finance* 37(11), p. 4510–4533.
- ONNELA, J-P, Anirban CHAKRABORTI et al. (2003). 'Dynamics of market correlations : Taxonomy and portfolio analysis'. In : *Physical Review E* 68(5), p. 056110.
- ONNELA, J-P, A CHAKRABORTI et al. (2002). 'Dynamic asset trees and portfolio analysis'. In : *The European Physical Journal B-Condensed Matter and Complex Systems* 30(3), p. 285–288.
- PARK, Jungwook et Ronald A. RATTI (2008). 'Oil price shocks and stock markets in the U.S. and 13 European countries'. In : *Energy Economics* 30(5), p. 2587–2608.
- PESARAN, M Hashem et Andreas PICK (2007). 'Econometric issues in the analysis of contagion'. In : *Journal of Economic Dynamics and Control* 31(4), p. 1245–1277.
- PINDYCK, Robert et Julio ROTEMBERG (1990). 'The excess co-movement of commodity prices'. In : *Economic Journal* 100.
- QUAN, Jing (1992). 'Twostep testing procedure for price discovery role of futures prices'. In : *Journal of Futures Markets* 12(2), p. 139–149.

- SARI, Ramazan, Shawkat HAMMOUDEH et Ugur SOYTAS (2010). 'Dynamics of oil price, precious metal prices, and exchange rate'. In : *Energy Economics* 32(2), p. 351–362.
- SCHWARCZ, Steven L. (2008). 'Systemic Risk'. In : *Georgetown Law Journal* 97, p. 193.
- SCHWARTZ, Eduardo S. (1997). 'The Stochastic Behavior of Commodity Prices : Implications for Valuation and Hedging'. In : *The Journal of Finance* 52(3), p. 923–973.
- SCHWARZ, Thomas V. et Andrew C. SZAKMARY (1994). 'Price discovery in petroleum markets : Arbitrage, cointegration, and the time interval of analysis'. In : *Journal of Futures Markets* 14(2), p. 147–167.
- SILVENNOINEN, Annastiina et Susan THORP (2013). 'Financialization, crisis and commodity correlation dynamics'. In : *Journal of International Financial Markets, Institutions and Money* 24, p. 42–65. ISSN : 1042-4431.
- SILVENNOINEN, Annastiina et Susan THORP (2016). 'Crude Oil and Agricultural Futures : An Analysis of Correlation Dynamics'. In : *Journal of Futures Markets* 36(6), p. 522–544.
- STOLL, Hans R. et Robert E. WHALEY (1990). 'The Dynamics of Stock Index and Stock Index Futures Returns'. In : *The Journal of Financial and Quantitative Analysis* 25(4), p. 441–468.
- TANG, Ke et Wei XIONG (2012). 'Index Investing and the Financialization of Commodities'. In : *Financial Analysts Journal* 68(6), p. 54–74.
- TELSER, Lester G (1958). 'Futures trading and the storage of cotton and wheat'. In : *Journal of Political Economy* 66(3), p. 233–255.
- TILL, Hilary (2006). 'A Review of the Historical Literature on Commodity Futures Contract Returns'. In : *Working paper*.
- WEST, Douglas B. (2001). *Introduction to graph theory*. Prentice Hall.
- WORKING, Holbrook (1949). 'The Theory of Price of Storage'. In : *The American Economic Review* 39(6), p. 1254–1262.
- ZHANG, Yue-Jun et al. (2008). 'Spillover effect of US dollar exchange rate on oil prices'. In : *Journal of Policy Modeling* 30(6), p. 973–991.

Résumé

Le phénomène de financiarisation des matières premières a créé un lien très fort entre ces marchés de commodités et les marchés financiers. L'évolution similaire de leurs prix a généré des craintes d'une influence de ces derniers sur les premiers, relevant alors du risque systémique. Nous mettons en avant deux de ses aspects : un choc simultané sur tous les marchés (choc systématique), lié à une forte intégration des marchés ; et la propagation d'un choc d'un marché à d'autres. C'est dans ce contexte que cette thèse se place.

Le premier chapitre vise à évaluer dans quelle mesure les aspects physiques des commodités se reflètent (encore) en termes de variations de prix et statistiques. Il met en avant une certaine sectorisation encore présente, mais s'effaçant au profit d'une intégration des commodités puis avec les marchés financiers.

Le deuxième chapitre étudie plus précisément cette intégration en grande dimension. Nous appliquons des outils de la théorie des graphes pour pouvoir analyser un large système et nous concentrons sur deux chocs financiers : le début de la crise des subprimes et le défaut de Lehman Brothers. Nous trouvons que seul le second a eu des conséquences à très court terme sur le système, mais n'a pas négativement affecté les commodités.

Le troisième chapitre distingue plus précisément les deux aspects du risque systémique en grande dimension. Nous appliquons pour cela des méthodes d'apprentissage automatique et de théorie des graphes. S'intéressant à un plus grand nombre d'événements financiers, nous trouvons que l'intégration a été prédominante, bien que le premier chapitre laissait penser à une propagation de chocs.

Mots Clés

Matières premières, financiarisation, risque systémique, grande dimension

Abstract

Due to their financialization, commodity markets have been more and more connected to financial markets. Since the former are essential in the real economy, when studying systemic risk, it is necessary to include both sectors in the analysis. One also needs to distinguish these two aspects of this risk: systematic risk (risk that all markets are affected by a shock), which increases with integration, and propagation risk. This is where this thesis tries to contribute.

The first chapter assesses how integrated commodity and financial markets are. The physical aspects of the former should translate into specific price variations distributions. We do find that there is a sectorization, but it fades away, showing the integration of commodity markets, then their integration with financial markets.

The second chapter investigates this integration in a large system of futures contracts. This high dimensional setting calls for the use of graph theory tools to keep the important information. We focus on two financial shocks: the start of the subprime crisis and the default of Lehman Brothers. Only the latter had significant impact over the short run, but it did not adversely affect commodity markets.

The third chapter differentiates the two aspects of systemic risk: propagation and (more) integration, in a high-dimension framework. We thus need machine learning and graph theory tools. We study more financial shocks and find that integration is prominent, even though the first chapter hinted for propagation for some shocks.

Keywords

Commodity markets, financialization, systemic risk, high dimension