ECOLE CENTRALE
DES ARTS ET MANUFACTURES

T H E S E

présentée par

Rémy CHICHEPORTICHE

pour l'obtention du

GRADE DE DOCTEUR

Specialité : Mathématiques appliquées
Laboratoire : Mathématiques Appliquées aux Systèmes (MAS) — EA 4037
Sujet : DÉPENDANCES NON LINÉAIRES EN FINANCE

soutenue le **27 juin 2013** devant un jury composé de:

Directeurs :	Frédéric ABERGEL	-	Ecole Centrale Paris
	Anirban CHAKRABORTI	-	Ecole Centrale Paris
Rapporteurs :	Yannick MALEVERGNE	-	Ecole de Management de Lyon
	Grégory SCHEHR	-	Université de Paris-Sud
Examinateurs :	Jean-Philippe BOUCHAUD	-	Capital Fund Management
	Jean-David FERMANIAN	-	CREST

CENTRALE

P A R I S

ECOLE CENTRALE
DES ARTS ET MANUFACTURES

PARIS

P H D   T H E S I S

to obtain the title of

PhD of Science
Specialty : Applied Mathematics

Non-linear Dependences in Finance

Defended by

Rémy Chicheportiche

on June 27 2013

Advisors :	Frédéric Abergel	-	Ecole Centrale Paris
	Anirban Chakraborti	-	Ecole Centrale Paris
Reviewers :	Yannick Malevergne	-	Ecole de Management de Lyon
	Grégory Schehr	-	Université de Paris-Sud
Examinators :	Jean-Philippe Bouchaud	-	Capital Fund Management
	Jean-David Fermanian	-	CREST

Dépendances non-linéaires en finance

Résumé: La thèse est composée de trois parties. La partie I introduit les outils mathématiques et statistiques appropriés pour l'étude des dépendances, ainsi que des tests statistiques d'adéquation pour des distributions de probabilité empiriques. Je propose deux extensions des tests usuels lorsque de la dépendance est présente dans les données, et lorsque la distribution des observations a des queues larges. Le contenu financier de la thèse commence à la partie II. J'y présente mes travaux concernant les dépendances transversales entre les séries chronologiques de rendements journaliers d'actions, c'est à dire les forces instantanées qui relient plusieurs actions entre elles et les fait se comporter collectivement plutôt qu'individuellement. Une calibration d'un nouveau modèle à facteurs est présentée ici, avec une comparaison à des mesures sur des données réelles. Finalement, la partie III étudie les dépendances temporelles dans des séries chronologiques individuelles, en utilisant les mêmes outils et mesures de corrélations. Nous proposons ici deux contributions à l'étude du "volatility clustering", de son origine et de sa description: l'une est une généralisation du mécanisme de rétro-action ARCH dans lequel les rendements sont auto-excitants, et l'autre est une description plus originale des auto-dépendances en termes de copule. Cette dernière peut être formulée sans modèle et n'est pas spécifique aux données financières. En fait, je montre ici aussi comment les concepts de récurrences, records, répliques et temps d'attente, qui caractérisent la dynamique dans les séries chronologiques, peuvent être écrits dans la cadre unifié des copules.

Mots-clés: dépendances statistiques, copules, tests d'adéquation, processus stochastiques, séries temporelles, models financiers, volatility clustering

Non-linear dependences in finance

Abstract: The thesis is composed of three parts. Part I introduces the mathematical and statistical tools that are relevant for the study of dependences, as well as statistical tests of Goodness-of-fit for empirical probability distributions. I propose two extensions of usual tests when dependence is present in the sample data and when observations have a fat-tailed distribution. The financial content of the thesis starts in Part II. I present there my studies regarding the "cross-sectional" dependences among the time series of daily stock returns, i.e. the instantaneous forces that link several stocks together and make them behave somewhat collectively rather than purely independently. A calibration of a new factor model is presented here, together with a comparison to measurements on real data. Finally, Part III investigates the temporal dependences of single time series, using the same tools and measures of correlation. I propose two contributions to the study of the origin and description of "volatility clustering": one is a generalization of the ARCH-like feedback construction where the returns are self-exciting, and the other one is a more original description of self-dependences in terms of copulas. The latter can be formulated model-free and is not specific to financial time series. In fact, I also show here how concepts like recurrences, records, aftershocks and waiting times, that characterize the dynamics in a time series can be written in the unifying framework of the copula.

Keywords: statistical dependences, copulas, goodness-of-fit tests, stochastic processes, time series, financial modeling, volatility clustering

Remerciements / Acknowledgments

Mes premiers remerciements vont évidemment à Jean-Philippe Bouchaud, qui m'a offert l'opportunité de m'engager sur la voie du doctorat à un moment où j'en avais abandonné l'idée. Je lui suis infiniment reconnaissant pour ses encouragements répétés, la profonde confiance qu'il m'a accordée et pour son invraisemblable disponibilité.

Je remercie Frédéric Abergel, directeur de la Chaire de finance quantitative de l'École Centrale, pour avoir accepté la supervision académique de mon doctorat, et mis à ma disposition un environnement intellectuel et matériel favorable. Merci à Anirban Chakraborti pour sa co-supervision et son enthousiasme à me proposer de nouveaux projets.

J'adresse un remerciement particulier à mon jury de thèse: à Yannick Maleverge et Grégory Schehr pour avoir immédiatement accepté la tâche de relecteur, et à Jean-David Fermanian qui a porté un intérêt à mon travail.

Je veux aussi exprimer ma gratitude aux dirigeants de Capital Fund Management pour s'être impliqués dans une convention CIFRE et m'avoir procuré des conditions de travail des plus confortables, ainsi qu'à tous les collaborateurs avec qui j'ai eu la chance d'interagir. Je garderai longtemps le souvenir de la "salle des stagiaires" et de ses occupants plus ou moins éphémères, à commencer par Romain Allez et Bence Tóth qui y ont séjourné avec moi le plus longtemps.

Mes passages à l'École ont été l'occasion d'échanges (scientifiques ou non) avec les membres de la chaire et du laboratoire MAS, en particulier avec Damien Challet et les autres doctorants: Alexandre Richard, Aymen Jedidi, Ban Zheng, Fabrizio Pomponio, Nicolas Huth, avec qui j'ai le plus interagi. Qu'ils soient ici cordialement salués, de même que mes anciens camarades Michael Kastoryano, Nicolas Cantale, Liliana Foletti, Stefanie Stantcheva et David Salfati.

Je n'aurais sans doute pas d'autre occasion d'exprimer ma gratitude à toutes les figures qui ont marqué mon éducation. Qu'il me soit donc permis de rendre ici un hommage tardif à Mesdames Nelly Piguet, Regula Krattenmacher et Larissa Shargorodsky, Messieurs Didier Deshusses, Christian Charvin, Frederic Hermann, Elie Prigent et Bernard Delez, Monsieur Paco Carbonell, Maîtres Raymond Hyvernaud, Georges Léger, Jean-Marc Cagnet et Yanaki Altanov.

Je dédicace cette thèse à ma grande famille, proche et lointaine, qu'il est superflu de remercier.

Paris, juin 2013

List of publications and submitted articles

Published articles:

[Chicheportiche 2011] R. Chicheportiche and J.-P. Bouchaud, Goodness-of-fit tests with dependent observations, *J. Stat. Mech.* **9**, P09003, 2011

[Chicheportiche 2012b] R. Chicheportiche and J.-P. Bouchaud, The joint distribution of stock returns is not elliptical, *Int. J. Theo. Appl. Fin.* **15**(3), 1250019, 2012

[Chicheportiche 2012c] R. Chicheportiche and J.-P. Bouchaud, Weighted Kolmogorov-Smirnov test: Accounting for the tails, *Phys. Rev. E* **4**(86), 1115, 2012

Pre-prints and submitted articles:

[Chicheportiche 2012a] R. Chicheportiche and J.-P. Bouchaud, The fine structure of volatility feedback, arXiv preprint qfin.ST/1206.2153, Aug. 2012

[Chicheportiche 2013d] R. Chicheportiche and A. Chakraborti, A model-free characterization of recurrences in stationary time series, arXiv preprint physics.data-an/1302.3704, Feb. 2013

[Chicheportiche 2013b] R. Chicheportiche and J.-P. Bouchaud, A minimal factor model for non-linear dependences in stock returns

[Chicheportiche 2013c] R. Chicheportiche and J.-P. Bouchaud, Some applications of first passage ideas in finance, to appear in *First-Passage Phenomena and Their Applications* (R. Metzler, G. Oshanin and S. Redner Eds.), World Scientific Publishers, May 2013

Work in progress, working papers and collaborations:

[Chicheportiche 2013e] R. Chicheportiche and A. Chakraborti, Non-linear and multi-points dependences in stationary time series: a copula approach

[Blanc 2013] P. Blanc, R. Chicheportiche and J.-P. Bouchaud, The fine structure of volatility II: intraday and overnight effects

[Richard 2013] A. Richard, R. Chicheportiche and J.-P. Bouchaud, The sup of the 2D pinned Brownian sheet

[Chicheportiche 2013a] R. Chicheportiche and J.-P. Bouchaud, Ironing the copula surface

[Gould 2013] M. Gould, R. Chicheportiche and J.-P. Bouchaud, Re-assessing the tail distribution of financial returns

vii

List of abbreviations

ARCH	Auto-Regressive Conditional Heteroskedastic
CDF, cdf	Cumulative distribution function
CLT	Central limit theorem
CvM, CM	Cramér-von Mises
d.o.f.	Degree(s) of freedom
FGN, fGn	Fractional Gaussian noise
FP	Fokker-Planck
FT	Fourier transform
GLS	Generalized least squares
GMM	Generalized method of moments
GoF	Goodness of fit
i.i.d.	Independent and identically distributed
KS	Kolmogorov-Smirnov
LHS	Left-hand side
MGF, mgf	Moment generating function
ML	Maximum likelihood
OU	Ornstein-Uhlenbeck
pBs	Pinned Brownian sheet
PCA	Principal components analysis
PDF, pdf	Probability distribution function
RHS	Right-hand side
RMS	Root mean square
s.d.	Standard deviation
SDE	Stochastic differential equation
TRI	Time-reversal invariance
WN	White noise

Contents

Introduction

Contents

1.1 Introduction en français

1.1.1 Les marchés financiers comme systèmes complexes

L'économie présente des traits caractéristiques des systèmes complexes: un grand nombre de variables (individus, entreprises, contrats, maturités, etc.), des agents hétérogènes (d'où une difficulté à isoler des groupes), de l'information partielle et de l'aléa, de l'auto-organisation et des phénomènes spontanés (interactions fortes, non-perturbativité), une dynamique de réseau (entres banques, pays, contreparties, entreprises, etc.) et des interactions à longue portée, un quasi-continuum d'échelles de temps, de la non-stationnarité, des trajectoires avec mémoire et de l'exposition à des forces extérieures (régulations, catastrophes, etc.).

Cette réalité contraste fortement avec les hypothèses des théories économique et financière classiques, comme l'homogénéité (l'existence d'un agent représentatif), une rationalité parfaite, une capacité de prévision à horizon infini, un monde à l'équilibre dont les perturbations sont instantanément ajustées, l'absence de mémoire (toute l'information passée est reflétée dans l'état actuel du monde), . . .

Pour traiter ces enjeux sous un angle plus empirique, des communautés de chercheurs avec des points de vue différents ont progressivement investi le champ des sciences économiques et sociales. Parmi elles, on trouve des sociologues quantitatifs, des éconophysiciens, des statisticiens (physiciens statistiques ou biostatisticiens), des mathématiciens appliqués. En effet, beaucoup des traits listés ci-dessus ont déjà été étudiés en divers champs des sciences naturelles, où des méthodes ont été mises au point pour prendre en compte l'aléa, les interactions de réseaux, les comportements collectifs, la persistance et la réversion, etc. [Anderson 1988, Arthur 1997, Blume 2006, Challet 2005, Galam 2008, Kirman 2010, Abergel 2013, Bouchaud 2013]

Parmi tous les champs de l'activité socio-économique, la finance de marché pourrait bien être le plus proche de ce que l'on peut raisonnablement étudier comme un système physique: du point de vue théorique, le nombre de degrés de liberté est quelque peu réduit par les règles des places de marché (mécanismes d'actions, carnets d'ordres, régulations, contrats standards, divulgation de l'information) alors que le nombre d'acteurs et d'actions reste assez large pour autoriser des moyennages dans des ensembles statistiques; et du point de vue empirique, la finance de marché génère des quantités astronomiques de données rendant possible la "répétition d'expériences" sous hypothèses, et le traitement des mesures avec des outils de statistique appropriés. En fait, il est établi que certaines propriétés des marchés financiers ressemblent à celles de systèmes étudiés en physique statistique: intermittence dans les écoulements turbulents, bruit de craquement de matériaux en réponse à un changement de conditions extérieures, lignes de fracture dans les solides, interactions à longue portée ou de champ moyen dans les systèmes de spins sur réseau, transitions de phases et criticalité spontanée, pour en citer quelques-unes.

1.1.2 Champ d'application de cette thèse

Cette thèse étudie les dépendances statistiques au sens général, aussi bien "transversalement" (entre variables aléatoires) que temporellement (réalisations successives d'une variable). Tous les résultats théoriques s'appliquent donc en principe à tout champ où plusieurs variables interagissent, ou où des processus stationnaires ont de la mémoire. Plus spécifiquement,

l'étude est restreinte à des processus discrets avec des intervalles de temps équidistants, ce qui signifie que les séries chronologiques correspondantes sont échantillonnées et n'ont pas besoin d'être "marquées" temporellement.

En ce qui concerne le contenu appliqué, le focus est sur les données financières, plus précisément les séries chronologiques de prix (ou rendements) d'actions à la fréquence relativement basse du jour. Il s'agit principalement des prix des actions les plus liquides négociées sur les marchés états-uniens, et dont les historiques peuvent être téléchargés librement de plusieurs sources.

Propriétés statistiques, dépendances et dynamique

Les caractéristiques statistiques d'un ensemble de variables aléatoires sont de plusieurs ordres: propriétés distributionnelles, corrélations, et dynamique.

En certains endroits de cette thèse, les propriétés distributionnelles sont à l'étude: la distribution de probabilité empirique de mesures répétées est typiquement la première chose qu'un modèle doit reproduire. Concrètement, les quatre premiers moments d'une distribution sont très informatifs quant au phénomène à l'œuvre, mais une perspective d'ensemble précise n'est atteignable que si des outils puissants sont disponibles pour comparer les fonctions de répartition. Développer de tels outils est un des buts de ce travail de doctorat.

Les corrélations et la dynamique peuvent être étudiées comme deux faces d'une même pièce, comme des dépendances spatiales et temporelles. Elles peuvent être vues respectivement comme une manifestation de propagation horizontale d'information, ou d'incorporation de l'information passée dans les réalisations présentes. Ces dépendances sont le sujet principal de cette thèse, avec des applications en finance qui pourraient avoir d'importantes conséquences en gestion du risque.

Non-linéarités

Pour des raisons historiques et de commodité mathématique, les corrélations linéaires sont souvent considérées par les professionnels (et quelques fois aussi dans le monde académique et l'enseignement) comme l'alpha et l'oméga des dépendances. Cette description est satisfaisante tant que l'on se trouve dans un régime où de petites perturbations ont de petites conséquences, mais l'expérience enseigne que ce n'est pas toujours le cas. Souvent, des effets de seuil, de la latence et de la mémoire (hystérèse), des effects collectifs (avalanches), du chaos (sensibilité extrême aux conditions initiales), etc. peuvent générer des comportements anormaux, et il est alors nécessaire de faire attention aux non-linéarités dans les dépendances, en étudiant par exemple les probabilités d'évènements de queue joints, les corrélations conditionnelles, corrélations quadratiques, etc. Dans des situations extrêmes, les corrélations linéaires peuvent même être trompeuses pour la compréhension des interactions sous-jacentes, et comme l'a écrit D. Helbing, "un modèle non-linéaire simple peut être capable d'expliquer des phénomènes, que même des modèles linéaires compliqués peuvent échouer à reproduire. Les modèles non-linéaires pourraient apporter un éclairage nouveau sur les phénomènes sociaux. Ils pourraient même conduire à un changement de paradigme dans la manière dont nous interprétons la société." [Helbing 2012]

1.1.3   Plan général de la thèse, et principales contributions originales

La thèse est composée de trois parties. La partie I introduit les outils mathématiques et statistiques appropriés pour l'étude des dépendances. La plupart de ce qui y est présenté est connu, mais je passe en revue les définitions et propriétés principales. C'est aussi dans cette partie que j'introduis les notations que j'ai essayé de maintenir cohérentes tout au long de la thèse. J'ai passé beaucoup de temps à travailler sur des tests statistiques d'adéquation ("Goodness of Fit", en anglais) pour des distributions de probabilité empiriques. Cela était originellement motivé par le besoin de tester les modèles bivariés finaux contre des hypothèses nulles. Ce "side project" initial s'est avéré être lié à un problème non-résolu et connu pour être difficile. Je n'apporte pas de solution définitive, mais discute en annexe de possibles directions pour des travaux futurs, qui ont émergé au cours de sessions de remue-méninges avec Jean-Philippe Bouchaud et Alexandre Richard. Pour autant, le sujet est fascinant, étant à la frontière de la théorie des probabilités (processus stochastiques) et la physique (particules quantiques dans des potentiels), et j'ai été intéressé de travailler sur d'autres extensions des tests d'adéquation usuels. Ce qui était censé n'être que le développement d'une boîte à outils pour un usage immédiat a finalement donné lieu à de très intéressants projets théoriques qui se sont concrétisés par la publication de deux articles et d'un chapitre de livre.

Le contenu financier de la thèse commence à la partie II. J'y présente mes travaux concernant les dépendances transversales entre les séries chronologiques de rendements journaliers d'actions, c'est à dire les forces instantanées qui relient plusieurs actions entre elles et les fait se comporter collectivement plutôt qu'individuellement. Ce vaste sujet avait déjà été approché dans mon travail de master [Chicheportiche 2010], où j'ai effectué une étude empirique des dépendances non-linéaires, et ai décrit une construction hiérarchique pour les modéliser. Une partie de ce travail a été publiée avec des développements consécutifs dans un journal de finance, et a servi de prémisse pour le modèle à facteur sur lequel j'ai finalement abouti au cours du doctorat. Une calibration de ce modèle est présentée ici, avec une comparaison aux données réelles.

Finalement, la partie III étudie les dépendances temporelles dans des séries chronologiques individuelles, en utilisant les mêmes outils et mesures de corrélations. Il est bien connu que les rendements d'actions ont une très faible auto-corrélation linéaire, ce qui se manifeste dans les alternances aléatoires de signes (rendements positifs ou négatifs), mais leurs amplitudes ont une longue mémoire. L'effet résultant de "volatility clustering" est une des plus anciennes énigmes en finance, et nous proposons ici deux contributions à l'étude de son origine et de sa description: l'une est une généralisation du mécanisme de rétro-action ARCH dans lequel les rendements sont auto-excitants, et l'autre est une description plus originale des auto-dépendances en termes de copule. Cette dernière peut être formulée sans modèle et n'est pas spécifique aux données financières. En fait, je montre ici aussi, après une longue collaboration avec Anirban Chakraborti, comment les concepts de récurrences, records, répliques et temps d'attente, qui caractérisent la dynamique dans les séries chronologiques, peuvent être écrits dans la cadre unifié des copules.

La thèse est rédigée au pluriel, car la plupart de son contenu est le fruit de collaborations (à tout le moins avec mes directeurs), et parce que j'y inclus beaucoup de matériel publié. Je renvoie aux articles listés en page v et aux remerciements qu'ils contiennent pour l'identité de mes co-auteurs et collaborateurs sur chaque sujet.

1.2 Introduction in English

1.2.1 The financial markets as complex systems

The economy presents features that are characteristic of complex systems: a large number of variables (agents, firms, contracts, maturities, etc.), heterogeneous agents (difficulty to isolate groups), partial information and randomness, self-organization and emergent phenomena (strong interactions, non-perturbativity), network dynamics (among banks, countries, counter-parties, companies, etc.) and long-ranged interactions, quasi-continuum of time scales, non-stationarity, history dependence and sensitivity to external forces (regulations, catastrophes, etc.).

This reality is in sharp contrast with the hypotheses of classical economic and financial theory, like homogeneity (representative agent), perfect rationality, infinite horizon forecasting capabilities, steady-state world where perturbations of the equilibrium are instantaneously adjusted, absence of memory (all past information is reflected in the current state of the world), ...

To address these issues with a more empirical perspective, different communities with diverse point of views have progressively invested the social and economic sciences. These include quantitative sociologists, econophysicists, statisticians (statistical physicists, biostatisticians), applied mathematicians and others. Indeed, many of the features listed above have already been encountered in various areas of natural sciences, where methods have been designed to account for randomness, network interactions, collective behaviors, persistence and reversion, etc. [Anderson 1988, Arthur 1997, Blume 2006, Challet 2005, Galam 2008, Kirman 2010, Abergel 2013, Bouchaud 2013]

Among all the fields of socio-economic activity, market finance may however be the closest to what can be reasonably studied like a physical system: on the theoretical side, the number of degrees of freedom is somewhat reduced by the rules of the exchanges (auctions, order books, regulations, standard contracts, information disclosure) while the number of "players" and "actions" is still very large allowing for averaging in statistical ensembles; and on the empirical side, it produces huge amounts of data making it possible to "repeat experiments" under assumptions, and treat the measurements with appropriate statistical tools. In fact there are evidences that some properties of financial markets resemble those of systems studied in statistical physics: intermittency in turbulent flows, crackling noise of materials responding to changing external conditions, lines of fractures in solids, long-ranged or mean-field interactions in Ising spins systems, phase transitions and self-organized criticality, to name a few.

1.2.2 Field covered in the thesis

This thesis studies statistical dependences in a general sense, both cross-sectionally (between random variables) and temporally (successive realizations of one variable). All theoretical results thus apply in principle to any field where many variables are interacting or where stationary processes have memory. More specifically, the study is restricted to discrete processes with equidistant intervals, meaning that the corresponding series are sampled and need not be "time-stamped".

As of the applied content, the focus is on financial data, more precisely time series of stock prices (or returns) at the rather low frequency of the day. Mainly, the prices of the most liquid stocks traded in US markets are used, whose histories can be downloaded freely from different sources.

Statistical properties, dependences and dynamics

The statistical characteristics of a set of random variables are of many kinds: distributional properties, correlations and dynamics.

In some places of this thesis, distributional properties are under the spotlights. As a matter of principle, the empirical probability distribution function of repeated measurement of a variable is the first thing a model should reproduce. Concretely, the first four moments of a distribution are much informative about the underlying phenomenon, but a precise global picture is only achieved if powerful tools are available to compare distribution functions. Developing such tools is one of the goals of this PhD work.

Correlations and dynamics can be studied as two faces of the same coin: they are spatial (or "cross-sectional") dependences and temporal dependences, and can be regarded as a manifestation of horizontal information propagation, or incorporation of past information into current realizations, respectively. These dependences are the main subject of the present thesis, with applications in finance that could have important consequences in risk management.

Non-linearities

For historical reasons and mathematical convenience, linear correlations are often seen by practitioners (and sometimes also in academia and teaching) as the "be-all and end-all" of dependences. This works fine as long as small perturbations have small effects, but experience teaches that this is not always the case. Often, threshold effects, latency and memory (hysteresis), collective effects (avalanches), chaos (sensitivity to initial conditions) generate abnormal behavior, and it is necessary to care for non-linearities in dependences by studying, for example, tail joint probabilities, conditional correlations, quadratic correlations, etc. In extreme situations, linear correlations may even be misleading the understanding of the underlying interactions, and, as D. Helbing wrote, "a simple nonlinear model may explain phenomena, which even complicated linear models may fail to reproduce. Nonlinear models are expected to shed new light on [...] social phenomena. They may even lead to a paradigm shift in the way we interpret society." [Helbing 2012]

1.2.3 Detailed outline of the thesis and main original contributions

The thesis is composed of three parts. Part I introduces the mathematical and statistical tools that are relevant for the study of dependences. Most of what is presented there is common knowledge, but I review the main definitions and properties for completeness and later usage. It is also the place where I introduce the notations, which I tried (but probably not succeeded completely) to keep coherent along the thesis. I have spent quite some time working on statistical tests of Goodness-of-fit for empirical probability distributions. This was originally motivated by the need to test the final multivariate models, or at least the bivariate marginals,

against null hypotheses. This initial "side project" turned out to be a tough one and no definite solution is provided here — I still give in appendix an account of the literature and of some possible directions for future endeavor, that came out during brainstorming sessions with Jean-Philippe Bouchaud and Alexandre Richard. Yet the subject was fascinating, being at the frontier of probability theory (stochastic processes) and plain physics (quantum particles in potentials), and I was interested in working on other extensions of the usual GoF tests. What was supposed to be only the development of a toolkit for immediate use gave rise to very interesting theoretical projects and in fact led to the publication of two articles and a textbook chapter.

The financial content of the thesis starts in Part II. I present there my studies regarding the "cross-sectional" dependences among the time series of daily stock returns, i.e. the instantaneous forces that link several stocks together and make them behave somewhat collectively rather than purely independently. Some work was already done in my Master's thesis [Chicheportiche 2010], where I performed an empirical study of non-linear dependences, and described a hierarchical construction to model them. Part of that work was published with subsequent developments in a financial journal, and served as a premise for the final factor model that I have come with. A calibration of this model is presented here, together with a comparison to measurements on real data.

Finally, Part III investigates the temporal dependences of single time series, using the same tools and measures of correlation. As is well known, stock returns have a vanishing linear autocorrelation that manifests itself in random sign alternations, but their magnitudes have a long memory. The resulting "volatility clustering" effect is one of the oldest puzzles in finance, and we propose here two contributions to the study of its origin and description: one is a generalization of the ARCH-like feedback construction where the returns are self-exciting, and the other one is a more original description of self-dependences in terms of copulas. The latter can be formulated model-free and is not specific to financial time series. In fact, I also show here, after a longstanding collaboration with Anirban Chakraborti and undergrad students at Ecole Centrale, how concepts like recurrences, records, aftershocks and waiting times, that characterize the dynamics in a time series can be written in the unifying framework of the copula.

After this brief non-technical overview of the landscape, I give below a more detailed outline with an abstract of each chapter. I quit now the first person and switch to the 'we' as most of what follows is the result of collaborations (at the very least with my advisors), and because I include much material of published work. I refer to the articles listed in page v and acknowledgments therein for the identity of my co-authors on each topic.

Part I

I.2. There exist many measures of dependence between random variables, some related to joint amplitudes, others related to joint occurrence probabilities. We recall the definition and properties of some of them, and introduce the unifying framework of the copula, an object that embeds all the linear and non-linear dependences that are invariant under a rescaling of the marginals, i.e. those that count joint occurrences. We then rewrite the important measures of dependences in terms of the copula, and suggest ways to efficiently visualize and compare them. An interesting point of comparison is

the elliptical copula, that includes the Gaussian and Student cases. We define this class of multivariate distributions, and summarize its main features.

The content of this chapter needs not be read straight from the beginning, but can be accessed "on demand" when referred to, later in the text. The original material is limited to one theorem characterizing the asymptotic tail dependence coefficient of a Student pair.

I.3. Usual goodness-of-fit (GoF) tests are designed for *independent* samples, and are not suited to investigate *tail regions*, because of the universal limit properties of the cumulative distribution functions. We extend the range of applicability of the GoF tests to these two cases. **Section 3.1**: Accurate tests for the extreme tails of empirical distributions is a very important issue, relevant in many contexts, including geophysics, insurance, and finance. We have derived exact asymptotic results for a generalization of the large-sample Kolmogorov-Smirnov test, well suited to testing a distribution with constant weight at any point of the domain, and in particular provide an improved resolution in the tail regions. In passing, we have rederived and made more precise the approximate limit solutions found originally in unrelated fields. **Section 3.2**: We revisit the Kolmogorov-Smirnov and Cramér-von Mises goodness-of-fit tests and propose a generalization to identically distributed, but dependent univariate random variables. We show that the dependence leads to a reduction of the "effective" number of independent observations. The generalized GoF tests are not distribution-free but rather depend on all the lagged bivariate copulas. Hence, a precise formulation of the test must rely on a modeled or estimated copula.

Part II

II.4. Using a large set of daily US and Japanese stock returns, we test in detail the relevance of Student models, and of more general elliptical models, for describing the joint distribution of returns. We find that while Student copulas provide a good approximation for strongly correlated pairs of stocks, systematic discrepancies appear as the linear correlation between stocks decreases, that rule out *all* elliptical models. Intuitively, the failure of elliptical models can be traced to the inadequacy of the assumption of a single volatility mode for all stocks. We suggest several ideas of methodological interest to efficiently visualize and compare different copulas. We identify the rescaled difference with the Gaussian copula and the central value of the copula as strongly discriminating observables. We insist on the need to shun away from formal choices of copulas with no financial interpretation.

II.5. We propose a non-hierarchical multi-factor model for the joint description of daily stock returns. The model intends to better reproduce linear and non-linear dependences of empirical time series. The non-Gaussianity of the factors is an important ingredient of the model, that allows for the empirically observed anomalous copula at the medial point and other stylized facts (quadratic covariances, copula diagonals). A spectral analysis of the factor series suggests that a structure is present in the *volatilities*, with a dominant mode clearly affecting all linear factors and residuals. The model embedding this feature is calibrated on US stocks over several periods, and reproduces qualitatively

all the non-trivial empirical observations. A systematic out-of-sample prediction over a long period confirms quantitatively the power of the model and of the proposed estimation methodology, both for linear and non-linear properties.

Part III

III.7. We attempt to unveil the fine structure of volatility feedback effects in the context of general quadratic autoregressive (QARCH) models, which assume that today's volatility can be expressed as a general quadratic form of the past daily returns. The standard ARCH or GARCH framework is recovered when the quadratic kernel is diagonal. The calibration of these models on US stock returns reveals several unexpected features. The off-diagonal (non ARCH) coefficients of the quadratic kernel are found to be highly significant both In-Sample and Out-of-Sample, but all these coefficients turn out to be one order of magnitude smaller than the diagonal elements. This confirms that daily returns play a special role in the volatility feedback mechanism, as postulated by ARCH models. The feedback kernel exhibits a surprisingly complex structure, incompatible with models proposed so far in the literature. Its spectral properties suggest the existence of volatility-neutral patterns of past returns. The diagonal part of the quadratic kernel is found to decay as a power-law of the lag, in line with the long-memory of volatility. Finally, QARCH models suggest some violations of Time Reversal Symmetry in financial time series, which are indeed observed empirically, although of much smaller amplitude than predicted. We speculate that a faithful volatility model should include both ARCH feedback effects and a stochastic component.

III.8. A discrete time series is a collection of successive realizations of a random variable, every realization being conditional on the previous state. Seen collectively and *ex ante*, it can also be seen as one realization of a random vector with (directed) dependence. We show how the copula introduced in Part I in the context of cross-sectional dependences can also describe appropriately the non-linear temporal dependences in time series: in this context, we call them "self-copulas".

III.9. We introduce a specific, log-normal model for these self-copulas, for which a number of analytical results are derived. An application to financial time series is provided. As is well known, the dependence is to be long-ranged in this case, a finding that we confirm using self-copulas. As a consequence, the acceptance rates for GoF tests are substantially higher than if the returns were iid random variables.

Appendices

A. Empirical counterparts of the copula (estimated together with the marginals on a sample of size T) have good properties only on discrete values in $[0, 1]$: for finite T, the naive copula estimates are biased. We provide an approximate bias-correction mechanism.

B. Gaussian sheets (the generalization of Gaussian process indexed by two "times") arise as limits of empirical multivariate cumulative distribution functions, when the sample size tends to infinity. Some properties of such sheets are of importance for distribution

testing, since the Kolmogorov-Smirnov test statistic is related to the law of the supremum of their absolute value. We study some time changes and discretization schemes that may allow to reach a better knowledge of this law.

C. Two contributions of methodological interest are made for models with a single common linear factor. In C.1 the estimator of optimal weights is found perturbatively in the limit of large dimension. In C.2 we reverse-engineer the constitution of the common factor, and embed the one-factor model in a hierarchical nested structure.

D. We provide here technical details for the perturbative expansion of the pseudo-elliptical copula with log-normal scale around the independence copula, when all dependence parameters are small.

E. Two appendices related to the continuous-time limit of the QARCH construction studied in Chapter III.6. In E.1 we compute the exact spectrum of the Borland-Bouchaud model of Ref. [Borland 2011], for three different kernel functions. In E.2 we relate the power-law behavior in the volatility correlation function, to the coefficient of the power-law decay in the kernel function of the FIGARCH model.

Part I

Mathematical and statistical tools

Characterizing the statistical dependence

Contents

2.1   Bivariate measures of dependence

In this first section, we recall several bivariate measures of statistical dependence between two random variables. Of course, as we discuss in the next section, many-points dependences among N variables do not reduce to the pairwise dependences in the general case. But in the perspective of empirically measuring the dependences on a dataset, at least four issues motivate the present restriction to 2-points coefficients: (i) the number of triplets, quadruplets, etc. that should be considered when measuring coefficients of dependence involving 3-points, 4-points, etc. gets rapidly huge when considering even as few as some dozens of variables, whereas the number of undirected 2-points dependence coefficients is "only" $N(N-1)/2$; (ii) error estimates on many-points coefficients are typically larger than on 2-points measures; (iii) in some classes of usual multivariate probability distributions, some information about the full structure can be inferred (or excluded) from the knowledge of the bivariate marginals; (iv) collective behavior and "effective" many-points dependences can be uncovered by collecting 2-points coefficients into a matrix or tensor and performing algebraic and spectral analysis.

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_N)^\dagger$ be a random vector with joint probability distribution function (pdf) $\mathcal{P}_{\boldsymbol{X}}(\boldsymbol{x})$ and joint cumulative distribution function (cdf) $\mathcal{P}_{<,\boldsymbol{X}}(\boldsymbol{x}) = \int \mathcal{P}_{\boldsymbol{X}}(\boldsymbol{z})\mathbb{1}\{\boldsymbol{z} \leq \boldsymbol{x}\}\,d\boldsymbol{z}$. Each component X_i is a random variable with marginal pdf $\mathcal{P}_i(x_i)$ and cdf $\mathcal{P}_{<,i}(x_i)$. Throughout the text (unless otherwise stated), we assume $\mathcal{P}_{<,\boldsymbol{X}}$ is continuous and $\mathcal{P}_{<,i}^{-1}$ are well-defined inverse marginal cdfs.

As the coefficients defined in this section are all measures of pairwise dependence, we define the pair $\boldsymbol{X_{ij}} = (X_i, X_j)^\dagger$. Similarly, all the transformations of $\boldsymbol{X}$ defined later in the text can be restricted to the 2-dimensional space describing the pair.

2.1.1   Usual correlation coefficients

Bravais-Pearson's correlation coefficient

The Bravais-Pearson's correlation coefficient of $\boldsymbol{X_{ij}}$ is the usual normalized linear covariance

$$\rho^{(\mathrm{P})}[\boldsymbol{X_{ij}}] = \frac{\mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j]}{\sqrt{\mathbb{V}[X_i]\mathbb{V}[X_j]}} \tag{2.1}$$

Often in what follows, it will be simply denoted ρ_{ij}. This correlation measure involves the probabilities *and* amplitudes of the possible realizations of $\boldsymbol{X_{ij}}$. It can also be understood as a measure of asymmetry between the sum and the difference of the normalized variables:

$$\rho^{(\mathrm{P})}[\boldsymbol{X_{ij}}] = \frac{\mathbb{V}\left[\frac{X_i}{\sqrt{\mathbb{V}[X_i]}} + \frac{X_j}{\sqrt{\mathbb{V}[X_j]}}\right] - \mathbb{V}\left[\frac{X_i}{\sqrt{\mathbb{V}[X_i]}} - \frac{X_j}{\sqrt{\mathbb{V}[X_j]}}\right]}{\mathbb{V}\left[\frac{X_i}{\sqrt{\mathbb{V}[X_i]}} + \frac{X_j}{\sqrt{\mathbb{V}[X_j]}}\right] + \mathbb{V}\left[\frac{X_i}{\sqrt{\mathbb{V}[X_i]}} - \frac{X_j}{\sqrt{\mathbb{V}[X_j]}}\right]} \tag{2.1$'$}$$

(the expression at the denominator is just an explicit way of writing the number 2).

Spearman's rho

Let $\boldsymbol{X}^{(i)} \overset{L}{=} (0, \ldots, X_i, \ldots, 0)$ be distributed like the i-th marginal and independent of $\boldsymbol{X}$. Then the vector $\sum_{i=1}^{N} \boldsymbol{X}^{(i)}$ has independent components with the same marginals as $\boldsymbol{X}$. The differ-

ence $\widetilde{\boldsymbol{X}} = \boldsymbol{X} - \sum_{i=1}^{N} \boldsymbol{X}^{(i)}$ is centered and has symmetric univariate marginals by construction. It allows to characterize the dependence in $\boldsymbol{X}$ by counting how often both components of $\boldsymbol{X_{ij}}$ are *simultaneously* below the realizations of the independent couple with same marginals. Spearman's rho is defined as

$$\rho^{(\mathrm{S})}[\boldsymbol{X_{ij}}] = 3\left(\mathbb{P}\left[\widetilde{X}_i\widetilde{X}_j > 0\right] - \mathbb{P}\left[\widetilde{X}_i\widetilde{X}_j < 0\right]\right) \tag{2.2}$$

and thus involves only the signs of $\widetilde{\boldsymbol{X}}$:

$$\rho^{(\mathrm{S})}[\boldsymbol{X_{ij}}] = 3\,\rho^{(\mathrm{P})}[\mathrm{sign}(\widetilde{\boldsymbol{X}_{ij}})] \tag{2.2$'$}$$

Kendall's tau

Let $\boldsymbol{X}' \overset{L}{=} \boldsymbol{X}$ be a random vector with same joint distribution as $\boldsymbol{X}$: compared to the previous case, $\boldsymbol{X}'$ has not only the same univariate marginals as $\boldsymbol{X}$, but also its dependence structure. Then $\widetilde{\boldsymbol{X}} = \boldsymbol{X} - \boldsymbol{X}'$ is the difference between two independent and identically distributed random vectors. It is centered and symmetric by construction. Kendall's tau is defined as

$$\tau^{(\mathrm{K})}[\boldsymbol{X_{ij}}] = \mathbb{P}\left[\widetilde{X}_i\widetilde{X}_j > 0\right] - \mathbb{P}\left[\widetilde{X}_i\widetilde{X}_j < 0\right] = \rho^{(\mathrm{P})}[\mathrm{sign}(\widetilde{\boldsymbol{X}_{ij}})] \tag{2.3}$$

and it measures concordances between realizations of identically distributed couples.

Blomqvist's beta

Denote now $\widetilde{\boldsymbol{X}} = \boldsymbol{X} - \mathrm{med}\,\boldsymbol{X}$. Blomqvist's beta coefficient of dependence is defined as

$$\beta^{(\mathrm{B})}[\boldsymbol{X_{ij}}] = \mathbb{P}\left[\widetilde{X}_i\widetilde{X}_j > 0\right] - \mathbb{P}\left[\widetilde{X}_i\widetilde{X}_j < 0\right] = \rho^{(\mathrm{P})}[\mathrm{sign}(\widetilde{\boldsymbol{X}_{ij}})]. \tag{2.4}$$

If all X_i's are centered around their median value (in particular if symmetric), then Blomqvist's beta is equal to the Pearson's correlation of the signs:

$$\beta_{ij}^{(\mathrm{B})} = \rho^{(\mathrm{P})}[\mathrm{sign}(\boldsymbol{X_{ij}})].$$

2.1.2 Non-linear correlation coefficients

Beyond the standard correlation coefficient, one can characterize the dependence structure through the correlation between functions $g(\boldsymbol{X_{ij}}) = (g_i(X_i), g_j(X_j))$ of the random variables. In particular, we define the following coefficients:

$$\rho_{ij}^{(d)} = \rho^{(\mathrm{P})}[g(\boldsymbol{X_{ij}})], \quad \text{when} \quad g_i(x) = g_j(x) = \mathrm{sign}(x)|x|^d \tag{2.5}$$

$$\zeta_{ij}^{(d)} = \rho^{(\mathrm{P})}[g(\boldsymbol{X_{ij}})], \quad \text{when} \quad g_i(x) = g_j(x) = |x|^d \tag{2.6}$$

provided the variances in the denominators are finite. The case $\rho^{(1)}$ corresponds to the usual linear correlation coefficient, for which we will use the standard notation ρ, whereas $\zeta^{(2)}$ is the correlation of the squared returns, that would appear in the Gamma-risk of a Δ-hedged option portfolio. However, high values of d are expected to be very noisy in the presence

of power-law tails (as is the case for financial returns) and one should seek for lower order moments, such as $\zeta^{(1)}$ which also captures the correlation between the *amplitudes* (or the volatility) of price moves, or even $\rho^{(s)} \equiv \rho^{(0)}$ that measures the correlation of the signs. Note that when $g_i(x) = \mathcal{P}_{<,i}(x)$, one recovers Spearman's rho:

$$\rho_{ij}^{(S)} = \rho^{(P)}[g(\boldsymbol{X_{\ddot{\imath}\!\jmath}})], \quad \text{when} \quad g_i(x) = \mathcal{P}_{<,i}(x). \tag{2.7}$$

Proof of Eq. (2.7). The LHS is decomposed onto four contributions:

$$\mathbb{P}[X_i > X_i^{(i)}, X_j > X_j^{(j)}] = \mathbb{P}[U_i > U_i^{(i)}, U_j > U_j^{(j)}]$$

$$= \iint_0^1 \mathbb{P}[u_i > U_i^{(i)}, u_j > U_j^{(j)}] \, \mathrm{d}C_{\boldsymbol{X_{ij}}}(u_i, u_j)$$

$$= \iint_0^1 u_i u_j \, \mathrm{d}C_{\boldsymbol{X_{ij}}}(u_i, u_j) \qquad \text{since } U_i^{(i)} \perp U_j^{(j)}$$

$$\mathbb{P}[X_i < X_i^{(i)}, X_j < X_j^{(j)}] = \iint_0^1 (1-u_i)(1-u_j) \, \mathrm{d}C_{\boldsymbol{X_{ij}}}(u_i, u_j)$$

$$\mathbb{P}[X_i > X_i^{(i)}, X_j < X_j^{(j)}] = \iint_0^1 u_i \, (1-u_j) \, \mathrm{d}C_{\boldsymbol{X_{ij}}}(u_i, u_j)$$

$$\mathbb{P}[X_i < X_i^{(i)}, X_j > X_j^{(j)}] = \iint_0^1 (1-u_i) \, u_j \, \mathrm{d}C_{\boldsymbol{X_{ij}}}(u_i, u_j).$$

Define then $U_i = \mathcal{P}_{<,i}(X_i) \sim \mathcal{U}[0,1]$; since $\mathbb{E}[U_i] = \frac{1}{2}$, the RHS is

$$\rho^{(P)}[\boldsymbol{U_{\ddot{\imath}\!\jmath}}] = \iint_0^1 (u_i - \tfrac{1}{2})(u_j - \tfrac{1}{2}) \, \mathrm{d}C_{\boldsymbol{X_{ij}}}(u_i, u_j).$$

The result follows, noting that

$$u_i \, u_j + (1-u_i)(1-u_j) - u_i(1-u_j) - (1-u_i)u_j = 4(u_i - \tfrac{1}{2})(u_j - \tfrac{1}{2})$$

and $\mathbb{V}[U_i] = \mathbb{V}[U_j] = \int_0^1 (u - \tfrac{1}{2})^2 \, \mathrm{d}u = \frac{1}{12}$. $\qquad\qquad\qquad\qquad\qquad \square$

Although the motivation of the definition (2.2) is quite intuitive (compare dependent and independent couples with same distributions), the definition of Spearman's rho as Pearson's correlation of the ranks $U_i = \mathcal{P}_{<,i}(X_i)$ makes it clear that it is invariant under any continuous stretching or shrinking of the axis, and that it involves the dependence structure regardless of the marginals.

2.1.3   Tail dependence

Another characterization of dependence, of great importance for risk management purposes, is the so-called *coefficient of tail dependence* which measures the joint probability of extreme events. More precisely, the upper tail dependence is defined as [Embrechts 2002b, Malevergne 2006]:

$$\tau_{ij}^{\text{UU}}(p) = \mathbb{P}\left[X_i > \mathcal{P}_{<,i}^{-1}(p) \,\middle|\, X_j > \mathcal{P}_{<,j}^{-1}(p)\right], \tag{2.8}$$

where $\mathcal{P}_{<,k}$ is the cumulative distribution function (cdf) of X_k, and p a certain probability level. In spite of its seemingly asymmetric definition, it is easy to show that τ^{UU} is in fact

symmetric in $X_i \leftrightarrow X_j$. When $p \to 1$, $\tau_*^{\mathrm{UU}} \equiv \tau^{\mathrm{UU}}(1)$ measures the probability that X_i takes a very large positive value knowing that X_j is also very large, and defines the asymptotic tail dependence. Random variables can be strongly dependent from the point of view of linear correlations, while being nearly independent in the extremes. For example, bivariate Gaussian variables are such that $\tau_*^{\mathrm{UU}} = 0$ for any value of $\rho < 1$. The lower tail dependence τ^{LL} is defined similarly:

$$\tau_{ij}^{\mathrm{LL}}(p) = \mathbb{P}\left[X_i < \mathcal{P}_{<,i}^{-1}(1-p) \,\middle|\, X_j < \mathcal{P}_{<,j}^{-1}(1-p)\right], \tag{2.9}$$

and is equal to $\tau^{\mathrm{UU}}(p)$ for symmetric bivariate distributions. One can also define mixed tail dependences:

$$\tau_{ij}^{\mathrm{LU}}(p) = \mathbb{P}\left[X_i < \mathcal{P}_{<,i}^{-1}(1-p) \,\middle|\, X_j > \mathcal{P}_{<,j}^{-1}(p)\right], \tag{2.10}$$

$$\tau_{ij}^{\mathrm{UL}}(p) = \mathbb{P}\left[X_i > \mathcal{P}_{<,i}^{-1}(p) \,\middle|\, X_j < \mathcal{P}_{<,j}^{-1}(1-p)\right], \tag{2.11}$$

with obvious interpretations.

2.2   Copulas

2.2.1   Definition and properties

A copula is a N-variate cdf $C(\boldsymbol{u})$ on the hypercube $[0,1]^N$, with uniform univariate marginals.

Theorem ([Sklar 1959]). A unique copula can be associated with any multivariate cdf $\mathcal{P}_{<,\boldsymbol{X}}$ by uniformizing its marginals:

$$C_{\boldsymbol{X}}(u_1,\ldots,u_N) = \mathcal{P}_{<,\boldsymbol{X}}\left(\mathcal{P}_{<,1}^{-1}(u_1),\ldots,\mathcal{P}_{<,N}^{-1}(u_N)\right) \tag{2.12}$$

According to this view, the copula of a random vector is the joint cdf of the marginal *ranks* $U_i = \mathcal{P}_{<,i}(X_i)$. It hence encodes all the dependence between the individual random variables that is invariant under increasing transformations. Since the marginals of U_i are uniform by construction, the copula only captures their degree of "entanglement".

Of course, the converse of Sklar's theorem is not true: infinitely many multivariate distributions can have the same copula ! In fact, it is in principle possible to build joint distributions with given copula and *any* univariate (continuous) marginals, for example a bivariate Gaussian copula with chi-2 marginals. Of course one sees that although possible, such exotic constructions barely make sense, as one would intuitively expect that random variables with positive support do not have a copula with the reversal symmetry $u_i \leftrightarrow 1 - u_i$.

Construction   A copula can be constructed in three manners:

1. From the very definition, by explicitly specifying a function $C(\boldsymbol{u})$ satisfying all required properties. Examples include

   - the independence (product) copula: $\Pi(\boldsymbol{u}) = u_1\, u_2\, \ldots\, u_N$;

   - the upper Fréchet-Hoeffding bound copula $M(\boldsymbol{u}) = \min\{u_1,\ldots,u_N\}$;

   - the lower Fréchet-Hoeffding bound $W(\boldsymbol{u}) = \max\{u_1 + \ldots + u_N - (N-1), 0\}$ which is a copula only for $N = 2$;

   - Archimedean copulas: $C(\boldsymbol{u}) = \phi^{-1}\left(\phi(u_1) + \ldots + \phi(u_N)\right)$,
     where ϕ (the generator function) is $N - 2$ times continuously differentiable, and such that $\phi(1) = 0$, $\lim_{u \to 0} \phi(u) = \infty$ and $\phi^{(N-2)}$ is decreasing convex (more on this is Sect. 2.2.4, page 21).

2. Applying Sklar's theorem to usual classes of parametric multivariate distributions. Examples include the Gaussian copula

$$C_{\mathrm{G}}(\boldsymbol{u}) = \Phi_\Sigma(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_N))$$

   where Σ is a symmetric positive definite matrix. And the more general Elliptically Contoured copulas, that we introduce in the coming Sect. 2.4.

3. Implicitly defined as the dependence function of a random vector described through a structural model. Example: factor models.

Fréchet-Hoeffding bounds Copulas are multivariate cumulative distribution functions, and hence take values between 0 and 1, but there are in fact stronger constraints accounting for the properties of probability distributions. It is obvious that a univariate cdf tends to 0 at the extreme low values of its domain, and to 1 at the extreme high values, reflecting roughly the facts that the probability of an empty set is 0 and the probability of the whole universe is 1. Much in this spirit, the extreme values of the copulas are "pinned" to $C(0, \ldots, 0) = 0$ and $C(1, \ldots, 1) = 1$. More precisely, if *any* component is 0, then the copula is 0, too. Furthermore, by reduction to the univariate marginals, if *all but one* components are 1, then the copula takes the value of the remaining component. Beyond these properties related to the marginals, the copula is subject to bounds imposed by the dependence structure: in short, a random pair cannot be more positively correlated than if it was built of twice the same variable, and similarly it cannot be more negatively correlated than if it was composed of a variable and its own opposite. This is more precisely stated by the following inequalities, for a copula of arbitrary dimension:

$$W(\boldsymbol{u}) \leq C(\boldsymbol{u}) \leq M(\boldsymbol{u}), \tag{2.13}$$

where the definitions of the upper and lower bounds were given in the previous page. We show in Fig. 2.1 how these bounds look like on the diagonal and anti-diagonal of a bivariate copula.

2.2.2 Dependence coefficients and the copula

The copula can be restricted to subsets of the studied random vector. For example, the bivariate copula of any random pair (X_i, X_j) is

$$C(u_i, u_j) = \mathbb{P}[\mathcal{P}_{<,i}(X_i) \leq u_i \text{ and } \mathcal{P}_{<,j}(X_j) \leq u_j]. \tag{2.14}$$

It is then natural to ask if and how all the coefficients of bivariate dependence introduced earlier are related to the description in terms of copula. It turns out that, whereas the $\rho^{(d)}$'s and $\zeta^{(d)}$'s depend on the marginal distributions, Spearman's rho, Kendall's tau and Blomqvist's beta can be expressed in terms of the copula only:

$$\rho^{(\mathrm{S})} = 12 \int_{[0,1]^2} u_i u_j \, \mathrm{d}C(u_i, u_j) - 3$$

$$\tau^{(\mathrm{K})} = 4 \int_{[0,1]^2} C(u_i, u_j) \, \mathrm{d}C(u_i, u_j) - 1$$

$$\beta^{(\mathrm{B})} = 4C(\tfrac{1}{2}, \tfrac{1}{2}) - 1.$$

Similarly, the tail dependence coefficients can be expressed, for $u \lesssim 1$, as:

$$\tau^{\mathrm{UU}}(u) = \frac{1 - 2u + C(u, u)}{1 - u} \qquad\qquad \tau^{\mathrm{LL}}(u) = \frac{C(1-u, 1-u)}{1-u} \tag{2.15a}$$

$$\tau^{\mathrm{UL}}(u) = \frac{1 - u - C(u, 1-u)}{1 - u} \qquad\qquad \tau^{\mathrm{LU}}(u) = \frac{1 - u - C(1-u, u)}{1-u}. \tag{2.15b}$$
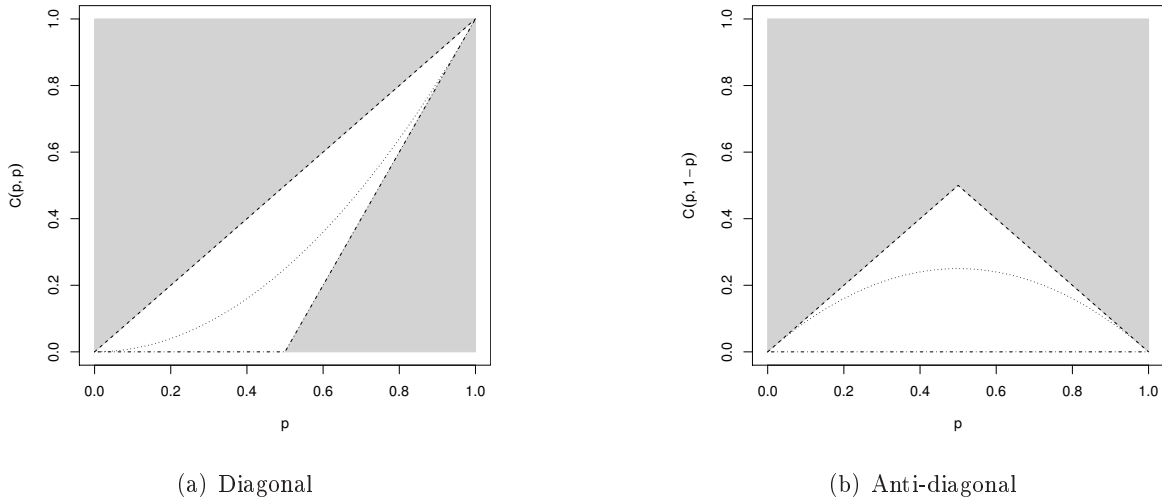
(a) Diagonal

(b) Anti-diagonal

Figure 2.1: Fréchet-Hoeffding bounds for the bivariate copula. The allowed space is the triangle delimited by dashed and dash-dotted lines, corresponding to the upper bound $M(\boldsymbol{u})$ (maximal dependence) and the lower bound $W(\boldsymbol{u})$ (maximal opposition), respectively. The product copula for independent variables is shown as a dotted line.

2.2.3 Visual representations

Copulas are not easy to visualize, first because they need to be represented as 3D plots of a surface in two dimensions and second because the bounds (2.13), within which $C(u,v)$ is constrained to live, compress the difference between two arbitrary copulas — the situation is even worse in higher dimensions. Estimating copula densities, on the other hand, is even more difficult than estimating univariate densities, especially in the tails. We therefore propose to focus on the diagonal of the copula, $C(u,u)$ and the anti-diagonal, $C(u, 1-u)$, that capture part of the information contained in the full copula, and can be represented as 1-dimensional plots. Furthermore, in order to emphasize the difference with the case of independent variables, it is expedient to consider the following quantities:

$$\frac{C(u,u) - u^2}{u\,(1-u)} = \tau^{\mathrm{UU}}(u) + \tau^{\mathrm{LL}}(1-u) - 1 \tag{2.16a}$$

$$\frac{C(u,1-u) - u\,(1-u)}{u\,(1-u)} = 1 - \tau^{\mathrm{UL}}(u) - \tau^{\mathrm{LU}}(1-u), \tag{2.16b}$$

where the normalization is chosen such that the tail correlations appear naturally. Another important reference point is the Gaussian copula $C_{\mathrm{G}}(u,v)$, and we will consider below the normalized differences along the diagonal and the anti-diagonal:

$$\Delta_{\mathrm{d}}(u) = \frac{C(u,u) - C_{\mathrm{G}}(u,u)}{u\,(1-u)}; \qquad \Delta_{\mathrm{a}}(u) = \frac{C(u,1-u) - C_{\mathrm{G}}(u,1-u)}{u\,(1-u)}. \tag{2.17}$$

The incentive to focus on the diagonals of the copula is motivated in the first place by the above argument that the whole copula is a particularly complex object that is hard to manipulate, estimate and visualize, and that some means have to be found in order to

investigate its subtleties while reducing the number of parameters. But it also has a profound meaning in terms of physical observables that can be empirically measured. In particular, the diagonal quantity (2.16a) tends to the asymptotic tail dependence coefficients τ_*^{UU} when $u \to 1$ (τ_*^{LL} when $u \to 0$). Similarly, the anti-diagonal quantity (2.16b) tends to $-\tau_*^{\mathrm{UL}}$ as $u \to 1$ and to $-\tau_*^{\mathrm{LU}}$ as $u \to 0$. This still holds for the alternative definition in Eq. (2.17): $\Delta_{\mathrm{d,a}}(u)$ tend to the asymptotic tail dependence coefficients in the limits $u \to 1$ and $u \to 0$, owing to the fact that these coefficients are all zero in the Gaussian case. As we show in Chapter III.7, the diagonal copula plays also an important role in the characterization of temporal dependences in time series: observables such as recurrence times, sequences lengths, waiting times, etc. are fundamentally many-points objects, and their statistics are characterized by the diagonal n-points copulas.

The center-point of the copula, $C(\frac{1}{2}, \frac{1}{2})$, is particularly interesting: it is the probability that both variables are simultaneously below their respective median. For bivariate Gaussian variables, one can show that:

$$C_{\mathrm{G}}(\tfrac{1}{2}, \tfrac{1}{2}) = \frac{1}{4} + \frac{1}{2\pi} \arcsin \rho \equiv \frac{1}{4} \left(1 + \rho^{(\mathrm{s})}\right), \tag{2.18}$$

where $\rho^{(\mathrm{s})}$ is the sign correlation coefficient defined on page 16. The trivial but remarkable fact is that the above expression holds for any elliptical models, that we define and discuss in the next section. This will provide a powerful test to discriminate between empirical data and a large class of copulas that have been proposed in the literature.

Generalized Blomqvist's beta   We report for completeness a possible multivariate generalization of the bivariate Blomqvist's beta, proposed in Ref. [Schmid 2007]:

$$\beta_{\boldsymbol{X}}(\boldsymbol{u}, \boldsymbol{v}) = \frac{C_{\boldsymbol{X}}(\boldsymbol{u}) - \Pi(\boldsymbol{u}) + \overline{C_{\boldsymbol{X}}}(\boldsymbol{v}) - \overline{\Pi}(\boldsymbol{v})}{M(\boldsymbol{u}) - \Pi(\boldsymbol{u}) + \overline{M}(\boldsymbol{v}) - \overline{\Pi}(\boldsymbol{v})}, \tag{2.19}$$

where $\overline{C_{\boldsymbol{X}}}$ is the survival copula defined similarly to Eq. (2.12) as

$$\overline{C_{\boldsymbol{X}}}(u_1, \ldots, u_N) = \mathcal{R}_{,\boldsymbol{X}} \left(\mathcal{P}_{>,1}^{-1}(1 - u_1), \ldots, \mathcal{P}_{>,N}^{-1}(1 - u_N)\right). \tag{2.20}$$

$\beta_{\boldsymbol{X}}$ has an immediate interpretation as a normalized distance between $C_{\boldsymbol{X}}$ and Π: it measures the effective global departure from independence normalized by the maximum departure from independence. When $N = 2$ and the copula is symmetric, it also nicely interpolates between several of the coefficients defined above, for example

$$\boldsymbol{u} = \boldsymbol{v} = (\tfrac{1}{2}, \tfrac{1}{2}) \Longrightarrow \beta_{\boldsymbol{X}}(\boldsymbol{u}, \boldsymbol{v}) = 4 C_{\boldsymbol{X}}(\tfrac{1}{2}, \tfrac{1}{2}) - 1 = \beta_{\boldsymbol{X}}^{(\mathrm{B})}$$

$$\boldsymbol{u} = (u, u), \boldsymbol{v} = (1, 1) \Longrightarrow \beta_{\boldsymbol{X}}(\boldsymbol{u}, \boldsymbol{v}) = \frac{C_{\boldsymbol{X}}(u, u) - u^2}{u(1 - u)} = \tau^{\mathrm{UU}}(u) + \tau^{\mathrm{LL}}(1 - u) - 1.$$

2.2.4   A word on Archimedean copulas

In the universe of all possible copulas, a particular family has become increasingly popular in finance: that of "Archimedean copulas". These copulas are defined as follows [Wu 2007]:

$$C_{\phi}(u, v) \equiv \phi^{-1}\left[\phi(u) + \phi(v)\right], \qquad u, v \in [0, 1] \tag{2.21}$$

where $\phi(u) : [0,1] \to [0,1]$ is a function such that $\phi(1) = 0$ and ϕ^{-1} is decreasing and completely monotone. For example, Frank copulas are such that $\phi_{\mathrm{F}}(u) = \ln[e^{\theta} - 1] - \ln[e^{\theta u} - 1]$ where θ is a real parameter, or Gumbel copulas, such that $\phi_{\mathrm{G}}(u) = (-\ln u)^{\theta}, 0 < \theta \leq 1$. The asymptotic coefficient of tail dependence are all zero for Frank copulas, whereas $\tau_*^{\mathrm{UU}} = 2 - 2^{\theta}$ (and all other zero) for the Gumbel copulas. The case of general multivariate copulas is obtained as natural generalization of the above definition.

One can of course attempt to fit empirical copulas with a specific Archimedean copula. By playing enough with the function ϕ, it is obvious that one will eventually reach a reasonable quality of fit. What we take issue with is the complete lack of intuitive interpretation or plausible mechanisms to justify why the returns of two correlated assets should be described with Archimedean copulas. This is particularly clear after recalling how two Archimedean random variables are generated: first, take two $\mathcal{U}[0,1]$ random variables s, w. Set $w' = K^{-1}(w)$ with $K(x) = x - \phi(x)/\phi'(x)$. Now, set:

$$u = \phi^{-1}\big[s\,\phi(w')\big]\,; \qquad v = \phi^{-1}\big[(1-s)\,\phi(w')\big]\,, \tag{2.22}$$

and finally write $X_1 = \mathcal{P}_{<,1}^{-1}(u)$ and $X_2 = \mathcal{P}_{<,2}^{-1}(v)$ to obtain the two Archimedean returns with the required marginals $\mathcal{P}_1$ and $\mathcal{P}_2$ [Genest 1993, Wu 2007]. Unless one finds a natural economic or micro-structural interpretation for such a labyrinthine construction, we content that such models should be discarded *a priori*, for lack of plausibility. In the same spirit, one should really wonder why the financial industry has put so much faith in Gaussian copulas models to describe correlation between *positive* default times, that were then used to price CDO's and other credit derivatives. We strongly advocate the need to build models bottom-up: mechanisms should come before any mathematical representation (see also [Mikosch 2006] and references therein for a critical view on the use of copulas).

2.3 Assessing temporal dependences

In the context of discrete time series, where i is a time index and $j = i + \ell$ ($\ell > 0$, say), the realization of the random variable X_j occurs after that of X_i. Nevertheless, before X_i is realized, the *a priori* probability distribution of X_j is not *conditional* but *joint* to that of X_j, and it is possible to use all the measures of dependence defined above in this context.

For example, ρ_{ij} is the coefficient of linear *auto*-correlation at lag $\ell = j - i$: it says in essence "how much of X_i is expected to be reproduced, ℓ time steps later, in the realization of X_j". When positive, it indicates persistence, while reversion is revealed by negative autocorrelation.

Blomqvist's beta (defined in Eq. (2.4), page 15) is somehow more intuitive, as it quantifies directly the difference in probability that the signs will persist or alternate.

When one wants to be more specific in the characterization of temporal dependences and study for example the persistence and reversion probabilities of one realization of arbitrary magnitude, one needs to resort to the copula. Typically, persistence of extreme events (p-th quantile, where p is close to 1) separated by a time lag ℓ is measured by the tail dependence coefficients $\tau_\ell^{\mathrm{UU}}(p)$ and $\tau_\ell^{\mathrm{LL}}(p)$, while reversion of extreme events is measured by $\tau_\ell^{\mathrm{UL}}(p)$ and $\tau_\ell^{\mathrm{LU}}(p)$. We will elaborate on the topic of copulas in time series in Chapter 7 of Part III.

All these coefficients will typically be of decreasing amplitude as ℓ gets larger, reflecting the loss of memory as time goes on and randomness comes in. There nevertheless exist systems where the ℓ-dependence is not monotonous but rather exhibits oscillations due to periodic activity (see for example the rising interest in electroencephalogram data analysis).

2.4   Review of elliptical models

2.4.1   General elliptical models

An elliptical distribution is characterized by a location parameter $\boldsymbol{\mu}$, a positive definite dispersion matrix Σ and a function ϕ with positive support: let us denote this class by $E_N(\boldsymbol{\mu}, \Sigma, \phi)$. Ellipticity of a random variable is a property of its characteristic function, as explicited in the following definition.

Definition.

$$\boldsymbol{X} \sim E_N(\boldsymbol{\mu}, \Sigma, \phi) \quad \Longleftrightarrow \quad \mathbb{E}\Big[\exp\Big(\mathrm{i}\boldsymbol{t}^\dagger(\boldsymbol{X} - \boldsymbol{\mu})\Big)\Big] = \phi(\boldsymbol{t}^\dagger \Sigma^{-1}\boldsymbol{t}) \tag{2.23}$$

The quadratic form in the argument of ϕ means that the points of constant characteristic function draw an ellipsoid in the Fourier space, whence the name (see [Cambanis 1981] for the construction and properties of elliptically contoured multivariate distributions). This definition in Fourier space is neither very convenient to build such variables, nor comfortable for its interpretation. The following theorem decomposes the elliptical random vector in terms of a uniformly distributed N-dimensional sphere and a random radial scale.

Theorem ([Cambanis 1981]). If A is a $(N \times k)$ matrix with $k \leq N$, and $\boldsymbol{U} \sim \mathcal{U}(S^k)$, then

$$\boldsymbol{X} = \boldsymbol{\mu} + \sigma \mathrm{A}\boldsymbol{U} \quad \Longrightarrow \quad \exists \phi \text{ s.t. } \boldsymbol{X} \sim E_N(\boldsymbol{\mu}, \mathrm{AA}^\dagger, \phi) \tag{2.24a}$$

The converse holds true and calls, for a positive definite Σ of rank k,

$$\boldsymbol{X} \sim E_N(\boldsymbol{\mu}, \Sigma, \phi) \quad \Longrightarrow \quad \exists \sigma \text{ and A s.t. } \mathrm{AA}^\dagger = \Sigma \quad \text{and} \quad \boldsymbol{X} = \boldsymbol{\mu} + \sigma \mathrm{A}\boldsymbol{U} \tag{2.24b}$$

The class of elliptical distributions comprises the Gaussian and Student distributions:

1. The Gaussian distribution: If the scaling variable has a chi-2 distribution with $k = \mathrm{rank}\,\Sigma$ degrees of freedom, i.e. $\sigma \sim \chi^2(k)$ then the resulting $\boldsymbol{X}$ is normally distributed:

$$\sigma \sim \chi^2(k) \quad \Longrightarrow \quad \sigma \mathrm{A}\boldsymbol{U} \sim \mathcal{N}(0, \Sigma)$$

2. The Student distribution: As is well known, the quotient of a Gaussian variable over a chi-2 variable is Student-distributed. Hence,

$$\sigma = \sqrt{\frac{\nu}{2}}\frac{\sigma_k}{\sigma_\nu} \quad \text{and} \quad \sigma_\nu \sim \chi^2(\nu) \quad \Longrightarrow \quad \sigma \mathrm{A}\boldsymbol{U} \sim t_\nu(0, \Sigma)$$

As a consequence of the theorem above and the of the example 1, all elliptical random variables X_i can be simply generated by multiplying standardized Gaussian random variables ϵ_i with a common random (strictly positive) factor σ, drawn independently from the ϵ's from an arbitrary distribution: [Embrechts 2002b]

$$\boldsymbol{X} = \boldsymbol{\mu} + \sigma\,\boldsymbol{\epsilon} \tag{2.25}$$

where μ is a location parameter, and $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \Sigma)$.

Property (stability). Let $\boldsymbol{X} \sim E_N(\boldsymbol{\mu}_X, \Sigma, \phi_X)$ and $\boldsymbol{Y} \sim E_N(\boldsymbol{\mu}_Y, c\Sigma, \phi_Y)$ be independent random vectors. Their sum is elliptical, too:

$$a\boldsymbol{X} + b\boldsymbol{Y} \sim E_N(a\boldsymbol{\mu}_X + b\boldsymbol{\mu}_Y, \Sigma, \phi) \quad \text{with} \quad \phi(\cdot) = \phi_X(a^2\cdot)\phi_Y(b^2 c\cdot)$$

As an example, consider the difference of two independent and elliptical random vectors: $\boldsymbol{X} - \boldsymbol{X}' \sim E_N(\boldsymbol{0}, \Sigma, \phi^2)$ Remark that the stability also holds whenever the random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ are dependent only through their radial part σ, see [Hult 2002].

Property (bivariate marginals). Let $\boldsymbol{X} \sim E_N(\boldsymbol{\mu}, \Sigma, \phi)$. Then, for all pair (i, j),

$$\boldsymbol{X_{ij}} \sim E_2(\boldsymbol{\mu_{ij}}, \Sigma_{(ij)}, \phi), \quad \text{with} \quad \Sigma_{(ij)} = \begin{pmatrix} \Sigma_{ii} & \Sigma_{ij} \\ \Sigma_{ij} & \Sigma_{jj} \end{pmatrix}$$

In words, the bivariate marginals of a multivariate elliptical random variable are also elliptical, and inherit the parameters of the joint distribution.

This is clear from either representations (2.24) or (2.25). This property has an important converse: if the bivariate marginals are not elliptical with the same radial characteristic function, then the joint probability density is not elliptical. We will make use of this corollary in our empirical study of multivariate distribution for stock returns in Chapter 4 of Part II.

2.4.2   Predicted dependence coefficients

The measures of dependence defined in Section 2.1 can be explicitly computed for elliptical models. Let $\boldsymbol{X} \sim E_N(\boldsymbol{\mu}, \Sigma, \phi)$ and $r_{ij} = \Sigma_{ij}/\sqrt{\Sigma_{ii}\Sigma_{jj}}$. Then

$$\rho_{ij}^{(1)} = r_{ij} \tag{2.26a}$$

$$\zeta_{ij}^{(1)} = \frac{f^{(1)} \cdot D(r_{ij}) - 1}{\frac{\pi}{2} f^{(1)} - 1} \tag{2.26b}$$

$$\zeta_{ij}^{(2)} = \frac{f^{(2)} \cdot (1 + 2r_{ij}^2) - 1}{3f^{(2)} - 1} \tag{2.26c}$$

$$\beta_{ij}^{(B)} = \rho_{ij}^{(0)} = \frac{2}{\pi} \arcsin r_{ij} \tag{2.26d}$$

$$\tau_{ij}^{(K)} = \frac{2}{\pi} \arcsin r_{ij} \qquad \text{proof in [Lindskog 2001]}, \tag{2.26e}$$

where $f^{(d)} = \mathbb{E}[\sigma^{2d}]/\mathbb{E}[\sigma^d]^2$ and $D(r) = \sqrt{1 - r^2} + r \arcsin r$. Some remarks regarding these equations are of importance:

- $\rho_{ij}^{(1)}$, $\beta_{ij}^{(B)}$ and $\tau_{ij}^{(K)}$ do not depend on ϕ (i.e. on σ): they are invariant in the class of elliptical distributions with continuous marginals and given dispersion matrix Σ.

- No such invariance exists for Spearman's rho $\rho^{(S)}$ ([Hult 2002]) and for $\rho^{(d)}$ and $\zeta^{(1)}$!

- For conclusions (2.26e) and (2.26d), $\boldsymbol{X} \sim E_N(\boldsymbol{\mu}, \Sigma, \phi)$ is a sufficient condition but not a necessary one ! Every random vector $\boldsymbol{X}$ with elliptical *copula* (see below) verifies (2.26e) and (2.26d), and is hence said to follow a multivariate meta-elliptical distribution.

- Relations (2.26e) and (2.26d) can be used to define alternative ways of measuring the correlation matrix elements r_{ij}. Indeed,

$$\rho^{(K)} \equiv \sin\left(\tfrac{\pi}{2}\tau^{(K)}\right) \qquad \text{and} \qquad \rho^{(B)} \equiv \sin\left(\tfrac{\pi}{2}\beta^{(B)}\right)$$

provide very low-moments (though slightly negatively biased) estimators — see [Lindskog 2001] for a study of $\rho^{(K)}$. Said differently, it is always possible to define the effective correlation $\rho^{(B)}$ which, according to Eq. (2.18) and the subsequent discussion on page 21, can also be written as

$$\rho^{(B)} = \cos\left(2\pi C(\tfrac{1}{2}, \tfrac{1}{2})\right). \tag{2.27}$$

Now Eq. (2.26d) states together with Eq. (2.26a) that for all elliptical models, this effective correlation is just equal to the usual linear correlation $\rho^{(1)}$! This property provides a very convenient and simple testable prediction to check whether a given copula is compatible with an elliptical model or not, and will be at the heart of our empirical study in Chapters 4 and 5 of part II.

- $f^{(2)}$ is related to the kurtosis of the X's through the relation $\kappa = 3(f^{(2)} - 1)$.

The calculation of the tail correlation coefficients depends on the specific form of the function ϕ or, said differently, on the distribution $\mathcal{P}_\sigma$ of σ in Eq. (2.25), for which several choices are possible. We will focus in the following on two of them, corresponding to the Student model and the log-normal model.

The Gaussian ensemble

In the case of bivariate Gaussian variables of variance σ_0^2, i.e. when

$$\mathcal{P}_\sigma(\sigma) = \delta(\sigma - \sigma_0),$$

all $f^{(d)} = 1$ and all the coefficients of dependence depend on ρ only ! For example, the quadratic correlation $\zeta^{(2)}$ is given by:

$$\zeta^{(2)} = \rho^2; \tag{2.28}$$

whereas the correlation of absolute values $\zeta^{(1)}$ is given by:

$$\zeta^{(1)} = \frac{D(\rho) - 1}{\frac{\pi}{2} - 1}, \tag{2.29}$$

For some other classes of distributions, the higher-order coefficients $\rho^{(d)}$ and $\zeta^{(d)}$ are explicit functions of the coefficient of linear correlation. This is for example the case of Student variables (see Fig. 2.2) and more generally for all elliptical distributions.

The Student ensemble

When the distribution of the square volatility is inverse Gamma, i.e. when

$$\mathcal{P}_{\sigma^2}(u) = \frac{1}{\Gamma(\frac{\nu}{2})} u^{-\frac{\nu}{2}-1} e^{-\frac{1}{u}},$$

the joint pdf of the returns turns out to have an explicit form [Demarta 2005, Embrechts 2002b, Malevergne 2006]:

$$t_\nu(\boldsymbol{x}) = \frac{1}{\sqrt{(\nu\pi)^N \det\Sigma}} \frac{\Gamma(\frac{\nu+N}{2})}{\Gamma(\frac{\nu}{2})} \left(1 + \frac{\boldsymbol{x}^\dagger \Sigma^{-1} \boldsymbol{x}}{\nu}\right)^{-\frac{\nu+N}{2}} \tag{2.30}$$

This is the multivariate Student distribution with ν degrees of freedom for N random variables with dispersion matrix Σ. Clearly, the marginal distribution of $t_\nu(\boldsymbol{x})$ is itself a Student distribution with a tail exponent equal to ν, which is well known to describe satisfactorily the univariate pdf of high-frequency returns (from a few minutes to a day or so), with an exponent in the range $[3, 5]$ (see e.g. Refs. [Bouchaud 2003, Cont 2001, Fuentes 2009, Plerou 1999]). The multivariate Student model is therefore a rather natural choice; its corresponding copula defines the Student copula. For $N = 2$, it is entirely parameterized by ν and the correlation coefficient $\rho = \Sigma_{12}$ ($\Sigma_{11} = \Sigma_{22} = 1$).

The moments of σ are easily computed and lead to the following expressions for the coefficients $f^{(d)}$:

$$f^{(1)} = \frac{2}{\nu-2} \left(\frac{\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu-1}{2})}\right)^2 \qquad\qquad f^{(2)} = \frac{\nu-2}{\nu-4} \tag{2.31a}$$

when $\nu > 2$ (resp. $\nu > 4$). Note that in the limit $\nu \to \infty$ at fixed N, the multivariate Student distribution boils down to a multivariate Gaussian. The shape of $\zeta^{(1)}(\rho)$ and $\zeta^{(2)}(\rho)$ for $\nu = 5$ is given in Fig. 2.2.

One can explicitly compute the coefficient of tail dependence for Student variables, which only depends on ν and on the linear correlation coefficient ρ. By symmetry, one has $\tau^{\mathrm{UU}}(p;\nu,\rho) = \tau^{\mathrm{LL}}(p;\nu,\rho) = \tau^{\mathrm{UL}}(p;\nu,-\rho) = \tau^{\mathrm{LU}}(p;\nu,-\rho)$. When $p = 1 - \epsilon$ with $\epsilon \to 0$, the result is given by the following theorem:[1]

Theorem ([Chicheportiche 2012b]). Let (X_1, X_2) follow a bivariate student distribution with ν degrees of freedom and correlation ρ, and denote by $T_\nu(x)$ the univariate Student cdf with ν degrees of freedom. The pre-asymptotic behavior of its tail dependence when $p \to 1$ is approximated by the following expansion in (rational) powers of $(1-p)$:

$$\tau^{\mathrm{UU}}(p;\nu,\rho) = \tau^{\mathrm{UU}}_*(\nu,\rho) + \beta(\nu,\rho) \cdot \epsilon^{\frac{2}{\nu}} + \mathcal{O}(\epsilon^{\frac{4}{\nu}}), \tag{2.32}$$

---

[1] This result was simultaneously found by Manner and Segers [Manner 2010] (in a somewhat more general context). We still sketch our proof because it follows a different route (uses the copula), and the final expression looks quite different, although of course numerically identical.
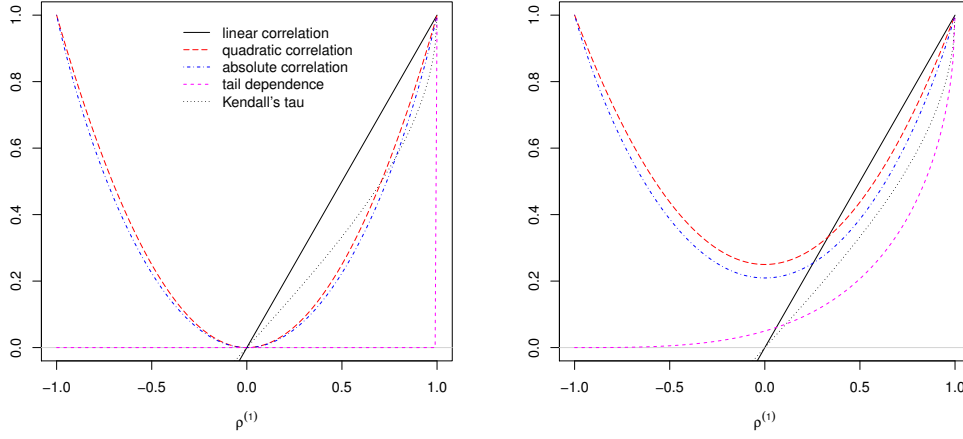
Figure 2.2: For all elliptical distributions, the different measures of dependence are functions of ρ. Illustration for a pair with a bivariate Gaussian distribution (left), and a bivariate Student distribution with $\nu = 5$ (right).

with

$$\tau_*^{\mathrm{UU}}(\nu, \rho) = 2 - 2\, T_{\nu+1}(k(1))$$

$$\beta(\nu, \rho) = \frac{\nu^{\frac{2}{\nu}+1}}{\frac{2}{\nu}+1}\, k(1)\, t_{\nu+1}\big(k(1)\big)\, L_\nu^{-\frac{2}{\nu}}$$

where $k(1) = \sqrt{(\nu+1)(1-\rho)}/\sqrt{1+\rho}$ and $L_\nu = \pi^{-\frac{1}{2}}\nu^{\frac{\nu}{2}}\Gamma(\frac{\nu+1}{2})/\Gamma(\frac{\nu}{2})$.

The following Lemma, is needed for the proof of the Theorem:

Lemma. Let (X_1, X_2) and $T_\nu(x)$ be like in the Theorem above, and $x_p \equiv T_\nu^{-1}(p)$. Define

$$K = \sqrt{\frac{\nu+1}{1-\rho^2}}\, \frac{X_1 - \rho x_p}{\sqrt{\nu + X_2^2}}$$

$$k(p) = \frac{\sqrt{\nu+1}\sqrt{1-\rho}}{\sqrt{1+\rho}}\, \big[\nu \cdot x_p^{-2} + 1\big]^{-\frac{1}{2}} = \frac{k(1)}{\sqrt{1 + \frac{\nu}{x_p^2}}}.$$

Then,

$$\mathbb{P}[K \leq k(p) \mid X_2 = x_p] = T_{\nu+1}\big(k(p)\big) \tag{2.33}$$

Proof of the Lemma. The proof proceeds straightforwardly by showing that $t_{\nu+1}(k) = \mathcal{P}_{X_1|X_2}(x_1|x_2)\frac{\partial x_1}{\partial k}$ when $x_2 = x_p$. The particular result for the limit case $p = 1$ is stated in [Embrechts 2002b]. $\square$

Proof of the Theorem. Recall from Eq. (2.15) that

$$\tau^{\mathrm{UU}}(p) = 2 - \frac{C(p,p) - C(1,1)}{p-1} = 2 - \frac{1}{1-p}\int_p^1 \frac{\mathrm{d}C(p,p)}{\mathrm{d}p}\, \mathrm{d}p \tag{2.34}$$

One easily shows that

$$\mathrm{d}C(p,p) = 2\,\mathbb{P}[X_1 \le x_p \mid X_2 = x_p]\,\mathrm{d}p = 2\,T_{\nu+1}\big(k(p)\big)\mathrm{d}p$$

where the second equality holds in virtue of the aforementioned lemma.
Now, for p close to 1, $k(p) = k(1)\left(1 - \frac{1}{2}\frac{\nu}{x_p^2}\right) + \mathcal{O}(x_p^{-4})$, and

$$\frac{\mathrm{d}C(p,p)}{\mathrm{d}p} = 2\,T_{\nu+1}\big(k(p)\big) \approx 2\,T_{\nu+1}\big(k(1)\big) - 2\big(k(1) - k(p)\big)\cdot t_{\nu+1}\big(k(1)\big)$$

But since the Student distribution behaves as a power-law precisely in the region $p \lesssim 1$, we write $t_\nu(x) \approx L_\nu/x^{\nu+1}$ and immediately get $1-p = x_p^{-\nu}L_\nu/\nu$. The result follows by collecting all the terms and performing the integration in (2.34). $\qquad\square$

The notable features of the above results are:

- For large ν, the exponent is almost zero and the correction term is of order $\mathcal{O}(1)$, and the expansion ceases to hold. This is particularly true for the Gaussian distribution ($\nu = \infty$) for which the behavior is radically different at the limit $p \to 1$ (where there's strictly no tail correlation) and at $p < 1$ where a dependence subsists, see Fig. 2.3(a).

- The asymptotic tail dependence $\tau_*^{\mathrm{UU}}(\nu,\rho)$ is strictly positive for all $\rho > -1$ and finite ν, and tends to zero in the Gaussian limit $\nu \to \infty$ (see Fig. 2.3(a)). The intuitive interpretation is quite clear: large events are caused by large occasional bursts in volatility. Since the volatility is common to all assets, the fact that one return is positive extremely large is enough to infer that the volatility is itself large. Therefore that there is a non-zero probability $\tau_*^{\mathrm{UU}}(\nu,\rho)$ that another asset also has a large positive return (except if $\rho = -1$ since in that case the return can only be large and negative!). It is useful to note that the value of $\tau_*^{\mathrm{UU}}(\nu,\rho)$ does not depend on the explicit shape of $\mathcal{P}_\sigma$ provided the asymptotic tail of $\mathcal{P}_\sigma$ decays as $L(\sigma)/\sigma^{1+\nu}$, where $L(\sigma)$ is a slow function.

- The coefficient $\beta(\nu,\rho)$ is also positive, indicating that estimates of $\tau_*^{\mathrm{UU}}(\nu,\rho)$ based on measures of $\tau^{\mathrm{UU}}(p;\nu,\rho)$ at finite quantiles (e.g. $p = 0.99$) are biased upwards. Note that the correction term is rapidly large because ϵ is raised to the power $2/\nu$. For example, when $\nu = 4$ and $\rho = 0.3$, $\beta \approx 0.263$ and one expects a first-order correction 0.026 for $p = 0.99$. This is illustrated in Figs. 2.3(a), 2.3(b). The form of the correction term (in $\epsilon^{2/\nu}$) is again valid as soon as $\mathcal{P}_\sigma$ decays asymptotically as $L(\sigma)/\sigma^{1+\nu}$.

- Not only is the correction large, but the accuracy of the first order expansion is very bad, since the ratio of the neglected term to the first correction is itself $\sim \epsilon^{2/\nu}$. The region where the first term is accurate is therefore exponentially small in ν — see Fig. 2.3(b).

Finally, we plot in Fig. 2.4 the rescaled difference between the Student copula and the Gaussian copula both on the diagonal and on the anti-diagonal, for several values of the linear correlation coefficient ρ and for $\nu = 5$. One notices that the difference is zero for $p = \frac{1}{2}$, as expected from the expression of $C^*(\rho)$ for general elliptical models. Away from $p = \frac{1}{2}$ on the diagonal, the rescaled difference has a positive convexity and non-zero limits when $p \to 0$ and $p \to 1$, corresponding to τ_*^{LL} and τ_*^{UU}.
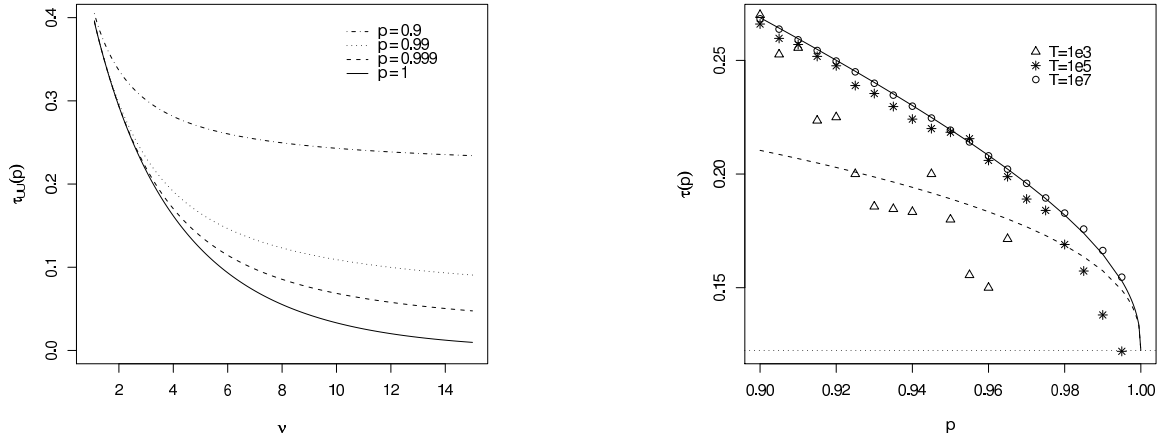
(a) τ^{UU} vs ν at several thresholds p for bivariate Student variables with $\rho = 0.3$. Note that $\tau_*^{\mathrm{UU}} \to 0$ when $\nu \to \infty$, but rapidly grows when $p \neq 1$.

(b) τ^{UU} vs p for bivariate Student variables with $\rho = 0.3$ and $\nu = 5$: exact curve (plain), first order power-law expansion (dashed) and simulated series with different lengths T (symbols)

Figure 2.3:

The log-normal ensemble

If we now choose $\sigma = \sigma_0 \, \mathrm{e}^\xi$ with $\xi \sim \mathcal{N}(0, s)$, the resulting multivariate model defines the log-normal ensemble and the log-normal copula. The factors $f^{(d)}$ are immediately found to be: $f^{(d)} = \mathrm{e}^{d^2 s^2}$ with no restrictions on d. The Gaussian case now corresponds to the limit $s \equiv 0$.

Although the inverse Gamma and the log-normal distributions are very different, it is well known that the tail region of the log-normal is in practice very hard to distinguish from a power-law. In fact, one may rewrite the pdf of the log-normal distribution as:

$$\widehat{\sigma}^{-1} \mathrm{e}^{-\frac{\ln^2 \widehat{\sigma}}{2s^2}} = \widehat{\sigma}^{-(1+\nu_{\mathrm{eff}})} \quad \text{with} \quad \nu_{\mathrm{eff}} = \frac{\ln \widehat{\sigma}}{2s^2}, \tag{2.35}$$

where we have introduced $\widehat{\sigma} = \sigma/\sigma_0$. The above equation means that for large $\widehat{\sigma}$ a log-normal is like a power-law with a slowly varying exponent. If the tail region corresponds to $\widehat{\sigma} \in [4, 8]$ (say), the effective exponent ν lies in the range $[0.7/s^2, 1.05/s^2]$. Another way to obtain an approximate dictionary between the Student model and the log-normal model is to fix $\nu_{\mathrm{eff}}(s)$ such that the coefficient $f^{(2)}$ (say) of the two models coincide. Numerically, this leads to $\nu_{\mathrm{eff}}(s) \approx 2 + 0.5/s^2$, leading to $s \approx 0.4$ for $\nu = 5$. In any case, our main point here is that from an empirical point of view, one does not expect striking differences between a Student model and a log-normal model — even though from a strict mathematical point of view, the models are very different. In particular, the asymptotic tail dependence coefficients τ_*^{UU} or τ_*^{LL} are always zero for the log-normal model (but $\tau^{\mathrm{LL}}(p)$ or $\tau^{\mathrm{UU}}(p)$ converge exceedingly slowly towards zero as $p \to 1$).
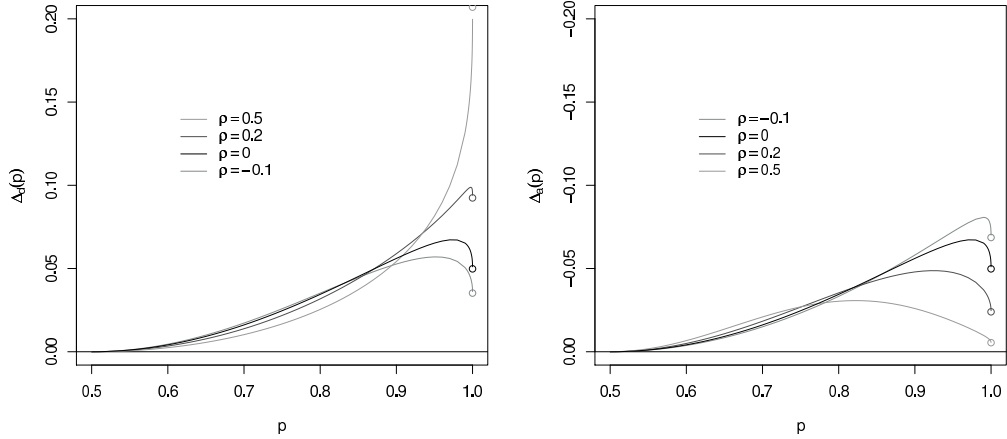
Figure 2.4: Diagonal and anti-diagonal of the Student copula: the quantities $\Delta_{\mathrm{d}}(p)$ and $\Delta_{\mathrm{a}}(p)$ defined in Eq. (2.17) are plotted versus p for several values of ρ and fixed $\nu = 5$. (The curves are identical in the range $p \in [0, 0.5]$ due to the symmetry $p \leftrightarrow 1 - p$).

2.4.3   Pseudo-elliptical generalization

In the previous elliptical description of the returns, all stocks are subject to the exact same stochastic volatility, what leads to non-linear dependences like tail effects and residual higher-order correlations even for $\rho = 0$ (see Fig. 2.2). In order to be able to fine-tune somewhat this dependence due to the common volatility, a simple generalization is to let each stock be influenced by an idiosyncratic volatility, thus allowing for a more subtle structure of dependence. More specifically, we write[2]:

$$X_i = \sigma_i \cdot \epsilon_i \qquad i = 1 \ldots N \tag{2.36}$$

where the Gaussian residuals ϵ_i have the same joint distribution as before (in particular with linear correlations $\mathbb{E}[\epsilon_i \epsilon_j] = r_{ij}$), and are independent of the σ_i, but we now generalize the definition of the ratios $f^{(d)}$. The random vector $\boldsymbol{\sigma}$ has a non trivial dependence structure which is partly described by the mixed d-moments

$$f_{ij}^{(d)} = \frac{\mathbb{E}[\sigma_i^d \sigma_j^d]}{\mathbb{E}[\sigma_i^d] \, \mathbb{E}[\sigma_j^d]}. \tag{2.37}$$

When the σ_i's are dependent but identically distributed, the diagonal elements $f_{ii}^{(d)}$ do not depend on i. Within this setting, the generalization of coefficients (2.26b) can be straightfor-

---

[2] In vectorial form, we collect the individual (yet dependent) stochastic volatilities σ_i on the diagonal of a matrix D, and $\boldsymbol{X} = \mathrm{D}\boldsymbol{\epsilon}$ can be decomposed as $\mathrm{X} = R\mathrm{DA}\boldsymbol{U}$ where $\boldsymbol{U}$ is a random vector uniformly distributed on the unit hypersphere, the radial component R is a chi-2 random variable independent of U with rank(A) degrees of freedom, and A is a matrix with appropriate dimensions such that $\mathrm{AA}^\dagger = \Sigma$. This description can be contrasted with the one proposed in Ref. [Kring 2009] under the term "Multi-tail Generalized Elliptical Distribution": $\boldsymbol{X} = R(\boldsymbol{u})\mathrm{A}\boldsymbol{U}$ with R now depending on the unit vector $\boldsymbol{u} = \mathrm{A}\boldsymbol{U}/||\mathrm{A}\boldsymbol{U}||$. In other words, the description in (2.36) provides a different radial amplitude for each component, whereas [Kring 2009] characterizes a direction-dependent radial part identical for every component. The latter allows for a richer phenomenology than the former, but lacks financial intuition.

wardly calculated:

$$\rho_{\ddot{y}}^{(1)} = r_{\ddot{y}}\frac{f_{\ddot{y}}^{(1)}}{f_{\ddot{u}}^{(1)}} \tag{2.38a}$$

$$\zeta_{\ddot{y}}^{(1)} = \frac{f_{\ddot{y}}^{(1)} \cdot D(r_{\ddot{y}}) - 1}{\frac{\pi}{2}f_{\ddot{u}}^{(1)} - 1} \tag{2.38b}$$

$$\zeta_{\ddot{y}}^{(2)} = \frac{f_{\ddot{y}}^{(2)} \cdot (1 + 2r_{\ddot{y}}^2) - 1}{3\,f_{\ddot{u}}^{(2)} - 1} \tag{2.38c}$$

Importantly, Kendall's tau and Blomqvist's beta remain invariant in this class of Pseudo-elliptical distributions !

As an explicit example, we consider the natural generalization of the log-normal model and write $\sigma_i = \sigma_{0i}\,\mathrm{e}^{\xi_i}$, with[3] $\xi_i \overset{\mathrm{id}}{\sim} \mathcal{N}(0, s)$, and some correlation structure of the ξ's: $\mathbb{E}[\xi_i\xi_j] = s^2 c_{ij}$. One then finds:

$$f_{ij}^{(d)} \equiv f^{(d)}(c_{ij}) = \mathrm{e}^{d^2 s^2 c_{ij}}. \tag{2.39}$$

Using the generic notation c for the correlation of the log-volatilities, and r for the correlation of the residuals (now different from ρ) we find:

$$\rho(r, c) = \frac{f^{(1)}(c)}{f^{(1)}(1)} \cdot r \tag{2.40a}$$

$$\zeta^{(2)}(r, c) = \frac{f^{(2)}(c)\,(1 + 2r^2) - 1}{3f^{(2)}(1) - 1} \tag{2.40b}$$

$$\zeta^{(1)}(r, c) = \frac{f^{(1)}(c)\,D(r) - 1}{\frac{\pi}{2}f^{(1)}(1) - 1} \tag{2.40c}$$

When c is fixed (e.g. for the elliptical case $c = 1$), ρ and r are proportional, and all measures of non-linear dependences can be expressed as a function of ρ. But this ceases to be true as soon as there exists some non trivial structure c in the volatilities. In that case, c and r are "hidden" underlying variables, that can only be reconstructed from the knowledge of ρ, $\zeta^{(2)}$, $C(\frac{1}{2}, \frac{1}{2})$, assuming of course that the model is accurate.

Notice that the result on $C(\frac{1}{2}, \frac{1}{2})$ is totally independent of the structure of the volatilities[4]. Indeed, what is relevant for the copula at the central point is not the amplitude of the returns, but rather the number of $+/-$ events, which is unaffected by any multiplicative scale factor as long as the median of the univariate marginals is nil. An important consequence of this result is that for all elliptical or pseudo-elliptical model, $\rho = 0$ implies that $C^*(\rho = 0) = \frac{1}{4}$.

2.5   Conclusion

We have suggested several ideas of methodological interest to efficiently visualize and compare different copulas. We recommend in particular the rescaled difference with the Gaussian copula along the diagonal $\Delta_{\mathrm{d,a}}(p)$ and the central value of the copula $C(\frac{1}{2}, \frac{1}{2})$ as strongly

---

[3] A further generalization that allows for stock dependent "vol of vol" s_i is also possible.

[4] This property holds even when σ_i depends on the sign of ϵ_i, which might be useful to model the leverage effect that leads to some asymmetry between positive and negative tails, as the data suggests.

discriminating observables. We have studied the dependence of these quantities, as well as other non-linear correlation coefficients like the higher-order correlations $\zeta^{(d)}$, with the linear correlation coefficient ρ.

The case of elliptical copulas is emphasized, with the lognormal and Student models as examples. We provide an original characterization of the pre-asymptotic tail dependence coefficient when $p \to 1$ as a development in rational powers of $(1-p)$.

An explicit prediction of elliptical models is empirically tested on multivariate financial data in Chapter 4 of part II.

Goodness-of-fit testing

Contents

The problem of testing whether a null-hypothesis theoretical probability distribution is compatible with the empirical probability distribution of a sample of observations is known as goodness-of-fit (GoF) testing and is ubiquitous in all fields of science and engineering. Goodness-of-Fit tests are designed to assess quantitatively whether a sample of N observations can statistically be seen as a collection of N realizations of a given probability law, or whether two such samples are drawn from the same hypothetical distribution.

The best known theoretical result is due to Kolmogorov and Smirnov (KS) [Kolmogorov 1933, Smirnov 1948], and has led to the eponymous statistical test for an *univariate* sample of *independent* draws. The major strength of this test lies in the fact that the asymptotic distribution of its test statistic is completely independent of the null-hypothesis cdf.

Several specific extensions have been studied (and/or are still under scrutiny), including: different choices of distance measures [Darling 1957], multivariate samples [Fasano 1987, Cabaña 1994, Cabaña 1997, Fermanian 2005a], investigation of different parts of the distribution domain [Anderson 1952, Deheuvels 2005, Noé 1968, Chicheportiche 2012c], dependence in the successive draws [Chicheportiche 2011], etc.

This class of problems has a particular appeal for physicists since the works of Doob [Doob 1949] and Khmaladze [Khmaladze 1982], who showed how GoF testing is related to stochastic processes. Finding the law of a test amounts to computing a survival probability in a dissipative system. In a Markovian setting, this is often achieved by treating a Fokker-Planck problem, which in turn maps into a Schrödinger equation for a particle in a certain potential confined by walls.

3.0.1 Empirical cumulative distribution and its fluctuations

Let $\boldsymbol{X}$ be a random vector of N independent and identically distributed variables, with marginal cumulative distribution function (cdf) F. One realization of $\boldsymbol{X}$ consists of a time series $\{x_1, \ldots, x_n, \ldots, x_N\}$ that exhibits no persistence (see Sect. 3.2 when some non trivial dependence is present). The empirical cumulative distribution function

$$F_N(x) = \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{\{X_n \le x\}} \tag{3.1}$$

converges to the true CDF F as the sample size N tends to infinity. For finite N, the expected value and fluctuations of $F_N(x)$ are

$$\mathbb{E}[F_N(x)] = F(x),$$

$$\text{Cov}(F_N(x), F_N(x')) = \frac{1}{N} \left[F(\min(x,x')) - F(x)F(x') \right].$$

The rescaled empirical CDF

$$Y_N(u) = \sqrt{N} \left[F_N(F^{-1}(u)) - u \right] \tag{3.2}$$

measures, for a given $u \in [0,1]$, the difference between the empirically determined cdf of the X's and the theoretical one, evaluated at the u-th quantile. It does not shrink to zero as $N \to \infty$, and is therefore the quantity on which any statistic for GoF testing is built.

Limit properties

One now defines the process $Y(u)$ as the limit of $Y_N(u)$ when $N \to \infty$. According to the Central Limit Theorem (CLT), it is Gaussian and its covariance function is given by:

$$I(u, v) = \min(u, v) - uv, \tag{3.3}$$

which characterizes the so-called Brownian bridge, i.e. a Brownian motion $Y(u)$ such that $Y(u{=}0) = Y(u{=}1) = 0$.

Interestingly, F does not appear in Eq. (3.3) anymore, so the law of any functional of the limit process Y is independent of the law of the underlying finite size sample. This property is important for the design of *universal* GoF tests.

Norms over processes

In order to measure a limit distance between distributions, a norm $||.||$ over the space of continuous bridges needs to be chosen. Typical such norms are the norm-2 (or 'Cramér-von Mises' distance)

$$||Y||_2 = \int_0^1 Y(u)^2 \mathrm{d}u,$$

as the bridge is always integrable, or the norm-sup (also called the Kolmogorov distance)

$$||Y||_\infty = \sup_{u \in [0,1]} |Y(u)|,$$

as the bridge always reaches an extremal value.

Unfortunately, both these norms mechanically overweight the core values $u \approx 1/2$ and disfavor the tails $u \approx 0, 1$: since the variance of $Y(u)$ is zero at both extremes and maximal in the central value, the major contribution to $||Y||$ indeed comes from the central region. To alleviate this effect, in particular when the GoF test is intended to investigate a specific region of the domain, it is preferable to introduce additional weights and study $||Y\sqrt{\psi}||$ rather than $||Y||$ itself. Anderson and Darling show in Ref. [Anderson 1952] that the solution to the problem with the Cramér-von Mises norm and arbitrary weights ψ is obtained by spectral decomposition of the covariance kernel, and use of Mercer's theorem. They design an eponymous test [Darling 1957] with the specific choice of weights $\psi(u) = 1/I(u, u)$ equal to the inverse variance, which equi-weights all quantiles of the distribution to be tested. We analyze here the case of the same weights, but with the Kolmogorov distance.

3.1   Weighted Kolmogorov-Smirnov tests

So again $Y(u)$ is a Brownian bridge, i.e. a centered Gaussian process on $u \in [0,1]$ with covariance function $I(u,v)$ given in Eq. (3.3). In particular, $Y(0) = Y(1) = 0$ with probability equal to 1, no matter how distant F is from the sample cdf around the core values. In order to put more emphasis on specific regions of the domain, we weight the Brownian bridge as follows: for given $a \in]0,1[$ and $b \in [a,1[$, we define

$$\tilde{Y}(u) = Y(u) \cdot \begin{cases} \sqrt{\psi(u)} & , a \leq u \leq b \\ 0 & , \text{otherwise.} \end{cases} \tag{3.4}$$

We will characterize the law of the supremum $K(a,b) \equiv \sup_{u \in [a,b]} \left| \tilde{Y}(u) \right|$:

$$\mathcal{P}_<(k|a,b) \equiv \mathbb{P}[K(a,b) \leq k] = \mathbb{P}\left[|\tilde{Y}(u)| \leq k, \forall u \in [a,b] \right].$$

3.1.1   The equi-weighted Brownian bridge: Kolmogorov-Smirnov

In the case of a constant weight, corresponding to the classical KS test, the probability $\mathcal{P}_<(k;0,1)$ is well defined and has the well known KS form [Kolmogorov 1933]:

$$\mathcal{P}_<(k;0,1) = 1 - 2 \sum_{n=1}^{\infty} (-1)^{n-1} e^{-2n^2 k^2}, \tag{3.5}$$

which, as expected, grows from 0 to 1 as k increases, see Fig. 3.5(a) on page 48. The value k^* such that this probability is 95% is $k^* \approx 1.358$ [Smirnov 1948]. This can be interpreted as follows: if, for a data set of size N, the maximum value of $|Y_N(u)|$ is larger than ≈ 1.358, then the hypothesis that the proposed distribution is a "good fit" can be rejected with 95% confidence.

Diffusion in a cage with fixed walls

The Brownian bridge Y is nothing else than a Brownian motion with imposed terminal condition, and can be written as $Y(u) = X(u) - u X(1)$ where X is a Brownian motion. The test law $\mathcal{P}_<(k|a,b) = \mathbb{P}[-k \leq Y(u) \leq k, \forall u \in [a,b]]$ is in fact the survival probability of Y in a cage with absorbing walls, and can be determined by counting the number of Brownian paths $X(u)$ that go from 0 to 0 without ever hitting the barriers. More precisely, the survival probability of the Brownian bridge in the stripe $[-k,k]$ can be computed as $f_1(0;k)/f_1(0;\infty)$, where $f_u(y;k)$ is the transition kernel of the Brownian motion within the allowed region, and it satisfies the simple Fokker-Planck equation

$$\begin{cases} \partial_u f_u(y;k) & = \dfrac{1}{2} \partial_y^2 f_u(y;k) \\ f_u(\pm k;k) & = 0 \end{cases} , \quad \forall u \in [0,1].$$

By spectral decomposition of the Laplace operator, the solution is found to be

$$f_u(y;k) = \frac{1}{k} \sum_{n \in \mathbb{Z}} e^{-E_n u} \cos\left(\sqrt{2E_n}\, y \right), \quad \text{where} \quad E_n = \frac{1}{2} \left(\frac{(2n-1)\pi}{2k} \right)^2$$

and the free propagator in the limit $k \to \infty$ is the usual

$$f_u(y; \infty) = \frac{1}{\sqrt{2\pi u}} \, e^{-\frac{y^2}{2u}},$$

so that the survival probability of the constrained Brownian bridge is

$$\mathcal{P}_<(k; 0, 1) = \frac{f_1(0; k)}{f_1(0; \infty)} = \frac{\sqrt{2\pi}}{k} \sum_{n \in \mathbb{Z}} \exp\left(-\frac{(2n-1)^2 \pi^2}{8k^2}\right). \tag{3.6}$$

Although it looks different from the historical solution Eq. (3.5), the two expressions can be shown to be exactly identical. First write the summand in Eq. (3.6) as the Fourier transform of $e^{-2k^2 x^2}$. Then, adding a small imaginary part to x allows to perform the summation of the series over n, what makes poles $[1 - e^{\pm i\pi x}]^{-1}$ appear in the integrand. Finally use Cauchy's residue theorem to perform the integral, obtaining a sum over all residues evaluated at integer values of x.

The computation of Kolmogorov's distribution $\mathcal{P}_<(k; 0, 1)$ performed above is way easier than the canonical ones [Anderson 1952].

Diffusion in a cage with moving walls

An appropriate change of variable and time

$$W(t) = (1 + t) \, Y\left(\frac{t}{1+t}\right), \quad t \in [0, \infty[$$

leads to the problem of a Brownian diffusion inside a box with walls *moving at constant velocity*. Since the walls expand as $\sim t$ faster than the diffusive particle can move ($\sim \sqrt{t}$), the survival probability converges to a positive value, which turns out to be given by the usual Kolmogorov distribution (3.5) [Krapivsky 1996].

A way of addressing this problem of a diffusing particle in an expanding cage was suggested in Refs. [Bray 2007a, Bray 2007b]. The time-dependent boundary conditions of the usual (forward) Fokker-Planck equation make it difficult to solve as such, and a nice way out is to consider instead the backward Fokker-Planck equation for the transition density, with the initial position of the absorbing wall as an additional independent variable ! This eliminates the time dependence in the boundary condition at the cost of introducing the initial wall parameter.

3.1.2 The variance-weighted Brownian bridge: accounting for the tails

The classical KS test suffers from an important flaw: the test is only weakly sensitive to the quality of the fit in the tails of the tested distribution [Mason 1983, Mason 1992], when it is often these tail events (corresponding to centennial floods, devastating earthquakes, financial crashes, etc.) that one is most concerned with (see, e.g., Ref. [Clauset 2009]).

A simple and elegant GoF test for the tails *only* can be designed starting with digital weights in the form $\psi(u; a) = \mathbb{1}_{\{u \geq a\}}$ or $\psi(u; b) = \mathbb{1}_{\{u \leq b\}}$ for upper and lower tail, respectively.
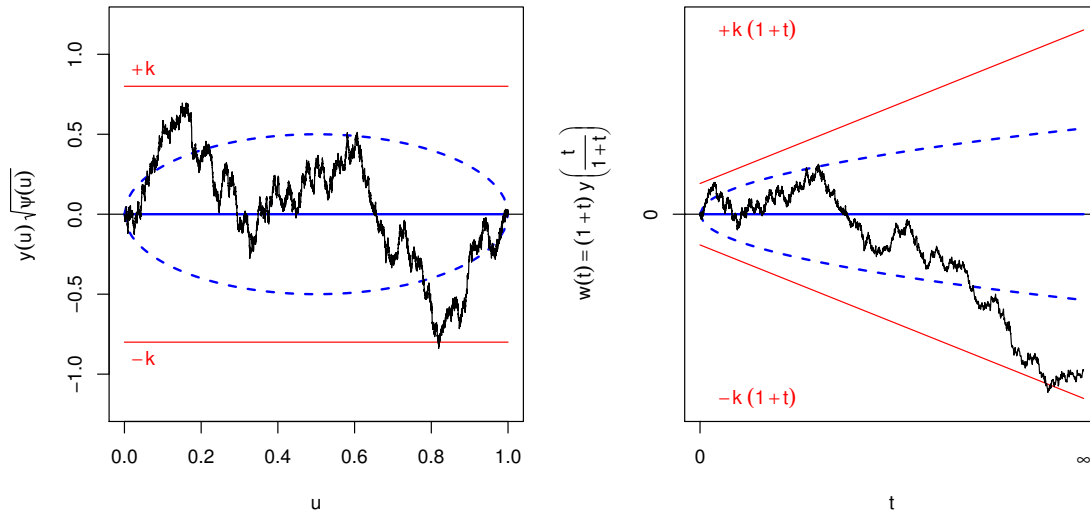
Figure 3.1: The equi-weighted Brownian bridge, $\psi(u) = 1$. The time-changed rescaled process lives in a geometry with boundaries receding at constant speed.

The corresponding test laws can be read off Eq. (5.9) in Ref. [Anderson 1952].[1] Investigation of both tails is attained with $\psi(u; q) = \mathbb{1}_{\{u \leq 1-q\}} + \mathbb{1}_{\{u \geq q\}}$ (where $q > \frac{1}{2}$).

Here we rather focus on a GoF test for a univariate sample of size $N \gg 1$, with the Kolmogorov distance but equi-weighted quantiles, which is equally sensitive to *all regions* of the distribution.[2] We unify two earlier attempts at finding asymptotic solutions, one by Anderson and Darling in 1952 [Anderson 1952] and a more recent, seemingly unrelated one that deals with "life and death of a particle in an expanding cage" by Krapivsky and Redner [Krapivsky 1996, Redner 2001]. We present here the exact asymptotic solution of the corresponding stochastic problem, and deduce from it the precise formulation of the GoF test, which is of a fundamentally different nature than the KS test.

So in order to zoom on the tiny differences in the tails of the Brownian bridge, we weight it as explained in the introduction, with its variance

$$\psi(u) = \frac{1}{u(1-u)}.$$

Solutions for the distributions of such variance-weighted Kolmogorov-Smirnov statistics were studied by Noé, leading to the laws of the one-sided [Noé 1968] and two-sided [Noé 1972] finite sample tests. They were later generalized and tabulated numerically by Niederhausen [Niederhausen 1981a, Niederhausen 1981b, Wilcox 1989]. However, although exact and appropriate for small samples, these solutions rely on recursive relations and are not in closed form. We instead come up with an analytic closed-form solution for large samples that relies on an elegant analogy from statistical physics.

---

[1]The quantity M appearing there is the volume under the normal bivariate surface between specific bounds, and it takes a very convenient form in the unilateral cases $\frac{1}{2} \leq a \leq u \leq 1$ and $0 \leq u \leq b \leq \frac{1}{2}$. Mind the missing j exponentiating the alternating (-1) factor.

[2]Other choices of ψ generally result in much harder problems.

Diffusion in a cage with moving walls

Define the time change $t = \frac{u}{1-u}$. The variable $W(t) = (1+t)\,Y\left(\frac{t}{1+t}\right)$ is then a Brownian motion (Wiener process) on $\left[\frac{a}{1-a}, \frac{b}{1-b}\right]$, since one can check that:

$$\mathrm{Cov}\big(W(t), W(t')\big) = \min(t, t').$$

$\mathcal{P}_<(k|a, b)$ can be now written as

$$\mathcal{P}_<(k|a, b) = \mathbb{P}\left[|W(t)| \le k\sqrt{t}, \forall t \in \left[\tfrac{a}{1-a}, \tfrac{b}{1-b}\right]\right].$$

The problem with initial time $\frac{a}{1-a} = 0$ and horizon time $\frac{b}{1-b} = T$ has been treated by Krapivsky and Redner in Ref. [Krapivsky 1996] as the survival probability $S(T; k = \sqrt{\frac{A}{2D}})$ of a Brownian particle diffusing with constant D in a cage with walls expanding as $\sqrt{At}$. Their result is that for large T,

$$S(T; k) \equiv \mathcal{P}_<(k|0, \tfrac{T}{1+T}) \propto T^{-\theta(k)}.$$

They obtain analytical expressions for $\theta(k)$ in both limits $k \to 0$ and $k \to \infty$. The limit solutions of the very same differential problem were found earlier by Turban for the critical behavior of the directed self-avoiding walk in parabolic geometries [Turban 1992].

We take here a slightly different route, suggested by Anderson and Darling in Ref. [Anderson 1952] but where the authors did not come to a conclusion. Our contributions are: (i) we treat the general case $a > 0$ for *any* k; (ii) we explicitly compute the k-dependence of both the exponent *and* the prefactor of the power-law decay; and (iii) we provide the link with the theory of GoF tests and compute the pre-asymptotic distribution of the weighted Kolmogorov-Smirnov test statistic for large sample sizes $N \to \infty$, i.e. when $]a, b[\to]0, 1[$.

Mean-reversion in a cage with fixed walls

Introducing now the new time change $\tau = \ln\sqrt{\frac{1-a}{a}\,t}$, the variable $Z(\tau) = W(t)/\sqrt{t}$ is a stationary Ornstein-Uhlenbeck (OU) process on $[0, T]$ where

$$T = \ln\sqrt{\frac{b\,(1-a)}{a\,(1-b)}}, \tag{3.7}$$

and

$$\mathrm{Cov}\big(Z(\tau), Z(\tau')\big) = \mathrm{e}^{-|\tau - \tau'|}.$$

Its dynamics is described by the stochastic differential equation (SDE)

$$\mathrm{d}Z(T) = -Z(T)\mathrm{d}T + \sqrt{2}\,\mathrm{d}B(T), \tag{3.8}$$

with $B(T)$ an independent Wiener process. The initial condition for $T = 0$ (corresponding to $b = a$) is $Z(0) = Y(a)/\sqrt{\mathbb{V}[Y(a)]}$, a random Gaussian variable of zero mean and unit
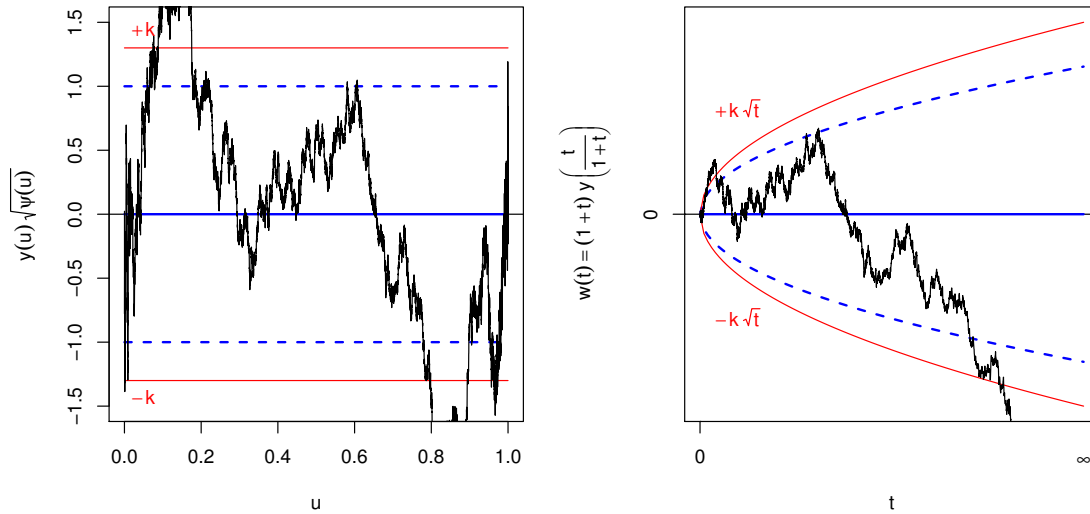
Figure 3.2: The variance-weighted Brownian bridge, $\psi(u) = 1/[u(1-u)]$. The time-changed rescaled process lives in a geometry with boundaries receding as $\sim \sqrt{t}$.

variance. The distribution $\mathcal{P}_<(k|a,b)$ can now be understood as the unconditional survival probability of a mean-reverting particle in a cage with fixed absorbing walls, see Fig. 3.2:

$$\mathcal{P}_<(k|T) = \mathbb{P}[-k \le Z(\tau) \le k, \forall \tau \in [0,T]]$$
$$= \int_{-k}^{k} f_T(z; k) \, \mathrm{d}z,$$

where

$$f_T(z; k) \, \mathrm{d}z = \mathbb{P}\big[Z(T) \in [z, z + \mathrm{d}z[|\{Z(\tau)\}_{\tau < T}\,\big]$$

is the density probability of the particle being at z at time T, when walls are in $\pm k$. Its dependence on k, although not explicit on the right hand side, is due to the boundary condition associated with the absorbing walls (it will be dropped in the following for the sake of readability)[3].

The Fokker-Planck equation governing the evolution of the density $f_T(z)$ reads

$$\partial_\tau f_\tau(z) = \partial_z \left[z \, f_\tau(z)\right] + \partial_z^2 \left[f_\tau(z)\right], \quad 0 < \tau \le T.$$

Calling $\mathcal{H}_{\mathrm{FP}}$ the second order differential operator $-\left[\mathbb{1} + z\partial_z + \partial_z^2\right]$, the full problem thus amounts to finding the general solution of

$$\begin{cases} -\partial_\tau f_\tau(z) &= \mathcal{H}_{\mathrm{FP}}(z) f_\tau(z) \\ f_\tau(\pm k) &= 0, \forall \tau \in [0, T] \end{cases}.$$

We have explicitly introduced a minus sign since we expect that the density decays with time in an absorption problem. Because of the term $z\partial_z$, $\mathcal{H}_{\mathrm{FP}}$ is not hermitian and thus

---

[3]In particular, $\mathcal{P}_<(k|0) = \mathrm{erf}\left(\frac{k}{\sqrt{2}}\right)$.

cannot be diagonalized. However, as is well known, one can define $f_\tau(z) = \mathrm{e}^{-\frac{z^2}{4}} \phi_\tau(z)$ and the Fokker-Planck equation becomes

$$\begin{cases} -\partial_\tau \phi_\tau(z) &= \left[-\partial_z^2 + \frac{1}{4}z^2 - \frac{1}{2}\mathbb{1} \right] \phi_\tau(z) \\ \phi_\tau(\pm k) &= 0, \forall \tau \in [0, T] \end{cases},$$

and its Green's function, i.e. the (separable) solution *conditionally on the initial position* $(z_\mathrm{i}, T_\mathrm{i})$, is the superposition of all modes

$$G_\phi(z, T \mid z_\mathrm{i}, T_\mathrm{i}) = \sum_\nu \mathrm{e}^{-\theta_\nu(T-T_\mathrm{i})} \widehat{\varphi}_\nu(z) \widehat{\varphi}_\nu(z_\mathrm{i}),$$

where $\widehat{\varphi}_\nu$ are the normalized solutions of the stationary Schrödinger equation

$$\begin{cases} \left[-\partial_z^2 + \frac{1}{4}z^2 \right] \varphi_\nu(z) &= \left(\theta_\nu + \frac{1}{2} \right) \varphi_\nu(z) \\ \varphi_\nu(\pm k) &= 0 \end{cases},$$

each decaying with its own energy θ_ν, where ν labels the different solutions with increasing eigenvalues, and the set of eigenfunctions $\{\widehat{\varphi}_\nu\}$ defines an orthonormal basis of the Hilbert space on which $\mathcal{H}_\mathrm{S}(z) = \left[-\partial_z^2 + \frac{1}{4}z^2 \right]$ acts. In particular,

$$\sum_\nu \widehat{\varphi}_\nu(z) \widehat{\varphi}_\nu(z') = \delta(z - z'), \tag{3.9}$$

so that indeed $G_\phi(z, T_\mathrm{i} \mid z_\mathrm{i}, T_\mathrm{i}) = \delta(z - z_\mathrm{i})$, and the general solution writes

$$f_T(z_T; k) = \int_{-k}^{k} \mathrm{e}^{\frac{z_\mathrm{i}^2 - z_T^2}{4}} G_\phi(z_T, T \mid z_\mathrm{i}, T_\mathrm{i}) f_0(z_\mathrm{i}) \, \mathrm{d}z_\mathrm{i},$$

where $T_\mathrm{i} = 0$, which corresponds to the case $b = a$ in Eq. (3.4), and f_0 is the distribution of the initial value z_i which is here, as noted above, Gaussian with unit variance.

$\mathcal{H}_\mathrm{S}$ figures out an harmonic oscillator of mass $\frac{1}{2}$ and frequency $\omega = \frac{1}{\sqrt{2}}$ within an infinitely deep well of width $2k$: its eigenfunctions are parabolic cylinder functions [Mei 1983, Gradshteyn 1980]

$$y_+(\theta; z) = \mathrm{e}^{-\frac{z^2}{4}} {}_1F_1\left(-\frac{\theta}{2}, \frac{1}{2}, \frac{z^2}{2} \right)$$

$$y_-(\theta; z) = z \, \mathrm{e}^{-\frac{z^2}{4}} {}_1F_1\left(\frac{1-\theta}{2}, \frac{3}{2}, \frac{z^2}{2} \right)$$

properly normalized. The only acceptable solutions for a given problem are the linear combinations of y_+ and y_- which satisfy orthonormality (3.9) and the boundary conditions: for periodic boundary conditions, only the integer values of θ would be allowed, whereas with our Dirichlet boundaries $|\widehat{\varphi}_\nu(k)| = -|\widehat{\varphi}_\nu(-k)| = 0$, real non-integer eigenvalues θ are allowed.[4] For instance, the fundamental level $\nu = 0$ is expected to be the symmetric solution $\widehat{\varphi}_0(z) \propto y_+(\theta_0; z)$ with θ_0 the smallest possible value compatible with the boundary condition:

$$\theta_0(k) = \inf_{\theta > 0} \left\{ \theta : y_+(\theta; k) = 0 \right\}. \tag{3.10}$$

---

[4] A similar problem with a *one-sided* barrier leads to a continuous spectrum; this case has been studied originally in Ref. [Mei 1983] and more recently in Ref. [Lladser 2000] (it is shown that there exists a quasi-stationary distribution for any θ) and generalized in Ref. [Aalen 2004].

In what follows, it will be more convenient to make the k-dependence explicit, and a hat will denote the solution with the normalization relevant to our problem, namely $\widehat{\varphi}_0(z; k) = y_+(\theta_0(k); z)/\|y_+\|_k$, with the norm

$$\|y_+\|_k^2 \equiv \int_{-k}^k y_+(\theta_0(k); z)^2 \, \mathrm{d}z,$$

so that $\int_{-k}^k \widehat{\varphi}_\nu(z; k)^2 \, \mathrm{d}z = 1$.

Asymptotic survival rate

Denoting by $\Delta_\nu(k) \equiv [\theta_\nu(k) - \theta_0(k)]$ the gap between the excited levels and the fundamental, the higher energy modes $\widehat{\varphi}_\nu$ cease to contribute to the Green's function when $\Delta_\nu T \gg 1$, and their contributions to the above sum die out exponentially as T grows. Eventually, only the lowest energy mode $\theta_0(k)$ remains, and the solution tends to

$$f_T(z; k) = A(k) \, \mathrm{e}^{-\frac{z^2}{4}} \, \widehat{\varphi}_0(z; k) \, \mathrm{e}^{-\theta_0(k)T},$$

when $T \gg (\Delta_1)^{-1}$, with

$$A(k) = \int_{-k}^k \mathrm{e}^{\frac{z_\mathrm{i}^2}{4}} \widehat{\varphi}_0(z_\mathrm{i}; k) f_0(z_\mathrm{i}) \, \mathrm{d}z_\mathrm{i}. \tag{3.11}$$

Let us come back to the initial problem of the weighted Brownian bridge reaching its extremal value in $[a, b]$. If we are interested in the limit case where a is arbitrarily close to 0 and b close to 1, then $T \to \infty$ and the solution is thus given by

$$\begin{aligned}
\mathcal{P}_<(k|T) &= A(k) \, \mathrm{e}^{-\theta_0(k)T} \int_{-k}^k \mathrm{e}^{-\frac{z^2}{4}} \widehat{\varphi}_0(z; k) \, \mathrm{d}z \\
&= \widetilde{A}(k) \, \mathrm{e}^{-\theta_0(k)T},
\end{aligned}$$

with $\widetilde{A}(k) \equiv \sqrt{2\pi} A(k)^2$.

We now compute explicitly the limit behavior of both $\theta_0(k)$ and $\widetilde{A}(k)$.

$\boxed{k \to \infty}$ As k goes to infinity, the absorption rate $\theta_0(k)$ is expected to converge toward 0: intuitively, an infinitely far barrier will not absorb anything. At the same time, $\mathcal{P}_<(k|T)$ must tend to 1 in that limit. So $\widetilde{A}(k)$ necessarily tends to unity. Indeed,

$$\theta_0(k) \xrightarrow{k \to \infty} \sqrt{\frac{2}{\pi}} k \, \mathrm{e}^{-\frac{k^2}{2}} \to 0, \tag{3.12}$$

$$\widetilde{A}(k) \xrightarrow{k \to \infty} \left(\int_{-\infty}^\infty \widehat{\varphi}_0(z; \infty)^2 \, \mathrm{d}z \right)^2 = 1.$$

In principle, we see from Eq. (3.11) that corrections to the latter arise both (and jointly) from the functional relative difference of the solution $\epsilon(z; k) = y_+(\theta_0(k); z)/y_+(0; z) - 1$, and from the finite integration limits ($\pm k$ instead of $\pm\infty$). However, it turns out that the correction of

the first kind is of second order in ϵ.[5] The correction to $A(k)$ is thus dominated by the finite integration limits $\pm k$, so that

$$\widetilde{A}(k \to \infty) \approx \mathrm{erf}\left(\frac{k}{\sqrt{2}}\right)^2.$$ (3.13)

$\boxed{k \to 0}$ For small k, the system behaves like a free particle in a sharp and infinitely deep well, since the quadratic potential is almost flat around 0. The fundamental mode becomes then

$$\widehat{\varphi}_0(z; k \to 0) = \frac{1}{\sqrt{k}} \cos\left(\frac{\pi z}{2k}\right),$$

and consequently

$$\theta_0(k) \xrightarrow{k \to 0} \frac{\pi^2}{4k^2} - \frac{1}{2},$$ (3.14)

$$\widetilde{A}(k) \xrightarrow{k \to 0} \left(\int_{-k}^{k} \frac{e^{-\frac{z^2}{4}}}{(2\pi)^{\frac{1}{4}}} \frac{1}{\sqrt{k}} \cos\left(\frac{\pi z}{2k}\right) \mathrm{d}z\right)^2$$

$$\approx \frac{1}{\sqrt{2\pi}\, k} \left(\frac{4k}{\pi}\right)^2 = \frac{16}{\pi^2 \sqrt{2\pi}} k.$$ (3.15)

We show in Fig. 3.3 the functions $\theta_0(k)$ and $\widetilde{A}(k)$ computed numerically from the exact solution, together with their asymptotic analytic expressions. In intermediate values of k (roughly between 0.5 and 3) these limit expressions fail to reproduce the exact solution.

Higher modes and validity of the asymptotic ($N \gg 1$) solution

Higher modes $\nu > 0$ with energy gaps $\Delta_\nu \lesssim 1/T$ must in principle be kept in the pre-asymptotic computation. This, however, is irrelevant in practice since the gap $\theta_1 - \theta_0$ is never small. Indeed, $\widehat{\varphi}_1(z; k)$ is proportional to the asymmetric solution $y_-(\theta_1(k); z)$ and its energy

$$\theta_1(k) = \inf_{\theta > \theta_0(k)} \left\{\theta : y_-(\theta; k) = 0\right\}$$

is found numerically to be very close to $1 + 4\theta_0(k)$. In particular, $\Delta_1 > 1$ (as we illustrate in Fig. 3.4) and thus $T\Delta_1 \gg 1$ will always be satisfied in cases of interest.

---

[5] From Eq. (3.11) we have, when $k \to \infty$,

$$A(k) = (2\pi)^{-1/2} \frac{\int_{-k}^{k} e^{-z^2/2}[1 + \epsilon(z; k)]\mathrm{d}z}{\sqrt{\int_{-k}^{k} e^{-z^2/2}[1 + \epsilon(z; k)]^2 \mathrm{d}z}}.$$

The result follows by keeping only the dominant terms in the expansion in powers of $\epsilon(z; k)$. A similar computation for the asymptotic analysis by expanding the wave function in θ was performed in Ref. [Krapivsky 1996]. Alternatively, algebraic arguments allow to understand that, to first order in the energy correction $\theta_0(k) - \theta_0(\infty)$, the perturbation of the wave function is orthogonal to $\widehat{\varphi}_0(z; \infty)$.
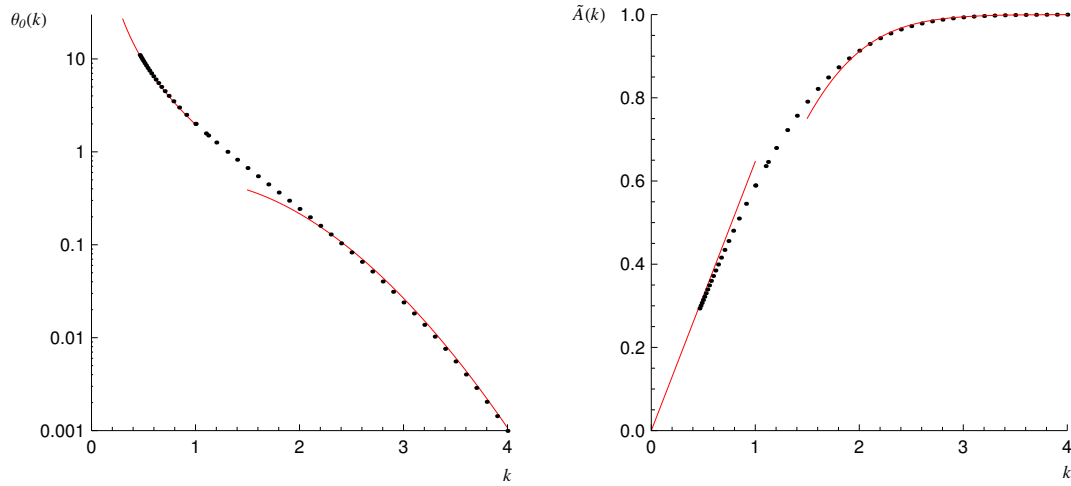
Figure 3.3:   **Left:**   Dependence of the exponent θ_0 on k; similar to Fig. 2 in Ref. [Krapivsky 1996], but in lin-log scale; see in particular Eqs. (9b) and (12) there. **Right:** Dependence of the prefactor $\widetilde{A}$ on k. The red solid lines illustrate the analytical behavior in the limiting cases $k \to 0$ and $k \to \infty$.

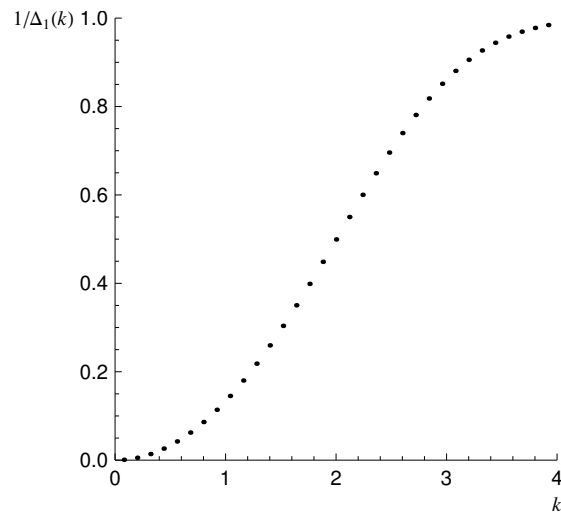Figure 3.4: $1/\Delta_1(k)$ saturates to 1, so that the condition $N \gg \exp[1/\Delta_1(k)]$ is virtually always satisfied.

3.1.3 Back to GoF testing and conclusion

Let us now come back to GoF testing. In order to convert the above calculations into a meaningful test, one must specify values of a and b. The natural choice would be $a = 1/N$, corresponding to the min of the sample series since $F(\min x_n) \approx F_N(\min x_n) = \frac{1}{N}$. Eq. (3.7) above motivates a slightly different value of $a = 1/(N+1)$ and $b = 1 - a$, such that the relevant value of T is given correspondingly by

$$T = \ln \sqrt{\frac{b\,(1-a)}{a\,(1-b)}} = \ln N.$$

This leads to our central result for the cdf of the weighted maximal Kolmogorov distance $K(\frac{1}{N+1}, \frac{N}{N+1})$ under the hypothesis that the tested and the true distributions coincide:

$$\boxed{S(N;k) = \mathcal{P}_<(k|\ln N) = \widetilde{A}(k)N^{-\theta_0(k)}}, \tag{3.16}$$

which is valid whenever $N \gg 1$ since, as we discussed above, the energy gap Δ_1 is greater than unity.

The final cumulative distribution function (the test law) is depicted in Fig. 3.5(b) for different values of the sample size N: As N grows toward infinity, the curve is shifted to the right, and eventually $S(\infty;k)$ is zero for any k. The red solid lines illustrate the analytical behavior in the limiting cases $k \to 0$ and $k \to \infty$. Contrarily to the standard KS case (Fig. 3.5(a)), this distribution *still depends on N*. In particular, the threshold value k^* corresponding to a 95% confidence level increases with N. Since for large N, $k^* \gg 1$ one can use the asymptotic expansion above, which soon becomes quite accurate, as shown in Fig. 3.5(b). This leads to:

$$\theta_0(k^*) \approx -\frac{\ln 0.95}{\ln N} \approx \sqrt{\frac{2}{\pi}} k^* \, \mathrm{e}^{-\frac{k^{*2}}{2}},$$

which gives $k^* \approx 3.439, 3.529, 3.597, 3.651$ for, respectively, $N = 10^3, 10^4, 10^5, 10^6$. For exponentially large N and to logarithmic accuracy, one has: $k^* \sim \sqrt{2\ln(\ln N)}$. This variation is very slow, but one sees that as a matter of principle, the "acceptable" maximal value of the weighted distance is much larger (for large N) than in the KS case.

In conclusion, we believe that accurate GoF tests for the extreme tails of empirical distributions is a very important issue, relevant in many contexts. We have derived exact asymptotic results for a generalization of the Kolmogorov-Smirnov test, well suited to testing the whole domain up to these extreme tails. Our final results are summarized in Eq. (3.16) and Fig. 3.5(b). In passing, we have rederived and made more precise the result of Krapivsky and Redner [Krapivsky 1996] concerning the survival probability of a diffusive particle in an expanding cage. It would be interesting to exhibit other choices of weight functions that lead to soluble survival probabilities. It would also be interesting to extend the present results to multivariate distributions, and to dependent observations, along the lines of Ref. [Chicheportiche 2011].
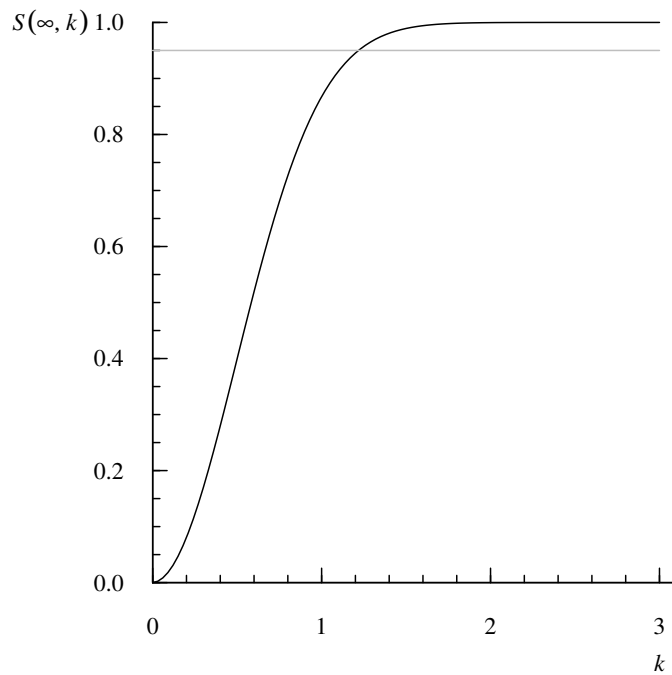
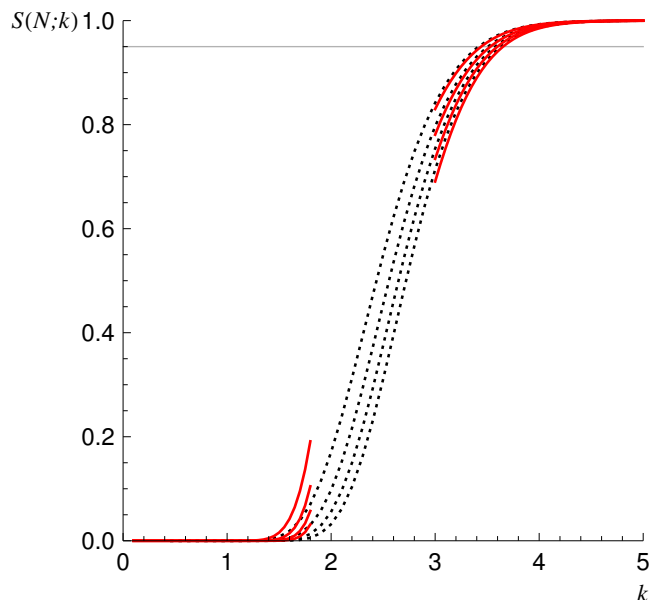(a) Kolmogorov asymptotic distribution.

(b) Dependence of $S(N; k)$ on k for $N = 10^3, 10^4, 10^5, 10^6$ (from left to right).

Figure 3.5: The horizontal grey line corresponds to a 95% confidence level.

3.2 Goodness-of-fit tests for a sample of dependent draws

Usual GoF tests, and even the generalization that we just worked out in the previous section of this chapter, are designed for samples of identically distributed *and independent* draws. It however so happens that in certain fields (physics, finance, geology, etc.) the random variables under scrutiny have some dependence, be it *spatial* or *temporal* (in which case one speaks of "memory"). Applying naively GoF tests to such samples leads to a pessimistic evaluation of the null hypothesis of the test, and eventually causes a too high rejection rate, as will be made more precise below. It is thus natural to ask how these dependences effectively affect the law of the test statistic. A minimal departure from independence is to allow for homogeneous dependences between draws i and j, for which the two-points copula is translationally invariant. In this chapter, we adopt the point of view of temporal dependences because it has immediate applications in financial time series (see later in Chapter III.8), but the theory applies as well to observables equally spaced on the line, and whose dependences rely on their respective distances.

Whereas the unconditional law of the variable may well be unique and independent of time, the conditional probability distribution of an observation *following* a previous observation exhibits specific patterns, and in particular long-memory, even when the linear correlation is short-ranged or trivial. Examples of such phenomena can be encountered in fluid mechanics (the velocity of a turbulent fluid) and finance (stock returns have small auto-correlations but exhibit strong volatility clustering, a form of heteroskedasticity). The long-memory nature of the underlying processes makes it inappropriate to use standard GoF tests in these cases. Still, the determination of the unconditional distribution of returns is a classic problem in quantitative finance, with obvious applications to risk control, portfolio optimization or derivative pricing. Correspondingly, the distribution of stock returns (in particular the behavior of its tails) has been the subject of numerous empirical and theoretical papers (see e.g. [Plerou 1999, Drăgulescu 2002] and for reviews [Bouchaud 2003, Malevergne 2006] and references therein). Clearly, precise statements are only possible if meaningful GoF tests are available [Weiss 1978].

In the rest of this section, we study theoretically how to account for general dependence in GoF tests: we first describe the statistical properties of the empirical cdf of a non-iid vector of observations of finite size, as well as measures of its difference with an hypothesized cdf. We then study the limit properties of this difference and the asymptotic distributions of two norms. Examples and financial applications are relegated to Chapter 8 in part III.

3.2.1 Empirical cumulative distribution and its fluctuations

Contrarily to section 3.0.1, we now consider $\boldsymbol{X}$ to be a random vector with N identically distributed but dependent variables, with marginal cdf F. One realization of $\boldsymbol{X}$ consists of a time series $\{x_1, \ldots, x_n, \ldots, x_N\}$ that exhibits some sort of persistence. Due to the persistence, the fluctuations of the empirical CDF (3.1) depend on the joint distribution of all couples (X_n, X_m), so that the covariance of the rescaled empirical CDF $Y_N(u)$ defined in Eq. (3.2) is now

$$\mathrm{Cov}(Y_N(u), Y_N(v)) = \min(u,v) - uv + \frac{1}{N} \sum_{n,m \neq n}^{N} \big(C_{nm}(u,v) - uv \big) \qquad (3.17)$$

and makes the pairwise copulas C_{nm} explicitly appear. It is more convenient to introduce

$$\Psi_N(u,v) = \frac{1}{N}\sum_{n,m\neq n}^{N} \Delta_{nm}(u,v), \quad \text{where} \quad \Delta_{nm}(u,v) = \frac{C_{nm}(u,v)-uv}{\min(u,v)-uv}, \tag{3.18}$$

which measures the relative departure from the independent case $\Psi_N(u,v) \equiv 0$, corresponding to the pairwise copula being independence $C_{nm}(u,v) = uv$. Note that decorrelated but dependent variables may lead to a non zero value of Ψ_N, since the whole pairwise copula enters the formula and not only the linear correlation coefficients. When the denominator is zero, the fraction should be understood in a limit sense; we recall in particular that [Chicheportiche 2012b]

$$\Delta_{nm}(u,u) = \frac{C_{nm}(u,u)-u^2}{u\,(1-u)} = \tau_{nm}^{\text{UU}}(u) + \tau_{nm}^{\text{LL}}(1-u) - 1 \tag{3.19}$$

tends to the upper/lower tail dependence coefficients $\tau_{nm}^{\text{UU}}(1)$ when u tends to 1, and to $\tau_{nm}^{\text{LL}}(1)$ when u tends to 0. Intuitively, the presence of $\Psi_N(u,v)$ in the covariance of Y_N above leads to a reduction of the number of effectively independent variables, but a more precise statement requires some further assumptions that we detail below.

In the following, we will restrict to the case of strong-stationary random vectors, for which the copula C_{nm} only depends on the lag $t = m - n$, i.e. $C_t \equiv C_{n,n+t}$. The average of Δ_{nm} over n, m can be turned into an average over t:

$$\Psi_N(u,v) = \sum_{t=1}^{N-1}(1-\frac{t}{N})\big(\Delta_t(u,v) + \Delta_{-t}(u,v)\big) \tag{3.20}$$

with $\Delta_t(u,v) = \Delta_{n,n+t}(u,v)$ and $\Delta_{-t}(u,v) = \Delta_{n+t,n}(u,v)$. Note that in general $\Delta_t(u,v) \neq \Delta_{-t}(u,v)$, but clearly $\Delta_t(u,v) = \Delta_{-t}(v,u)$, which implies that $\Psi_N(u,v)$ is symmetric in $u \leftrightarrow v$.

We will assume in the following that the dependence encoded by $\Delta_t(u,v)$ has a limited range in time, or at least that it decreases sufficiently fast for the above sum to converge when $N \to \infty$. If the characteristic time scale for this dependence is T, we assume in effect that $T \ll N$. In the example worked out in Section 8.1 below, one finds:

$$\Delta_t(u,v) = f\left(\frac{t}{T}\right)\frac{A(u,v)}{I(u,v)}, \qquad I(u,v) \equiv \min(u,v) - uv$$

where $f(\cdot)$ is a certain function. If $f(r)$ decays faster than r^{-1}, one finds (in the limit $T \gg 1$):

$$\Psi_\infty(u,v) = \lim_{N\to\infty}\Psi_N(u,v) = T\,\frac{A(u,v)+A(v,u)}{I(u,v)}\int_0^\infty \mathrm{d}r f(r),$$

with corrections at least of the order of T/N when $N \gg T$.

3.2.2 Limit properties

We now define the process $\tilde{y}(u)$ as the limit of $Y_N(u)$ when $N \to \infty$. For a given u, it represents the asymptotics of the difference between the empirically determined cdf of the

underlying X's and the theoretical one, at the u-th quantile. According to the Central Limit Theorem under weak dependences, it is Gaussian as long as the strong mixing coefficients,

$$\alpha_{\mathrm{SM}}(t) = \sup_{\tau} \sup_{A,B} \left\{ \left| \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) \right| : A \in \sigma(\{Z_n(u)\}_{n \leq \tau}), B \in \sigma(\{Z_n(v)\}_{n \geq \tau+t}) \right\}$$

associated to the sequence $\{Z_n(u)\} = \{\mathbb{1}_{\{F(X_n) \leq u\}} - u\}$, vanish at least as fast as $\mathcal{O}(t^{-5})^6$. We will assume this condition to hold in the following. For example, this condition is met if the function $f(r)$ defined above decays exponentially, or if $f(r \geq 1) = 0$.

The covariance of the process $\tilde{y}(u)$ is given by:

$$H(u,v) = I(u,v) \left[1 + \Psi_\infty(u,v)\right] \tag{3.21}$$

and characterizes a Gaussian bridge since $\mathbb{V}[\tilde{y}(0)] = \mathbb{V}[\tilde{y}(1)] = 0$, or equivalently $\mathbb{P}[\tilde{y}(0) = y] = \mathbb{P}[\tilde{y}(1) = y] = \delta(y)$. Indeed, $I(u,v) = \min(u,v) - uv$ is the covariance function of the Brownian bridge, and $\Psi_\infty(u,v)$ is a non-constant scaling term.

By Mercer's theorem, the covariance $H(u,v)$ can be decomposed on its eigenvectors and $\tilde{y}(u)$ can correspondingly be written as an infinite sum of Gaussian variables:

$$\tilde{y}(u) = \sum_{j=1}^{\infty} U_j(u)\sqrt{\lambda_j}\, z_j \tag{3.22}$$

where z_j are independent centered, unit-variance Gaussian variables, and the functions U_j and the numbers λ_j are solutions to the eigenvalue problem:

$$\int_0^1 H(u,v)U_i(v)\,\mathrm{d}v = \lambda_i\, U_i(u) \quad \text{with} \quad \int_0^1 U_i(u)U_j(u)\,\mathrm{d}u = \delta_{ij}. \tag{3.23}$$

In practice, for every given problem, the covariance function in Equation (3.21) has a specific shape, since $\Psi_\infty(u,v)$ is copula-dependent. Therefore, contrarily to the case of independent random variables, the GoF tests will not be characterized by universal (problem independent) distributions.

3.2.3 Law of the norm-2 (Cramér-von Mises)

The norm-2 of the limit process is the integral of $\tilde{y}^2$ over the whole domain:

$$CM = \int_0^1 \tilde{y}(u)^2\,\mathrm{d}u. \tag{3.24a}$$

In the representation (3.22), it has a simple expression:

$$CM = \sum_{j=1}^{\infty} \lambda_j z_j^2. \tag{3.24b}$$

---

[6]This condition means that the occurrence of any two realizations of the underlying variable can be seen as independent for sufficiently long time between the realizations. Since the copula induces a measure of probability on the Borel sets, it amounts in essence to checking that $|C_t(u,v) - uv|$ converges quickly towards 0. See Refs. [Bradley 2007, Chen 2010, Beare 2010] for definitions of $\alpha-$, $\beta-$, $\rho-$mixing coefficients and sufficient conditions on copulas for geometric mixing (fast exponential decay) in the context of copula-based stationary Markov chains.

and its law is thus the law of an infinite sum of squared independent Gaussian variables weighted by the eigenvalues of the covariance function. Diagonalizing H is thus sufficient to find the distribution of CM, in the form of the Fourier transform of the characteristic function

$$\phi(t) = \mathbb{E}\left[e^{it\,CM}\right] = \prod_j \left(1 - 2it\lambda_j\right)^{-\frac{1}{2}}. \tag{3.25}$$

The hard task consists in finding the infinite spectrum of H (or some approximations, if necessary).

Ordering the eigenvalues by decreasing amplitude, Equation (3.24b) makes explicit the decomposition of CM over contributions of decreasing importance so that, at a wanted level of precision, only the most relevant terms can be retained. In particular, if the top eigenvalue dominates all the others, we get the chi-square law with a single degree of freedom:

$$\mathbb{P}[CM \leq k] = \operatorname{erf}\sqrt{\frac{k}{\lambda_0}}. \tag{3.26}$$

Even if the spectrum cannot easily be determined but $H(u,v)$ is known, all the moments of the distribution can be computed exactly. For example:

$$\mathbb{E}[CM] = \operatorname{Tr} H = \int_0^1 H(u,u)\,\mathrm{d}u, \tag{3.27a}$$

$$\mathbb{V}[CM] \equiv 2 \operatorname{Tr} H^2 = 2 \iint_0^1 H(u,v)^2\,\mathrm{d}u\,\mathrm{d}v. \tag{3.27b}$$

3.2.4 Law of the supremum (Kolmogorov-Smirnov)

The supremum of the difference between the empirical cdf of the sample and the target cdf under the null-hypothesis has been used originally by Kolmogorov and Smirnov as the measure of distance. The variable

$$KS = \sup_{u\in[0,1]} |\tilde{y}(u)| \tag{3.28}$$

describes the limit behavior of the GoF statistic. In the case where $1 + \Psi_\infty(u,v)$ can be factorized as $\sqrt{\psi(u)}\sqrt{\psi(v)}$, the procedure for obtaining the limiting distribution was worked out in [Anderson 1952], and leads to a problem of a diffusive particle in an expanding cage, for which some results are known. There is however no general method to obtain the distribution of KS for an arbitrary covariance function H.

Nevertheless, if H has a dominant mode, the relation (3.22) becomes approximately: $\tilde{y}(u) = U_0(u)\sqrt{\lambda_0}\,z_0 \equiv \kappa_0(u)z_0$, and

$$KS = \sqrt{\lambda_0}\,|z_0| \sup_{u\in[0,1]} |U_0(u)| \equiv \kappa_0(u_0^*)|z_0|. \tag{3.29}$$

The cumulative distribution function is then simply

$$\mathbb{P}[KS \leq k] = \operatorname{erf}\left(\frac{k}{\sqrt{2}\,\kappa_0(u_0^*)}\right), \qquad k \geq 0. \tag{3.30}$$

This approximation is however not expected to work for small values of k, since in this case z_0 must be small, and the subsequent modes are not negligible compared to the first one. A perturbative correction — working also for large k only — can be found when the second eigenvalue is small, or more precisely when $\tilde{y}(u) = \kappa_0(u)z_0 + \kappa(u)z_1$ with $\epsilon = \kappa/\kappa_0 \ll 1$. The first thing to do is find the new supremum

$$u^* = \arg\sup(\tilde{y}(u)^2) = u_0^* + \frac{\kappa'(u_0^*)}{|\kappa_0''(u_0^*)|}\frac{z_1}{z_0}. \tag{3.31}$$

Notice that it is dependent upon z_0, z_1 so that KS is no longer exactly the absolute value of a Gaussian. However it can be shown (after lengthy but straightforward calculation) that, to second order in ϵ, $\tilde{y}(u^*)$ *remains Gaussian*, albeit with a new width

$$\kappa^* \approx \sqrt{\kappa_0^2 + \kappa^2} > \kappa_0, \tag{3.32}$$

where all the functions are evaluated at u_0^*. In fact, this approximation works also with more than two modes, provided

$$\kappa(u)^2 \equiv \sum_{j\neq 0} \lambda_j U_j(u)^2 \ll \kappa_0(u)^2 = \lambda_0 U_0(u)^2, \tag{3.33}$$

in which case:

$$\kappa^* \approx \sqrt{\sum_j \lambda_j U_j(u_0^*)^2}, \tag{3.34}$$

where the sum runs from 0 to the farthest mode still satisfying the inequality in Eq. (3.33).

3.2.5 Conclusion

We have introduced a framework for the study of statistical tests of Goodness-of-Fit with dependent samples, that heavily relies on the notion of bivariate copulas.

In summary, GoF testing on persistent series cannot be universal as is the case for i.i.d. variables, but requires a careful estimation of the self-copula at all lags. Correct asymptotic laws for the test statistics can be found as long as dependences are short ranged, i.e. $T \ll N$. These laws depend on the spectral properties of the covariance kernel. The resulting distribution of the Cramér-von Mises statistic is given in Eq. (3.25) (in Fourier form), and the law of the Kolmogorov-Smirnov test can be read off Eq. (3.30) when a single eigenmode is dominant, or substituting Eq. (3.34) to take into account the contribution of subsequent modes.

An application of the modified Goodness-of-Fit tests developed in this chapter is provided in Chapter 8 of part II, where we test the distribution of stock returns.

We conclude with two remarks of methodological interest.

1) The method presented for dealing with self-dependences while using statistical tests of Goodness-of-Fit is computationally intensive in the sense that it requires to estimate empirically the self-copula for all lags over the entire unit square. In the non-parametric setup, discretization of the space must be chosen so as to provide a good approximation of the continuous distance measures while at the same time not cause too heavy computations. Considering that fact, it is often more appropriate to use the Cramér-von Mises-like test rather

than the Kolmogorov-Smirnov-like, as numerical error on the evaluation of the integral will typically be much smaller than on the evaluation of the supremum on a grid, more so when the grid size is only about $\frac{1}{M} \approx \frac{1}{100}$.

2) The case with long-ranged dependence $T \gg N \gg 1$ cannot be treated in the framework presented here. First because the Central Limit Theorem does not hold in that case, and finding the limit law of the statistic may require more advanced mathematics. But even pre-asymptotically, summing the lags over the available data up to $t \approx N$ means that a lot of noise is included in the determination of $\Psi_N(u,v)$ (see Equation 3.18). This, in turn, is likely to cause the empirically determined kernel $H(u,v)$ not to be positive definite. One way of addressing this issue is to follow a semi-parametric procedure: the copula C_t is still estimated non-parametrically, but the kernel H sums the lagged copulas C_t only up to a scale where the linear correlations and leverage correlations vanish, and only one long-ranged dependence mode remains. This last contribution can be fitted by an analytical form, that can then be summed up to its own scale, or even to infinity.

Part II

Cross-sectional dependences

The joint distribution of stock returns is not elliptical

Contents

4.1   Introduction

The most important input of portfolio risk analysis and portfolio optimization is the correlation matrix of the different assets. In order to diversify away the risk, one must avoid allocating on bundles of correlated assets, an information in principle contained in the correlation matrix. The seemingly simple mean-variance Markowitz program is however well known to be full of thorns. In particular, the empirical determination of large correlation matrices turns out to be difficult, and some astute "cleaning" schemes must be devised before using it for constructing optimal allocations [Ledoit 2004]. This topic has been the focus of intense research in the recent years, some inspired by Random Matrix Theory (for a review see [El Karoui 2009, Potters 2009]) or clustering ideas [Marsili 2002, Tola 2008, Tumminello 2007].

There are however many situations of practical interest where the (linear) correlation matrix is inadequate or insufficient [Bouchaud 2003, Malevergne 2006, Mashal 2002]. For example, one could be more interested in minimizing the probability of large negative returns rather than the variance of the portfolio. Another example is the Gamma-risk of option portfolios, where the correlation of the squared-returns of the underlyings is needed. Credit derivatives are bets on the probability of simultaneous default of companies or of individuals; again, an estimate of the correlation of tail events is required (but probably very hard to ascertain empirically) [Frey 2001], and for a recent interesting review [Brigo 2010].

Apart from the case of multivariate Gaussian variables, the description of non-linear dependence is not reducible to the linear correlation matrix. The general problem of parameterizing the full joint probability distribution of N random variables can be "factorized" into the specification of all the marginals on the one hand, and of the dependence structure (called the 'copula') of N standardized variables with uniform distribution in $[0, 1]$, on the other hand. The nearly trivial statement that all multivariate distributions can be represented in that way is called Sklar's Theorem [Embrechts 2002b, Sklar 1959]. Following a typical pattern of mathematical finance, the introduction of copulas ten years ago has been followed by a calibration spree, with academics and financial engineers frantically looking for copulas to best represent their pet multivariate problem. But instead of trying to understand the economic or financial mechanisms that lead to some particular dependence between assets and construct copulas that encode these mechanisms, the methodology has been — as is sadly often the case — to brute force calibrate on data copulas straight out from statistics handbooks [Durrleman 2000, Fermanian 2005b, Fischer 2009, Fortin 2002, Patton 2001]. The "best" copula is then decided from some quality-of-fit criterion, irrespective of whether the copula makes any intuitive sense (some examples are given below). This is reminiscent of the 'local volatility models' for option markets [Dupire 1994]: although the model makes no intuitive sense and cannot describe the actual dynamics of the underlying asset, it is versatile enough to allow the calibration of almost any option smile (see [Hagan 2002a]). This explains why this model is heavily used in the financial industry. Unfortunately, a blind calibration of some unwarranted model (even when the fit is perfect) is a recipe for disaster. If the underlying reality is not captured by the model, it will most likely derail in rough times — a particularly bad feature for risk management! Another way to express our point is to use a Bayesian language: there are families of models for which the 'prior' likelihood is clearly extremely small — we discuss below the case of Archimedean copulas. Statistical tests are

not enough — intuition and plausibility are mandatory.

The aim of this chapter is to study in depth the family of elliptical copulas, in particular Student copulas, that have been much used in a financial context and indeed have a simple intuitive interpretation [Embrechts 2002b, Frahm 2003, Hult 2002, Luo 2010, Malevergne 2006, Shaw 2007]. We investigate in detail whether or not such copulas can faithfully represent the joint distribution of the returns of US stocks. (We have also studied other markets as well). We unveil clear, systematic discrepancies between our empirical data and the corresponding predictions of elliptical models. These deviations are qualitatively the same for different tranches of market capitalizations, different time periods and different markets. Based on the financial interpretation of elliptical copulas, we argue that such discrepancies are actually expected, and propose the ingredients of a generalization of elliptical copulas that should capture adequately non-linear dependences in stock markets. The full study of this generalized model, together with a calibration procedure and stability tests, is provided in the next chapter.

4.2 Empirical study of the dependence of stock pairs

We carefully compare our comprehensive empirical data on stock markets with the predictions of elliptical models, with special emphasis on Student copulas and log-normal copulas. We conclude that elliptical copulas fail to describe the full dependence structure of equity markets.

4.2.1 Methodology: what we do and what we do not do

As this chapter's title stresses, the object of this study is to dismiss the elliptical copula as description of the multivariate dependence structure of stock returns. Concretely, the null-hypothesis "H_0: the joint distribution is elliptical" admits as corollary "all pairwise bivariate marginal copulas are elliptical and differ only by their linear correlation coefficient ". In other words, all pairs with the same linear correlation ρ are supposed to have identical values of non-linear dependence measures, and this value is predicted by the elliptical model. Focusing the empirical study on the pairwise measures of dependence, we will reject H_0 by showing that their average value over all pairs with a given ρ is different from the value predicted by elliptical models.

Our methodology differs from usual hypothesis testing using statistical tools and goodness of fit tests, as can be encountered for example in [Malevergne 2003] for testing the Gaussian copula hypothesis on financial assets. Indeed, the results of such tests are often not valid because financial time series are *persistent* (although almost not linearly autocorrelated), so that successive realizations cannot be seen as independent draws of an underlying distribution, see the recent discussion on this issue in [Chicheportiche 2011].

4.2.2 Data set and time dependence

The dataset we considered is composed of daily returns of 1500 stocks labeled in USD in the period 1995–2009 (15 full years). We have cut the full period in three sub-periods (1995–1999, 2000–2004 and 2005–2009), and also the stock universe into three pools (large caps, mid-caps and small caps). We have furthermore extended our analysis to the Japanese stock markets.

Qualitatively, our main conclusions are robust and do not depend neither on the period, nor on the capitalization or the market. Some results, on the other hand, do depend on the time period, but we never found any strong dependence on the capitalization. All measures of dependence are calculated pairwise with non-parametric estimators, and using all trading dates shared by both equities in the pair.

We first show the evolution of the linear correlation coefficients (Fig. 4.1) and the upper and lower tail dependence coefficients for $p = 0.95$ (Fig. 4.2) as a function of time for the large-cap stocks. One notes that (a) the average linear correlation $\overline{\rho}(t)$ fluctuates quite a bit, from around 0.1 in the mid-nineties to around 0.5 during the 2008 crisis; (b) the distribution of correlation coefficients shifts nearly rigidly around the moving average value $\overline{\rho}(t)$; (c) there is a marked, time dependent asymmetry between the average upper and lower tail dependence coefficients; (d) overall, the tail dependence tracks the behavior of $\overline{\rho}(t)$. More precisely, we show the time behavior of the tail dependence assuming either a Gaussian underlying copula or a Student ($\nu = 5$) copula with the same average correlation $\overline{\rho}(t)$. Note that the former model works quite well when $\overline{\rho}(t)$ is small, whereas the Student model fares better when $\overline{\rho}(t)$ is large. We will repeatedly come back to this point below.

We computed the quadratic $\zeta^{(2)}$ and absolute $\zeta^{(1)}$ correlation coefficients for each pair in the pool, as well as the whole rescaled diagonal and anti-diagonal of the empirical copulas (this includes the coefficients of tail dependence[1], see Part I). We then average these observables over all pairs with a given linear coefficient ρ, within bins of varying width in order to take account of the frequency of observations in each bin. We show in Fig. 4.3 $\zeta^{(1)}$ and $\tau^{\mathrm{UU}}(0.95), \tau^{\mathrm{LL}}(0.95)$ as a function of ρ for all stocks in the period 2000–2004, together with the prediction of the Student copula model for various values of ν, including the Gaussian case $\nu = \infty$. For both quantities, we see that the empirical curves systematically cross the Student predictions, looking more Gaussian for small ρ's and compatible with Student $\nu \approx 6$ for large ρ's, echoing the effect noticed in Fig. 4.2 above. The same effect would appear if one compared with the log-normal copula: elliptical models imply a residual dependence due to the common volatility, *even when the linear correlation goes to zero*. This property is not observed empirically, since we rather see that higher-order correlations almost vanish together with ρ. The assumption that a common volatility factor affects all stocks is therefore too strong, although it seems to make sense for pairs of stocks with a sufficiently large linear correlation. This result is in fact quite intuitive and we will expand on this idea in Sect. 4.2.5 below.

The above discrepancies with the predictions of elliptical models are qualitatively the same for all periods, market caps and is also found for the Japan data set. Besides, elliptical models predict a symmetry between upper and lower tails, whereas the data suggest that the tail dependence is asymmetric. Although most of the time the lower tail dependence is stronger, there are periods (such as 2002) when the asymmetry is reversed.

4.2.3   Copula diagonals

As argued in Chapter I.2 above, it is convenient to visualize the rescaled difference $\Delta_{\mathrm{d,a}}(p)$ between the empirical copula and the Gaussian copula, along the diagonal (p, p) and the

---

[1] The empirical relative bias for $p = 0.95$ is $\lesssim 1\%$ even for series with only 10^3 points — typically daily returns over 4 years.
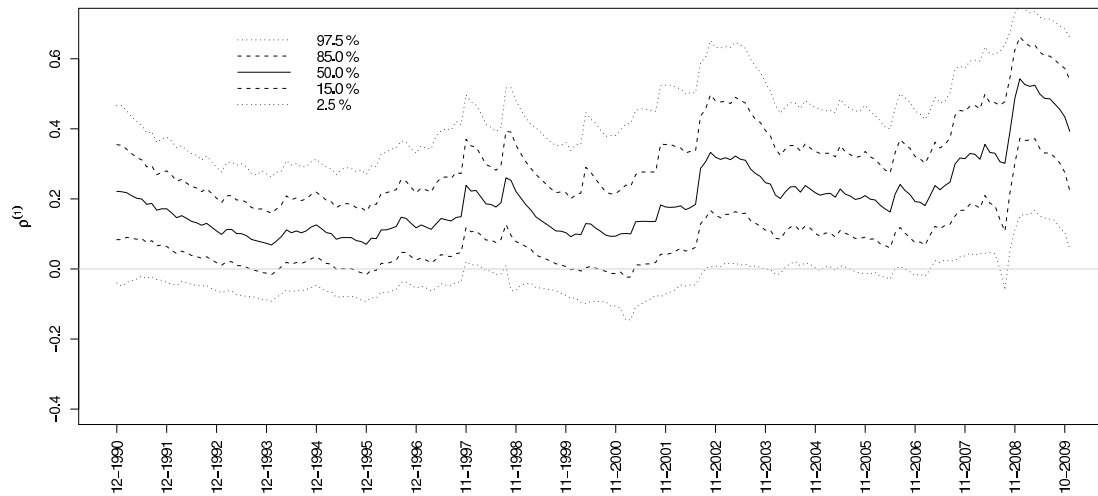
Figure 4.1: Time evolution of the coefficient of linear correlation, computed for the stocks in the S&P500 index, with an exponentially moving average of 125 days from January 1990 to December 2009. Five quantiles of the ρ distribution are plotted according to the legend, showing that this distribution moves quasi-rigidly around its median value.
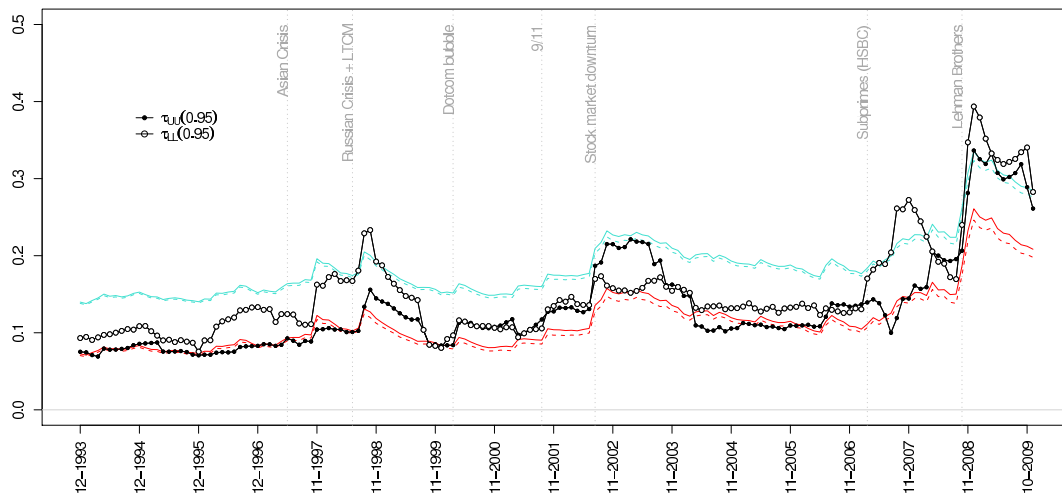

Figure 4.2: Time evolution of the coefficient of tail dependence at $p = 0.95$, computed for the stocks in the S&P500 index. A sliding flat window of 250 days moves monthly (25 days) from January 1990 to December 2009. We depict in turquoise and red the predictions for Student pair with $\nu = 5$ and $\nu = \infty$ respectively, based on the evolution of the mean value $\bar{\rho}(t)$ of the cross-sectional linear correlation (dashed) or averaged over the full instantaneous distribution of $\rho(t)$ (plain), with little difference.
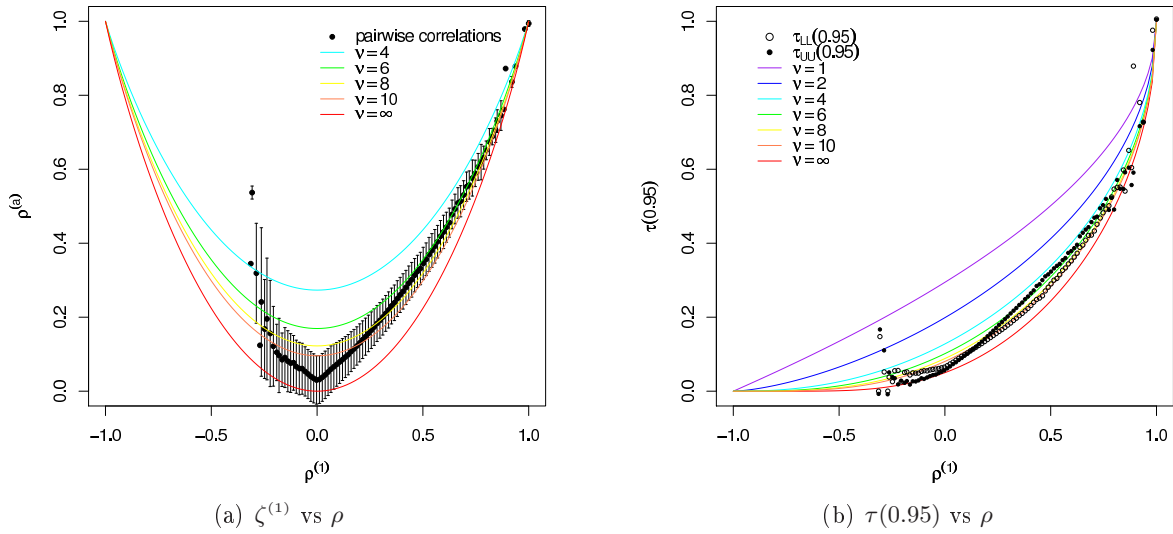
(a) $\zeta^{(1)}$ vs ρ

(b) $\tau(0.95)$ vs ρ

Figure 4.3: Empirical (2000–2004) and elliptical. The fact that the empirical points "cross" the theoretical lines corresponding to elliptical models reveals that there is a dependence structure in the volatility, more complex than what is suggested by a single common stochastic scale factor.

anti-diagonal $(p, 1-p)$, see Eqs. (2.17) page 20. The central point $p = \frac{1}{2}$ is special: $\Delta_{\mathrm{d,a}}(\frac{1}{2}) = \beta^{(\mathrm{B})} - \frac{2}{\pi} \arcsin \rho^{(1)}$ is zero for all pseudo-elliptical models.

We show in Fig. 4.5 the diagonal and anti-diagonal copulas for all pairs of stocks in the period 2000–2004, and for various values of ρ (similar plots for other periods are available in Ref. [Chicheportiche 2010]). We also show the prediction of the Student $\nu = 5$ model and of a Frank copula model with Student $\nu = 5$ marginals and the appropriate value of ρ. What is very striking in the data is that $\Delta_{\mathrm{d}}(p)$ is concave for small ρ's and becomes convex for large ρ's, whereas the Student copula diagonal is *always convex*. This trend is observed for all periods, and all caps, and is qualitatively very similar in Japan as well for all periods between 1991 and 2009. We again find that the Student copula is a reasonable representation of the data only for large enough ρ. The Frank copula is always a very bad approximation – see the wrong curvature along the diagonal and the inaccurate behavior along the anti-diagonal.

Let us now turn to the central point of the copula, $C^* = C(\frac{1}{2}, \frac{1}{2})$. We plot in Fig. 4.4 the quantity $-\cos(2\pi C^*) - \rho$ as a function of ρ, which should be zero for all elliptical models, as discussed in pages 25 and 32. The data here include the 284 equities constantly member of the S&P500 index in the period 2000–2009.

We again find a clear systematic discrepancy, that becomes stronger for smaller ρ's: the empirical value of C^* is too large compared with the elliptical prediction. In particular, for stocks with zero linear correlation ($\rho = 0$), we find $C^* > \frac{1}{4}$, i.e. even when two stocks are uncorrelated, the probability that they move in the same direction[2] is larger than $\frac{1}{4}$. The bias shown in Fig. 4.4 is again found for all periods and all market caps, and for Japanese stocks as well. Only the amplitude of the effect is seen to change across the different data

---

[2] A more correct statement is that both stocks have returns below their median with probability larger than $\frac{1}{4}$. However, the median of the distributions are very close to zero, justifying our slight abuse of language.

sets.

Statistical errors in each bin are difficult to estimate since pairs containing the same asset are mechanically correlated. In order to ascertain the significance of the previous finding, we have compared the empirical value of $C^*(\rho)$ with the result of a numerical simulation, where we generate time series of Student ($\nu = 5$) returns using the empirical correlation matrix. In this case, the expected result is that all pairs with equal correlation have the same bivariate copula and thus the same C^* so that the dispersion of the results gives an estimate of measurement noise. We find that, as expected, $C^*(\rho)$ is compatible with the Student prediction, at variance with empirical results. We also find that the dispersion of the empirical points is significantly larger than that of the simulated elliptical pairs with identical linear correlations, suggesting that all pairs *cannot be described by the same bivariate copula* (and definitively not an elliptical copula, as argued above).

All these observations, and in particular the last one, clearly indicate that Student copulas, or any elliptical copulas, are inadequate to represent the full dependence structure in stock markets.Because this class of copulas has a transparent interpretation, we in fact know why this is the case: assuming a common random volatility factor for all stocks is oversimplified. This hypothesis is indeed only plausible for sufficiently correlated stocks, in agreement with the set of observations we made above. As a first step to relax this hypothesis, we now turn to the pseudo-elliptical log-normal model, that allows further insights but still has unrecoverable failures.

4.2.4 Interpretation

The incentive to focus on bivariate measures comes from the theoretical property that all the marginals, including bivariate, of a multivariate elliptical distribution are themselves elliptical. In turn, a motivated statement that the pairwise distributions are not elliptical is enough to claim the non-ellipticity of the joint multivariate distribution. This is the basic line of argumentation of the present paper.

In a financial context, elliptical models are basically stochastic volatility models with arbitrary dynamics such that the ergodic distribution is $\mathcal{P}_\sigma$ (since we model only single-time distributions, the time ordering is irrelevant here). The important assumption, however, is that the random amplitude factor σ is *equal for all stocks* (or all assets in a more general context). In other words, one assumes that the mechanisms leading to increased levels of volatility affect all individual stocks identically. There is a unique market volatility factor. Of course, this is a very restrictive assumption since one should expect *a priori* other, sector specific, sources of volatility. This, we believe, is the main reason for the discrepancies between elliptical models and the empirical data reported below.

4.2.5 Pseudo-elliptical log-normal model

As we just showed, the fact that the central value of the copula C^* does not obey the relation $-\cos\left(2\pi C^*\right) = \rho$ rules out all elliptical models. One possible way out is to consider a model where random volatilities are stock dependent, as explained in Sect. 2.4.3. Choosing for simplicity a log-normal model of correlated volatilities and inserting (2.39) into (2.40), the
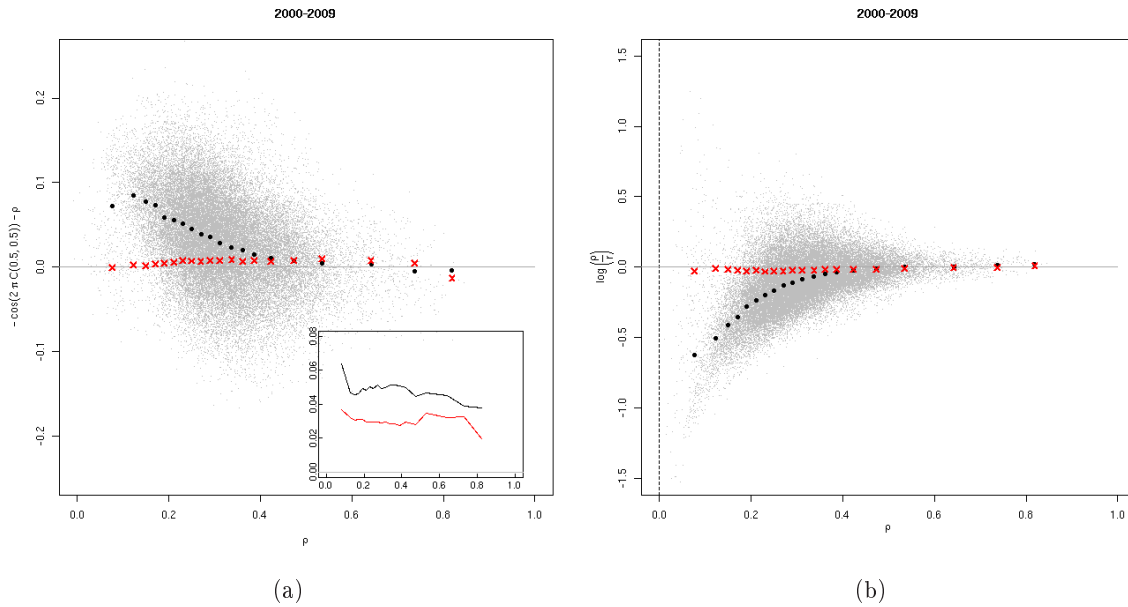
(a)   (b)

Figure 4.4: **Left:** $-\left(\cos\left(2\pi C^*\right) + \rho\right)$ versus the coefficient ρ of linear correlation. The grey cloud is a scatter plot of all pairs of stocks. Black dots correspond to the average value of empirical measurements within a correlation bin, and red crosses correspond to the numerically generated elliptical (Student, $\nu = 5$) data, compatible with the expected value zero. Statistical error are not shown, but their amplitudes is of the order of the fluctuations of the red crosses, so that empirics and elliptical prediction don't overlap for $\rho \lesssim 0.4$. This representation allows to visualize very clearly the systematic discrepancies for small values of ρ, beyond the average value. Inset: the 1 s.d. dispersion of the scattered points inside the bins. **Right:** $z = \ln\left(\rho/|\cos\left(2\pi C^*\right)|\right)$ as a function of ρ, with the same data and the same conventions. This shows more directly that the volatility correlations of weakly correlated stocks is overestimated by elliptical models, whereas strongly correlated stocks are compatible with the elliptical assumption of a common volatility factor.
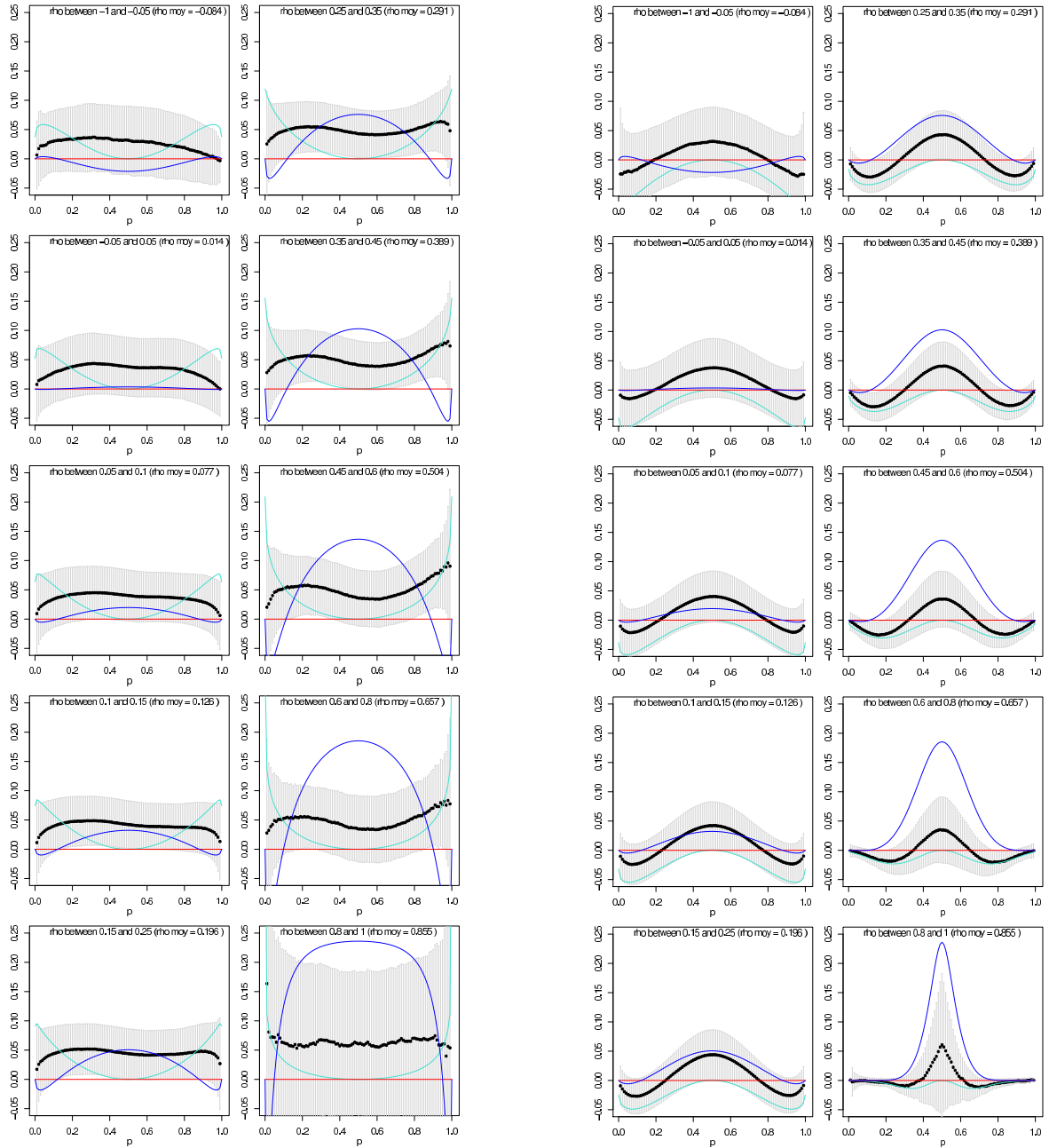
Figure 4.5: Diagonal (left) and anti-diagonal (right) of the copula vs p. Real data (2000-2004) averaged in 10 bins of ρ. The vertical lines corresponds to 1-sigma dispersion of the results (i.e. not the statistical error bar). The turquoise line corresponds to an elliptical model (Student with $\nu = 5$): it predicts too low a value at the center point and too large tail dependences (left and right limits). The blue line depicts the behavior of Frank's copula. It is worth noticing that the tail dependence coefficient is always measured empirically at some $p < 1$, for example $p = 0.95$. In this case, the disagreement between data and an elliptical model may be accidentally small. It is important to distinguish the interpretation of the limit coefficient from that of a penultimate value [6].

new prediction is:

$$\rho = -e^{s^2(c-1)} \cos\left(2\pi C^*\right), \tag{4.1}$$

where c is the correlation of the log-volatilities. This suggests to plot $z = \ln\left(\rho/|\cos\left(2\pi C^*\right)|\right)$ as a function of ρ, as shown in Fig. 4.4(b). For a purely elliptical model, z should be identically zero, corresponding to perfectly correlated volatilities ($c = 1$). What we observe in Fig. 4.4(b), on the other hand, is that c is indeed close to unity for large enough ρ's, but systematically decreases as ρ decreases; in other words, the volatilities of weakly correlated stocks are themselves weakly correlated. This is, again, in line with the conclusion we reached above.

However, this pseudo-elliptical model still predicts that $C^* = \frac{1}{4}$ for stocks with zero linear correlations, in disagreement with the data shown in Fig. 4.4(a). This translates, in Fig. 4.4(b), into a negative divergence of z when $\rho \to 0$. This finding means that we should look for other types of constructions. How can one have at the same time $\rho = 0$ and $C^* > \frac{1}{4}$? A toy-model, that serves as the basis for a much richer model that we will report elsewhere [Chicheportiche 2013b], is the following. Consider two independent, symmetrically distributed random factors ψ_1 and ψ_2 with equal volatilities, and construct the returns of assets 1 and 2 as:

$$X_{1,2} = \psi_1 \pm \psi_2. \tag{4.2}$$

Clearly, the linear correlation is zero. Using a cumulant expansion, one easily finds that to the kurtosis order, the central value of the copula is equal to:

$$C^* = \frac{1}{4} + \frac{\kappa_2 - \kappa_1}{24\pi} + \ldots \tag{4.3}$$

Therefore, if the kurtosis κ_1 of the factor to which both stocks are positively exposed is smaller than the kurtosis κ_2 of the spread, one does indeed find $C^*(\rho = 0) > \frac{1}{4}$.

We will investigate in Chapter 5 an additive model such as the above one, with random volatilities affecting the factors ψ_1 and ψ_2, that generalize elliptical models in a way to capture the presence of several volatility modes.

4.3   Conclusion

The object of this chapter was to discuss the adequacy of Student copulas, or more generally elliptical copulas, to describe the multivariate distribution of stock returns. We have shown, using a very large data set, with the daily returns of 1500 US stocks over 15 years, that elliptical models fail to capture the detailed structure of stock dependences.

In a nutshell, the main message elicited by our analysis is that Student copulas provide a good approximation to describe the joint distribution of strongly correlated pairs of stocks, but badly miss their target for weakly correlated stocks. We believe that the same results hold for a wider class of assets: it is plausible that highly correlated assets do indeed share the same risk factor. Intuitively, the failure of elliptical models to describe weakly correlated assets can be traced to the inadequacy of the assumption of a single "market" volatility mode. We expect that exactly as for returns, several factors are needed to capture sectorial volatility modes and idiosyncratic modes as well. The precise way to encode this idea into a workable model that faithfully captures the non-linear dependence in stock markets is

reported in the next chapter. We strongly believe that such a quest cannot be based on a formal construction of mathematically convenient models (such as Archimedean copulas that, in our opinion, cannot be relevant to describe asset returns). The way forward is to rely on intuition and plausibility and come up with models that make financial sense. This is, of course, not restricted to copula modeling, but applies to all quarters of quantitative finance.

A minimal factor model for volatility dependences in stock returns

Contents

5.1 Introduction

Dependences among financial assets or asset classes stand at the heart of modern portfolio selection theories. Whatever the (concave) utility of an investor and its risk measure, diversification is profitable but optimal diversification is only reached if the underlying dependence structure is well understood.

Financial risk and portfolio selection: the linear covariances

The standard Markowitz theory [Markowitz 1952, Markowitz 1959, Bouchaud 2003] of optimal portfolio design aims at finding the optimal weights w_i to attribute to each stock of a pool. It assumes that stock returns are correlated random variables x_i, and that the optimizing agent has a "mean-variance" quadratic utility function in the form $U(\boldsymbol{w}) = \mathbb{E}[\boldsymbol{x} \cdot \boldsymbol{w}] - \mu \mathbb{V}[\boldsymbol{x} \cdot \boldsymbol{w}]$. It hence relies on the linear covariance matrix $\rho = \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^\dagger]$ of the stock returns, and more importantly on its inverse ρ^{-1}. Indeed, with no further constraints (budget, transaction costs, operational risk constraint, prohibition of short selling, etc.), the optimal weights are given by

$$\boldsymbol{w}_\rho^* \propto \rho^{-1}\boldsymbol{g} = \boldsymbol{g} + \mathrm{V}\left(\Lambda^{-1} - \mathbb{1}\right)\mathrm{V}^\dagger\boldsymbol{g}$$

where $\boldsymbol{g}$ is the vector of gain targets for the assets in the basket, and $\rho = \mathrm{V}\Lambda\mathrm{V}^\dagger$ is the spectral decomposition of the covariance matrix with V being the square matrix of eigenvectors and Λ the diagonal matrix of eigenvalues. Empirical estimates of ρ and its spectrum Λ are typically very noisy, and cleaning schemes need to be applied before inversion if one wants to avoid artificially enhancing the weights of low-risk in-sample modes (i.e. with eigenvalue smaller than 1) that turn into high-risk realized out-of-sample modes.

All this is fairly standard practice now, and several cleaning schemes have been designed, in view of modeling either the signal (parametric models, factor models, Principal Components Analysis), or the noise (RMT-based [Laloux 1999, Laloux 2000, Ledoit 2004, Potters 2005, Bartz 2012]).

Non-linear dependences and risk

However, it is now established that markets operate beyond the linear regime, that implicitly assumes gaussianity.

For one thing, individual stock returns are known to be non-Gaussian, and moments beyond the mean and variance have gained considerable interest (e.g. the excess kurtosis, or low-moment estimates thereof). But more importantly, stock returns are *jointly* not Gaussian: the structure of dependence between pairs of stocks is not compatible with the Gaussian copula, and as a consequence the penalty in the utility function should be more subtle than just the portfolio variance and include non-linear measures of risk (like tail events, quadratic correlations, etc.) in order to better fit the agent's risk aversion profile. Only in a multivariate Gaussian setting do these non-linear dependences resume to the linear correlations.

Non-linear dependences are also very important in the pricing and risk management of structured products and portfolios of derivatives. For example, the payoff of a hedged option has a V-shape with linear asymptotes and quadratic core, see Fig. 5.1. A portfolio of several such hedged options has thus a variance characterized by the absolute and quadratic
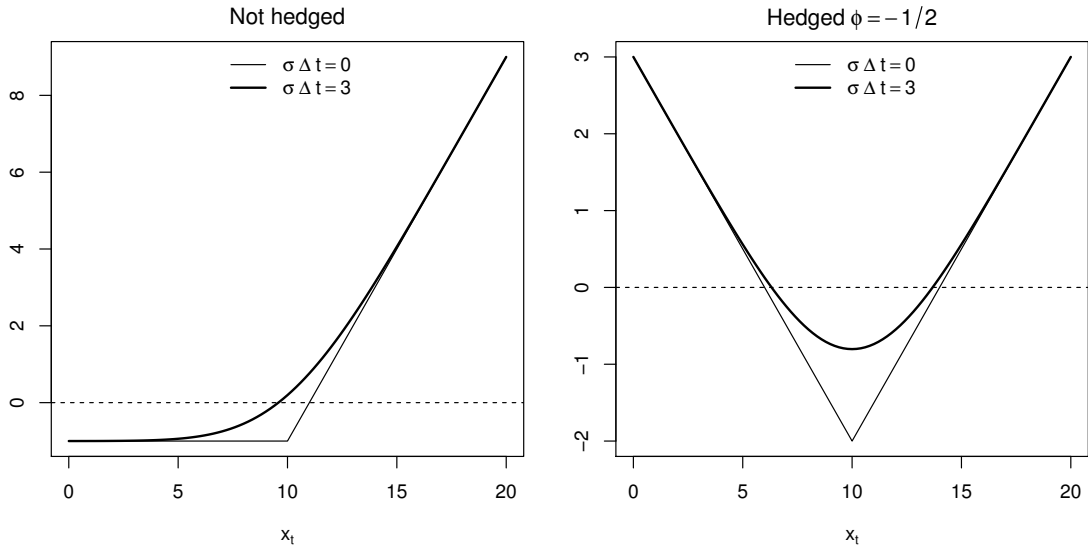
Figure 5.1: Expected payoff of an option as a function of the current price of the underlying stock. **Left:** unhedged; **Right:** hedged by short selling $\phi = 1/2$ shares of the underlying. The illustration is for a call option of strike $x_K = 10$, price $\mathcal{C} = 1$ on a stock of initial price $x_0 = 8$ following a Bachelier diffusion with volatility σ. The thin line is the payoff at expiry ($\sigma \Delta t = 0$) and the thick curve is the expected value of the payoff before expiry ($\sigma \Delta t = 3$).

correlations of the underlying stocks (gamma risk). These correlations of amplitudes are noisier than linear correlations, whence the need for a reliable model of both linear and non-linear dependences.

In Chapter 4, we showed that the daily returns of stocks are not exposed to a unique mode of volatility affecting all individuals at once. We thus ruled out all models with a single stochastic volatility σ, of the form

$$x_i = \sigma \, \epsilon_i, \tag{5.1}$$

with jointly Gaussian (and correlated) residuals ϵ_i's. This, we argued, was revealing a finer structure in the non-linear dependences, and opened the way for a description taking into account several modes of volatility. However, we also showed that any description in the form of individual volatilities

$$x_i = \sigma_i \, \epsilon_i, \tag{5.2}$$

with possible dependences between the σ_i's, would not be able to explain successfully the departure of $C(\frac{1}{2}, \frac{1}{2})$ from the elliptic prediction either. We called models like Eq. (5.2) "pseudo-elliptical", with one typical example being the multiplicative decomposition of σ_i onto the market volatility σ, a sectorial volatility $\hat{\sigma}_s$ (where stock i belongs to sector s), and a residual volatility $\widetilde{\sigma}_i$:

$$\sigma_i = \sigma \hat{\sigma}_s \widetilde{\sigma}_i.$$

It hence turned out that tuning only the radial amplitudes of a Gaussian vector would not be enough to reproduce some non-linear dependences relying on the *ranks* of the realizations in the data. Instead, additive non-Gaussian factors are thought to be able to generate anomalous

copula values, because of the interplay of factor kurtosis and residual kurtosis, as motivated by the toy model for $C(\frac{1}{2}, \frac{1}{2}) > 0$ with uncorrelated variables, presented in the previous chapter, page 63.

All pseudo-empirical models have a simple prediction linking the medial copula to the coefficient of linear correlation, see the discussion in page 32:

$$C(\tfrac{1}{2}, \tfrac{1}{2}) = \frac{1}{4} + \frac{1}{2\pi} \arcsin \rho \tag{5.3}$$

Said differently, the effective correlation[1]

$$\rho^{(\mathrm{B})} \equiv \cos\big(2\pi C(\tfrac{1}{2}, \tfrac{1}{2})\big) \tag{5.4}$$

is equal to ρ for these models, whereas in general it is not.

Summary of empirical stylized facts to be addressed by the model

More generally, any model for a joint description of stock returns must address the stylized facts:

- Generate fat-tailed return series, and even non-Gaussian factors and residuals

- Reproduce the structure of linear correlations with a reduced number of factors. In particular exhibit a market mode of linear dependences.

- Allow for a dependence between the volatilities of the residuals and the volatilities of the factors [Cizeau 2001, Allez 2011].

- Reproduce the anomalous copula structure (diagonals, and in particular medial point), see Fig. 5.2. It was already noted in chapter 4 that stock pairs with a high correlation are "more elliptical", and that in periods of high turmoil like the financial crisis (b) stock pairs are both more correlated and more elliptical, revealing a strong exposure to a common mode of volatility.

- Predict the structure of non-linear (typically quadratic) correlations with a reduced number of new parameters, in order to clean the empirically measured (and much noisy) corresponding dependence coefficients.

We show below, that this can be achieved with a factor structure in the factor-volatilities themselves, with again a common mode and idiosyncratic volatility modes for the linear factors.

Data set

We study daily close-to-close log-returns of stock prices of companies that are present in the S&P500 index during the whole of the period studied. We will be considering three periods for the empirical study and the model calibration: before the financial crisis (Jan,

---

[1] The superscript (B) stands for "Blomqvist", as $\rho^{(\mathrm{B})}$ is related to Blomqvist's beta coefficient. The properties of $\rho^{(\mathrm{B})}$ were discussed in Chapter I.2, see in particular the definition (2.27) and the comments in page 26.
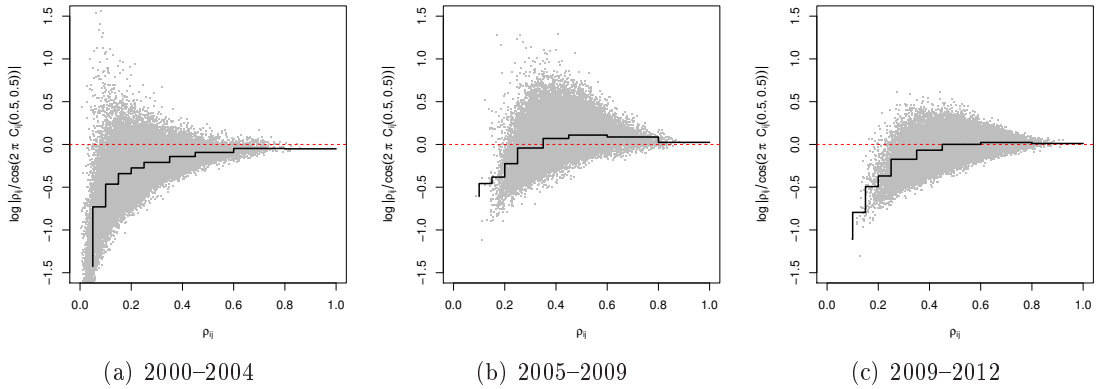
(a) 2000–2004   (b) 2005–2009   (c) 2009–2012

Figure 5.2: Empirics: scatter plot of $\ln|\rho/\rho^{(\mathrm{B})}|$ vs ρ for each stock pair, see Eq. (5.4). This figure quantifies the departure of the medial point of the copula $C(\frac{1}{2}, \frac{1}{2})$ from the elliptical benchmark, for which the prediction is a straight horizontal line at 0 (dashed red).

Bloomberg sector	Code	2000–04	2005–09	2009–12	2000–09
Communications	# 3	33	25	29	18
Consumer, Cyclical	# 4	60	49	33	40
Consumer, Non-Cyclical	# 5	67	75	75	53
Diversified	# 6	0	0	1	0
Energy	# 7	19	21	34	15
Financial	# 8	57	55	75	37
Industrial	#11	51	50	50	42
Technology	#13	38	43	35	33
Utilities	#14	27	27	28	24
Total number of firms (N)		352	345	360	262
Total number of days (T)		1255	1258	755	2514

Table 5.1: Economic sectors according to Bloomberg classification, with corresponding number of individuals for each period.

2000 – Dec, 2004); during the financial crisis (Jan, 2005 – Dec, 2009); after the financial crisis (Aug, 2009 – Dec, 2012). A longer dataset is used for the sliding windows procedure of In-sample/Out-of-sample testing in the last section: there we consider the ten years period 2000–2009.

It is enlightening to group the companies according to their sector of activity, in order to see if patterns appear. When needed, we will make use of Bloomberg's classification as summarized in Tab. 5.1.

Outline

This chapter is made of four sections. In Section 2, we study the linear correlations of pairs of stocks and discuss the design and estimation of a factor model for their description. The factors and residuals generated by the calibration of the model are studied in Section 3, and motivate the specification of the volatility content of the model that we present in Sec-

tion 4. The resulting non-Gaussian model is calibrated, and Section 5 is dedicated to an Out-of-sample stability analysis, validating its usefulness for the description of non-linear dependences.

5.2   Linear factors

We study a simple one-level (not hierarchical) model for the joint description of the stock returns x_i of N firms, as a combination of M shared factors f_k:

$$x_i = \sum_{k=1}^{M} \beta_{ki} f_k + e_i. \tag{5.5}$$

The weight β_{ki} parameterize the linear exposure of stock i to factor k. At this stage, we do not yet specify the statistical properties of the factors f_k and residuals e_i, except that we impose their linear uncorrelatedness:

$$\langle f_k f_l \rangle = \delta_{kl} \tag{5.6a}$$

$$\langle e_i e_j \rangle = \delta_{ij} \left(1 - \sum_l \beta_{li}^2 \right) \tag{5.6b}$$

$$\langle f_k e_j \rangle = 0. \tag{5.6c}$$

This way, the residuals can be understood as idiosyncrasies and all the linear dependence is supposed to be taken into account ant the factors. This, as we will shortly see, is already not trivial, for statistical factor models usually rely on the Principal Components Analysis (PCA)decomposition which does not generate orthogonal residuals.

5.2.1   Linear correlations

The predictions of the model in terms of covariances of the returns x_i do not need additional assumptions, and only depend on the matrix of linear weights β:

$$\rho_{ij} = \langle x_i x_j \rangle = \begin{cases} (\beta^\dagger \beta)_{ij} & , i \neq j \\ 1 & , i = j \end{cases} \tag{5.7}$$

When calibrating the model on real data, we want to find the M most relevant *uncorrelated* and *common* unit-variance factors F $(T \times M)$, and the exposures β $(M \times N)$ of every stock to every factor, such that

$$X = F\beta + E, \tag{5.5$'$}$$

and at the same time, E be orthogonal $T \times (N-M)$. Notice that the factors series F_{tk} are *not* inputs of the estimation problem, but rather are obtained as a byproduct. This is at contrast with the usual econometric determination of elasticities β in linear regressions of the explanatory (thus known) variables F onto the explained variables X.

Positing that the model (5.5$'$) is a faithful representation of real returns, we are able to uncover the factor structure and calibrate the corresponding weights β. Then, the estimate of the correlations (5.7) is a cleaned version of the brute force and noisy sample covariances $\frac{1}{T} X^\dagger X$.

A proxy: the Principal Components Analysis (PCA)

Diagonalization of the sample correlation matrix yields

$$\frac{1}{T}X^\dagger X = V\Lambda V^\dagger$$

where Λ is the diagonal matrix of eigenvalues, and the columns of V are the corresponding eigenvectors. Hence, there always exist (linearly orthogonal) factor series $\widetilde{F}$ such that the return series X can be decomposed as

$$X = \widetilde{F}\Lambda^{\frac{1}{2}}V^\dagger \qquad \text{where} \qquad \frac{1}{T}\widetilde{F}^\dagger\widetilde{F} = \mathbb{1}_N. \tag{5.8}$$

In order to reconciliate this decomposition in terms of statistical uncorrelated modes $\widetilde{F}$ with the factors F of the model, the PCA solution (5.8) needs to be identified with Eq. (5.5').

The factors should explain as much as possible of the returns covariances (thus of the portfolio variance), leaving only idiosyncratic residual volatility to be explained by the e_i's. Said differently, only those eigenvalues having a significant signal-to-noise ratio (compared f.ex. to an RMT benchmark) should be kept in the identification of the spectral decomposition with the factor model. This procedure is known as "eigenvalue clipping", and typically the most relevant eigenvalues are the largest ones and the smallest ones, on the right and the left of the noise bulk in the empirical spectrum. Since the smallest eigenvalues do not contribute much to the portfolio variance despite being statistically significant, we will focus on the modes with large eigenvalues. Ordering the eigenvalues in decreasing order, and splitting the first M (subscript $_{M|}$) from the last $(N-M)$ (subscript $_{|N-M}$), it is straightforward to obtain the identification

$$\beta_{\text{PCA}} = \Lambda_{M|}^{\frac{1}{2}}V_{M|}^\dagger. \tag{5.9}$$

At this stage, the series of factors can be formally identified as the first M spectral modes

$$F = \widetilde{F}_{M|} = (XV\Lambda^{-\frac{1}{2}})_{M|} \tag{5.10}$$

such that indeed $\frac{1}{T}F^\dagger F = \mathbb{1}_M$. But the residuals are *not* orthogonal:

$$E = \widetilde{F}_{|N-M}\Lambda_{|N-M}^{\frac{1}{2}}V_{|N-M}^\dagger \quad \text{s.t.} \quad \frac{1}{T}E^\dagger E = V_{|N-M}\Lambda_{|N-M}V_{|N-M}^\dagger \tag{5.11}$$

and thus cannot be understood as idiosyncrasies of the returns series.

Factor weights

The PCA can alternatively be thought of as the solution of

$$\frac{1}{T}X^\dagger X = \beta^\dagger\beta \quad \text{with} \quad \beta\beta^\dagger \text{ diagonal}$$

where importantly β is full rank, and the M modes with largest amplitude $(\beta\beta^\dagger)_{kk}$ are kept *after* the equation is solved. When one rather wants to optimize the contribution of *only* M modes, one can try to minimize a distance between the LHS and the RHS with a matrix of weights of rank M:

$$\arg\min \left\|\frac{1}{T}X^\dagger X - \beta^\dagger\beta\right\|$$

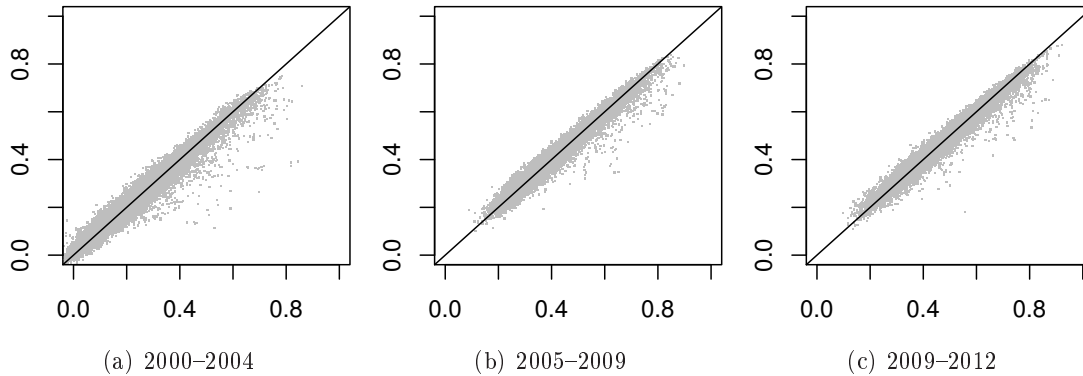(a) 2000–2004   (b) 2005–2009   (c) 2009–2012

Figure 5.3: Linear correlations: calibrated ($M = 10$) vs sample correlations. Outliers that stand far from the identity line (in particular for the period (a)) are strongly correlated pairs, typically in the same sector. Such pair correlations are poorly taken into account by factor models.

In this case an exact solution does not exist as there are NM variables for $N(N + 1)/2$ equations, yet one does not throw away information of subsequent modes, so that in the end the information content of that solution is typically bigger than that of the first M columns of β_{PCA}.

On the top of this, the estimation of the β's can be further improved if one *knows* the standard deviation of the series of returns X, and there only remains to estimate the correlations (rather than the covariances). After renormalization, the matrix $\frac{1}{T}\mathrm{X}^\dagger\mathrm{X}$ has only ones on the diagonal, and since this is also what the model predicts according to Eq. (5.7), the optimal estimation can be performed on the off-diagonal content only:

$$\arg\min \left\| \frac{1}{T}\mathrm{X}^\dagger\mathrm{X} - \beta^\dagger\beta \right\|_{\text{off-diag}}. \tag{5.12}$$

By lowering the number of equations to $N(N - 1)/2$ while keeping the same number of variables NM, one improves the identifiability of the model parameters.

The underlying view of this estimation method as of the collective dynamics is that variances and correlations are better estimated separately than jointly through the covariance matrix. As soon as one claims having a "good" estimator of the variances, one is better off estimating the covariances of the normalized time series, rather than having estimated the covariances in the first place. Notice that orthogonality of the lines of β obtained with this method is not granted, as opposed to the PCA, but what matters is rather uncorrelatedness of the factors.

Linear model calibration on financial data

The linear weights β are calibrated on the three data sets by solving Eq. (5.12) with $M = 10$, corresponding to the number of Bloomberg sectors plus one. The resulting correlations $(\beta^\dagger\beta)_{ij}$ are shown as a scatter plot versus the sample correlation $\frac{1}{T}(\mathrm{X}^\dagger\mathrm{X})_{ij}$ on Fig. 5.3. Although the calibration seems to provide a "good fit" (whatever this might mean), a precise characteriza-

tion of the benefit of using the method discussed above over the PCA is only achieved by a criterion of model selection, or by out-of-sample risk measurements.

We discuss this in detail in Sect. 5.5, where we use Markowitz' portfolio selection framework to compute the *realized* risk of optimal baskets of stocks. We show there that when using the proposed calibration method, the out-of-sample risk can be more than 25% lower than PCA[2], with an in-sample risk almost unchanged, see Fig. 5.19 below.

Orthogonal residuals

As already said, the byproduct of the PCA identification (5.10) is the matrix E of overlapping residuals given in Eq. (5.11). When the weights β are known, it is however possible to design a different identification scheme that generates orthogonal residuals. Consider indeed the date-by-date regression

$$X_{t\cdot} = F_{t\cdot}\beta + E_{t\cdot}. \tag{5.13}$$

where the explained variables are the input returns, the explanatory variables are the (freshly estimated) weights β and the regression parameters to be estimated are the factors. Then a GLS solution of the regression yields the wanted factors and residual series. It is only approximate in the sense that $\frac{1}{T}E^{\dagger}E$ is only "as close as can be" to a diagonal matrix, and $\frac{1}{T}F^{\dagger}F$ is only approximately $\mathbb{1}_M$.

In the next section, we discuss in details the properties of the factors series $F_{t\cdot}$ and residual series $E_{t\cdot}$, obtained after calibration of the model on the datasets.

5.3   Properties of the reconstructed factors and residuals

After the calibration of the weights β, we are able to perform a statistical analysis on the reconstructed series of factors and residuals. Because we have designed a calibration procedure improving over the PCA, the obtained factors are not pure eigenmodes of the correlation matrix. Still, we can interpret them looking at their statistical signatures. In particular, the first factor, f_1 is clearly a market mode (the "index", a collective mode of fluctuations). Indeed, introducing the average daily return $\overline{X} = \frac{1}{N}X\mathbf{1}$, we find that it has an almost perfect overlap with the first factor: $\mathrm{Cor}(F_{t1}, \overline{X}_t) = 99.9\%$!

The non-linear properties of the reconstructed factors and residuals can be investigated through the higher-order correlations

$$\frac{1}{p^2} \ln \frac{\langle|F_{tk}F_{tl}|^p\rangle}{\langle|F_{tk}|^p\rangle\langle|F_{tl}|^p\rangle} \tag{5.14a}$$

$$\frac{1}{p^2} \ln \frac{\langle|E_{ti}E_{tj}|^p\rangle}{\langle|E_{ti}|^p\rangle\langle|E_{tj}|^p\rangle} \tag{5.14b}$$

$$\frac{1}{p^2} \ln \frac{\langle|F_{tk}E_{tj}|^p\rangle}{\langle|F_{tk}|^p\rangle\langle|E_{tj}|^p\rangle} \tag{5.14c}$$

for any value of $p > 0$. As an example, we show in Fig. 5.4 the off-diagonal matrix elements of the factor-factor correlations (5.14a) for a value of $M = 10$ (again, the factors series are

---

[2]The PCA method is also known as eigenvalue clipping in the context of cleaning schemes for matrix inversion, and is one of the best generic cleaning scheme known so far, see Ref. [Potters 2009]
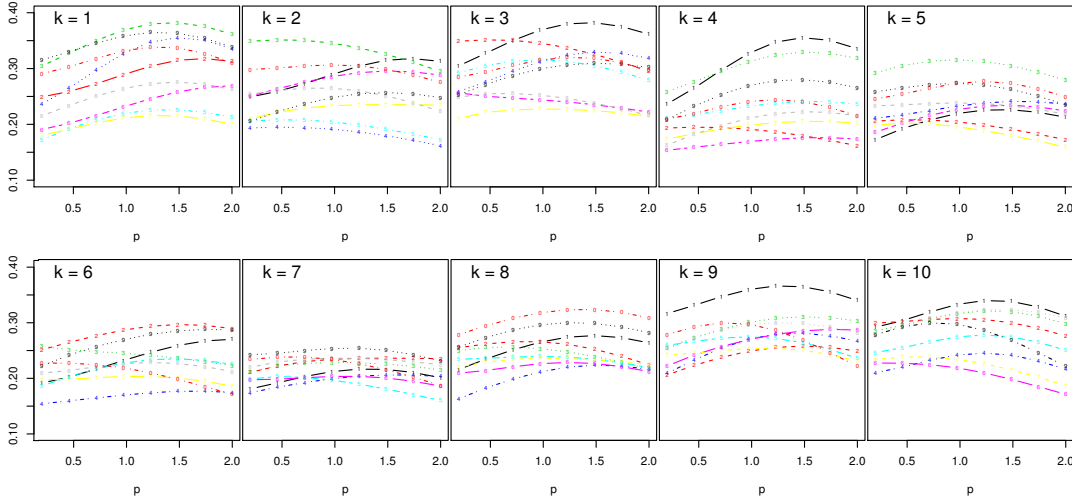
Figure 5.4: Visual representation of the estimated factors-factors dependences, for $M = 10$, on the period 2000–2009. The correlation (5.14a) is shown for every factor k with all other factors $l \neq k$, as a function of the order p of the absolute moment considered.

obtained from the procedure described in the previous section). On each figure corresponding to a value of k, the curves represent the values of Eq. (5.14a) for all $l \neq k$, as a function of p. We observe that the curves depart from the Gaussian benchmark for which the correlation computed and shown would all be 0. Anticipating over the forthcoming theoretical description, we also see that the curves are not compatible with factors having a lognormal volatility, in which case all the curves would be exactly horizontal. The concavity of the curves is a signature of non-Gaussianity in log-volatilities, while their splitting (in particular as $p \to 0$) reveals a complex structure that we will uncover using a model in Section 5.4 below.

This "naked eye" analysis cannot be reproduced for large matrices like the factors-residuals and residuals-residuals correlations, for which it turns much more convenient to use a spectral approach. Spectral decomposition of the corresponding matrices of coefficients reveals eigenmodes of amplitude fluctuations. Fig. 5.5(a) illustrates the content of the spectra for the factor-factor coefficients (top), the residual-residual coefficients (middle), and the singular values of the rectangular matrix of the factors-residuals cross-correlations (bottom), for the period 2000–2009. The first three values of each spectrum are shown in dark, while the others are represented as a grey bulk. Notice how the bulk widens for the res-res spectrum as the moment p increases: this is due to the high dimensionality of this matrix, and the large resulting noise.

This model-free description highlights the presence of few dominant modes of *amplitude* (not to be confused with the factors of *linear* correlations), in particular a common mode. This is confirmed by the study of the corresponding eigenvectors on Fig. 5.5(b): the first eigenvector of both the factor-factor amplitudes and the residual-residual amplitudes has all components with same sign, revealing a collective mode driving the amplitudes of the factors and the residuals. The second eigenvector of res-res reveals "finance against all" behavior in the given period. The third eigenvector of res-res reveals "energy against all" behavior (energy + utilities). This mode is not observed on the fact-res correlations, thus indicating
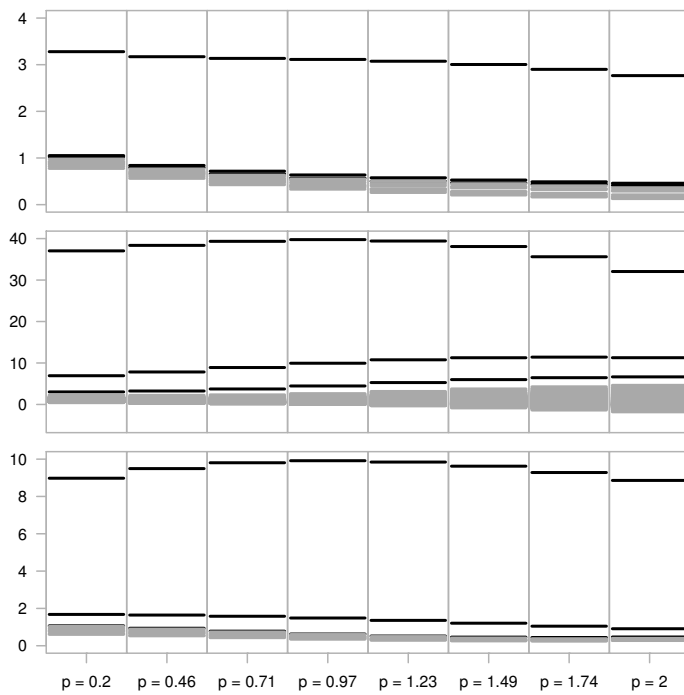
(a) Singular values spectra (vertical axis), for several values of p (horizontal axis). The three largest values are in black, while the remaining values are shown as a grey bulk.
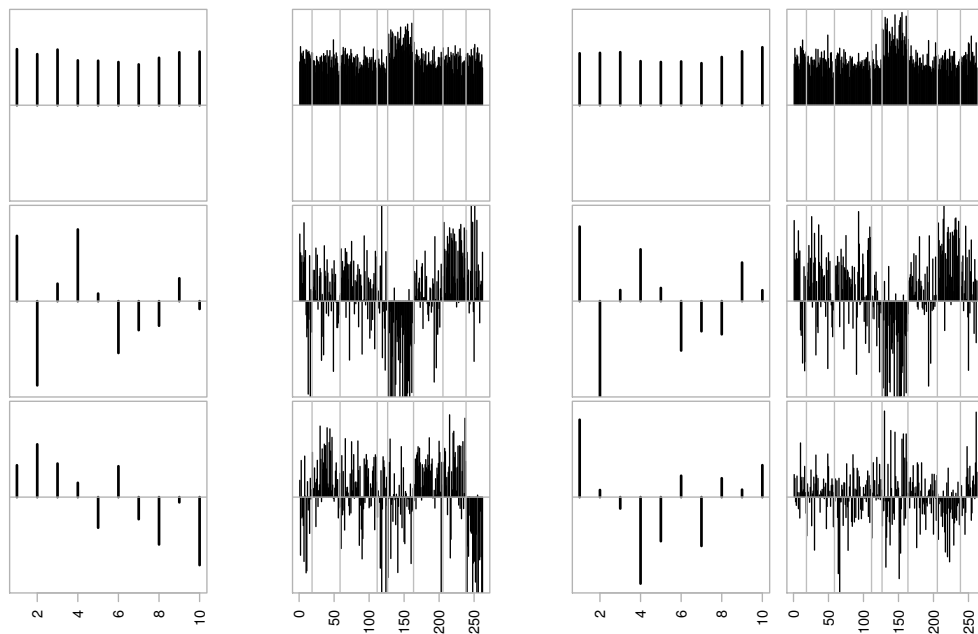
(b) Corresponding eigenvectors: first three, from top to bottom, averaged over p.

Figure 5.5: $M = 10$, 2000-2009. Spectral decomposition of the factor-factor, residual-residual, and factor-residual correlations.
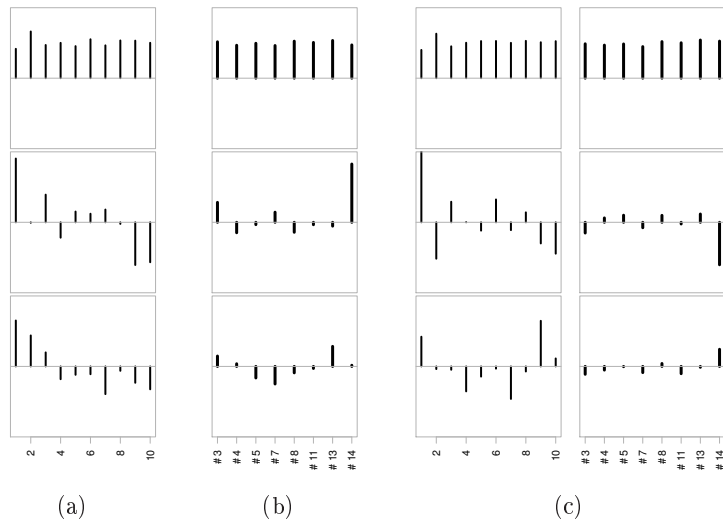
Figure 5.6: $M = 10$, 2000-2004.
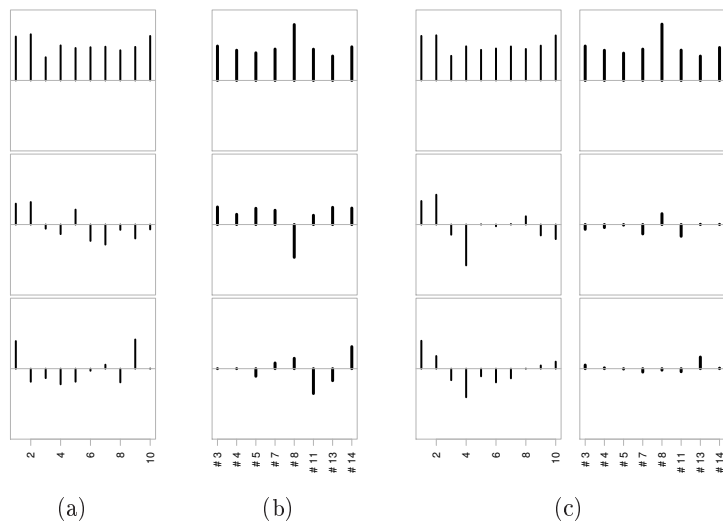
Figure 5.7: $M = 10$, 2005-2009
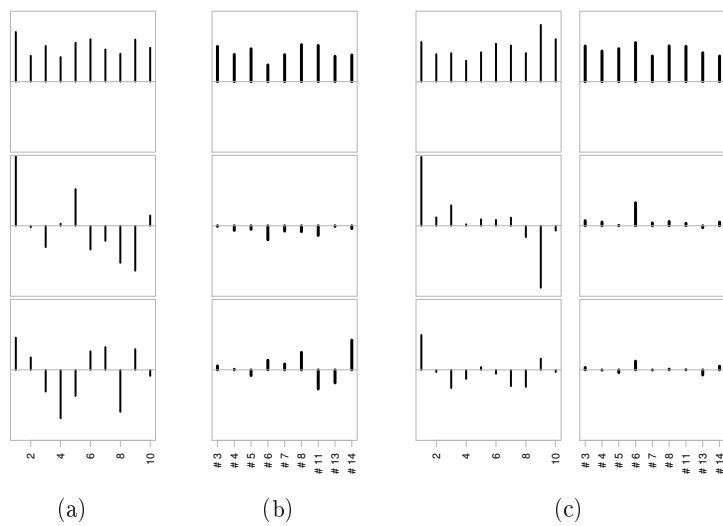
Figure 5.8: $M = 10$, 2009-2012

that the impact of energy on the volatility is only present in the residuals, and not shared by the linear factors.

Of course this structure is not stable over time, where it is expected and observed that the number and meaning of significant modes varies according to the studied period. The spectral decomposition for the three sub-periods 2000–2004, 2005–2009, 2009–2012 is shown in Figs. 5.6, 5.7, 5.8, respectively. We represent the eigenvectors of the factor-factor (a), residual-residual (b), and the left/right singular vectors of the factor-residual (c). When applicable, we have averaged the components of the vectors over Bloomberg sector in order to see if a structure emerges (see Tab. 5.1 for the labels). As a general observation, the collective mode of amplitude fluctuations is stable over time, and strongly dominant over any other mode. As can be seen in Fig. 5.7, it seems to be impacting/impacted by the financial sector more strongly in the 2005–2009 period (containing the financial crisis) whereas in other periods it is almost uniformly spread over sectors. This financial sector is also a stand-alone mode of fluctuations in the crisis period, as revealed by the structure of the second eigenvector in Figs. 5.7(b) and 5.7(c), where the stocks in that sector oscillate with opposing phase with respect to the other sectors. In the pre-crisis period, the second relevant mode is rather composed of commodities. Indeed, the second eigenvector features the opposition of utilities (# 14), energy (# 7) and communications (# 3) against the rest. This already illustrates that the structure is not stable over time, what is confirmed by the post-crisis period where there is not even a clear signature for the existence of a significant second mode (the apparent structure in the second mode of Fig. 5.8(c) is not significant because the sector # 6 is populated with a single individual).

5.4 Modeling the volatility content

The univariate and multivariate pseudo-empirical properties of the reconstructed factors series motivate the construction of a model that integrates the following characteristics:

- all factors amplitudes are driven by a common mode;

- all residuals amplitudes are driven by *that same* common mode;

- factors and residuals series are non-Gaussian, and their amplitudes are not lognormal either.

The spectral analysis suggests the presence of two volatility drivers σ_0 and $\sigma_{0'}$ in the factors and the residuals.[3] Concretely, we aggregate the volatility determinants multiplicatively as follows:

$$f_k = \epsilon_k \exp(A_{k0}\omega_0) \exp(A_{k0'}\omega_{0'}) \exp(A_{kk}\omega_k) \qquad (5.15a)$$

$$e_j = \eta_j \exp(B_{j0}\omega_0) \exp(B_{j0'}\omega_{0'}) \exp(B_{jj}\widetilde{\omega}_j), \qquad (5.15b)$$

where the ω's are stochastic log-volatilities (all independent of each other and independent of the ϵ's and η's), and the parameters A's and B's weight the contribution of every volatility

---

[3] It is very easy to add a possible third driver $\sigma_{0''}$ in the equation (5.15b) for the residuals, but its statistical significance is not granted and the estimation of the associated parameters could turn quite noisy.

mode. In particular, we expect A_{k0} and B_{k0} to reflect approximately the coordinates of the first eigenvectors of the "log-abs" correlations on Fig. 5.6–5.8(a) and Fig. 5.6–5.8(b) respectively, and $A_{k0'}$ and $B_{k0'}$ the coordinates of the second eigenvectors. A_{kk} and B_{jj} are the standard-deviations of the remaining log-volatilities not explained by the common drivers (if any).

In the next subsection, we estimate a restricted model where only a single volatility driver is active.

5.4.1 A dominant volatility mode

The simplest improvement over the independent factors assumption, while keeping uncorrelatedness, is to allow for a common fluctuation of amplitudes. The multiplicative structure of Eqs. (5.15) is better rendered using the summation in exponential notation:

$$f_k = \epsilon_k \exp(A_{k0}\omega_0 + A_{kk}\omega_k) \tag{5.16a}$$

$$e_j = \eta_j \exp(B_{j0}\omega_0 + B_{jj}\widetilde{\omega}_j) \tag{5.16b}$$

with ϵ_k, η_j Gaussian with variance such that Eqs. (5.6) hold, and we have dropped the tildes and hats when no confusion was possible. It is important to keep the intuitive meaning of the log-volatility factors in mind:

$\omega_0 = $ dominant and common driver of log-volatility across factors and residuals

$\omega_1 = $ log-volatility of the index (market linear mode f_1), net of ω_0 contribution

and the subsequent $\omega_k, \widetilde{\omega}_k, (k > 1)$ characterize the "residual volatility" not explained by the common driver ω_0 in the amplitude of the factors f_k and residuals e_j.

The model is completely characterized from a probabilistic point of view when the law of the log-volatilities is specified. Rather than using their Probability Distribution Functions, we resort to the Moment Generating Function (MGF) $M_\omega(p) \equiv \mathbb{E}[\exp(p\omega)]$, as we find it more appropriate to characterize the ω's by their moments, in the perspective of calibration. The p-dependence of the curves in Fig. 5.4 suggests that the non-Gaussianity in the log-volatilities is approximately homogeneous across the factors, and thus possibly due to the common volatility driver ω_0 only, while the residual volatilities ω_k and $\widetilde{\omega}_j$ can be taken as Gaussian. For those, the MGF is $M_{\mathrm{G}}(p) = \exp(p^2/2)$. But in the general case, developing in cumulants, M_{ω_0} is the exponential of a polynomial. Typically, with

$$\mathbb{E}[\omega_0] = 0 \qquad \mathbb{E}[\omega_0^2] = 1 \qquad \mathbb{E}[\omega_0^3] = \zeta_0 \qquad \mathbb{E}[\omega_0^4] = 3 + \kappa_0$$

one gets

$$M_{\omega_0}(p) = \exp\left(\frac{1}{2}p^2 + \frac{\zeta_0}{6}p^3 + \frac{\kappa_0}{24}p^4\right). \tag{5.17}$$

Parameters estimation

The model defined by Eqs. (5.5,5.16) contains the following parameters: MN linear weights β_{ki} (previously estimated, see section 5.2), M coefficients A_{k0} and N coefficients B_{j0} of

exposure to the common volatility mode, the standard-deviations A_{kk} and B_{jj} of the residual log-volatilities, and the skewness ζ_0 and kurtosis κ_0 of the volatility mode ω_0. So there are overall $NM + 2(N + M) + 2$ parameters, for a dataset of size NT. More importantly the number of parameters is only marginally enhanced with respect to a typical linear factor model (where only the NM linear weights enter into account): only $2(N + M + 1)$ new parameters intended to improve the description of all pairwise dependences coefficients.

In order to estimate the latter we express the model predictions for the observables defined in Eq. (5.14). It is convenient to introduce the following ratio of MGF's:

$$\Phi_0(a, b) = \frac{M_{\omega_0}(a+b)}{M_{\omega_0}(a) M_{\omega_0}(b)}, \tag{5.18}$$

as well as the Gaussian equivalent $\Phi_{\mathrm{G}}(a, b) = \exp(ab)$. In logarithmic form, $\phi_0(a, b; p) = \frac{1}{p^2} \ln \Phi_0(pa, pb)$ is a polynomial in p when expanding in cumulants. Indeed, with Eq. (5.17),

$$\phi_0(a, b; p) = ab + \frac{p}{2}\zeta_0(a^2 b + ab^2) + \frac{p^2}{12}\kappa_0(2a^3 b + 3a^2 b^2 + 2ab^3)$$

and $\phi_{\mathrm{G}}(a, b; p) = ab$ is independent of p. Then the theoretical prediction for the matrix elements can be computed analytically:

$$\frac{1}{p^2} \ln \frac{\mathbb{E}[|f_k|^p |f_l|^p]}{\mathbb{E}[|f_k|^p]\,\mathbb{E}[|f_l|^p]} = \phi_0(A_{k0}, A_{l0}; p) + \left(\gamma(p) + A_{kk}^2\right)\delta_{kl} \tag{5.19a}$$

$$\frac{1}{p^2} \ln \frac{\mathbb{E}[|f_k|^p |e_i|^p]}{\mathbb{E}[|f_k|^p]\,\mathbb{E}[|e_i|^p]} = \phi_0(A_{k0}, B_{i0}; p) \tag{5.19b}$$

$$\frac{1}{p^2} \ln \frac{\mathbb{E}[|e_i|^p |e_j|^p]}{\mathbb{E}[|e_i|^p]\,\mathbb{E}[|e_j|^p]} = \phi_0(B_{i0}, B_{j0}; p) + \left(\gamma(p) + B_{ii}^2\right)\delta_{ij} \tag{5.19c}$$

where

$$\gamma(p) = \frac{1}{p^2} \ln \frac{\mathbb{E}\big[|\epsilon|^{2p}\big]}{\mathbb{E}[|\epsilon|^p]^2} = \frac{1}{p^2} \ln \left(\sqrt{\pi} \frac{\Gamma(\frac{1}{2} + p)}{\Gamma(\frac{1+p}{2})^2} \right)$$

is the normalized $2p$-moment of the absolute value of Gaussian variables.

Clearly, if even ω_0 was Gaussian, and the diagonal elements were regular, the matrices described in Eqs. (5.19a) and (5.19c) would be trivially of rank 1, and the identification of $A_{\cdot 0}$ and $B_{\cdot 0}$ with the first eigenvectors of the corresponding matrices would be straightforward. Non-Gaussianity and specificities on the diagonal perturb this identification, but the overall picture is essentially the same story, as we will shortly show with the calibration results.

The model estimation procedure is as follows (the linear weights β are previously estimated). As discussed just above, there are $2(N + M + 1)$ parameters to be estimated. Because the equations (5.19) are coupled through (5.19b), all parameters should in principle be estimated jointly. The corresponding optimization program would however be computer intensive, and the stability of the solution would not be granted in such a large dimensional space. We proceed stepwise instead, by first estimating the parameters $A_{k0}, A_{kk}, \zeta_0, \kappa_0$ using the fac-fac predictions (5.19a), and then estimate the remaining parameters B_{i0}, B_{ii} from the res-res *and* fac-res correlations for consistence. Note that there is an overall sign degeneracy, as well as a sign indetermination for the idiosyncratic parameters A_{kk}, B_{ii} (which we then arbitrarily take positive).

	2000–2004	2005–2009	2009–2012
ζ_0	-0.072	-1.492	0.607
κ_0	-0.129	-1.916	-0.608

Table 5.2: $M = 10$. Estimated non-Gaussianity parameters.

1. Estimate A_{k0}, A_{kk} and the non-Gaussianity parameters from Eq. (5.19a):

$$\min \left\{ \sum_p \sum_{k,l} \left(\frac{1}{p^2} \ln \frac{\langle |F_{tk}F_{tl}|^p \rangle}{\langle |F_{tk}|^p \rangle \langle |F_{tl}|^p \rangle} - \frac{1}{p^2} \ln \frac{\mathbb{E}[|f_k|^p |f_l|^p]}{\mathbb{E}[|f_k|^p] \, \mathbb{E}[|f_l|^p]} \right)^2 \right\} \tag{5.20}$$

The sum on p runs over eight values between $p = 0.2$ and $p = 2$ (see x-axis of Fig. 5.5(a)) and is crucial here to the estimation of the non-Gaussianity parameters, since the loss function is independent of p for Gaussian variables. This amounts to performing a best (joint!) quadratic fit of the curves similar to Fig. 5.4, for each period.

2. Estimate B_{i0} from Eq. (5.19b):

$$\min \left\{ \sum_{k,i} \left(\frac{1}{p^2} \ln \frac{\langle |F_{tk}E_{ti}|^p \rangle}{\langle |F_{tk}|^p \rangle \langle |E_{ti}|^p \rangle} - \frac{1}{p^2} \ln \frac{\mathbb{E}[|f_k|^p |e_i|^p]}{\mathbb{E}[|f_k|^p] \, \mathbb{E}[|e_i|^p]} \right)^2 \right\} \tag{5.21a}$$

or jointly with B_{ii} from Eq. (5.19c), as the vector solution of

$$\min \left\{ \sum_{i,j} \left(\frac{1}{p^2} \ln \frac{\langle |E_{ti}E_{tj}|^p \rangle}{\langle |E_{ti}|^p \rangle \langle |E_{tj}|^p \rangle} - \frac{1}{p^2} \ln \frac{\mathbb{E}[|e_i|^p |e_j|^p]}{\mathbb{E}[|e_i|^p] \, \mathbb{E}[|e_j|^p]} \right)^2 \right\} \tag{5.21b}$$

(here it is too intensive to calculate the optimum in the N-dimensional space for all values of p so we take a single value, typically $p = 1$ if we intend to reproduce best absolute correlations, or $p = 2$ if we favor quadratic correlations).

The convergence is ensured by starting close to the solution, namely taking as prior the first eigenvector of the corresponding matrices. The calibration results are given graphically on Figs. 5.9, 5.10, 5.11, where we show, separately for each sub-period, the estimated parameters A_{0k} and B_{0j}. As expected and discussed above, they are very close to the first eigenvector of the corresponding matrix of "log-abs" correlations shown in Figs. 5.6, 5.7 and 5.8. We also show the resulting parameters A_{kk} and B_{jj}: notably, some factors k seem to have all their volatility explained by the common driver ω_0 so that there is no left residual volatility. The estimated values of the non-gaussianity parameters of the log-volatilities are reported in Tab. 5.2. noticeably, the kurtosis of the common driver ω_0 is found to be negative in every period: the log-volatility is less kurtic than a Gaussian. This was already revealed by the concavity of the curves in Fig. 5.4.

Reconstructing the volatility factors

We are interested in reconstructing the series of the common volatility mode ω_{t0} out of the model equations and the estimated parameters. Similarly to what was done in Sect. 5.2 to

(a) A_{k0} and A_{kk}
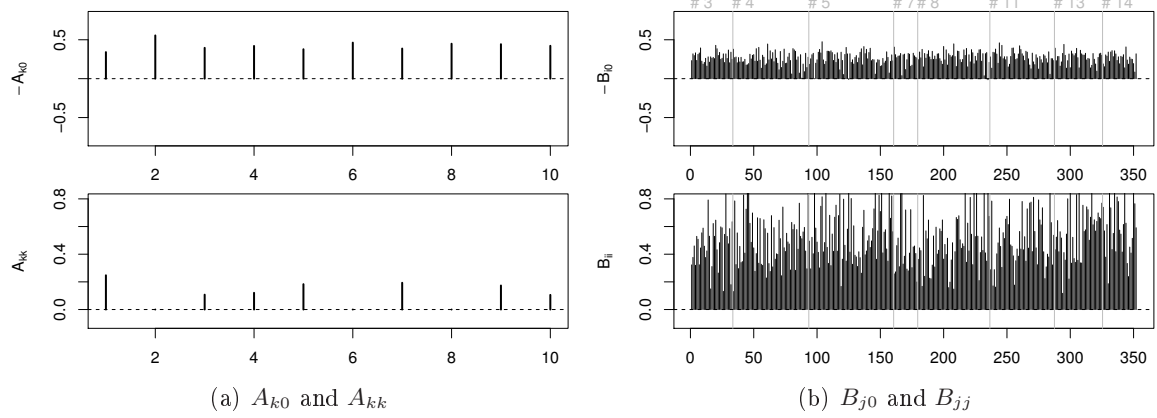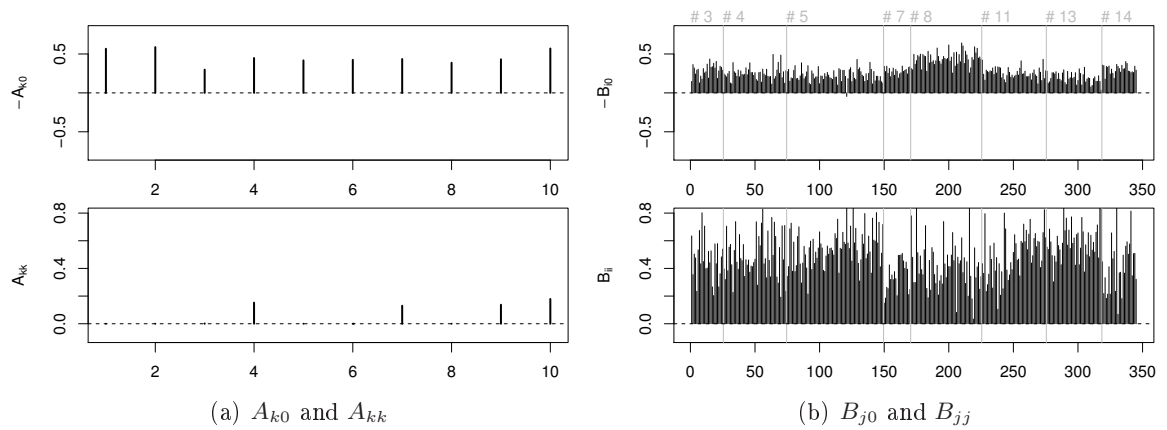
(b) B_{j0} and B_{jj}

Figure 5.9: $M = 10$, 2000-2004.

(a) A_{k0} and A_{kk}

(b) B_{j0} and B_{jj}

Figure 5.10: $M = 10$, 2005-2009

(a) A_{k0} and A_{kk}

(b) B_{j0} and B_{jj}

Figure 5.11: $M = 10$, 2009-2012
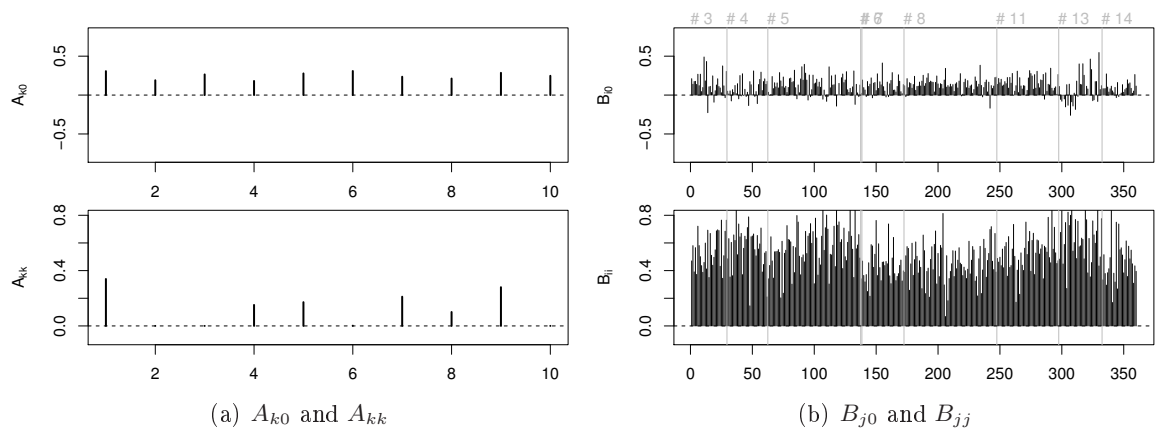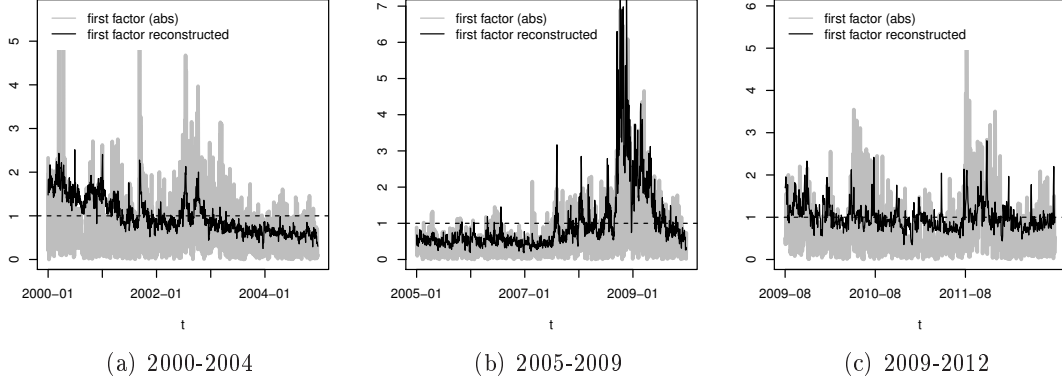
(a) 2000-2004   (b) 2005-2009   (c) 2009-2012

Figure 5.12: Original and reconstructed first factor ($M = 10$)

recover the series of linear factors, we perform here two date-by-date regressions motivated by the Eqs. (5.16):

$$\ln |\mathrm{F}_{tk}| - \langle \ln |\mathrm{F}_{tk}| \rangle = \omega_{t0} A_{k0} + A_{kk} \omega_{tk}$$
$$\ln |\mathrm{E}_{tj}| - \langle \ln |\mathrm{E}_{tj}| \rangle = \omega_{t0} B_{j0} + B_{jj} \widetilde{\omega}_{tj}$$

Whereas the first regression is performed over only M variables A_{k0}, the second one is realized over the N variables B_{j0} and thus much less noisy (we will always use the second determination in the following). The overlap of the series of ω_0 estimated with the two regressions is good, with a correlation coefficient between 0.55 and 0.75 depending on the period studied. We show in Fig. 5.12 the series of first factor straight from the linear program F_{t1}, and also reconstructed from the procedure above as $\exp(A_{10}\omega_{t0})$, after estimation of the parameters.

Predicted non-linear dependences

As a consistency check of the estimation procedure, we analyze the model prediction with estimated parameters and compare them with empirical measurements of the same quantities. Of particular interest are the quadratic correlations and the diagonal copulas, whose anomalies observed in a previous study [Chicheportiche 2012b] motivated the present model.

The quadratic correlations can be computed from the model definition and write

$$
\begin{aligned}
\mathbb{E}[x_i^2 x_j^2] = & \sum_{kl} \left(\beta_{ki}^2 \beta_{lj}^2 + 2\beta_{ki}\beta_{kj}\beta_{li}\beta_{lj} \right) \Phi_0(A_{k0}, A_{l0}; 2) \left(\tfrac{1}{3} \cdot 3 \cdot \Phi_{\mathrm{G}}(A_{kk}, A_{ll}; 2) \right)^{\delta_{kl}} \\
& + (1 + 2\delta_{ij}) \left(1 - \sum_l \beta_{li}^2 \right) \sum_k \beta_{kj}^2 \Phi_0(A_{k0}, B_{i0}; 2) \\
& + (1 + 2\delta_{ij}) \left(1 - \sum_l \beta_{lj}^2 \right) \sum_k \beta_{ki}^2 \Phi_0(A_{k0}, B_{j0}; 2) \\
& + \left(1 - \sum_l \beta_{li}^2 \right) \left(1 - \sum_l \beta_{lj}^2 \right) \Phi_0(B_{i0}, B_{j0}; 2) \left(3\Phi_{\mathrm{G}}(B_{ii}, B_{jj}; 2) \right)^{\delta_{ij}}.
\end{aligned}
\tag{5.22}
$$

When all parameters A, B are zero, the prediction for Gaussian factors and residuals is retrieved: $\mathbb{E}[x_i^2 x_j^2] = 1 + 2\mathbb{E}[x_i x_j]^2$. We illustrate in the left panel of Fig. 5.14 a scatter
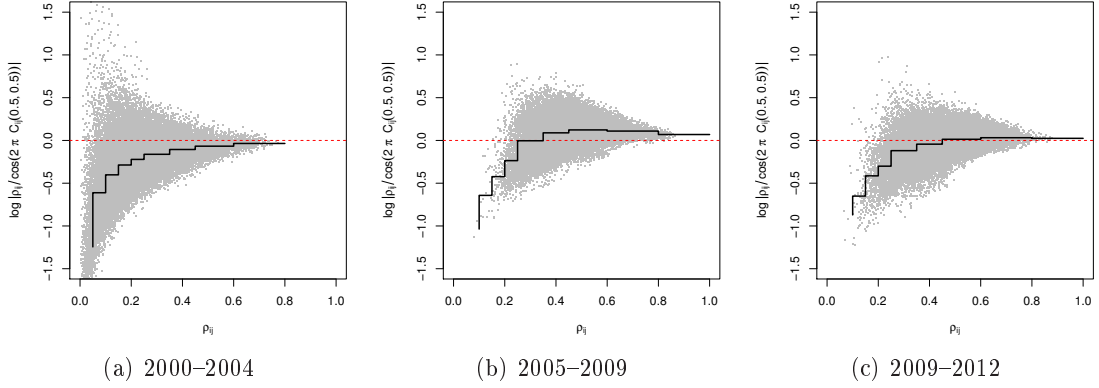
(a) 2000–2004   (b) 2005–2009   (c) 2009–2012

Figure 5.13: Calibrated model. $\ln|\rho/\rho^{(\mathrm{B})}|$ vs ρ, see Eq. (5.4).

plot of the left-hand side (calibrated) versus the right-hand side (empirical) of Eq. (5.22), for all periods. They show a good agreement of model and sample quadratic correlations. Furthermore, the middle and right panels of the same figure illustrate the fact that the pairs of stock returns cannot be described by a bivariate Student distribution, for which a regular curve should be observed instead of the scattered cloud in the plane of quadratic vs linear correlations, a conclusion already drawn in chapter II.4 but which the present model highlights.[4]

The middle point $C(\frac{1}{2}, \frac{1}{2})$ of the copula, which was shown on Fig. 5.2 to be incompatible with any elliptical prediction, is also better explained by the model. Although an analytical expression linking $C(\frac{1}{2}, \frac{1}{2})$ to the model parameters is out of reach, it is possible to reproduce its predicted value by simulating long time series according to the model with estimated parameters.[5] Fig. 5.13 shows the obtained coefficients in the same form as the empirical measurements of Fig. 5.2, which they have to be compared with. The results are in very good agreement, and emphasize the capacity of our non-Gaussian factor model to cope with the non-trivial behavior of the medial point of the copula.

This is confirmed and even generalized by the reproduction of the copula along the whole diagonals. Figs. 5.15, 5.16 and 5.17 compare empirically measured and model-predicted values of the quantities

$$\Delta_{\mathrm{d}}(p) = \frac{C(p, p) - C_{\mathrm{G}}(p, p)}{p\,(1 - p)} \quad \text{and} \quad \Delta_{\mathrm{a}}(p) = \frac{C(p, 1 - p) - C_{\mathrm{G}}(p, 1 - p)}{p\,(1 - p)}$$

versus p, for several values of the linear correlation and over the three periods. A direct visual comparison reveals that the main non-trivial qualitative features of the empirical diagonal

---

[4] Notice that the choice of p in the estimation procedure of the parameters B_{i0}, $B_{i0'}$ and B_{ii} is important here. Although the model has identical predictions for any p, estimation biases and errors are in practice different for low moments $p \approx 0.2$ or high moments $p \approx 2$. Obviously, best fits for the quadratic correlations are obtained with $p = 2$ since in this case the same quantities appear in Eq. (5.22) and in the loss function (5.21).

[5] The non-Gaussian series of log-volatility ω_{t0} is generated as independent realizations of a Beta distribution whose coefficients are determined so that the first four moments match those of ω_0. This class of distributions allows for negative kurtosis. It is known that the realizations of volatility exhibit strong persistence, a characteristic that our simulated series do not reproduce. This however does not generate a bias in the obtained coefficients, but rather makes them "too un-noisy".
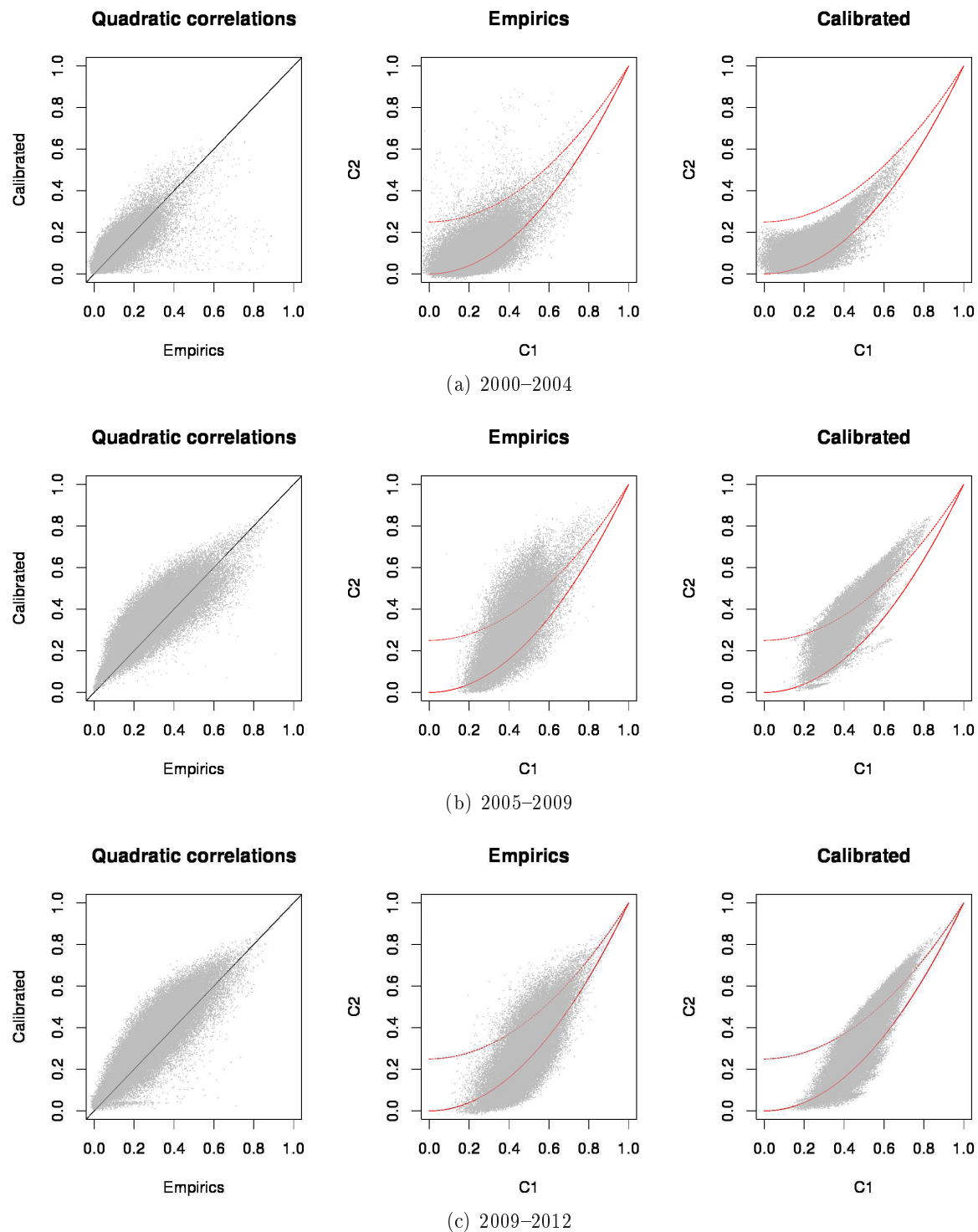
(a) 2000–2004

(b) 2005–2009

(c) 2009–2012

Figure 5.14: **Left:** calibrated vs sample quadratic correlations. **Middle:** sample quadratic correlations vs sample linear correlations; **Right:** calibrated quadratic correlations vs calibrated linear correlations. Two benchmark curves are added in red: the Gaussian case (solid) and the Student case with $\nu = 5$ d.o.f. (dashed)

copulas are reproduced by our model: the evolution of the concavity as ρ changes, some of the tail behaviors, and the middle-point behavior as already discussed above. One may note however that asymmetries $u \leftrightarrow 1 - u$ are not reproduced by our fully symmetric model.

Before proceeding to the out-of-sample evaluation of the model in Sect. 5.5, we briefly present how the model is improved (though marginally) by introducing a second volatility driver $\omega_{0'}$, as suggested by the spectral analysis of Sect. 5.3.

5.4.2 Another volatility driver

The spectral analysis of the factor and residual absolute correlations has revealed that there exists a small but significant second mode of volatility. The model (5.16) can be improved accordingly in order to account for this additional source of collective amplitude fluctuations:

$$f_k = \epsilon_k \exp(A_{k0}\omega_0 + A_{k0'}\omega_{0'} + A_{kk}\omega_k) \tag{5.23}$$

$$e_j = \eta_j \exp(B_{j0}\omega_0 + B_{j0'}\omega_{0'} + B_{jj}\widetilde{\omega}_j), \tag{5.24}$$

and as a consequence the RHS of Eqs. (5.19) get an additional term each, namely $\phi_{0'}(A_{k0'}, A_{l0'}; p)$, $\phi_{0'}(A_{k0'}, B_{i0'}; p)$ and $\phi_{0'}(B_{i0'}, B_{j0'}; p)$, respectively.

The whole estimation procedure runs identically. However, for the determination of the parameters A_{k0} and $A_{k0'}$, the reduced number of observations ($M(M-1)/2$ factor-factor correlations, times 8 values of p) provides only a low resolution, and the minimization program does not succeed in distinguishing the two volatility drivers: it outputs an hybrid where both ω_0 and $\omega_{0'}$ contribute to the same mode (this issue is not relevant in the determination of the parameters B_{j0}, $B_{j0'}$ thanks to the sufficient number of observations). In order to break the degeneracy and "orthogonalize" the modes, we add an overlap term $(\sum_k A_{k0}A_{k0'})^2$ in the cost function Eq. (5.20).

As an example, we report in Fig. 5.18 the results for the period 2000–2004. As expected, the parameters A_{k0} and $A_{k0'}$ are very close to the first two eigenvectors of the factor-factor "log-abs" correlation matrix (Fig. 5.6(a)), and the parameters B_{j0}, $B_{j0'}$ look like the first two eigenvectors of the residual-residual matrix (Fig. 5.6(b), averaged over sectors). Clearly, taking this additional second volatility driver into consideration improves the theoretical description of the returns. We illustrate this on Fig. 5.18(c) where we show how ω_0 et $\omega_{0'}$ contribute respectively to the volatility of the market mode of linear correlation, f_1.

5.5 Out-of-sample analysis

All the results presented above are "in-sample", in the sense that we have shown the predicted dependence coefficients with estimated parameters on a period and compared them to the realized coefficients in that same period. The ultimate test for a model that aims at describing joint financial returns (and more generally of any risk model), is to perform good "out-of-sample", i.e. use a model calibrated on a period to predict a (average) behavior in a *subsequent* period. Refer to Ref. [Potters 2009]!

We consider a long period 2000–2009 on which we perform an In-sample/Out-of-sample analysis over sliding windows ($N = 262$ returns series are kept, see Tab. 5.1). We rely on the procedure used in Ref. [Potters 2009].

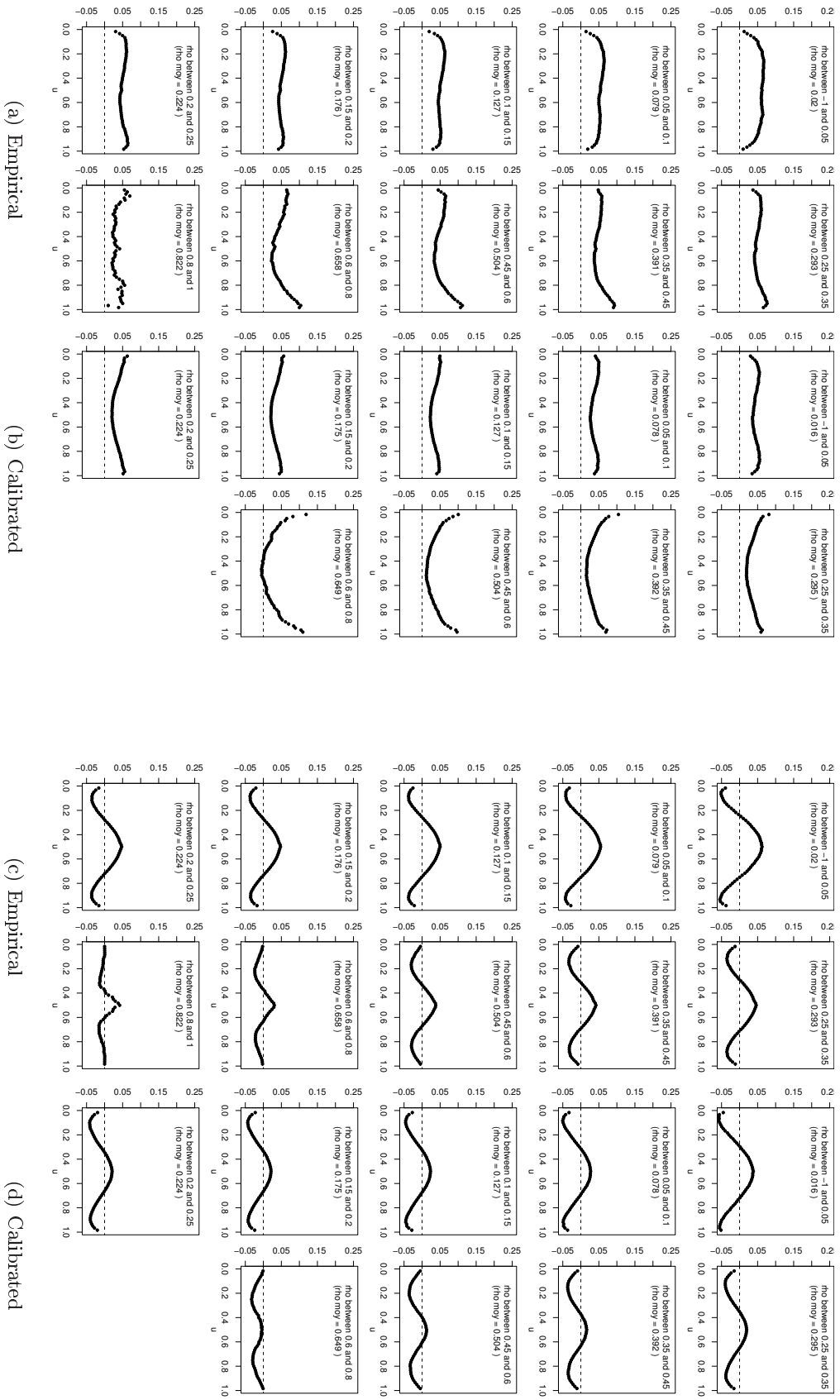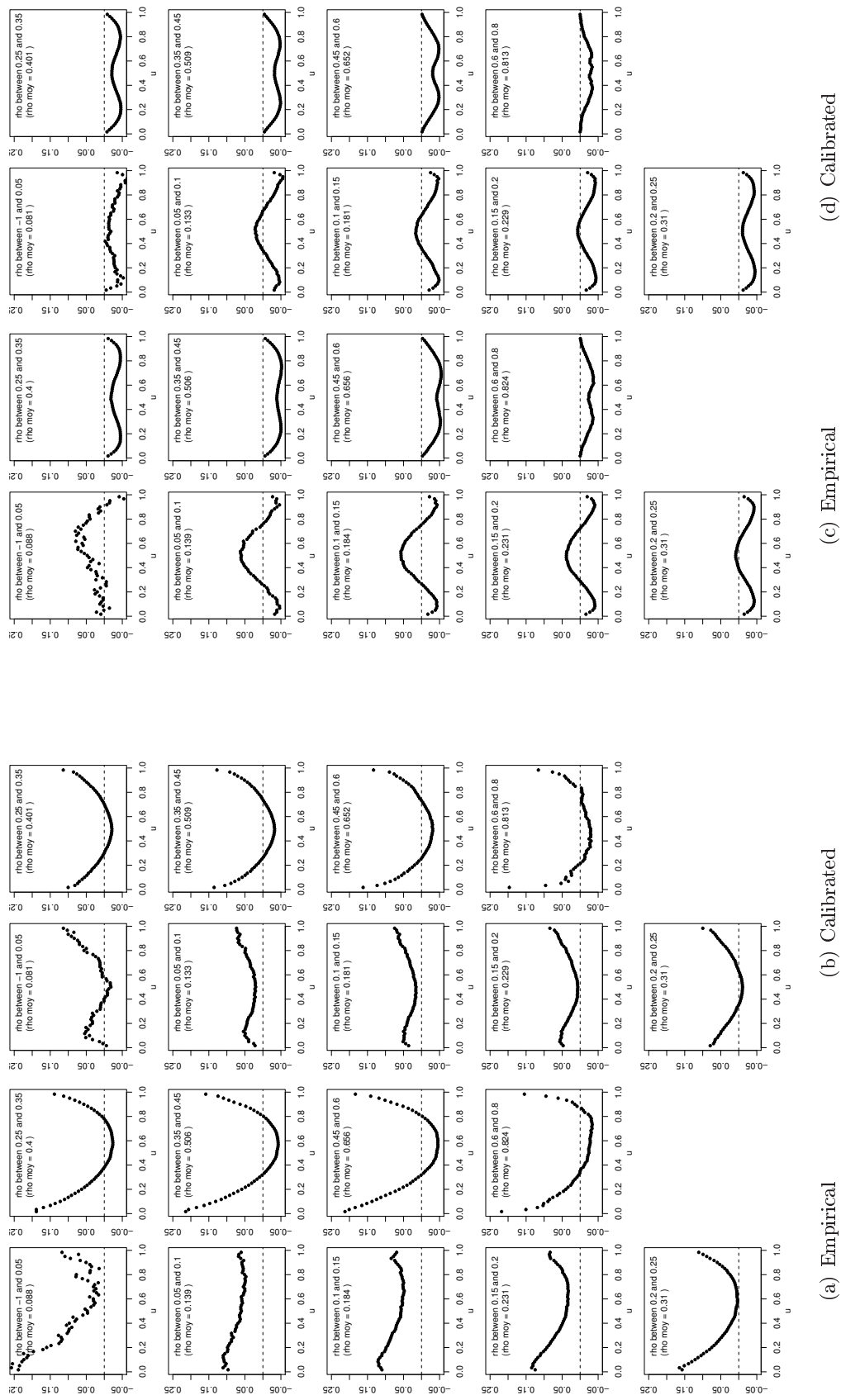Figure 5.15: 2000–2004, diagonal (left) and anti-diagonal (right)

Figure 5.16: 2005–2009, diagonal (left) and anti-diagonal (right)
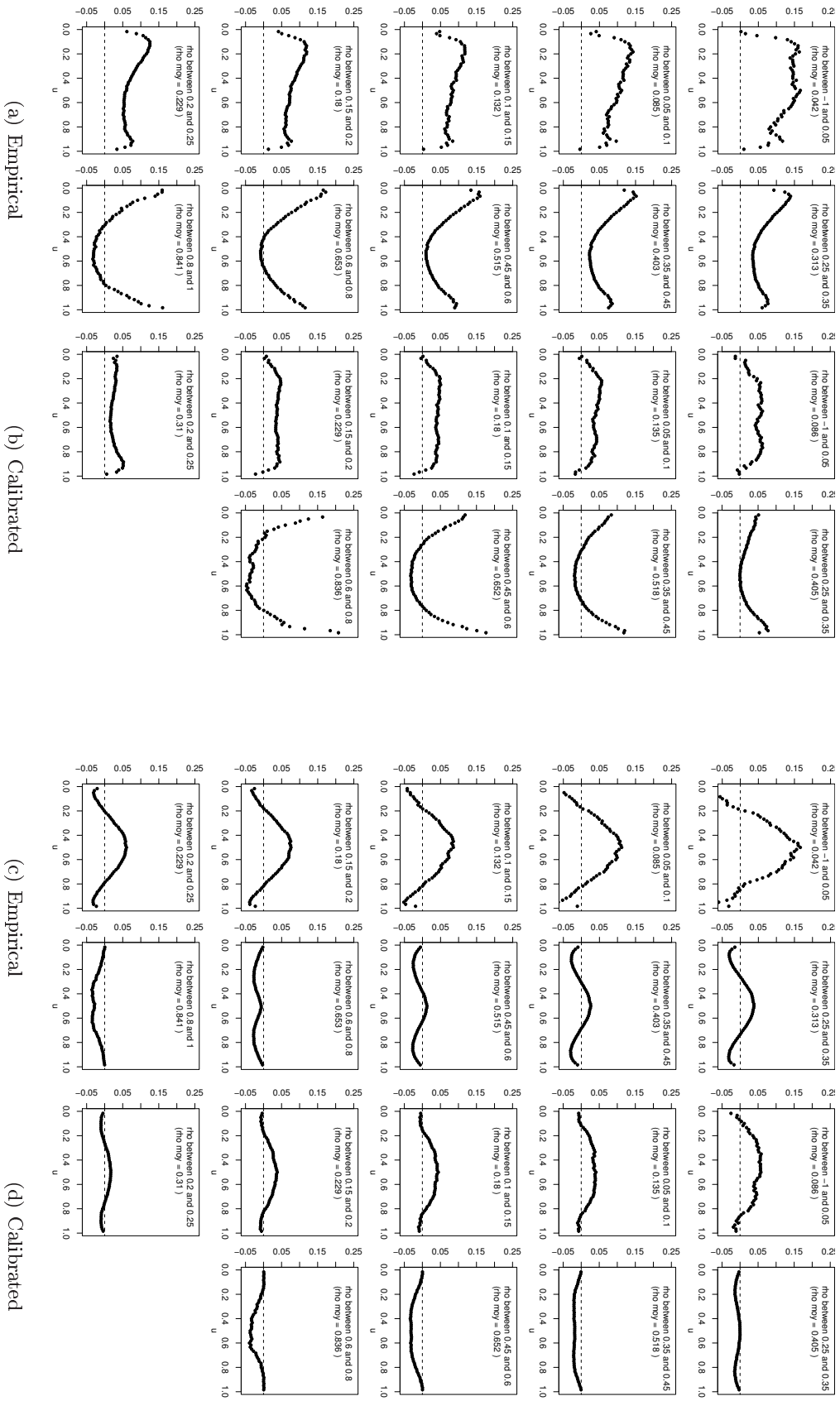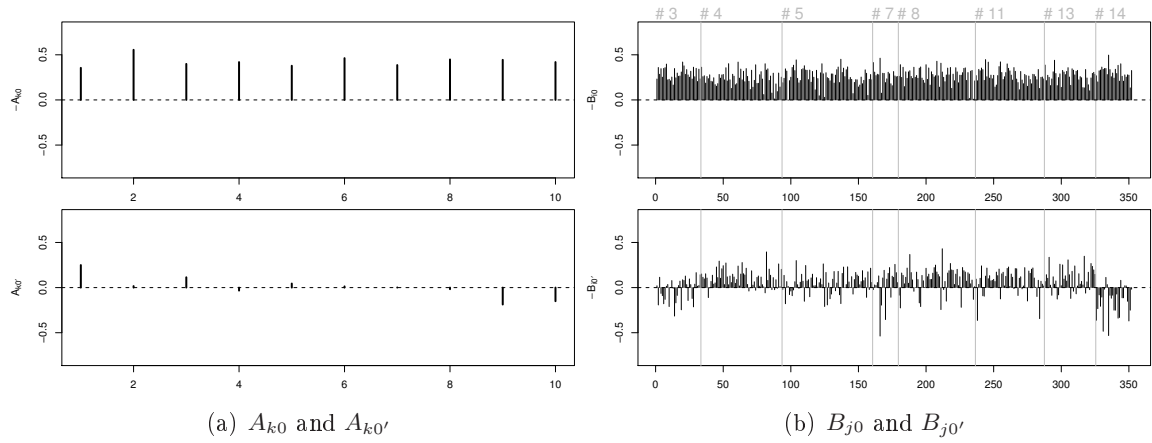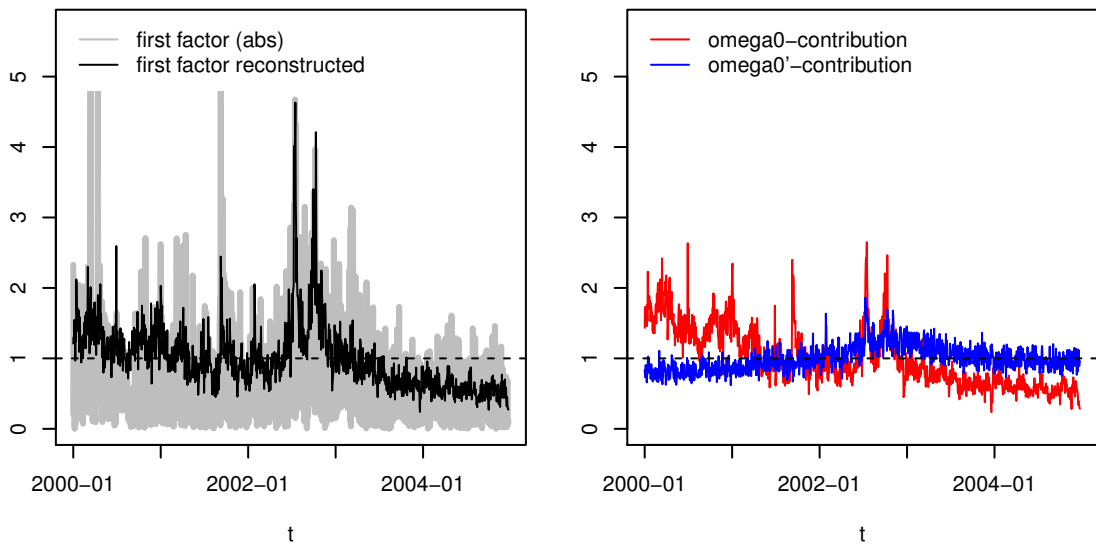
Figure 5.17: 2009–2012, diagonal (left) and anti-diagonal (right)

(a) A_{k0} and $A_{k0'}$

(b) B_{j0} and $B_{j0'}$

(c) Index amplitudes reproduced with two volatility drivers ω_0 and $\omega_{0'}$, to be compared with Fig. 5.12(a)

Figure 5.18: $M = 10$, 2000–2004

1. The model is calibrated in windows of $T^{\text{IS}} = 2N = 524$ days. An optimal portfolio is built and a corresponding risk measure is computed over the window used for estimation: this is the In-sample risk. We consider below two kinds of risks corresponding to two different portfolios: (i) the quadratic risk of a basket of returns, that will assess the quality of the *linear* elements of the model; and (ii) the quadratic risk of a basket of (centered and normalized) absolute returns, that will assess the quality of the *volatility* description of the model.

2. The same risk measures are computed Out-of-sample on a small period of $T^{\text{OS}} = 59$ days (three months) following the estimation period.

3. The sliding lags are chosen so that the control samples are non-overlapping, i.e. at dates $\tau = T^{\text{IS}} + n \times T^{\text{OS}} + 1$, $n = 0, 1, 2, \ldots$. Sliding windows will be indexed with parenthesis notation '(τ)', in order to avoid confusion with regular time stamps t of the running dates.

5.5.1 Linear correlations

For a given covariance matrix ρ, optimal portfolio weights can be computed in the sense of Markowitz:

$$\boldsymbol{w}^*(\tau) = \frac{\rho^{-1} \boldsymbol{g}(\tau)}{\boldsymbol{g}(\tau)^\dagger \rho^{-1} \boldsymbol{g}(\tau)} \tag{5.25}$$

where we consider an omniscient stationary predictor of returns

$$g_i(\tau) = \frac{\text{X}_{\tau i}}{\sqrt{\frac{1}{N} \sum_j \text{X}_{\tau j}^2}} \tag{5.26}$$

and a unit total gain $\mathcal{G} \equiv \boldsymbol{g}^\dagger \boldsymbol{w}^* = 1$. This means that the in-sample/out-of-sample test procedure applied below is intended to measure only risk and not the risk-return trade-off (Sharp ratio) as is usual e.g. when back-testing financial strategies. Indeed what we ultimately want to conclude is whether our model of stock returns allows to have a better view of dependences and thus to better diversify away the risk (since we work with normalized returns, we are not concerned with individual variances but only care for dependences).

Quadratic risk, is essentially a measure of expected small fluctuations of the portfolio value:

$$\mathcal{R}^2(\tau) = \frac{1}{T'} \sum_{t'} \frac{1}{N} \sum_{i=1}^{N} \left[\text{X}_{t'i} \frac{w_i^*(\tau)}{\sigma_i^{\text{IS}}(\tau)} \right]^2 \tag{5.27}$$

where, for later convenience, the returns are normalized by a rolling in-sample estimate of their volatility $\sigma^{\text{IS}}(\tau)$ (although the returns have been normalized over the whole period, they may not be close to unit-variance in-sample because of low-frequency regime switches in the volatility). This risk is computed both in-sample (in which case $T' = T^{\text{IS}}$ and $t' = \tau - T', \ldots, \tau - 1$) and out of sample ($T' = T^{\text{OS}}$ and $t' = \tau + 1, \ldots, \tau + T'$), for different input correlation matrices in Eq. (5.25):

- Empirical: the in-sample raw correlation matrix,

$$\rho_{\text{Emp}}^{(1)}(\tau) = \frac{1}{T^{\text{IS}}} \sum_{t'=\tau-T^{\text{IS}}}^{\tau-1} \text{X}_{t'\cdot} \cdot \text{X}_{t'\cdot};$$

- Clipped: the PCA solution of Eq. (5.9), $\rho_{\text{Clip}}^{(1)}(\tau) = \beta_{\text{PCA}}(\tau)^\dagger \beta_{\text{PCA}}(\tau)$, for several values of the number of retained modes M with largest eigenvalues;

- MultiFactor: the improved solution of Eq. (5.12), $\rho_{\text{MF}}^{(1)}(\tau) = \beta(\tau)^\dagger \beta(\tau)$, for several values of the number of factors M.

All these scenarios can furthermore be compared to the benchmark of a full-rank pure noise Wishart correlation matrix. In this case, Random Matrix Theory predicts the values of the average in-sample and out-of-sample risks, in the limit of large matrices with quality factor $q = N/T^{\text{IS}}$:

$$\langle \mathcal{R}_{\text{RMT}}^2 \rangle_{\text{IS}} = \mathcal{R}_{\text{True}}^2 \cdot (1 - q) \qquad \text{and} \qquad \langle \mathcal{R}_{\text{RMT}}^2 \rangle_{\text{OS}} = \mathcal{R}_{\text{True}}^2 \, / \, (1 - q).$$

Moreover, the true risk (i.e. the value of $\mathcal{R}^2$ when the optimal weights are determined using the correlation matrix of the process that generates the realized returns $\text{X}_{\cdot i}/\sigma_i^{\text{IS}}$) can be shown to be $\mathcal{R}_{\text{True}}^2 = 1$ with the definition (5.27).

We show graphically the results of the testing procedure on Fig. 5.19: In-sample and Out-of-sample average risks of every cleaning scheme are plotted versus the control parameter $\alpha = M/N$, where averages are performed over the sliding windows (τ).

When only a very reduced number of factors ($M \approx 1, 2, 3$) is kept, eigenvalue clipping performs better (although quite bad), and similarly when keeping also the very last modes: this is because the linear factors are only good when the eigenmodes are statistically significant, on the left and right of the RMT noise bulk. In the limit $\alpha \to 1$ (i.e. $M = N$) all cleaning schemes collapse to the risk values associated with the raw sample correlation matrix (black solid level line). The RMT predictions are shown for reference as dashed grey lines.

In intermediate values instead (where the Out-of-sample risk is minimal because of marginal gain in signal being higher than marginal risk increase due to added "false positive information"), the procedure worked out in Sect. 5.2.1 to better estimate the linear correlations provides an improved determination of average Out-of-sample risk. In fact, Fig. 5.20 shows that the relative gain

$$[\langle \mathcal{R}_{\text{Clip}}^2 \rangle - \langle \mathcal{R}_{\text{MF}}^2 \rangle] / [\langle \mathcal{R}_{\text{Clip}}^2 \rangle - \mathcal{R}_{\text{True}}^2] \tag{5.28}$$

can reach up to 25%, while not dramatically increasing over-fitting (the In-sample risk is only slightly artificially lowered).

5.5.2 Absolute correlations

We now turn to the core properties of the model: the non-linear dependences. We have already shown that the model is able to reproduce, after calibration, several empirically observed quantities like the copula, and want now to perform an out-of-sample assessment of the volatility-driven dependence in the absolute correlations $\langle \text{Y}_{ti} \text{Y}_{tj} \rangle_t$, where

$$\text{Y}_{t_i} = \frac{|\text{X}_{ti}| - \langle |\text{X}_{ti}| \rangle_t}{\sqrt{\left(|\text{X}_{ti}| - \langle |\text{X}_{ti}| \rangle_t \right)^2}}$$

The definitions of the gain predictor $\boldsymbol{g}$ and the risk measure $\mathcal{R}^2(\tau)$ are identical to Eqs. (5.26) and (5.27) respectively, with Y in place of X. The different cleaning schemes considered are:

Figure 5.19: Linear correlations: In-sample risk (lower curves) and out-of-sample risk (upper curves) defined in Eq. (5.27) and averaged over sliding windows in 2000–2009, for two cleaning schemes: eigenvalue clipping (red circles) and calibrated multi-factor model (blue crosses), both with $M = \alpha N$ linear factors.

Figure 5.20: Linear correlations: Out-of-sample risk is lowered by more than 25%. The location of the minimum coincides with that of minimum risk ($\alpha \approx 0.1$).

- Empirical: the in-sample raw correlation matrix,

$$\rho_{\text{Emp}}^{(\text{a})}(\tau) = \frac{1}{T^{\text{IS}}} \sum_{t'=\tau-T^{\text{IS}}}^{\tau-1} \mathbf{Y}_{t'\cdot} \cdot \mathbf{Y}_{t'\cdot};$$

- Clipped: the PCA solution $\rho_{\text{Clip}}^{(\text{a})}(\tau) = \mathbf{V}_{M|}\Lambda_{M|}\mathbf{V}_{M|}^{\dagger}$, keeping the M eigenmodes of $\rho_{\text{Emp}}^{(\text{a})}$ with largest eigenvalues;

- Gaussian factors: the Gaussian prediction $\rho_{\text{MFG}}^{(\text{a})}(\tau)$ obtained as the sample absolute correlations of long time series simulated according to the M-factors model (5.23) where all volatility parameters A and B are set to 0.

- Multifactor (model): the model prediction $\rho_{\text{MFnG}}^{(\text{a})}(\tau)$ obtained as the sample absolute correlations of long time series simulated according to the M-factors model (5.23) calibrated with the volatility parameters (for each of the two drivers) turned on.

Notice that the meaning of M is not comparable in all cleaning schemes: while for the "clipped eigenvalues" it corresponds to the number of relevant modes in the matrix of *absolute correlations*, for the multi-factor models it instead counts the the number of *linear* factors. This can be seen immediately on Fig. 5.21, where the red curve corresponding to "Clipping" has the usual U-shape discussed above, while the blue curves corresponding to "multi-factor" saturate as α increases above ≈ 0.1, a threshold above which letting additional linear factors barely affects the volatility dependences.

More importantly, this figure shows that multi-factor models offer a better optimal Out-of-sample risk together with less In-sample over-fitting. (Notice that the RMT benchmark is not justified in the case of absolute returns, which are much skewed). The role of volatility dependences is put forward by the better performance of the final non-Gaussian multi-factor level over the Gaussian multi-factor cleaning scheme: Fig. 5.22 shows that while the Gaussian model offers an Out-of-sample risk about 20–25% lower than eigenvalue clipping, the non-Gaussian model performs up to 50% better, at the location of the minimal risk ($\alpha \approx 0.1$). These figures are calculated using the risk difference defined in Eq. (5.28), but with a reference $\mathcal{R}_{\text{ref}}^2 = 1.5$ (arbitrarily chosen between the highest average In-sample risk and the lowest Out-of-sample risk), since the true risk for absolute returns is not 1. The absolute numbers are not informative, but we also show in inset of Fig. 5.22 the relative over-performance of the non-Gaussian multi-factor model over the Gaussian multi-factor model:

$$[\langle \mathcal{R}_{\text{MFnG}}^2 \rangle - \langle \mathcal{R}_{\text{MFG}}^2 \rangle]/[\langle \mathcal{R}_{\text{MFnG}}^2 \rangle - \mathcal{R}_{\text{ref}}^2].$$

The improvement is maximal again around $\alpha \approx 0.1$.

5.5.3   How many factors to keep in the model ?

The number M of linear factors in the description (5.5) is an important input of the model. The intuition that statistical factors are somewhat related to economic sectors does not stand the identification of algebraic modes of fluctuations to sectorial or other macroeconomic factors, beyond the first two or three modes. Still, even if there is no one-to-one identification,
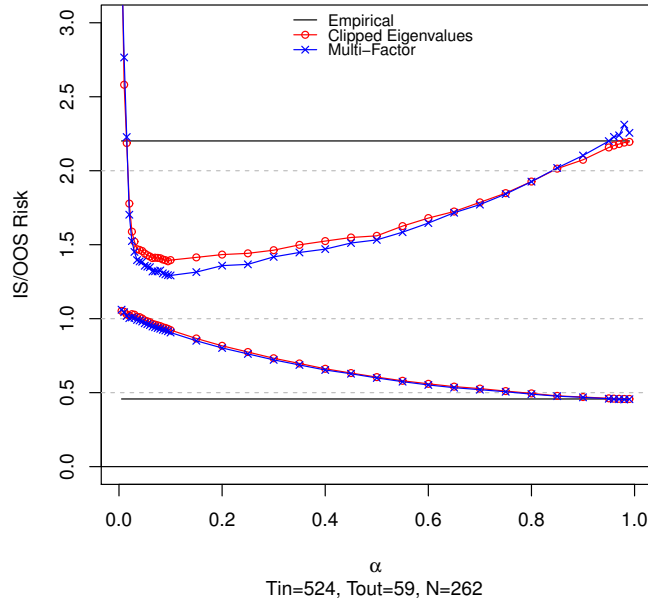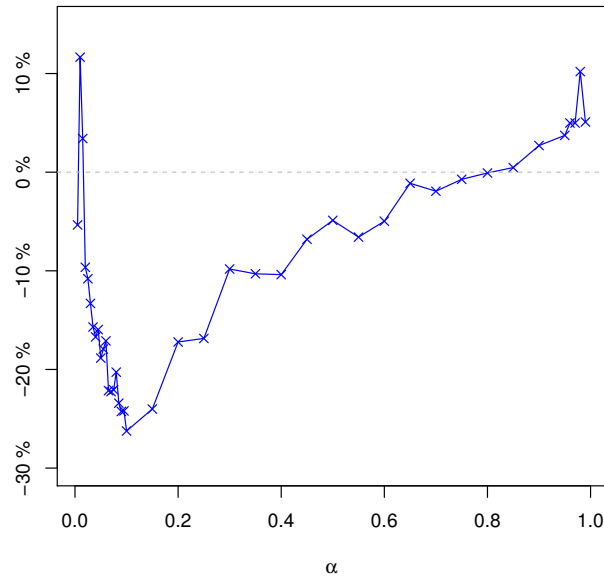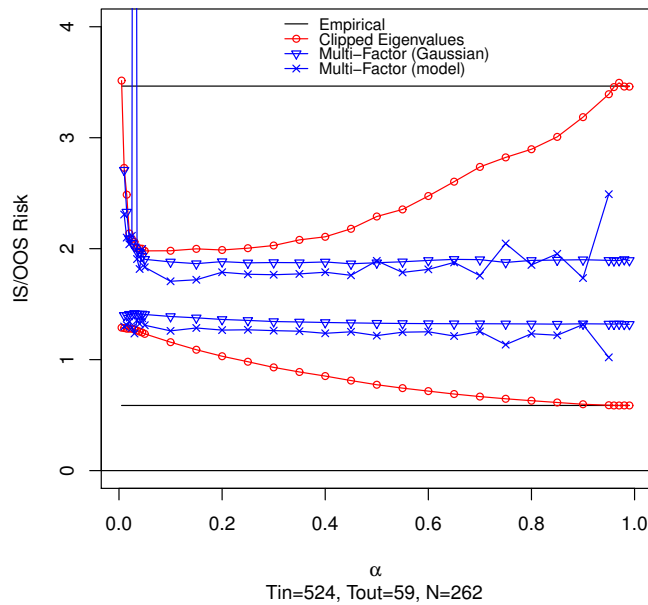
Figure 5.21: Absolute correlations: In-sample risk (lower curves) and out-of-sample risk (upper curves) defined in Eq. (5.27) and averaged over sliding windows in 2000–2009, for three cleaning schemes: eigenvalue clipping of $M = \alpha N$ modes of quadratic correlation (red circles), Gaussian multi-factor (blue triangles) and calibrated multi-factor model (blue crosses), both with $M = \alpha N$ linear factors.
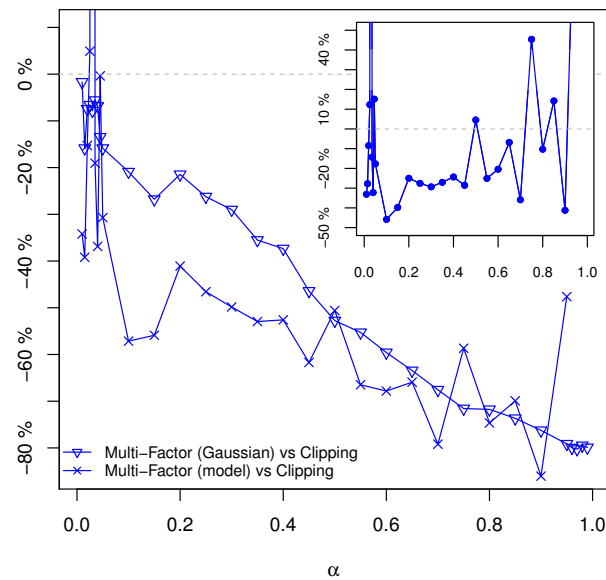
Figure 5.22: Absolute correlations: Out-of-sample risk is lowered by more than 50% where the risk is minimal ($\alpha \approx 0.1$). The role of volatility dependences is put forward by the better performance of the non-Gaussian multi-factor level over the Gaussian multi-factor cleaning scheme: the relative risk difference is shown in the inset.

the number of sectors can be regarded as a reasonable prior for M. In our calibration, we have retained $M = 10$ factors corresponding to the number of Bloomberg sectors plus one, with satisfactory results at reproducing the main empirical stylized facts.

A more justified determination of M is reached by considering the In-Sample / Out-of-sample risk measurement described just above. The In-sample risk is always "too good" because it is measured using a model calibrated so as to match the particular realizations of the noise in the same period: this phenomenon is known as "over-fitting" and is more pronounced the higher the number of parameters is. The Out-of-sample risk is always "too bad" because it is measured using a model that can never account completely for the reality, and unexplained noise remains. A too small number of parameters makes the model oversimplified and unrealistic, and leaves room for much uncertainty that manifests itself in a high risk. On the other hand, calibrating too many parameters will cause the model to over-fit the data of the estimation period, and depart significantly from a "typical" realization of noise; when applying this over-fitted model out-of-sample, the realized risk can be very large as well. Therefore, there typically exists an optimal number of parameters, for which the model fits reasonably the data and is stable when applied out-of-sample.

Adjusting the trigger parameter $\alpha = M/N$ allowed us to find an optimal configuration where the out-of-sample risk is minimized while the in-sample risk is not artificially lowered. A unique value of $\alpha \approx 0.1$ is found to be optimal both for the quadratic risk and for the risk associated with absolute returns, and thus suggests a number of $M \approx 24$ factors to be kept.

5.6   Conclusions

We have presented a factor model for stock returns aiming at reproducing observed non-linear dependences like the quadratic correlations and the copula diagonals, the latter being sensitive to the relative ranks of the joint realizations. Simple generalizations of elliptical model, with idiosyncratic radial parts, were still unable to explain important features like the medial copula departing from a standard value. Such properties can only be accounted for by an interplay of the kurtosis of factors and residual parts.

The main features of our model are the following: (i) A linear factor structure, that is able to generate very accurately the linear correlations, up to pair effects that are not to be explained by any factor model; (ii) independent residuals which allow for an interpretation in terms of indiosyncracies; (iii) non-Gaussianity of the factors with specific kurtosis; (iv) non-Gaussianity of the residuals with specific kurtosis; (v) volatility correlations among the factors generated by common modes of log-volatilities, in particular a "volatility market mode"; (vi) volatility correlations between factors and residuals.

Precise predictions are formulated, and a calibration of all parameters is performed over different periods. An original calibration scheme is provided, with powerful results both in the linear sector and in the volatility parameters. The calibrated models exhibit dependences that show a very good overall fit to empirical observations, even for the non-trivial observables like the copula medial point. The stability analysis performed in the last section illustrates the ability of the model not only to fit the data in-sample but also to reproduce the typical behavior of out-of-sample realizations of the noise. It points toward an optimal number of factors $M \approx 24$.

To conclude this part about the cross-sectional dependences and provide a link to the next part dedicated to temporal dependences, notice that the model studied in this chapter has no temporal content. In particular, the description of the log-volatilities does not exhibit the time-dependence responsible for the so called "volatiilty clustering" effect that will be discussed in several places in the next part. This however was not crucial here for the characterization of dependences among stocks. In terms of inference and parameters estimation, it generates larger uncertainties, but no bias is expected as long as the time series still have a self-averaging property.

Part III

Temporal dependences

CHAPTER 6

Volatility dynamics

Contents

6.1   Introduction

One of the most striking universal stylized facts of financial returns is the volatility clustering effect, which was first reported by Mandelbrot as early as 1963 [Mandelbrot 1963]. He noted that ... *large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes.* The first quantitative description of this effect was the ARCH model proposed by Engle in 1982 [Engle 1982]. It formalizes Mandelbrot's hunch in the simplest possible way, by postulating that returns r_t are conditionally Gaussian random variables, with a time dependent volatility (RMS) σ_t that evolves according to:

$$\sigma_t^2 = s^2 + g r_{t-1}^2. \tag{6.1}$$

In words, this equation means that the (squared) volatility today is equal to a baseline level s^2, plus a self-exciting term that describes the influence of yesterday's *perceived* volatility r_{t-1}^2 on today's activity, through a feedback parameter g. Note that this ARCH model was primarily thought of as an econometric model that needs to be calibrated on data, while a more ambitious goal would be to *derive* such a model from a more fundamental theory — for example, based on behavioural reactions to perceived risk.

It soon became apparent that the above model is far too simple to account for empirical data. For one thing, it is unable to account for the long memory nature of volatility fluctuations. It is also arbitrary in at least two ways:

- First, there is no reason to limit the feedback effect to the previous day only. The Generalized ARCH model (GARCH) [Bollerslev 1986], which has become a classic in quantitative finance, replaces r_{t-1}^2 by an exponential moving average of past squared returns. Obviously, one can further replace the exponential moving average by any weighting kernel $k(\tau) \geq 0$, leading to a large family of models such as:

$$\sigma_t^2 = s^2 + \sum_{\tau=1}^{\infty} k(\tau) r_{t-\tau}^2, \tag{6.2}$$

  which includes all ARCH and GARCH models. For example, ARCH(q) corresponds to a kernel $k(\tau)$ that is strictly zero beyond $\tau = q$. A slowly (power-law) decaying kernel $k(\tau)$ is indeed able to account for the long memory of volatility — this corresponds to the so-called FIGARCH model (for Fractionally Integrated GARCH) [Bollerslev 1994].

- Second, there is no *a priori* reason to single out the day as the only time scale to define the returns. In principle, returns over different time scales could also feedback on the volatility today [Müller 1997, Borland 2011, Lynch 2003], leading to another natural extension of the GARCH model as:

$$\sigma_t^2 = s^2 + \sum_{\ell} \sum_{\tau=1}^{\infty} g_\ell(\tau) R_{t-\tau}^{(\ell)2}, \tag{6.3}$$

  where $R_t^{(\ell)}$ is the cumulative, ℓ day return between $t - \ell$ and t. The first model in that category is the HARCH model of the Olsen group [Müller 1997], where the first "H" stands for Heterogeneous. The authors had in mind that different traders are

sensitive to and react to returns on different time scales. Although this behavioural interpretation was clearly expressed, there has been no real attempt[1] to formalize such an intuition beyond the hand-waving arguments given in [Borland 2011].

The common point to the zoo of generalizations of the initial ARCH model is that the current level of volatility σ_t^2 is expressed as a quadratic form of past realized returns. The most general model of this kind, called QARCH (for Quadratic ARCH), is due to Sentana [Sentana 1995], and reads:

$$\sigma_t^2 = s^2 + \sum_{\tau=1}^{\infty} L(\tau)\, r_{t-\tau} + \sum_{\tau,\tau'=1}^{\infty} K(\tau,\tau')\, r_{t-\tau} r_{t-\tau'}, \tag{6.4}$$

where $L(\tau)$ and $K(\tau,\tau') = K(\tau',\tau)$ are some kernels that should satisfy technical conditions for σ_t^2 to be always positive (see below and [Sentana 1995]). The QARCH can be seen as a general discrete-time model for the dependence of σ_t^2 on all past returns $\{r_{t'}\}_{t'<t}$, truncated to second order. The linear contribution, which involves $L(\tau)$, captures a possible dependence of the volatility on the sign of the past returns. For example, negative past returns tend to induce larger volatility in the future — this is the well-known leverage effect [Black 1976, Bouchaud 2001a, Bekaert 2000], see also [Reigneron 2011] and references therein.[2] The quadratic contribution, on the other hand, contains through the matrix $K(\tau,\tau')$ all ARCH models studied in the literature. For example, ARCH(q), GARCH and FIGARCH models all correspond to a purely *diagonal* kernel, $K(\tau,\tau') = k(\tau)\delta_{\tau,\tau'}$ where $\delta_{\tau,\tau'}$ is Kronecker's delta.

In view of the importance of ARCH modelling in finance, it is somewhat surprising that the general framework provided by QARCH has not been fully explored. Only versions with very short memories, corresponding to at most 2×2 matrices for K, seem to have been considered in the literature. In fact, Sentana's contribution is usually considered to be the introduction of the linear contribution in the GARCH framework, rather than unveiling the versatility of the quadratic structure of the model. The aim of this chapter is to explore in detail the QARCH framework, both from a theoretical and empirical point of view. Of particular interest is the empirical determination of the structure of the feedback kernel $K(\tau,\tau')$ for the daily returns of stocks, which we compare with several proposals in the literature, including the multi-scale model of [Borland 2011] and the trend-induced volatility model of [Zumbach 2010]. Quite surprisingly, we find that while the off-diagonal elements of $K(\tau,\tau')$ are significant, they are at least an order of magnitude smaller than the diagonal elements $k(\tau) := K(\tau,\tau)$. The latter are found to decay very slowly with τ, in agreement with previous discussions. Therefore, in a first approximation, the dominant feedback effect comes from the amplitude of *daily returns* only, with minor corrections coming from returns computed on large time spans, at variance with the assumption of the model put forward in [Borland 2011]. We believe that this finding is unexpected and far from trivial. It is a strong constraint on any attempt to justify the ARCH feedback mechanism from a more fundamental point of view.

---

[1]See however the very recent stochastic volatility model with heterogeneous time scales of [Delpini 2012].

[2](G)QARCH and alternative names such as Asymmetric (G)ARCH, Nonlinear (G)ARCH, Augmented ARCH, etc. often refer to this additional leverage (asymmetry) contribution, whereas the important innovation of QARCH is in fact the possibility of off-diagonal terms in the kernel K.

In parallel with ARCH modelling, stochastic volatility models represent another strand of the literature that has vigorously grown in the last twenty years. Here again, a whole slew of models has emerged [Henry-Labordère 2008], with the Heston model [Heston 1993] and the SABR model [Hagan 2002b] as the best known examples. These models assume that the volatility itself is a random process, governed either by a stochastic differential equation (in time) or an explicit cascade construction in the case of more recent multifractal models [Muzy 2000, Calvet 2008, Lux 2008] (again initiated by Mandelbrot as early as 1974! [Mandelbrot 1974]). There is however a fundamental difference between most of these stochastic volatility models and the ARCH framework: while the former are *time-reversal invariant* (TRI), the latter is explicitly *backward looking*. This, as we shall discuss below, implies that certain correlation functions are not TRI within QARCH models, but are TRI within stochastic volatility models. This leads to an empirically testable prediction; we report below that TRI is indeed violated in stock markets, as also documented in [Zumbach 2009].

The outline of this chapter is as follows. We first review in Section 6.2 some general analytical properties of QARCH models, in particular about the existence of low moments of the volatility. We then introduce in Section 6.3 several different sub-families of QARCH, that we try to motivate intuitively. The consideration of these sub-families follows from the necessity of reducing the dimensionality of the problem, but also from the hope of finding simple regularities that would suggest a plausible interpretation (behavioural or else) of the model, beyond merely best fit criteria. In Section 6.4, we attempt to calibrate "large" QARCH models on individual stock returns, first without trying to impose any a priori structure on the kernel $K(\tau, \tau')$, and then specializing to the various sub-families mentioned above. The same analysis is done in Section 6.5 for the returns of the stock index. We isolate in Section 6.6 the discussion on the issue of TRI for stock returns, both from a theoretical/modeling and an empirical point of view. We give our conclusions in Section 6.7, and relegate to Appendices E.1 and E.2 more technical issues.

6.2   General properties of QARCH models

Some general properties of QARCH models are discussed in Sentana's seminal paper [Sentana 1995]. We review them here and derive some new results. The QARCH model for the return at time t, r_t, is such that:

$$\ln p_t - \ln p_{t-1} = r_t = \sigma_t \xi_t, \tag{6.5}$$

where p_t is the price at time t, σ_t is given by the QARCH specification, Eq. (6.4) above, while the ξ's are i.i.d. random variables, of zero mean and variance equal to unity. While many papers take these ξ's to be Gaussian, it is preferable to be agnostic about their univariate distribution. In fact, several studies including our own (see below), suggest that the ξ's themselves have fat-tails: asset returns are *not* conditionally Gaussian and "true jumps" do occur.[3]

In this section, we will focus on the following non-linear correlation functions (other

---

[3]There seems to be a slowly growing consensus on this point (see e.g. [Ait-Sahalia 2009]): Gaussian processes with stochastic volatility cannot alone account for the discontinuities observed in market prices.

correlations will be considered below, when we turn to empirical studies):

$$\mathcal{C}^{(2)}(\tau) \equiv \langle \left(r_t^2 - \langle r_{t'}^2 \rangle_{t'} \right) r_{t-\tau}^2 \rangle_t \tag{6.6a}$$

$$\widetilde{\mathcal{C}}^{(2)}(\tau) \equiv \langle \left(\sigma_t^2 - \langle \sigma_{t'}^2 \rangle_{t'} \right) r_{t-\tau}^2 \rangle_t \tag{6.6b}$$

$$\mathcal{D}(\tau', \tau'') \equiv \langle \left((r_t^2 - \langle r_{t'}^2 \rangle_{t'} \right) r_{t-\tau'} r_{t-\tau''} \rangle_t \tag{6.6c}$$

$$\widetilde{\mathcal{D}}(\tau', \tau'') \equiv \langle \left((\sigma_t^2 - \langle \sigma_{t'}^2 \rangle_{t'} \right) r_{t-\tau'} r_{t-\tau''} \rangle_t. \tag{6.6d}$$

Here and below, we assume stationarity and correspondingly $\langle \ldots \rangle_t$ refers to a sliding average over t. The following properties are worth noticing: by definition, $\mathcal{D}(\tau, \tau) \equiv \mathcal{C}^{(2)}(\tau)$ and $\widetilde{\mathcal{D}}(\tau, \tau) \equiv \widetilde{\mathcal{C}}^{(2)}(\tau)$. Furthermore, whereas $\mathcal{C}^{(2)}(\tau) = \mathcal{C}^{(2)}(-\tau)$ by construction, the same is not true in general for $\widetilde{\mathcal{C}}^{(2)}(\tau)$. However, using the QARCH causal construction and the independence of the ξ's, one can easily convince oneself that when $\tau > 0$, $\widetilde{\mathcal{C}}^{(2)}(\tau) \equiv \mathcal{C}^{(2)}(\tau)$. Similarly, for $\tau' > \tau'' > 0$, $\widetilde{\mathcal{D}}(\tau', \tau'') \equiv \mathcal{D}(\tau', \tau'')$, while in general, $\mathcal{D}(\tau', \tau'') \neq \mathcal{D}(-\tau', -\tau'') \equiv 0$.

6.2.1 Second moment of the volatility and stationarity

QARCH models only make sense if the expected volatility does not diverge to infinity. The criterion for stability is easy to establish if the ξ's are IID and of zero mean, and reads:

$$\mathrm{Tr}\, K < 1. \tag{6.7}$$

In this case, the volatility is a stationary process such that $\langle \sigma^2 \rangle \equiv \mathbb{E}[\sigma^2] = s^2/(1 - \mathrm{Tr}\, K)$: the feedback-induced increase of the volatility only involves the diagonal elements of K. Note also that the leverage kernel $L(\tau)$ does not appear in this equation. As an interesting example, we consider kernels with a power-law decaying diagonal: $K(\tau, \tau) = g\, \tau^{-\alpha} \mathbb{1}_{\{\tau \leq q\}}$. For a given decay parameter α, the amplitude g must be smaller than a certain value $g_c(\alpha, q)$ for $\langle \sigma^2 \rangle$ to be finite. Fig. 6.1 shows the critical frontier $g_c(\alpha, q)$ for $q = 1, 32, 256$ and $q \to \infty$. The critical frontier in the limit case $q = \infty$ is given by $g_c = 1/\zeta(\alpha)$, where $\zeta(\alpha)$ is Riemann's zeta function (solid red). Note in particular that the model is always unstable when $\alpha < 1$, i.e. when the memory of past realized volatility decays too slowly.[4] At the other extreme, $q = 1$, the constraint is well known to be $g = k(1) \leq 1$ (solid red).

Within a strict interpretation of the QARCH model, there are additional constraints on the kernels K and L that arise from the fact that σ_t^2 should be positive for any realization of price returns. This imposes that a) all the eigenvalues of K should be non-negative, and b) that the following inequality holds:

$$\sum_{\tau, \tau'=1}^{q} L(\tau) K^{-1}(\tau, \tau') L(\tau') \leq 4\, s^2, \tag{6.8}$$

where K^{-1} is the matrix inverse of K. However, these constraints might be too strong if one interprets the QARCH model as a generic expansion of σ_t^2 in powers of past returns, truncated to second order [Sentana 1995]. It could well be that higher order terms are stabilizing and lead to a meaningful, stable model beyond the limits quoted here.

---

[4] In the context of fractionally integrated processes $I(d)$, the condition $\alpha \leq 1$ is equivalent to the 'difference parameter' $d = \alpha - 1$ being positive.
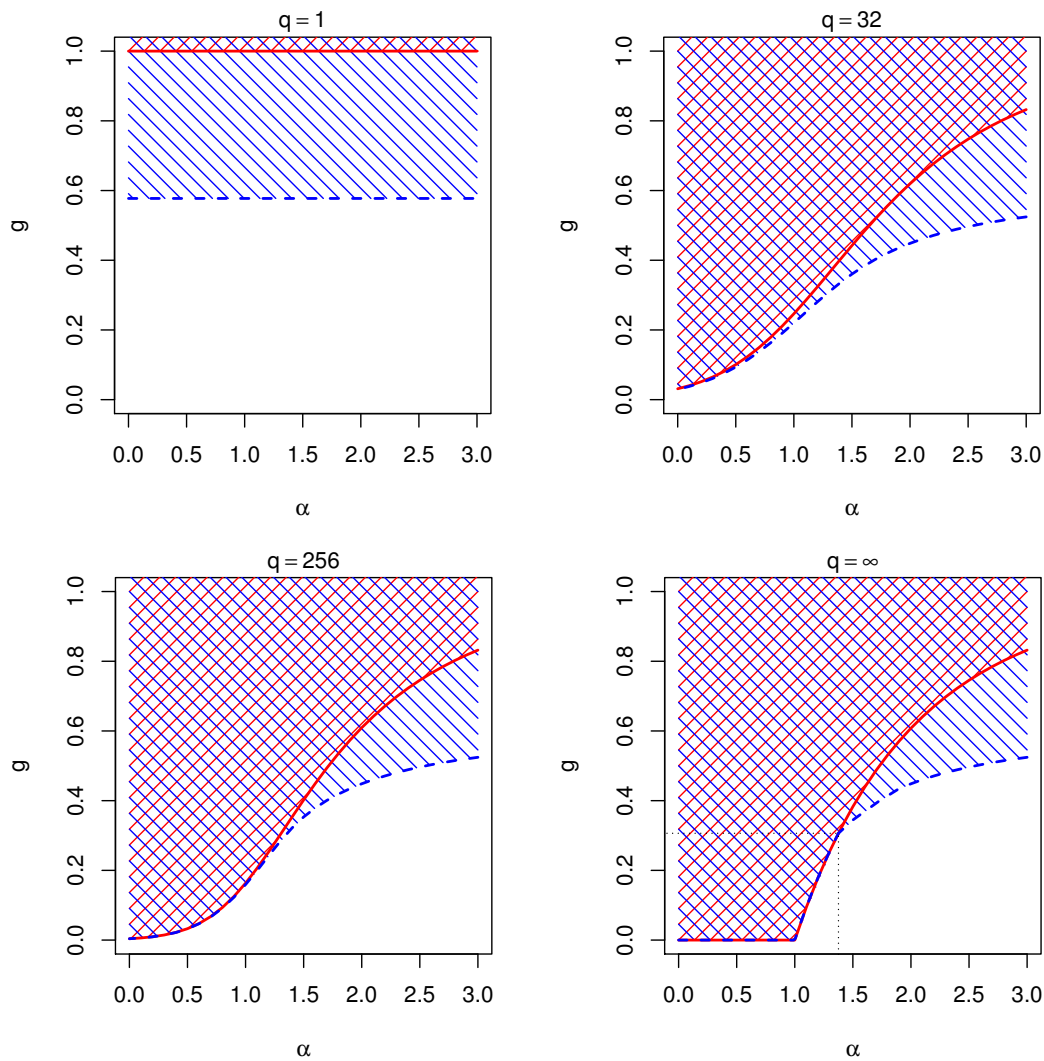
Figure 6.1: Allowed region in the parameter space for $K(\tau, \tau) = g\,\tau^{-\alpha}\mathbb{1}_{\{\tau \leq q\}}$ and $L(\tau) = 0$, according to the finiteness of $\langle\sigma^2\rangle$ and $\langle\sigma^4\rangle$. Divergence of $\langle\sigma^2\rangle$ is depicted by $45°$ (red) hatching, while divergence of $\langle\sigma^4\rangle$ is depicted by $-45°$ (blue) hatching. In the wedge between the dashed blue and solid red lines, $\langle\sigma^2\rangle < \infty$ while $\langle\sigma^4\rangle$ diverges.

6.2.2 Fourth moment of the volatility

The existence of higher moments of σ can also be analyzed, leading to more and more cumbersome algebra [He 1999, Ling 2002, Teyssière 2010]. In view of its importance, we have studied in detail the conditions for the existence of the fourth moment of σ, which allows one to characterize the excess kurtosis κ of the returns, traditionally defined as:

$$\kappa = \frac{\langle r^4 \rangle}{\langle r^2 \rangle^2} - 3 \equiv \frac{\langle \sigma^4 \rangle \langle \xi^4 \rangle}{\langle \sigma^2 \rangle^2} - 3. \tag{6.9}$$

In the general case, $\langle \sigma^4 \rangle, \mathcal{C}^{(2)}(\tau)$ and $\mathcal{D}(\tau', \tau'')$ are related by the following set of self-consistent equations:

$$\langle \sigma^4 \rangle - \langle \sigma^2 \rangle^2 = \langle \sigma^2 \rangle^2 \Big(\operatorname{Tr}(K^2) - \operatorname{Tr}(K^{\bullet 2}) \Big) + \sum_{\tau > 0} K(\tau, \tau) \mathcal{C}^{(2)}(\tau) \tag{6.10a}$$

$$+ 2 \sum_{0 < \tau_2 < \tau_1} K(\tau_1, \tau_2) \left[\mathcal{D}(\tau_1, \tau_2) - \sum_{0 < \tau < \tau_2} K(\tau, \tau) \mathcal{D}(\tau_1 - \tau, \tau_2 - \tau) \right]$$

$$\mathcal{C}^{(2)}(\tau > 0) = K(\tau, \tau) \Big(\langle \sigma^4 \rangle \langle \xi^4 \rangle - \langle \sigma^2 \rangle^2 \Big) + \sum_{\tau' \neq \tau} K(\tau', \tau') \mathcal{C}^{(2)}(\tau - \tau') \tag{6.10b}$$

$$+ 2 \sum_{\tau' > \tau'' = \tau + 1}^{q} K(\tau', \tau'') \mathcal{D}(\tau' - \tau, \tau'' - \tau)$$

$$\mathcal{D}(\tau_1 > 0, \tau_2 > 0) = 2 K(\tau_1, \tau_2) \Big(\mathcal{C}^{(2)}(\tau_1 - \tau_2) + \langle \sigma^2 \rangle^2 \Big) \tag{6.10c}$$

$$+ \sum_{0 < \tau' < \tau_2} K(\tau', \tau') \mathcal{D}(\tau_1 - \tau', \tau_2 - \tau')$$

$$+ 2 \sum_{\tau' > \tau_2, \tau' \neq \tau_1} K(\tau', \tau_2) \mathcal{D}(\tau' - \tau_2, |\tau_1 - \tau_2|).$$

where we assume for simplicity here that the leverage effect is absent, i.e. $L(\tau) \equiv 0$, and $K^{\bullet 2}$ means the square of K in the Hadamard sense (i.e. element by element). For a QARCH with maximum horizon q, we have thus a set of $1 + q + q(q-1)/2$ linear equations for $k(\tau \geq 0)$ that can be numerically solved for an arbitrary choice of the kernel K. These equations simplify somewhat in the case of a purely diagonal kernel $K(\tau, \tau') = k(\tau) \delta_{\tau, \tau'}$. One finds:

$$\langle \sigma^4 \rangle - \langle \sigma^2 \rangle^2 = \sum_{\tau' > 0} k(\tau') \mathcal{C}^{(2)}(\tau') \tag{6.11a}$$

$$\mathcal{C}^{(2)}(\tau) = \sum_{\tau' > 0} k(\tau') \mathcal{C}^{(2)}(\tau' - \tau) \tag{6.11b}$$

By substituting $\langle \sigma^4 \rangle$, it is easy to explicit the linear system in matrix form $\nabla \mathcal{C}^{(2)} = S$ with

$$\nabla(\tau, \tau') = \delta_{\tau\tau'} - \langle \xi^4 \rangle k(\tau) k(\tau') - \big[k(\tau - \tau') + k(\tau + \tau') \big] \tag{6.12a}$$

$$S(\tau) = k(\tau) \langle \sigma^2 \rangle^2 \big(\langle \xi^4 \rangle - 1 \big) \tag{6.12b}$$

and the convention that $k(\tau) = 0, \forall \tau \leq 0$.

Let us examine this in more detail for ARCH(q). For simplicity, we assume here that ξ is Gaussian ($\langle \xi^4 \rangle = 3$) and s is chosen such that $\langle \sigma^2 \rangle = 1$. The condition on $k(\tau)$ for which $\langle \sigma^4 \rangle$ diverges is given by $\det \nabla = 0$, where ∇ is the matrix whose entries are defined in Eq. 6.12a. For different q's, this reads:

- for $q = 1$, one recovers the well known result that ARCH(1) has a finite fourth moment only when $k_1 < 1/\sqrt{3}$.

- for $q = 2$, the stability line is given by $k_1 + k_2 = 1$, while the existence of a finite fourth moment is given by the condition $k_1^2 < (1/3 - k_2^2)(1 - k_2)/(1 + k_2)$.

- for $q \to \infty$, we again assume the τ dependence of $k(\tau)$ to be a power-law, $g\,\tau^{-\alpha}$ (corresponding to the FIGARCH model). The critical line for which the fourth moment diverges is shown in dashed blue in Fig. 6.1. After a careful extrapolation to $q = \infty$, we find that whenever $1 < \alpha < \alpha_c \approx 1.376$, the fourth moment exists as soon as the model is stationary, i.e. when $g < 1/\zeta(\alpha) < 1/\zeta(\alpha_c) \approx 0.306$.

The last result is quite interesting and can be understood from Eq. (6.11), which shows that to lowest order in g, one has:

$$\frac{\langle \sigma^4 \rangle}{\langle \sigma^2 \rangle^2} - 1 \approx (\langle \xi^4 \rangle - 1)g^2 \sum_{\tau > 0} \frac{1}{\tau^{2\alpha}}. \tag{6.13}$$

The above expression only diverges if $\alpha < 1/2$, but this is far in the forbidden region $\alpha < 1$ where $\langle \sigma^2 \rangle$ itself diverges. Therefore, perhaps unexpectedly, a FIGARCH model with a long memory (i.e. $\alpha < 1.376$) cannot lead to a large kurtosis of the returns, *unless the ξ variables have themselves fat tails.* We will come back to this important point below. By the way, FIGARCH models with long memory and $1 < \alpha < 3/2$ are able to generate power-law correlations of the volatility $\mathcal{C}^{(2)}(\tau) \propto \tau^{-\beta}$, with a slow decay characterized by an exponent $\beta = 3 - 2\alpha$, as we demonstrate in Appendix E.2.

As we alluded to in the introduction, ARCH(q) models posit that today's volatility is only sensitive to past daily returns. This assumption can be relaxed in several natural ways, each of which leading to a specific structure of the feedback kernel K. We will present these extensions in increasing order of complexity.

6.2.3 Returns over different time scales

Let us define the ℓ-day return between $t - \ell$ and t as $R_t^{(\ell)}$, such that:

$$R_t^{(\ell)} = \sum_{\tau=1}^{\ell} r_{t-\tau}; \qquad R_t^{(1)} = r_{t-1} = \ln p_{t-1} - \ln p_{t-2}, \tag{6.14}$$

where p_t is the price at time t. The simplest extension of ARCH(q) is to allow all past 2-day returns to play a role as well, i.e.:

$$\sigma_t^2 = s^2 + \sum_{\tau=0}^{q-1} g_1(\tau)[R_{t-\tau}^{(1)}]^2 + \sum_{\tau=0}^{q-2} g_2(\tau)[R_{t-\tau}^{(2)}]^2, \tag{6.15}$$
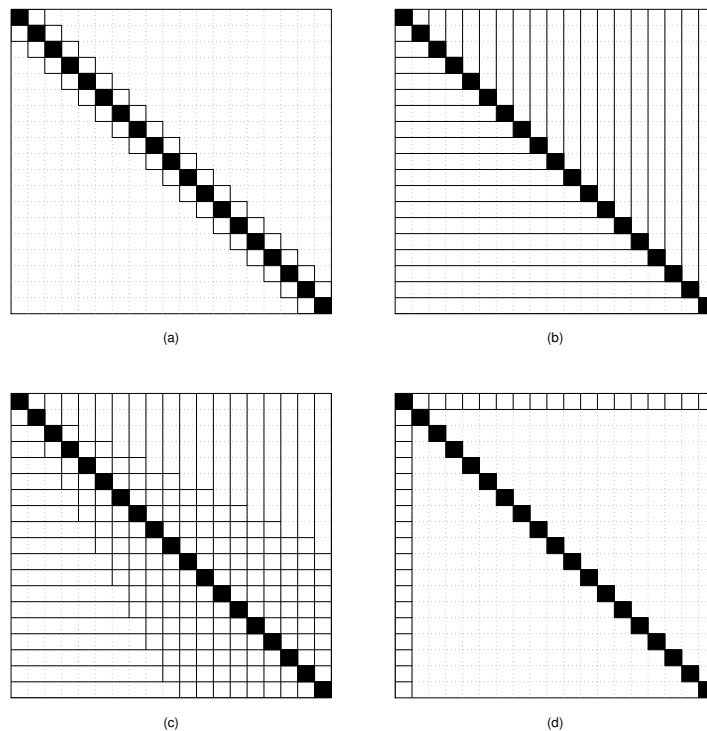
Figure 6.2: Matrix structures. **(a)** Overlapping two-scales; **(b)** Borland-Bouchaud multi-scale; **(c)** Zumbach; **(d)** Long-Trend.

where $g_1(\tau)$ and $g_2(\tau)$ are coefficients, all together $2q - 1$ of them. Upon identification with the QARCH kernel, one finds:

$$
\begin{aligned}
K(\tau, \tau) &= g_1(\tau - 1) + g_2(\tau - 1) + g_2(\tau - 2), \\
K(\tau, \tau + 1) &= g_2(\tau - 1) \\
K(\tau, \tau + \ell) &= 0 \ \text{ for } \ \ell \geq 2,
\end{aligned}
\tag{6.16}
$$

with the convention that $g_2(-1) = 0$. The model can thus be re-interpreted in the following way: the square volatility is still a weighted sum of past *daily* squared returns, but there is an extra contribution that picks up the realized one-day covariance of returns. If $g_2(\tau) \geq 0$, the model means that the persistence of the same trend two days in a row leads to increased volatilities. A schematic representation of this model is given in Fig. 6.2(a).

One can naturally generalize again the above model to include 2-day, 3-day, ℓ-day returns, with more coefficients $g_1(\tau), g_2(\tau), \ldots, g_\ell(\tau)$, with a total of $\ell(2q + 1 - \ell)/2$ parameters. Obviously, when $\ell = q$, all possible time scales are exhausted, and the number of free parameters is $q(q + 1)/2$, i.e. exactly the number of independent entries of the symmetric $q \times q$ feedback kernel K.

6.3   Some special families of QARCH models

6.3.1   Multi-scale, cumulative returns

Another model, proposed in [Müller 1997, Borland 2011], is motivated by the idea that traders may be influenced not only by yesterday's return, but also by the change of price between today and 5-days ago (for weekly traders), or today and 20-days ago (for monthly traders), etc. In this case, the natural extension of the ARCH framework is to write:

$$\sigma_t^2 = s^2 + \sum_{\ell=1}^{q} g_{\mathrm{BB}}(\ell)[R_t^{(\ell)}]^2, \tag{6.17}$$

where the index BB refers to the model put forward in [Borland 2011]. The BB model requires a priori q different parameters. It is simple to see that in this case, the kernel matrix can be expressed as:

$$K_{\mathrm{BB}}(\tau', \tau'') = G[\max(\tau', \tau'')], \quad \text{with} \quad G[\tau] = \sum_{\ell=\tau}^{q} g_{\mathrm{BB}}(\ell). \tag{6.18}$$

The spectral properties of these matrices are investigated in detail in Appendix E.1. One can also consider a mixed model where both cumulative returns and daily returns play a role. This amounts to taking the off-diagonal elements of K as prescribed by the above equation, but to specify the diagonal elements $K(\tau, \tau)$ completely independently from $G[\tau]$. This leads to a matrix structure schematically represented in Fig. 6.2(b), parameterized by $2q - 1$ independent coefficients.

6.3.2   Zumbach's trend effect (ARTCH)

Zumbach's model [Zumbach 2010] is another particular case in the class of models described by Eq. (6.4). It involves returns over different lengths of time and characterizes the effect of past trending aggregated returns on future volatility. The quadratic part in the volatility prediction model is

$$\mathrm{ARCH} + \sum_{\ell=1}^{\lfloor q/2 \rfloor} g_{\mathrm{Z}}(\ell)\, R_t^{(\ell)} R_{t-\ell}^{(\ell)} \tag{6.19}$$

When relevant, only specific time scales like the day ($\ell = 1$), the week ($\ell = 5$), the month ($\ell = 20$), etc. can be retained in the summation. The off-diagonal elements of the kernel K now take the following form:

$$K_{\mathrm{Z}}(\tau', \tau'' > \tau') = \sum_{\ell=\max(\tau', \frac{\tau''}{2})}^{\min(\tau''-1, \lfloor q/2 \rfloor)} g_{\mathrm{Z}}(\ell) \tag{6.20}$$

Since it is upper triangular by construction, we symmetrize it as $\frac{1}{2}(K + K^{\dagger})$, and the diagonal is filled with the ARCH parameters. This model contains $q + \lfloor q/2 \rfloor$ independent coefficients, and is schematically represented in Fig. 6.2(c).

6.3.3 A generalized trend effect

In Zumbach's model, the trend component is defined by comparing returns computed over the same horizon ℓ. This of course is not necessary. As an extreme alternative model, we consider a model where the volatility today is affected by the last return r_{t-1} confirmation (or the negation) of a long trend. In more formal terms, this writes:

$$\text{ARCH} + r_{t-1} \times \sum_{\ell=1}^{q-1} g_{\text{LT}}(\ell) r_{t-1-\ell}, \tag{6.21}$$

where $g_{\text{LT}}(\ell)$ is the sequence of weights that defines the past "long trend" (hence the index LT). This now corresponds to a kernel K with diagonal elements corresponding to the ARCH effects and a single non trivial row (and column) corresponding to the trend effect: $K(1, \tau > 1) = g_{\text{LT}}(\tau - 1)$. This model has again $2q - 1$ free parameters.

Of course, one can consider QARCH models that encode some, or all of the above mechanisms — for example, a model that schematically reads $\text{ARCH} + \text{BB} + \text{LT}$ would require $3q - 2$ parameters.

6.3.4 Spectral interpretation of the QARCH

Another illuminating way to interpret QARCH models is to work in the diagonal basis of the K matrix, where the quadratic term in Eq. (6.4) reads:

$$\sum_{\tau', \tau''=1}^{q} \left(\sum_n \lambda_n v_n(\tau') v_n(\tau'') \right) r_{t-\tau'} \, r_{t-\tau''} \equiv \sum_n \lambda_n \, \langle r | v_n \rangle_t^2 \tag{6.22}$$

with (λ_n, v_n) being, respectively, the n-th eigenvalue and eigenvector of K, and $\langle r | v_n \rangle_t = \sum_{\tau=1}^{q} v_n(\tau) \, r_{t-\tau}$ the projection of the pattern created by the last q returns on the n-th eigenvector. One can therefore say that the square volatility σ_t^2 picks up contributions from various past returns eigenmodes. The modes associated to the largest eigenvalues λ are those which have the largest contribution to volatility spikes.

The $\text{ARCH}(q)$ model corresponds to a diagonal matrix; in this case the modes are trivially individual daily returns. Another trivial case is when K is of rank one and its spectral decomposition is simply

$$K(\tau', \tau'') = \lambda v(\tau') v(\tau'') \tag{6.23}$$

where $\lambda = \text{Tr}(K)$ is the only non-null eigenvalue, and $v(\tau) = \sqrt{K(\tau, \tau)/\text{Tr}(K)}$ the eigenvector associated with this non-degenerate mode. The corresponding contribution to the increase in volatility (6.22) is therefore $\lambda \widehat{R}_t^2$, where

$$\widehat{R}_t = \langle r | v \rangle_t = \sum_{\tau=1}^{q} v(\tau) \, r_{t-\tau}, \tag{6.24}$$

can be interpreted as an average return over the whole period, with a certain set of weights $v(\tau)$.

The pure BB model (without extra ARCH contributions) can also be diagonalized analytically in the large q limit for certain choices of the function $g_{\text{BB}}(\tau)$. We detail these calculations (which are mostly of theoretical interest) in Appendix E.1.

6.4   Empirical study: single stocks

We now turn to the calibration on real data of "large" QARCH models, i.e. models that take into account q past returns with q large (20 or more). The difficulty is that the full calibration of the matrix K requires the determination of $q(q+1)/2$ parameters, which is already equal to 210 when $q = 20$! This is why the discussion of the previous section is important: imposing some *a priori* structure on the matrix K may help limiting the number of parameters, and gaining robustness and transparency. However, perhaps surprisingly, we will find that none of the above model seem to have enough flexibility to reproduce the subtle structure of the empirically determined K matrix.

6.4.1   Dataset and methodology

Equation (6.4) is a prediction model for the predicted variable σ_t with explanatory variables past returns r at all lags. The dataset we will use to calibrate the model is composed of daily stock prices (Open, Close, High, Low) for $N = 280$ names present in the S&P-500 index from Jan. 2000 to Dec. 2009 ($T = 2515$ days), without interruption. The reference price for day t is defined to be the close of that day C_t, and the return r_t is given by $r_t = \ln C_t - \ln C_{t-1}$. The true volatility is of course unobservable; we replace it by the standard Rogers-Satchell (RS) estimator [Rogers 1991, Floros 2009]:

$$\widehat{\sigma}_t^2 = \ln(H_t/O_t)\ln(H_t/C_t) + \ln(L_t/O_t)\ln(L_t/C_t). \tag{6.25}$$

As always in this kind of studies over extended periods of time, our dataset suffers from a selection bias since we have retained only those stock names that have remained alive over the whole period.

There are several further methodological points that we need to discuss right away:

- *Universality hypothesis.* We assume that the feedback matrix K and the leverage kernel L are identical for all stocks, once returns are standardized to get rid of the idiosyncratic average level of the volatility. This will allow us to use the whole data set (of size $N \times T$) to calibrate the model. Some dependence of K and L on global properties of firms (such as market cap, liquidity, etc.) may be possible, and we leave this for a later study. However, we will see later that the universality hypothesis appears to be a reasonable first approximation.

- *Removal of the market-wide volatility.* We anticipate that the volatility of a single stock has a market component that depends on the return of the index, and an idiosyncratic component that we attempt to account for with the returns of the stock itself. As a proxy for the instantaneous market volatility, we take the cross-sectional average of the squared returns of individual stocks, i.e.

$$\Sigma_t = \sqrt{\frac{1}{N}\sum_{i=1}^{N} r_{i,t}^2} \tag{6.26}$$

    and redefine returns and volatilities as r_t/Σ_t and $\widehat{\sigma}_t/\Sigma_t$. Finally, as announced above, the return time series are centered and standardized, and the RS volatility time series

are standardized such that $\langle \widehat{\sigma}_{i,t}^2 \rangle = 1$ for all is. (This also gets rid of the multiplicative bias of the Rogers-Satchell estimator when used with non-Gaussian returns.)

- *Calibration strategy.* The standard procedure used to calibrate ARCH models is maximum-likelihood, which relies on the choice of a family of distributions for the noise component ξ, often taken to be Student-t distributions (which include, in a limit, the Gaussian distribution). However, this method cannot be used directly in our case because there are far too many parameters and the numerical optimization of the log-likelihood is both extremely demanding in computer time and unreliable, as many different optima are expected in general. An alternative method, the Generalized Method of Moments (GMM), is to determine the $1 + q + q(q+1)/2$ parameters of the model using empirically determined correlation functions that depend on s^2, $L(\tau)$ and $K(\tau, \tau')$. This latter method is however sensitive to tail events and can lead to biases. We will therefore use a hybrid strategy, where a first estimate of these parameters, obtained using the GMM, serves as the starting point of a one-step likelihood maximization, which determines the set of most likely parameters in the vicinity of the GMM estimate (more details on this below).

- *Choice of the horizon q.* In principle, the value of the farthest relevant lag q is an additional free parameter of the model, and should be estimated jointly with all the others. However, this would lead to a huge computational effort and to questionable conclusions. In fact, we will find that the diagonal elements $K(\tau, \tau)$ decay quite slowly with τ (in line with many previous studies) and can be accurately determined up to large lags using the GMM. Off-diagonal elements, on the other hand, turn out to be much smaller and rather noisy. We will therefore restrict the horizon for these off-diagonal elements to $q = 10$ (two weeks) or $q = 20$ (four weeks). Longer horizons, although possibly still significant, lead to very small out-of-sample extra predictability (but note that longer horizons *are* needed for the diagonal elements of K).

6.4.2 GMM estimation based on correlation functions

On top of the already defined four-point correlation functions $\mathcal{C}^{(2)}(\tau)$ and $\mathcal{D}(\tau', \tau'')$ (and their corresponding "tilde" twins), we will introduce two- and three-point correlation functions that turn out to be useful (note that the r_ts are assumed to have zero mean):

$$\mathcal{C}^{(1)}(\tau) \equiv \langle r_t r_{t-\tau} \rangle_t \tag{6.27a}$$

$$\mathcal{C}^{(a)}(\tau) \equiv \langle (r_t^2 - \langle r^2 \rangle) |r_{t-\tau}| \rangle_t \tag{6.27b}$$

$$\widetilde{\mathcal{C}}^{(a)}(\tau) \equiv \langle (\sigma_t^2 - \langle \sigma^2 \rangle) |r_{t-\tau}| \rangle_t \tag{6.27c}$$

$$\mathcal{L}(\tau) \equiv \langle (r_t^2 - \langle r^2 \rangle) r_{t-\tau} \rangle_t \tag{6.27d}$$

$$\widetilde{\mathcal{L}}(\tau) \equiv \langle (\sigma_t^2 - \langle \sigma^2 \rangle) r_{t-\tau} \rangle_t \tag{6.27e}$$

$$\mathcal{L}^{(a)}(\tau) \equiv \langle |r_t| r_{t-\tau} \rangle_t \tag{6.27f}$$

$$\mathcal{D}^{(a)}(\tau', \tau'') \equiv \langle (|r_t| - \langle |r| \rangle) r_{t-\tau'} r_{t-\tau''} \rangle. \tag{6.27g}$$

The $\mathcal{C}^{(1)}(\tau)$ correlation function is by definition equal to $\langle r_t^2 \rangle_t = 1$ for $\tau = 0$, and is usually considered to be zero for $\tau > 0$. However, as is well known, there are small anti-correlations

of stock returns. On our data set, we find that these linear correlations are very noisy but significant, and can be fitted by:

$$\mathcal{C}^{(1)}(\tau \geq 1) \approx -0.04\, \mathrm{e}^{-0.39\tau}, \qquad (6.28)$$

corresponding to a decay time of ≈ 2.5 days. The values of $\mathcal{C}^{(a)}$ characterize volatility correlations and are similar in spirit to $\mathcal{C}^{(2)}$, but they only involve third order moments of r, instead of fourth order moments, and are thus more robust to extreme events. The $\mathcal{L}$ correlations, on the other hand, characterize the leverage effect, i.e. the influence of the *sign* of past returns on future volatilities.

These correlation functions allow us to define a well-posed problem of solving a system with $1 + q + \frac{q(q+1)}{2}$ unknowns $\left(s^2, L(\tau), K(\tau', \tau'')\right)$ using the following $1 + q + q + \frac{q(q-1)}{2}$ equations that involve empirically measured correlation functions (in calligraphic letters), for $1 \leq \tau \leq q$ and $1 \leq \tau_2 < \tau_1 \leq q$:

$$\langle \sigma^2 \rangle = s^2 + \sum_{\tau',\tau''} K(\tau',\tau'')\mathcal{C}^{(1)}(\tau'-\tau'') \qquad (6.29\mathrm{a})$$

$$\widetilde{\mathcal{L}}(\tau) = \sum_{\tau'} L(\tau')\mathcal{C}^{(1)}(\tau-\tau') \qquad (6.29\mathrm{b})$$

$$+ \sum_{\tau'} K(\tau',\tau')\mathcal{L}(\tau-\tau') + 2\sum_{\tau'\neq\tau} K(\tau',\tau)\mathcal{L}(\tau'-\tau) \qquad (6.29\mathrm{c})$$

$$\widetilde{\mathcal{C}}^{(a)}(\tau) \approx \sum_{\tau'} L(\tau')\mathcal{L}^{(a)}(\tau'-\tau) \qquad (6.29\mathrm{d})$$

$$+ \sum_{\tau'} K(\tau',\tau')\mathcal{C}^{(a)}(\tau - \tau') + 2\sum_{\tau'>\tau''>\tau>0} K(\tau',\tau'')\mathcal{D}^{(a)}(\tau'-\tau,\tau''-\tau)$$

$$\widetilde{\mathcal{D}}(\tau_1,\tau_2) \approx L(\tau_2)\mathcal{L}(\tau_1-\tau_2) + L(\tau_1)\mathcal{L}(\tau_2-\tau_1) \qquad (6.29\mathrm{e})$$

$$+ 2\sum_{\tau'>\tau_2} K(\tau',\tau_2)\left[\mathcal{D}(\tau_1-\tau_2,\tau'-\tau_2) + \mathcal{C}^{(1)}(\tau_1-\tau') - \mathcal{C}^{(1)}(\tau'-\tau_2)\mathcal{C}^{(1)}(\tau_1-\tau_2)\right]$$

$$+ \sum_{\tau'\leq\min(\tau_1,\tau_2)} K(\tau',\tau')\mathcal{D}(\tau_1-\tau',\tau_2-\tau'),$$

where all the sums only involve positive τs. These equations are exact if all 3-point and 4-point correlations that involve rs at 3 (resp. 4) distinct times are strictly zero. But since the linear correlations $\mathcal{C}^{(1)}(\tau > 0)$ are very small, it is a safe approximation to neglect these higher order correlations.

Note that the above equations still involve fourth order moments (the off-diagonal elements of $\mathcal{D}$), that in turn lead to very noisy estimators of the off-diagonal of K. In order to improve the accuracy of these estimators, we have cut-off large events by transforming the returns r_t into $r_{\mathrm{cut}}\tanh(r_t/r_{\mathrm{cut}})$, which leaves small returns unchanged but caps large returns. We have chosen to truncate events beyond $3 - \sigma$, i.e. $r_{\mathrm{cut}} = 3$. In any case, we will use the above equations in conjunction with maximum likelihood (for which no cut-off is used) to obtain more robust estimates of these off-diagonal elements.

When solving the set of equations (6.29), we find that the diagonal elements $K(\tau, \tau)$ are an order of magnitude larger than the corresponding off-diagonal elements $K(\tau, \tau' \neq \tau)$. This was not expected *a priori* and is in fact one of the central result of this study, and

confirms that *daily returns* indeed play a special role in the volatility feedback mechanism, as postulated by ARCH models. Returns on different time scales, while significant, can be treated as a perturbation of the main ARCH effect.

6.4.3 The diagonal kernels

The above remark suggests to calibrate the model in two steps. We first neglect off-diagonal effects altogether, and determine the $2q + 1$ parameters s^2, $L(\tau)$ and $k(\tau) = K(\tau, \tau)$ through the following reduced set of equations:

$$\langle \sigma^2 \rangle = 1 = s^2 + \sum_\tau k(\tau) \tag{6.30a}$$

$$\widetilde{\mathcal{L}}(\tau) = L(\tau) + \sum_{\tau' \neq \tau} L(\tau') \mathcal{C}^{(1)}(\tau - \tau') + \sum_{\tau'} k(\tau') \mathcal{L}(\tau - \tau') \tag{6.30b}$$

$$\widetilde{\mathcal{C}}^{(a)}(\tau) \approx \sum_{\tau'} L(\tau') \mathcal{L}^{(a)}(\tau' - \tau) + \sum_{\tau'} k(\tau') \mathcal{C}^{(a)}(\tau - \tau'). \tag{6.30c}$$

The input empirical correlation functions $\mathcal{L}(\tau)$ and $\widetilde{\mathcal{C}}^{(a)}(\tau)$ are plotted in Fig. 6.3, together with, respectively, a double-exponential fit and a truncated power-law fit (see legend for parameters values). $\widetilde{\mathcal{L}}$ and $\mathcal{L}^{(a)}$ look very similar to $\mathcal{L}(\tau)$; note that all these functions are approximately zero for $\tau < 0$. The above equations are then solved using these analytical fits, which leads to the kernels $k(\tau)$ and $L(\tau)$ that we report in bold in Fig. 6.4. Using the raw data — instead of the fits — for all the correlation functions results in more noisy $L(\tau)$ and $k(\tau)$ (thin lines), but still very close to the bold curves shown in Fig. 6.4. As expected, $L(\tau)$ is very close to $\widetilde{\mathcal{L}}(\tau)$: there is a weak, but significant leverage effect for individual stocks.

We then show in Fig. 6.5 a plot of $s^2(q) = 1 - \sum_{\tau=1}^q k(\tau)$ as a function of q. Including the feedback of the far away past progressively decreases the value of the baseline level s^2. In order to extrapolate to $q \to \infty$, we have found that the following fit is very accurate:

$$s^2(q) = s_\infty^2 + g \frac{q^{1-\alpha}}{\alpha - 1} e^{-q/q_0}, \tag{6.31}$$

with $s_\infty^2 \approx 0.21$, $\alpha \approx 1.11$, $g \approx 0.081$ and $q_0 \approx 53$. Several comments are interesting at this stage:

- The asymptotic value s_∞^2 is equal to 20% of the observed squared volatility, meaning that volatility feedback effects increase the volatility by a factor ≈ 2.25 (but remember that we have scaled out the market wide volatility). Such a strong amplification of the volatility resonates with Shiller's "excess volatility puzzle" and gives a quantitative estimate of the role of self-reflexivity and feedback effects in financial markets [Shiller 1981, Cutler 1989, Fair 2002, Soros 1994, Bouchaud 2011, Filimonov 2012].

- The above fit corresponds to a power-law behavior, $k(\tau) \approx g\tau^{-\alpha}$ for $\tau \ll q_0$, and an exponential decay for larger lags. Therefore, a characteristic time scale of $q_0 \approx 3$ months appears, beyond which volatility feedback effects decay more rapidly.
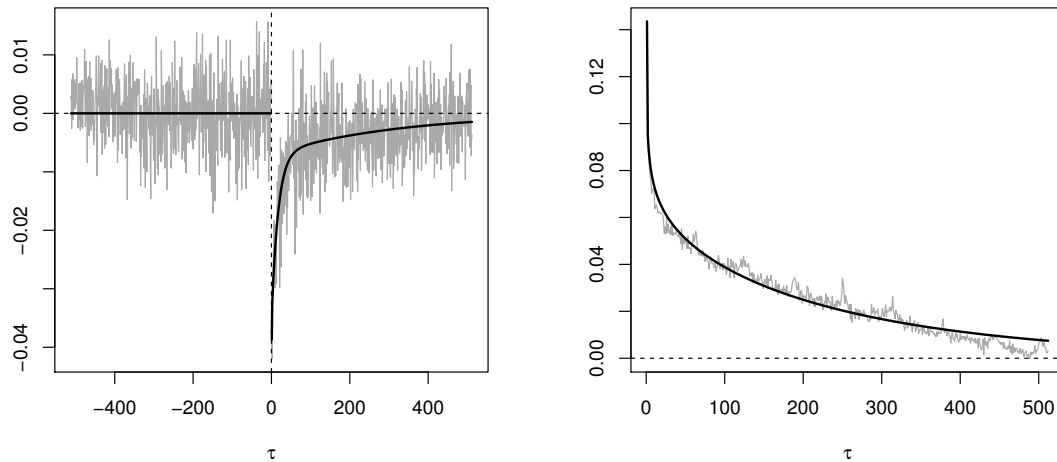
Figure 6.3: Two empirical correlation functions: the leverage $\mathcal{L}(\tau)$ and the correlation of amplitudes $\widetilde{C}^{(a)}(\tau)$, together with their fits. $\mathcal{L}(\tau)$ is fitted by the sum of two exponentials $-a\,\mathrm{e}^{-\tau/b} - c\,\mathrm{e}^{-\tau/d}$, with $a = 0.007$, $b = 327$ days, $c = 0.029$, $d = 17$ days; whereas $\widetilde{C}^{(a)}(\tau)$ is fitted by a power-law truncated by an exponential: $B\tau^{-\beta}\mathrm{e}^{-\tau/\tau_0}$, with $\beta = 0.14, B = 0.106, \tau_0 = 290$ days.
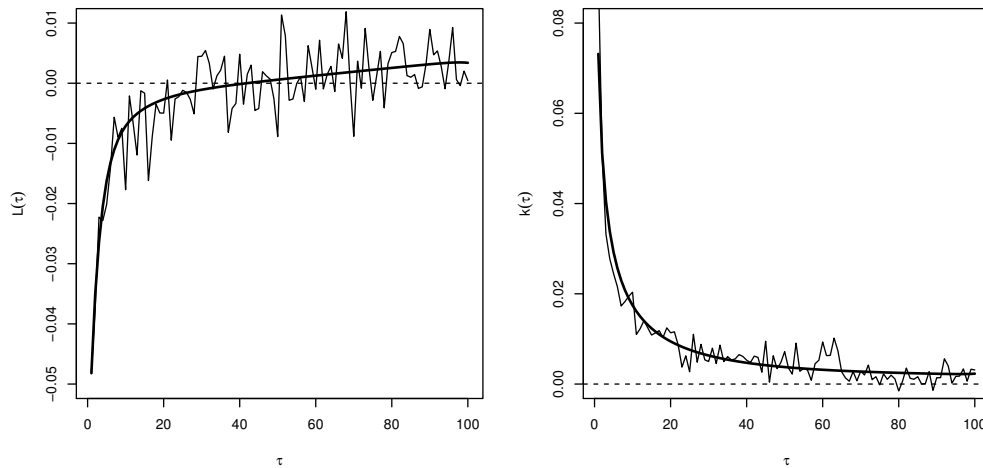
Figure 6.4: Calibration of the diagonal kernels for stocks, with $q = 100$. **Left:** $L(\tau)$. **Right:** $K(\tau, \tau)$. The thick curves are obtained with fitted instead of raw input correlation functions, see Fig. 6.3.
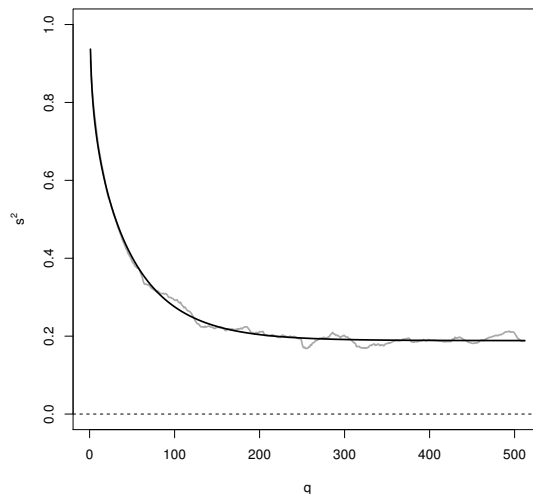
Figure 6.5: s^2 as a function of the farthest lag; the solid line is a fit according to the formula (6.31) (see text for parameter values).

- With a diagonal positive kernel K, the condition for positive definiteness (6.8) of the quadratic form reads $\sum_\tau L(\tau)^2/k(\tau) \leq 4s^2$. The estimated values of L and k yield a left-hand side equal to 0.595, while the right-hand side amounts to 0.823.

- Using the results of Sect. 6.2.2, one can compute the theoretical value of $\langle\sigma^4\rangle$ that corresponds to the empirically determined $k(\tau)$. As expected from the fact that g is small and α close to unity, one finds that the fluctuations of volatility induced by the long-memory feedback are weak: $\langle\sigma^4\rangle = 1.156$ (see Eq. (6.13) above). Including the contribution of the leverage kernel $L(\tau)$ to $\langle\sigma^4\rangle$ does not change much the final numerical value, that shifts from 1.156 to 1.161.

- The smallness of $\langle\sigma^4\rangle - \langle\sigma^2\rangle^2$ shows that most of the kurtosis of the returns $r_t = \sigma_t\xi_t$ must come from the noise ξ_t, which cannot be taken as Gaussian. Using the diagonal ARCH model with the kernels determined as above to predict σ_t, one can study the distribution of $\xi_t = r_t/\sigma_t$ and find the most likely Student-t distribution that accounts for it. We find that the optimal number of degrees of freedom is $\nu \approx 6.4$, and the resulting fit is shown in Fig. 6.6. Note that while the body and 'near-tails' of the distribution are well reproduced by the Student-t, the far-tails are still fatter than expected. This is in agreement with the commonly accepted tail index of $\nu_{\text{tail}} \approx 4$, significantly smaller than 6.4.

Assuming ξ_t to be a Student-t random variable with $\nu = 6.4$ degrees of freedom, we have re-estimated $k(\tau)$ and $L(\tau)$ using maximum-likelihood (see below). The final results are more noisy, but close to the above ones after fitting. Our strategy is thus to fix both $k(\tau)$ and $L(\tau)$, and only focus on the off-diagonal elements of K henceforth.
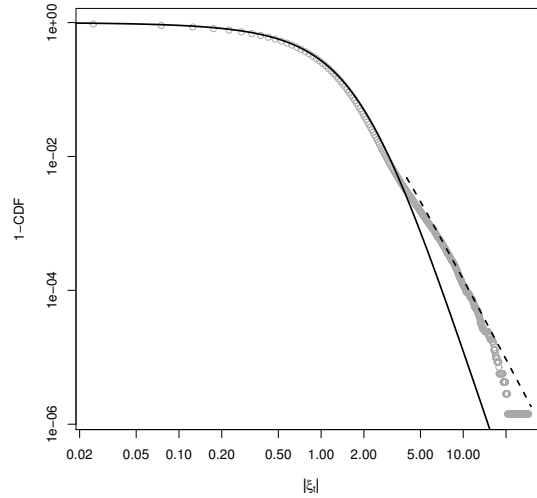
Figure 6.6: Cumulative distribution function of the residuals $\xi_t = r_t/\sigma_t$. The plain line is the Student distribution with best (maximum-likelihood) fitting tail parameter $\nu = 6.4$. Far tails suggest a fatter distribution with a smaller value of $\nu_{\text{tail}} \approx 4$ (dashed).

6.4.4   The off-diagonal kernel, GMM & maximum likelihood estimation

We can now go back to Eq. (6.29e) that allows one to solve for $K(\tau', \tau'' > \tau')$, once $k(\tau)$ and $L(\tau)$ are known. As announced above, we choose $q = 20$ for the time being. Because $\mathcal{D}$ involves the fourth moment of the returns, this procedure is not very stable, even with a lot of data pooled together, and even after the truncation of large returns. Maximum likelihood estimates would be more adapted here, but the dimensionality of the problem prevents a brute force determination of the $q(q-1)/2$ parameters. In order to gain some robustness, we use the following strategy. The Student log-likelihood per point $\mathcal{I}$, with ν degrees of freedom, is given by:[5]

$$\mathcal{I}_\nu(L, K, \{r_t\}) = \frac{1}{2T} \sum_t \left[\nu \ln a_t^2 - (\nu + 1) \ln(a_t^2 + r_t^2) \right], \quad a_t^2 \equiv (\nu - 2)\,\sigma_t^2, \qquad (6.32)$$

where $r_t \equiv \sigma_t \xi_t$ and σ_t^2 is given by the QARCH model expression, Eq. (6.4), and in this section K is a notation for the off-diagonal content only. We fix $\nu = 6.4$ and determine numerically the gradient $\partial \mathcal{I}_\nu$ and the Hessian $\partial\partial \mathcal{I}_\nu$ of $\mathcal{I}_\nu$ as a function of *all* the $q(q-1)/2$ off-diagonal coefficients $K(\tau', \tau'' > \tau')$, computed either around the GMM estimates of these parameters, or around the ARCH point where all these coefficients are zero. Note that $\partial \mathcal{I}_\nu$ is a vector with $q(q-1)/2$ entries and $\partial\partial \mathcal{I}_\nu$ is a $q(q-1)/2 \times q(q-1)/2$ matrix. It so happens that the eigenvalues of the Hessian are all found to be *negative*, i.e. the starting point is in the vicinity of a local maximum. This allows one to find easily the values of the $q(q-1)/2$

---

[5]In the following we do *not* truncate the large returns, but completely neglect the weak linear correlations $\mathcal{C}^{(1)}(\tau)$ that are present for small lags, and that should in principle be taken into account in the likelihood estimator.

parameters that maximize the value of $\mathcal{I}_\nu$; they are (symbolically) given by:

$$K^* = K_0 - \left(\overline{\partial\partial\mathcal{I}_\nu} \right)^{-1} \cdot \overline{\partial\mathcal{I}_\nu}, \tag{6.33}$$

where K_0 is the chosen starting point — either the GMM estimate $K_0 = K_{\mathrm{GMM}}$ based on Eq. (6.29e), or simply $K_0 = 0$ if one starts from a diagonal ARCH model — and the overline on top of $\partial\partial\mathcal{I}, \partial\mathcal{I}$ indicates averaging over stocks. This one step procedure is only approximate but can be iterated; it however assumes that the maximum is close to the chosen initial point, and would not work if some eigenvalues of the Hessian become positive. In our case, both starting points ($K_0 = 0$ or $K_0 = K_{\mathrm{GMM}}$) lead to nearly exactly the same solution; furthermore the Hessian recomputed at the solution point is very close to the initial Hessian, indicating that the likelihood is a locally quadratic function of the parameters, and the gradient evaluated at the solution point is very close to zero in all directions, confirming that a local maximum as been reached.

Before exposing our results, we briefly go through a digression to discuss the bias and error on the estimated parameters K^* as well as on the resulting maximal average likelihood $\overline{\mathcal{I}}^*$. The likelihood $\mathcal{I}$, its gradient $\partial\mathcal{I}$ and Hessian matrix $\partial\partial\mathcal{I}$ are generic functions of the set of parameters K to be estimated, and of the dataset, of size n. As the number n of observations tends to infinity, the covariance matrix of the ML estimator of the parameters is well known to be $(nI)^{-1}$, where I is the Fisher Information matrix

$$I = \mathbb{E}[-\partial\partial\mathcal{I}] \approx -\overline{\partial\partial\mathcal{I}}(K^*)$$

while the asymptotic bias scales as n^{-1} and is thus much smaller than the above error ($\sim n^{-1/2}$). As a consequence, ML estimates of K exceeding $\pm\,\mathrm{diag}(-n\,\overline{\partial\partial\mathcal{I}}^*)^{-1/2}$ will be deemed significant. The *bias* on the average in-sample (IS) value of the maximum likelihood itself can be computed to first order in $\frac{1}{n}$, and is very generally found to be $+M/2n$, where M is the number of parameters to be determined. Similarly, the bias on the average out-of-sample value of the maximum likelihood is $-M/2n$.[6] Since each sampling of our data set will contain $n = T \cdot N/2 \approx 350{,}000$ observations, differences of likelihood smaller than $M/2n \approx 3 \cdot 10^{-4}$ are insignificant when $M = 190$ (corresponding to all off-diagonal elements when $q = 20$). This number is ≈ 5 times smaller when one considers the restricted models introduced above (which contain ≈ 40 parameters).

The most likely off-diagonal coefficients of K^* are found to be highly significant (see Tab. 6.1): the IS increase of likelihood from the purely diagonal ARCH(q) model is $\Delta\mathcal{I} \approx 10^{-3}$ per point. This is confirmed by an out-of-sample (OS) experiment, where we determine K^* on half the pool of stocks and use it to predict the volatility on the other half (whence the above factor $1/2$ in the numerical estimation of n). The experiment is performed over $N_{\mathrm{samp}} = 150$ random pool samplings. The average OS likelihood is very significantly better for the full off-diagonal kernel K^* than for the diagonal ARCH(q), itself being better than the GMM estimate K_{GMM} based on Eq. (6.29e), and probably subject to biases due to the truncation procedure. Note that the full off-diagonal kernel K^* has many more parameters than the

---

[6]These corrections to the likelihoods lead to the (per point) Akaike Information Criterion [Akaike 1974] $AIC = -2(\mathcal{I} - M/n)$, that trades off the log-likelihood and the number of parameters. AIC is used for model selection purposes mainly. When comparing parametric models with the same number of parameters, AIC is not more powerful than the likelihood.
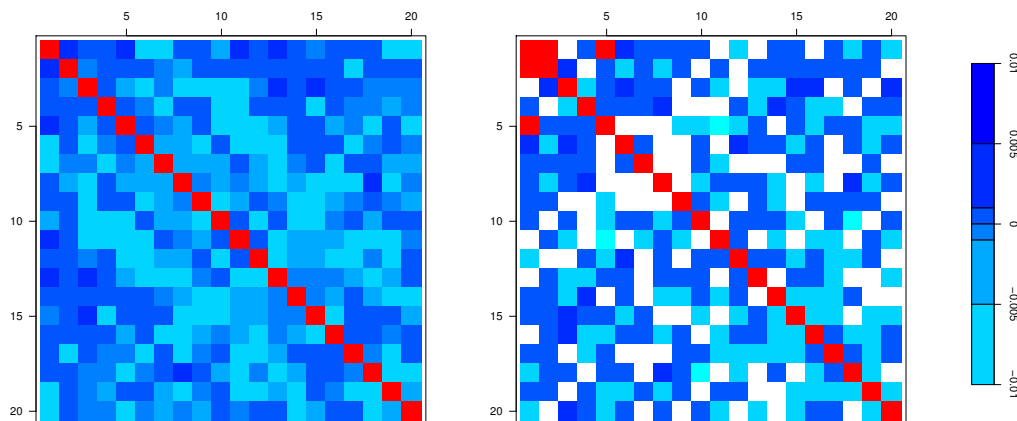
Figure 6.7: Heatmap of the unconstrained model. **Left:** GMM estimation; **Right:** ML estimation (white spots correspond to values smaller than their corresponding error margin). Red spots correspond to values larger than 0.01. Note the negative streaks at large lags, and the significant off-diagonal entry for $\tau = 1, \tau' = 2$ days.

diagonal ARCH(q); it therefore starts with a handicap out-of-sample since the bias on the OS likelihood is, as recalled above, $-M/2n$.

However, as announced above, the off-diagonal elements of K^* are a factor ten smaller than the corresponding diagonal values. We give a heat-map representation of the matrix K^* in Fig. 6.7. Various surprises are immediately apparent.

First, while the off-diagonal elements are mostly positive for small τ', τ'', clear negative streaks appear for intermediate and large τs. This is unexpected, since one would have naively guessed that any trend (i.e. positive realized correlations between returns) should *increase* future volatilities. Here we see that some quadratic combinations of past returns contribute negatively to the volatility. This will show up in the spectral properties of K^* (see Sect. 6.4.5 below).

The second surprise is that there does not seem to be any obvious structure of the matrix, that would be reminiscent of one of the simple models represented in Fig. 6.2. This means that the fine structure of volatility feedback effects is much more subtle than anticipated.

We have nevertheless implemented a *restricted* maximum-likelihood estimation that imposes the structure of one of the models considered in Sect. 6.3. We find that all these models are equally "bad" — although they lead to a significant increase of likelihood compared to the pure ARCH case, both IS and OS, they are all superseded by the unconstrained model shown in Fig. 6.7, again both IS and OS (see Tab. 6.1). The best OS model is "Long-trend", with a kernel $g_{LT}(\ell)$ shown in Fig. 6.8, together with the functions $g_2(\ell)$, $g_{BB}(\ell)$, $g_Z(\ell)$. While $g_{LT}(\ell)$ looks roughly like an exponential with memory time 10 days, the two-day return kernel $g_2(\ell)$ reveals intriguing oscillations. Two-day returns influence the volatility quite differently from one day returns! On the other hand, we do not find any convincing sign of the multi-scale "BB" structure postulated in [Borland 2011].

Note that the structure shown in Fig. 6.8 is found to be stable when q is changed. It would be interesting to subdivide the pool of stocks in different categories (for example, small
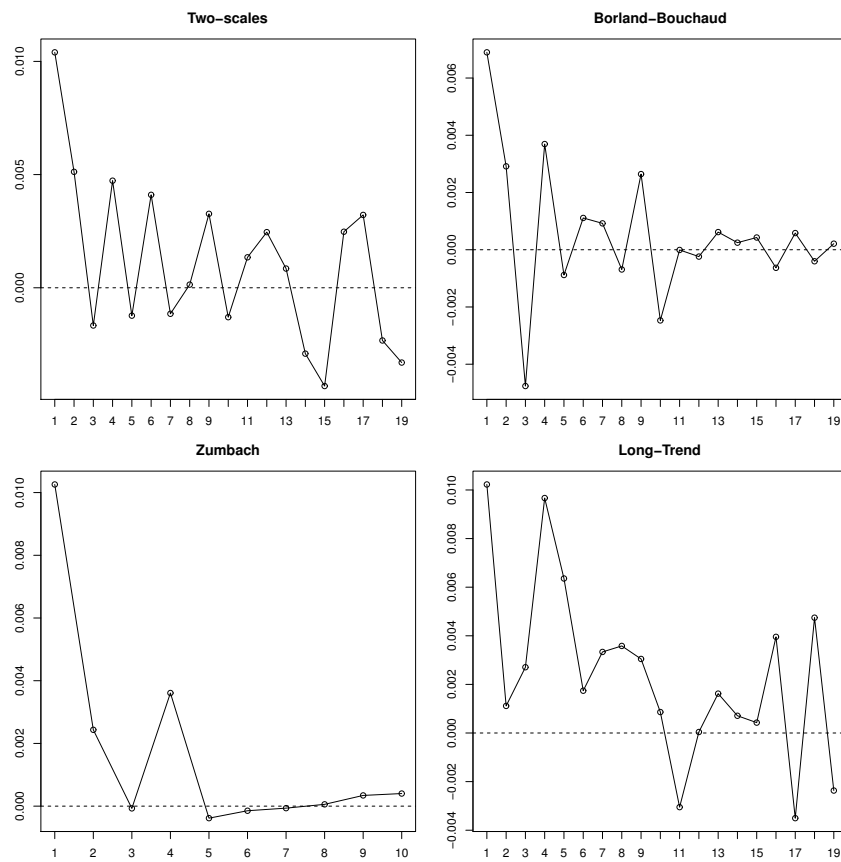
Figure 6.8: Plot of the empirically determined kernels $g_2(\ell)$, $g_{BB}(\ell)$, $g_Z(\ell)$ and $g_{LT}(\ell)$ for the restricted models of Sect. 6.3.

caps/large caps) or in different sub-periods, and study how the off-diagonal structure of K is affected. However, we note that the dispersion of likelihood over different samplings of the pool of stocks is only 50% larger than the "true" dispersion, due only to random samplings of a fixed QARCH model with parameters calibrated to the data (see caption of Tab. 6.1). This validates, at least as a first approximation, the assumption of homogeneity among all the stock series that have been averaged over.

To conclude this empirical part, we have performed several ex-post checks to be sure that our assumptions and preliminary estimations are justified. First, we have revisited the most likely value of the Student parameter ν (tail index of the distribution of the residuals $\xi(t) = r(t)/\sigma_{\mathrm{QARCH}}(t)$) with now the full matrix K^*, plus diagonal terms up to $q = 100$, and found again $\nu = 6.4$. This shows that our procedure is consistent from that point of view. Second, we have computed the quadratic correlation of the residuals ξ_t, which are assumed in the model to be IID random variables with, in particular, no variance autocorrelation: $\langle \xi_t^2 \xi_{t-\tau}^2 \rangle - 1 = 0$ for $\tau \neq 0$. Empirically, we observe a negative correlation of weak magnitude exponentially decaying with time. This additional dependence of the amplitude of the residuals, together with the excess fat tails of their probability distribution, is probably a manifestation of the truly exogenous events occurring in financial markets that have different statistical properties [Joulin 2008] and not captured by the endogenous feedback mechanism. Finally, about the universality hypothesis, we discuss in the caption of Tab. 6.1 how the assumption of homogeneous stocks is validated by comparing the cross-sectional dispersion of the likelihoods obtained empirically and on simulated series.

6.4.5   Spectral properties of the empirical kernel K

As discussed in Sect. 6.3.4, another way to decipher the structure of K is to look at its eigenvalues and eigenvectors. We show in Fig. 6.9 the eigenvalues of K^* as a function of the eigenvalues of the purely diagonal ARCH model. We see that a) the largest eigenvalue is clearly shifted upwards by the off-diagonal elements; b) the structure of the top eigenvector is non-trivial, and has positive contributions at all lags (up to noise); c) the unconstrained estimations — both GMM and ML — lead to 6 very small eigenvalues (perhaps even slightly negative) that all constrained models fail to reproduce.

The positiveness of all eigenvalues is not granted a priori, because nothing in the calibration procedure imposes the positivity of the matrix K^*. Although we would naively expect that past excitation could only lead to an amplification of future volatility (i.e. that only strictly positive modes should appear in the feedback kernel), we observe that quasi-neutral modes do occur. This is clearly related to the negative streaks noted above at large lags, but we have no intuitive interpretation for this effect at this stage.

## 6.5   Empirical study: stock index

We complete our analysis by a study of the returns of the S&P-500 index in the QARCH framework. We use a long series of more than 60 years, from Oct. 1948 to Sept. 2011 (15 837 days).

---

[8]Note that the true likelihood is not necessarily larger than the realized one under a misspecified model.

		GMM	ARCH(20)	ARCH+ML
IS		-1.31533	-1.31503	**-1.31405**
OS		-1.32003	-1.31971	**-1.31914**

ARCH+	BB	Z	2s	LT	2s+Z	2s+LT
IS	-1.31486	-1.31490	-1.31490	-1.31487	-1.31488	-1.31489
OS	-1.31960	-1.31957	-1.31962	-1.31957	-1.31956	-1.31957

Table 6.1: Log-likelihoods, according to Eq. (6.32). In-sample and out-of-sample likelihoods are computed as follows: for each of $N_{\text{samp}} = 150$ iterations, half of the stock names are randomly chosen for the calibration of K, L and the likelihood is computed with the obtained K^*, L^* on each series of the same sample ('In-sample' likelihoods). Then, the likelihood is again computed with the same parameters but on the series of the other sample ('Out-of-sample' likelihoods). While the former quantify how much the estimated model succeeds in reproducing the given sample, the latter measure the reliability of the model on *other* similar datasets. In order to quantify the validity of the model in an absolute way, the likelihood can be compared with the "true" value, obtained with simulated data (since an analytical treatment is out of reach). The average likelihood per point $\overline{\mathcal{I}}^*(r_t)$ with r_t simulated as a QARCH with parameters K^*, L^*, and $\nu = 6.4$ is equal to -1.34019, which is 1.5% away from the empirical values[8]. The likelihoods reported in the table are averages over all samplings, and the corresponding 1-s.d. dispersion is found to be $\approx 3 \cdot 10^{-3}$ in all cases, to be compared to $2 \cdot 10^{-3}$ for random samplings of a fixed QARCH model with the same parameters.
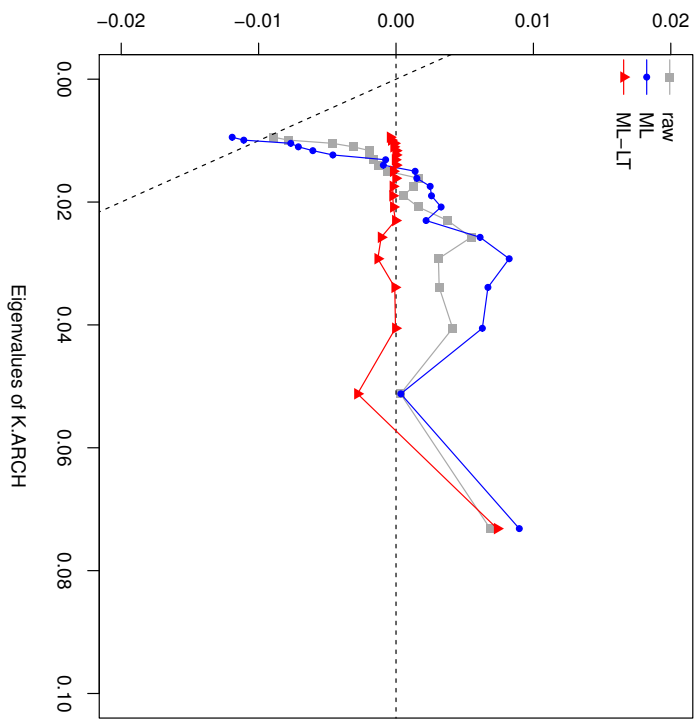
(a) The difference between the ranked eigenvalues of the estimated kernels and those of K_{ARCH} as a function of the latter (again ranked). The dashed oblique line has slope -1 and separates positive eigenvalues from negative ones.
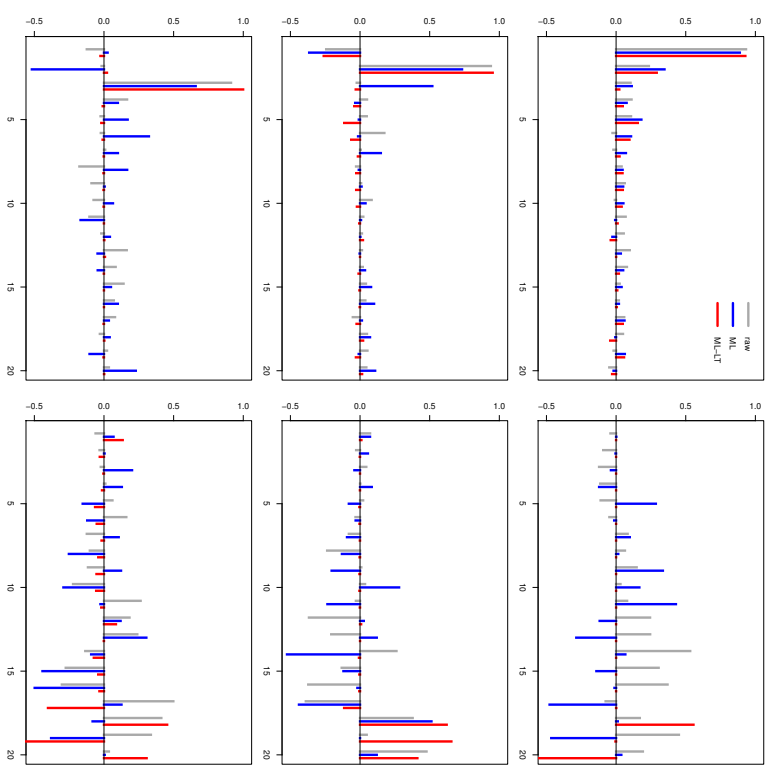
(b) Structure of the first three and last three eigenvectors. Whatever the estimation method, the first eigenvector has a non trivial structure, with mostly positive components, indicating a genuine departure from the diagonal ARCH benchmark, for which we would find a single peak at $\tau = 1$.

Figure 6.9: Spectral decomposition of the feedback kernel K, for the GMM, ML and ML+LT estimates.
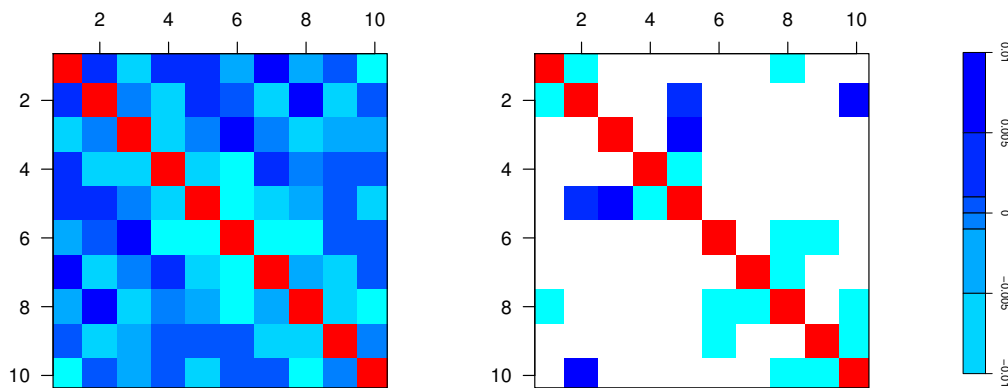
Figure 6.10: Heatmap of the $q = 10$ kernel for the index volatility. **Left:** the GMM estimation; **Right:** the ML estimation with GMM prior (again, we have checked that the ARCH prior leads to very close results).

The computation of the correlation functions and the GMM calibration of a long ARCH(512) yields a $s^2(q)$ that can be fitted with the formula (6.31) and the following parameters: $s^2_\infty \approx 0.20, \alpha \approx 1.28, g \approx 0.162$ and $q_0 \approx 262$. Contrarily to single stocks, the One-step Maximum Likelihood estimation failed with $q = 20$, as the gradient and Hessian matrix evaluated at the arrival point are, respectively, large and not negative definite. Although the starting point appears to be close to a local maximum (the Hessian matrix is negative definite), the one-step procedure does not lead to that maximum.

In order to control better the Maximum Likelihood estimation, we lower the dimensionality of the parameter space and estimate a QARCH(10), although still with a long memory diagonal ($q = 50$). Here the procedure turns out to be legitimate, and the resulting kernel K is depicted in Fig. 6.10 (right). Interestingly, the off-diagonal content in the QARCH model for index returns is mostly not significant (again, white regions correspond to values not exceeding their theoretical uncertainty) apart from a handful of negative values around $\tau = 8 - 10$ and one row/column at $\tau = 5$. The contribution of the latter to the QARCH feedback can be made explicit as

$$2\, r_{t-5} \sum_{\tau < 5} K(5, \tau) r_{t-\tau}$$

and describes a trend effect between the daily return last week r_{t-5} and the (weighted) cumulated return since then $\sum_{\tau<5} K(5,\tau)r_{t-\tau}$. It would be interesting to know whether this finding is supported by some intuitive feature of the trading activity on the S&P-500 index. Note that, again, none of the "simple" structures discussed in Sect. 6.3 is able to account for the structure of K^* (compare Fig. 6.10 with Fig. 6.2).

The spectral analysis reveals a large eigenvalue much above the ARCH prediction, and a couple of very small eigenvalues, similarly to what was found for the stock data. However, the eigenvectors associated with them exhibit different patterns: the first eigenvector does not reveal the expected collective behavior, but rather a dominant $\tau = 1$ component, with a

	GMM	ARCH(10)	ARCH+ML
IS	-1.16750	-1.16666	-1.16522
OS	-1.16972	-1.16704	-1.17079

Table 6.2: Average log-likelihoods, according to Eq. (6.32), for the stock index.

significant $\tau = 2$ component of opposite sign. The other modes do not show a clear signature and are hard to interpret.

The procedure for computing in-sample/out-of-sample likelihoods is similar to the case of the stock data, but the definitions of the universes differs somewhat since we only have a single time series at our disposal. Instead of randomly selecting half of the series, we select half of the dates (in block, to avoid breaking the time dependences) to define the in-sample universe Ω, on which the correlation functions are computed and both estimation methods (GMM, and one-step maximum likelihood) are applied. Then we evaluate the likelihoods of the calibrated kernels, first on Ω to obtain the 'in-sample' likelihoods, and then on the complement of Ω to get the 'out-of-sample' likelihoods. We iterate $N_{\text{samp}} = 150$ times and draw a random subset of dates every time, then average the likelihoods, that we report in the Table 6.2. The 1-s.d. systematic dispersion of the samplings is now $\approx 7 \cdot 10^{-3}$. Because of the fewer number of observations in the index data compared to the stock data, corrections for the bias induced by the number of parameters M become relevant. Adjusting the out-of-sample likelihood by subtracting the bias $-M/2n \approx 3 \cdot 10^{-3}$ (with $n \approx T/2 = 7.5 \cdot 10^3$ and $M = q(q-1)/2 = 45$), brings the ARCH+ML result to a level competitive with ARCH (but not obviously better), and certainly better than the GMM estimate.

6.6   Time-reversal invariance

As noticed in the introduction, QARCH models violate, by construction, time-reversal invariance. Still, the correlation of the squared returns $\mathcal{C}^{(2)}(\tau)$ is trivially invariant under time-reversal, i.e. $\mathcal{C}^{(2)}(\tau) = \mathcal{C}^{(2)}(-\tau)$. However, the correlation of the true squared volatility with past squared returns $\widetilde{\mathcal{C}}^{(2)}(\tau)$ is in general not (see [Pomeau 1985, Zumbach 2009] for a general discussion). A measure of TRI violations is therefore provided by the integrated difference $\Delta(\tau)$:

$$\Delta(\tau) = \sum_{\tau'=1}^{\tau} \left[\widetilde{\mathcal{C}}^{(2)}(\tau') - \widetilde{\mathcal{C}}^{(2)}(-\tau') \right]. \tag{6.34}$$

The empirical determination of $\widetilde{\mathcal{C}}^{(2)}(\tau)$ and $\Delta(\tau)$ for stock returns is shown in Fig. 6.11. Although less strong than for simulated data (see Fig. 6.12), we indeed find a clear signal of TRI violation for stock returns, in agreement with a related study by Zumbach [Zumbach 2009]. We compare in Fig. 6.12 the quantity $\Delta(\tau)$ obtained from a *bona fide* numerical simulation of the model, with previously estimated parameters. Note that any measurement noise on the volatility σ_t^2 tends to reduce the TRI violations, but we have performed the numerical simulation in a way to reproduce this measurement noise as faithfully as possible.

However, the alert reader should worry that the existence of asymmetric leverage correlations $\mathcal{L}(\tau > 0) \neq 0$ between past returns and future volatilities is in itself a TRI-violating

mechanism, which has nothing to do with the ARCH feedback mechanism. In order to ascertain that the effect we observe is not a spurious consequence of the leverage effect, we have also computed the contribution of $\mathcal{L}(\tau)$ to $\Delta(\tau)$, which reads to lowest order and schematically:

$$\Delta(\tau) = \sum_{\tau'=1}^{\tau} L(\tau') \left[\mathcal{L}(\tau' - \tau) - \mathcal{L}(\tau' + \tau)\right] + K \text{ contributions.} \qquad (6.35)$$

The first term on the right-hand side is plotted in the inset of Fig. 6.11, and is found to have a *negative* sign, and an amplitude much smaller than $\Delta(\tau)$ itself. It is therefore quite clear that the TRI-violation reported here is genuinely associated to the ARCH mechanism and not to the leverage effect, a conclusion that concurs with that of [Zumbach 2009].

Still, the smallness of the empirical asymmetry compared with the simulation results suggests that the ARCH mechanism is "too deterministic". It indeed seems reasonable to think that the baseline volatility s^2 has no reason to be constant, but may contain an extra random contribution. Writing

$$\sigma^2(t) = \sigma_A^2(t) + \omega(t); \qquad \langle \omega \rangle = 0; \qquad \langle \omega(t)\omega(t-\tau) \rangle \equiv \mathcal{C}_\omega(\tau) = \mathcal{C}_\omega(-\tau)$$

with ω_t a noise contribution and σ_A the ARCH volatility[9] (i.e. deterministic when conditioned on past returns), then the asymmetry is found to be given by:

$$\Delta(\tau) = \Delta_A(\tau) - \sum_{\tau''=1}^{\tau} \sum_{\tau'=1}^{q} k(\tau') \left[\mathcal{C}_\omega(\tau' - \tau'') - \mathcal{C}_\omega(\tau' + \tau'')\right].$$

If one assumes that the correlation function $\mathcal{C}_\omega$ is positive and decays with time, the extra contribution to the asymmetry is negative, and reduces the observed TRI. This conclusion speaks in favor of a mixed approach to volatility modeling, bringing together elements of autoregressive QARCH models with those of stochastic volatility models. It would in fact be quite surprising that (although unobservable) the volatility should be a purely deterministic function of past returns. Although the behavioral interpretation of the above construction is not clear at this stage, the uncertainty on the baseline volatility level s^2 could come, for example, from true exogenous factors that mix in with the volatility feedback component described by the QARCH framework.

6.7 Conclusion, extensions

We have revisited the QARCH model, which postulates that the volatility today can be expressed as a general quadratic form of the past daily returns r_t. The standard ARCH or GARCH framework is recovered when the quadratic kernel is diagonal, which means that only past squared daily returns feedback on today's volatility. This is a very restrictive *a priori* assumption, and the aim of the present study was to unveil the possible influence of other quadratic combinations of past returns, such as, for example, square weekly returns. We have defined and studied several sub-families of QARCH models that make intuitive sense, and with a restricted number of free parameters.

---

[9]For the sake of clarity we consider here a diagonal ARCH framework, but the argument is straightforwardly generalized for a complete QARCH.
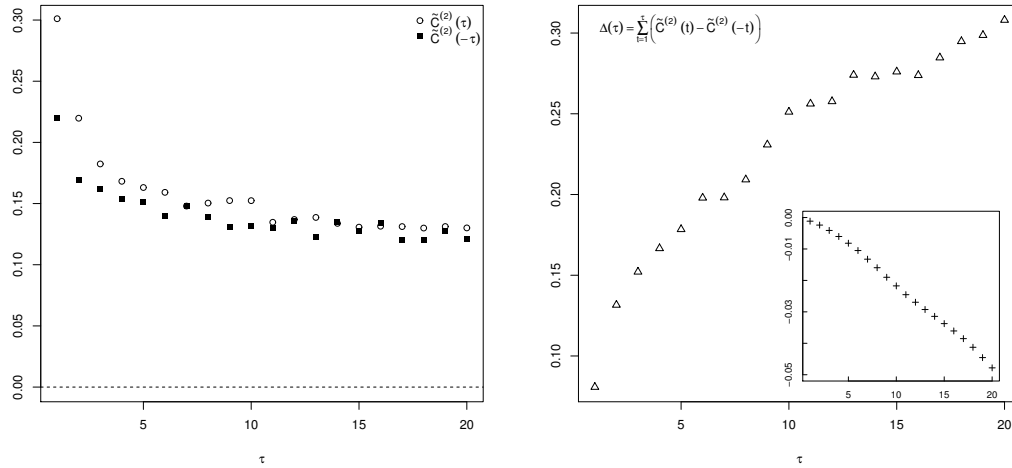
Figure 6.11: Measure of time-reversal asymmetry (Eq. 6.34) for the stock data. **In-set:**   The contribution to $\Delta(\tau)$ stemming from the leverage, i.e. the quantity $\sum_{\tau'=1}^{\tau} L(\tau')\left[\mathcal{L}(\tau'-\tau) - \mathcal{L}(\tau'+\tau)\right]$. Note that this contribution is negative, and an order of magnitude smaller than $\Delta(\tau)$ itself.
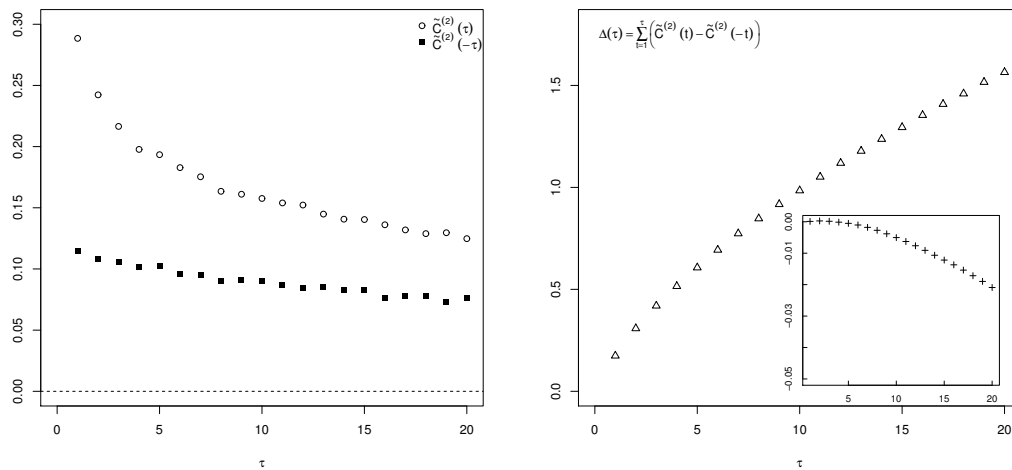
Figure 6.12: Measure of time-reversal asymmetry (Eq. 6.34) for a simulated ARCH model with Student ($\nu = 4$) residuals on the 5 minute scale. The parameters of the simulation are the estimated kernel $k^*(\tau)$ and $L^*(\tau)$ for stocks, with $q = 20$. At each date, 100 intraday prices are simulated (corresponding to the number of 5-minutes bins inside 8 hours) with the *same* σ_t^2 given by the QARCH model. The volatility is then computed using Rogers-Satchell's estimator, exactly as for empirical data.

The calibration of these models on US stock returns has revealed several features that we did not anticipate. First, although the off-diagonal (non ARCH) coefficients of the quadratic kernel are found to be highly significant both in-sample and out-of-sample, they are one order of magnitude smaller than the diagonal elements. This confirms that daily returns indeed play a special role in the volatility feedback mechanism, as postulated by ARCH models. Returns on different time scales can be thought of as a perturbation of the main ARCH effect. The second surprise is that the structure of the quadratic kernel does not abide with any of the simpler QARCH models proposed so-far in the literature. The fine structure of volatility feedback is much more subtle than anticipated. In particular, neither the model proposed in [Borland 2011] (where returns over several horizons play a special role), nor the trend model of Zumbach in [Zumbach 2010] are supported by the data. The third surprise is that some off-diagonal coefficients of the kernel are found to be negative for large lags, meaning that some quadratic combinations of past returns contribute negatively to the volatility. This also shows up in the spectral properties of the kernel, which is found to have very small eigenvalues, suggesting the existence of unexpected volatility-neutral patterns of past returns.

As for the diagonal part of the quadratic kernel, our results are fully in line with all past studies: the influence of the past squared-return $r_{t-\tau}^2$ on today's volatility σ_t^2 decays as a power-law $g\,\tau^{-\alpha}$ with the lag τ, at least up to $\tau \approx 2$ months, with an exponent α close to unity ($\alpha \approx 1.11$), which is the critical value below which the volatility diverges and the model becomes non-stationary. As emphasized in [Borland 2011], markets seem to operate close to criticality (this was also noted in different contexts, see [Bouchaud 2004, Bouchaud 2006, Bacry 2012, Tóth 2011, Filimonov 2012, Parisi 2012] for example). The smallness of $\alpha - 1$ has several important consequences: first, this leads to long-memory in the volatility; second, the average square volatility is a factor 5 higher than the baseline volatility, in line with the excess volatility story [Shiller 1981]: most of the market volatility appears to be endogenous and comes from self-reflexive, feedback effects (see e.g. [Soros 1994, Bouchaud 2011, Filimonov 2012] and references therein). Third, somewhat paradoxically, the long memory nature of the kernel leads to *small* fluctuations of the volatility. This is due to a self-averaging mechanism occurring in the feedback sum, that kills fat tails. This means that the high kurtosis of the returns in ARCH models cannot be ascribed to volatility fluctuations but rather to leptokurtic residuals, also known as unexpected price jumps.

Related to price jumps, we should add the following interesting remark that stresses the difference between endogenous jumps and exogenous jumps within the ARCH framework. Several studies have revealed that the volatility relaxes after a jump as a power-law, akin to Omori's law for earthquake aftershocks: $\sigma_\tau^2 \sim \sigma_0^2 \tau^{-\theta}$, where $t = 0$ is the time of the jump. The value of the exponent θ seems to depend on the nature of the initial price jump. When the jump occurs because of an exogenous news, $\theta \approx 1$ [Lillo 2003, Joulin 2008], whereas when the jump appears to be of endogenous origin, the value of θ falls around $\theta \approx \frac{1}{2}$ [Zawadowski 2006, Joulin 2008]. In other words, as noted in [Joulin 2008], the volatility seems to resume its normal course faster after a news than when the jump seems to come from nowhere. A similar difference in the response to exogenous and endogenous shocks was also reported in [Sornette 2004] for book sales. Now, if one simulates long histories of prices generated using an ARCH model with a diagonal kernel decaying as $g\,\tau^{-\alpha}$, one can measure the exponent θ by conditioning on a large price jump (which can only be endogenous, by definition!). One finds that θ varies continuously with the amplitude of the initial jump, and

saturates around $\theta \approx \frac{1}{2}$ for large jumps (we have not found a way to show this analytically). A similar behavior is found within multifractal models as well [Sornette 2003]. If on the other hand an exogenous jump is artificially introduced in the time series by imposing a very large value of $\xi_{t=0}$, one expects the volatility to follow the decay of the kernel and decay as $g\,\tau^{-\alpha}\xi_0^2$, leading to $\theta = \alpha \approx 1$. Therefore, the dichotomy between endogenous and exogenous shocks seem to be well reproduced within the ARCH framework.

Finally, we have emphasized the fact that QARCH models are by construction backward looking, and predict clear Time-Reversal Invariance (TRI) violations for the volatility/square-return correlation function. Such violations are indeed observed empirically, although the magnitude of the effect is quite a lot smaller than predicted. This suggests that QARCH models, which postulate a *deterministic* relation between volatility and past returns, discard another important source of fluctuations. We postulate that "the" grand model should include both ARCH-like feedback effects and stochastic volatility effects, in such a way that TRI is only weakly broken. The stochastic volatility component could be the source of the extra kurtosis of the residuals noted above.[10]

This study is, to the best of our knowledge, the first attempt at unveiling the fine structure volatility feedback effects in autoregressive models. We believe that it is a step beyond the traditional econometric approach of postulating a convenient mathematical model, which is then brute-force calibrated on empirical data. What we really need is to identify the *underlying mechanisms* that would justify, at a deeper level, the use of a QARCH family of model rather than any another one, for example the multifractal random walk model. From this point of view, we find remarkable that the influence of daily returns is so strongly singled out by our empirical results, when we expected that other time scales would emerge as well. The quandary lies in the unexpectedly complex structure of the off-diagonal feedback component, for which we have no interpretation.

A natural extension of our study that should shed further light on a possible behavioral interpretation of volatility feedback is to decompose daily returns into higher frequency components, for example overnight and intraday returns. Preliminary investigations in this direction point toward a clear distinction of the intraday and overnight mechanisms of volatility dynamics.[11] In particular, the self-excitement at night is characterized by both a smaller baseline intensity and fatter-tailed residuals than during the day, indicating that there are extreme exogenous shocks occurring overnight, which however are not propagated since the dynamics results even more from the endogenous feedback at night than intraday.

The feedback mechanism of the volatility at the intraday scale (e.g. 5 minute returns) might be related to the self-excitement of order arrivals in the book [Bacry 2012]. Many other remaining questions should be addressed empirically, for example the dependence of the feedback effects on market capitalization, average volatility, etc. We have chosen to scale out the market-wide volatility, but other choices would be possible, such as a double regression on the past returns of stocks and of the index. Finally, other financial assets, such as currencies or commodities, should be studied as well. Stocks, however, offer the advantage that the data is much more abundant, specially if one chooses to invoke some structural universality, and to treat all stocks as different realizations of the same underlying process.

---

[10]This discussion might be related to the interesting observation made by Virasoro in [Virasoro 2011].

[11]A detailed study of intraday and overnight effects in the feedback mechanisms of volatility dynamics will be reported elsewhere [Blanc 2013]

Copulas in time series

Contents

7.1   Introduction

As a tool to study the — possibly highly non-linear — correlations between returns, "copulas" have long been used in actuarial sciences and finance to describe and model cross-dependences of assets, often in a risk management perspective [Embrechts 2003, Embrechts 2002a, Malevergne 2006]. Although the widespread use of simple analytical copulas to model multivariate dependences is more and more criticized [Mikosch 2006, Chicheportiche 2012b], copulas remain useful as a tool to investigate empirical properties of multivariate data [Chicheportiche 2012b].

More recently, copulas have also been studied in the context of self-dependent univariate time series, where they find yet another application range: just as Pearson's ρ coefficient is commonly used to measure both linear cross-dependences and temporal correlations, copulas are well-designed to assess non-linear dependences both across assets or in time [Beare 2010, Ibragimov 2008, Patton 2009] — we will speak of "self-copulas" in the latter case.

Notations

We consider a time series of length T, that can be seen as one realization of a corresponding discrete stochastic process $\{X_t\}_{t=1...T}$. The joint cumulative distribution function (CDF) of n occurrences ($1 \le t_1 < \ldots < t_n < T$) of the process is

$$\mathcal{F}_{t_1,\ldots,t_n}(\mathbf{x}) = \mathbb{P}[X_{t_1} < x_{t_1}, \ldots, X_{t_n} < x_{t_n}]. \tag{7.1}$$

We assume that the generating process $\{X_t\}$ of the series has a strong-stationary marginal CDF F, and let the joint distribution $\mathcal{F}$ have long-ranged dependences. Strictly speaking, most of the discussion that follows only relies on *time-homogeneity* of the copula, i.e. translational invariance of the dependence structure, and does not need stationarity of the marginal. However, in this class of problems much related to empirical data, it is more convenient (although sometimes not justified) to assume at least a weak form of stationarity, in order to give sense to the notion of empirical CDF. In the following, we will always assume stationarity of the marginal and homogeneity of the copula.

A realization of X_t at date t will be called an "event" when its value exceeds a threshold $X^{(\pm)}$: "negative event" when $X_t < X^{(-)}$, and "positive event" when $X_t > X^{(+)}$. The probability p_- of such a 'negative event' is $F(X^{(-)})$, and similarly, the probability that X_t is above a threshold $X^{(+)}$ is the tail probability $p_+ = 1 - F(X^{(+)})$.

If a unique threshold $X^{(+)} = X^{(-)}$ is chosen, then obviously $p_+ = 1 - p_-$. This is appropriate when the distribution is one-sided, typically for positive only signals, and one wishes to distinguish between two regimes: extreme events (above the unique threshold), and regular events (below the threshold). This case is illustrated schematically in Fig. 7.1(b). When the distribution is two-sided, it is more convenient to define, $X^{(+)}$ as the q-th quantile of F, and $X^{(-)}$ as the $(1-q)$-th quantile, for a given $q \in [\frac{1}{2}, 1]$, so that $p_+ = p_- = 1 - q$. This allows to investigate persistence and reversion effects in *signed* extreme events, while excluding a neutral zone of regular events between $X^{(-)}$ and $X^{(+)}$, see Fig. 7.1(a)

When the threshold for the recurrence is defined in terms of quantiles like above (a *relative* threshold), stationarity is not needed theoretically but much wanted empirically as already said (otherwise the height of the threshold might change every time). In contrast, when the

(a) $F(X^{(+)}) = 1 - F(X^{(-)}) = q$ (b) $F(X^{(+)}) = F(X^{(-)}) = q$
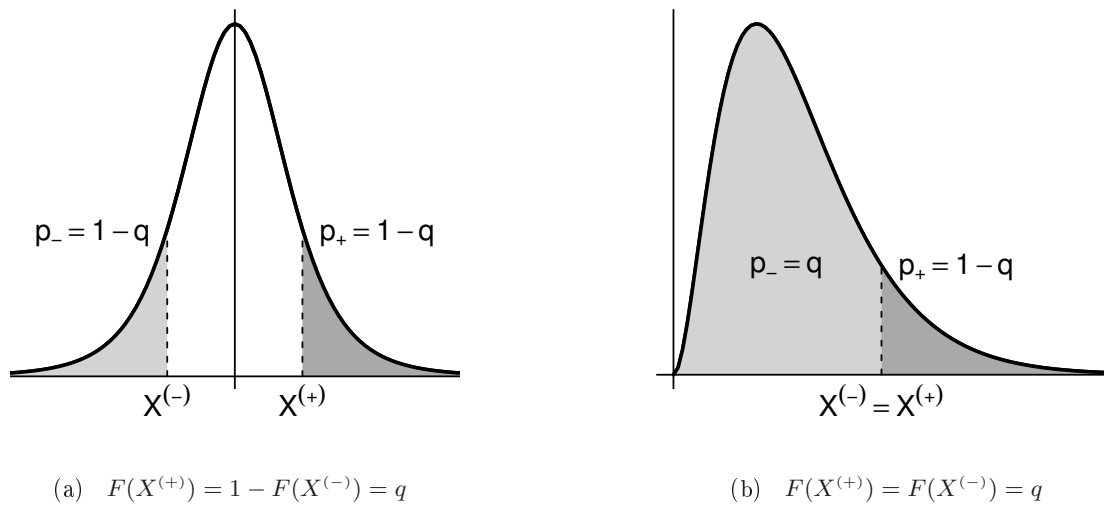
Figure 7.1: Two possible definitions of events: either p_- and p_+ are probabilities of extremes (negative and positive, respectively), or only p_+ is a probability of extreme and $p_- = 1 - p_+$.

threshold is set as a number (an *absolute* threshold), there's no issue on the empirical side, but the theoretical discussion makes sense only under stationary marginal.

7.2 Discrete time

In this section, we consider the case where the discrete times t_n in the definition (7.1) are equidistant ("regularly sampled"). A short discussion of the case with arbitrary (continuous) time stamps is provided in the next section.

7.2.1 Two-points dependence measures

Typical measures of dependences in stationary processes are two-points expectations that only involve one parameter: the lag ℓ separating the points in time. For example, the usefulness of the linear correlation function

$$\rho(\ell) = \mathbb{E}[X_t X_{t+\ell}] - \mathbb{E}[X_t]\,\mathbb{E}[X_{t+\ell}] \tag{7.2}$$

is rooted in the analysis of Gaussian processes, as those are completely characterized by their covariances, and multi-linear correlations are reducible to all combinations of 2-points expectations. Some non-linear dependences, like the tail-dependence for example (defined in part I), are however not expressed in terms of simple correlations, but involve the whole bivariate copula:

$$C_\ell(u,v) = \mathcal{F}_{t,t+\ell}(F^{-1}(u), F^{-1}(v)), \tag{7.3}$$

where $(u,v) \in [0,1]^2$. C_ℓ contains the full information on bivariate dependence that is invariant under increasing transformations of the marginal. For example, the conditional

probability

$$p_{++}^{(\ell)} = \mathbb{P}[X_{t+\ell} > X^{(+)} | X_t > X^{(+)}],$$

which is a measure of *persistence* of the "positive" events, can be written in terms of copulas, together with all three other cases of conditioning

$$p_{++}^{(\ell)} = [2p_+ - 1 + C_\ell(1-p_+, 1-p_+)]/p_+, \tag{7.4a}$$

$$p_{--}^{(\ell)} = C_\ell(p_-, p_-)/p_-, \tag{7.4b}$$

$$p_{-+}^{(\ell)} = [p_- - C_\ell(p_-, 1-p_+)]/p_-, \tag{7.4c}$$

$$p_{+-}^{(\ell)} = [p_- - C_\ell(1-p_+, p_-)]/p_+. \tag{7.4d}$$

When $X^{(+)} = X^{(-)} = 0$ and $\ell = 1$, this is exactly the definition of Bogùna and Masoliver in Ref. [Boguná 2004], with accordingly $p_- = p_+ = F(0)$, see Fig. 7.1. Note also that $p_{\pm\pm}^{(\ell)}$ and $p_{\pm\mp}^{(\ell)}$ are straightforwardly related to the tail dependence coefficients through the relations (2.16) of page 20.

   In addition to caring for frequencies of conditional events, one can try to characterize their magnitude. For a single random variable with cdf F, this can be quantified for example by the Expected Shortfall (or Tail Conditional Expectation), i.e. the average loss when conditioning on large events:

$$\text{ES}(p_-) = \mathbb{E}[X_t | X_t < X^{(-)}] = \frac{1}{p_-} \int_{-\infty}^{F^{-1}(p_-)} x \, \mathrm{d}F(x) = \frac{1}{p_-} \int_0^{p_-} F^{-1}(p) \, \mathrm{d}p$$

In the same spirit, for bivariate distributions, the mean return conditionally on preceding return 'sign' is defined:

$$\langle X \rangle_+^{(\ell)} = \mathbb{E}[X_t | X_{t-\ell} > X^{(+)}] \tag{7.5a}$$

$$\langle X \rangle_-^{(\ell)} = \mathbb{E}[X_t | X_{t-\ell} < X^{(-)}]. \tag{7.5b}$$

As an example, consider the Gaussian bivariate copula of the pair $(X_t, X_{t+\ell})$, whose whole ℓ-dependence is in the linear correlation coefficient $\rho(\ell)$. Fig. 7.2(a) illustrates the conditional probabilities (7.4) as a function of the threshold when $p_+ = p_- = 1 - q$, and Fig. 7.3 shows the conditional Expected Shortfall that can be computed exactly from Eqs. (7.5), and is proportional to the inverse Mill ratio:

$$\langle X \rangle_\pm = \pm \rho(\ell) \frac{\Phi'(X^{(\pm)})}{p_\pm},$$

where Φ denotes the CDF of the univariate standard normal distribution.

Aftershocks

Omori's law characterizes the ℓ dependence of $p_{++}^{(\ell)}$, i.e. the average frequency of events occurring ℓ time steps after a main event. It has been stated first in the context of earthquakes occurrences [Omori 1895], where this time dependence is power-law:

$$p_{++}^{(\ell)} = \lambda \cdot \ell^{-\alpha}. \tag{7.6}$$
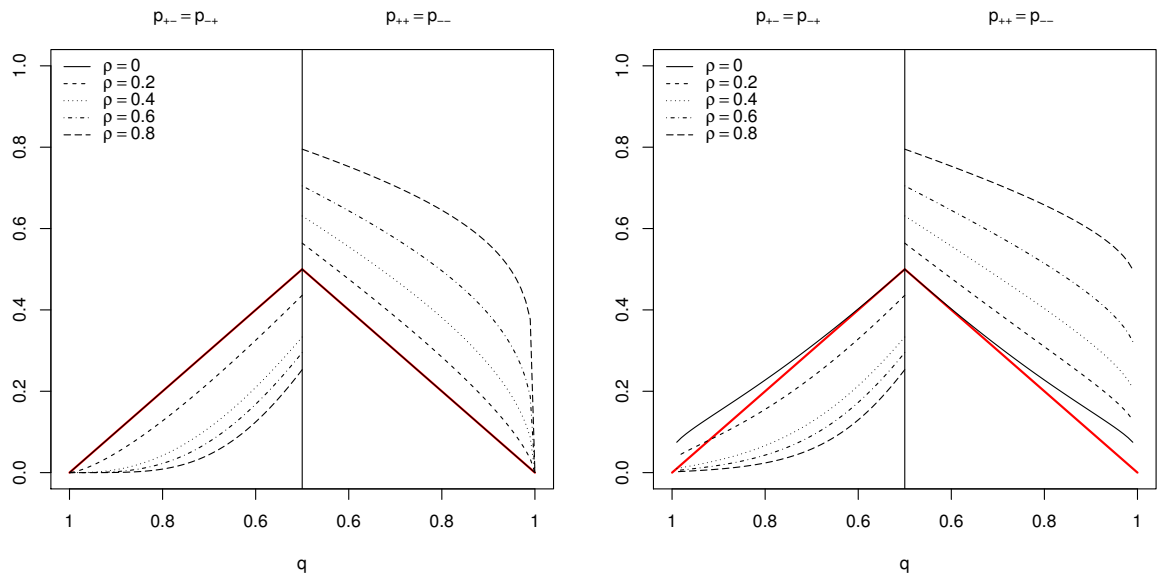
(a) Gaussian copula

(b) Student copula with $\nu = 5$ (see also Fig. 2.4 on page 31)

Figure 7.2: Conditional probabilities $p_{\pm\mp}^{(\ell)}$ and $p_{\pm\pm}^{(\ell)}$ for different values of $\rho(\ell)$ with thresholds at $p_+ = p_- = 1 - q$. The value at $q = 0.5$ is $\frac{1}{2} + \frac{1}{\pi} \arcsin \rho(\ell)$.
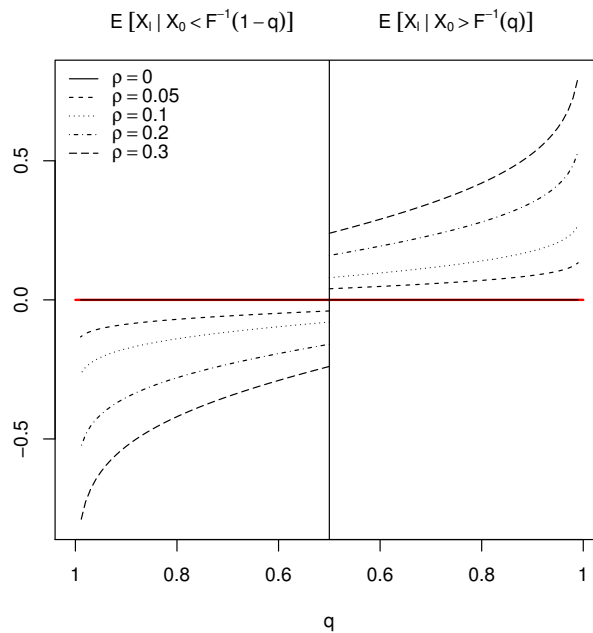
Figure 7.3: Conditional Expected Shortfall of a Gaussian pair (X_0, X_ℓ) for different values of $\rho(\ell)$. The value at $q = 0.5$ is $\sqrt{\frac{2}{\pi}} \, \rho(\ell)$.

Notice that any dependence on the threshold must be hidden in λ according to this description. The average *cumulated* number N_ℓ of these aftershocks until ℓ is thus

$$\langle N_\ell \rangle_+ = \lambda \cdot \ell^{1-\alpha}/(1-\alpha),$$

with in fact $\lambda \equiv p_+$ since, when $\alpha \to 0$, N_ℓ has no time-dependence i.e. it counts independent events (white noise), and $p_{++}^{(\ell)}$ must thus tend to the unconditional probability.

In order to give a phenomenological grounding to this empirical law also later observed in finance [Weber 2007, Petersen 2010], the authors of Ref. [Lillo 2003] model the aftershock volatilities in financial markets as a decaying scale $\sigma(\ell)$ times an independent stochastic amplitude r_ℓ with CDF ϕ. As a consequence, $p_{++}^{(\ell)} \sim 1 - \phi(X^{(+)}/\sigma(\ell))$ and the power-law behavior of Omori's law results from (i) power-law marginal $\phi(r) \sim r^{-\gamma}$, and (ii) scale decaying as power-law $\sigma(t) \sim t^{-\beta}$, so that relation (7.6) is recovered with $\alpha = \beta\gamma$. The non-stationarity described by σ is only introduced in a *conditional* sense, and might be appropriate for aging systems or financial markets, but we believe that Omori's law can be accounted for in a stationary setting and without necessarily having power-law distributed amplitudes.

The scaling of $p_{++}^{(\ell)}$ with the magnitude of the main shock is encoded in the prefactor $\lambda \equiv p_+$, which, for example, accounts for the exponentially distributed magnitudes of earthquakes (Gutenberg-Richter law [Gutenberg 1936]). The linear dependence of $p_{++}^{(\ell)}$ on p_+ shall be reflected in the diagonal of the underlying copula:

$$C_\ell(p,p) = p^2\,\ell^{-\alpha},$$

a prediction that can be tested empirically.

Note that Omori's law is a measure involving only the two-points probability. In the next subsection, we show what additional information many-points probability can reflect.

7.2.2 Multi-points dependence measures

Although the n-points expectations of Gaussian processes reduce to all combinations of 2-points expectations (7.2), their full dependence structure is not reducible to the bivariate distribution, unless the process is also Markovian (i.e. only in the particular case of exponential correlation). Furthermore, when the process is not Gaussian, even the multi-linear correlations are irreducible. In the general case, the whole multivariate CDF is needed, but many measures of dependence that we introduce below only involve the diagonal n-points copula:[1]

$$\mathcal{C}_n(p) = \mathcal{F}_{t+[\![1,n]\!]}(F^{-1}(p),\dots,F^{-1}(p)), \tag{7.7}$$

which measures the joint probability that all $n \geq 1$ consecutive variables $X_{t+1},\dots,X_{t+n}$ are below the upper p-th quantile of the stationary distribution ($p \in [0,1]$, and $t + [\![1,n]\!]$ is a shorthand for $\{t+1,\dots,t+n\}$). Clearly, $\mathcal{C}_1(p) = p$ and we set by convention $\mathcal{C}_0(p) \equiv 1$.

As an example, the Gaussian diagonal copula is

$$\mathcal{C}_n(p) = \Phi_\rho\big(\Phi^{-1}(p),\dots,\Phi^{-1}(p)\big) \tag{7.8}$$

---

[1] We use a calligraphic $\mathcal{C}$ in order to make it clearly distinct from the bivariate copula discussed in the previous section.

where Φ^{-1} is the univariate inverse CDF, and Φ_ρ denotes the multivariate CDF with $(n \times n)$ covariance matrix ρ, which is Toeplitz with symmetric entries

$$\rho_{tt'} = \rho(|t - t'|), \quad t, t' = 1, \ldots, n.$$

The White Noise (WN) product copula $\mathcal{C}_n(p) = p^n$ is recovered in the limit of vanishing correlations $\rho(\ell) = 0 \, \forall \ell$, and other examples include the exponentially correlated Markovian Gaussian noise, the power-law correlated (thus scale-free) fractional Gaussian noise, and the logarithmically correlated multi-fractal Gaussian noise.

Empirically, the n-points probabilities are very hard to measure due to the large noise associated with such rare joint occurrences. However, there exist observables that embed many-points properties and are more easily measured, such as the length of sequences (clusters) of thresholded events, and the recurrence times of such events, that we study next.

Recurrence intervals

The probability $\pi(\tau)$ of observing a recurrence interval τ between two events is the conditional probability of observing a sequence of $\tau - 1$ "non-events" bordered by two events:

$$\pi(\tau) = \mathbb{P}[X_\tau > X^{(+)}, X_{[\![1;\tau[\![} < X^{(+)}|X_0 > X^{(+)}].$$

(We focus on positive events, but the recurrence of negative events can be studied with the substitution $X \to -X$, and the case of recurrence in amplitudes with the substitution $X \to |X|$). After a simple algebraic transformation flipping all '>' signs to '<', it can be written in the language of copulas as:

$$
\begin{aligned}
\pi(\tau) &= \frac{\mathbb{P}[X_\tau > X^{(+)}, X_{[\![1;\tau[\![} < X^{(+)}; X_0 > X^{(+)}]}{\mathbb{P}[X_0 > X^{(+)}]} \\
&= \frac{\mathbb{P}[X_{[\![1;\tau[\![} < X^{(+)}]}{p_+} - \frac{\mathbb{P}[X_\tau < X^{(+)}, X_{[\![1;\tau[\![} < X^{(+)}]}{p_+} \\
&\quad - \frac{\mathbb{P}[X_{[\![1;\tau[\![} < X^{(+)}; X_0 < X^{(+)}]}{p_+} + \frac{\mathbb{P}[X_\tau < X^{(+)}, X_{[\![1;\tau[\![} < X^{(+)}; X_0 < X^{(+)}]}{p_+} \\
\pi(\tau) &= \frac{\mathcal{C}_{\tau-1}(1-p_+) - 2\,\mathcal{C}_\tau(1-p_+) + \mathcal{C}_{\tau+1}(1-p_+)}{p_+}.
\end{aligned}
\tag{7.9}
$$

The cumulative distribution

$$\Pi(\tau) = \sum_{n=1}^{\tau} \pi(n) = 1 - \frac{\mathcal{C}_\tau(1-p_+) - \mathcal{C}_{\tau+1}(1-p_+)}{p_+} \tag{7.10}$$

is more appropriate for empirical purposes, being less sensitive to noise. These exact expressions make clear — almost straight from the definition — that (i) the distribution of recurrence times *depends only on the copula* of the underlying process and not on the stationary law, in particular its domain or its tails (this is because we take a relative definition of the threshold as a quantile); (ii) *non-linear* dependences are highly relevant in the statistics of recurrences, so that linear correlations can in the general case by no means explain alone the properties of $\pi(\tau)$ [Altmann 2005]; and (iii) recurrence intervals have a *long memory* revealed by the $(\tau+1)$-points copula being involved, so that only when the underlying process

X_t is Markovian will the recurrences themselves be memoryless.[2] Hence, when the copula is known (Eq. (7.8) below for Gaussian processes), the distribution of recurrence times is exactly characterized by the analytical expression in Eq. (7.9).

The average recurrence time is found straightforwardly by summing the series

$$\mu_\pi = \langle \tau \rangle = \sum_{\tau=1}^{\infty} \tau\, \pi(\tau) = \frac{1}{p_+}, \tag{7.11}$$

and is *universal* whatever the dependence structure. This result was first stated and proven by Kac in a similar fashion [Kac 1947]. It is intuitive as, for a given threshold, the whole time series is the succession of a fixed number $p_+ T$ of recurrences whose lengths τ_i necessarily add up to the total size T, so that $\langle \tau \rangle = \sum_i \tau_i/(p_+ T) = 1/p_+$. Note that Eq. (7.11) assumes an infinite range for the possible lags τ, which is achieved either by having an infinitely long time series, or more practically when the translational-invariant copula is periodic at the boundaries of the time series, as is typically the case for artificial data which are simulated using numerical Fourier Transform methods. Introducing the copula allows to emphasize the validity of the statement even in the presence of non-linear long-term dependences, as Eq. (7.11) means that the average recurrence interval is *copula*-independent.

More generally, the m-th moment can be computed as well by summing $\tau^m \pi(\tau)$ over τ:

$$\langle \tau^m \rangle = \frac{1}{p_+} \left[1 + \sum_{\tau=1}^{\infty} \left(|\tau+1|^m - 2\tau^m + |\tau-1|^m \right) \mathcal{C}_\tau(1-p_+) \right]. \tag{7.12}$$

In particular, the variance of the distribution is

$$\sigma_\pi^2 \equiv \langle \tau^2 \rangle - \mu_\pi^2 = \frac{2}{p_+} \sum_{\tau=1}^{\infty} \mathcal{C}_\tau(1-p_+) - \frac{1-p_+}{p_+^2}, \tag{7.13}$$

It is not universal, in contrast with the mean, and can be related to the average unconditional waiting time, see below. Notice that in the independent case the variance $\sigma_\pi^2 = (1 - p_+)/p_+^2$ is not equal to the mean $\mu_\pi = 1/p_+$ (as would be the case for a continuous-time Poisson process) because of discreteness effects.

Waiting times

The conditional mean residual time to next event, when sitting τ time steps after a (positive) event, is

$$\langle w|\tau \rangle = \sum_{w=1}^{\infty} w\, \pi(\tau+w) = \frac{1}{p_+} \mathcal{C}_\tau(1-p_+). \tag{7.14}$$

One is often more concerned with unconditional waiting times, which is equivalent to asking what the size of a sequence of w 'non-events' starting now will be, *regardless of what happened previously*. The distribution of these waiting times is

$$\phi(w) = \mathcal{C}_w(1-p_+) - \mathcal{C}_{w+1}(1-p_+),$$

---

[2]It may be mentioned that in a non-stationary context, renewal processes are also able to produce independent consecutive recurrences [Santhanam 2008, Sazuka 2009].

and its expected value is

$$\mu_\phi = \langle w \rangle = \sum_{w=1}^{\infty} \mathcal{C}_w (1 - p_+),$$

consistently to what would be obtained by averaging $\langle w | \tau \rangle$ over all possible elapsed times in Eq. (7.14). From Eq. (7.13), we have the following relationship between the variance of the distribution π of recurrence intervals, and the mean waiting time:

$$\sigma_\pi^2 = \mu_\pi \left[2\mu_\phi + 1 \right] - \mu_\pi^2 \tag{7.15}$$

Sequences lengths

The dependence in the process is also revealed by the distribution of sequences sizes. The probability that a sequence of consecutive negative events[3], starting just after a 'non-event', will have a size n is

$$\psi(n) = \frac{\mathcal{C}_n(p_-) - 2\mathcal{C}_{n+1}(p_-) + \mathcal{C}_{n+2}(p_-)}{p_- \left(1 - p_- \right)}$$

and the average length of a random sequence

$$\mu_\psi = \langle n \rangle = \sum_{n=1}^{\infty} n\, \psi(n) = \frac{1}{1 - p_-} \tag{7.16}$$

is universal, just like the mean recurrence time. This property rules out the analysis of [Boguná 2004] who claim to be able to distinguish the dependence in processes according to the average sequence size.

Conditional recurrence intervals, clustering

The dynamics of recurrence times is as important as their statistical properties, and in fact impacts the empirical determination of the latter[4]. It is now clear, both from empirical evidences and analytically from the discussion on Eq. (7.9), that recurrence intervals have a long memory. In dynamic terms, this means that their occurrences show some clustering. The natural question is then: "Conditionally on an observed recurrence time, what is the probability distribution of the next one?" This probability of observing an interval τ' immediately following an observed recurrence time τ is

$$\mathbb{P}[X_{\tau+\tau'} > X^{(+)}, X_{\tau + [\![1;\tau'[\![} < X^{(+)} | X_\tau > X^{(+)}, X_{[\![1;\tau[\![} < X^{(+)}, X_0 > X^{(+)}]. \tag{7.17}$$

Again, flipping the '>' to '<' allows to decompose it as

$$\frac{\mathcal{C}_{\tau-1;\tau'-1} - \mathcal{C}_{\tau;\tau'-1} - \mathcal{C}_{\tau-1;\tau'} + \mathcal{C}_{\tau;\tau'}}{\mathcal{C}_{\tau-1} - 2\mathcal{C}_\tau + \mathcal{C}_{\tau+1}} - \frac{\pi(\tau + \tau')}{\pi(\tau)},$$

---

[3] We consider the case of "negative" events, i.e. those with $X_t < X^{(-)}$ because it expresses simply in terms of diagonal copulas. The mirror case with "positive" events has the exact same expression but $\mathcal{C}_n$ must be inverted around the median. For a symmetric F, this distinction is irrelevant.

[4] Distribution testing for $\pi(\tau)$ involving Goodness-of-fit tests [Ren 2010] should be discarded because those are not designed for dependent samples and rejection of the null cannot be relied upon. See Section 3.2 of Chapter II.3 for an extension of GoF tests when some dependence is present.

$\pi_+(\tau)$	$\langle\tau\rangle_+$	$\phi_+(w)$	$\langle w\rangle_+$	$\psi_-(n)$	$\langle n\rangle_-$	$R(t)$
$(1-q)\,q^{\tau-1}$	$1/(1-q)$	$(1-q)\,q^w$	$q/(1-q)$	$q\,(1-q)^{n-1}$	$1/q$	$1/t$

Table 7.1: Different probabilities introduced, with thresholds defined as $F(X^{(+)}) = q = 1 - F(X^{(-)})$, for the White Noise process.

where the $(\tau+\tau')$-points probability

$$\mathcal{C}_{\tau;\tau'}(p) = \mathcal{F}_{[\![0;\tau+\tau']\!]\setminus\{\tau\}}(F^{-1}(p),\dots,F^{-1}(p))$$

shows up. Of course, this exact expression has no practical use, again because there is no hope of empirically measuring any many-points probabilities of extreme events with a meaningful signal-to-noise ratio. We rather want to stress that non-linear correlations and multi-points dependences are relevant, and that a characterization of clustering based on the autocorrelation of recurrence intervals is an oversimplified view of reality.

Record statistic

We conclude this theoretical section on multi-points non-linear dependences by mentioning that the diagonal n-points copula $\mathcal{C}_n$ can alternatively be understood as the distribution of the maximum of n realizations of X in a row, since

$$\mathbb{P}\left[\max_{\tau\le n}\{X_{t+\tau}\} < F^{-1}(p)\right] = \mathbb{P}\left[X_{t+[\![1,\tau]\!]} < F^{-1}(p)\right] = \mathcal{C}_n(p).$$

Studying the statistics of such "local" maximas in short sequences (see e.g. [Eichner 2006]) can thus provide information on the multi-points properties of the underlying process. The CDF of the running maximum, or *record*, is $\mathcal{C}_t(F(x))$ and the probability that $t > 1$ will be a record-breaking time is the joint probability

$$R(t) = \mathbb{P}\left[\max_{\tau<t}\{X_\tau\} < X_t\right],$$

which is *irrespective of the marginal law* !

7.3 Continuous time

The n-points copula of a continuous-time process X_t is the counting associated with the corresponding arrival process, whose events are defined like Bernoulli variables with a latent condition $\mathbb{1}_{\{F(X_t)\leq u_t\}}$. Indeed,

$$\mathbb{E}[\mathbb{1}_{\{F(X_{t_1})\leq u_{t_1}\}}, \ldots, \mathbb{1}_{\{F(X_{t_n})\leq u_{t_n}\}}]$$

is a straightforward generalization of the discrete n-points copula

$$\mathcal{F}_{[\![1,n]\!]}(F^{-1}(u_{t_1}), \ldots, F^{-1}(u_{t_n})),$$

when the time-stamps t_i are not equidistant.

More generally, one can thus study either directly the arrivals times t_i of events to be defined, or more fundamentally the latent physical process X_t underlying the firing of the events. The latter description can be modeled by diffusions or any kind of continuous processes (see Sect. 7.3.2 below), while the former is achieved using point processes and is studied next.

7.3.1 Arrival times: the duration process

In the context of arrivals of a point process, a recurrence interval R_t is larger than τ if the number of events $N_{t+\tau}$ has not increased since last event time t:

$$\mathbb{P}[R_t > \tau] = \mathbb{P}[N_{t+\tau} - N_t = 0].$$

When the arrivals are independent and occur with a deterministic intensity $\lambda(t)$ per unit of time, then the process is Poisson

$$\mathbb{P}[N_{t+\tau} - N_t = 0|\lambda] = e^{-\int_t^{t+\tau} \lambda(s)\mathrm{d}s}, \tag{7.18}$$

and the associated durations have a distribution $\pi(\tau|t) \equiv -\mathrm{d}\mathbb{P}[R_t > \tau]/\mathrm{d}\tau$ equal to

$$\pi(\tau|t) = \lambda(t+\tau)\, e^{-\int_t^{t+\tau} \lambda(s)\mathrm{d}s}.$$

When the underlying process is stationary, the intensity λ is constant and in fact equal to the unconditional probability of events. The usual exponential distribution of recurrence times for independent arrivals is retrieved, and one verifies in particular that the average duration is equal to the inverse intensity:

$$\langle \tau \rangle = 1/\lambda.$$

When $\lambda(\cdot)$ is a stochastic process, Eq. (7.18) must be understood conditionally, and the correct probabilities are found averaging over all possible $\Lambda_{t,t+\tau} = \int_t^{t+\tau} \lambda(s)\mathrm{d}s$. Such doubly stochastic Poisson processes were initially introduced by [Cox 1955]. Of course, Λ being an "arrival rate" it must always remain positive, but for the sake of the example, let us examine the case of Gaussian processes, and in which limits this holds. As is know from exponentials of Gaussian variables, the correction to the probability (7.18) is

$$\mathbb{P}[R_t > \tau] = e^{-\mathbb{E}[\Lambda_{t,t+\tau}] + \frac{1}{2}\mathbb{V}[\Lambda_{t,t+\tau}]}.$$

For example, with the integrated intensity following an Ornstein-Uhlenbeck process

$$\Lambda_{t,t+\tau} = \overline{\lambda}\tau + \sigma \int_0^{t+\tau} e^{-\theta(\tau+t-s)} dz_s, \tag{7.19}$$

the drift is $\mathbb{E}[\Lambda_{t,t+\tau}] = \overline{\lambda}\tau$ and the variance $\mathbb{V}[\Lambda_{t,t+\tau}] = \frac{\sigma^2}{2\theta}\left(1 - e^{-\theta(2t+\tau)}\right)$. Since Λ must always remain positive, the Gaussian case can only be a valid approximation as long as $\mathbb{E}[\Lambda] \gg \mathbb{V}[\lambda]$, i.e. $\tau \gg \sigma^2/(2\theta\overline{\lambda})$ in the stationary OU example.

The intensity process (7.19) is Gaussian and Markovian, but can be generalized to incorporate long-memory by replacing the exponential integration kernel by one with a slow decay, and/or choosing positive-only innovations in place of the Wiener dz_t [Basu 2002]. The translational invariance is broken at short times $t \to \infty$, if the intensity depends upon the past realizations of the counting process N_t, but at long times, stationarity is able to settle down. Such so-called Hawkes processes are obtained replacing the independent innovation dz_t by the counting process itself dN_t, and are thus self-exciting. They have been recently the focus of intense research in earthquakes studies — see also the related Epidemic-Type Aftershock Sequence (ETAS) model of triggered seismicity [Saichev 2007, Sornette 2008] — and in finance [Muni Toke 2012, Bacry 2012, Filimonov 2012]. The latter studies point toward a strong self-exciting mechanism that brings the system close to criticality, what was also the conclusion of feedback QARCH model of Chapter 6.

7.3.2 Continuous processes

The recurrence times problem for a continuous process can be formulated as a first-passage problem [Bray 2013]: indeed, the probability density

$$\pi(\tau|t)d\tau = d\mathbb{P}[X_\tau = X^{(+)}, X_{t'} < X^{(+)} \forall t' < \tau | X_t \geq X^{(+)}]$$

can be simply written in terms of the transition probability $f(X^{(+)}, \tau; x, t)$ of a particle starting at $X_t = x$ at the initial time t and ending at $X_\tau = X^{(+)}$ without ever hitting the wall at $X^{(+)}$ inbetween. Of course, since there are no jumps in this continuous setting, the initial value must be conditioned to be $X_t = X^{(+)}$ for all the subsequent values to be lower than $X^{(+)}$, and thus

$$\pi(\tau|t) = f(X^{(+)}, \tau; X^{(+)}, t).$$

Once the dynamics of the process is specified (by a Stochastic Differential Equation, a Langevin equation or a non-Markovian generalization thereof), the transition density f can be found by path-integral methods, or solving Fokker-Planck-like equations with appropriate boundary conditions: $f(X^{(+)}, t'|X^{(+)}, t) = 0 \forall t' < \tau$.

Importantly, specific methods of non-Markovian Fokker-Planck equations need to be used in case the process has memory: f.ex. convolution of usual Fokker-Planck with a memory kernel [Sokolov 2002], fractional Fokker-Planck (see [Metzler 1999] and many later works of the same authors), continuous limits of Master equations, etc.

In the case of the Brownian Motion, the moments of the distribution $f(x, \tau - t; x_0, 0)$ of first passages satisfy a recursive relation [Ebeling 2005].

Stock Index	Country	From	To	T
S&P-500	USA	Jan. 02, 1970	Dec. 23, 2011	10 615
KOSPI-200	South Korea	Jan. 03, 1990	Dec. 26, 2011	5 843
CAC-40	France	Jul. 09, 1987	Dec. 23, 2011	6 182
DAX-30	Germany	Jan. 02, 1970	Dec. 23, 2011	10 564
SMI-20	Switzerland	Jan. 07, 1988	Dec. 23, 2011	5 902

Table 7.2: Description of the dataset used: time series of daily returns of international stock indices.

7.4 Financial self-copulas

We illustrate some of the quantities introduced above on series of daily index returns. The properties of the time series used are summarized in Tab. 7.2.

7.4.1 Conditional probabilities and 2-points dependences

We reproduce the study of Ref. [Boguná 2004] on the statistic of price changes conditionally on previous return sign, and extend the analysis to any threshold $|X^{(\pm)}| \geq 0$ and to remote lags. We first illustrate on Fig. 7.4 the conditional probabilities $p_{\pm\pm}^{(\ell)}$ (filled symbols) and $p_{\pm\mp}^{(\ell)}$ (empty symbols) with varying threshold $q = F(X^{(+)}) = 1 - F(X^{(-)})$, for $\ell = 1$. To study the departure from the independent case, it is more convenient to subtract the White Noise contribution, to get the corresponding *excess* probabilities. Several features can be immediately observed: positive events (upward triangles) trigger more subsequent positive (filled) than negative (empty) events; negative events (downward triangles) trigger more subsequent negative (filled) than positive (empty) events, except in the far tails $q \gtrsim 0.9$ where reversion is stronger than persistence after a negative event. Both these effects dominate the WN benchmark, but the latter effect is however much stronger. This overall behavior is similar for all time series.

The shapes of $p_{\pm\pm}$ and $p_{\pm\mp}$ versus q are not compatible with the Student copula benchmarks (correlation $\rho = 0.05$ and d.o.f. $\nu = 5$) shown in dashed and dotted lines, respectively. Notice that, due to its non-trivial tail-correlations, see part I, the Student copula does generate increased persistence with respect to WN, lower reversion in the core and higher reversion in the tails. But empirically the reversion is asymmetric and typically stronger when conditioning on negative events rather than on positive events, a property reminiscent of the leverage effect which cannot be accounted for by a pure Student copulas. Some of the indices exhibit more pronounced reversion and persistence effects. Interestingly, the CAC-40 returns have a regime $0.5 \leq q \lesssim 0.9$ close to a white noise (with in particular a value of $p_{\pm\pm}^{(1)} = p_{\pm\mp}^{(1)}$ very close to the 0 at $q = 0.5$, revealing an inefficient conditioning, i.e. as many positive and negative returns immediately following positive or negative returns), but the extreme positive events $q \gtrsim 0.9$ show a very strong persistence, and the extreme negative events a very strong reversion.

In the next chapter, we study in detail the p and ℓ dependence of $[C_\ell(p,p) - p^2]$ and $[C_\ell(p,1-p) - p(1-p)]$ — which are straightforwardly related to $p_{\pm\pm}^{(\ell)}$ and $p_{\pm\mp}^{(\ell)}$, respectively — and find that the self-copula of stock returns can be modeled with a high accuracy by a

(a) SP500

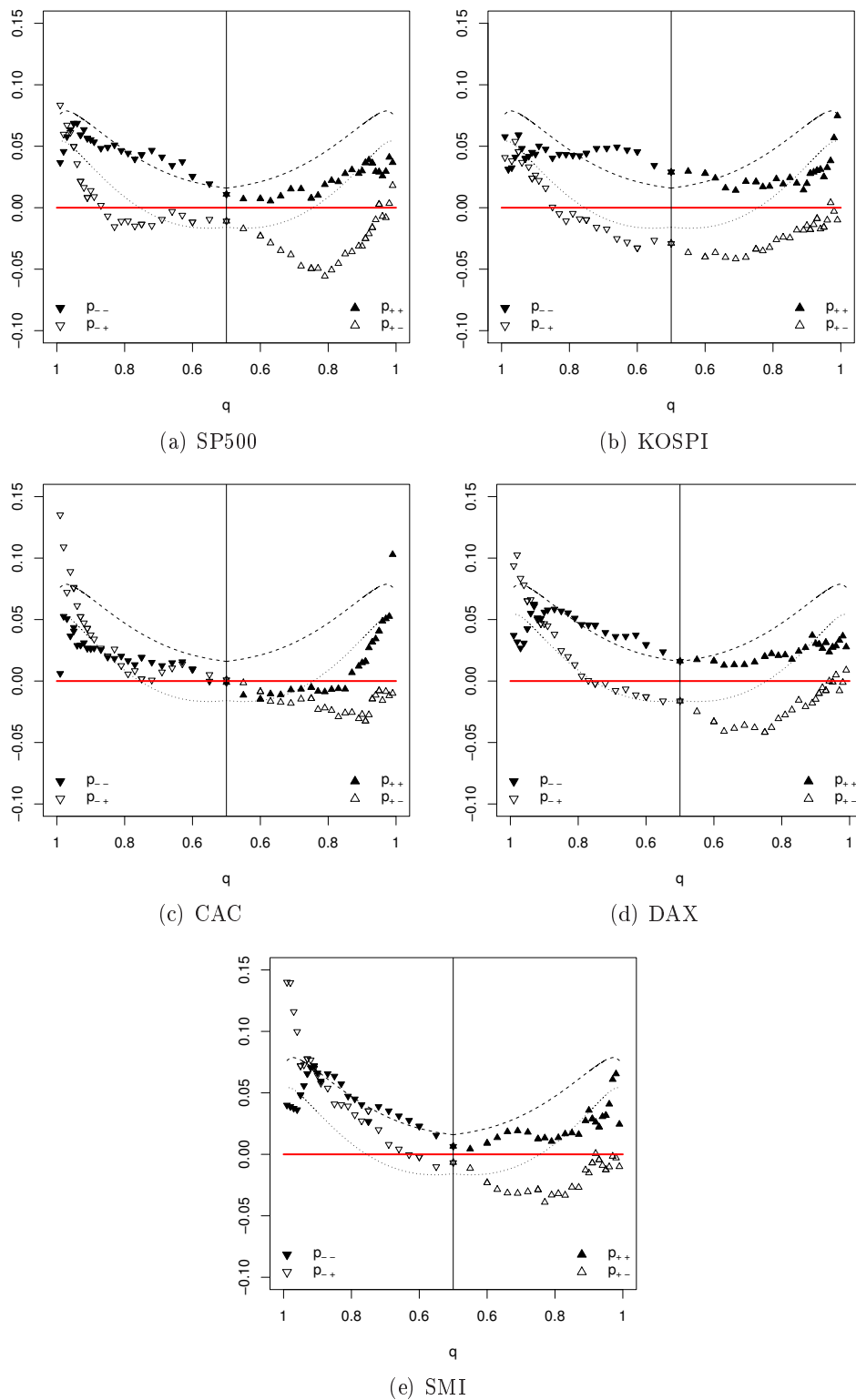(b) KOSPI

(c) CAC

(d) DAX

(e) SMI

Figure 7.4: Conditional extreme probabilities at $\ell = 1$ (the WN contribution has been subtracted). Filled symbols are for persistence, and empty symbols for reversion. Upward pointing triangles are conditioned on positive jumps, and downward pointing triangles are conditioned on negative jumps.
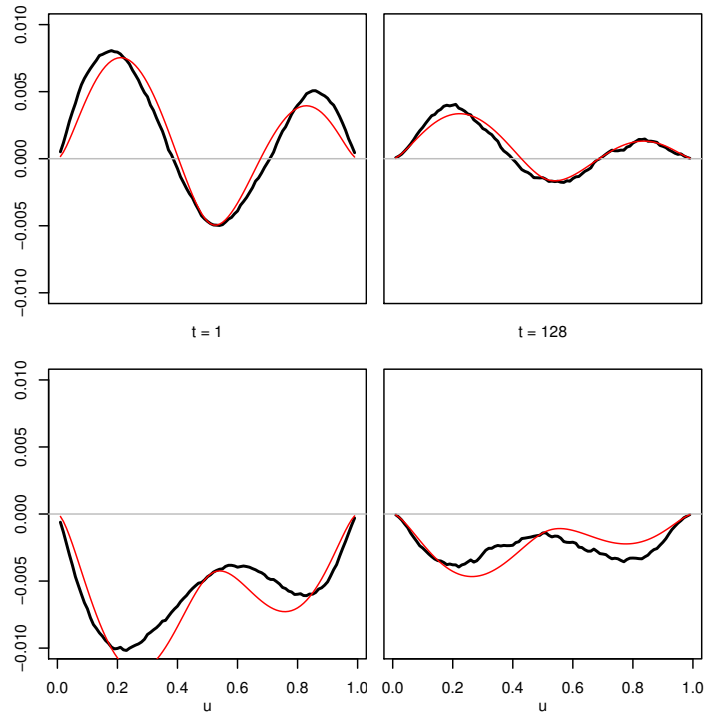
Figure 7.5: Diagonal (top) and anti-diagonal (bottom) of the self-copula for lags $\ell = 1$ and $\ell = 128$; the product copula has been subtracted. The copula determined empirically on stock returns is in bold black, and a fit with the model of Chapter 8 is shown in thin red.

log-normal volatility with log-decaying correlation, in agreement with multifractal volatility models. To anticipate over the discussion there, we give a preview of the results in Fig. 7.5, for the aggregated copula of all stocks in the S&P500 in 2000–2004. We are able to show precisely how every kind of dependence present in the underlying process materializes itself in $p_{++}^{(\ell)}$ for different q's: short ranged linear anti-correlation accounts for the central part $(p \approx 0.5)$ departing from the WN prediction, long-ranged amplitude clustering is responsible for the "M" and "W" shapes that reveal excess persistence and suppressed reversion, and the leverage effect can be observed in the asymmetric heights of the "M" and "W".

Fig. 7.6 displays the behavior of $\langle X \rangle_{\pm}$ versus q (we also show the median $\mathrm{med}(X)_{\pm}$) at lags corresponding to one day ($\ell = 1$), one week ($\ell = 5$) and one month ($\ell = 20$). The conditional amplitudes $\langle X \rangle_{\pm}$ measure "how large" a realization will be on average after an event at a given threshold, whereas the conditional probabilities $p_{\pm\pm}$ and $p_{\pm\mp}$ quantify "how often" repeated such events occur. Mind the *unconditional* mean and median values, both above zero and distinct from each other. At $\ell = 1$, the reversion of *extreme* events is revealed again by the change of monotonicity from $q \approx 0.8$ on, and more strongly for $q > 0.9$ where $\langle X \rangle_{-}$ has an opposite sign than the preceding return; this corroborates the observation made on conditional probabilities above. Beyond the next day, the general picture is that dependences tend to vanish and the empirical measurements get more concentrated around the WN prediction. However, tail effects are strongly present, with unexpectedly a typical behavior opposite to that of $\ell = 1$: weekly monthly reversion of extreme positive jumps. See the caption for a detailed discussion of the specificities of each stock index at every lag ℓ.

(a) SP500

(b) KOSPI
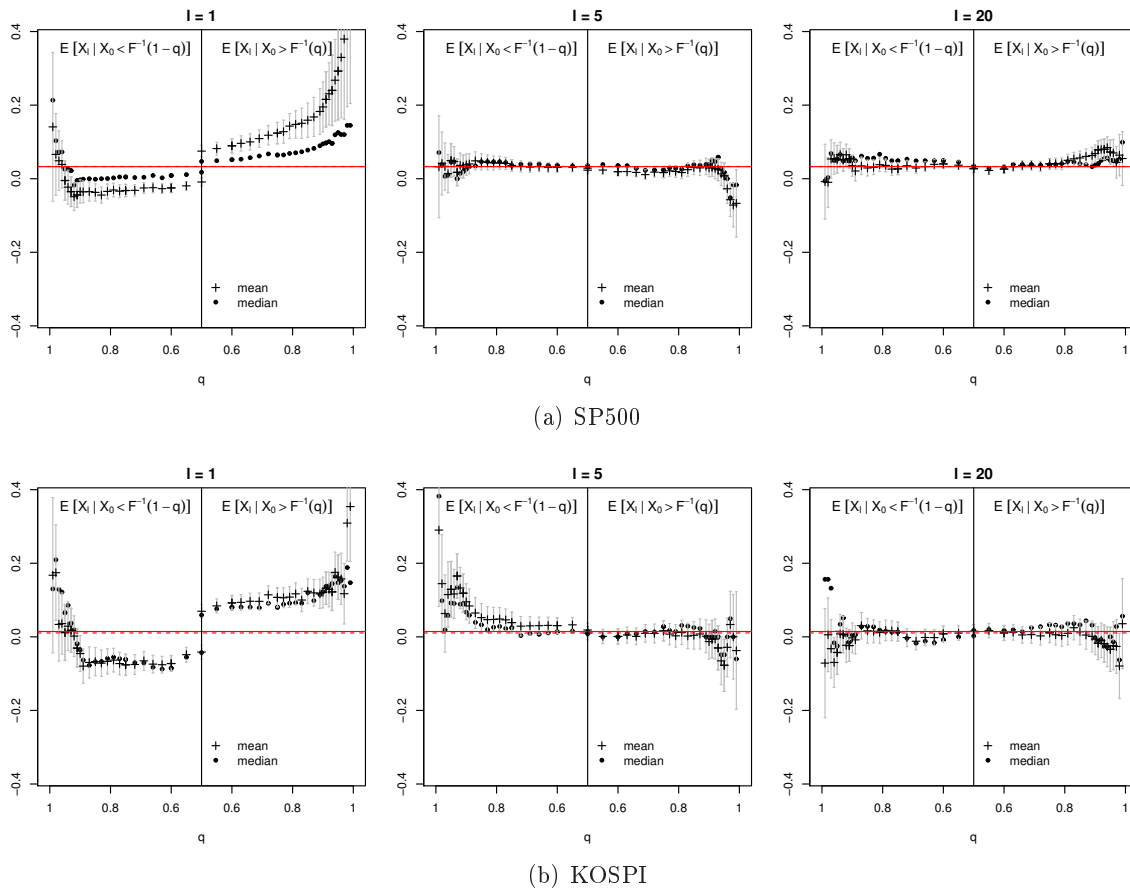
Figure 7.6: Conditional extreme amplitudes, at lags $\ell = 1, 5, 20$. The upper-right and lower-left quadrants express persistence, while the upper-left and lower-right quadrants reveal reversion. For a scale-free dependence structure, one would expect the magnitudes to decrease with the lag ℓ but the global shape to be conserved. What we instead observe is important changes of configuration at different lags: For example, the strong reversion of negative tail events at $\ell = 1$ vanishes at farther lags, and even turns into strong persistence for the CAC and DAX indices. That is to say that these indices tend to mean-revert after a negative event at the daily frequency, but to trend on the weekly scale. Similarly, the strong persistence of positive events at $\ell = 1$ converts to a strong reversion in the tails at $\ell = 20$ for the European indices (CAC, DAX, SMI); a weaker reversion is observed at intermediate scale ($\ell = 5$) for most indices (including US and Korean).
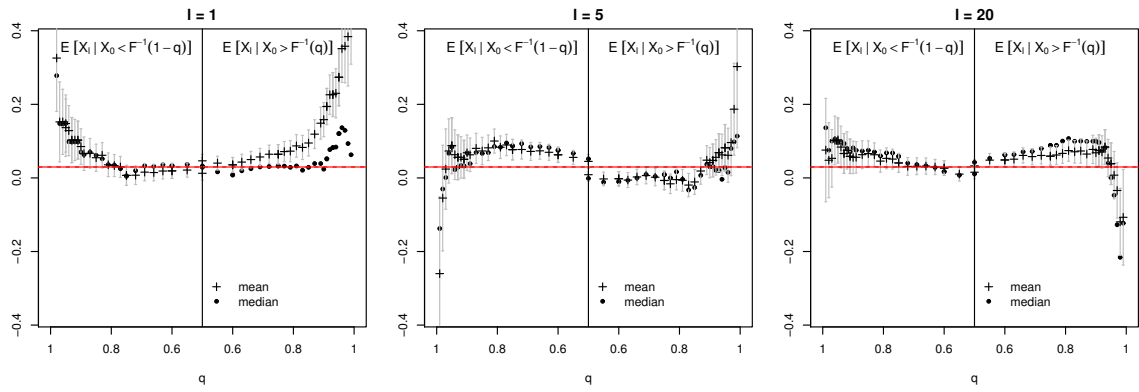
(c) CAC
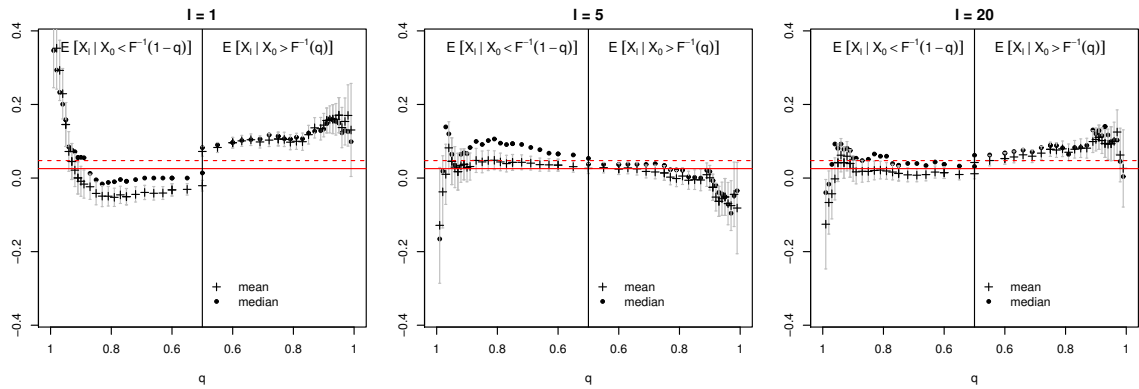
(d) DAX
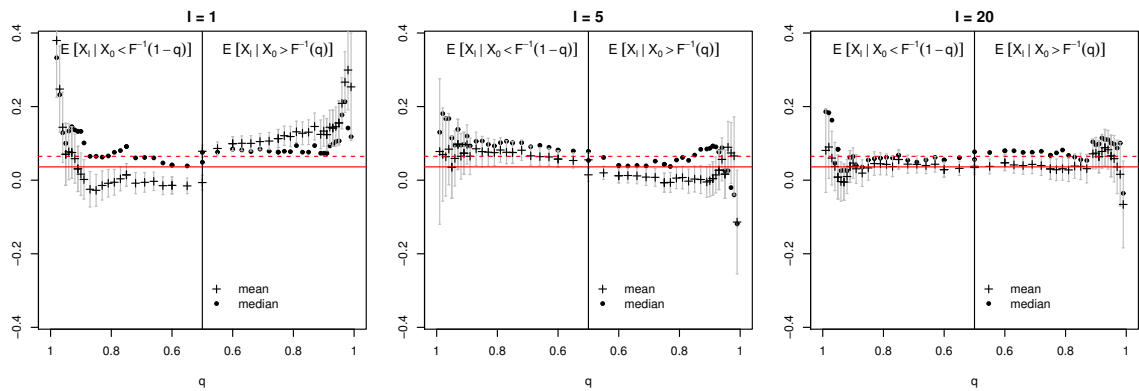
(e) SMI

Figure 7.6: (continued)
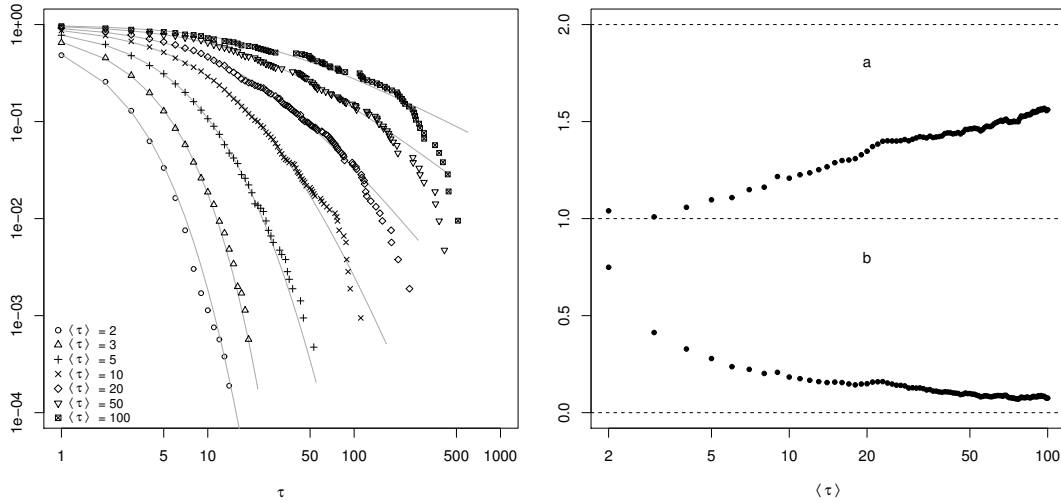
Figure 7.7: DAX index returns. **Left:** tail probability $1 - \Pi(\tau)$ of the recurrence intervals, at several thresholds $p_+ = 1/\langle\tau\rangle$, in log-log scale. Grey curves are best fits to $[1 + b\,(a - 1)\,\tau]^{(2-a)/(a-1)}$ suggested in Ref. [Ludescher 2011]. **Right:** estimated parameters a and b of the best fit.

7.4.2 Recurrence intervals and many-points dependences

Even the simple, two-points measures of self-dependence studied up to now show that non-linearities and multi-scaling are two ingredients that must be taken into account when attempting to describe financial time series; we now examine their many-points properties. As an example, we compute the distribution of recurrence times of returns above a threshold $X^{(+)} = F^{-1}(1 - p_+)$.

Fig. 7.7 shows the tail cumulative distribution $1 - \Pi(\tau)$ of the recurrence intervals of DAX returns, at several thresholds $p_+ = 1/\langle\tau\rangle$ — the distribution for other indices is very similar. In the log-log representation used, an exponential distribution (corresponding to independent returns) would be concave and rapidly decreasing, while a power-law would decay linearly. The empirical distributions fit neither of those, and the authors of Ref. [Ludescher 2011] suggested a parametric fit of the form

$$1 - \Pi(\tau) = [1 + b\,(a-1)\,\tau]^{(2-a)/(a-1)}.$$

However, important deviations are present in the tail regions for thresholds at $X^{(+)} \gtrsim F^{-1}(0.9)$, i.e. $\langle\tau\rangle \gtrsim 1/(1 - 0.9) = 10$: as a consequence, there is no hope that the curves for different threshold collapse onto a single curve after a proper rescaling [Chicheportiche 2013d], as is the case e.g. for seismic data. A more fundamental determination of the form of $\Pi(\tau)$ should rely on Eq. (7.10) and a characterization of the τ-points copula.

Similarly to the statistic of the recurrence intervals, their dynamics must be studied carefully. We have shown that the conditional distribution of recurrence intervals after a previous recurrence is very complex and involves long-ranged non-linear dependences, so that a simple assessment of recurrence times auto-correlation may not be informative enough for a deep understanding of the mechanisms at stake.

7.5   Conclusion

We report several properties of recurrence times and the statistic of other observables (waiting times, cluster sizes, records, aftershocks) in light of their decription in terms of the diagonal copula, and hope that these studies can shed light on the n-points properties of the process by assessing the statistics of simple variables rather than positing an *a priori* dependence structure.

The exact universality of the mean recurrence interval imposes a natural scale in the system. A scaling relation in the distribution of such recurrences is only possible in absence of any other characteristic time. When such additional characteristic times are present (typically in the non-linear correlations), no such scaling is expected, in contrast with time series of earthquake magnitudes.

We also stress that recurrences are intrinsically multi-points objects related to the non-linear dependences in the underlying time-series. As such, their autocorrelation is not a reliable measure of their dynamics, for their conditional occurence probability is much history dependent.

In a continuous-time setting, the role of the n-points copula is played by a counting process. The events to be counted can either be triggered by an underlying continuous process crossing a threshold, or more directly be modeled as a self-exciting point process, like a Hawkes process.

Ultimately, recurrences may be used to characterize risk in a new fashion. Instead of — or in addition to — caring for the amplitude and probability of adverse events at a given horizon, one should be able to characterize the risk in a dynamical point of view. In this sense, an asset A_1 could be said to be "more risky" than another asset A_2 if its distribution of recurrence of adverse events has such and such "bad" properties that A_1 does not share. This amounts to characterizing the disutility by "When ?" shocks are expected to happen, in addition to the usual "How often ?" and "How large ?".

The generalization of the n-points copula for continuous-time processes has been rapidly introduced, and it would be interesting to study many-points dependences in continuous processes, for example transaction times in a Limit Order Book.

The next chapter uses the copula approach to address the long-range nature of volatility-dependences and incorporate multi-scaling and non-linearities. The theory of Goodness-of-fit tests for dependent observations developed in part I is put in practice to characterize the unconditional distribution of stock returns.

The long-ranged nature of volatility dependences

Contents

When trying to extend GoF tests to dependent variables, self-copulas appear naturally, as we carefully showed in Section 3.2 of Chapter I.3. It is thus very important to estimate and/or model precisely the underlying dependence structure in order to apply the theory presented there. This chapter proposes two applications of the GoF tests with dependent observations.

In Section 8.1 we go through a detailed example when the dependences are described by a pseudo-elliptical copula exhibiting log-amplitudes correlations. We first investigate the Markovian case of weak, short-range dependences, and then relax the second assumption to investigate long-memory; we also add linear correlations and leverage effect in order to be as close as possible to financial copulas. Indeed, Section 8.2 is dedicated to an application of the theory to the case of financial data: after defining our data set, we perform an empirical study of dependences in series of stock returns, and interpret the results in terms of the "self-copula"; implication of the dependences on GoF tests are illustrated for this special case using Monte-Carlo simulations. In our empirical study of financial self-copulas, we rely on a non-parametric estimation rather than imposing, for example, a Markovian structure of the underlying process, as in e.g. [Darsow 1992, Ibragimov 2008].

8.1   An explicit example: the log-normal volatility

In order to illustrate the above general formalism, we focus on the specific example of the product random variable $X = \sigma \xi$, with i.i.d. Gaussian residuals ξ and log-normal stochastic amplitude $\sigma = e^\omega$ (we denote generically by F the cdf of X). Such models are common in finance to describe stock returns, as will be discussed in the next section. For the time being, let us consider the case where the ω's are of zero mean, and covariance given by:

$$\mathrm{Cov}(\omega_n \omega_{n+t}) = \Sigma^2 f\left(\frac{t}{T}\right), \qquad (t > 0). \tag{8.1}$$

where T is a typical cutoff scale.

The pairwise copulas can be explicitly written in the limit of weak correlations, $\Sigma^2 \to 0$. One finds:

$$C_t(u,v) - uv = \Sigma^2 f\left(\frac{t}{T}\right) \tilde{A}(u)\tilde{A}(v) + \mathcal{O}(\Sigma^4) \tag{8.2}$$

$$\text{with} \quad \tilde{A}(u) = \int\limits_{-\infty}^{\infty} \varphi(\omega)\varphi'(\frac{F^{-1}(u)}{e^\omega})\mathrm{d}\omega \tag{8.3}$$

where here and in the following $\varphi(\cdot)$ denotes the univariate Gaussian pdf, and $\Phi(\cdot)$ the Gaussian cdf. The spectrum of $A(u,v) = \tilde{A}(u)\tilde{A}(v)$ consists in a single non-degenerate eigenvalue $\lambda^A = \mathrm{Tr}\, A = \int_0^1 \tilde{A}(u)^2 \,\mathrm{d}u$, and an infinitely degenerate null eigenspace. Assuming $f_\infty = \sum_{r=1}^{\infty} f(r) < +\infty$ (the memory might be long-ranged, yet must decay sufficiently fast), the covariance kernel reads:

$$H(u,v) = I(u,v) + 2\,T\,\Sigma^2 f_\infty\, A(u,v).$$

Depending on the value of the parameters, the first term or the second term may be dominant. Note that one can be in the case of weak correlations ($\Sigma^2 \to 0$) but long range memory $T \gg 1$,

such that the product $T\Sigma^2$ can be large (this is the case of financial time series, see below). If $T\Sigma^2$ is small, one can use perturbation theory around the Brownian bridge limit (note that $\mathrm{Tr}\,I \approx 10\,\mathrm{Tr}\,A$, see Tab. D.1 in appendix), whereas if $T\Sigma^2$ is large, it is rather the Brownian term $I(u,v)$ that can be treated as a perturbation. Elements of the algebra necessary to set up these perturbation theories are given in Appendix D.

We now turn to a numerical illustration of our theory, in this simple case where only volatility correlations are present.

8.1.1 Short-range memory

In order to remain in the framework of the theory exposed in Chapter I.3 and satisfy in particular the conditions for the validity of the CLT under weak dependences, we assume a Markovian dynamics for the log-volatilities:

$$X_n = \xi_n\,\mathrm{e}^{\omega_n - \mathrm{V}[\omega]}, \quad \text{with} \quad \omega_{n+1} = g\omega_n + \Sigma\eta_n, \tag{8.4}$$

where $g < 1$ and η_n are i.i.d. Gaussian variables of zero mean and unit variance. In this case,

$$\alpha_t = \mathrm{Cov}(\omega_n\omega_{n+t}) = \frac{\Sigma^2}{1-g^2}\,g^t. \tag{8.5}$$

In the limit where $\Sigma^2 \ll 1$, the weak dependence expansion holds and one finds explicitly:

$$H(u,v) = I(u,v) + 2\frac{g\Sigma^2}{(1-g)^2(1+g)}\,A(u,v). \tag{8.6}$$

In order to find the limit distribution of the test statistics, we proceed by Monte-Carlo simulations. The range $[0,1]^2$ of the copula is discretized on a regular lattice of size $(M \times M)$. The limit process is described as a vector with M components and built from Equation (3.22), page 51, as $\tilde{\mathbf{y}} = U\Lambda^{\frac{1}{2}}\mathbf{z}$ where the diagonal elements of Λ are the eigenvalues of H (in decreasing order), and the columns of U are the corresponding eigenvectors. Clearly, $\mathrm{Cov}(\tilde{\mathbf{y}}, \tilde{\mathbf{y}}) = U\Lambda U^\dagger = H$.

For each Monte-Carlo trial, M independent random values are drawn from a standard Gaussian distribution and collected in $\mathbf{z}$. Then $\mathbf{y}$ is computed using the above representation. This allows one to determine the two relevant statistics:

$$KS = \max_{u=1...M} |\tilde{y}_u|$$

$$CM = \frac{1}{M}\sum_{u=1}^{M}\tilde{y}_u^2 = \frac{1}{M}\tilde{\mathbf{y}}^\dagger\tilde{\mathbf{y}} = \frac{1}{M}\mathbf{z}^\dagger\Lambda\mathbf{z}.$$

The empirical cumulative distribution functions of the statistics for a large number of trials are shown in Figure 8.1 together with the usual Kolmogorov-Smirnov and Cramér-von-Mises limit distributions corresponding to the case of independent variables.

In order to check the accuracy of the obtained limit distribution, we generate 350 series of $N = 2500$ dates according to Equation (8.4). For each such series, we perform two GoF tests, namely KS and CM, and calculate the corresponding p-values. By construction, the p-values of a test should be uniformly distributed if the underlying distribution of the
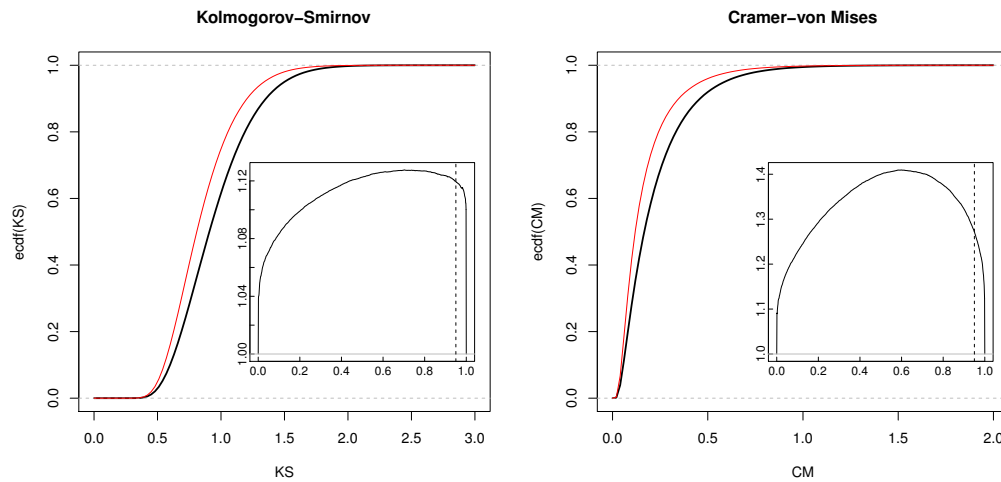
Figure 8.1: Markovian model. **Left:** Cumulative distribution function of the supremum of $\tilde{y}(u)$. **Right:** Cumulative distribution of the norm-2 of $\tilde{y}(u)$. The cases of independent draws (thin red) and dependent draws (bold black) are compared. The dependent observations are drawn according to the weak-dependence kernel (8.6) with parameters $g = 0.88, \Sigma^2 = 0.05$. **Insets:** The effective reduction ratio $\sqrt{\frac{N}{N_{\text{eff}}(u)}} = \frac{\text{ecdf}^{-1}(u)}{\text{cdf}_L^{-1}(u)}$ where L = KS,CM. The dashed vertical line is located at the 95-th centile and thus indicates the reduction ratio corresponding to the p-value $p = 0.05$ (as the test is unilateral).
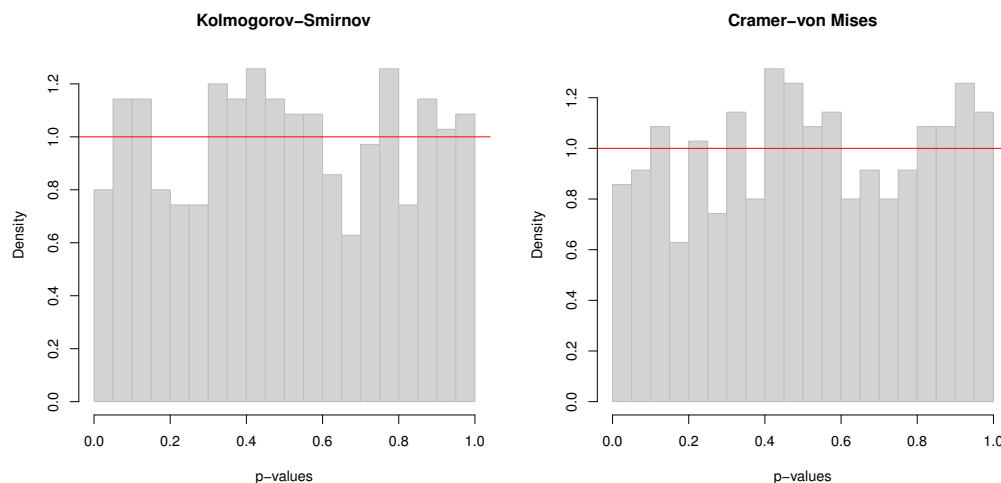
Figure 8.2: Histogram of the p-values in the GoF test on simulated data, according to Equation (8.4). Uniform distribution of the p-values of a test indicates that the correct law of the statistic is used.

simulated data is indeed the same as the hypothesized distribution. In our case, when using the usual KS and CM distributions for independent data, the p-values are much too small and their histogram is statistically not compatible with the uniform distribution. Instead, when using the appropriate limit distribution found by Monte-Carlo and corresponding to the correlation kernel (8.6), the calculated p-values are uniformly distributed, as can be visually seen on Figure 8.2, and as revealed statistically by a KS test (on the KS test!), comparing the 350 p-values to $H_0 : p \sim \mathcal{U}[0,1]$.

8.1.2 Long-range memory

If, instead of the AR(1) (Markovian) prescription (8.4), the dynamics of the ω_n is given by a Fractional Gaussian Noise (i.e. the increments of a fractional Brownian motion) [Mandelbrot 1968] with Hurst index $\frac{2-\nu}{2} > \frac{1}{2}$, the log-volatility has a long ranged auto-covariance

$$\alpha_t = \text{Cov}(\omega_n, \omega_{n+t}) = \frac{\Sigma^2}{2} \left(|t+1|^{2-\nu} - 2t^{2-\nu} + |t-1|^{2-\nu} \right), \quad t \geq 0 \qquad (8.7)$$

that decays as a power law $\propto (2 - 3\nu + \nu^2)t^{-\nu}$ as $t \to \infty$, corresponding to long-memory, therefore violating the hypothesis under which the above theory is correct. Still, we want to illustrate that the above methodology leads to a meaningful improvement of the test, even in the case of long-ranged dependences. The corresponding covariance kernel of the Xs,

$$H(u,v) = I(u,v) + 2\Sigma^2 A(u,v) \sum_{t=1}^{N} \left(1 - \frac{t}{N} \right) \alpha_t, \qquad (8.8)$$

is used in a Monte-Carlo simulation like in the previous case in order to find the appropriate distribution of the test statistics KS and CM (shown in Figure 8.3, see caption for the choice of parameters). We again apply the GoF tests to simulated series, and compute the p-values according to the theory above. As stated above, our theory is designed for short-ranged dependences whereas the fGn process is long-ranged. The p-values are therefore not expected to be uniformly distributed. Nevertheless, the distribution of the p-values is significantly corrected toward the uniform distribution, see Figure 8.3(a): with the naive CM distribution (middle), the obtained p-values are localized around zero, suggesting that the test is strongly negative. If instead we use our prediction for short-range dependences (right), we find a clear improvement, as the p-values are more widely spread on $[0,1]$ (but still not uniformly).

8.1.3 Accounting for linear correlation and time reversal asymmetry

It is interesting to generalize the above model to account for weak dependence between the residuals ξ and between the residual and the volatility, without spoiling the log-normal structure of the model. We therefore write:

$$X_0 = \xi_0 e^{\omega_0}; \quad X_t = \xi_t e^{\alpha_t \omega_0 + \beta_t \xi_0 + \sqrt{1 - \alpha_t^2 - \beta_t^2} \omega_t} \quad \text{with} \quad \mathbb{E}[\xi_0 \xi_t] = r_t$$

Cramer–von Mises
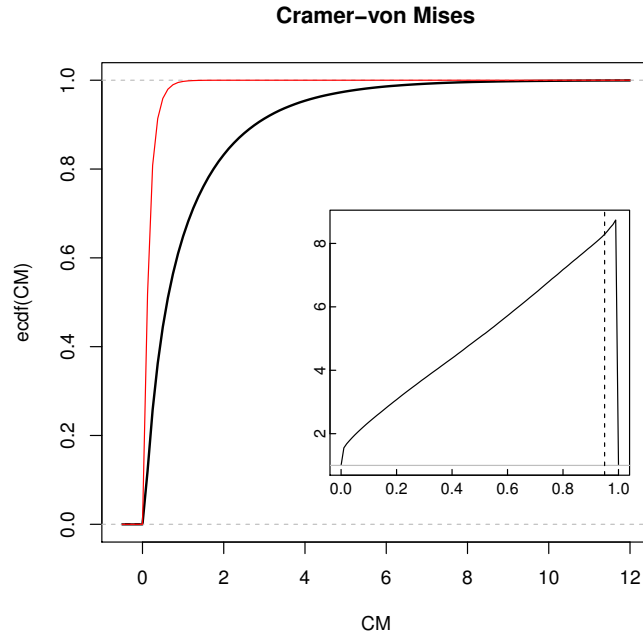
(a) Cumulative distribution function of the norm-2 of $\tilde{y}(u)$, see Fig. 8.1 for full caption.
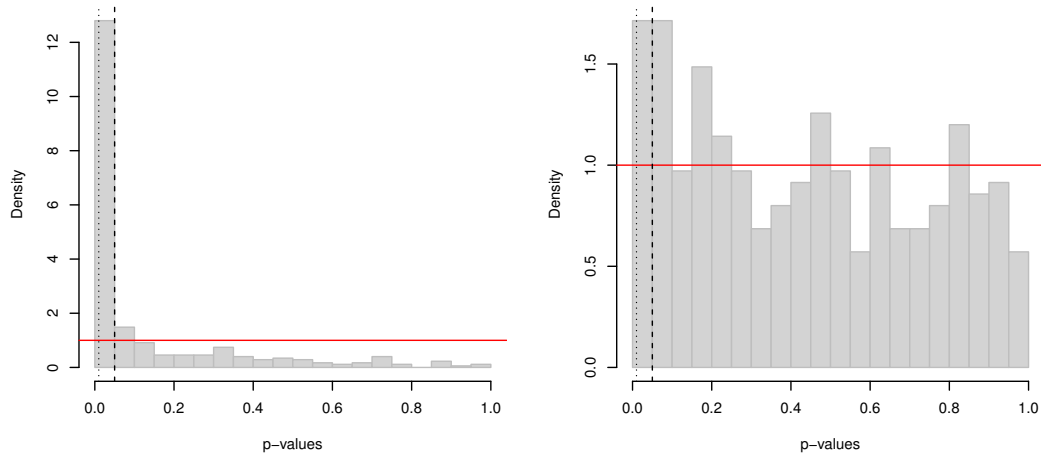
(b) Histogram of the p-values in the CM test on simulated data, using the i.i.d. CM distribution (left) and the prediction obtained assuming short range correlations (right). The dependent observations are drawn according to Eq. (8.7) with parameters $\nu = \frac{2}{5}$, $\Sigma^2 = 1$, $N = 1500$.

Figure 8.3: Fractional Brownian Motion.
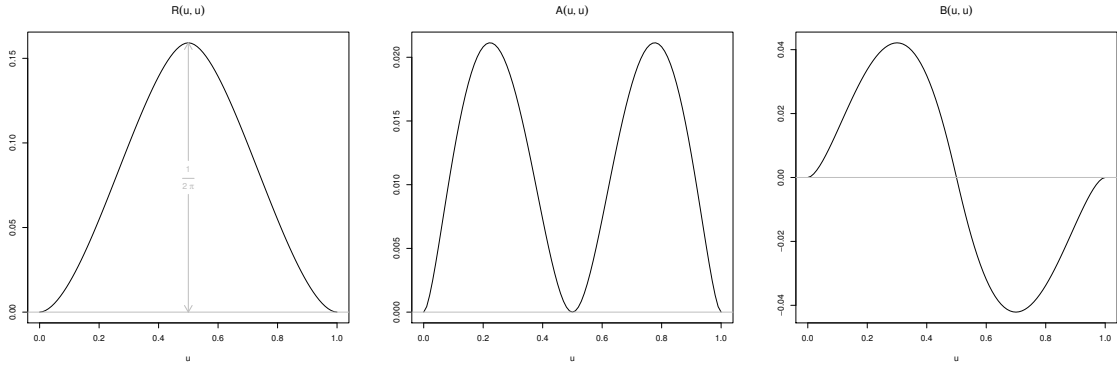
Figure 8.4: Copula diagonal of the log-normal volatility model: linear corrections to independence. **Left:** correction $R(u,u)$ due to correlation of the residuals (vertical axis in multiples of ρ) **Middle:** correction $A(u,u)$ due to correlation of the log-vols (vertical axis in multiples of α) **Right:** correction $B(u,u)$ due to leverage effect (vertical axis in multiples of $-\beta$)

where all the variables are $\mathcal{N}(0,1)$, so that in particular

$$\rho_t = \mathrm{Corr}(X_0, X_t) = r_t(1+\beta_t^2)\mathrm{e}^{\alpha_t - 1}$$
$$\mathrm{Corr}(X_0^2, X_t^2) = \frac{\left(1 + 2r_t^2(1 + 10\beta_t^2 + 8\beta_t^4) + 4\beta_t^2\right)\mathrm{e}^{4\alpha_t} - 1}{3\mathrm{e}^4 - 1}$$
$$\mathrm{Corr}(X_0, X_t^2) = 2\beta_t \frac{\left(1 + 2r_t^2(1 + 2\beta_t^2)\right)\mathrm{e}^{2\alpha_t - \frac{1}{2}}}{\sqrt{\mathrm{e}^4 - 1}}$$

The univariate marginal distributions of X_0 and X_t are identical and their cdf is given by the integral

$$F(x) = \int_{-\infty}^{\infty} \varphi(\omega)\Phi(\frac{x}{\mathrm{e}^\omega})\mathrm{d}\omega. \tag{8.9}$$

Expanding the bivariate cdf (or the copula) in the small dependence parameters $\alpha_t, \beta_t, \rho_t$ around $(0,0,0)$, we get

$$C_t(u,v) - uv \approx \alpha_t A(u,v) - \beta_t B(u,v) + \rho_t R(u,v) \tag{8.10}$$
$$\approx \alpha_t \tilde{A}(u)\tilde{A}(v) - \beta_t \tilde{R}(u)\tilde{A}(v) + \rho_t \tilde{R}(u)\tilde{R}(v)$$

where $\tilde{A}(u)$ was defined above in Equation (8.3), and

$$\tilde{R}(u) = \int_{-\infty}^{\infty} \varphi(\omega)\varphi(\frac{F^{-1}(u)}{\mathrm{e}^\omega})\mathrm{d}\omega = \tilde{R}(1-u). \tag{8.11}$$

The contributions of $A(u,v), B(u,v)$ and $R(u,v)$ on the diagonal are illustrated in Figure 8.4. Notice that the term $B(u,v)$ (coming from cross-correlations between ξ_0 and ω_t, i.e. the so-called leverage effect, see below) breaks the symmetry $C_t(u,v) \neq C_t(v,u)$.

8.2   Application to financial time series

8.2.1   Stylized facts of daily stock returns

One of the contexts where long-ranged persistence is present is time series of financial asset returns. At the same time, the empirical determination of the distribution of these returns is of utmost importance, in particular for risk control and derivative pricing. As we will see, the volatility correlations are so long-ranged that the number of effectively independent observations is strongly reduced, in such a way that the GoF tests are not very tight, even with time series containing thousands of raw observations.

It is well-known that stock returns exhibit dependences of different kinds:

- at relatively high frequencies (up to a few days), returns show weak, but significant negative linear auto-correlations (see e.g. [Avramov 2006]);

- the absolute returns show persistence over very long periods, an effect called multi-scale volatility clustering and for which several interesting models have been proposed in the last ten years [Pasquini 1999, Calvet 2008, Muzy 2000, Zumbach 2001, Lynch 2003, Borland 2005];

- past negative returns favor increased future volatility , an effect that goes under the name of "leverage correlations" in the literature [Bouchaud 2001b, Perelló 2004, Pochart 2002, Eisler 2004, Ahlgren 2007, Reigneron 2011].

Our aim here is neither to investigate the origin of these effects and their possible explanations in terms of behavioral economics, nor to propose a new family of models to describe them. We rather want to propose a new way to characterize and measure the full structure of the temporal dependences of returns based on copulas, and extract from this knowledge the quantities needed to apply GoF tests to financial times series.

Throughout this section, the empirical results are based on a data set consisting of the daily returns of the stock price of listed large cap US companies. More precisely we keep only the 376 names present in the S&P-500 index constantly over the five years period 2000–2004, corresponding to $N = 1256$ days. The individual series are standardized, but this does not change the determination of copulas, that are invariant under increasing and continuous transformations of the marginals.

8.2.2   Empirical self-copulas

For each (u, v) on a lattice, we determine the lag dependent "self-copula" $C_t(u, v)$ by assuming stationarity, i.e. that the pairwise copula $C_{nm}(u, v)$ only depends on the time lag $t = m - n$. We also assume that all stocks are characterized by the same self-copula, and therefore an average over all the stock names in the universe is done in order to remove noise. Both these assumptions are questionable and we give at the end of this section an insight on how non-stationarities as well as systematic effects of market cap, liquidity, tick size, etc, can be accounted for.

The self-copulas are estimated non-parametrically with a bias correction[1], then fitted to the parametric family of log-normal copulas introduced in the previous section. We assume

---

[1]Details on the copula estimator and the bias issue are given in appendix.
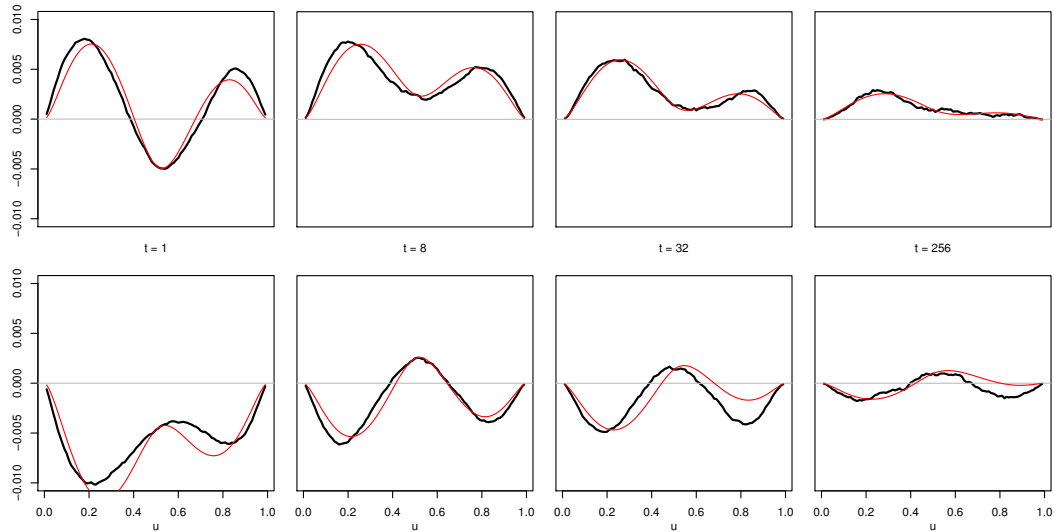
Figure 8.5: Diagonal (top) and anti-diagonal (bottom) of the self-copula for different lags; the product copula has been subtracted. A fit with Equation (8.10) is shown in thin red. Note that the y scale is small, confirming that the weak dependence expansion is justified. The dependence is still significant even for $t \sim 500$ days!

(and check a posteriori) that the weak dependence expansion holds, leaving us with three functions of time, α_t, β_t and ρ_t, to be determined. We fit for each t the copula diagonal $C_t(u, u)$ to Equation (8.10) above, determine α_t, β_t and ρ_t, and test for consistency on the anti-diagonal $C_t(u, 1 - u)$. Alternatively, we could determine these coefficients to best fit $C_t(u, v)$ in the whole (u, v) plane, but the final results are not very different. The results are shown in Figure 8.5 for lags $t = 1, 8, 32, 256$ days. Fits of similar quality are obtained up to $t = 512$.

Before discussing the time dependence of the fitted coefficients α_t, β_t and ρ_t, let us describe how the different effects show up in the plots of the diagonal and anti-diagonal copulas. The contribution of the linear auto-correlation can be directly observed at the central point $C_t(\frac{1}{2}, \frac{1}{2})$ of the copula. It is indeed known [Chicheportiche 2012b] that for any pseudo-elliptical model (including the present log-normal framework) one has:

$$C_t(\tfrac{1}{2}, \tfrac{1}{2}) = \frac{1}{4} + \frac{1}{2\pi} \arcsin \rho_t.$$

Note that this relation holds beyond the weak dependence regime. If $\beta_t^{(\mathrm{B})} = C_t(\frac{1}{2}, \frac{1}{2}) - \frac{1}{4}$ — this is in fact Blomqvist's beta coefficient [Blomqvist 1950] — the auto-correlation is measured by $\rho_t = \sin(2\pi \beta_t^{(\mathrm{B})})$.

The volatility clustering effect can be visualized in terms of the diagonals of the self-copula; indeed, the excess (unconditional) probability of large events following previous large events of the same sign is $\left[C_t(u, u) - u^2 \right]$ with $u < \frac{1}{2}$ for negative returns, and $u > \frac{1}{2}$ for positive ones. On the anti-diagonal, the excess (unconditional) probability of large positive events following large negative ones is, for small $u < \frac{1}{2}$, the upper-left volume $\left[C_t(u, 1) - u \cdot 1 \right] - \left[C_t(u, 1-u) - u(1-u) \right] = u(1-u) - C_t(u, 1-u)$ and similarly the excess probability of large negative events following large positive ones is the same expression for large $u > \frac{1}{2}$ (lower-
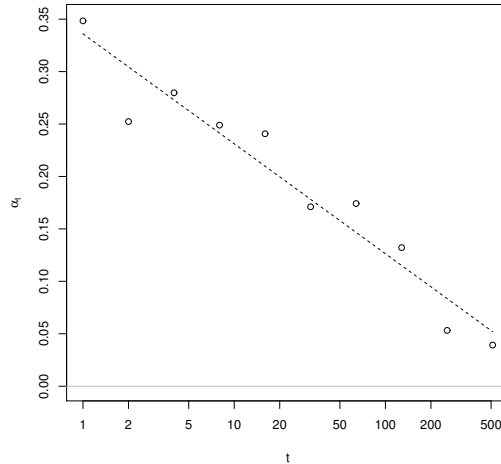
Figure 8.6: Auto-correlation of the volatilities, for lags ranging from 1 to 768 days. Each point represents the value of α_t extracted from a fit of the empirical copula diagonal at a given lag to the relation (8.10). We also show the fit to a multifractal model, $\alpha_t = -\Sigma^2 \ln \frac{t}{T}$, with $\Sigma^2 = 0.046$ and $T = 1467$ days.

right volume). As illustrated on Figure 8.5, these four quadrants exceed the independent case prediction, suggesting a genuine clustering of the amplitudes, measured by α_t. Finally, an asymmetry is clearly present: the effect of large negative events on future amplitudes is stronger than the effect of previous positive events. This is an evidence for the leverage effect: negative returns cause a large volatility, which in turn makes future events (positive or negative) to be likely larger. This effect is captured by the coefficient β_t.

The evolution of the coefficients α_t, β_t and ρ_t for different lags reveals the following properties: i) the linear auto-correlation ρ_t is short-ranged (a few days), and negative; ii) the leverage parameter β_t is short-ranged and, as is well known, negative, revealing the asymmetry discussed above; iii) the correlation of volatility is *long-ranged* and of relatively large positive amplitude (see Figure 8.6), in line with the known long range volatility clustering. More quantitatively, we find that the parameter α_t for lags ranging from 1 to 768 days is consistent with an effective relation well fitted by the "multifractal" [Muzy 2001, Calvet 2008, Lux 2008, Muzy 2000] prediction for the volatility autocorrelations: $\alpha_t = -\Sigma^2 \ln \frac{t}{T}$, with an amplitude $\Sigma^2 = 0.046$ and a horizon $T = 1467$ days consistent, in order of magnitude, with previous determinations.

The remarkable point, already noticed in previous empirical works on multifractals [Muzy 2001, Duchon 2008], is that the horizon T, beyond which the volatility correlations vanish, is found to be extremely long. In fact, the extrapolated horizon T is larger than the number of points of our sample N! This long correlation time has two consequences: first, the parameter $2T\Sigma^2 f_\infty$ that appears in the kernel $H(u,v)$ is large, ≈ 135. This means that the dependence part $T\Sigma^2 f_\infty A(u,v)$ is dominant over the independent Brownian bridge part $I(u,v)$. This is illustrated in Figure 8.7, where we show the first eigenvector of $H(u,v)$, which we compare to the non-zero eigenmode of $A(u,v)$, and to the first eigenvector of $I(u,v)$. Sec-
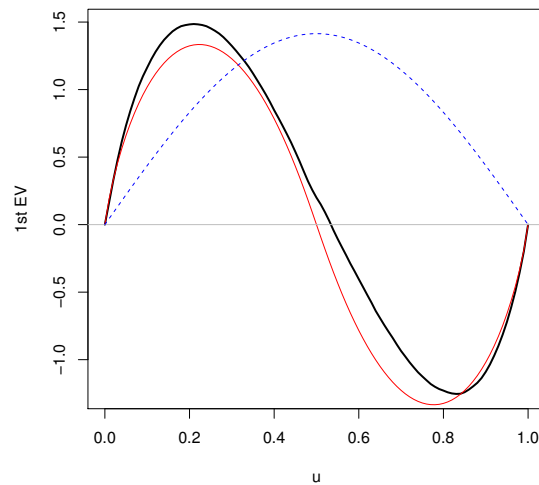
Figure 8.7: **Bold black:** The first eigenvector of the empirical kernel $H(u,v) = I(u,v)\,[1 + \Psi_N(u,v)]$. **Plain red:** The function $\tilde{A}(u)$ (normalized), corresponding to the pure effect of volatility clustering in a log-normal model, in the limit where the Brownian bridge contribution $I(u,v)$ becomes negligible. **Dashed blue:** The largest eigenmode $|1\rangle = \sqrt{2}\sin(\pi u)$ of the independent kernel $I(u,v)$.

ond, the hypothesis of a stationary process, which requires that $N \ll T$, is not met here, so we expect important pre-asymptotic corrections to the above theoretical results.

8.2.3 Monte-Carlo estimation of the limit distributions

Since $H(u,v)$ is copula-dependent, and considering the poor analytical progress made about the limit distributions of KS and CM in cases other than independence, the asymptotic laws will be computed numerically by Monte-Carlo simulations (like in the example of Section 8.1) with the empirically determined $H(u,v)$.

The empirical cumulative distribution functions of the statistics for a large number of trials are shown in Figure 8.8 together with the usual Kolmogorov-Smirnov and Cramér-von-Mises limit distributions corresponding to the case of independent variables. One sees that the statistics adapted to account for dependences are stretched to the right, meaning that they accept higher values of KS or CM (i.e. measures of the difference between the true and the empirical distributions). In other words, the outcome of a test based on the usual KS or CM distributions is much more likely to be negative, as it will consider "high" values (like 2–3) as extremely improbable, whereas a test that accounts for the strong dependence in the time series would still accept the null-hypothesis for such values.

As an illustration, we apply the test of Cramér-von Mises to our dataset, comparing the empirical univariate distributions of stock returns to a simple model of log-normal stochastic
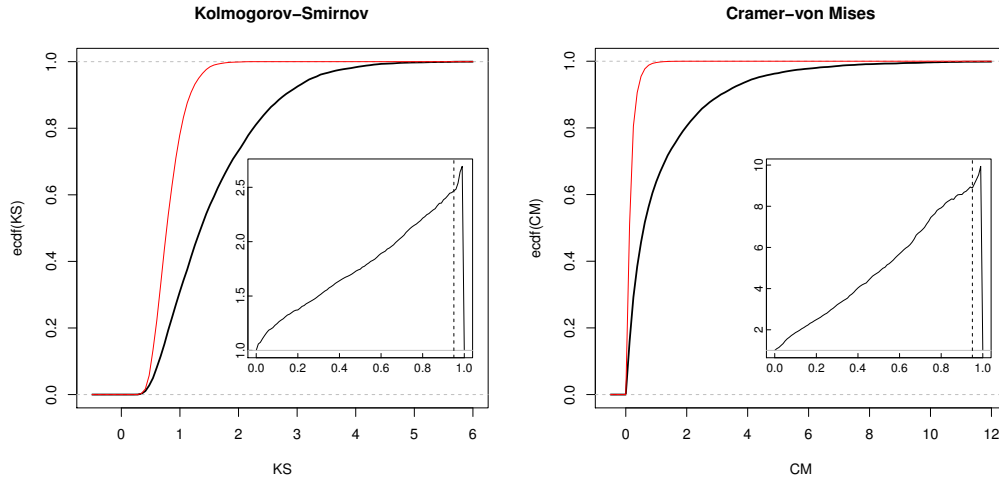
Figure 8.8: **Left:** Cumulative distribution function of the supremum of $\tilde{y}(u)$. **Right:** Cumulative distribution of the norm-2 of $\tilde{y}(u)$. The cases of independent draws (thin red) and dependent draws (bold black) are compared. The dependent observations are drawn according to the empirical average self-copula of US stock returns in 2000-2004. **Insets:** The effective reduction ratio $\sqrt{\frac{N}{N_{\text{eff}}(u)}} = \frac{\text{ecdf}^{-1}(u)}{\text{cdf}_{\text{L}}^{-1}(u)}$ where L = KS,CM.

volatility, i.e. that the null-hypothesis cdf is similar to Eq. (8.9):

$$F(x) = \int_{-\infty}^{\infty} \varphi(\omega)\Phi\left(\frac{x}{\mathrm{e}^{s\omega-s^2}}\right)\mathrm{d}\omega. \tag{8.12}$$

The volatility of volatility parameter s can be calibrated from the time series $\{x_t\}_t$ as

$$s^2 = \ln\left(\frac{2}{\pi}\frac{\langle x_t^2\rangle_t}{\langle x_t\rangle_t^2}\right). \tag{8.13}$$

We want to test the hypothesis that the log-normal model with a unique value of *s for all stocks* is compatible with the data. In practice, for each stock i, s_i is chosen as the average of (8.13) over all *other* stocks in order to avoid endogeneity issues and biases in the calculations of the p-values of the test. s_i is found to be ≈ 0.5 and indeed almost identical for all stocks. Then the GoF statistic *CM* is computed for each stock i and the corresponding p-value is calculated.[2]

Figure 8.9 shows the distribution of the p-values, as obtained by using the usual asymptotic Cramér-von Mises distribution for independent samples (left) and the modified version allowing for dependence (right). We clearly observe that the standard Cramér-von Mises

---

[2]Another source of endogeneity is the estimated copula (either parametric or parameter-free) used in the asymptotic test law, and is similar to the bias induced when applying usual univariate GoF test with an estimated null. This issue can be dealt with by either estimating the copula on a subset and apply the test on another subset, or by using a parametric form for the copula, where the parameters are not estimated but scanned over given ranges. A p-value is obtained for each combination of parameters, and optimal parameters can be determined among the accepted tests. This of course can be very burdensome when the parameter space is large, and in addition it relies on a model for the copula.
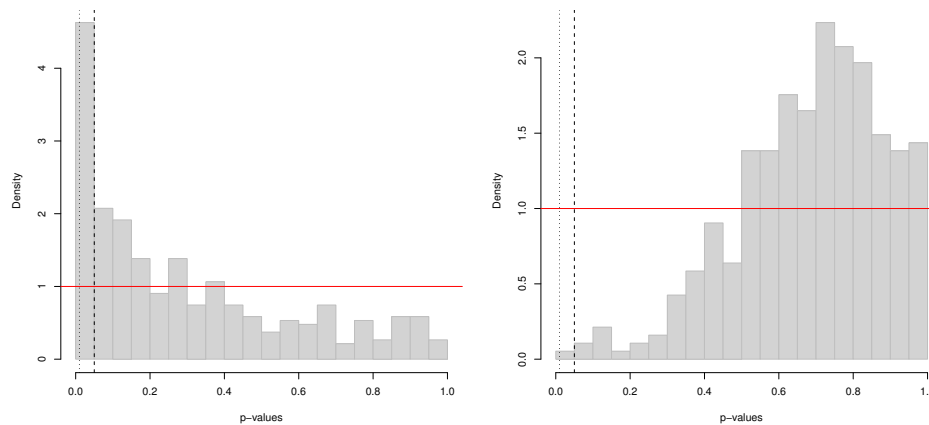
Figure 8.9: Histogram of the p-values in a Cramér-von Mises-like test, see the text for the test design. **Left:** when using the standard Cramér-von Mises law, the obtained p-values are far from uniformly distributed and strongly localized under the threshold $p = 0.05$ (dashed vertical line) occasioning numerous spurious rejections. **Right:** when using the modified law taking dependences into account, the test rejects the hypothesis of an identical distribution for all stocks much less often.

test strongly rejects the hypothesis of a common log-normal model, as the corresponding p-values are concentrated around zero, which leads to an excessively high rejection rate. The use of the generalized Cramér-von Mises test for dependent variables greatly improves the situation, with in fact now too many high values of p. Therefore, *the hypothesis of a common log-normal model for all stocks cannot be rejected* when the long-memory of volatility is taken into account. The overabundant large values of p may be due to the fact that all stocks are in fact exposed to a common volatility factor (the "market mode"), which makes the estimation of s somewhat endogenous and generates an important bias. Another reason is that the hypothesis that the size of the sample N is much larger than the correlation time T does not hold for our sample, and corrections to our theoretical results are expected in that case.[3] It would be actually quite interesting to extend the above formalism to the long-memory case, where $T \gg N \gg 1$.

8.2.4   Beyond stationarity and universality

The assumption that financial time series are stationary is far from granted. In fact, several observations suggest that financial markets operate at the frontier between stationarity and non-stationarity. The multifractal model, for example, assumes that correlations are decaying marginally slowly (as a logarithm), which technically corresponds to the boundary between stationary models (when decay is faster) and non-stationary models. Furthermore, as stated above, the horizon T that appears in the model is empirically found to be very long (on the order of several years) and therefore very difficult to measure precisely. Other models, like the "multi-scale" GARCH [Zumbach 2001, Borland 2011, Chicheportiche 2012a], point

---

[3]Note that in practice, we have estimated $\Psi_N(u,v)$ by summing the empirically determined copulas up to $t_{\max} = 512$, which clearly underestimates the contribution of large lags.

in the same direction: when these models are calibrated on financial data, the parameters are always found to be very close to the limit of stability of the model (see [Borland 2011] for a discussion of this point). Several very recent studies on high-frequency data point in the same direction, and even consider non-stationary extensions of the "critically stationary" multifractal or self-exciting models [Morales 2012, Muzy 2013, Hardiman 2013].

Furthermore, what is relevant in the present context is strong stationarity, i.e. the stationarity of the full self-copula. Hence, testing for stationarity amounts to comparing copulas, which amounts to a GoF for two-dimensional variables . . . for which general statistical tools are missing, even in the absence of strong dependence! So in a sense this problem is like a snake chasing its own tail. A simple (and weak) argument is to perform the above study in different periods. When fitting the multifractal model to the self-copulas, we find the following values for $(\Sigma^2, \ln T)$: $(0.032, 6.39)$ in 1995-1999, $(0.046, 7.29)$ in 2000-2004, $(0.066, 8.12)$ in 2000-2009 and $(0.078, 7.86)$ in 2005-2009. These numbers vary quite a bit, but this is exactly what is expected with the multifractal model itself when the size of the time series N is not much larger than the horizon T! (see [Muzy 2006, Bacry 2006] for a detailed discussion of the finite N properties of the model).

As of the universality of the self-copula across the stocks, it is indeed a reasonable assumption that allows one to average over the stock ensemble. We compared the averaged self-copula on two subsets of stocks obtained by randomly dividing the ensemble, and found the resulting copulas to be hardly distinguishable. This can be understood if we consider that all stocks are, to a first approximation, driven by a common force — the "market" — and subject to idiosyncratic fluctuations. We know that this picture is oversimplified, and that systematic effects of sector, market cap, etc. are expected and could in fact be treated separately inside the framework that we propose. Such issues are related to the cross-sectional non-linear dependence structure of stocks, and are far beyond the objectives of this article.

8.3   Conclusion

From the empirical estimation of the self-copula of US stock returns, long-ranged volatility clustering with multifractal properties is observed as the dominant contribution to self-dependence, in line with previous studies. However, sub-dominant modes are present as well and a precise understanding of those involves an in-depth study of the spectral properties of the correlation kernel H.

One of the remarkable consequences of the long-memory nature of the volatility is that the number of effectively independent observations is significantly reduced, as both the Kolmogorov-Smirnov and Cramér-von Mises tests accept much larger values of the deviations (see Figure 8.8). As a consequence, it is much more difficult to reject the adequation between any reasonable statistical model and empirical data. We suspect that many GoF tests used in the literature to test models of financial returns are fundamentally flawed because of the long-ranged volatility correlations. In intuitive terms, the long-memory nature of the volatility can be thought of as a sequence of volatility regime shifts, each with a different lifetime, and with a broad distribution of these lifetimes. It is clear that in order to fully sample the unconditional distribution of returns, all the regimes must be encountered several times. In the presence of volatility persistence, therefore, the GoF tests are much less

stringent, because there is always a possibility that one of these regimes was not, or only partially, sampled. The multifractal model is an explicit case where this scenario occurs.

In terms of financial developments, we believe that an empirical exploration of the self-copulas for series of diverse asset returns and at diverse frequencies is of primordial importance in order to grasp the complexity of the non-linear time-dependences. In particular, expanding the concept of the self-copula to pairs of assets is likely to reveal subtle dependence patterns. From a practitioner's point of view, a multivariate generalization of the self-copula could lead to important progresses on such issues as causality, lead-lag effects and the accuracy of multivariate prediction.

Conclusion

General conclusion and outlooks

Partial conclusions were given at the end of every chapter, summarizing the latter's content, discussing its conclusions in light of recent developments in the literature, and suggesting directions for possible extensions.

Here I want to give a general overview of the topics studied during the last three years — without reviewing in detail the original contributions in the thesis[1] — and situate them in the current state of the research.

Stochastic processes and applications

The link between the theory of Goodness-of-fit tests and stochastic processes is a beautiful one, yet a poorly known one possibly due to the distance between the two concerned communities of statisticians and physicists. Understanding the law of the Kolmogorov-Smirnov statistics as the survival probability of a randomly walking particle in an expanding absorbing cage allows to use techniques from statistical physics to solve problems in probability theory, which otherwise necessitate lengthy proofs. More importantly, the analogy turns particularly useful in variants of the initial problem: weighting the Kolmogorov-Smirnov statistics in order to increase the test resolution in specific ranges of the domain in fact amounts to switching on a confining potential inside the cage. In higher dimensions (i.e. when designing tests for multivariate distributions), the link to random walkers is made nonintuitive by the trajectories being parameterized by more than one time index. The lack of an obvious meaning to Markovianity or even to 'causality' in such settings, is responsible for seemingly simple associated problems to remain unsolved: for example, the propagator of the 2D Brownian sheet (pinned or not) constrained between floor and ceiling is still unknown. However, the discretization of one time direction suggests new perspectives to addressing the 2D GoF tests via the survival of a random walker in an infinite-dimensional absorbing cage, see Appendix B.

Other properties of stochastic processes like first-passage times can be encountered in financial applications. Typically, problems related to optimal selling times of an asset, exercising of an American option, barriers and stop-loss criteria, etc. explicitly involve the statistics of first-passages and/or recurrence intervals in threshold-crossing processes. Less obvious is the link between optimal control (e.g. in trading systems) and first-passage problems: it turns out that the stationary solution of the Bellman equation can sometimes be addressed, in a continuous limit, as a Kolmogorov or Fokker-Planck equation with vanishing Dirichlet boundary conditions. [Chicheportiche 2013c]

---

[1] See, for this, the abstracts in the Introduction, pages 7–10.

Multivariate statistics and time series analysis

The study of multivariate data analysis typically begins with the structure of linear correlations. Disentangling signal and noise in estimated covariance matrices is an old yet always actual problem. Factor models, spectral analysis, and Random Matrix Theory are some of the tools that were used in this thesis to address the linear properties of transversal dependences. There is a growing interest in the "sparse" symmetric matrices $\Sigma = \frac{1}{T} X^{\dagger} X$, with the $(T \times N)$ random entries X_{ti} distributed according to an asymptotically power-law distribution, or even as Bernoulli variables with very low probability p *not scaling with N*. As opposed to the GOE and GUE ensembles, such matrices have localized eigenvectors with rich properties. As a consequence of this non-invariance under rotations, methods relying on free matrices and R-transforms to find the spectrum of Σ in the limit of large matrices are not applicable. The replica trick provides partial results, but new methods from statistical condensed-matter physics like the 'cavity approach' for interacting particles on graphs has proven successful in obtaining precise results [Rogers 2008]: yet another bridge between cutting edge research in probability theory and statistical physics.

This thesis dissertation presented many more theoretical and applied topics in connection to multivariate time series modeling, including extreme value statistics and tail dependences, ellipticity, optimal portfolio design, non-linear dependences as measured by Blomqvist's, Kendall's or Spearman's coefficients. In this context, the fashionable 'copula' has proved a useful unifying theoretical framework, though issues remain on the estimation and interpretation of this somewhat complex object.

Other modern ways of addressing inter-dependences are attracting interest: graphs and complex networks are an alternative to simple undirected pairwise links. The higher and higher connectedness of the world has obvious illustrations in the Web 2.0 and the global financial crisis, to name only two. Social networks, systemic risk, inter-bank lending structure, shareholding capital and voting rights webs, corporate client-contractor relationships, genetic and proteomic bipartite interactions, etc. can all be studied and modeled on graphs. The emergence of huge databases and the trendy field of "Big Data" will certainly allow to complement theoretical constructions by empirical evidences and help the global pictures rely on microscopical foundations, be it in the fields of finance, economics, sociology, medicine or others.

Finance

Beside the distributional properties of financial returns, many of the most important research subjects of quantitative finance deal with dependences: the cross-sectional dependences must be well understood and modeled for optimal portfolio design and risk management, and the temporal dependences are exploited for signal prediction and optimal trading execution.

My first attempt at describing cross-sectional non-linear correlations was with hierarchical models, where each stock name is the leaf of a tree sharing branches with other leaves. The dependences between the leaves is hence caused by a "common factor" mechanism. Although such a construction allowed for a powerful interpretation of the dependence structure and reproduced many required properties of dependences among stock return series, the fine structure of dependences could not be well described. The alternative that was then finally

chosen is a multi-factor model with a flat structure. This construction, although less transparent from the point of view of economic interpretation, has several advantages: first it builds upon the traditional Principal Components Analysis, what provides a convenient benchmark; then it does not involve such a hidden structures like a "tree" (although the statistical factors do not have an obvious meaning either); and more importantly it conveniently reproduces many linear and non-linear properties of financial datasets, as studied at length in Part II.

As of the temporal dependence, all the studies in the literature point toward the same conclusion: financial time series have a long memory despite their being linearly uncorrelated! In my own studies (Part III) this showed up in many places: (i) in the long-memory kernel of the auto-regressive mechanism of the volatility dynamics; (ii) in the waiting time and recurrence intervals statistics; and (iii) in the estimation of the self-copula in connection with the effective sample size diminution in GoF tests. These many different approaches can be complemented by at least two others: the multifractal volatility models and non-stationary recent extensions thereof, and the Hawkes self-exciting mechanism. A convergence of (some of) these models is expected in the near future: first, ARCH and Hawkes description are conceptually similar, and in fact are equally able to generate power-law correlations with a power-law decreasing feedback kernel, see Appendix E.2 and Refs. [Bacry 2012, Filimonov 2013]; and secondly auto-regressive deterministic constructions must be augmented by a stochastic volatility component *à la* multifractal in order to reconciliate the Time Reversal Asymmetry with empirical low (but finite) values.

Present and future developments in econophysics and financial data analysis include understanding the same properties of temporal and cross-sectional dependences but at higher frequency. In this respect, I am involved in collaborations in view of assessing the finer structure of volatility dynamics at a scale lower than the day and develop intraday (Q)ARCH models [Blanc 2013], and also apply GoF tests on high frequency data to address the distributional properties of order books [Gould 2013].

Appendices

Non-parametric copula estimator

The copula $C(u, v)$ of a random pair (X, Y) is

$$C(u, v) = \mathbb{P}[F_X(X) \leq u, F_Y(Y) \leq v] = \mathbb{E}\left[\mathbb{1}_{\{X \leq F_X^{-1}(u)\}} \mathbb{1}_{\{Y \leq F_Y^{-1}(v)\}}\right].$$

If the univariate marginals F_X, F_Y are known, the usual empirical counterpart to the expectation operator can be used to define the empirical copula over a sample of N i.d. realizations (X_n, Y_n) of the random pair:

$$\widehat{C}(u, v) = \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{\{X_n \leq F_X^{-1}(u)\}} \mathbb{1}_{\{Y_n \leq F_Y^{-1}(v)\}} \tag{A.1}$$

which is clearly unbiased for any N. But if the marginals are unknown, they have to be themselves estimated, for example by their usual empirical counterpart. Since $F_X(x) = \mathbb{P}[X \leq x] = \mathbb{E}[\mathbb{1}_{\{X \leq x\}}]$, an unbiased ecdf is obtained as $\widehat{F_X}(x) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{X_i \leq x\}}$, but now the expected value of

$$\widehat{C}(u, v) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{\widehat{F_X}(X_i) \leq u\}} \mathbb{1}_{\{\widehat{F_Y}(Y_i) \leq v\}} \tag{A.2}$$

is not $C(u, v)$ anymore (only asymptotically is $\widehat{C}(u, v)$ unbiased), but rather

$$\mathbb{E}[\widehat{C}(u, v)] = \int \mathrm{d}F_{XY}(x, y) \, \mathbb{P}[B(x) \leq Nu - 1, B(y) \leq Nv - 1],$$

where $B_X(x) = \sum_{n < N} \mathbb{1}_{\{X_n \leq x\}}$ has a binomial distribution $\mathcal{B}(p, N-1)$ with $p = F_X(x)$ and is not independent of $B_Y(y)$. As an example, the expected value for the estimator of the independence (product) copula $C(u, v) = \Pi(u, v) = uv$ is

$$\mathbb{E}[\widehat{\Pi}(u, v)] = \frac{\lfloor Nu \rfloor}{N} \frac{\lfloor Nv \rfloor}{N}$$

resulting in a relative bias $b(u, v) = \frac{\lfloor Nu \rfloor}{Nu} \frac{\lfloor Nv \rfloor}{Nv} - 1$ vanishing only asymptotically.

As $\mathbb{E}[\widehat{C}(u, v)]$ may not be computable in the general case, we define as

$$\frac{1}{b(u, v) + 1} \widehat{C}(u, v) = \frac{Nu}{\lfloor Nu \rfloor} \frac{Nv}{\lfloor Nv \rfloor} \widehat{C}(u, v)$$

our non-parametric estimator of the copula with bias correction, even when C is not the independence copula. Therefore, our estimator is technically biased at finite N but with a good bias correction, and asymptotically unbiased.

Notice that the copula estimator $\widehat{C}$ as defined in Eq. (A.1) *is a copula* as long as F_X^{-1} and F_Y^{-1} are continuous, whereas $\widehat{C}$ as defined in Eq. (A.2) *is not a copula* on $[0, 1]^2$ for

$N < \infty$. Only on the discrete grid $\{\frac{1}{N}, \frac{2}{N}, \ldots, 1\}^2$ does Eq. (A.2) define a copula, provided the empirical inverse marginals define the ranks (or inverse "order statistic") [Deheuvels 1979, Deheuvels 1980]; in this case the bias $b(u, v)$ is zero. Continuous interpolations are not trivial, see Sect. 5.1 in [Malevergne 2006]. In practice the copula is numerically estimated on a grid with a resolution typically coarser than $\frac{1}{N}$.
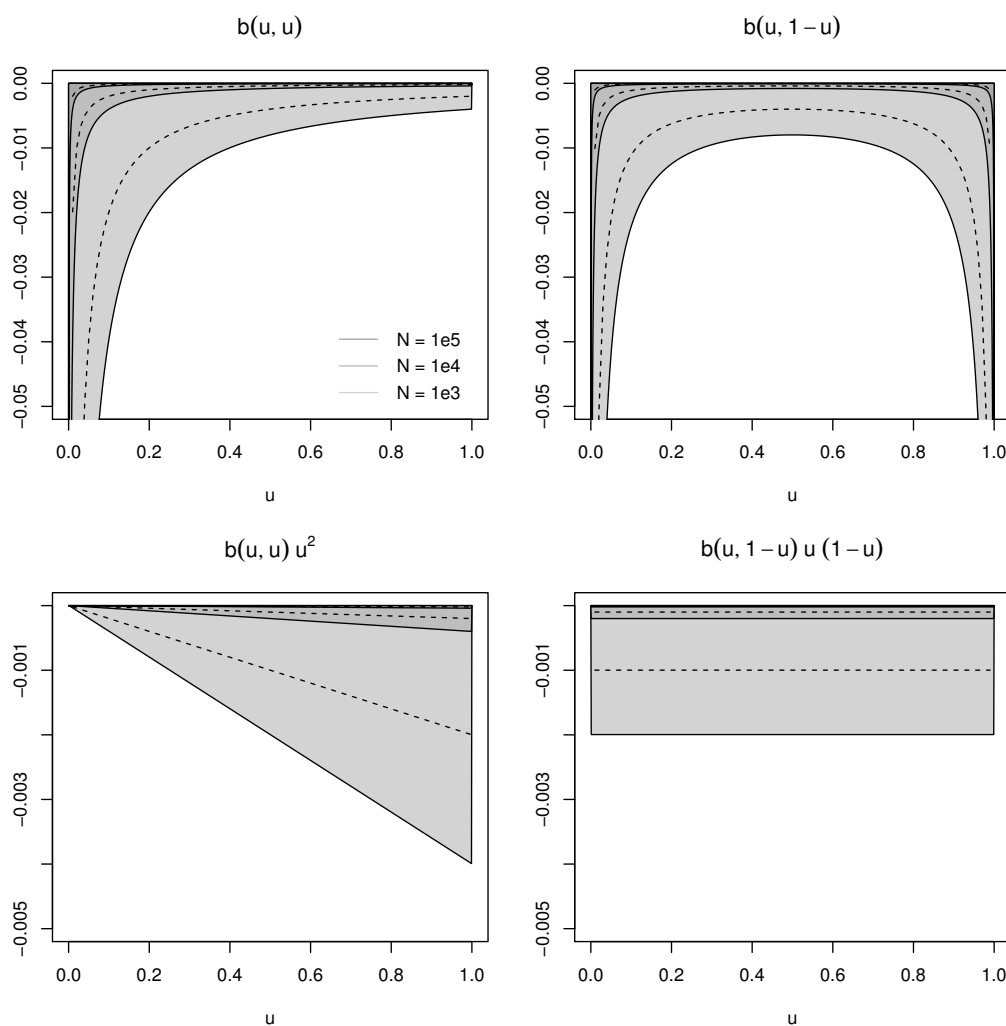
Figure A.1: Bias of the non-parametric naive estimator of the independence copula with unknown marginals, for different sample sizes (see legend). **Top:** Relative bias. **Bottom:** Absolute bias. For a given sample size N, $\lfloor Nu \rfloor$ satisfies the following relations $Nu \geq \lfloor Nu \rfloor \approx Nu - 1 \geq Nu - 2$, where the equalities hold for some values of $u \in [0, 1]$! The dashed lines represent the approximate value, and the plain lines show the bounding envelope.

Gaussian sheets and bivariate distributions testing

Goodness-of-fit testing for univariate probability distributions has been addressed in Chapter 3 of Part I, both for independent and dependent samples. We address in this appendix the generalization to multivariate samples.

We assume that the marginals were previously tested, and can work on the ranks only.

Empirical cumulative distribution and its fluctuations

Consider a bivariate sample of N joint observations (U_n, V_n) with uniform marginals. The empirical estimate of its copula,

$$C_N(u, v) = \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{\{U_n \leq u\}} \mathbb{1}_{\{V_n \leq v\}}, \tag{B.1}$$

converges to the true copula C as the sample size N tends to infinity, but is biased for finite N. In the following, we only care for asymptotic and pre-asymptotic (large $N \gg 1$) properties of C_N so that the bias is not relevant, but the bias-corrected estimator discussed in Appendix A should be considered when turning to practical applications with finite samples, i.e. when the ranks U_n and V_n are estimated rather than known. In particular, the bias correction is exact for the product copula $C = \Pi$, what may be of practical interest for the independence test worked out in Section B.1 below.

The rescaled empirical copula

$$Y_N(u, v) = \sqrt{N} \left[C_N(u, v) - C(u, v) \right] \tag{B.2}$$

measures, for a given $(u, v) \in [0, 1]^2$, the difference between the empirically determined copula and the theoretical one, evaluated at the (u, v)-th quantiles. It does not shrink to zero as $N \to \infty$, and is therefore an appropriate quantity on which to build a statistic for bivariate GoF testing.

Limit properties

One now defines the process $Y(u,v)$ as the limit of $Y_N(u,v)$ when $N \to \infty$. According to the Central Limit Theorem (CLT), it is Gaussian and its covariance function is given by:

$$H_C(u,v;u',v') = C(\min(u,u'),\min(v,v')) - C(u,v)C(u',v') \qquad (B.3)$$

$$= \begin{cases} C(u,v)\,(1 - C(u',v')) & ,u \le u', v \le v' \\ C(u,v') - C(u,v)C(u',v') & ,u \le u', v' \le v \\ C(u',v) - C(u,v)C(u',v') & ,u' \le u, v \le v' \\ C(u',v')\,(1 - C(u,v)) & ,u' \le u, v' \le v \end{cases},$$

and characterizes a pinned Gaussian sheet, i.e. a two-times Gaussian process $Y(u,v)$ such that $Y(u{=}0,v) = Y(u,v{=}0) = Y(u{=}1,v{=}1) = 0$. More precisely, H_C is the covariance function of the so-called *pinned C-Brownian sheet*, and characterizes the statistical fluctuations of the copula estimate C_N on an infinitely large sample. Importantly, Eq. (B.3) makes clear the fact that these asymptotic fluctuations still depend on the null copula C as opposed to the univariate case, where GoF tests are universal.

A global, scalar measure of distance between C_N and C is provided for example by the Cramér-von Mises statistic

$$CM = \iint_0^1 \psi(u,v)\,Y(u,v)^2 \mathrm{d}u\mathrm{d}v, \qquad (B.4a)$$

or by the Kolmogorov-Smirnov statistic

$$KS = \sup_{(u,v)\in[0,1]^2} \sqrt{\psi(u,v)}\,|Y(u,v)|, \qquad (B.4b)$$

where $\psi(u,v)$ is a weight function accounting for possible inhomogeneities to be flattened in the domain.

Finding the laws of the functionals (B.4a) and (B.4b) (and many others) of the pinned C-Brownian sheet has been the focus of intense research, but their exact distributions are still not known, even in the simplest cases when $\psi(u,v) = 1$ and/or $C = \Pi$.

To overcome the intrinsic difficulty — related in large part to the covariance (B.3) being not factorizable in u, v — two main approaches have been followed:

- Either changing the functional of $\widetilde{Y}$: computing other functionals [Cabaña 1994] or investigating separate subdomains of $[0,1]^2$ [Fasano 1987]. In this respect, there have been attemps at building single-time transformations, along a path $(u, \vartheta(u))$, or introducing more complex processes like Kendall's process [Genest 1993, Fermanian 2012];

- Or alternatively use another specification of the test that leads not to the pinned Brownian sheets, but rather to more convenient processes. This means take into account not only the global spreads of the copula $|C_N - C|$ like above, but also integrate information of the marginals. This approach is used by Deheuvels in Ref. [Deheuvels 2005] to design a Cramer-von Misés test for bivariate copulas, and is in the spirit of what was done by Justel, Peña, Zamra [Justel 1997] who build upon the Probability Integral Transformation suggested by Rosenblatt [Rosenblatt 1952], see also [Genest 2006, Fermanian 2012].

We choose to remain in the traditional setup of the Brownian sheets (B.3), and to take advantage of the weighting possibilities offered by $\psi(u,v)$ to compute the Kolmogorov-smirnov like statistic (B.4b). We focus on the KS case because of the link with first-passages in stochastic processes, see Chapter I.3 and Ref. [Chicheportiche 2013c]. Indeed, the law of the KS statistic can be written as the survival probability of a constrained process $\widetilde{Y}(u,v) = \sqrt{\psi(u,v)}\,Y(u,v)$:

$$
\begin{aligned}
S(k) &= \mathbb{P}\left[\sup_{(u,v)\in[0,1]^2} |\widetilde{Y}(u,v)| \leq k\right] \\
&= \mathbb{P}\left[-k \leq \widetilde{Y}(u,v) \leq k, \forall (u,v) \in [0,1]^2\right].
\end{aligned}
\tag{B.5}
$$

B.1   Test of independence

Consider first the case where the null-hypothesis copula is the product copula $C(u,v) = \Pi(u,v) = uv$, which is appropriate when one wants to test for independence of the variables. As a particular case of Eq. (B.3), $Y(u,v)$ has the covariance function of a standard pinned Brownian sheet (pBs):

$$I(u,v;u',v') \equiv H_\Pi(u,v;u',v') = \min(u,u')\min(v,v') - uu'vv'.$$

The unweighted Kolmogorov-Smirnov-like statistic $\sup|Y(u,v)|$ has an unknown distribution

$$S(k) = \mathbb{P}\left[\sup_{(u,v)\in[0,1]^2} |\sqrt{\psi(u,v)}Y(u,v)| \le k\right], \qquad \psi(u,v) = 1.$$

Still, there exist bounds and approximations [Adler 1986, Adler 1987], in particular:

$$1 - S(k) \ge (1 + 2k^2)\,\mathrm{e}^{-2k^2} \tag{B.6}$$

$$1 - S(k) \xrightarrow{k \gg \infty} 4\ln(2)\,k^2\,\mathrm{e}^{-2k^2}, \tag{B.7}$$

and similarly, for the C-pBS only approximations and numeric solutions are available [Adler 1990, Aldous 1989, Greenberg 2006, Fermanian 2005a].

B.1.1   Pinning the Brownian sheet ($\psi(u,v) = 1$)

The pBs is not the simplest 2-times process one could imagine. Indeed, like any pinned C-Brownian sheet it has strong constraints at the borders

$$Y(u,0) = Y(0,v) = Y(1,1) = 0.$$

However, it can be simply written as $Y(u,v) = W(u,v) - uv\,W(1,1)$, where $W(u,v)$ is a standard free Brownian sheet (two-times Wiener process) with covariance

$$\mathbb{E}[W(u,v)W(u',v')] = \min(u,u')\min(v,v'), \tag{B.8}$$

and $W(u,0) = W(0,v) = 0$. The possible trajectories (or surfaces) of $Y(u,v)$ are thus a subset of all possible paths gone by $W(u,v)$, whose free propagator is

$$\mathcal{P}_{(u,v)}(w;\infty) = \frac{1}{\sqrt{2\pi uv}}\exp\left(-\frac{1}{2}\frac{w^2}{uv}\right).$$

But of course the survival probability (B.5) involves constrained rather than free processes:

$$S(k) = \mathbb{P}\big[-k \le Y(u,v) \le k, \forall(u,v) \in [0,1]^2\big] = \frac{\mathcal{P}_{(1,1)}(0;k)}{\mathcal{P}_{(1,1)}(0;\infty)}, \tag{B.9}$$

where

$$\mathcal{P}_{(u,v)}(w;k) = \mathbb{P}\big[|W(u',v')| \le k, \forall u < u', v < v' \ \text{ and } \ W(u,v) = w\big].$$

Equation (B.9) means that the law $S(k)$ of the supabs of the standard unweighted pinned Brownian sheet is equal, up to a constant $\sqrt{2\pi}$, to the propagator of the Brownian sheet constrained in a stripe $[-k,k]$.

Unfortunately, we are not aware of a relevant Markov property that would make it possible to write (and eventually solve) a Fokker-Planck-like equation with two times. Instead, we present next an original way out, by discretizing one of the time directions.

B.1.1.1 Law of the supabs of the Brownian sheet on $[0,1]^2$: discretizing the time

$W(u,v)$ can be seen as a continuous accumulation of innovations along the two time directions. Discretizing the time in one of the dimensions, f.ex. forcing v to take values in the sequence $(v_j)_{j \leq N}$ of length N, it is possible to write $W(u,v_j) = W_j(u)$, and collect all the components in a N-dimensional vector $\boldsymbol{W}$ whose dynamics is driven by the joint SDE

$$\mathrm{d}\boldsymbol{W}(u) = \rho^{\frac{1}{2}} \, \mathrm{d}\widetilde{\boldsymbol{W}}(u),$$

where $\widetilde{\boldsymbol{W}}$ is a vector of independent Brownian motions, and ρ is the tensor of diffusion, with elements

$$\rho_{jj'} = \min(v_j, v'_j).$$

Then, the transition density $f_u(\boldsymbol{w};k)$ of the vector process $\boldsymbol{W}$ with all components constrained in $[-k,k]$ is solution of the multivariate Fokker-Planck equation

$$\frac{\partial f_u(\boldsymbol{w};k)}{\partial u} = \frac{1}{2}\nabla^\dagger \rho \nabla f_u(\boldsymbol{w};k), \quad 0 \leq u \leq 1$$

with the condition that $f_u(\boldsymbol{w};k)$ vanishes at the borders of the square box

$$\mathcal{D}_N = \big\{\boldsymbol{w} : |w_j| \leq k, \forall j \in [\![1,N]\!] \big\}.$$

Finally, the constrained two-times propagator writes

$$\mathcal{P}_{(1,1)}(w;k) = \lim_{N \to \infty} \mathbb{P}[\forall u \in [0,1], \forall j = 1 \ldots N, |W_j(u)| \leq k \ \text{ and } \ W_N(1) = w]$$

$$= \lim_{N \to \infty} \int_{-k}^{k} \mathrm{d}w_1 \cdots \int_{-k}^{k} \mathrm{d}w_N \, f_u(\boldsymbol{w};k) \, \delta(w_N - w).$$

The problem is thus equivalent to the survival of a particle diffusing in an N-dimensional cage with absorbing walls at $\pm k$, and anisotropic diffusion tensor. It can alternatively be stated as an isotropic diffusion toward the absorbing walls of an irregular box, by appropriately changing the system of coordinates.

Changing base: the isotropic problem. We first diagonalize ρ in order to obtain an isotropic and rescaled diffusion, albeit in a cage with a new geometry. The corresponding eigenvalues are collected on the diagonal of Λ and the eigenvectors in the columns of U. Changing base according to $U\Lambda^{-1/2}$, any point writes $\hat{\boldsymbol{w}} = (U\Lambda^{-1/2})^\dagger \boldsymbol{w}$ in the new coordinates. When the discretized time stamps are equidistant ($v_j = j/N$), the diffusion coefficients are $\rho_{jj'} = \frac{1}{N}\min(j,j')$ and $Q \equiv \frac{1}{N}\rho^{-1} = \frac{1}{N}U\Lambda^{-1}U^\dagger$ turns out to be the tridiagonal matrix corresponding to the discrete N-dimensional second-order finite difference operator: 2 on the main diagonal, -1 on both secondary diagonals, 1 at position (N,N), and 0 elsewhere. The eigenvectors and eigenvalues of ρ are easily found to be

$$U_{ja} = \sqrt{\frac{2}{N}}\sin\left(\frac{(2a-1)\pi j}{2N}\right) \tag{B.10a}$$

$$\lambda_a \equiv \Lambda_{aa} = \frac{4N}{(2a-1)^2\pi^2} \tag{B.10b}$$

In this new basis, the diffusion is isotropic

$$\frac{\partial f_u(\hat{\boldsymbol{w}}; k)}{\partial u} = \frac{1}{2}\nabla^2 f_u(\hat{\boldsymbol{w}}; k), \quad 0 \leq u \leq 1 \tag{B.11}$$

but the box has a new reshaped geometry

$$-k\sqrt{\lambda_j} \leq \sum_a \mathrm{U}_{aj}\hat{w}_a \leq k\sqrt{\lambda_j}, \quad \forall j = 1\ldots N. \tag{B.12}$$

The point on the boundary hypersurface that is the closest to the center is found minimizing the distance $\hat{d}^2 = \hat{\boldsymbol{w}} \cdot \hat{\boldsymbol{w}} = \boldsymbol{w}^\dagger \rho^{-1}\boldsymbol{w}$ to the center, under the constraint that at least one of the $|w_j|$ is equal to k (or equivalently maximizing the effective directional diffusion rate). This special proximal point is found to have coordinates $w_j = k\,j/N = k\,\rho_{iN}$ and be at a distance $\hat{d} = k\sqrt{N}$ proportional to the square root of the number of faces (but recall that the box expands as $\sim \sqrt{N}$ too, see Eqs. (B.12) and (B.10b)).

	Original box	Deformed geometry
$\boldsymbol{w}_{\min}/\hat{\boldsymbol{w}}_{\min}$	$k\,(1/N, 2/N, \ldots, 1)$	$k\,(\mathrm{U}_{N1}/\sqrt{\lambda_1}, \ldots, \mathrm{U}_{NN}/\sqrt{\lambda_N})$
$d_{\min}/\hat{d}_{\min}$	$\frac{k}{N}\sqrt{\frac{N^3}{3} + \frac{N^2}{2} + \frac{N}{6}}$	$k\sqrt{N}$
$\sigma_{\min}/\hat{\sigma}_{\min}$	$\sqrt{\frac{w_{\min}^\dagger w_{\min}}{w_{\min}^\dagger \rho^{-1} w_{\min}}} \approx \sqrt{\frac{N}{3}}$	1

In fact, it is possible to calculate the distance from the center to every face. Looking for the vector $w^{(n)}$ such that

$$w^{(n)} = \arg\min\{\hat{w}^\dagger\hat{w} \ \text{ s.t. } \ w_n = k\} = \arg\min\{w^\dagger\rho^{-1}w \ \text{ s.t. } \ w_n = k\}$$

$$= \arg\min\left\{k^2\mathrm{Q}_{nn} + 2k\sum_{m\neq n}\mathrm{Q}_{nm}w_m + \sum_{m,m'\neq n}w_m'\mathrm{Q}_{m'm}w_m\right\}$$

$$= \arg\left\{\sum_{m\neq n}\mathrm{Q}_{m'm}w_m = -k\,\mathrm{Q}_{m'n}, \ m' \neq n\right\},$$

one finds

$$w_j^{(n)} = k \cdot \begin{cases} \frac{j}{n} & ,j \leq n \leq N \\ 1 & ,n \leq j \leq N \end{cases},$$

meaning that the closest point to *any* face is on an edge, since $w_j^{(n)}$ touches $N - n + 1$ faces. In particular, the closest exit in any direction is on the N-th face. We already know that $\hat{d}_{\min} = \hat{d}^{(N)} = k\sqrt{N}$, and we now find more generally that $\hat{d}^{(n)} = \sqrt{\hat{w}^{(n)\dagger}\hat{w}^{(n)}} = kN/\sqrt{n}$.

B.2 2D Kolmogorov-smirnov test of Goodness-of-fit

The strategy:

- Unweighted: Look for the time change that transforms any C-pBs into the standard pBs. This is achieved in Sect. B.2.1 just below. Then, the 2D-Kolmogorov-Smirnov test is essentially related to the law of the supabs of the pBs, Sect. B.1.

- Weighted: Change of time and variable to recover nice processes: 2D Brownian Motion (Brownian sheet) or 2D Ornstein-Uhlenbeck, or mixed Brownian Motion/Ornstein-Uhlenbeck.

B.2.1 Flattening an arbitrary copula

Let $C(u, v)$ be a two-dimensional copula function. We consider the following problem: find a parametrization $s(u, v)$ increasing in u for all v, and $t(u, v)$ increasing in v for all u, that brings the copula $C(u, v)$ to the independence copula $\Pi(s, t) = st$. A solution to this problem in *any dimension* with no assumption but absolute continuity of the joint cdf is known as Rosenblatt's transform [Rosenblatt 1952], but it relies on an arbitrary ordering of the variables. We instead propose a solution for the 2D case with no particular order of the variables. This is better stated combining integral and differential representations, and thus necessitates the appropriate differentiability conditions:

$$\begin{cases} C(u, v) & = \Pi(s(u, v), t(u, v)) = s(u, v)t(u, v) \\ c(u, v)\, \mathrm{d}u\mathrm{d}v & = \pi(s, t)\, \mathrm{d}s\mathrm{d}t = \left| \dfrac{\partial(s, t)}{\partial(u, v)} \right| \mathrm{d}u\mathrm{d}v \end{cases} \tag{B.13}$$

with c and $\pi \equiv 1$ the densities of C and Π respectively: we want to stretch the probability space in a particular fashion so that the copula density in that space is uniform. Eliminating $t(u, v)$ from the system above, the solution satisfies the Monge-Ampère partial differential equation

$$c(u, v) = \left| \frac{1}{s}\frac{\partial s}{\partial u}\frac{\partial C}{\partial v} - \frac{1}{s}\frac{\partial s}{\partial v}\frac{\partial C}{\partial u} \right| = ||\nabla \ln s(u, v) \wedge \nabla C(u, v)||. \tag{B.14}$$

This equation for $\ln s$ can be solved by the "method of characteristics": first, define the parametric trajectory — the "characteristic curve" — $\boldsymbol{U}(\gamma) \equiv (u(\gamma), v(\gamma))$ satisfying

$$\begin{cases} \dfrac{\mathrm{d}u}{\mathrm{d}\gamma} & = \dfrac{\partial C(u, v)}{\partial v} \\ \dfrac{\mathrm{d}v}{\mathrm{d}\gamma} & = -\dfrac{\partial C(u, v)}{\partial u} \end{cases} \quad \text{along which} \quad c(u(\gamma), v(\gamma)) = \left| \dfrac{\mathrm{d}s(u(\gamma), v(\gamma))}{\mathrm{d}\gamma} \right| \tag{B.14'}$$

The algebraic interpretation is the following:

$$\mathrm{d}C(u(\gamma), v(\gamma)) = \frac{\partial C}{\partial u}(u(z), v(z))\, \mathrm{d}u(z) + \frac{\partial C}{\partial v}(u(z), v(z))\, \mathrm{d}v(z) = 0$$
$$\nabla C(\boldsymbol{U}(\gamma)) \cdot \mathrm{d}\boldsymbol{U}(\gamma) = 0$$

so $C(u(\gamma), v(\gamma)) = \mathcal{C} = \text{const}$, what defines the function $v_{\mathcal{C}}(u')$ on $u' \in [\mathcal{C}, 1]$ as imposed by the Frechet bounds, see page 19. Notice that $v_{\mathcal{C}}(u')$ is decreasing in u' along a characteristic

of fixed $\mathcal{C}$, increasing in $\mathcal{C}$ for a given $u' > 0$, and that $\lim_{\mathcal{C} \to 1} v_{\mathcal{C}}(u') = 1$, but the limit as $\mathcal{C} \to 0$ is degenerate since the copula is zero at any point on the edges $(u, 0)$ and $(0, v)$. Then, the solution of Eq. (B.14') *along a given characteristic* is found integrating

$$\mathrm{d} \ln s = c(u(\gamma), v(\gamma)) \mathrm{d}\gamma = c(u', v_{\mathcal{C}}(u')) \frac{\mathrm{d}\gamma}{\mathrm{d}u'} \, \mathrm{d}u',$$

yielding

$$\ln s_{\mathcal{C}}(u) - \ln s_{\mathcal{C}}(u^*) = \int_{u*}^{u} c(u', v_{\mathcal{C}}(u')) \left[\frac{\partial C}{\partial v}(u', v_{\mathcal{C}}(u')) \right]^{-1} \mathrm{d}u'. \qquad (\text{B.15a})$$

Since the density $c(u, v)$ and the first derivative $\partial C / \partial v$ are always positive, $s_{\mathcal{C}}(u)$ is increasing in u at given $\mathcal{C}$. Because the problem is symmetric in $s \leftrightarrow t$, one could as well eliminate s from the initial system of equations, and solve it to obtain

$$\ln t_{\mathcal{C}}(v) - \ln t_{\mathcal{C}}(v^*) = \int_{v*}^{v} c(u_{\mathcal{C}}(v'), v') \left[\frac{\partial C}{\partial u}(u_{\mathcal{C}}(v'), v') \right]^{-1} \mathrm{d}v'. \qquad (\text{B.15b})$$

The final solution of Eq. (B.14) is

$$s(u, v) = s_{C(u,v)}(u) \qquad\qquad\qquad\qquad\qquad (\text{B.16a})$$

$$t(u, v) = t_{C(u,v)}(v) = C(u, v)/s(u, v), \qquad\qquad (\text{B.16b})$$

and it is necessary to check that it satisfies the conditions set at the begining, namely that $s(u, v)$ be increasing in u and $t(u, v)$ be increasing in v.

The integration constants $s_{\mathcal{C}}(u^*)$ and $t_{\mathcal{C}}(v^*)$ are determined as follows. Evaluating simultaneously Eq. (B.16a) at a point $(u^*, v_{\mathcal{C}}(u^*))$ and Eq. (B.16b) at a point $(u_{\mathcal{C}}(v^*), v^*)$, both on the characteristic curve, and taking the product, one gets

$$s(u^*, v_{\mathcal{C}}(u^*)) \, t(u_{\mathcal{C}}(v^*), v^*) = \mathcal{C}.$$

But since one also has generically $s(u^*, v^*) \, t(u^*, v^*) = \mathcal{C}$, the identification yields

$$u_{\mathcal{C}}(v^*) = u^* \qquad \text{and} \qquad v_{\mathcal{C}}(u^*) = v^* \qquad\qquad (\text{B.17})$$

For a symmetric copula, one finds that $s_{\mathcal{C}}(u^*) = t_{\mathcal{C}}(v^*) = \sqrt{\mathcal{C}}$ for $u^* = v^*$ solution of $C(u^*, u^*) = \mathcal{C}$. In this case, Eq. (B.16b) makes also clear that $t(u, v) = s(v, u)$.
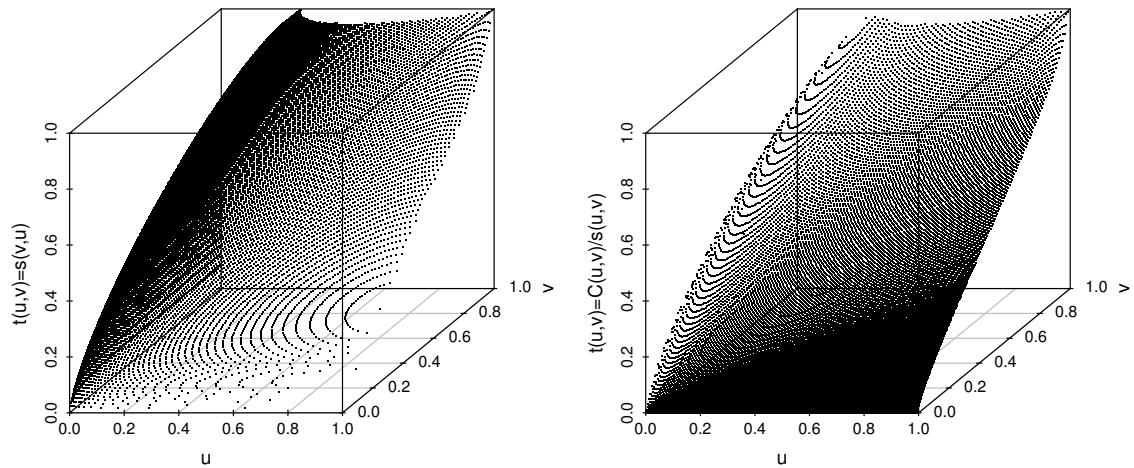
Figure B.1: Flattening the Gaussian copula, example with a correlation coefficient $\rho = 0.5$. The time change $s(u, v)$ is computed from Eqs. (B.15a,B.16a), then $t(u, v)$ is checked for consistency: **Left:** $t(u, v) = s(v, u)$; **Right:** $t(u, v) = C(u, v)/s(u, v)$.

B.2.2   Variance-weighting the C-pBs $\left(\psi(u,v) = 1/H_C(u,v;u,v)\right)$

Let $Y(u,v)$ be the pinned C-Brownian sheet on $(u,v) \in [\epsilon, 1-\epsilon]^2$, so that, in particular, $\mathbb{V}[Y(u,v)] = H_C(u,v;u,v)$ as defined in Eq. (B.3). The time-changed rescaled process

$$Z(s,t) = Y(f(s,t), g(s,t))/\sqrt{\mathbb{V}[Y(f(s,t), g(s,t))]}.$$

is a 2D Ornstein-Uhlenbeck on $(s,t) \in [0, T_\epsilon]^2$, if

$$\begin{cases} C(f(s,t), g(s,t)) &= \dfrac{\sqrt{\frac{C(\epsilon,\epsilon)}{1-C(\epsilon,\epsilon)}} \, \mathrm{e}^{s+t}}{\left[\sqrt{\frac{C(\epsilon,\epsilon)}{1-C(\epsilon,\epsilon)}} \, \mathrm{e}^{s+t}\right]^{-1} + \sqrt{\frac{C(\epsilon,\epsilon)}{1-C(\epsilon,\epsilon)}} \, \mathrm{e}^{s+t}} \\[1em] f(s,t) \le f(s',t') &\Leftrightarrow g(s,t) \le g(s',t') \end{cases},$$

since one shows that the correlations are

$$\mathbb{E}[Z(s,t)Z(s',t')] = \mathrm{e}^{-|s-s'|} \, \mathrm{e}^{-|t-t'|},$$

and decay exponentially with the time lags, as expected for a *stationary* OU process with distribution $\mathcal{N}(0,1)$ (as long as $0 < \epsilon < 1$).

Obviously, since the pBs Y is not Markovian, the description in terms of the Markovian process Z is only possible in a limit sense, where the corner $(u,v) = (1,1)$ is never reached, thanks to a singular change of time: $(u,v) = (\epsilon, \epsilon)$ maps onto $(s,t) = (0,0)$, and $(u,v) = (1-\epsilon, 1-\epsilon)$ maps onto $(s,t) = (T_\epsilon, T_\epsilon)$ where

$$T_\epsilon = \ln \sqrt[4]{\frac{C(1-\epsilon, 1-\epsilon)}{1 - C(1-\epsilon, 1-\epsilon)} \frac{1 - C(\epsilon, \epsilon)}{C(\epsilon, \epsilon)}} \xrightarrow{\epsilon \to 0} \infty.$$

The diagonal and anti-diagonal copula can be written in terms of tail dependence coefficients:

$$T_\epsilon = -4 \ln\left[\epsilon^2 \, \tau^{\mathrm{LL}}(1-\epsilon) \left(2 - \tau^{\mathrm{UU}}(1-\epsilon)\right)\right] \xrightarrow{\epsilon \to 0} \infty,$$

and the whole space $[0,1]^2$ tends to be covered only when T_ϵ diverges to ∞.

In the notations of page 180, $u = f(s,t)$, $v = g(s,t)$ and the weights are

$$\sqrt{\psi(f(s,t), g(s,t))} = \left[\sqrt{\frac{\epsilon}{1-\epsilon}} \, \mathrm{e}^{s+t}\right]^{-1} + \sqrt{\frac{\epsilon}{1-\epsilon}} \, \mathrm{e}^{s+t}.$$

The law of the corresponding KS statistic (B.4b) writes now

$$S(k) = \lim_{\epsilon \to 0} \mathbb{P}\left[-k \le Z(s,t) \le k, \forall (s,t) \in [0, T_\epsilon]^2\right]$$

$$= \int_{-k}^{k} \mathrm{d}z \int_{-k}^{k} \mathrm{d}z_0 \, \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-z_0^2/2} \, \mathcal{P}_{(T_\epsilon, T_\epsilon)}(z - z_0; k)$$

and is the survival probability of the 2D mean-reverting process. Note that we have averaged over the initial position z_0 since $Z(0) = Y(\epsilon, \epsilon)/[\epsilon (1-\epsilon)] \sim \mathcal{N}(0,1)$. Unfortunately, although the free propagator of the OU process $Z(u,v)$ is known to be

$$\mathcal{P}_{(s,t)}(z; \infty) = \frac{1}{\sqrt{2\pi st (1 - st)}} \exp\left(-\frac{1}{2} \frac{z^2}{st (1 - st)}\right),$$

the propagator of the constrained process is unknown. Again, there is no known Partial Derivatives Equation (PDE) of the Fokker-Planck type for two times, that would describe the solution with arbitrary boundary conditions, and whose free solution would be $\mathcal{P}_{(s,t)}(z; \infty)$.

B.2.2.1 Discretizing one time direction

Define $Z_n(t) = Z(s_n, t)$ for a $\mathbb{R}$-valued increasing sequence $(s_n)_{n \in [\![0, N]\!]}$. The vector $\boldsymbol{Z}(t)$ represents a collection of $(N+1)$ OU processes with covariance matrix

$$\rho_{nm} = \mathbb{E}[Z_n(t)Z_m(t)] = \mathrm{e}^{-|s_n - s_m|}.$$

The vectorial SDE writes $\mathrm{d}\boldsymbol{Z}(t) = -\boldsymbol{Z}(t)\,\mathrm{d}t + \rho^{\frac{1}{2}}\,\mathrm{d}\boldsymbol{W}(t)$ with the initial condition $\boldsymbol{Z}(0) \sim \mathcal{N}(0, \rho)$, and the FPE is

$$\partial_t f_t(\boldsymbol{z}; k) = f_t(\boldsymbol{z}; k) + \boldsymbol{z} \cdot \nabla f_t(\boldsymbol{z}; k) + \frac{1}{2} \nabla^\dagger \rho \nabla f_t(\boldsymbol{z}; k),$$

where $\boldsymbol{z}$ lives in the interior of the $(N+1)$-dimensional square box

$$\mathcal{D}_N = \big\{ \boldsymbol{z} : |z_n| \leq k, \forall n \in [\![0, N]\!] \big\},$$

and the density $f_t(\boldsymbol{z}; k)$ vanishes at the border $\partial \mathcal{D}_N$.

If the sequence (s_n) covers the entire interval $[0, T_\epsilon]$ as $N \to \infty$, then

$$S(k) = \lim_{\epsilon \to 0} \lim_{N \to \infty} \mathbb{P}[-k \leq Z(s_n, t) \leq k, \forall n \in [\![0, N]\!], \forall t \in [0, T_\epsilon]]$$

$$= \lim_{\epsilon \to 0} \lim_{N \to \infty} \int_{\mathcal{D}_N} \mathrm{d}\boldsymbol{z} \int_{-k}^{k} \mathrm{d}\boldsymbol{z}_0 \, \frac{1}{\sqrt{(2\pi)^N \det \rho}} \mathrm{e}^{-\boldsymbol{z}_0 \rho^{-1} \boldsymbol{z}_0 / 2} \, f_{T_\epsilon}(\boldsymbol{z} - \boldsymbol{z}_0; k)$$

The order of the limits matters, because one is likely to be interested in the preasymptotic regime where $\epsilon > 0$!

B.2.3 Unpinning the C-pBs

We saw above that it is possible to perform a change of time and variable in order to write the unweighted pBs as a Brownian sheet that is eventually pinned at (1,1) (Section B.1.1), and that the variance-weighted pBs could as well be rewritten as an unpinned process, namely the 2D Ornstein-Uhlenbeck (Section B.2.2). Moreover, by discretizing one time direction, the problems reduced to those of a collection of N processes (Brownian or OU, respectively). We now intend to write the C-pBs as a collection of Brownian Motions by a change of time.

We consider a pinned C-Brownian sheet, and look for a mapping between the coordinates $(u, v) \in [\epsilon_s, 1-\epsilon_s] \times [\epsilon_t, 1-\epsilon_t]$ and $(s, t) \in [0, T_{\epsilon_s}] \times [0, T_{\epsilon_t}]$. We write this mapping as

$$u = f(s, t) \quad v = g(s, t),$$

and define simultaneously $X(s, t) = \widetilde{Y}(f(s, t), g(s, t))$. For every s, the covariance of $X(s, t)$ is that of a Brownian motion

$$\mathbb{E}[X(s, t)X(s, t')] = \gamma(s) \, \min(t, t'), \tag{B.18}$$

if there exist characteristic times $\tau(s) > 0$ and increasing mappings f, g, such that

$$\psi(f(s, t), g(s, t)) = \gamma(s) \left(\tau(s) + 2t + t^2/\tau(s) \right) = \frac{\gamma(s)}{\tau(s)} \left(\tau(s) + t \right)^2$$

and

$$C(f(s,t), g(s,t)) = \frac{t}{\tau(s) + t}, \quad \forall s. \tag{B.19}$$

Along two trajectories $s \neq s'$ indexed by the same time $t' = t$, we find

$$\mathbb{E}[X(s,t)X(s',t')] = \sqrt{\gamma(s)\gamma(s')}\,\rho(s,s')\,\min(t,t')$$

where $\rho(s,s') = \sqrt{\tau_{\max}/\tau_{\min}}$.

Example: By choosing $\gamma(s) = 1$ and $\tau(s) = \tau_0\,\mathrm{e}^{-2s}$, the equal-time correlation has the exponential form $\rho(s,s') = \mathrm{e}^{-|s-s'|}$, so $X(s,t)$ is a Gauss-Markov process in s at every t ! We get back a case similar to that treated in Sect. B.1.1: a Brownian motion in t and an Ornstein-Uhlenbeck in s, although there we considered a regular unpinned Brownian sheet on $[0,1]^2$, and we discretized one time direction uniformly. According to Eq. (B.19), we can also write inversely

$$\mathrm{e}^{2s}\,t/\tau_0 = \frac{C(u,v)}{1 - C(u,v)} \tag{B.20}$$

$$T_{\epsilon_s} + \ln\sqrt{\frac{T_{\epsilon_t}}{\tau_0}} = \ln\sqrt{\frac{C(1-\epsilon_s, 1-\epsilon_t)}{1 - C(1-\epsilon_s, 1-\epsilon_t)}} \tag{B.21}$$

Since the times s and t are of different "nature" (one is exponential, the other is linear), it is convenient to choose the horizon such that

$$T_{\epsilon_t} = \tau_0\,\mathrm{e}^{2T_{\epsilon_s}} = \tau_0\,\sqrt{\frac{C(1-\epsilon, 1-\epsilon)}{1 - C(1-\epsilon, 1-\epsilon)}} \equiv T_\epsilon,$$

where we simply impose $\epsilon_s = \epsilon_t \equiv \epsilon$.

Example (a path in the plane): By choosing $\tau(s) = \tau_0$, the equal-time correlation is simply $\rho(s,s') = 1$, meaning that all the $X(s,\cdot)$ are equal: this is a "Gaussian front". We can take arbitrarily $s = t$, which is equivalent to considering the joint mappings $(u,v) \equiv (u, \vartheta(u))$, s.t. $\vartheta(1-\epsilon_s) = 1-\epsilon_t \equiv 1-\epsilon$, i.e.

$$u \equiv f(t,t) \quad \text{and} \quad v \equiv g(t,t) \equiv (\vartheta \circ f)(t,t).$$

From Eq. (B.19), we can also write inversely

$$\psi(u, \vartheta(u)) = \tau_0/[1 - C(u, \vartheta(u))]^2 \tag{B.22}$$

$$t/\tau_0 = \frac{C(u, \vartheta(u))}{1 - C(u, \vartheta(u))} \tag{B.23}$$

$$T_\epsilon = \tau_0\frac{C(1-\epsilon, \vartheta(1-\epsilon))}{1 - C(1-\epsilon, \vartheta(1-\epsilon))} \tag{B.24}$$

This is very similar to the unidimensional case, since we investigate only a path in the space $[0,1]^2$.

B.2.3.1 Special case $\vartheta(u) = u^\alpha$ for independence copula

We now look more carfully at the above example where $v = \vartheta(u)$, in the particular case where the copula is $C(u, v) = \Pi(u, v) = uv$. The weights are

$$\psi(u, u^\alpha) = \tau_0 / [1 - u^{\alpha+1}]^2 \tag{B.25a}$$

$$\psi(f(t,t), f(t,t)^\alpha) = \tau_0 \left[1 + t/\tau_0\right]^2 \tag{B.25b}$$

and the time changes f, g are

$$f_\alpha(t) = \left(\frac{t}{t + \tau_0}\right)^{\frac{1}{1+\alpha}} \quad \text{et} \quad g_\alpha(t) = \left(\frac{t}{t + \tau_0}\right)^{\frac{\alpha}{1+\alpha}}$$

whatever the exponent $\alpha \neq -1$, where $t \leq T_\epsilon(\alpha) \equiv (1 - \epsilon)^{\alpha+1} / [1 - (1 - \epsilon)^{\alpha+1}]$. We exhibit a sequence $\{\alpha(n; N)\}_{-N < n < N}$ such that the whole space is covered in the limit $N \to \infty$, and such that the functions f_n, g_n have the required properties:

$$\alpha(n; N) = \begin{cases} \dfrac{N - n}{N} & \text{si } 0 \leq n < N \\ \dfrac{N}{N - |n|} & \text{si } -N < n \leq 0 \end{cases}$$

One notices that there is no condition on τ_0, and it is possible to take $\tau_0 = 1$ for all n ! The time change specific to each n is performed through α_n, and the characteristic time scale is not important (as long as the change of time remains singular). Finally we get $\psi(n, t) = (1 + t)^2$, and

$$f(n, t; N) = \left(\frac{t}{t + 1}\right)^{\frac{1}{1+\alpha(n;N)}} \quad \text{et} \quad g(n, t; N) = \left(\frac{t}{t + 1}\right)^{\frac{\alpha(n;N)}{1+\alpha(n;N)}} .$$

It is important to notice that the weights B.25 do not depend at all on the function $\vartheta(\cdot)$ chosen to parameterize the second time. They define a curve, and not a surface in $[0, 1]^2$. Nevertheless, by following several curves $\vartheta(u) = u^\alpha(n; N)$ (with $-N < n < N$), it is possible to cover $[0, 1]^2$ in the limit $N \to \infty$, but with different weights ψ for every n.

B.3 Perspectives

The diagram on Fig. B.2 summarizes the possible strategies outlined in this chapter to address the problem of finding the sup of a pinned C-Brownian bridge. Appropriate time changes allow to transform the problem and handle stochastic processes that are more common and whose properties are better known, like Brownian motions (BM), 1D or 2D Ornstein-Uhlenbeck (OU) processes, Brownian sheets (Bs). We have provided the explicit transformations needed (change of time and change of variable, weighting scheme) for every strategy.

The most important original contribution of this chapter is however to suggest a discretization of one time direction, that ultimately leads to the problem of finding the density of presence of a stochastic particle in an infinite-dimensional cage with absorbing walls.

This opens a new way of addressing stochastic processes indexed by two times, and we hope that future efforts can lead to a better characterization of the law of the supremum.

C–pBs

time change

pBs
(B.2.1)

specific weights
C–pBs

variance–weighted
C–pBs

time change

time change

Bs on $[0, 1]^2$
(B.1.1)

BM + ? on $[0, \infty]^2$
(B.2.3)

OU on $[0, \infty]^2$
(B.2.2)

BM+OU

BM along path
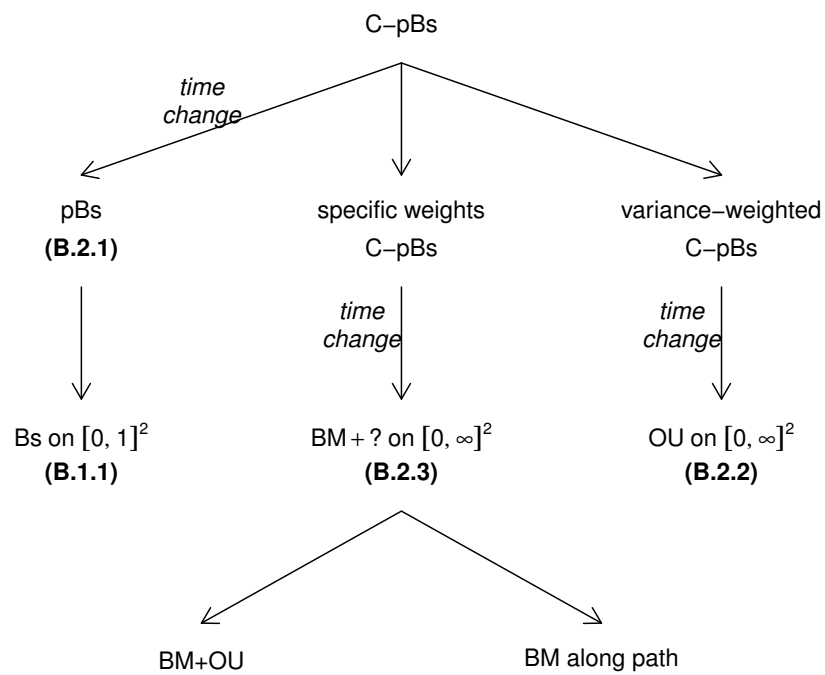
Figure B.2: The different strategies to address the problem of finding the sup of the pinned C-Brownian bridge.

One-factor linear model

C.1 Diagonally perturbed rank-1 operator

Let $|\beta\rangle$ be a vector whose N components are required to satisfy the following $N(N+1)/2$ equations

$$\mathfrak{C}_{ij} = \epsilon\delta_{ij} + (1 - \epsilon\delta_{ij})\,\beta_i\beta_j, \tag{C.1}$$

or in matrix form (and using Dirac's bra-ket notation):

$$\mathfrak{C} = \epsilon\left[\mathbb{1} - \mathrm{diag}(\beta^2)\right] + |\beta\rangle\langle\beta| \tag{C.2}$$

where $\mathfrak{C}$ is a given symmetric positive definite matrix. When $\epsilon = 1$, this amounts to trying to explain the correlation structure of N random variables with only N parameters. This case is typically encountered in factor models with linear exposition of every individual to a common hidden variable.

Of course this problem does not have a solution in the general case, due to the number of variables being much lower than the number of constraints. It is then better stated in a "best quadratic fit" (least-squares estimation) sense: minimize the loss function (Hamiltonian) given by the Mean Squared Error (MSE)

$$|\beta^*\rangle = \arg\min_\beta\left\{\frac{1}{N(N-1)}\sum_{i,j\neq i}(\mathfrak{C}_{ij} - \beta_i\beta_j)^2\right\}.$$

Differentiation yields the set of equations for the optimal solution (we drop the *):

$$\sum_{i=1}^{N}\mathfrak{C}_{ki}\beta_i = \left[\langle\beta|\beta\rangle + \epsilon\left(1 - \beta_k^2\right)\right]\beta_k. \tag{C.3}$$

$|\beta\rangle$ can be decomposed onto contributions longitudinal and orthogonal to the first (normalized) eigenvector $|V\rangle$ of $\mathfrak{C}$ (corresponding to the largest eigenvalue λ):

$$|\beta\rangle = v|V\rangle + \epsilon\left[A_\parallel|V\rangle + |V^\perp\rangle\right]. \tag{C.4}$$

The correction is explicitly written in multiples of ϵ, since to zero-th order one indeed recovers the eigenvalue equation $\mathfrak{C}|\beta\rangle = \langle\beta|\beta\rangle|\beta\rangle$ with solution $|\beta\rangle = \sqrt{\lambda}\,|V\rangle$.[1] At this order, $\langle\beta|\beta\rangle = \lambda \sim N$, so it dominates any ϵ-correction.

---

[1] In fact any couple of eigenvalue/eigenvector of $\mathfrak{C}$ is solution, but the projector $\lambda|V\rangle\langle V|$ associated to the largest eigenvalue has the largest information content in the class of rank-one operators.

In full generality, the squared norm of $|\beta\rangle$ is

$$\langle\beta|\beta\rangle = (v + \epsilon A_\parallel)^2 + \epsilon^2\langle V^\perp|V^\perp\rangle = v^2 + 2v\epsilon A_\parallel + \epsilon^2[A_\parallel^2 + \langle V^\perp|V^\perp\rangle]$$

and, to first order in ϵ/N, Eq. (C.3) writes

$$A_\parallel\lambda V_k + \sum_i \mathfrak{C}_{ki}V_k^\perp = \left(\lambda A_\parallel + 2\lambda A_\parallel + \sqrt{\lambda}(1 - \lambda V_k^2)\right)V_k + \lambda V_k^\perp$$

Now, decomposing

$$V_k^3 = \langle 1|V^4\rangle V_k + W_k^\perp$$

longitudinal and transverse components must satisfy

$$2\lambda A_\parallel + \sqrt{\lambda} = \lambda^{3/2}\langle 1|V^4\rangle$$
$$(\lambda\mathbb{1} - \mathfrak{C})|V^\perp\rangle = \lambda^{3/2}|W^\perp\rangle,$$

i.e.

$$A_\parallel = \frac{1}{2}\left(\langle 1|V^4\rangle\lambda^{1/2} - \lambda^{-1/2}\right) \qquad\qquad \sim 1/\sqrt{N} \qquad\qquad \text{(C.5)}$$
$$|V^\perp\rangle = \lambda^{3/2}(\lambda\mathbb{1} - \mathfrak{C})^{-1}|W^\perp\rangle \qquad\qquad\qquad \text{(C.6)}$$
$$\approx \sqrt{\lambda}\left(|V^3\rangle - \langle 1|V^4\rangle|V\rangle\right) \qquad\qquad \sim 1 \qquad\qquad \text{(C.7)}$$

so that the final solution is

$$\boxed{|\beta\rangle/\sqrt{\lambda} = \left[1 - \frac{\epsilon}{2}\left(\langle 1|V^4\rangle + 1/\lambda\right)\right]|V\rangle + \epsilon|V^3\rangle}$$

C.2 Nested single-factor linear model

One-factor model

Let $\boldsymbol{r}$ be an S-variate random vector $\boldsymbol{r}$, and $\mathfrak{C} = \mathbb{E}[\boldsymbol{r}\boldsymbol{r}^\dagger]$ its correlation matrix — everything is normalized. The linear correlation structure of the S random variables r_s is fully determined by their exposure to a common factor $r^{(0)}$ of zero mean and unit variance:

$$r_s = \beta_s r^{(0)} + e_s. \tag{C.8}$$

The residuals are assumed uncorrelated[2] with variances $\mathbb{V}[e_s] = 1 - \beta_s^2$.

Estimating the model

We address in this section the question of the estimation of the parameters $\boldsymbol{\beta}$. This simple setup is intended to be the building block of a more complex auto-similar description of hierarchically organized variables.

When the factor $r^{(0)}$ is known

In this case, Eq. (C.8) is seen as a regression, and the OLS estimates of $\boldsymbol{\beta}$ are

$$\hat{\beta}_s = \langle r_s r^{(0)} \rangle, \tag{C.9}$$

where $\langle \cdot \rangle$ means sample average over *realizations* of the random variables.

The theoretical bias is $\mathbb{E}[\hat{\beta}_s] - \beta_s = \mathbb{E}[e_s r^{(0)}]$, so that the OLS estimator is unbiased only when the correlation of the residual with the explanatory variable is nil. When the residuals are independent Gaussian, the OLS has optimal properties (minimal asymptotic variance).

When the factor $r^{(0)}$ is not known

In this case another set of model predictions should be used to estimate the exposure parameters, for example the correlation structure

$$\mathbb{E}[r_s r_u] = \beta_s \beta_u + \beta_s \mathbb{E}[e_u r^{(0)}] + \beta_u \mathbb{E}[e_s r^{(0)}] + \mathbb{E}[e_s e_u]. \tag{C.10}$$

With uncorrelated residuals, we can build an estimator which leaves (C.10) unbiased under the same conditions as (C.9):

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \mathcal{L}_\beta(\boldsymbol{\beta}|\boldsymbol{r}) \tag{C.11}$$

where we have introduced the loss function $\mathcal{L}_\beta(\boldsymbol{\beta}|\boldsymbol{r}) = \left|\left|\mathfrak{C} - \boldsymbol{\beta}\boldsymbol{\beta}^\dagger\right|\right|$ with an appropriate "norm" taking care of the peculiarities of the correlation matrix, f.ex. the RMS over off-diagonal elements of a symmetric matrix:

$$||\mathrm{A}|| = \sqrt{\frac{1}{S(S-1)} \sum_{j<i} \mathrm{A}_{ij}^2}.$$

In this case, the perturbative method of Sect. C.1 can be used, otherwise a numerical solution must be looked for.

---

[2]They need not be independent. If they are, the whole non-linear dependence structure is fully determined by Eq. (C.8).

Reconstructing the factor   For some applications, it might be necessary to have in hands the series of the residuals, or even that of the factor itself. The factor has then to be reconstructed from the individual variables with some weights w_s: $\hat{r}^{(0)}(\boldsymbol{w})$. It is clear that any such attempt will unavoidably introduce correlations of the form that bias the estimates of $\boldsymbol{\beta}$. But this in turn precisely provides an optimality criterion for the choice of the weights:

$$\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \underbrace{\left|\left|\langle \boldsymbol{e}\hat{r}^{(0)}(\boldsymbol{w})^{\dagger}\rangle\right|\right|}_{\mathcal{L}_w(\boldsymbol{w}|\boldsymbol{r},\boldsymbol{\beta})} = \arg\min_{\boldsymbol{w}} \sqrt{\left(\langle\boldsymbol{r}\hat{r}^{(0)}(\boldsymbol{w})\rangle - \boldsymbol{\beta}\right)^2} \tag{C.12}$$

For consistency, the residual correlation $\mathcal{L}_w(\hat{\boldsymbol{w}}|\boldsymbol{r},\hat{\boldsymbol{\beta}})$ has to be very small !

A simple and intuitive expression for the factor variable is a linear combination of its constituents: $\hat{r}^{(0)}(\boldsymbol{w}) = \boldsymbol{r}^{\dagger}\boldsymbol{w}/(\boldsymbol{w}^{\dagger}\mathfrak{C}\boldsymbol{w})^{\frac{1}{2}}$. In this case, the loss function writes

$$\mathcal{L}_w(\boldsymbol{w}|\boldsymbol{r},\hat{\boldsymbol{\beta}})^2 = \frac{\boldsymbol{w}^{\dagger}\mathfrak{C}\mathfrak{C}\boldsymbol{w}}{\boldsymbol{w}^{\dagger}\mathfrak{C}\boldsymbol{w}} - 2\frac{\hat{\boldsymbol{\beta}}^{\dagger}\mathfrak{C}\boldsymbol{w}}{\sqrt{\boldsymbol{w}^{\dagger}\mathfrak{C}\boldsymbol{w}}} + \hat{\boldsymbol{\beta}}^{\dagger}\hat{\boldsymbol{\beta}}.$$

For computational convenience, the normalization of the weights can even be performed *ex post* so that the minimization program (C.12) is quadratic in $\boldsymbol{w}$.

Alternatively to the two-steps procedure exposed above, a single optimization program can be run to estimate both sets of parameters $\boldsymbol{\beta}$ and $\boldsymbol{w}$ in a self-consistent way: considering the system

$$\begin{cases} r_s = \beta_s \hat{r}^{(0)}(\boldsymbol{w}) + e_s, & \mathbb{V}[e_s] = 1 - \beta_s^2 \\ \hat{r}^{(0)}(\boldsymbol{w}) = \sum_s w_s r_s \\ \boldsymbol{w}^{\dagger}\mathfrak{C}\boldsymbol{w} = 1 \end{cases}$$

the joint estimator is

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{w}}) = \arg\min_{(\boldsymbol{\beta},\boldsymbol{w})} \left\{ \mathcal{L}(\boldsymbol{\beta},\boldsymbol{w}|\boldsymbol{r}) \,\Big|\, \boldsymbol{w}^{\dagger}\mathfrak{C}\boldsymbol{w} = 1 \right\}$$

with

$$\mathcal{L}(\boldsymbol{\beta},\boldsymbol{w}|\boldsymbol{r})^2 = \frac{1}{S(S-1)} \sum_{u<s} \left(\mathfrak{C}_{su} - \beta_s\beta_u - \frac{\beta_s w_u(1-\beta_u^2) + \beta_u w_s(1-\beta_s^2)}{1 - \sum_k \beta_k w_k} \right)^2$$

Nested one-factor model

We now consider a model where two levels of the previous structure are nested:

$$r_s^{(1)} = \beta_s r^{(0)} + e_s, \qquad \mathbb{V}[e_s] = 1 - \beta_s^2 \tag{C.13a}$$

$$r_i^{(2)} = \gamma_i r_s^{(1)} + e_i, \qquad \mathbb{V}[e_i] = 1 - \gamma_i^2, \qquad i \in s \tag{C.13b}$$

In each branch s there are N_s leaves, and the trunk/branch indices are constructed as linear combinations of their respective constituents:

$$\hat{r}^{(0)}(\boldsymbol{w}) = \sum_s w_s r_s^{(1)} \quad \text{and} \quad \hat{r}_s^{(1)}(\boldsymbol{z}_{[s]}) = \sum_{i\in s} z_i r_i^{(2)} \tag{C.14}$$

Estimation algorithms when all sectorial and global factors are unobservable

All the estimation algorithms start with a prior classification, typically Bloomberg sectors.

Upward estimation algorithm:

- In each sector s, apply the one-factor calibration procedure (either 2-steps or self-consistent)

  - compute the leaves' correlation: $\mathfrak{C}^{(2)}_{[s]} = \langle \boldsymbol{r}^{(2)}_{[s]} \boldsymbol{r}^{(2)\dagger}_{[s]} \rangle$
  - estimate $\boldsymbol{\gamma}$ and the weights $\boldsymbol{z}_{[s]}$
  - define the branch index $\hat{r}^{(1)}_s(\hat{\boldsymbol{z}}_{[s]})$ and compute the leaves' residuals $\boldsymbol{e}^{(2)}_{[s]} = \boldsymbol{r}^{(2)}_{[s]} - \hat{\boldsymbol{\gamma}}_{[s]} r^{(1)}_s, i \in s$

- Apply the one-factor calibration procedure to the root (either 2-steps or self-consistent)

  - compute the branches' correlation $\mathfrak{C}^{(1)} = \langle \boldsymbol{r}^{(1)} \boldsymbol{r}^{(1)\dagger} \rangle$
  - estimate $\boldsymbol{\beta}$ and the weights $\boldsymbol{w}$
  - define the root index $\hat{r}^{(0)}(\hat{\boldsymbol{w}})$ and compute the branches' residuals $\boldsymbol{e}^{(1)} = \boldsymbol{r}^{(1)} - \hat{\boldsymbol{\beta}} r^{(0)}$

Pluses: auto-similar recursive procedure (appropriate for many levels); quadratic minimization problems very efficient and quick. *Minuses*: biases are propagated (and possibly amplified) from leaves to root; trouble with very small sectors (1 to 3 individuals).

Integrated estimation algorithm:

- compute all leaves' correlations $\mathfrak{C}^{(2)} = \langle \boldsymbol{r}^{(2)} \boldsymbol{r}^{(2)\dagger} \rangle$

- the corresponding theoretical prediction is

$$Q_{ij} = \delta_{ij} + (1 - \delta_{ij})\gamma_i \gamma_j \left[\delta_{s_i s_j} + (1 - \delta_{s_i s_j})\beta_{s_i}\beta_{s_j} \right]$$

- estimate all exposure parameters jointly: $(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}}) = \arg\min \|\mathfrak{C}^{(2)} - Q\|$

- In each sector s,

  - estimate the weights $\boldsymbol{z}_{[s]}$
  - define the branch index $\hat{r}^{(1)}_s(\hat{\boldsymbol{z}})$ and compute the leaves' residuals $\boldsymbol{e}^{(2)}_{[s]} = \boldsymbol{r}^{(2)}_{[s]} - \hat{\boldsymbol{\gamma}}_{[s]} r^{(1)}_s, i \in s$

- compute the branches' correlation $\mathfrak{C}^{(1)} = \langle \boldsymbol{r}^{(1)} \boldsymbol{r}^{(1)\dagger} \rangle$ and check that $\left\| \mathfrak{C}^{(1)} - \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^{\dagger} \right\| \approx 0$

- At root level,

  - estimate the weights $\boldsymbol{w}$
  - define the root index $\hat{r}^{(0)}(\hat{\boldsymbol{w}})$ and compute the branches' residuals $\boldsymbol{e}^{(1)} = \boldsymbol{r}^{(1)} - \hat{\boldsymbol{\beta}} r^{(0)}$

Pluses: dissociation of the estimation of the exposure parameters and that of the weights; one single optimization program for all levels simultaneously; a posterior check of the correctness of intermediate correlations. *Minuses*: non-quadratic minimization with a large number of parameters !

Mixed estimation algorithm:

- compute all leaves' correlations $\mathfrak{C}^{(2)} = \langle \boldsymbol{r}^{(2)} \boldsymbol{r}^{(2)\dagger} \rangle$

- In each sector s, apply the one-factor calibration procedure (either 2-steps or self-consistent)

    - estimate $\boldsymbol{\gamma}_{[s]}$ and the weights $\boldsymbol{z}_{[s]}$, using intra-sector correlations: $\mathfrak{C}_{[s]}$
    - define the branch index $\hat{r}_s^{(1)}(\hat{\boldsymbol{z}}_{[s]})$ and compute the leaves' residuals $\boldsymbol{e}_{[s]}^{(2)} = \boldsymbol{r}_{[s]}^{(2)} - \hat{\boldsymbol{\gamma}}_{[s]} r_s^{(1)}, i \in s$

- Use inter-sector correlations to calibrate $\boldsymbol{\beta}$:

    - Semi-theoretical prediction: $Q_{ij} = \hat{\gamma}_i \hat{\gamma}_j \beta_{s_i} \beta_{s_j}$ for $s_i \neq s_j$
    - $\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} ||\mathfrak{C}^{(2)} - Q||$

- define the root index $\hat{r}^{(0)}(\hat{\boldsymbol{w}})$ and compute the branches' residuals $\boldsymbol{e}^{(1)} = \boldsymbol{r}^{(1)} - \hat{\boldsymbol{\beta}} r^{(0)}$

- compute the branches' correlation $\mathfrak{C}^{(1)} = \langle \boldsymbol{r}^{(1)} \boldsymbol{r}^{(1)\dagger} \rangle$ and check that $\left|\left| \mathfrak{C}^{(1)} - \hat{\boldsymbol{\beta}} \hat{\boldsymbol{\beta}}^{\dagger} \right|\right| \approx 0$

Pluses: only quadratic minimization problems; dissociation of the estimation of the exposure parameters and that of the weights; a posterior check of the correctness of intermediate correlations. *Minuses*: trouble with very small sectors (1 to 3 individuals).

Improving over the initial classification

Allocating each leave to its correct branch   The correlation of a leave variable $r_i^{(2)}$ with a sector residual $e_s^{(1)}$ is

$$\mathbb{E}[r_i^{(2)} e_s^{(1)}] = \begin{cases} \gamma_i \left(1 + \frac{\beta_s w_s}{1 - \sum_u \beta_u w_u} \right) (1 - \beta_s^2) & , i \in s \\ \gamma_i \frac{\beta_{s_i} w_s}{1 - \sum_u \beta_u w_u} & , i \notin s \end{cases} \tag{C.15a}$$

which means that, up to noise and the small bias introduced by endogenizing the factor variables, we should observe

$$\langle r_i^{(2)} e_s^{(1)} \rangle \approx \begin{cases} \gamma_i \left(1 - \beta_s^2 \right) & , i \in s \\ 0 & , i \notin s \end{cases} \tag{C.15b}$$

Significant departure from this prediction indicates a bad initial classification . . . and suggests the correct one.

Exposure to several factors   It happens that projecting the leave i onto the residuals of every branch s reveals several non trivial correlations, ruling out the pure 1-factor model and the corresponding prediction (C.15). In particular, two distinct factors (according to the initial classification) can be simultaneously influential over given variables. The relation (C.13b) needs to be corrected accordingly:

$$r_i^{(2)} = \gamma_{i;a} r_a^{(1)} + \gamma_{i;b} r_b^{(1)} + e_i, \qquad \mathbb{V}[e_i] = 1 - \gamma_{i;a}^2 - \gamma_{i;b}^2 - 2\gamma_{i;a}\gamma_{i;b}\beta_a\beta_b, \qquad i \in a \cap b$$

and the corresponding prediction in terms of correlations updated: for $i \in a \cap b, j \in s$,

$$\mathbb{E}[r_i^{(2)} r_j^{(2)}] = \gamma_{i;a} \gamma_j \left[\delta_{as} + (1 - \delta_{as})\beta_a \beta_s\right] + \gamma_{i;b} \gamma_j \left[\delta_{bs} + (1 - \delta_{bs})\beta_b \beta_s\right]$$

Another norm: overlap distance

Starting from a prior Q^0 obtained e.g. by the previous estimation method, we want to find the optimal corrections Q^1 that minimize the overlap distance of the eigenvectors Ψ of $Q = Q^0 + Q^1$ with the eigenvectors Φ of $\mathfrak{C}$. The M-overlap distance is

$$d_M(\mathfrak{C}, Q) = -\frac{1}{M} \ln |G_{M|}|$$

where $G = \Phi^\dagger \Psi$ and the subscript denotes the restriction to the first M columns and lines. Minimizing the distance over the parameters Q^1 necessitates, for a higher convergence rate, the gradient. We compute it explicitly to first order in perturbation theory,

$$|\psi_n\rangle = |\psi_n^0\rangle + \sum_{m \neq n} |\psi_m^0\rangle \frac{\langle \psi_m^0 | Q^1 | \psi_n^0 \rangle}{\lambda_n^0 - \lambda_m^0}$$

$$= |\psi_n^0\rangle + \Psi_{-n}^0 (\lambda_n^0 - \Lambda_{-n}^0)^{-1} \Psi_{-n}^{0\dagger} Q^1 |\psi_n^0\rangle, \text{ or}$$

$$G = \underbrace{\Phi^\dagger \Psi^0}_{G^0} + G^1$$

$$d_p(\mathfrak{C}, Q) = -\frac{1}{M} \ln |G_{M|}^0| - \text{Tr}\left[G_{M|}^1 (G_{M|}^0)^{-1}\right]$$

$$\partial d_p(\mathfrak{C}, Q) = -\text{Tr}\left[(\partial G_{M|}^1)(G_{M|}^0)^{-1}\right]$$

Pseudo-elliptical copula: expansion around independence

We compute here the spectrum and eigenvectors of the kernel $H(u,v)$ in the case of pseudo-elliptical copula with weak dependences, starting from the expansion (8.10) on page 159.

The situation is better understood in terms of operators acting in the Hilbert space of continuous functions on $[0,1]$ vanishing in the border. Using Dirac's bra-ket notations, $A = |\tilde{A}\rangle\langle\tilde{A}|$, $B = |\tilde{R}\rangle\langle\tilde{A}|$, $R = |\tilde{R}\rangle\langle\tilde{R}|$. The sine functions $|j\rangle = \sqrt{2}\sin(j\pi u)$ build a basis of this Hilbert space, and interestingly they are the eigenvectors of the independent kernel $I(u,v)$ (I stands for 'Independence' and is the covariance matrix of the Brownian bridge, see page 37: $I = M - \Pi$ where M denotes the bivariate upper Fréchet-Hoeffding copula and Π the bivariate product copula defined in page 18).

It is then easy to find the spectra: rank-one operators have at most one non-null eigenvalue. Using the parities of $\tilde{A}(u)$ and $\tilde{R}(u)$ with respect to $\frac{1}{2}$ and imposing orthonormality of the eigenvectors, we can sketch the following table of the non zero eigenvalues and eigenvectors of the different operators:

$$
\begin{aligned}
\lambda_j^I &= (j\pi)^{-2} & U_j^I(u) &= |j\rangle \\
\lambda^R &= \langle\tilde{R}|\tilde{R}\rangle = \operatorname{Tr} R & |U_0^R\rangle &= |\tilde{R}\rangle/\sqrt{\operatorname{Tr} R} \\
\lambda^A &= \langle\tilde{A}|\tilde{A}\rangle = \operatorname{Tr} A & |U_0^A\rangle &= |\tilde{A}\rangle/\sqrt{\operatorname{Tr} A}.
\end{aligned}
$$

For the pseudo-elliptical copula with weak dependence, H has the following general form:

$$
H = I + \tilde{\rho}R + \tilde{\alpha}A - \frac{\tilde{\beta}}{2}(B + B^\dagger). \tag{D.1}
$$

The operator $B + B^\dagger$ has two non zero eigenvalues $\pm\sqrt{\lambda^R\lambda^A}$, with eigenvectors $[|U_0^R\rangle \pm |U_0^A\rangle]/\sqrt{2}$. In order to approximately diagonalize H, it is useful to notice that in the present context A and R are close to commuting with I. More precisely, it turns out that $|U_0^A\rangle$ is very close to $|2\rangle$, and $|U_0^R\rangle$ even closer to $|1\rangle$. Indeed, $a_2 = \langle U_0^A|2\rangle \approx 0.9934$ and $r_1 = \langle U_0^R|1\rangle \approx 0.9998$. Using the symmetry of A and R, we can therefore write:

$$
\begin{aligned}
|U_0^A\rangle &= a_2|2\rangle + \epsilon_a|2_\perp\rangle & \text{with} \quad \langle 2|2_\perp\rangle = \langle 2j-1|2_\perp\rangle = 0, \forall j \geq 1 \\
|U_0^R\rangle &= r_1|1\rangle + \epsilon_r|1_\perp\rangle & \text{with} \quad \langle 1|1_\perp\rangle = \langle 2j|1_\perp\rangle = 0, \forall j \geq 1
\end{aligned}
$$

where $\epsilon_a = \sqrt{1 - a_2^2} \ll 1$ and $\epsilon_r = \sqrt{1 - r_1^2} \ll 1$. The components of $|2_\perp\rangle$ on the even eigenvectors of I are determined as the projection:

$$
\langle 2_\perp|2j\rangle = \frac{\langle U_0^A|2j\rangle}{\epsilon_a} \qquad j \geq 2,
$$

I	A	R	I^2	IA	IR
16.667	1.176	7.806	111.139	2.948	79.067

Table D.1: Traces of the operators appearing in the covariance functions (multiples of 10^{-2}). Traces of the powers of the rank-one operators A and R equal powers of their traces. The trace of $B + B^\dagger$ is zero.

and similarly:

$$\langle 1_\perp | 2j-1 \rangle = \frac{\langle U_0^R | 2j-1 \rangle}{\epsilon_r} \qquad j \geq 2.$$

Using the definition of the coefficients α_t, β_t and ρ_t given in section 8.1 of chapter III.8 (page 159), we introduce the following notations:

$$\tilde{\alpha} = 2 \operatorname{Tr} A \lim_{N \to \infty} \sum_{t=1}^{N-1} \left(1 - \frac{t}{N}\right) \alpha_t$$

$$\tilde{\rho} = 2 \operatorname{Tr} R \lim_{N \to \infty} \sum_{t=1}^{N-1} \left(1 - \frac{t}{N}\right) \rho_t$$

$$\tilde{\beta} = 2 \sqrt{\operatorname{Tr} A \operatorname{Tr} R} \lim_{N \to \infty} \sum_{t=1}^{N-1} \left(1 - \frac{t}{N}\right) \beta_t$$

so that H writes:

$$
\begin{aligned}
H =\ & I + \tilde{\alpha} \, |U_0^A\rangle\langle U_0^A| + \tilde{\rho} \, |U_0^R\rangle\langle U_0^R| - \tilde{\beta} \, \overleftrightarrow{|U_0^R\rangle\langle U_0^A|} \\
=\ & H_0 + \epsilon_a \left(\tilde{\alpha} \, a_2 \, \overleftrightarrow{|2\rangle\langle 2_\perp|} - \tilde{\beta} \, r_1 \, a_\perp \overleftrightarrow{|1\rangle\langle 2_\perp|} \right) \\
& + \epsilon_r \left(\tilde{\rho} \, r_1 \, \overleftrightarrow{|1_\perp\rangle\langle 1|} - \tilde{\beta} \, a_2 \, \overleftrightarrow{|1_\perp\rangle\langle 2|} \right) \\
& + \left(\tilde{\alpha} \, \epsilon_a^2 \, |2_\perp\rangle\langle 2_\perp| + \tilde{\rho} \, \epsilon_r^2 \, |1_\perp\rangle\langle 1_\perp| - \tilde{\beta} \, \epsilon_a \, \epsilon_r \, \overleftrightarrow{|2_\perp\rangle\langle 1_\perp|} \right),
\end{aligned}
$$

where $\overleftrightarrow{|\psi_1\rangle\langle\psi_2|} = \frac{1}{2}\left[|\psi_1\rangle\langle\psi_2| + |\psi_2\rangle\langle\psi_1|\right]$ and H_0 is the unperturbed operator (0-th order in both ϵs)

$$H_0 = \sum_{j \geq 3} \lambda_j^I \, |j\rangle\langle j| + (\lambda_2^I + \tilde{\alpha} \, a_2^2) \, |2\rangle\langle 2| + (\lambda_1^I + \tilde{\rho} \, r_1^2) \, |1\rangle\langle 1| - \tilde{\beta} \, r_1 \, a_2 \, \overleftrightarrow{|2\rangle\langle 1|}.$$

The spectrum of the latter is easy to determine as:

$$\lambda_1^{H_0} = \lambda_- \xrightarrow{\tilde{\rho}, \tilde{\beta} \to 0} \lambda_1^I \qquad\qquad |U_1^{H_0}\rangle = -\frac{|-\rangle}{\sqrt{\langle -|-\rangle}} \xrightarrow{\tilde{\rho}, \tilde{\beta} \to 0} |1\rangle$$

$$\lambda_2^{H_0} = \lambda_+ \xrightarrow{\tilde{\rho}, \tilde{\beta} \to 0} \lambda_2^I + \tilde{\alpha} a_2^2 \qquad |U_2^{H_0}\rangle = \frac{|+\rangle}{\sqrt{\langle +|+\rangle}} \xrightarrow{\tilde{\rho}, \tilde{\beta} \to 0} |2\rangle$$

$$\lambda_j^{H_0} = \lambda_j^I \qquad\qquad\qquad |U_j^{H_0}\rangle = |j\rangle \qquad (j \geq 3)$$

where

$$\lambda_\pm = \frac{\lambda_1^I + \tilde{\rho} \, r_1^2 + \lambda_2^I + \tilde{\alpha} \, a_2^2 \pm \sqrt{(\lambda_1^I + \tilde{\rho} \, r_1^2 - \lambda_2^I - \tilde{\alpha} \, a_2^2)^2 + 4 \, (\tilde{\beta} \, r_1 \, a_2)^2}}{2}$$

and $|\pm\rangle$ the corresponding eigenvectors, which are linear combination of $|1\rangle$ and $|2\rangle$ only. Therefore, $\langle 1_\perp|\pm\rangle = \langle 2_\perp|\pm\rangle = 0$, which implies that there is no corrections to the eigenvalues of H_0 to first order in the ϵs. At the next order, instead, some corrections appear. We call:

$$
\begin{aligned}
V_{i,j} &= \left[\tilde\rho\, r_1\,\langle 1|U_i^{H_0}\rangle - \tfrac{1}{2}\tilde\beta\, a_2\langle 2|U_i^{H_0}\rangle\right]\langle j|1_\perp\rangle\,\epsilon_r \\
&\quad + \left[\tilde\alpha\, a_2\,\langle 2|U_i^{H_0}\rangle - \tfrac{1}{2}\tilde\beta\, r_1\langle 1|U_i^{H_0}\rangle\right]\langle j|2_\perp\rangle\,\epsilon_a
\end{aligned}
$$

the matrix elements of the first order perturbation of H, whence

$$
\lambda_1^H = \lambda_1^{H_0} + \sum_{j\geq 3}\frac{V_{1,j}^2}{\lambda_1^{H_0} - \lambda_j^{H_0}}
$$

$$
\lambda_2^H = \lambda_2^{H_0} + \sum_{j\geq 3}\frac{V_{2,j}^2}{\lambda_2^{H_0} - \lambda_j^{H_0}}
$$

$$
\lambda_j^H = \lambda_j^{H_0} + \sum_{i=1,2}\frac{V_{i,j}^2}{\lambda_j^{H_0} - \lambda_i^{H_0}}
$$

$$
+ \tilde\alpha\,\epsilon_a^2\,\langle j|2_\perp\rangle^2 + \tilde\rho\,\epsilon_r^2\,\langle j|1_\perp\rangle^2 - \tilde\beta\,\epsilon_a\,\epsilon_r\,\langle j|1_\perp\rangle\langle j|2_\perp\rangle.
$$

As of the eigenvectors, it is enough to go to first order in ϵs to get a non-trivial perturbative correction:

$$
|U_1^H\rangle = |U_1^{H_0}\rangle + \sum_{j\geq 3}\frac{V_{1,j}}{\lambda_1^{H_0} - \lambda_j^{H_0}}|j\rangle
$$

$$
|U_2^H\rangle = |U_2^{H_0}\rangle + \sum_{j\geq 3}\frac{V_{2,j}}{\lambda_2^{H_0} - \lambda_j^{H_0}}|j\rangle
$$

$$
|U_j^H\rangle = |j\rangle + \sum_{i=1,2}\frac{V_{i,j}}{\lambda_j^{H_0} - \lambda_i^{H_0}}|U_i^{H_0}\rangle.
$$

The special case treated numerically in chapter III.8 corresponds to $\tilde\rho = \tilde\beta = 0$, such that the above expressions simplify considerably, since in that case $V_{1,j} \equiv 0$ and $V_{2,2j-1} = 0$, while $V_{2,2j} = \tilde\alpha\, a_2\langle U_0^A|2j\rangle$. To first order in the ϵs, the spectrum is not perturbed and calls $\lambda_i^H = \lambda_i^{H_0} = \lambda_i^I + \tilde\alpha\, a_2^2\,\delta_{i2}$, so that the characteristic function of the modified CM distribution is, according to Eq. (3.25),

$$
\phi(t) = \prod_j\left(1 - 2\mathrm{i}t/(j\pi)^2\right)^{-\frac{1}{2}} \times \sqrt{\frac{1 - 2\mathrm{i}t\lambda_2^I}{1 - 2\mathrm{i}t\lambda_2^{H_0}}}.
$$

Its pdf is thus the convolution of the Fourier transform of $\phi_I(t)$ (characteristic function associated to the usual CM distribution $\mathcal{P}_I$ [Anderson 1952]) and the Fourier transform (FT) of the correction $\phi_c(t) = \sqrt{1 - 2\mathrm{i}t\lambda_2^I}/\sqrt{1 - 2\mathrm{i}t\lambda_2^{H_0}}$. Noting that $(1 - 2\mathrm{i}\sigma^2 t)^{-\frac{1}{2}}$ is the characteristic function of the chi-2 distribution, it can be shown that for $k > 0$, and with $\mu \equiv \lambda_2^{H_0}$

for the sake of readability:

$$\frac{1}{\sqrt{2\pi}}\mathrm{FT}[\phi_c] = \delta(k) - \int_{\lambda_2^I}^{\mu} \mathrm{d}\lambda \frac{\partial}{\partial k}\left(\chi^2(k;\mu) * \chi^2(k;\lambda)\right)$$

$$= \delta(k) - \int_{\lambda_2^I}^{\mu} \mathrm{d}\lambda \frac{\mathrm{e}^{-\frac{\lambda+\mu}{4\lambda\mu}k}}{8(\lambda\mu)^{\frac{3}{2}}}\left((\mu - \lambda)\,I_1(\frac{\mu - \lambda}{4\lambda\mu}k) - (\mu + \lambda)\,I_0(\frac{\mu - \lambda}{4\lambda\mu}k)\right)$$

$$\approx \delta(k) + \mathrm{e}^{-k/2\lambda}\frac{\tilde{\alpha}a_2^2}{4\lambda^2}\,I_0(\frac{\tilde{\alpha}a_2^2}{4\lambda^2}k),$$

where $\chi^2(k;\sigma^2) = (2\pi\sigma^2 k\,\mathrm{e}^{k/\sigma^2})^{-\frac{1}{2}}$ is the pdf of the chi-2 distribution, I_n are the modified Bessel functions of the first kind, and $*$ denotes the convolution operation. The approximation on the last line holds as long as $\tilde{\alpha} \ll \lambda_2^I = (2\pi)^{-2}$ and in this regime we obtain finally the distribution of the CM statistic with asymptotic kernel H, $\mathcal{P}_H(k) = \mathrm{d}\mathbb{P}[CM \le k]/\mathrm{d}k$:

$$\mathcal{P}_H(k) = \sqrt{2\pi}\,\mathrm{FT}[\phi](k) = (\mathrm{FT}[\phi_I] * \mathrm{FT}[\phi_c])(k)$$

$$= \mathcal{P}_I(k) + 4\tilde{\alpha}a_2^2\pi^4 \int_0^k \mathcal{P}_I(z)\mathrm{e}^{-2\pi^2(k-z)}I_0(4\tilde{\alpha}a_2^2\pi^4(k-z))\,\mathrm{d}z$$

$$= \mathcal{P}_I(k) + 4\tilde{\alpha}a_2^2\pi^4 \int_0^k \mathcal{P}_I(k-z)\mathrm{e}^{-2\pi^2 z}I_0(4\tilde{\alpha}a_2^2\pi^4 z)\,\mathrm{d}z.$$

The second term characterizes the perturbative correction due to the departure of H_0 from the independent kernel I.

Appendices to Chapter III.6

E.1 The exact spectrum of the Borland-Bouchaud model

In this appendix to chapter 6 of part III, we compute the spectral properties of a continuous time kernel with the BB structure of Ref. [Borland 2011]:

$$K(\tau', \tau'') = k(\max(\tau', \tau'')), \qquad \varepsilon \leq \tau', \tau'' \leq q.$$

The eigenvalue equation

$$\int_\varepsilon^q K(\tau, \tau')v(\tau')\,\mathrm{d}\tau' = \lambda v(\tau)$$

is differentiated to obtain a second order linear differential equation with appropriate boundary conditions:

$$\begin{cases} \lambda V''(\tau) &= k'(\tau)V(\tau) \\ \lambda V'_\lambda(q) &= k(q)V_\lambda(q) \\ V(\varepsilon) &= 0 \end{cases} \tag{E.1}$$

where $V(q) \equiv \int_\varepsilon^q v(\tau)\mathrm{d}\tau$. The resolution of this differential problem depends on the choice of $k(\tau)$, and we investigate below three particular cases: a linearly, exponentially and power-law decreasing kernel.

For the problem in continuous time, ε can be taken as 0, but in case this differential problem is seen as an approximation to a discrete time problem, it is important to keep $\varepsilon = 1$. For the sake of simplicity of the solutions, we set $\varepsilon = 0$ in the following, but more precise numerical results for the first eigenmodes are obtained with $\varepsilon = 1$ (see Fig. E.1), although for higher modes and at large q the choice of ε is hardly relevant.

Linearly decreasing

When $k(\tau) = g \cdot (1 - \tau/q)\mathbb{1}_{\{\tau \leq q\}}$, the general solution is a superposition of the two linearly independent solutions:

$$V^s_\lambda(\tau) = \sin\left(\sqrt{\frac{gq}{\lambda}}\frac{\tau}{q}\right) \quad \text{and} \quad V^c_\lambda(\tau) = \cos\left(\sqrt{\frac{gq}{\lambda}}\frac{\tau}{q}\right).$$

The boundary condition $V(0) = 0$ disqualifies V^c, and $V'(q) = 0$ selects only those values of λ which satisfy

$$\lambda_n = \frac{gq}{\pi^2} \cdot \frac{1}{(n-\frac{1}{2})^2}, \quad n \in \mathbb{N}\backslash\{0\},$$

so that finally

$$v_n(\tau) = V'_{\lambda_n}(\tau) \propto \cos\left((n-\tfrac{1}{2})\pi\frac{\tau}{q}\right).$$

In fact, we see straight from Eq. (E.1) that when k' is constant, the problem for V amounts to that of a free quantum particle in a box with absorbing left wall and reflecting right wall.

Exponentially decreasing

When $k(\tau) = g\,e^{-\alpha\tau}$, the general solution is a superposition of the stretched Bessel functions

$$Z_0\Big(\pm\gamma\,e^{-\alpha\tau/2}\Big)\,,$$

where $\gamma = 2\sqrt{g(\lambda\alpha)}$. Bessel functions with negative argument are complex, so we keep only $+\gamma > 0$. The Bessel functions J_ν and Y_ν of the first and second kind respectively, are linearly independent, but the lower boundary condition imposes the coefficients of the combination:

$$V(\tau) = Y_0(\gamma)\,J_0\Big(\gamma\,e^{-\alpha\tau/2}\Big) - J_0(\gamma)\,Y_0\Big(\gamma\,e^{-\alpha\tau/2}\Big)\,.$$

Using recursion formulas for Bessel functions [Gradshteyn 1980], the upper boundary condition becomes

$$Y_0(\gamma_n)\,J_2\Big(\gamma_n e^{-\alpha q/2}\Big) = J_0(\gamma_n)\,Y_2\Big(\gamma_n e^{-\alpha q/2}\Big)\,,$$

and the eigenvalues λ_n are quantized according to the corresponding zeros.

Power-law decreasing

Taking $k(\tau) = g\,\tau^{-\alpha}$, the solutions are given in terms of rescaled Bessel functions Z_ν (formula 8.491.12 in [Gradshteyn 1980]):

$$V^\pm(\tau) = \tau^{\frac{1}{2}}\,Z_\nu\Big(\pm\gamma\tau^{\frac{1-\alpha}{2}}\Big)\,, \tag{E.2}$$

with $\gamma = \frac{2}{|1-\alpha|}\sqrt{g\alpha/\lambda}$ and $\nu = 1/(1-\alpha)$. Bessel functions with negative argument are complex, so we keep only $+\gamma > 0$. In the cases we consider, $\alpha > 1$ (see discussion in page 107) so that $\nu < 0$. Although the Bessel function of the first kind $J_\nu(x)$ with negative non-integer ν diverges at the origin (as $\sim x^\nu$), $V(\tau)$ vanishes linearly in 0^+ since $V(\tau) \to \sqrt{\tau}(\tau^{\frac{1}{2\nu}})^\nu = \tau$. The first boundary condition is thus satisfied for $Z_\nu = J_\nu$. Using recursion formulas for Bessel functions, the boundary condition becomes

$$J_{\nu-2}\Big(\gamma_n q^{\frac{1-\alpha}{2}}\Big) = 0,$$

and the eigenvalues λ_n are quantized according to the corresponding zeros.

(a) Linear decay ($g = 0.1$)

(b) Exponential decay ($g = 0.1, \alpha = 0.15$)

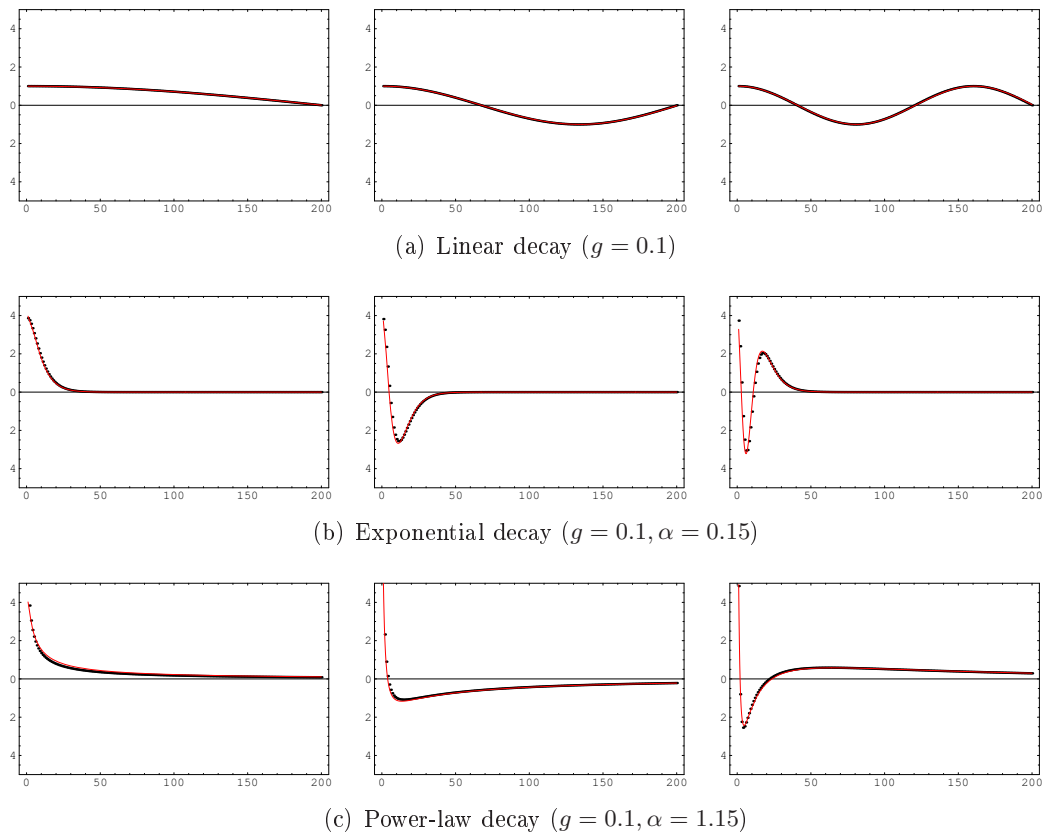(c) Power-law decay ($g = 0.1, \alpha = 1.15$)

Figure E.1: First three eigenvectors of the BB Kernel ($q = 200$), from left to right. The theoretical solutions (red line) are hardly distinguishable from the results of the numerical diagonalization (black dots).

E.2 Power-law volatility correlations in FIGARCH

In Chapter 6 of Part III, we studied theoretical and empirical properties of Quadratic ARCH models. In particular, 2nd and 4th order stationarity conditions were summarized, for the power-law decreasing diagonal case, in the parameter space of Fig. 6.1, page 108, by numerically solving Eq. (6.11).

We want here to get a more precise insight on the behavior of the quadratic correlation $\mathcal{C}^{(2)}(\tau)$ when the input kernel $k(\tau)$ is long-ranged asymptotically power-law:

$$k(\tau) \xrightarrow{\tau \to \infty} A/\tau^{1+\epsilon}, \quad 0 < \epsilon,$$

where the bound on ϵ ensures the integrability of the kernel $\int_0^\infty k(\tau)\mathrm{d}\tau = 1 - s^2$.

In order to address this question analytically, we assume that the feedback kernel is infinitely-ranged, and consider the continuous-time approximation. The sum in Eq. (6.11) is approximated by an integral, and decomposed as follows

$$\mathcal{C}^{(2)}(\tau) = \mathcal{C}^{(2)}(-\tau) = \int_0^\infty k(\tau')\mathcal{C}^{(2)}(\tau - \tau')\,\mathrm{d}\tau'$$

$$= \underbrace{\int_0^\tau k(\tau')\mathcal{C}^{(2)}(\tau - \tau')\,\mathrm{d}\tau'}_{\mathcal{C}_-^{(2)}(\tau)} + \underbrace{\int_\tau^\infty k(\tau')\mathcal{C}^{(2)}(\tau' - \tau)\,\mathrm{d}\tau'}_{\mathcal{C}_+^{(2)}(\tau)}$$

The behavior at large $\tau \to \infty$ is studied by taking the Laplace transform and investigating $\omega \to 0$ while keeping a non-diverging product $\omega\tau$:

$$\widehat{k}(\omega) = \int_0^\infty 1 - \left(1 - \mathrm{e}^{-\omega\tau}\right)k(\tau)\,\mathrm{d}\tau = \widehat{k}(0) - \int_0^\infty \left(1 - \mathrm{e}^{-x}\right)k(x/\omega)\,\mathrm{d}x/\omega$$

$$= \widehat{k}(0) - \int_0^\infty \left(1 - \mathrm{e}^{-x}\right)A\left(\frac{\omega}{x}\right)^{1+\epsilon}\mathrm{d}x/\omega$$

$$= \widehat{k}(0) - A\omega^\epsilon \int_0^\infty \left(1 - \mathrm{e}^{-x}\right)x^{-1-\epsilon}\,\mathrm{d}x = \widehat{k}(0) + A\omega^\epsilon \int_0^\infty \mathrm{e}^{-x}\frac{x^{-\epsilon}}{-\epsilon}\,\mathrm{d}x$$

$$= \widehat{k}(0) + A\omega^\epsilon \frac{\Gamma(1 - \epsilon)}{-\epsilon} = \widehat{k}(0) + A\,\Gamma(-\epsilon)\omega^\epsilon$$

Power-law resulting correlation

Empirical observations motivate the following Ansatz for the quadratic correlation:

$$\mathcal{C}^{(2)}(\tau) \xrightarrow{\tau \to \infty} B/\tau^{\beta}, \quad 0 < \beta < 1.$$

We hope to be able to reconciliate the "fast" decay of k (since integrable) with a very slow asymptotic decay of the solution, by finding a relationship between β and ϵ. In Laplace space, we have

$$\widehat{\mathcal{C}^{(2)}}(\omega) = \int_0^{\infty} e^{-\omega\tau}\, \mathcal{C}^{(2)}(\tau)\, \mathrm{d}\tau = \int_0^{\infty} \mathcal{C}^{(2)}(x/\omega)\, e^{-x}\, \mathrm{d}x/\omega$$

$$\xrightarrow{\omega \to 0} \int_0^{\infty} B\left(\frac{\omega}{x}\right)^{\beta} e^{-x}\frac{\mathrm{d}x}{\omega} = B\,\Gamma(1-\beta)\,\omega^{\beta-1}$$

$$\widehat{\mathcal{C}_-^{(2)}}(\omega) = \widehat{k}(\omega)\,\widehat{\mathcal{C}^{(2)}}(\omega)$$

$$\xrightarrow{\omega \to 0} B\,\Gamma(1-\beta)\,\omega^{\beta-1}\left[(1-s^2) - A\frac{\Gamma(1-\epsilon)}{\epsilon}\omega^{\epsilon}\right]$$

$$\mathcal{C}_+^{(2)}(\tau) = \int_0^{\infty} k(\tau+u)\,\mathcal{C}^{(2)}(u)\,\mathrm{d}u \qquad \text{dominated by the large } u \sim \tau$$

$$\xrightarrow{\tau \to \infty} \int_0^{\infty} \frac{B}{u^{\beta}}\frac{A}{(\tau+u)^{1+\epsilon}}\,\mathrm{d}u = \tau^{-\beta-(1+\epsilon)+1}\int_0^{\infty}\frac{B}{x^{\beta}}\frac{A}{(1+x)^{1+\epsilon}}\mathrm{d}x$$

$$= \tau^{-(\beta+\epsilon)}AB\int_0^1\left(\frac{y}{1-y}\right)^{\beta}y^{1+\epsilon}\frac{\mathrm{d}y}{y^2} = \tau^{-(\beta+\epsilon)}AB\frac{\Gamma(\epsilon+\beta)\,\Gamma(1-\beta)}{\Gamma(1+\epsilon)}$$

$$\widehat{\mathcal{C}_+^{(2)}}(\omega) = \int_0^{\infty} e^{-\omega\tau}\,\mathcal{C}_+^{(2)}(\tau)\,\mathrm{d}\tau = \int_0^{\infty} \mathcal{C}_+^{(2)}(x/\omega)\, e^{-x}\, \mathrm{d}x/\omega$$

$$\xrightarrow{\omega \to 0} AB\frac{\Gamma(\epsilon+\beta)\,\Gamma(1-\beta)}{\epsilon\Gamma(\epsilon)}\,\omega^{\epsilon+\beta-1}\,\Gamma(1-\epsilon-\beta)$$

Collecting all the terms, we finally get

$$B\,\Gamma(1-\beta)\,\omega^{\beta-1} = B\,\Gamma(1-\beta)\,\omega^{\beta-1}\left[(1-s^2) - A\frac{\Gamma(1-\epsilon)}{\epsilon}\omega^{\epsilon}\right]$$

$$+ AB\,\frac{\Gamma(\epsilon+\beta)\,\Gamma(1-\beta)}{\epsilon\Gamma(\epsilon)}\,\omega^{\epsilon+\beta-1}\,\Gamma(1-\epsilon-\beta)$$

$$s^2 B\,\Gamma(1-\beta)\,\omega^{\beta-1} = AB\,\omega^{\epsilon+\beta-1}\,\Gamma(1-\beta)\left[\frac{\Gamma(\epsilon+\beta)\,\Gamma(1-\epsilon-\beta)}{\epsilon\,\Gamma(\epsilon)} - \frac{\Gamma(1-\epsilon)}{\epsilon}\right]$$

Very surprisingly, the functional dependence is perfectly equalized in the limit $s^2 \to 0$: the different terms have compatible powers of ω. For the relationship to be an equality, it is necessary that $s^2 \to 0$ (i.e. the model is at the critical limit of quadratic non-stationarity), and simultaneously the RHS term vanishes:

$$\Gamma(\epsilon+\beta)\,\Gamma(1-\epsilon-\beta) = \Gamma(1-\epsilon)\,\Gamma(\epsilon)$$

$$\sin\left(\pi(\epsilon+\beta)\right) = \sin\left(\pi\epsilon\right)$$

$$(\epsilon+\beta) - \epsilon = 2n \quad \text{ou} \quad (\epsilon+\beta) + \epsilon = 1 + 2n, \qquad n \in \mathbb{Z}$$

For $\widehat{k}(0) = 1 - s^2 \to 1$ to hold, ϵ must be close to 0, and in this case there is only one solution

$$\boxed{\beta = 1 - 2\epsilon, \qquad 0 < \epsilon < \tfrac{1}{2}}.$$

Exponential resulting correlation

We now check whether there might exist a crossover to an exponentially vanishing asymptotic correlation:

$$\mathcal{C}^{(2)}(\tau) \xrightarrow{\tau \to \infty} B\,e^{-\beta\tau}, \quad 0 < \beta.$$

Again, the Laplace transform are computed asymptotically:

$$\widehat{\mathcal{C}^{(2)}}(\omega) = \frac{B}{\omega + \beta} \xrightarrow{\omega \to 0} \frac{B}{\beta}\left(1 - \frac{\omega}{\beta}\right)$$

$$\widehat{\mathcal{C}_{-}^{(2)}}(\omega) \xrightarrow{\omega \to 0} \frac{B}{\omega + \beta}\left[(1 - s^2) + A\,\Gamma(-\epsilon)\omega^\epsilon\right]$$

$$\mathcal{C}_{+}^{(2)}(\tau) = \int_0^\infty k(\tau + u)\,\mathcal{C}^{(2)}(u)\,\mathrm{d}u \qquad \text{dominated by the large } u \sim \tau$$

$$\xrightarrow{\tau \to \infty} \int_0^\infty B\,e^{-\beta u}\frac{A}{(\tau + u)^{1+\epsilon}}\,\mathrm{d}u = \tau^{-(1+\epsilon)+1}\int_0^\infty B\,e^{-\beta\tau u}\frac{A}{(1 + x)^{1+\epsilon}}\,\mathrm{d}x$$

$$= \tau^{-\epsilon}AB\,e^{\beta\tau}\int_1^\infty y^{-1-\epsilon}\,e^{-\beta\tau y}\,\mathrm{d}y$$

$$\widehat{\mathcal{C}_{+}^{(2)}}(\omega) = \int_0^\infty e^{-\omega\tau}\mathcal{C}_{+}^{(2)}(\tau)\,\mathrm{d}\tau = \int_0^\infty \mathcal{C}_{+}^{(2)}(x/\omega)e^{-x}\,\mathrm{d}x/\omega$$

$$\xrightarrow{\omega \to 0} \omega^{-1}\int_0^\infty \mathrm{d}x\,e^{-x}\left(\frac{x}{\omega}\right)^{-\epsilon}AB\int_1^\infty \mathrm{d}y\,y^{-1-\epsilon}e^{-\beta\frac{x}{\omega}y}$$

$$= \omega^{\epsilon-1}AB\int_1^\infty \mathrm{d}y\,y^{-1-\epsilon}\int_0^\infty \mathrm{d}x\,x^{-\epsilon}e^{-x(1+\frac{\beta}{\omega}y)}$$

$$= \omega^{\epsilon-1}AB\int_1^\infty \mathrm{d}y\,y^{-1-\epsilon}\left(1 + \beta\frac{y}{\omega}\right)^{-(1-\epsilon)}\Gamma(1 - \epsilon), \quad 1 - \epsilon > 0$$

$$= AB\,\Gamma(1 - \epsilon)\int_1^\infty \mathrm{d}y\,y^{-1-\epsilon}(\omega + \beta y)^{\epsilon-1}, \quad 1 - \epsilon > 0$$

$$= AB\,\Gamma(1 - \epsilon)\,\beta^\epsilon\int_\beta^\infty \mathrm{d}z\,\frac{(\omega + z)^{\epsilon-1}}{z^{\epsilon+1}}$$

$$= AB\,\Gamma(1 - \epsilon)\frac{(\beta + \omega)^\epsilon - \beta^\epsilon}{\omega\epsilon}$$

Bringing everything together,

$$\frac{\omega^\epsilon}{\omega + \beta} = \frac{(\omega + \beta)^\epsilon - \beta^\epsilon}{\omega},$$

which only works if $\beta = 0$.

Bibliography

[Aalen 2004] Odd O. Aalen and Håkon K. Gjessing. *Survival models based on the Ornstein-Uhlenbeck process*. Lifetime Data Analysis, vol. 10, no. 4, pages 407–423, 2004. (Cited on page 43.)

[Abergel 2013] Frédéric Abergel, Bikas K. Chakrabarti, Anirban Chakraborti and Asim Ghosh, editeurs. Econophysics of systemic risk and network dynamics. New Economic Windows. Springer, 2013. (Cited on pages 2 and 5.)

[Adler 1986] Robert J. Adler and Lawrence D. Brown. *Tail Behaviour for Suprema of Empirical Processes*. The Annals of Probability, vol. 14, no. 1, pages 1–30, 1986. (Cited on page 182.)

[Adler 1987] Robert J. Adler and Gennady Samorodnitsky. *Tail behaviour for the suprema of Gaussian processes with applications to empirical processes*. The Annals of Probability, vol. 15, no. 4, pages 1339–1351, 1987. (Cited on page 182.)

[Adler 1990] Robert J. Adler. An introduction to continuity, extrema, and related topics for general Gaussian processes. IMS Lecture Notes-Monograph Series, 1990. (Cited on page 182.)

[Ahlgren 2007] Peter T. H. Ahlgren, Mogens H. Jensen, Ingve Simonsen, Raul Donangelo and Kim Sneppen. *Frustration driven stock market dynamics: Leverage effect and asymmetry*. Physica A: Statistical Mechanics and its Applications, vol. 383, no. 1, pages 1–4, 2007. (Cited on page 160.)

[Ait-Sahalia 2009] Yacine Ait-Sahalia and Jean Jacod. *Analyzing the Spectrum of Asset Returns: Jump and Volatility Components in High Frequency Data*. to appear in The Journal of Economic Literature, 2009. (Cited on page 106.)

[Akaike 1974] Hirotugu Akaike. *A new look at the statistical model identification*. IEEE Transactions on Automatic Control, vol. 19, no. 6, pages 716–723, 1974. (Cited on page 121.)

[Aldous 1989] David J. Aldous. Probability approximations via the Poisson clumping heuristic. Springer Verlag, New York, 1989. (Cited on page 182.)

[Allez 2011] Romain Allez and Jean-Philippe Bouchaud. *Individual and collective stock dynamics: intra-day seasonalities*. New Journal of Physics, vol. 13, no. 2, page 025010, 2011. (Cited on page 72.)

[Altmann 2005] Eduardo G. Altmann and Holger Kantz. *Recurrence time analysis, long-term correlations, and extreme events*. Physical Review E, vol. 71, no. 5, page 056106, 2005. (Cited on page 139.)

[Anderson 1952] Theodore W. Anderson and Donald A. Darling. *Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes*. The Annals of Mathematical Statistics, vol. 23, no. 2, pages 193–212, 1952. (Cited on pages 36, 37, 39, 40, 41, 52 and 203.)

[Anderson 1988] Philip W. Anderson, Kenneth Arrow and David Pines, editeurs. The economy as an evolving complex system, volume 5. Santa Fe Institute, Westview Press, 1988. studies in the sciences of complexity. (Cited on pages 2 and 5.)

[Arthur 1997] W. Brian Arthur, Steven N. Durlauf and David A. Lane, editeurs. The economy as an evolving complex system II, volume 27. Santa Fe Institute, Addison-Wesley, 1997. studies in the sciences of complexity. (Cited on pages 2 and 5.)

[Avramov 2006] Doron Avramov, Tarun Chordia and Amit Goyal. *Liquidity and autocorrelations in individual stock returns*. The Journal of Finance, vol. 61, no. 5, pages 2365–2394, 2006. (Cited on page 160.)

[Bacry 2006] Emmanuel Bacry, Alexey Kozhemyak and Jean-François Muzy. *Are asset return tail estimations related to volatility long-range correlations?* Physica A: Statistical Mechanics and its Applications, vol. 370, no. 1, pages 119–126, 2006. (Cited on page 166.)

[Bacry 2012] Emmanuel Bacry, Khalil Dayri and Jean-François Muzy. *Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data*. The European Physical Journal B, vol. 85, pages 1–12, 2012. (Cited on pages 131, 132, 144 and 173.)

[Bartz 2012] Daniel Bartz, Kerr Hatrick, Christian W. Hesse, Klaus-Robert Müller and Steven Lemm. *Directional Variance Adjustment: improving covariance estimates for high-dimensional portfolio optimization*. arXiv preprint q-fin.ST/1109.3069, 2012. (Cited on page 70.)

[Basu 2002] Sankarshan Basu and Angelos Dassios. *A Cox process with log-normal intensity*. Insurance: Mathematics and Economics, vol. 31, no. 2, pages 297 – 302, 2002. (Cited on page 144.)

[Beare 2010] Brendan K. Beare. *Copulas and temporal dependence*. Econometrica, vol. 78, no. 1, pages 395–410, 2010. (Cited on pages 51 and 134.)

[Bekaert 2000] Geert Bekaert and Guojun Wu. *Asymmetric volatility and risk in equity markets*. Review of Financial Studies, vol. 13, no. 1, pages 1–42, 2000. (Cited on page 105.)

[Berd 2011] Arthur Berd, editeur. Lessons from the 2008 crisis. Risk Publications, 2011. (Cited on page 213.)

[Black 1976] F. Black. *Studies of stock price volatility changes*. In Proceedings of the 1976 American Statistical Association, Business and Economical Statistics Section, page 177, 1976. (Cited on page 105.)

[Blanc 2013] Pierre Blanc, Rémy Chicheportiche and Jean-Philippe Bouchaud. In preparation. In preparation, 2013. (Cited on pages v, 132 and 173.)

[Blomqvist 1950] Nils Blomqvist. *On a measure of dependence between two random variables.* The Annals of Mathematical Statistics, vol. 21, no. 4, pages 593–600, 1950. (Cited on page 161.)

[Blume 2006] Lawrence E. Blume and Steven N. Durlauf, editeurs. The economy as an evolving complex system III: Current perspectives and future directions. Santa Fe Institute, Oxford University Press, 2006. studies in the sciences of complexity. (Cited on pages 2 and 5.)

[Boguná 2004] Marián Boguná and Jaume Masoliver. *Conditional dynamics driving financial markets.* The European Physical Journal B - Condensed Matter and Complex Systems, vol. 40, no. 3, pages 347–352, 2004. (Cited on pages 136, 141 and 145.)

[Bollerslev 1986] Tim Bollerslev. *Generalized Autoregressive Conditional Heteroskedasticity.* Journal of Econometrics, vol. 31, no. 3, pages 307–327, 1986. (Cited on page 104.)

[Bollerslev 1994] Tim Bollerslev, Robert F. Engle and Daniel B. Nelson. Arch models, pages 2959–3038. Volume 4 of Engle & McFadden [Engle 1994], 1994. (Cited on page 104.)

[Borland 2005] Lisa Borland, Jean-Philippe Bouchaud, Jean-François Muzy and Gilles Zumbach. *The Dynamics of Financial Markets — Mandelbrot's multifractal cascades, and beyond.* Wilmott Magazine, pages 86–96, March 2005. (Cited on page 160.)

[Borland 2011] Lisa Borland and Jean-Philippe Bouchaud. *On a multi-timescale statistical feedback model for volatility fluctuations.* The Journal of Investment Strategies, vol. 1, no. 1, pages 65–104, December 2011. (Cited on pages 10, 104, 105, 112, 122, 131, 165, 166 and 205.)

[Bouchaud 2001a] Jean-Philippe Bouchaud, Andrew Matacz and Marc Potters. *Leverage Effect in Financial Markets: The Retarded Volatility Model.* Physical Review Letters, vol. 87, page 228701, Nov 2001. (Cited on page 105.)

[Bouchaud 2001b] Jean-Philippe Bouchaud and Marc Potters. *More stylized facts of financial markets: leverage effect and downside correlations.* Physica A: Statistical Mechanics and its Applications, vol. 299, no. 1-2, pages 60–70, 2001. (Cited on page 160.)

[Bouchaud 2003] Jean-Philippe Bouchaud and Marc Potters. Theory of Financial Risk and Derivative Pricing: from Statistical Physics to Risk Management. Cambridge University Press, 2003. (Cited on pages 27, 49, 58 and 70.)

[Bouchaud 2004] Jean-Philippe Bouchaud, Yuval Gefen, Marc Potters and Matthieu Wyart. *Fluctuations and response in financial markets: the subtle nature of 'random' price changes.* Quantitative Finance, vol. 4, no. 2, pages 176–190, 2004. (Cited on page 131.)

[Bouchaud 2006] Jean-Philippe Bouchaud, Julien Kockelkoren and Marc Potters. *Random walks, liquidity molasses and critical response in financial markets.* Quantitative Finance, vol. 6, no. 2, pages 115–123, 2006. (Cited on page 131.)

[Bouchaud 2011] Jean-Philippe Bouchaud. The endogenous dynamics of markets: Price impact, feedback loops and instabilities. In Berd [Berd 2011], 2011. (Cited on pages 117 and 131.)

[Bouchaud 2013] Jean-Philippe Bouchaud. *Crises and Collective Socio-Economic Phenomena: Simple Models and Challenges.* Journal of Statistical Physics, pages 1–40, 2013. (Cited on pages 2 and 5.)

[Bradley 2007] Richard C. Bradley. Introduction to strong mixing conditions, volume 1–3. Kendrick Press, Heber City, Utah, 2007. (Cited on page 51.)

[Bray 2007a] Alan J. Bray and Richard Smith. *Survival of a diffusing particle in an expanding cage.* Journal of Physics A: Mathematical and Theoretical, vol. 40, no. 36, page 10965, 2007. (Cited on page 39.)

[Bray 2007b] Alan J. Bray and Richard Smith. *The survival probability of a diffusing particle constrained by two moving, absorbing boundaries.* Journal of Physics A: Mathematical and Theoretical, vol. 40, no. 10, page F235, 2007. (Cited on page 39.)

[Bray 2013] Alan J. Bray, Satya N. Majumdar and Grégory Schehr. Persistence and first-passage properties in non-equilibrium systems. To appear in Advances in Physics, 2013. (Cited on page 144.)

[Brigo 2010] Damiano Brigo, Andrea Pallavicini and Roberto Torresetti. Credit Models and the Crisis: A journey into CDOs, Copulas, Correlations and Dynamic Models. John Wiley & Sons, Chichester, 2010. (Cited on page 58.)

[Cabaña 1994] Alejandra Cabaña and Enrique M. Cabaña. *Goodness-of-Fit and Comparison Tests of the Kolmogorov-Smirnov Type for Bivariate Populations.* The Annals of Statistics, vol. 22, no. 3, pages 1447–1459, 1994. (Cited on pages 36 and 180.)

[Cabaña 1997] Alejandra Cabaña and Enrique M. Cabaña. *Transformed empirical processes and modified Kolmogorov-Smirnov tests for multivariate distributions.* The Annals of Statistics, vol. 25, no. 6, pages 2388–2409, 1997. (Cited on page 36.)

[Calvet 2008] Laurent E. Calvet and Adlai Fisher. Multifractal Volatility: Theory, Forecasting, and Pricing. Academic Press, 2008. (Cited on pages 106, 160 and 162.)

[Cambanis 1981] Stamatis Cambanis, Steel Huang and Gordon Simons. *On the theory of elliptically contoured distributions.* Journal of Multivariate Analysis, vol. 11, no. 3, pages 368–385, 1981. (Cited on page 24.)

[Challet 2005] Damien Challet, Matteo Marsili and Yi-Cheng. Zhang. Minority games: Interacting agents in financial markets. Oxford finance. Oxford University Press, 2005. (Cited on pages 2 and 5.)

[Chen 2010] Xiaohong Chen, Lars P. Hansen and Marine Carrasco. *Nonlinearity and temporal dependence.* Journal of Econometrics, vol. 155, no. 2, pages 155–169, 2010. (Cited on page 51.)

[Chicheportiche 2010] Rémy Chicheportiche. Dépendances entre actions US: une modélisation en arbre. Master's thesis, ENSAE ParisTech, Jan 2010. (Cited on pages 4, 7 and 62.)

[Chicheportiche 2011] Rémy Chicheportiche and Jean-Philippe Bouchaud. *Goodness-of-fit tests with dependent observations.* Journal of Statistical Mechanics: Theory and Experiment, vol. 2011, no. 09, page P09003, 2011. (Cited on pages v, 36, 47 and 59.)

[Chicheportiche 2012a] Rémy Chicheportiche and Jean-Philippe Bouchaud. *The fine-structure of volatility feedback.* arXiv preprint q-fin.ST/1206.2153, 2012. (Cited on pages v and 165.)

[Chicheportiche 2012b] Rémy Chicheportiche and Jean-Philippe Bouchaud. *The joint distribution of stock returns is not elliptical.* International Journal of Theoretical and Applied Finance, vol. 15, no. 3, May 2012. (Cited on pages v, 27, 50, 86, 134 and 161.)

[Chicheportiche 2012c] Rémy Chicheportiche and Jean-Philippe Bouchaud. *Weighted Kolmogorov-Smirnov test: Accounting for the tails.* Physical Review E, vol. 86, page 041115, Oct 2012. (Cited on pages v and 36.)

[Chicheportiche 2013a] Rémy Chicheportiche and Jean-Philippe Bouchaud. Ironing the copula surface. In preparation, 2013. (Cited on page v.)

[Chicheportiche 2013b] Rémy Chicheportiche and Jean-Philippe Bouchaud. A minimal factor model for non-linear dependences in stock returns. In preparation, 2013. (Cited on pages v and 66.)

[Chicheportiche 2013c] Rémy Chicheportiche and Jean-Philippe Bouchaud. Some applications of first passage ideas to finance. To appear in First-Passage Phenomena and Their Applications (R. Metzler, G. Oshanin and S. Redner Eds.), World Scientific Publishers, 2013. (Cited on pages v, 171 and 181.)

[Chicheportiche 2013d] Rémy Chicheportiche and Anirban Chakraborti. *A model-free characterization of recurrences in stationary time series.* arXiv preprint physics.data-an/1302.3704, 2013. (Cited on pages v and 150.)

[Chicheportiche 2013e] Rémy Chicheportiche and Anirban Chakraborti. Non-linear and multi-points dependences in stationary time series: a copula approach. In preparation, 2013. (Cited on page v.)

[Cizeau 2001] Pierre Cizeau, Marc Potters and Jean-Philippe Bouchaud. *Correlation structure of extreme stock returns.* Quantitative Finance, vol. 1, no. 2, pages 217–222, 2001. (Cited on page 72.)

[Clauset 2009] Aaron Clauset, Cosma Rohilla Shalizi and M. E. J. Newman. *Power-law distributions in empirical data.* SIAM review, vol. 51, no. 4, pages 661–703, 2009. (Cited on page 39.)

[Cont 2001] Rama Cont. *Empirical properties of asset returns: stylized facts and statistical issues.* Quantitative Finance, vol. 1, no. 2, pages 223–236, 2001. (Cited on page 27.)

[Cont 2010] Rama Cont, editeur. *Encyclopedia of quantitative finance.* John Wiley & Sons, Chichester, 2010. (Not cited.)

[Cox 1955] David Roxbee Cox. *Some statistical methods connected with series of events.* Journal of the Royal Statistical Society. Series B (Methodological), pages 129–164, 1955. (Cited on page 143.)

[Cutler 1989] David M. Cutler, James M. Poterba and Lawrence H. Summers. *What moves stock prices?* The Journal of Portfolio Management, vol. 15, no. 3, pages 4–12, 1989. (Cited on page 117.)

[Darling 1957] Donald A. Darling. *The Kolmogorov-Smirnov, Cramer-von Mises Tests.* The Annals of Mathematical Statistics, vol. 28, no. 4, pages 823–838, 1957. (Cited on pages 36 and 37.)

[Darsow 1992] William F. Darsow, Bao Nguyen and Elwood T. Olsen. *Copulas and Markov processes.* Illinois Journal of Mathematics, vol. 36, no. 4, pages 600–642, 1992. (Cited on page 154.)

[Deheuvels 1979] Paul Deheuvels. *La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance.* Bulletin de la Classe des Sciences, 5eme série, vol. 65, no. 6, pages 274–292, 1979. (Cited on page 178.)

[Deheuvels 1980] Paul Deheuvels. *Non parametric tests of independence.* In Jean-Pierre Raoult, editeur, Statistique non Paramétrique Asymptotique, volume 821 of *Lecture Notes in Mathematics*, pages 95–107. Springer Verlag, Berlin Heidelberg, 1980. (Cited on page 178.)

[Deheuvels 2005] Paul Deheuvels. *Weighted Multivariate Cramér-von Mises-type Statistics.* Afrika Statistika, vol. 1, no. 1, page 1:14, 2005. (Cited on pages 36 and 180.)

[Delpini 2012] Danilo Delpini and Giacomo Bormetti. *Stochastic Volatility with Heterogeneous Time Scales.* arXiv preprint q-fin.ST/1206.0026, 2012. (Cited on page 105.)

[Demarta 2005] Stefano Demarta and Alexander J. McNeil. *The t copula and related copulas.* International Statistical Review, vol. 73, no. 1, pages 111–129, 2005. (Cited on page 27.)

[Dempster 2002] Michael Alan Howarth Dempster, editeur. Risk Management: Value at Risk and Beyond. Cambridge University Press, Cambridge, 2002. (Cited on page 217.)

[Doob 1949] Joseph Leo Doob. *Heuristic approach to the Kolmogorov-Smirnov theorems.* The Annals of Mathematical Statistics, vol. 20, no. 3, pages 393–403, 1949. (Cited on page 36.)

[Drăgulescu 2002] Adrian A. Drăgulescu and Victor M. Yakovenko. *Probability distribution of returns in the Heston model with stochastic volatility.* Quantitative Finance, vol. 2, no. 6, pages 443–453, 2002. (Cited on page 49.)

[Duchon 2008] Jean Duchon, Raoul Robert and Vincent Vargas. *Forecasting volatility with the multifractal random walk model*. Mathematical Finance, 2008. (Cited on page 162.)

[Dupire 1994] Bruno Dupire. *Pricing with a smile*. Risk, vol. 7, no. 1, pages 18–20, 1994. (Cited on page 58.)

[Durrleman 2000] Valdo Durrleman, Ashkan Nikeghbali and Thierry Roncalli. *Which copula is the right one*. Ssrn working paper, Groupe de Recherche Opérationnelle, Crédit Lyonnais, 2000. (Cited on page 58.)

[Ebeling 2005] Werner Ebeling and Igor M. Sokolov. Statistical thermodynamics and stochastic theory of nonequilibrium systems, volume 8. World Scientific Publishing Company Incorporated, 2005. (Cited on page 144.)

[Eichner 2006] Jan F. Eichner, Jan W. Kantelhardt, Armin Bunde and Shlomo Havlin. *Extreme value statistics in records with long-term persistence*. Physical Review E, vol. 73, page 016130, Jan 2006. (Cited on page 142.)

[Eisler 2004] Zoltán Eisler and János Kertesz. *Multifractal model of asset returns with leverage effect*. Physica A: Statistical Mechanics and its Applications, vol. 343, pages 603–622, 2004. (Cited on page 160.)

[El Karoui 2009] Nicole El Karoui. *High-dimensionality effects in the Markowitz problem and other quadratic programs with linear equality constraints: risk underestimation*. Working paper, University of Berkeley, 2009. (Cited on page 58.)

[Embrechts 2002a] Paul Embrechts, Alexander McNeil and Daniel Straumann. *Correlation and dependence in risk management: properties and pitfalls*. Risk management: value at risk and beyond, pages 176–223, 2002. (Cited on page 134.)

[Embrechts 2002b] Paul Embrechts, Alexander J. McNeil and Daniel Straumann. Correlation and dependence in risk management: Properties and pitfalls, chapitre 7, pages 176–223. In Dempster [Dempster 2002], 2002. (Cited on pages 16, 24, 27, 28, 58 and 59.)

[Embrechts 2003] Paul Embrechts, Filip Lindskog and Alexander McNeil. *Modelling dependence with copulas and applications to risk management*. Handbook of heavy tailed distributions in finance, vol. 8, pages 329–384, 2003. (Cited on page 134.)

[Engle 1982] Robert F. Engle. *Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation*. Econometrica: Journal of the Econometric Society, pages 987–1007, 1982. (Cited on page 104.)

[Engle 1994] Robert F. Engle and Daniel L. McFadden, editeurs. Handbook of econometrics, volume 4. Elsevier/North-Holland, Amsterdam, 1994. (Cited on page 213.)

[Fair 2002] Ray C. Fair. *Events that shook the market*. Journal of Business, vol. 75, pages 713–732, 2002. (Cited on page 117.)

[Fasano 1987] Giovanni Fasano and Alberto Franceschini. *A multidimensional version of the Kolmogorov-Smirnov test*. Monthly Notices of the Royal Astronomical Society, vol. 225, pages 155–170, 1987. (Cited on pages 36 and 180.)

[Fermanian 2005a] Jean-David Fermanian. *Goodness-of-fit tests for copulas.* Journal of Multivariate Analysis, vol. 95, no. 1, pages 119 – 152, 2005. (Cited on pages 36 and 182.)

[Fermanian 2005b] Jean-David Fermanian and Olivier Scaillet. Some Statistical Pitfalls in Copula Modeling for Financial Applications, chapitre 4, pages 57–72. In Klein [Klein 2005], 2005. (Cited on page 58.)

[Fermanian 2012] Jean-David Fermanian. An overview of the goodness-of-fit test problem for copulas. to appear, 2012. (Cited on page 180.)

[Filimonov 2012] Vladimir Filimonov and Didier Sornette. *Quantifying reflexivity in financial markets: Toward a prediction of flash crashes.* Physical Review E, vol. 85, page 056108, May 2012. (Cited on pages 117, 131 and 144.)

[Filimonov 2013] Vladimir Filimonov, Spencer Wheatley and Didier Sornette. *Effective Measure of Endogeneity for the Autoregressive Conditional Duration Point Processes via Mapping to the Self-Excited Hawkes Process.* arXiv preprint q-fin.ST/1306.2245, 2013. (Cited on page 173.)

[Fischer 2009] Matthias Fischer, Christian Köck, Stephan Schlüter and Florian Weigert. *An empirical analysis of multivariate copula models.* Quantitative Finance, vol. 9, no. 7, pages 839–854, 2009. (Cited on page 58.)

[Floros 2009] Christos Floros. *Modelling volatility using high, low, open and closing prices: evidence from four S&P indices.* International Research Journal of Finance and Economics, vol. 28, pages 198–206, 2009. (Cited on page 114.)

[Fortin 2002] Ines Fortin and Christoph Kuzmics. *Tail-dependence in stock-return pairs.* International Journal of Intelligent Systems in Accounting, Finance & Management, vol. 11, no. 2, pages 89–107, 2002. (Cited on page 58.)

[Frahm 2003] Gabriel Frahm, Markus Junker and Alexander Szimayer. *Elliptical copulas: applicability and limitations.* Statistics & Probability Letters, vol. 63, no. 3, pages 275–286, 2003. (Cited on page 59.)

[Frey 2001] Rüdiger Frey, Alexander J. McNeil and Mark A. Nyfeler. *Modelling Dependent Defaults: Asset Correlations Are Not Enough!* Rapport technique, Swiss Banking Institute, University of Zurich, ETHZ, March 2001. (Cited on page 58.)

[Fuentes 2009] Miguel A. Fuentes, Austin Gerig and Javier Vicente. *Universal Behavior of Extreme Price Movements in Stock Markets.* Public Library of Science ONE, vol. 4, no. 12, page e8243, 2009. (Cited on page 27.)

[Galam 2008] Serge Galam. *Sociophysics: A review of Galam models.* International Journal of Modern Physics C, vol. 19, no. 03, pages 409–440, 2008. (Cited on pages 2 and 5.)

[Genest 1993] Christian Genest and Louis-Paul Rivest. *Statistical inference procedures for bivariate Archimedean copulas.* Journal of the American Statistical Association, vol. 88, no. 423, pages 1034–1043, 1993. (Cited on pages 22 and 180.)

[Genest 2006] Christian Genest, Jean-François Quessy and Bruno Rémillard. *Goodness-of-fit procedures for copula models based on the Probability Integral Transformation.* Scandinavian Journal of Statistics, vol. 33, no. 2, pages 337–366, 2006. (Cited on page 180.)

[Gould 2013] Martin Gould, Rémy Chicheportiche and Jean-Philippe Bouchaud. Reassessing the tail distribution of financial returns. In preparation, 2013. (Cited on pages v and 173.)

[Gradshteyn 1980] Israil S. Gradshteyn and Iosif M. Ryzhik. Table of integrals, series and products. corrected and enlarged edition. Academic Press, New York, London, Toronto and Tokyo, 1980. (Cited on pages 43 and 206.)

[Greenberg 2006] Simon L. Greenberg. Bivariate Goodness-of-fit Tests Based on Kolmogorov-Smirnov Type Statistics. University of Johannesburg, 2006. (Cited on page 182.)

[Gutenberg 1936] Beno Gutenberg and Charles Francis Richter. *Magnitude and Energy of earthquakes.* Science, vol. 83, no. 2147, pages 183–185, 1936. (Cited on page 138.)

[Hagan 2002a] Patrick S. Hagan, Deep Kumar, Andrew S. Lesniewski and Diana E. Woodward. *Managing Smile Risk.* Wilmott Magazine, pages 84–108, September 2002. (Cited on page 58.)

[Hagan 2002b] Patrick S. Hagan, Deep Kumar, Andrew S. Lesniewski and Diana E. Woodward. *Managing smile risk.* Wilmott magazine, pages 84–108, September 2002. (Cited on page 106.)

[Hardiman 2013] Stephen J. Hardiman, Nicolas Bercot and Jean-Philippe Bouchaud. *Critical reflexivity in financial markets: a Hawkes process analysis.* arXiv preprint q-fin.ST/1302.1405, 2013. (Cited on page 166.)

[He 1999] Changli He and Timo Teräsvirta. *Fourth moment structure of the GARCH(p,q) process.* Econometric Theory, vol. 15, no. 6, pages 824–846, 1999. (Cited on page 109.)

[Helbing 2012] Dirk Helbing, editeur. Social self-organization: Agent-based simulations and experiments to study emergent social behavior. Understanding Complex Systems. Springer Verlag, Berlin Heidelberg, 2012. (Cited on pages 3 and 6.)

[Henry-Labordère 2008] Pierre Henry-Labordère. Analysis, geometry, and modeling in finance: Advanced methods in option pricing, volume 13 of *Financial Mathematics Series.* Chapman & Hall/CRC Press, 2008. (Cited on page 106.)

[Heston 1993] Steven L. Heston. *A closed-form solution for options with stochastic volatility with applications to bond and currency options.* Review of Financial Studies, vol. 6, no. 2, pages 327–343, 1993. (Cited on page 106.)

[Hult 2002] Henrik Hult and Filip Lindskog. *Multivariate extremes, aggregation and dependence in elliptical distributions.* Advances in Applied probability, vol. 34, no. 3, pages 587–608, 2002. (Cited on pages 25 and 59.)

[Ibragimov 2008] Rustam Ibragimov and George Lentzas. Copulas and long memory. Harvard Institute of Economic Research discussion paper, 2008. (Cited on pages 134 and 154.)

[Joulin 2008] Armand Joulin, Augustin Lefevre, Daniel Grunberg and Jean-Philippe Bouchaud. *Stock price jumps: News and volume play a minor role*. Wilmott Magazine, pages 1–7, September/October 2008. (Cited on pages 124 and 131.)

[Justel 1997] Ana Justel, Daniel Peña and Rubén Zamar. *A multivariate Kolmogorov-Smirnov test of goodness of fit*. Statistics & Probability Letters, vol. 35, no. 3, pages 251–259, 1997. (Cited on page 180.)

[Kac 1947] Mark Kac. *On the notion of recurrence in discrete stochastic processes*. Bulletin of the American Mathematical Society, vol. 53, no. 10, pages 1002–1010, 1947. (Cited on page 140.)

[Khmaladze 1982] Estate V. Khmaladze. *Martingale Approach in the Theory of Goodness-of-Fit Tests*. Theory of Probability and its Applications, vol. 26, page 240, 1982. (Cited on page 36.)

[Kirman 2010] Alan Kirman. *The economic crisis is a crisis for economic theory*. CESifo Economic Studies, vol. 56, no. 4, pages 498–535, 2010. (Cited on pages 2 and 5.)

[Klein 2005] Edith Klein, editeur. Capital formation, governance and banking. Nova Publishers, 2005. (Cited on page 217.)

[Kolmogorov 1933] Andreï N. Kolmogorov. *Sulla determinazione empirica di una legge di distribuzione*. Giornale dell'Istituto Italiano degli Attuari, vol. 4, no. 1, pages 83–91, 1933. (Cited on pages 36 and 38.)

[Krapivsky 1996] Paul L. Krapivsky and Sidney Redner. *Life and death in an expanding cage and at the edge of a receding cliff*. American Journal of Physics, vol. 64, no. 5, pages 546–551, 1996. (Cited on pages 39, 40, 41, 45, 46 and 47.)

[Kring 2009] Sebastian Kring, Svetlozar T. Rachev, Markus Höchstötter, Frank J. Fabozzi and Michele L. Bianchi. *Multi-tail generalized elliptical distributions for asset returns*. Econometrics Journal, vol. 12, no. 2, pages 272–291, 2009. (Cited on page 31.)

[Laloux 1999] Laurent Laloux, Pierre Cizeau, Jean-Philippe Bouchaud and Marc Potters. *Noise Dressing of Financial Correlation Matrices*. Physical Review Letters, vol. 83, pages 1467–1470, Aug 1999. (Cited on page 70.)

[Laloux 2000] Laurent Laloux, Pierre Cizeau, Marc Potters and Jean-Philippe Bouchaud. *Random matrix theory and financial correlations*. International Journal of Theoretical and Applied Finance, vol. 3, no. 3, pages 391–398, 2000. (Cited on page 70.)

[Ledoit 2004] Olivier Ledoit and Michael Wolf. *A well-conditioned estimator for large-dimensional covariance matrices*. Journal of multivariate analysis, vol. 88, no. 2, pages 365–411, 2004. (Cited on pages 58 and 70.)

[Lillo 2003] Fabrizio Lillo and Rosario N. Mantegna. *Power-law relaxation in a complex system: Omori law after a financial market crash*. Physical Review E, vol. 68, page 016119, Jul 2003. (Cited on pages 131 and 138.)

[Lindskog 2001] Filip Lindskog, Alexander Mcneil and Uwe Schmock. *Kendall's tau for elliptical distributions*. Rapport technique, RiskLab, Departement Mathematik, ETH Zentrum, Zurich, Jun 2001. (Cited on pages 25 and 26.)

[Ling 2002] Shiqing Ling and Michael McAleer. *Stationarity and the existence of moments of a family of GARCH processes*. Journal of Econometrics, vol. 106, no. 1, pages 109–117, 2002. (Cited on page 109.)

[Lladser 2000] Manuel Lladser and Jaime San Martín. *Domain of attraction of the quasi-stationary distributions for the Ornstein-Uhlenbeck process*. Journal of Applied Probability, vol. 37, no. 2, pages 511–521, 2000. (Cited on page 43.)

[Ludescher 2011] Josef Ludescher, Constantino Tsallis and Armin Bunde. *Universal behaviour of interoccurrence times between losses in financial markets: An analytical description*. EPL (Europhysics Letters), vol. 95, no. 6, page 68002, 2011. (Cited on page 150.)

[Luo 2010] Xiaolin Luo and Pavel V. Shevchenko. *The t copula with multiple parameters of degrees of freedom: bivariate characteristics and application to risk management*. Quantitative Finance, vol. 10, no. 9, pages 1039–1054, 2010. (Cited on page 59.)

[Lux 2008] Thomas Lux. *The Multi-Fractal Model of Asset Returns: Its Estimation via GMM and Its Use for Volatility Forecasting*. Journal of Business and Economic Statistics, vol. 26, page 194, 2008. (Cited on pages 106 and 162.)

[Lynch 2003] Paul E. Lynch and Gilles O. Zumbach. *Market heterogeneities and the causal structure of volatility*. Quantitative Finance, vol. 3, no. 4, pages 320–331, 2003. (Cited on pages 104 and 160.)

[Malevergne 2003] Yannick Malevergne and Didier Sornette. *Testing the Gaussian copula hypothesis for financial assets dependences*. Quantitative Finance, vol. 3, no. 4, pages 231–250, 2003. (Cited on page 59.)

[Malevergne 2006] Yannick Malevergne and Didier Sornette. Extreme financial risks: From dependence to risk management. Springer Verlag, Berlin Heidelberg New York, 2006. (Cited on pages 16, 27, 49, 58, 59, 134 and 178.)

[Mandelbrot 1963] Benoit Mandelbrot. *The variation of certain speculative prices*. The Journal of Business, vol. 36, no. 4, pages 394–419, 1963. (Cited on page 104.)

[Mandelbrot 1968] Benoit B. Mandelbrot and John W. Van Ness. *Fractional Brownian motions, fractional noises and applications*. SIAM review, vol. 10, no. 4, pages 422–437, 1968. (Cited on page 157.)

[Mandelbrot 1974] Benoit B. Mandelbrot. *Intermittent turbulence in self-similar cascades-Divergence of high moments and dimension of the carrier.* Journal of Fluid Mechanics, vol. 62, no. 2, pages 331–358, 1974. (Cited on page 106.)

[Manner 2010] Hans Manner and Johan Segers. *Tails of correlation mixtures of elliptical copulas.* Insurance: Mathematics and Economics, 2010. (Cited on page 27.)

[Markowitz 1952] Harry Markowitz. *Portfolio Selection.* The Journal of Finance, vol. 7, no. 1, pages 77–91, 1952. (Cited on page 70.)

[Markowitz 1959] Harry M. Markowitz. Portfolio selection: efficient diversification of investments. John Wiley & Sons, Berlin Heidelberg, 1959. (Cited on page 70.)

[Marsili 2002] Matteo Marsili. *Dissecting financial markets: sectors and states.* Quantitative Finance, vol. 2, page 297, 2002. (Cited on page 58.)

[Mashal 2002] Roy Mashal and Assaf Zeevi. Beyond correlation: Extreme co-movements between financial assets. SSRN working paper, 2002. (Cited on page 58.)

[Mason 1983] David M. Mason and John H. Schuenemeyer. *A Modified Kolmogorov-Smirnov Test Sensitive to Tail Alternatives.* The Annals of Statistics, vol. 11, no. 3, pages 933–946, 1983. (Cited on page 39.)

[Mason 1992] David M. Mason and John H. Schuenemeyer. *Correction: A Modified Kolmogorov-Smirnov Test Sensitive to Tail Alternatives.* The Annals of Statistics, vol. 20, no. 1, pages 620–621, 1992. (Cited on page 39.)

[Mei 1983] W. N. Mei and Y. C. Lee. *Harmonic oscillator with potential barriers — exact solutions and perturbative treatments.* Journal of Physics A: Mathematical and General, vol. 16, page 1623, 1983. (Cited on page 43.)

[Metzler 1999] Ralf Metzler, Eli Barkai and Joseph Klafter. *Anomalous diffusion and relaxation close to thermal equilibrium: A fractional Fokker-Planck equation approach.* Physical Review Letters, vol. 82, no. 18, pages 3563–3567, 1999. (Cited on page 144.)

[Mikosch 2006] Thomas Mikosch. *Copulas: Tales and facts.* Extremes, vol. 9, no. 1, pages 3–20, 2006. (Cited on pages 22 and 134.)

[Morales 2012] Raffaello Morales, Tiziana Di Matteo and Tomaso Aste. *Non stationary multifractality in stock returns.* arXiv preprint q-fin.ST/1212.3195, 2012. (Cited on page 166.)

[Müller 1997] Ulrich A. Müller, Michel M. Dacorogna, Rakhal D. Davé, Richard B. Olsen, Olivier V. Pictet and Jacob E. von Weizsäcker. *Volatilities of different time resolutions — analyzing the dynamics of market components.* Journal of Empirical Finance, vol. 4, no. 2, pages 213–239, 1997. (Cited on pages 104 and 112.)

[Muni Toke 2012] Ioane Muni Toke and Fabrizio Pomponio. *Modelling Trades-Through in a Limit Order Book Using Hawkes Processes.* Economics: The Open-Access, Open-Assessment E-Journal, vol. 6, no. 22, 2012. (Cited on page 144.)

[Muzy 2000] Jean-François Muzy, Emmanuel Bacry and Jean Delour. *Modelling fluctuations of financial time series: from cascade process to stochastic volatility model*. The European Physical Journal B, vol. 17, no. 3, pages 537–548, 2000. (Cited on pages 106, 160 and 162.)

[Muzy 2001] Jean-François Muzy, Didier Sornette, Jean Delour and Alain Arneodo. *Multifractal returns and hierarchical portfolio theory*. Quantitative Finance, vol. 1, no. 1, pages 131–148, 2001. (Cited on page 162.)

[Muzy 2006] Jean-François. Muzy, Emmanuel Bacry and Alexey Kozhemyak. *Extreme values and fat tails of multifractal fluctuations*. Physical Review E, vol. 73, page 066114, Jun 2006. (Cited on page 166.)

[Muzy 2013] Jean-François Muzy, Rachel Baïle and Emmanuel Bacry. *Random cascade model in the limit of infinite integral scale as the exponential of a non-stationary $1/f$ noise. Application to volatility fluctuations in stock markets*. arXiv preprint q-fin.ST/1301.4160, 2013. (Cited on page 166.)

[Niederhausen 1981a] Heinrich Niederhausen. *Sheffer polynomials for computing exact Kolmogorov-Smirnov and Rényi type distributions*. The Annals of Statistics, vol. 9, no. 5, pages 923–944, 1981. (Cited on page 40.)

[Niederhausen 1981b] Heinrich Niederhausen. *Tables of Significance Points for the Variance-Weighted Kolmogorov-Smirnov Statistics*. Rapport technique, Stanford University, Department of Statistics, February 1981. (Cited on page 40.)

[Noé 1968] Marc Noé and Georges Vandewiele. *The calculation of distributions of Kolmogorov-Smirnov type statistics including a table of significance points for a particular case*. The Annals of Mathematical Statistics, vol. 39, no. 1, pages 233–241, 1968. (Cited on pages 36 and 40.)

[Noé 1972] Marc Noé. *The Calculation of Distributions of Two-Sided Kolmogorov-Smirnov Type Statistics*. The Annals of Mathematical Statistics, vol. 43, no. 1, pages 58–64, 1972. (Cited on page 40.)

[Omori 1895] Fusakichi Omori. *On the aftershocks of earthquakes*. Journal of the College of Science, no. 7, pages 111–200, 1895. (Cited on page 136.)

[Parisi 2012] Daniel R. Parisi, Didier Sornette and Dirk Helbing. *Universality class of balanced flows with bottlenecks: granular flows, pedestrian fluxes and financial price dynamics*. arXiv preprint q-fin.ST/1205.2915, 2012. (Cited on page 131.)

[Pasquini 1999] Michele Pasquini and Maurizio Serva. *Multiscaling and clustering of volatility*. Physica A: Statistical Mechanics and its Applications, vol. 269, no. 1, pages 140–147, 1999. (Cited on page 160.)

[Patton 2001] Andrew Patton. *Estimation of Copula Models for Time Series of Possibly Different Length*. Economics working paper series, University of California at San Diego, 2001. (Cited on page 58.)

[Patton 2009] Andrew J. Patton. *Copula-based models for financial time series.* Handbook of financial time series, pages 767–785, 2009. (Cited on page 134.)

[Perelló 2004] Josep Perelló, Jaume Masoliver and Jean-Philippe Bouchaud. *Multiple time scales in volatility and leverage correlations: a stochastic volatility model.* Applied Mathematical Finance, vol. 11, no. 1, pages 27–50, 2004. (Cited on page 160.)

[Petersen 2010] Alexander M. Petersen, Fengzhong Wang, Shlomo Havlin and H. Eugene Stanley. *Market dynamics immediately before and after financial shocks: Quantifying the Omori, productivity, and Bath laws.* Physical Review E, vol. 82, no. 3, page 6114, Sep 2010. (Cited on page 138.)

[Plerou 1999] Vasiliki Plerou, Parameswaran Gopikrishnan, Luís A. Nunes Amaral, Martin Meyer and H. Eugene Stanley. *Scaling of the distribution of price fluctuations of individual companies.* Physical Review E, vol. 60, pages 6519–6529, Dec 1999. (Cited on pages 27 and 49.)

[Pochart 2002] Benoît Pochart and Jean-Philippe Bouchaud. *The skewed multifractal random walk with applications to option smiles.* Quantitative Finance, vol. 2, no. 4, pages 303–314, 2002. (Cited on page 160.)

[Pomeau 1985] Y. Pomeau. *Symétrie des fluctuations dans le renversement du temps.* Journal de Physique, vol. 43, no. 6, pages 859–867, 1985. (Cited on page 128.)

[Potters 2005] Marc Potters, Jean-Philippe Bouchaud and Laurent Laloux. *Financial Applications of Random Matrix Theory: Old Laces and New Pieces.* Acta Physica Polonica B, vol. 36, page 2767, 2005. (Cited on page 70.)

[Potters 2009] Marc Potters and Jean-Philippe Bouchaud. *Financial applications of random matrix theory: a short review.* arXiv preprint q-fin.ST/0190.1205, 2009. (Cited on pages 58, 77 and 89.)

[Redner 2001] Sidney Redner. *A guide to first-passage processes.* Cambridge University Press, Cambridge, UK, 2001. (Cited on page 40.)

[Reigneron 2011] Pierre-Alain Reigneron, Romain Allez and Jean-Philippe Bouchaud. *Principal Regression Analysis and the index leverage effect.* Physica A: Statistical Mechanics and its Applications, vol. 390, no. 17, pages 3026–3035, 2011. (Cited on pages 105 and 160.)

[Ren 2010] Fei Ren and Wei-Xing Zhou. *Recurrence interval analysis of high-frequency financial returns and its application to risk estimation.* New Journal of Physics, vol. 12, no. 7, page 5030, 2010. (Cited on page 141.)

[Richard 2013] Alexandre Richard, Rémy Chicheportiche and Jean-Philippe Bouchaud. The sup of the 2D pinned Brownian sheet. In preparation, 2013. (Cited on page v.)

[Rogers 1991] L. Christopher G. Rogers and Stephen E. Satchell. *Estimating variance from high, low and closing prices.* The Annals of Applied Probability, pages 504–512, 1991. (Cited on page 114.)

[Rogers 2008] Tim Rogers, Isaac Pérez Castillo, Reimer Kühn and Koujin Takeda. *Cavity approach to the spectral density of sparse symmetric random matrices*. Physical Review E, vol. 78, no. 3, page 031116, 2008. (Cited on page 172.)

[Rosenblatt 1952] Murray Rosenblatt. *Remarks on a multivariate transformation*. The Annals of Mathematical Statistics, vol. 23, no. 3, pages 470–472, 1952. (Cited on pages 180 and 185.)

[Saichev 2007] Alexander Saichev and Didier Sornette. *Theory of earthquake recurrence times*. Journal of Geophysical Research: Solid Earth (1978–2012), vol. 112, no. B4, 2007. (Cited on page 144.)

[Santhanam 2008] M. S. Santhanam and Holger Kantz. *Return interval distribution of extreme events and long-term memory*. Physical Review E, vol. 78, no. 5, page 1113, Nov 2008. (Cited on page 140.)

[Sazuka 2009] Naoya Sazuka, Jun-Ichi Inoue and Enrico Scalas. *The distribution of first-passage times and durations in FOREX and future markets*. Physica A: Statistical Mechanics and its Applications, vol. 388, no. 14, pages 2839 – 2853, 2009. (Cited on page 140.)

[Schmid 2007] Friedrich Schmid and Rafael Schmidt. *Nonparametric inference on multivariate versions of Blomqvist's beta and related measures of tail dependence*. Metrika, vol. 66, no. 3, pages 323–354, 2007. (Cited on page 21.)

[Sentana 1995] Enrique Sentana. *Quadratic ARCH models*. The Review of Economic Studies, vol. 62, no. 4, page 639, 1995. (Cited on pages 105, 106 and 107.)

[Shaw 2007] W. T. Shaw and K. T. A. Lee. *Copula Methods vs Canonical Multivariate Distributions: the multivariate Student T Distribution with general degrees of freedom*. Working paper, King's College London, 2007. (Cited on page 59.)

[Shiller 1981] Robert J. Shiller. *Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?* American Economic Review, vol. 71, no. 3, pages 421–436, 1981. (Cited on pages 117 and 131.)

[Sklar 1959] Abe Sklar. *Fonctions de répartition à n dimensions et leurs marges*. Publ. Inst. Statist. Univ. Paris, vol. 8, pages 229–231, 1959. (Cited on pages 18 and 58.)

[Smirnov 1948] Nikolai Smirnov. *Table for estimating the goodness of fit of empirical distributions*. The Annals of Mathematical Statistics, vol. 19, no. 2, pages 279–281, 1948. (Cited on pages 36 and 38.)

[Sokolov 2002] Igor M. Sokolov. *Solutions of a class of non-Markovian Fokker-Planck equations*. Physical Review E, vol. 66, no. 4, page 041101, 2002. (Cited on page 144.)

[Sornette 2003] Didier Sornette, Yannick Malevergne and Jean-François Muzy. *What causes crashes*. Risk Magazine, vol. 67, February 2003. (Cited on page 132.)

[Sornette 2004] Didier Sornette, Fabrice Deschâtres, Thomas Gilbert and Yann Ageon. *Endogenous Versus Exogenous Shocks in Complex Networks: An Empirical Test Using Book Sale Rankings*. Physical Review Letters, vol. 93, page 228701, Nov 2004. (Cited on page 131.)

[Sornette 2008] D Sornette, S Utkin and A Saichev. *Solution of the nonlinear theory and tests of earthquake recurrence times*. Physical Review E, vol. 77, no. 6, page 066109, 2008. (Cited on page 144.)

[Soros 1994] Georges Soros. The alchemy of finance: Reading the mind of the market. Wiley Audio. John Wiley & Sons, Ney York, 1994. (Cited on pages 117 and 131.)

[Teyssière 2010] Gilles Teyssière and Alan P. Kirman. Long memory in economics. Springer Verlag, Berlin Heidelberg New York, 2010. (Cited on page 109.)

[Tilak 2013] Gayatri Tilak, Tamás Széll, Rémy Chicheportiche and Anirban Chakraborti. *Study of Statistical Correlations in Intraday and Daily Financial Return Time Series*. In Frédéric Abergel, Bikas K. Chakrabarti, Anirban Chakraborti and Asim Ghosh, editeurs, Econophysics of Systemic Risk and Network Dynamics, New Economic Windows, pages 77–104. Springer, Milan, 2013. (Not cited.)

[Tola 2008] Vincenzo Tola, Fabrizio Lillo, Mauro Gallegati and Rosario N. Mantegna. *Cluster analysis for portfolio optimization*. Journal of Economic Dynamics and Control, vol. 32, no. 1, pages 235–258, 2008. (Cited on page 58.)

[Tóth 2011] Bence Tóth, Yves Lempérière, Cyril Deremble, Joachim De Lataillade, Julien Kockelkoren and Jean-Philippe Bouchaud. *Anomalous Price Impact and the Critical Nature of Liquidity in Financial Markets*. Physical Review X, vol. 1, page 021006, Oct 2011. (Cited on page 131.)

[Tumminello 2007] Michele Tumminello, Fabrizio Lillo and Rosario N. Mantegna. *Hierarchically nested factor model from multivariate data*. EPL (Europhysics Letters), vol. 78, page 30006, 2007. (Cited on page 58.)

[Turban 1992] Loïc Turban. *Anisotropic critical phenomena in parabolic geometries: the directed self-avoiding walk*. Journal of Physics A: Mathematical and General, vol. 25, page L127, 1992. (Cited on page 41.)

[Virasoro 2011] Miguel Angel Virasoro. *Non-Gaussianity of the Intraday Returns Distribution: its evolution in time*. arXiv preprint q-fin.ST/1112.0770, 2011. (Cited on page 132.)

[Weber 2007] Philipp Weber, Fengzhong Wang, Irena Vodenska-Chitkushev, Shlomo Havlin and H. Eugene Stanley. *Relation between volatility correlations in financial markets and Omori processes occurring on all scales*. Physical Review E, vol. 76, no. 1, page 6109, Jul 2007. (Cited on page 138.)

[Weiss 1978] Marc S. Weiss. *Modification of the Kolmogorov-Smirnov statistic for use with correlated data*. Journal of the American Statistical Association, vol. 73, no. 364, pages 872–875, 1978. (Cited on page 49.)

[Wilcox 1989] Rand R. Wilcox. *Percentage points of a weighted Kolmogorov-Smirnov statistic.* Communications in Statistics - Simulation and Computation, vol. 18, no. 1, pages 237–244, 1989. (Cited on page 40.)

[Wu 2007] Florence Wu, Emiliano A. Valdez and Michael Sherris. *Simulating Exchangeable Multivariate Archimedean Copulas and its Applications.* Communications in Statistics - Simulation and Computation, vol. 36, no. 5, pages 1019–1034, 2007. (Cited on pages 21 and 22.)

[Zawadowski 2006] Ádám G. Zawadowski, György Andor and János Kertész. *Short-term market reaction after extreme price changes of liquid stocks.* Quantitative Finance, vol. 6, no. 4, pages 283–295, 2006. (Cited on page 131.)

[Zumbach 2001] Gilles Zumbach and Paul Lynch. *Heterogeneous volatility cascade in financial markets.* Physica A: Statistical Mechanics and its Applications, vol. 298, no. 3-4, pages 521–529, 2001. (Cited on pages 160 and 165.)

[Zumbach 2009] Gilles O. Zumbach. *Time reversal invariance in finance.* Quantitative Finance, vol. 9, no. 5, pages 505–515, 2009. (Cited on pages 106, 128 and 129.)

[Zumbach 2010] Gilles O. Zumbach. *Volatility conditional on price trends.* Quantitative Finance, vol. 10, no. 4, pages 431–442, 2010. (Cited on pages 105, 112 and 131.)