UNIVERSITÉ PARIS-DAUPHINE
ÉCOLE DOCTORALE DE DAUPHINE
Laboratoire d'Économie de Dauphine (LEDa-SDFi)

No. attribué par la bibliothèque

STRESS TESTING AND FINANCIAL RISKS

Thèse

Pour l'obtention du titre de

Docteur en Sciences Économiques

(Arrêté du 7 août 2006)

Présentée et soutenue publiquement par

Lyes KOLIAI

Le 27 octobre 2014

Sous la direction de Monsieur le Professeur Sanvi AVOUYI-DOVI

Jury de thèse

Directeur de thèse	Pr. Sanvi AVOUYI-DOVI	Professeur associé – Université Paris-Dauphine Conseiller – Banque de France
Rapporteurs	Pr. Bertrand CANDELON	Professeur – Maastricht University
	Pr. Christophe HURLIN	Professeur – Université d'Orléans
Suffragants	Pr. Olivier de BANDT	Professeur associé – Université Paris-Ouest Directeur des Études – ACPR
	Pr. Anton BRENDER	Professeur honoraire – Université Paris-Dauphine Chef économiste – Candriam Investors Group
	Dr. Julio RAMOS-TALLADA	Senior économiste – Banque de France

Paris-Dauphine University is not to provide any approval or disapproval regarding the opinions contained in this thesis. These opinions must be considered as being solely those of their author.

Acknowledgements

It would not have been possible to write this doctoral thesis without the help and support of the kind people around me, to only some of whom it is possible to give a particular mention here.

This thesis would not have been possible without the help, support, and patience of my supervisor, Prof. Sanvi Avouyi-Dovi, not to mention his advice and unsurpassed knowledge. He has been invaluable on both an academic and a personal level. I could not have imagined having a better advisor and mentor for my thesis. I am extremely grateful.

Besides my advisor, I would like to thank the rest of my dissertation committee: Profs. Christpothe HURLIN, Bertrand CANDELON, Anton BRENDER, and Olivier de BANDT, and Dr. Julio RAMOS-TALLADA for having accepted participating in the jury.

I would like to thank Virginie FAJON, Marie-Christelle ATOHOUN and Cristophe PRAT for their kindness, friendship and administrative support, together with the other assistants of Prof. Avouyi-Dovi at Banque de France.

I thank my fellow labmates in Paris-Dauphine University for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last four years.

A big thank you to my colleagues and friends for their support and encouragement throughout. A special note for Profs. Eike Christian BRECHMANN and Helmut ELSINGER for their precious help.

Last but not the least, I would like to thank my family for supporting me spiritually throughout my life, and for which my mere expression of thanks likewise does not suffice.

Contents

Acknowledgements . . . 5
List of Tables . . . 11
List of Figures . . . 13

1 Introduction **1**

Résumé en français des chapitres de la thèse **8**

2 Financial Stress Tests
Lessons from a Decade of Practice **31**

2.1 Introduction . . . 32
2.2 Relevance criteria of a stress test . . . 33
2.2.1 Severity . . . 33
2.2.2 Plausibility . . . 33
2.2.3 Usefulness . . . 35
2.3 Conceptual issues . . . 36
2.3.1 Stakeholders and prior arrangements . . . 36
2.3.2 Setting the starting assumptions . . . 37
2.4 Applications . . . 39
2.4.1 Objectives . . . 39
2.4.2 Stress testing and risk management . . . 39
2.5 Scope . . . 41
2.5.1 Portfolios . . . 41
2.5.2 Risk factors . . . 41
2.5.3 Horizon . . . 42
2.6 Design of scenarios . . . 43
2.6.1 Typology and selection criteria . . . 44
2.6.2 Parameter setting . . . 45
2.7 Results . . . 47
2.7.1 Decision-making . . . 47

2.7.2 Complementing other management tools . . . 48
2.8 Conclusion . . . 49

Appendices **53**
2.A Value-at-Risk (VaR) . . . 53
2.B A unified framework VaR – stress tests . . . 54

3 Extreme Risk Modelling
A Sequential EVT – Pair-copulas Approach **57**
3.1 Introduction . . . 58
3.2 Financial extreme returns modelling . . . 59
3.2.1 The marginal distributions . . . 59
3.2.2 The dependence structure . . . 60
3.3 Data and descriptive statistics . . . 61
3.3.1 Data . . . 61
3.3.2 Statistical evidence of stylized facts . . . 62
3.4 The model . . . 65
3.4.1 The marginal model . . . 65
3.4.2 The dependence structure model . . . 67
3.5 Empirical results . . . 68
3.5.1 The EVT-EGARCH models . . . 68
3.5.2 The R-vine model . . . 70
3.5.3 Validation tests . . . 71
3.6 An application to financial stress tests . . . 72
3.7 Conclusion . . . 74

Appendices **81**
3.A Stylized facts in financial time series . . . 81
3.B Generalized hyperbolic distributions (GH) . . . 82
3.C The Extreme value theory (EVT) . . . 83
3.D The direct multivarite modelling approach . . . 85
3.E Copulas . . . 87
3.F The dependence structure . . . 89
3.G The R-vine model for risk management . . . 91

4 Systemic Risk Assessment
Stress Testing the French Banking System **113**
4.1 Introduction . . . 114

4.2 Systemic risk modelling . . . 115
4.2.1 Systemic risk and systemic risk factors . . . 115
4.2.2 Systemic risk valuation models . . . 116
4.3 The model . . . 118
4.3.1 Overview . . . 118
4.3.2 The risk model . . . 119
4.3.3 First-round effects . . . 120
4.3.4 Private response functions . . . 123
4.3.5 Contagion effects . . . 125
4.3.6 Public response functions . . . 128
4.4 Data and estimation results . . . 128
4.5 Simulation and stress testing results . . . 130
4.6 Conclusion . . . 135

5 Systematic Identification of Stress Scenarios A Bayesian Approach 159

5.1 Introduction . . . 160
5.2 The systematic research of stress scenarios . . . 161
5.3 Distance metrics for stress scenario selection . . . 162
5.4 Stress scenarios' selection: Bayesian methods . . . 164
5.4.1 Nested sampling parameter inference . . . 165
5.4.2 Worst case scenario selection . . . 167
5.5 Experimental results . . . 168
5.6 Stress tests in risk management . . . 170
5.6.1 A unified framework for stress tests and VaR . . . 170
5.6.2 Identification of core risk factors . . . 171
5.6.3 Portfolio allocation . . . 171
5.6.4 Reverse stress tests . . . 172
5.7 Conclusion . . . 173

6 Conclusion 181

List of Tables

3.1 Data description . . . 93
3.2 Stationarity tests . . . 94
3.3 Descriptive statistics . . . 95
3.4 Nonparametric extreme dependence coefficients and tests of asymmetry . . . 97
3.5 Statistical features of the bivariate copulas . . . 98
3.6 Estimation and diagnostic tests of the ARMA-EGARCH models . . . 100
3.7 Etimation and GoF tests for the marginal distributions . . . 101
3.8 Sequential estimation of the R-vine model . . . 103
3.9 Assessment of the static properties of the R-vine model . . . 104
3.10 Assessment of the dynamic properties of the R-vine model (Baktesting) . . . 106
3.11 Estimation of the initial shock in the stress scenarios . . . 107
3.12 Impact of the scenarios in terms of portfolios' return . . . 107

4.1 Classification and valuation models for balance sheet items . . . 141
4.2 Data description . . . 142
4.3 Estimation results . . . 143
4.4 Solvability and leverage ratios: unweighted vs. weighted values . . . 144
4.5 First-round effects: A decomposition by portfolio exposures (last period) . . . 145
4.6 First-round effects: classification of the banking population (last period) . . . 145
4.7 Second-round effects (last period) . . . 148
4.8 Contagion effects . . . 149
4.9 Outcomes evolution across stages and over the simulation horizon . . . 152

5.1 Descriptive statistics of the marginal posterior distributions . . . 177
5.2 Outputs of the worst case scenario . . . 177

List of Figures

2.1 A simplified overview of a stress testing exercise 55

3.1 Empirical exceedance correlations (equity indices) 108
3.2 Vine-copulas models' structure 109
3.3 Graphical representation of the R-vine estimation results 110
3.4 Out-of-sample performance of multivariate risk models (equity indices) 111
3.5 Impact of the stress scenarios in terms of regulatory capital 112

4.1 Overview of the valuation model's main stages 153
4.2 Simulated paths of the risk factors over the stress horizon 154
4.3 First-round effects over the stress horizon 155
4.4 First-round effects: distribution of results (last period) 156
4.5 Private response functions 157

5.1 The portfolio's value profile 178
5.2 Nested Sampling estimation results for the BVAR model 179
5.3 Expected empirical density functions for the risk factors 180

1. Introduction

The succession of financial and economic crises over the last few decades has caused unexpected losses both on the individual and the systemic levels. Within the financial sector, private portfolios as well as regional and global systems have suffered widespread and complex repercussions from these events. At the same time, the undertaken responses have not always been optimal. The ensuing discussion surrounding these events has been structured around three central issues: (i) an identification of the core risk factors behind financial crises, (ii) an accurate assessment of their potential impact on the financial sector and the real economy, and (iii) a definition of appropriate private and public measures allowing to prevent their occurrence and to react properly when they do. This thesis is part of the enhancing literatures related to the last two topics.

Until quite recently, predicting the impact of extreme events on financial portfolios and the design of related management actions have been led through a series of conventional tools, such as the Value at Risk (VaR) and Financial Soundness Indicators (FSIs). These tools are usually designed by regulators and financial institutions and made available for risk managers and public policy-makers. They are often based on simple hypotheses deemed reliable and valid at all times. However, their consistency has been seriously questioned after a recurrence of extreme events whose impact exceeds – sometimes even contradicts – all expectations. These failures are mainly due to the incoherence of some tools as risk measures and to the unsuitability of some underlying hypotheses. In most cases, these tools also omit some realistic aspects, such as endogenous behaviours within the financial system. This explains several failures in the management actions considered, before, during, and after the occurrence of extreme events.

The need for a significant change in the existing practices of (extreme) risk management had become increasingly evident, hastened by higher costs in terms of economic stability and social welfare. It is in this context that the first financial "stress testing exercises" - or simply "stress tests" – have been proposed. These are designed to complement the conventional risk management framework. Stress testing consists in estimating the likely impact of *harmful yet plausible* events (the stress scenario) on some variables of interest. This impact is assessed over a given period of time (the stress horizon) using an adapted valuation model. Compared to other pure statistical tools, stress tests allow personifying all the events that make the scenario. By doing so, the portfolio's exposures are explicitly identified and each potential outcome is associated to the generating scenario; hence guiding decision-making. Moreover, the scenario and the valuation model can be augmented to account for more realistic features, such as second-round and feedback effects, private investment strategies, public response functions, etc. This allows improving the knowledge of the financial sector functioning, particularly during turmoil episodes.

Given the expected contribution of stress testing exercises, these have been widely disseminated by major international regulators (see, e.g., IIF, 2008; CRMPG III, 2008; BCBS, 2009;

CEBS, 2010a). A series of recommendations have been made to inspire portfolio managers and national supervisors to review their practices, by implementing stress tests on a regular basis. Later on, the inclusion of the tests' results has been required for both active and prudential risk management processes. However, the first realized exercises have been dotted with several limits both on methodological and practical aspects (BCBS, 2012; IMF, 2012; Borio et al., 2012; Greenlaw et al., 2012). Most criticism have been levelled against a subjective choice of stress scenarios, insufficiencies in the used valuation models, and higher implementation costs (e.g. data, expertise, and time requirements, communication issues, etc.). Due to reservations voiced even by the concerned decision-makers, the results of the first stress tests have repeatedly been neglected. Quite often, regulatory constraints and market pressures (customers, investors, rating agencies, etc.) are the main incentives to conduct stress tests and to capitalize on their results.

The recent financial crises have enhanced the role of these external factors. Accordingly, the number of realized stress tests has increased considerably. In most cases, these are carried out under the auspice of supervision authorities. US banks (FRB, 2009, 2011, 2012, 2013a,b, 2014) and EU-wide banks (CEBS, 2009, 2010b; EBA, 2011, 2014a,b) stress tests are the among most popular examples. However, the absence of a clear and reliable assessment of these exercises often makes them ambiguous to say the least. A lack of consistent literature on their conceptual and practical aspects has often led to divergent viewpoints when it comes to assess the underlying assumptions (e.g. scenarios, models, etc.) and to comment the outcomes. Even though the publication of recent stress testing results is sometimes accompanied by information folders, these still lack of an objective assessment of the strengths and weaknesses of these exercises. Hence, while the publication of the results may impact the behaviour of one part of market players, the apprehension is still dominating for the remaining part, including specialists and the general public.

The existing literature on financial stress tests can be split into four main topics. First papers, going back to the early 2000s, have been dedicated to a general presentation of the conceptual aspects of this tool, at the time considered as relatively new in financial risk management (Berkowitz, 2000; Blaschke et al., 2001; Čihák, 2007). A second body of the literature has focused on practical guidelines for the so-called individual or portfolio stress tests (Kupiec, 1998; Breuer and Krenn, 1999; Bee, 2001; Kim and Finger, 2000; Breuer et al., 2002; Alexander and Sheedy, 2008; Breuer and Csiszár, 2013; McNeil and Smith, 2012). Later on, due to the widespread repercussions of the 2007-2009 financial crisis, researches have been more concerned about systemic stress tests (Boss, 2008; Alessandri et al., 2009; Aikman et al., 2009; van den End, 2010, 2012; Engle et al., 2014; Acharya et al., 2014). A final wave of papers have made a first diagnostic of the realized exercises since the 2007-2009 financial crisis, by outlining their main limits and the remaining challenges (Haldane, 2009; Borio and Drehmann, 2009; Hirtle et al., 2009; IMF, 2012; Greenlaw et al., 2012; Borio et al., 2012).

The objective of this thesis is twofold:

(i) Introducing a comprehensive framework aiming to set the bases of an objective assessment of the conceptual and the practical aspects of stress testing exercises. This allows identifying and preventing the main shortcomings of the existing tests.

(ii) Developing a robust and flexible framework that takes into account the main stages of a stress testing exercise, namely risk factors' modelling, stress scenarios' selection, and valuation model. This allows extending the current literature – when it does exist – and

validating the potential of stress tests for a better use of their results in financial risk management.

Our basic premise is that an additional flexibility in the underlying hypotheses of a stress test enhances the credibility of the exercise. This, in turn, improves the utility of the results when it comes to decision-making. However, the added flexibility may also lead to additional costs, in terms of implementation, communication, and results' treatment. Hence, the central challenge is to design stress testing exercises associating consistent theoretical framework with practical handling.

The first chapter performs a comprehensive analysis of financial stress tests, based on a qualitative and quantitative assessment of the main stages of the exercise. We compare, for each stage, the underlying theoretical and practical aspects with respect to international regulatory standards in terms of risk management and stress testing. This method is applied to the stress tests carried out on major financial institutions and financial systems of developed countries over the period 2000-2014. The results allow assessing the strengths and weaknesses of current practices, the relevance of the related results, and the remaining challenges.

The second chapter presents a semi-parametric approach to model the risk factors that make a stress testing scenario. Extreme changes in financial returns are considered as core risk factors. The proposed risk model is based on a flexible and sequential specification of the distribution and the dependence structure of the returns. The Extreme Value Theory (EVT) and the theorem of pair-copulas are used to put a specific emphasis on extreme returns. The model performance is assessed for three sets of assets: equity indices, nominal exchange rates, and commodity prices. Empirical results support a better static and dynamic properties of the presented model compared to most common specifications used in practice. Within a stress testing framework, the use of a wide range of risk model specifications is proven to cause a significant disparity in the obtained results. Hence, adopting a flexible and powerful specification ensures a better credibility of the stress scenario and enhances the usefulness of the test.

The third chapter presents a valuation model adapted to systemic stress testing exercises. The model allows assessing the impact of stress scenarios on a banking system within a bottom-up approach. A modular and sequential specification is used to capture the direct impact of the scenario on the individual bank balance sheets, as well as the resulting dynamic process over an arbitrary horizon. The direct impact is assessed on the balance sheet items exposed to market, credit, and interest rate risks. Changes on a solvency ratio are recorded accordingly. Given the obtained value of this ratio, a dynamic process is engaged, including one or more of the following features: (i) individual bank reactions to the shock, (ii) the shock transmission across banks, through interbank networks, the financial market, and the money market channels, (iii) second-round effects, and (iv) public response functions represented by the central bank and the Treasury. The model is estimated and simulated quarterly for the French banking system. The results show a high vulnerability of the trading portfolios compared to other banking activities. Second-round effects seem to be relatively limited except in the presence of a severe stress scenario. Public responses turned to be crucial, given that scenarios where this has been omitted result in a failure of a significant part of the banking sector after only two quarters of simulation.

The final chapter adapts the Nested Sampling Bayesian approach (Skilling, 2004, 2006) to carry out a systematic selection of stress testing scenarios. This approach supports scenarios consisting of changes in risk factor values and uncertainties in the parameters of the underlying

risk model. From the estimated posterior distribution of parameters, the set of plausible scenarios is bounded using extreme quartiles. The impact of each plausible scenario on the portfolio is then assessed using the Maximum Loss (MaxLoss) measure. This allows, for a given plausibility threshold, identifying the most severe scenario(s) for the portfolio. This method improves the commonly used subjective and ad hoc methods to set stress scenarios. Moreover, due to their flexibility, Bayesian methods are more suited to do so than alternative methods based on distance metrics. Finally, the stochastic framework of Bayesian methods allows for useful applications of stress testing exercises in risk management (e.g. core risk factors identification, portfolio allocation, etc.) and for a more complementarity between stress tests and conventional risk measures.

Different readings may be made for the last three chapters of this thesis. Firstly, the chapters can be seen to as focusing on specific stages of the stress testing exercise. Hence, the second chapter studies the risk factors composing the stress scenario, the following chapter analyses the impact of these risk factors on financial institutions, and the last chapter proposes to take account of the outcomes in the risk management process. Secondly, the chapters study, respectively, the design of individual (or micro-) stress tests, systemic (or macro-) stress tests, and the combination of stress tests and conventional risk management tools. Finally, given the diversified knowledge required to carry out a stress testing exercise, this thesis is related to several fields of the literature. With this respect, the second chapter performs microeconomic and econometric analyses, the third chapter relies on macroeconomic and financial accounting, and the final chapter uses statistics and mathematics.

References

Acharya, A., R. Engle, and D. Pierret (2014): Testing Macroprudential Stress Tests: The Risk of Regulatory Risk Weights, *Journal of Monetary Economics*, 65(2014), 36-53.

Aikman, D., P. Alessandri, B. Eklund, P. Gai, S. Kapadia, E. Martin, N. Mora, G. Sterne, and M. Willison (2009): Funding Liquidity Risk in a Quantitative Model of Systemic Stability, *Bank of England*, Working Paper No. 372.

Alessandri, P., P. Gai, S. Kapadia, N. Mora, and C. Puhr (2009): Towards a Framework for Quantifying Systemic Stability, *International Journal of Central Banking*, 5(3), 47-81.

Alexander, C. and E. Sheedy (2008): Developing a Stress Testing Framework Based on Market Risk Models, *Journal of Banking and Finance*, 32(10), 2220-36.

Basel Committee on Banking Supervision (BCBS) (2012): Peer Review of Supervisory Authorities' Implementation of Stress Testing Principles, April 13.

Basel Committee on Banking Supervision (BCBS) (2009): Principles for Sound Stress Testing Practices and Supervision, Consultative Document, May 20.

Bee, M. (2001): Mixture Models for VaR and Stress Testing, *ALEA Tech Reports*, Università di Trento, No. 12.

Berkowitz, J. (2000): A Coherent Framework for Stress-Testing, *Journal of Risk*, 2(2), 5-15.

Blaschke, W., M.T. Jones, G. Majnoni, and S. Martinez Peria (2001): Stress Testing of Financial Systems: an Overview of Issues, Methodologies, and FSAP Experiences, *International Monetary Fund*, Working Paper No. 01/88.

Board of Governors of the Federal Reserve (FRB) (2014): Dodd-Frank Act Stress Test 2014: Supervisory Stress Test Methodology and Results, March 24.

Board of Governors of the Federal Reserve (FRB) (2013a): Comprehensive Capital Analysis and Review 2013: Assessment Framework and Results, March 14.

Board of Governors of the Federal Reserve (FRB) (2013b): Dodd-Franck Act Stress Test 2013: Supervisory Stress Test Methodology and Results, March 7.

Board of Governors of the Federal Reserve (FRB) (2012): Comprehensive Capital Analysis and Review 2012: Methodology and Results for Stress Scenario Projections, March 13.

Board of Governors of the Federal Reserve (FRB) (2011): Comprehensive Capital Analysis and Review: Objectives and Overview, March 18.

Board of Governors of the Federal Reserve (FRB) (2009): The Supervisory Capital Assessment Program: Overview of Results, May 7.

Borio, C., M. Drehmann, and K. Tsatsaronis (2012): Stress Testing Macro Stress Testing: Does it Live Up to Expectations? *Bank for International Settlements*, Working Paper No. 369.

Borio, C. and M. Drehmann (2009): Towards an Operational Framework for Financial Stability: "Fuzzy" Measurement and its Consequences, *Central Bank of Chile*, Working Paper No. 544.

Boss, M.G. (2008): Stresstests Für das Osterreichische FSAP-Update 2007: Methodik, Szenarien und Ergebnisse, Oesterreichische Nationalbank, *OeNB Finanzmarktsstabilitätsbericht*, 15, 72-99.

Breuer, T. and I. Csiszár (2013): Systematic Stress Tests with Entropic Plausibility Constraints, *Journal of Banking and Finance*, 37(5), 1552-9.

Breuer, T., G. Krenn, and F. Pistovčák (2002): Stress Tests, Maximum Loss, and Value at Risk, in: B. Britzelmaier, S. Geberl, M. Menichetti, and H.-R. Kaufmann (Eds.), Regulierung oder Deregulierung der Finanzmärkte: Strategieimplikationen für Finanzdienstleister, Physica Verlag (Springer), 301-14.

Breuer, T. and G. Krenn (1999): Stress Testing: Guidelines on Market Risk, Volume 5, Oesterreichische Nationalbank.

Čihák, M. (2007): Introduction to Applied Stress Testing, Technical Report, *International Monetary Fund*, Working Paper No. 07/59.

Committee of European Banking Supervisors (CEBS) (2010a): CEBS Guidelines on Stress Testing, August 26.

Committee of European Banking Supervisors (CEBS) (2010b): Results of 2010 EU-Wide Stress Testing Exercise, July 23.

Committee of European Banking Supervisors (CEBS) (2009): CEBS Press Release on the Results of the EU-Wide Stress Testing Exercise, October 1.

Counterparty Risk Management Policy Group (CRMPG III) (2008): Containing Systemic Risk: The Road to Reform – The Report of the CRMPG III, August 6.

Engle, R., E. Jondeau, and M. Rockinger (2014): Systemic Risk in Europe, *Review of Finance*, forthcoming.

European Banking Authority (EBA) (2014a): Final Templates for the 2014 EU-Wide Stress Test, August 20.

European Banking Authority (EBA) (2014b): Common Methodology and Scenario for 2014 EU-Banks Stress Test, April 29.

European Banking Authority (EBA) (2011): Results of 2011 EU-Wide Stress Testing Exercise, July 15.

Greenlaw, D., A. Kashyap, K. Schoenholtz, and H. Shin. (2012): Stressed Out: Macroprudential Principles for Stress Testing, *Chicago Booth School of Business*, Working Paper No. 71.

Haldane, A. (2009): Why Banks Failed the Stress Test, Speech to Marcus-Evans Conference on Stress Testing, London, February 9-10.

Hirtle, B., T. Schuermann, and K. Stiroh (2009): Macroprudential Supervision of Financial Institutions: Lessons from the SCAP, *Federal Reserve Bank of New York*, Staff Report No. 409.

International Monetary Fund (IMF) (2012): Macrofinancial Stress Testing, Principles and Practices, Working Paper, August 22.

Institute of International Finance (IIF) (2008): Final Report of the IIF Committee on Market Best Practices: Principles of Conduct and Best Practice Recommendations, Financial Services Industry Response to the Market Turmoil of 2007-2008, July 21.

Kim, J. and C.C. Finger (2000): A Stress Test to Incorporate Correlation Breakdown, *RiskMetrics Journal*, 1, 61-75.

Kupiec, P.H. (1998): Stress-Testing in a Value at Risk Framework, *Journal of Derivatives*, 6(1), 7-24.

McNeil, A.J. and A.D. Smith (2012): Multivariate Stress Scenarios and Solvency, *Insurance: Mathematics and Economics*, 50(3), 299–308.

Skilling, J. (2006): Nested Sampling for General Bayesian Computation, *Bayesian Analysis*, 1(4), 833–60.

Skilling, J. (2004): Nested Sampling, in: R. Fischer, R. Preuss, U.V. Toussaint (Eds.), American Institute of Physics Conference Series, 395–405.

van den End, J.W. (2012): Do Basel III and Unconventional Monetary Policy Work? *Applied Financial Economics*, 22(15), 1233–57.

van den End, J.W. (2010): Liquidity Stress-Tester: A Macro Model for Stress-Testing Banks' Liquidity Risk, *CESifo Economic Studies*, 56(1), 38–69.

Résumé du premier chapitre
Stress Tests : Bilan d'une Décennie de Pratique

Au sens restreint, « un stress test » désigne un ensemble d'outils permettant d'évaluer l'impact potentiel d'événements « sévères mais plausibles » sur un portefeuille financier. La nature de ce portefeuille détermine la catégorie du test en question. On distingue ainsi : (i) les stress tests individuels ou « micro stress tests », portant sur les portefeuilles détenus par les unités d'une même institution financière et (ii) les stress tests systémiques ou « macro stress tests », portant sur des institutions qui relèvent d'un même système financier. Le test permet une réévaluation du portefeuille, à partir d'un ensemble de « scénarios » portant sur les trajectoires futures des facteurs de risque sous-jacents. Cette définition peut être étendue pour inclure l'intégration des résultats du test dans le processus de gestion des risques lié au portefeuille stressé. Ce processus inclut l'analyse des résultats et de l'information obtenus, l'évaluation des réactions éventuelles et la prise de décision. De ces deux définitions, trois éléments peuvent être retenus pour désigner « les critères de pertinence » d'un stress test, à savoir : les conditions de sévérité et de plausibilité, requises pour le scénario de stress, ainsi que la portée et le mode d'utilisation des résultats.

Un scénario de stress est dit sévère lorsqu'il permet de capter les risques les plus défavorables au portefeuille stressé. La sévérité varie d'un portefeuille à l'autre, en fonction de la composition du portefeuille, du temps et des risques potentiels. La définition de scénarios sévères constitue un premier défi dans la mise en œuvre de stress tests pertinents. Toutefois, quelle que soit la nature des mouvements retenus dans le scénario de stress (prix, rendement, volatilité, corrélation, etc.), il est souvent possible de concevoir des mouvements encore plus sévères. Quel(s) scénario(s) peut-on alors retenir parmi toutes ces alternatives ; ou du moins, comment restreindre leur nombre ? Le critère de plausibilité apporte une première réponse à cette question.

La plausibilité d'un scénario désigne le niveau de confiance associé aux mouvements de facteurs de risque considérés. Il est souvent mesuré par la probabilité d'occurrence des mouvements de facteurs. Plus cette probabilité est élevée, plus le scénario est plausible. Cette définition permet d'écarter les scénarios qui présentent une faible probabilité. Elle suscite, toutefois, deux autres questions :

(i) comment définir la probabilité du scénario, puisque différentes probabilités peuvent être associées à un même événement, en fonction de la distribution sous-jacente retenue ?

(ii) à partir de quel niveau de probabilité un scénario de stress peut être considéré comme plausible ?

La réponse à la première question consiste à mettre en place des tests de spécification des fonctions de distribution. La réponse à la seconde question permet de définir l'ensemble des scénarios dont la probabilité d'occurrence est supérieure ou égale à un « seuil de plausibilité » prédéfini. Cependant, les scénarios à forte probabilité – et donc, les plus plausibles – sont parfois les moins sévères – au moins, pour des portefeuilles linéaires. Dans ce cas, un arbitrage sévérité-plausibilité est nécessaire. Il peut être opéré via trois approches alternatives :

(i) définir un mouvement de facteurs de risque et calculer sa probabilité d'occurrence ;

(ii) définir une probabilité et déduire les mouvements de facteurs correspondants ; ou

(iii) définir un seuil de plausibilité et chercher, parmi l'ensemble des mouvements de facteurs correspondants, celui (ou ceux) qui présente(nt) l'impact le plus sévère sur le portefeuille stressé. Cette méthode est appelée « recherche systématique des pires scénarios ».

Les deux premières approches partagent deux limites essentielles : (i) le risque d'omettre des scénarios plus sévères et/ou plus plausibles et (ii) la possibilité d'associer différentes probabilités au même scénario ou différents mouvements pour une même probabilité, en fonction de la distribution considérée. En outre, la deuxième approche présente une ambigüité liée au choix d'un seul scénario parmi l'ensemble des possibles. La troisième approche, fondée sur l'arbitrage sévérité-plausibilité, permet d'éviter la plupart de ces inconvénients. Elle reste néanmoins tributaire de l'incertitude liée au choix de la distribution. A cela, s'ajoutent les incertitudes liées au choix du modèle d'évaluation du portefeuille et de l'algorithme numérique de recherche du pire scénario. Nous appellerons « hypothèses de départ », toutes les hypothèses retenues pour définir les niveaux de sévérité et de plausibilité du scénario de stress. Un scénario de stress alliant sévérité et plausibilité est alors dit « crédible » ; une désignation également appliqué au stress test en question.

Le stress test devient « utile » à la gestion du portefeuille stressé lorsqu'il permet d'intégrer les résultats obtenus au processus de décision. Ce troisième critère de pertinence est d'autant plus important que les coûts de mise en œuvre de l'exercice soient élevés (en termes de temps, de moyens humains et matériels, etc.). Il continue toutefois de faire défaut à la plupart des tests réalisés, malgré les recommandations de régulateurs nationaux et internationaux. Ci-après, nous analysons les aspects méthodologiques et fonctionnels de la mise en place d'un stress test, en présentant les principales étapes et en mettant l'accent sur les obstacles récurrents, les solutions proposées et les défis subsistants. L'analyse est appliquée aux principaux exercices menés sur des institutions et systèmes financiers des pays développés sur la période 2000-2014

Dans un scénario de stress, la conception des hypothèses de départ est plus décisive que les applications numériques qui permettant leur implémentation au cours des phases empiriques de l'exercice. Ainsi, la mise en place d'hypothèses plus objectives permet d'asseoir la crédibilité du scénario, d'entériner les résultats du test, de réduire les écarts d'appréciation et d'améliorer son utilité.

Un micro stress test est conduit par l'institution concernée ou par l'une de ses unités ou branches d'activité. Dans ce cas, le test peut s'inscrire : (i) dans une démarche volontaire et dans le cadre des processus internes de gestion des risques, (ii) en réponse à une demande émanant d'une autorité de tutelle ou de supervision, ou (iii) dans un cadre mixte regroupant des intervenants internes et externes. Les tests conduits sous l'égide du CEBS sont un exemple de ces derniers, où les banques centrales ou les superviseurs établissent les scénarios et laissent aux institutions privées le soin de choisir les modèles et les données appropriés. Un macro stress test peut être directement mené par la banque centrale ou le superviseur, en utilisant leur propres modèles, ainsi que des données individuelles ou agrégées des institutions concernés. Au-delà des difficultés liées à la conception de scénarios crédibles, ces exercices sont confrontés à plusieurs défis relatifs à la collecte et à l'harmonisation des données individuelles, la spécification des liens intra-sectoriels, la gestion des aspects dynamiques des événements de stress (ex. contagion, effets de retour et de second tour), l'agrégation des résultats individuels, etc.

La réussite du test est liée aux hypothèses de départ. Cependant, compte tenu du nombre d'intervenants dans cette phase, la mise en place de ces hypothèses peut s'avérer délicate et

souvent longue. Des critères subjectifs peuvent ainsi être retenus par les parties prenantes. Les gestionnaires de risques ont souvent tendance à surestimer les risques potentiels, en associent une forte probabilité aux scénarios les plus sévères. En face, la Direction est plus optimiste, considérant inappropriée la prise en compte de scénarios jugés peu vraisemblables. Le gestionnaire doit alors convaincre sa Direction et les autres intervenants du bien-fondé et de l'objectivité des hypothèses retenues. Cela peut souvent s'avérer délicat, en raison du caractère souvent subjectif de certaines hypothèses et des implications pratiques du choix des scénarios - en termes de coûts d'opportunité notamment. A ces problèmes internes, s'ajoutent des considérations externes. Les effets liés à la publication des résultats et à leur communication aux autorités de supervision peuvent constituer des critères pouvant influencer le choix des hypothèses de départ.

Afin de confirmer et/ou de modifier les hypothèses de départ, le concepteur du test est appelé à réagir quasi-spontanément aux résultats du test. Les pratiques montrent, toutefois, une relative lenteur dans les réactions, en raison notamment des limites des systèmes d'information et des techniques utilisées dans la plupart des institutions. Pour être pertinent, le processus doit ainsi être itératif. La tenue d'une documentation appropriée sur chaque phase de l'exercice est un élément crucial pour réduire les coûts et gagner en efficacité. La crise de 2007-2009 a confirmé les limites d'intégration des résultats dans le processus de gouvernance, notamment par le manque de débats internes sur les facteurs de risque, les opportunités et les coûts d'investissement, les stratégies de cession et de couverture, etc. A ces problèmes s'ajoutent des limites liées aux aspects organisationnels des stress tests. En effet, dans la plupart des cas, ces tests sont souvent considérés de façon désagrégée entre les unités d'activités et les classes de risque.

Les répercussions du crash boursier de 1987 ont révélé l'urgence de la mise en place d'outils plus performants pour la gestion des risques financiers, pouvant garantir une couverture suffisante contre des événements de pareille ampleur. C'est ainsi qu'est née la *Value-at-Risk* ou VaR, comme nouvel outil de gestion, alliant simplicité et flexibilité. Cependant, après plusieurs années de pratique, il s'est avéré que des erreurs ont été induites par l'utilisation de la VaR, notamment en périodes de forte agitation des marchés. C'est dans ce cadre que sont introduits les premiers stress tests, conçus pour une gestion des risques extrêmes. Ces exercices apportent deux principales innovations : (i) explorer en détail les queues de distribution des rendements de portefeuilles et de facteurs de risque, au-delà des seuils de confiance retenus dans le calcul de la VaR et (ii) relier, de façon flexible et explicite, tout scénario de stress à son impact potentiel sur le portefeuille stressé. Ce dernier point permet d'identifier les sources potentielles de risque et d'évaluer la contribution de chaque scénario.

Le champ d'analyse d'un stress test comprend le choix des portefeuilles, des facteurs de risque et de l'horizon du test. Dans la pratique, plus des trois quarts des tests se rapportent à des portefeuilles de négociation. L'utilisation de techniques relativement simples (*market-to-market*) dans l'évaluation du profil de risque de ce type de portefeuilles, ainsi que la fréquence et la régularité des données constituent les principaux arguments en faveur de ce choix. A l'inverse, les tests visant des portefeuilles de produits structurés ou des positions hors-bilan sont moins répandus. La modélisation du profit, des taxes et des dividendes est également marginalisée dans les tests conduits, en raison de la granularité et du poids relativement faible de ces aspects dans le profil du portefeuille.

Par ailleurs, la plupart des études et enquêtes conduites auprès des institutions financières révèlent un usage récurrent de certains facteurs risque. Ainsi, pour l'actif du bilan, le risque de marché est souvent ventilé en facteurs de risque de prix, (actions, matières premières, produits

dérivés, propriétés, etc.) et en facteurs de risque de taux (intérêt, change, primes de risque). Certains exercices, macro stress tests essentiellement, incluent aussi la liquidité de marché comme facteur de risque. Toutefois, l'introduction du risque de liquidité présente des difficultés pratiques ; ce qui explique le nombre réduit des tests de cette nature. Les scénarios fondés sur des hypothèses portant sur des variations de la volatilité ou des corrélations des facteurs de risque sont également rares.

Au passif des bilans, les facteurs de risque sont souvent assimilés à des risques de liquidité de financement. Ces risques se concrétisent notamment via : le comportement du passif à court terme (ex. dépôts, emprunts au jour le jour, lignes d'engagements interbancaires), la qualité des contreparties (dégradation des notations, liquidité de marché), les coûts de refinancement (taux d'intérêt, primes de risque), la qualité du collatéral (montants exigés et décotes appliquées), les conditions d'accès à la monnaie centrale (restrictions spécifiques ou généralisées), et l'arsenal réglementaire et juridique en place (assurance-dépôts). Cette classe de risque est mieux explorée que la liquidité de marché, grâce notamment à l'introduction de méthodes d'évaluation relativement simples. Les modèles afférant aux spirales de liquidité restent, eux, insuffisamment développés. Enfin, contrairement au risque marchand de crédit, traité comme risque de marché, la prise en compte du risque non marchand de crédit est moins fréquente. Il en est de même pour le risque opérationnel.

L'impact des facteurs de risque sur le portefeuille est évalué sur une période déterminée. L'horizon d'un stress test consiste à définir : (i) la durée de persistance des mouvements appliqués aux facteurs de risque, durant laquelle de nouveaux éléments peuvent être introduits et (ii) la période durant laquelle l'impact de ces mouvements sera enregistré. En l'absence de conventions sur l'horizon optimal, on considère des horizons plus courts pour le risque de marché. Pour les risques de crédit, l'horizon est généralement plus important, permettant de prendre en compte la cristallisation des événements de stress, leur transmission au portefeuille de crédit et la prise de décisions adéquates. Pour un macro stress test, l'horizon est généralement de plusieurs trimestres ou de plusieurs années. Cela est en partie dû à la fréquence des données et des modèles utilisés. Un horizon long permet également de capter l'évolution des expositions et la transmission des différentes classes de risque aux portefeuilles, les temps de réaction publique et privée et les externalités éventuelles (effets de second tour, contagion, effet de retour sur l'économie réelle, etc.).

L'étape suivante dans la mise en œuvre du test consiste à définir, pour un ou plusieurs facteurs, la taille et la probabilité du mouvement à appliquer et l'évolution de ce dernier à travers l'horizon du test. Il faut ensuite mesurer l'impact de ces mouvements sur le portefeuille, à l'aide de modèles d'évaluation adaptés. Le premier point regroupe les principales composantes du scénario, définies par un « modèle de risque ». Le deuxième, désigne « le modèle d'évaluation ». Cette phase est, en général, la plus délicate et la plus controversée dans un stress test. La mise en œuvre d'un scénario consiste à définir de façon quantitative l'amplitude des variations à appliquer – par simulation – aux facteurs de risque (le choc initial), ainsi que la façon dont ces mouvements persistent, évoluent et/ou interagissent durant l'horizon du test (l'après-choc). Dans certains tests, la probabilité d'occurrence du scénario est également définie.

Le choix des scénarios de stress repose sur des critères théoriques et opérationnels. La définition des paramètres du scénario peut être établie à partir : des variations historiques passées des facteurs de risque (scénarios historiques), d'hypothèses théoriques (scénarios hypothétiques), ou d'une combinaison des deux (scénarios hybrides). Le choix entre un scénario historique et hypothétique relève notamment de : (i) l'adéquation au portefeuille et au contexte considérés (lié à l'arbitrage sévérité-plausibilité établi), (ii) la conformité du scénario aux

objectifs fixés (permettant une meilleure mise en valeur du critère d'utilité), et (iii) les coûts de conception (en termes de temps et de moyens alloués). Ainsi, un scénario satisfaisant les deux premiers critères doit également présenter un coût de conception raisonnable, favorisant son adoption et sa reconduction.

Reproduire un scénario historique sur la base d'épisodes passés peut permettre de réduire les coûts d'analyse et la subjectivité de l'exercice. Ces scénarios peuvent, néanmoins, souffrir d'un manque d'adéquation au portefeuille présent (nouveaux produits, nouvelles expositions, etc.), au contexte spatio-temporel dans lequel est conduit le test (changements structurels, nouvelles phases d'interdépendance des facteurs, etc.) ou aux objectifs fixés. Un scénario hypothétique, fondé sur des hypothèses flexibles, offre alors une meilleure prise en compte de ces éléments. Toutefois, le coût de ces scénarios représente une contrainte matérielle de taille, à laquelle s'ajoute le défi humain consistant à démontrer l'objectivité des hypothèses considérées. Un arbitrage flexibilité-coût est alors souvent nécessaire. Une première solution peut provenir des scénarios hybrides. Ces derniers sont conçus sur la base d'événements historiques, sans pour autant constituer une réplique conforme d'un épisode spécifique.

Les résultats du test fournissent une information supplémentaire sur les sources potentielles de risque pour le portefeuille, ainsi que le rapport entre le rendement espéré et le niveau de risque affectant chaque position. Même si leur usage est encore limité dans la pratique, ces informations peuvent être exploitées à des fins de gestion interne au niveau commercial, managérial et de la direction générale. Les résultats peuvent aussi être exploités par d'autres acteurs économiques, et notamment les superviseurs, les partenaires (sociaux, économiques et commerciaux) de l'institution financière et les agences de notation.

Ainsi, la communication des résultats peut être exploitée à différentes finalités par les autorités de supervision :

(i) dans un cadre de supervision individuelle, en permettant d'identifier les diverses expositions et vulnérabilités. Les résultats du test permettent, par exemple, de révéler la situation des ratios réglementaires et d'inciter l'institution à redresser ces différents indicateurs de risque. La culture de risque d'une institution apparait également à travers les scénarios réalisés et les résultats obtenus. Enfin, un résultat de stress peut être utilisé par le superviseur pour évaluer la solidité financière d'une institution et sa capacité à respecter ses engagements. Ces critères détermineront, en partie, les conditions de refinancement sur le marché monétaire, par exemple, tout en renforçant l'efficience de ce marché.

(ii) dans un cadre de régulation, de gestion prudentielle et de maintien de la stabilité systémique, en identifiant, à partir des expositions individuelles, les risques auxquelles sont confrontés le système financier et l'économie réelle.

(iii) dans l'évaluation de l'impact et de la pertinence des différentes politiques et programmes de gestion, établis par l'Etat, le superviseur et les institutions privées. Des mesures supplémentaires ou alternatives peuvent ainsi être suggérées.

En fonction des résultats des tests et des expositions révélées, certains investisseurs peuvent décider de réallouer leur fonds vers des valeurs moins exposées aux risques considérés. D'autres préfèreront céder les titres dont ils disposent, ou encore refuser de financer les institutions fortement exposées. Une communication maîtrisée autour du stress tests est primordiale face à ce genre de public, parfois peu averti, pour se prémunir contre une exploitation inappropriée des résultats.

Enfin, les résultats peuvent être exploités par les agences de notation, en intégrant les conclusions des tests aux critères de notation des portefeuilles stressés et de leurs contreparties. Ainsi, par exemple, des résultats révélant une forte vulnérabilité vis-à-vis d'un risque particulier peuvent conduire à dévaluer la note de l'institution pour ce type d'exposition et/ou la note d'une contrepartie particulière. Rappelons cependant, que le résultat du test n'est pas la conséquence d'une situation prévue, mais celle d'une situation possible. Ainsi, au lieu de dévaluer la note d'une contrepartie, l'agence de notation peut maintenir cette dernière et la placer en perspective négative, car présentant un (assez fort) risque potentiel de défaut. Ceci peut permettre d'éviter les variations brutales de notations observées sur la dernière période.

Au-delà des difficultés relatives à la conception de scénarios, l'exploitation des résultats du test n'est pas facile à mettre en œuvre au sein même de l'institution concernée. En effet, le stress test est une mesure de risque qui s'ajoute aux outils déjà en place. En présence de résultats d'exercices de stress, un gestionnaire de risques peut ainsi se retrouver face à une situation ambigüe : faut-il tenir compte des résultats de stress tests fondés sur des événements sévères, des résultats de la VaR valables en périodes normales des marchés, ou des deux mesures à la fois ? Si les résultats des stress tests sont pris en compte, par quels moyens associer les deux approches ? Dans la pratique, la plupart des tests sont conduits de façon autonome et leurs résultats exploités séparément de ceux des autres outils de gestion des risques.

L'élément central dans la réalisation d'un stress test réside dans la conception des scénarios. Cette phase, qui consiste à adapter une modélisation adéquate des facteurs de risque et de leur impact sur les portefeuilles stressés, constitue encore le principal défi de tests réalisés. Dans les deux prochains chapitres, nous présenterons deux cadres flexibles de modélisation des trois paramètres des scénarios de stress. Les deux premiers paramètres sont analysés au chapitre 2, avec un modèle de risque appliqué à un portefeuille financier privé. Le troisième paramètre fera l'objet du chapitre 3, avec un modèle d'évaluation du risque systémique porté par le secteur bancaire français. L'intégration des résultats du test au processus interne de gestion des risques est explorée au chapitre 4.

Résumé du chapitre 2
Modélisation des Risques Extrêmes
Une Approche Séquentielle TVE – Copules-paires

Les paramètres du scénario de stress sont définis à l'aide d'un modèle de risque. A partir d'une série d'hypothèses, ce model permet : (i) de spécifier et d'estimer la distribution, la dynamique et la structure de dépendance des facteurs de risque et (ii) de générer un ensemble de scénarios par la simulation du modèle estimé. Une procédure de sélection permet ensuite de définir les scénarios crédibles parmi l'ensemble des scénarios. L'impact des scénarios sur le portefeuille stressé est, enfin, capté par un modèle d'évaluation. Ainsi, le résultat du test est conditionné par le choix préalable du modèle de risque regroupant un ensemble de spécifications théoriques supposées reproduire les propriétés empiriques des facteurs de risque sous-jacents. Dans un portefeuille financier, les rendements des actifs – contenus dans le portefeuille ou sous-jacents aux actifs contenus – constituent les principaux facteurs de risque. L'analyse des séries de rendements revêt donc un caractère central dans les stress tests appliqués aux institutions financières.

La littérature a mis en valeur un ensemble de propriétés statistiques empiriques ou « faits stylisés », liées aux rendements financiers et à leur structure de dépendance. Pour capter quantitativement ces faits stylisés, des modèles alternatifs sont retenus et estimés sur des échantillons de données. Des exercices de simulations sont alors conduits, parmi lesquels les prévisions et les scénarios de stress tests constituent deux applications particulières. Sur la période récente, des erreurs de prévision récurrentes et de plus en plus conséquentes ont été relevées. Les rendements supposés sévères par les scénarios de stress se sont, par ailleurs, avérés anodins, alors que des pertes non-anticipées et plus sévères ont été enregistrées. Ces erreurs révèlent les limites des modèles en place à reproduire les propriétés des rendements. Les rendements extrêmes sont particulièrement affectés par ces erreurs. Le manque de flexibilité des modèles utilisés constitue la principale raison pour expliquer ces échecs. En effet, pour des considérations pratiques (maniabilité, interprétation, disponibilité des données, coûts en termes de temps et de moyens, etc.), des spécifications simples ont souvent été considérées, conduisant à négliger une ou plusieurs caractéristiques des séries financières. La nécessité d'une modélisation plus adaptée aux données est alors plus que d'actualité.

Alors que l'inadéquation de la loi normale à reproduire les distributions empiriques des séries financières est maintenant admise, aucun critère objectif ne permet de distinguer les distributions alternatives proposées par la littérature. L'évolution de la performance de ces distributions, en fonction de l'actif et de la période considérés, constitue le principal obstacle à un tel choix. La Théorie des Valeurs Extrêmes (TVE) apporte une solution consistante à ce problème, en proposant une spécification directe des queues de distribution au-delà d'un quantile suffisamment élevé. D'après le théorème central limite, et sous certaines conditions, les distributions des valeurs extrêmes d'une variable aléatoire présentent un comportement asymptotique analogue quelle(s) que soi(en)t la variable et la distribution considérée(s). Ainsi, au lieu de spécifier l'ensemble de la distribution – au risque d'une mauvaise spécification – seuls les comportements asymptotiques sont modélisés. Des distributions de valeurs extrêmes sont utilisées à cet effet. Elles présentent de meilleures performances pour estimer les queues de distribution par rapport aux distributions paramétriques.

Dans un cadre multivarié, la structure de dépendance entre plusieurs variables (ex. rendements) est modélisée par leur distribution jointe. Cette dernière, peut être construite de

deux façons. La première, dite méthode directe, est une généralisation au cadre multivarié des distributions uni-variées. Elle consiste à imposer une forme analytique particulière de la distribution jointe et à estimer les paramètres correspondants. La matrice de variance-covariance fait partie des paramètres à estimer. L'approche alternative, dite méthode indirecte, consiste à décomposer la distribution jointe en : (i) distributions uni-variées ou marginales et (ii) une fonction multivariée appelée « copule ». Les distributions marginales captent les comportements individuels des variables, tandis que la copule capte la structure de dépendance du système multivarié. Cette approche permet, notamment, de dissocier la forme de la structure de dépendance de celles des distributions marginales et de généraliser le concept de corrélation linéaire à celle de dépendance – qui est une fonction non-linéaire.

La modélisation de la structure de dépendance repose sur la spécification de la copule multivariée (également appelée « méta-copule »). Cependant, la plupart des formes existantes manquent de performance quand il s'agit de reproduire la structure de dépendance des rendements financiers extrêmes. Certaines copules sont peu flexibles, alors que d'autres présentent des problèmes pratiques dus à leur complexité. Pour contourner ces difficultés, une approche alternative a été proposée pour décomposer la copule multivariée en copules bi-variées (conditionnelles et non conditionnelles) dites « copules-paires ». Cette technique personnalise la spécification des structures de dépendance à chaque paire de variables, tout en facilitant l'estimation – désormais fondée sur des distributions bi-variées. Cette flexibilité supplémentaire apporte une meilleure performance à cette approche, par rapport aux modèles fondés sur des méta-copules multivariées ou sur la méthode directe.

Le modèle R-vine représente une spécification généralisée des modèles fondés sur les copules-paires. Dans ce chapitre, nous considérons le modèle R-vine pour modéliser la structure de dépendance des séries de rendements financiers. Via l'approche copula-GARCH, nous associons le modèle R-vine à un modèle EGARCH relatif aux distributions marginales des rendements individuels. Pour une modélisation spécifique des rendements extrêmes, nous avons étendu cette approche, en deux points centraux, en introduisant : (i) la TVE, pour une meilleure précision dans la modélisation des queues de distribution des séries de rendements et (ii) un modèle ARMA-EGARCH, pour une spécification flexible des processus de moyenne-variance conditionnelles des séries ; ce qui permet de capter la plupart des faits stylisés évoqués ci-dessus.

Notre échantillon de données est constitué de trois classes de séries journalières : deux classes de séries financières et une classe de séries non financières. La première classe inclut le CAC40 (France), le DAX30 (Allemagne), le FTSE100 (Royaume-Uni), le SP500 (Etats-Unis) et le SPTSX (Canada). La deuxième, inclut les taux de change nominaux des six paires de devises les plus négociées sur le marché des changes au comptant. Il s'agit des cotations à l'incertain, contre le dollar américain (USD) de l'euro (USD/EUR), du yen (USD/JPY), de la livre sterling (USD/GBP), du dollar australien (USD/AUD), du franc suisse (USD/CHF) et du dollar canadien (USD/CAD). La troisième série comporte les prix au comptant de cinq matières premières, relevant de cinq catégories distinctes : le Brent de la mer du Nord (énergies), l'or (métaux précieux), le cuivre (métaux de base), le blé (matières agricoles) et le maïs (bioénergies). Les séries d'indices boursiers sont collectées du 9/7/1987 au 18/7/2013 ; celles des taux de change, entre le 4/1/1999 et le 18/7/2013; et les prix de matières premières, entre 5/7/1993 le 18/7/2013.

Pour chaque série, nous avons considéré deux sous-périodes : la première est réservée à l'estimation ; la seconde, formée des 250 dernières observations, pour la prévision et l'évaluation *out-of-sample*.

Après avoir vérifié et confirmé la présence des faits stylisés dans l'échantillon, nous présentons un modèle séquentiel à deux niveaux pour spécifier le comportement des séries de rendements financiers. Dans un premier temps, nous présentons le modèle TVE-EGARCH, offrant une spécification semi-paramétrique des distributions marginales des rendements individuels. Nous définissons, ensuite, la structure de dépendance du système multivarié, à l'aide de copules-paires associées aux distributions marginales ainsi constituées.

L'application de la TVE requiert un filtre préalable des séries de rendements. A cet effet, un modèle ARMA-EGARCH, avec une distribution – temporaire – normale pour les résidus est retenu pour spécifier la moyenne et la variance conditionnelles des rendements. Pour une modélisation précise des queues de distribution, nous appliquons, ensuite, la TVE aux résidus filtrés du modèle ARMA-EGARCH, via la méthode Peaks Over Threshold (POT), selon laquelle, au-delà d'un seuil suffisamment élevé, la distribution des excès suit une distribution de Pareto Généralisée (GPD). Les seuils extrêmes, supérieur et inférieur, sont fixés, tel que la proportion des excès représente, de chaque côté de la distribution, 10% du nombre total d'observations. Les observations restantes sont spécifiées par une fonction non-paramétrique, obtenue par une interpolation linéaire. La distribution marginale semi-paramétrique est alors obtenue par un mélange : (i) des distributions de valeurs extrêmes et (ii) d'une distribution empirique pour les valeurs situées entre les seuils extrêmes.

A partir des distributions marginales obtenues, nous modélisons la structure de dépendance du système multivarié par un modèle R-vine. Ce dernier est estimé en adaptant la méthode de sélection/estimation séquentielle, qui permet de définir, simultanément, les copules-paires et le schéma de décomposition appropriés aux données utilisées. Plusieurs formes de dépendance sont testées à l'aide de douze copules elliptiques et archimédiennes bi-variées.

Pour identifier les ordres de retard dans les modèles ARMA-EGARCH, nous avons estimé l'ensemble des combinaisons possibles pour un ordre de retard maximal de 3. Le critère d'information d'Akaike a été retenu pour choisir la meilleure spécification parmi toutes les combinaisons. Pour les modèles retenus, la plupart des paramètres estimés sont significatifs au seuil de confiance de 99%. Les moyennes non-conditionnelles sont nulles partout. L'hétéroscédasticité des séries est confirmée, avec : (i) des effets ARCH présents dans toutes les séries et (ii) des effets GARCH présents dans plus des deux tiers des cas. Dans l'analyse empirique, nous avons montré que l'autocorrélation de la plupart des séries est non-significative pour des ordres de retard de 25 et 50. Il semble toutefois, que celle-ci soit présente pour des ordres de retard inférieurs. En effet, les paramètres du 2e et 3e ordre de retard des processus AR et MA sont significatifs ; mieux, les modèles correspondants constituent les meilleures spécifications parmi l'ensemble des combinaisons possibles. Même si de meilleures spécifications peuvent encore être trouvées en relaxant la contrainte imposée aux ordres de retard, nous avons retenu ces modèles pour deux considérations principales : un souci de parcimonie et parce qu'ils s'avèrent suffisants pour filtrer les séries de rendements. En effet, le test de Ljung-Box (LB) rejette l'hypothèse nulle d'autocorrélation et d'hétéroscédasticité des résidus filtrés, alors que le test ARCH rejette l'hypothèse nulle de présence d'effets ARCH. Ces résultats sont valables avec des niveaux de confiance allant de 10% à 100% et pour des ordres de retard allant de 1 à 50.

Avant d'estimer les distributions marginales à partir des résidus filtrés, nous avons analysé l'impact de la présence d'autocorrélation et d'hétéroscédasticité sur les distributions des séries de rendements. Nous avons ainsi comparé, pour chaque série, les valeurs empiriques des coefficients de skewness et de kurtosis, relatifs à la distribution des résidus filtrés d'un côté, et à la distribution des rendements de l'autre. Trois conclusions notables peuvent être soulignées. Premièrement, les valeurs respectives des coefficients de skewness et de kurtosis varient sensible-

ment mais convergent vers 0 et 3 respectivement. Ce résultat confirme le fait stylisé selon lequel les distributions conditionnelles sont plus proches de la loi normale que les distributions non-conditionnelles correspondantes. Deuxièmement, seuls les rendements boursiers, qui présentaient les distributions non-conditionnelles les plus éloignées de la loi normale conservent des distributions conditionnelles leptokurtiques. Le reste des séries présentent désormais une distribution conditionnelle mésokurtique (voire platikurtique pour les taux de change). Troisièmement, les distributions conditionnelles demeurent, pour la plupart, asymétriques, même si le coefficient de skewness est souvent proche de zéro pour la plupart des rendements autres que boursiers. Ces résultats confirment le caractère élusif des formes de distribution des séries de rendements. Celles-ci varient entre les différentes classes de données, entre les séries d'une même classe, voire pour une même série. On voit ainsi le risque lié à toute démarche consistant à considérer une forme paramétrique (commune) préalablement définie. C'est ainsi que nous avons modélisé les distributions des séries de rendements à l'aide d'une approche semi-paramétrique.

Celle-ci est appliquée aux résidus filtrés des modèles ARMA-EGARCH. Le résultat confirme la forme leptokurtique et asymétrique de l'ensemble des distributions de rendements boursiers. Il est validé par les valeurs du paramètre de dispersion qui affiche en général, une valeur supérieure sur la queue inférieure. Les rendements de taux de change présentent, dans quatre cas sur six, des paramètres de forme négatifs (respectivement, positifs) sur la queue inférieure (respectivement, supérieure). Pour les distributions marginales de cette classe de rendements, on note une absence de queue épaisse inférieure et/ou un sens d'asymétrie instable. L'instabilité de la symétrie est également valable pour les rendements de matières premières, même si ceux-ci présentent au moins une queue épaisse.

L'adéquation des distributions marginales estimées aux propriétés des résidus filtrés est évaluée par quatre tes tests de *Goodness of fit* (GoF) : trois versions du test de GoF de Kolmogorov-Smirnov (KS), pour lesquels les hypothèses alternatives sont, respectivement, une distribution des PIT différente (KS), supérieure (KS+) et inférieure (KS-) à la loi uniforme, et le test de Anderson-Darling (AD) surpondérant les résultats pour les queues de distribution. Les résultats montrent un rejet de l'hypothèse alternative avec des niveaux de confiance allant de 64% à 100%, quels que soient le test et la série considérés. On conclut, ainsi, sur une (forte) performance des distributions considérées pour modéliser les distributions de rendements financiers. Par ailleurs, ces performances ont été comparées à celles de cinq distributions paramétriques, souvent utilisées dans la pratique, à savoir : normale, Student, Student asymétrique, hyperbolique généralisée et l'une de ses principales formes spécifiques, la distribution t asymétrique. Les résultats montrent une nette infériorité de ces distributions paramétriques par rapport aux distributions obtenues via la méthode semi-paramétrique, et ce quels que soient le test GoF et la série considérés. Certaines distributions paramétriques ne performent nulle part. La performance de certaines autres distributions varient en fonction de la classe de l'actif considéré.

Trois modèles R-vine sont estimés pour les trois classes de données. Les résultats confirment l'hétérogénéité des structures de dépendance bi-variées et multivariées, en fonction des séries considérées. Toutefois, quelle que soit la série, les résultats montrent que les rendements extrêmes présentent une forte dépendance (*à la* Student). Cette dernière est souvent symétrique, sauf pour deux paires de rendements de matières premières. Hormis, pour les séries de taux de change (copule Student), la dépendance conditionnelle extrême des rendements est souvent nulle (copules Gaussienne, Frank et BB8). Ainsi, seuls les rendements de taux de changes sont conjointement corrélés en périodes de stress et d'euphorie. Ce résultat est confirmé pour les structures de dépendance jointes des trois systèmes multivariés. Ces dernières sont symétriques, quel que soit la classe de données. Elles présentent une (forte) dépendance extrême jointe pour

les rendements de taux de change (copule Student) et une dépendance extrême jointe nulle pour les autres classes (copules Gaussienne et Frank). Ainsi, un portefeuille de devises, peu volatile en périodes d'accalmie, devient, en périodes de stress, plus risqué qu'un portefeuille d'actions ou de matières premières pourtant plus volatiles par nature.

Pour vérifier la performance du modèle de R-vine, nous avons simulé un échantillon de 1 million de réalisations multivariées pour les trois modèles estimés. Pour chaque échantillon, les coefficients de corrélation extrêmes ont été calculés. Ces derniers ont, ensuite, été comparés à : (i) des coefficients de corrélation empiriques calculées pour les résidus filtrés, et (ii) des coefficients de corrélation calculés pour un 1 million de réalisations simulées à partir de deux modèles à méta-copules (Student et gaussien) estimé pour les résidus filtrés. Trois principaux résultats peuvent être tirés. Premièrement, les estimations obtenues par les modèles R-vine et Student sont assez similaires et sont clairement distingués de ceux du modèle gaussien. Ce résultat est dû au fait que le modèle R-vine est essentiellement composé de copules-paires Student. Deuxièmement, le modèle R-vine semble être plus efficace pour capter la dépendance empirique des rendements extrêmes négatifs. Ainsi, il surpasse souvent (resp. toujours) le modèle Student (resp. gaussien). Troisièmement, dans la plupart des cas, le modèle gaussien semble dominer les deux autres sur les quartiles extrêmes supérieurs. Ce résultat peut s'expliquer comme suit. Certaines paires de rendement ont une structure de dépendance asymétrique. Toutefois, étant donné que cette asymétrie est souvent faible, elle est rarement captée par les tests statistiques qui fondent la procédure de sélection/estimation du modèle R-vine. Cette procédure conduit alors, dans la plupart des cas, à sélectionner des copules-paires symétriques. L'utilisation de copules symétriques peut donc conduire à une surestimation de la dépendance pour les quartiles extrêmes supérieurs (modèles R-vine et Student) ou à une sous-estimation des quartiles extrêmes inférieurs (modèles Student et gaussiens). La dynamique de la structure de dépendance peut également être une explication de ce résultat.

La performance des modèles à copules-paires dépasse celles des modèles basés sur des méta-copules. Ceci est vérifié en considérant, d'un côté, le modèle R-vine et deux de ces principales versions restreintes et, de l'autre, les modèles Student et gaussien. La comparaison est menée sur la base des log-vraisemblances des estimations, des critères d'information usuels et deux tests de spécification pour modèles imbriqués. Le modèle R-vine présente les meilleures propriétés statiques, parmi toutes les spécifications retenues. Nous avons, ensuite, comparé les propriétés dynamiques des modèles cités. Deux modèles DCC-GARCH, appartenant à l'approche directe de modélisation, ont également été inclus dans cette étape. La comparaison est basée sur les procédures de *backtesting* usuelles, pour des portefeuilles analytiques. Là encore, les résultats montrent de faibles performances des modèles gaussiens et des résultats conservateurs pour modèles DCC Student et à méta-copules Student au-dessus d'un niveau de confiance de 99%. L'utilisation de ces derniers pour la gestion des risques peut ainsi générer d'importants coûts d'opportunité. Dans l'ensemble, les modèles à copules sont meilleurs que les modèles DCC, alors que les modèles à méta-copules demeurent moins performants que les modèles à copules-paires. Enfin, le modèle R-vine domine légèrement ses deux versions restreintes.

La performance confirmée du modèle présenté justifie son utilisation dans un cadre de gestion des risques extrêmes, et particulièrement pour concevoir des scénarios de stress. Par rapport aux scénarios historiques, les scénarios hypothétiques sont moins répandus dans la littérature. Les rares travaux se rapportant à des scénarios statiques, et souvent uni-variés. Nous apportons donc une extension de ces travaux aux cadres dynamique et multivarié.

Nous analysons l'impact des scénarios de stress sur des portefeuilles analytiques linéaires. Un portefeuille équipondéré est considéré pour chacun de nos trois ensembles de données. Le

rendement du portefeuille est donné comme une fonction linéaire des rendements individuels pris comme facteurs de risque. La spécification du choc initial sur les facteurs de risque et la probabilité relative est basée sur les distributions marginales de la série de rendement. L'évolution du choc sur un horizon de stress arbitraire (c'est à dire l'après-choc) est donnée par les modèles ARMA-EGARCH et R-vine. Le résultat du test est évalué en termes de perte de rendement du portefeuille et de capital réglementaire requis pour couvrir ces pertes. Pour montrer l'importance du modèle de risque sur le résultat d'un stress test, nous avons également mesuré les résultats obtenus par d'autres modèles. Ceux-ci sont les mêmes que déjà examinés à des fins de validation dans la section précédente. Nous avons comparé les résultats obtenus par l'utilisation de quatre distributions marginales, en plus de la distribution semi-paramétrique présentée. Les résultats montrent une différence significative dans la taille des chocs correspondants. Or, nous avons montré, au chapitre 1, que l'un des principaux défis d'un stress test réside dans le choix de la distribution qui déterminera le niveau de plausibilité du scénario – et donc sa crédibilité et celle du test en question. Dans ce chapitre, nous avons démontré la faible performance de la plupart des spécifications paramétriques pour capturer les propriétés empiriques des séries de rendement. Par conséquent, l'utilisation de ces distributions dans un exercice de stress test peut générer une surestimation de la probabilité du scénario. Les tests basés sur ce genre de distributions peuvent ainsi souffrir d'un manque de crédibilité. En revanche, la spécification semi-paramétrique s'est avéré être plus flexible, moins subjective quant au choix de la forme de distribution et conduit à une meilleure performance pour les rendements extrêmes. Elle semble donc plus appropriée pour concevoir les scénarios de stress (extrêmes).

Par ailleurs, les résultats montrent une différence significative entre les modèles multivariés, en termes d'impact généré sur le portefeuille. Par exemple, en utilisant le modèle à méta-copule Student, présentant de fortes dépendances extrêmes, l'essentiel du choc initial est transmis à travers le système multivarié. Cela génère des pertes importantes sur le rendement du portefeuille, au fur et à mesure que l'horizon progresse. Or, cette perte est surestimée, car seuls quelques paires de rendements présentent effectivement une dépendance *à la* Student. Par ailleurs, l'utilisation d'un modèle DCC peut être source de confusion, car il ne permet pas de distinguer entre la forme de la structure de dépendance et celle des distributions marginales sous-jacentes. L'utilisation de modèles à copules-paires permet d'éviter ces inconvénients, en se basant sur une modélisation explicite de toutes les dépendances paires qui constituent la structure de dépendance multivariée. Exploiter l'extra-flexibilité et la performance avérée de ce modèle permet donc d'asseoir la crédibilité du test, conduisant à une meilleure utilisation de ses résultats.

Résumé du chapitre 3
Évaluation du Risque Systémique
Stress Testing du Système Bancaire Français

La crise des *subprimes*, amorcée aux Etats-Unis durant l'été 2007, s'est brusquement convertie en crise financière, de dimension planétaire, aux conséquences lourdes sur les économies réelles. Ces développements imprévus, tant en termes quantitatifs que qualitatifs, ont suscité, de la part des autorités de régulation et de supervision financières et de celle des décideurs de politique publique, une volonté de mieux appréhender les mécanismes de fonctionnement des systèmes financiers, afin de prévenir l'avènement de tels événements. Plusieurs outils d'analyse ont ainsi été mobilisés pour identifier les sources de vulnérabilité d'un système financier, évaluer leur impact potentiel sur le système en question et sur l'économie réelle et définir les modes de prévention et de gestion appropriés aux crises.

Les stress tests systémiques (également dits, macro stress tests) font partie de cet arsenal de gestion. Sur un plan conceptuel, ils répondent précisément aux attentes des régulateurs et des décideurs publics. Ils ont été, à cet effet, bien encadré par les régulations internationales. Néanmoins, l'engouement suscité par leur introduction a été modéré par les limites relevées sur les premiers exercices réalisés. Ces limites se traduisent, notamment, par : (i) des échecs dans l'identification des sources de vulnérabilité systémique (i.e. les facteurs du risque systémique), (ii) l'insuffisance des modèles d'évaluation retenus pour capter les canaux de transmission de ces facteurs au système financier et à l'économie réelle (i.e. le risque systémique), et (iii) le manque de crédibilité des scénarios de stress considérés. Ces aspects ont souvent mené à des conclusions hâtives et à l'illusion de robustesse des systèmes stressés. Plusieurs cas concrets illustrent ces erreurs d'appréciation.

Les coûts économiques, sociaux et politiques de ces erreurs renforcent la nécessité d'une meilleure maîtrise de ces exercices, en vue d'une gestion plus optimale. Ce chapitre apporte une réponse à la deuxième limite des stress tests systémiques – au sens de la classification présentée ci-dessus. Nous proposons un modèle de stress test qui permet d'évaluer le risque systémique porté par un système bancaire. Le modèle est fondé sur une spécification modulaire et séquentielle des canaux de transmission d'un mouvement (sévère) de facteurs de risque au système considéré. Associé à un modèle de risque, qui définit le comportement joint des facteurs de risque systémiques, le modèle d'évaluation proposé mesure le risque systémique en tenant compte de : (i) l'impact direct d'un mouvement de facteurs de risque (ci-après, le choc) sur les bilans bancaires individuels, et (ii) de la dynamique de cet impact sur un horizon temporel défini. L'impact direct est mesuré en termes de variation dans la valeur des positions de bilan exposées aux risques de marché, de crédit, de taux et de réputation. La dynamique du modèle, qui dépend de l'impact direct, inclut : (iii) les réactions privées au choc, (iv) la transmission du choc à travers les différentes banques du système, via les marchés financier et monétaire et le réseau de paiements interbancaires, (v) les effets de second tour, et (vi) l'intervention publique de la banque centrale et du Trésor public. Le modèle, qui s'inscrit dans une approche *bottom-up*, est appliqué aux bilans des six principaux groupes du système bancaire français.

La spécification des différents modules est simple et extensible. Celle-ci est menée via des estimations de forme-réduite et des règles générales (ou *rules of thumb*). Ce choix permet de capter, de façon flexible et opérationnelle, l'évolution du choc et le poids de chaque phase dans le résultat final. L'analyse détaillée et l'interprétation des résultats sont ainsi renforcées. En plus de permettre une estimation du risque systémique potentiel, le modèle proposé peut être

également considéré comme un outil quantitatif d'évaluation des stress tests conduits par les institutions privées et les autorités de supervision.

Notre étude contribue, à plusieurs égards, à la littérature existante. Tout d'abord, par l'introduction d'une nouvelle classe d'effets de premier tour, à savoir un coût de la réputation idiosyncratique pour une banque ayant subi une perte à la suite du scénario de stress. Deuxièmement, nous avons explicitement considéré les fonctions de réponse en matière de politique publique, impliquant la banque centrale et le Trésor public. Troisièmement, un nouvel algorithme de contagion est introduit, incluant trois canaux interactifs représentés par le marché financier, le réseau de paiements interbancaires et le marché monétaire. Quatrièmement, le modèle proposé repose uniquement sur l'information publique. Cela permet une implémentation répandue de ce modèle par les acteurs du marché financier, en présence d'une restriction d'accès aux données privées des banques. Enfin, au meilleur de notre connaissance, cette étude est la première à mener une analyse *bottom-up* de stress tests systémiques pour le système bancaire français depuis la crise de 2007-09.

Nous modélisons le risque porté par un système bancaire à l'aide d'un modèle dynamique, fondé sur une spécification modulaire qui permet de capter les mouvements de facteurs de risque et leur impact sur le système considéré. Le modèle est estimé à fréquence trimestrielle. Sa flexibilité permet son utilisation à des fins de prévision et de simulation, pour un horizon indéfini. Ces propriétés seront mises à profit dans ce chapitre pour conduire des stress tests systémiques. Le cadre général du modèle est développé ci-après.

Un modèle de risque définit les valeurs des facteurs de risque exogènes au début de la période. Compte tenu de ces valeurs, quatre effets directs sont modélisées sur les bilans des banques individuelles : les gains et pertes sur le portefeuille de négociation, les pertes sur créances, le revenu net d'intérêts et les coûts de réputation. Après l'enregistrement de ces effets, la rentabilité de chaque banque est évaluée. Deux situations peuvent se présenter : toutes les banques sont encore rentables, ou au moins une banque fait des pertes. Dans le premier cas, les bilans sont « réajustés », selon quatre règles de base (*rules of thumb*) connus pour refléter un comportement commun des banques commerciales dans la pratique. Ce réajustement marque la fin de la période et le début de la période suivante.

Si une ou plusieurs banques font une perte, leur solvabilité est évaluée. Les ratios réglementaires définis par les normes de Bâle III sont utilisés à cette fin. Deux situations peuvent se présenter : le ratio de solvabilité est encore au-dessus du minimum réglementaire pour toutes les banques non rentables, ou ce seuil est violé pour au moins une banque. Dans le premier cas, un coût de réputation est subi par les banques non rentables. Ce coût se manifeste par un resserrement de leurs conditions d'accès au refinancement, représenté par une augmentation du taux d'intérêt sur les nouveaux emprunts contractés par ces banques. Ce coût sera comptabilisé dans le revenu net d'intérêts de la période suivante. Cela marque la fin de la période et le passage à la suivante, après réajusté les bilans des banques rentables et « rééquilibré » ceux des banques non rentables pour tenir compte des pertes. Ces deux actions produisent des effets de second tour sur les banques rentables et non rentables.

Quand une banque non rentable enregistre un ratio de solvabilité inférieur au seuil réglementaire, elle est supposée être en défaut de paiement. Il s'ensuit alors une procédure de faillite complexe, définie en plusieurs étapes. Tout d'abord, les actifs de la banque défaillante perdent une partie de leur valeur, correspondant à ce qu'on appelle « les coûts de la faillite ». Pour respecter les engagements de la banque, les actifs restants sont vendus sur les marchés financiers. Cette action affecte les banques restantes via trois canaux de contagion complémentaires. Tout

d'abord, la vente massive d'actifs crée un choc d'offre sur le marché concerné, conduisant à une diminution des prix et une dévaluation des actifs marchands. Le portefeuille de négociation de l'ensemble du secteur est affecté via le canal de la liquidité de marché. Deuxièmement, le résultat des ventes d'actifs des banques défaillantes peut être insuffisant pour payer la totalité de la dette. Dans ce cas, les banques créancières subissent une perte supplémentaire liée à leur risque de crédit de contrepartie. C'est le canal de contagion via le réseau de paiements interbancaires. Enfin, la défaillance d'une ou plusieurs banques affecte la confiance du marché envers le système bancaire dans son ensemble. Les banques restantes seront, dans ce cas, confrontées à un durcissement des conditions de refinancement, faisant que leurs emprunts futurs soient majorés d'une prime de risque supplémentaire. C'est le canal de la liquidité de financement. La somme des trois effets de contagion sur les banques non défaillantes est enregistrée. Leur rentabilité et leur solvabilité sont réévaluées à nouveau, comme expliqué ci-dessus. Si de nouvelles faillites apparaissent, la procédure est répétée, causant d'avantage d'effets sur les banques restantes. Cette procédure se poursuit jusqu'à ce qu'aucun cas de faillite ne soit observé ou, moins probable, jusqu'à ce que l'ensemble du système tombe en faillite.

Lorsque la chaîne de contagion prend fin, et avant de rééquilibrer les bilans des banques restantes, un dernier module permet d'introduire les fonctions de réaction publique. En effet, pour rendre plus réaliste la structure du modèle, nous avons considéré deux fonctions de réaction pour chacun de la banque centrale et du Trésor public. La réponse de la banque centrale consiste à rétablir les prix des actifs (sur le marché financier) et le taux d'intérêt interbancaire (sur le marché monétaire) à leur niveau du début de période. Ces deux actions permettent, pour les banques restantes, d'absorber les pertes résultant du premier et du dernier canal de contagion. La réaction du Trésor consiste à réduire les coûts de faillite découlant d'une défaillance bancaire et à participer à l'augmentation du capital des banques non rentables restantes. La période se termine par un rééquilibrage des bilans, pour tenir compte des gains et pertes de l'ensemble de la période

L'estimation du modèle requiert deux catégories de données : des données macro-financières, pour estimer le modèle de risque et des données de bilan, pour estimer le modèle d'évaluation. Les modules sont estimés séparément, en fréquence trimestrielle. Huit indicateurs sont retenus pour caractériser les facteurs de risque exogènes du secteur bancaire français : le produit intérieur brut réel (PIB), l'indice des prix à la consommation hors énergie (IPC), l'indice des prix de logements anciens dans Paris (IPH), l'indice boursier CAC 40 (CAC), le taux interbancaire au jour le jour (TC), le taux des obligations d'Etat à 30 ans (TL), le taux de change nominal euro/dollar (EX) et le cours du baril de pétrole brut (Brent). Les données sont collectées à la fin de chaque trimestre, sur la période 1992:T1–2013:T4 (soit 88 observations). Les séries du PIB, de l'IPC et de l'IPH, ajustées des variations saisonnières, sont tirées du site de l'INSEE. La série du taux de change est obtenue à partir du site de la Banque d'Angleterre. Le reste des données sont collectées à partir de Datastream.

Les données de bilan sont constituées des bilans des six principaux groupes bancaires de droit français, dont le produit net bancaire (PNB) est supérieur à 10 milliards d'euros, au terme de l'exercice 2012. Il s'agit donc de : BNP Paribas (BNPP), Société Générale (SG), le Groupe Crédit Agricole (GCA), le Groupe Banques Populaires Caisses d'Epargnes (BPCE), le Groupe Crédit mutuel-CIC (GCM) et La Banque Postale (LBP). Le PNB de ces groupes réunis représente 94% du volume d'activité du secteur bancaire français. Quatre des cinq groupes sont issus d'opérations de fusions-acquisitions, réalisées dans les années 1990 et 2008, pour consolider les positions des banques françaises à l'échelle européenne et internationale. Pour contourner la complexité liée à ces montages et vu l'absence de données (détaillées) pour les anciennes entités,

nous avons opté pour une estimation du modèle d'évaluation à partir des données de bilan du groupe SG. Il s'agit, en effet, de la seule banque à ne pas avoir subi d'opérations significatives de fusion-acquisition au cours de cette période. Les données retenues pour l'estimation couvrent la période 2000:T1–2013:T4. Elles sont obtenues à partir des rapports trimestriels publiés par chaque banque. Les paramètres estimés du modèle sont ensuite utilisés pour mener des exercices de stress. Un scénario hypothétique est appliqué à l'ensemble des banques, sur la base de leurs bilans bancaires au 4e trimestre 2013.

Nous rapportons les principaux résultats d'un scénario hypothétique. Ces résultats ne sont en aucun cas exhaustifs. Ils ont été choisis en raison de contraintes d'espace, tandis que d'autres résultats sont encore disponibles sur demande auprès de l'auteur. Nous fournissons également un code Matlab avec des fonctions pour estimer le modèle, pour concevoir et appliquer des scénarios de stress stochastiques et déterministes, et pour récupérer les résultats. Ce code est assez simple et flexible. De plus, d'autres fonctions peuvent être introduites, pour tenir compte de nouvelles données (ex. des informations privées ou plus détaillés), pour ajuster la spécification du modèle et pour extraire d'autres sorties. La procédure d'estimation est assez rapide, dans la mesure où elle peut facilement être répliquée sur une base quotidienne. Cette caractéristique permet de prendre en compte toute nouvelle information pour mettre à jour les données et les équations du modèle.

Nous avons utilisé le modèle pour évaluer l'impact d'un scénario de stress non-stochastique sur le système bancaire français. Même si la conception du modèle permet également de considérer des scénarios stochastiques, nous nous limitons à des scénarios non stochastiques, compte tenu de l'objet de ce chapitre, qui se focalise sur l'impact des scénarios de stress plutôt que sur la conception proprement dite des scénarios (voir Chapitre 4). Le scénario consiste en un ensemble de variations appliquées aux valeurs des facteurs de risque eu égard des données de 2013:T4. Le choc considéré est inspiré des exercices de simulation de crise systémique couramment menées par les autorités de supervisions. Ici, nous considérons : (i) une baisse de 1% du PIB français, (ii) une augmentation de 10% de l'indice IPH, (iii) une baisse de 15% de la valeur réelle de l'indice CAC 40, (iv) une augmentation de 50 points de base du taux de prêt interbancaire, (v) une augmentation de 30% de la probabilité de défaut des ménages et des entreprises non financières, et (vi) un assèchement de la liquidité du marché financier. La plausibilité de ce scénario peut être justifiée par l'observation d'événements similaires au cours la crise financière de 2007-2009 ou au cours de la période considérée dans notre échantillon de données.

Le choc est appliqué au début de la période de simulation (2013:T4), correspondant à la date 0. La dynamique du choc et son impact sur les bilans bancaires sont évalués sur un horizon de trois ans sur une base trimestrielle. Par conséquent, l'horizon de stress correspond à 2014:T1-2016:T4, soit douze périodes. A partir du choc initial, sa dynamique est obtenue, pour les périodes suivantes, en utilisant 1 million de réalisations simulées du modèle de risque. Pour chacune des périodes de l'horizon, et pour chaque vecteur de facteurs de risque simulé, l'impact correspondant sur les bilans individuels est mesuré à l'aide des équations estimées du modèle d'évaluation. Les bilans 2013:T4 sont considérés pour mesurer l'impact du scénario à la première période. Au cours de l'horizon du test, seuls les changements endogènes basés sur les modèles considérés sont introduits sur les bilans initiaux. Les modifications exogènes sont donc exclues, à des fins de simplicité. Dans le même esprit, aucun choc supplémentaire n'est introduit au cours de la période de simulation. Là encore, ces deux hypothèses sont aisément modifiables dans le code fourni.

Les résultats montrent une forte relation entre la dynamique du profit et le portefeuille de

négociation. Toutefois, malgré un bénéfice en baisse, les bilans des banques sont en constante augmentation au cours de la période de simulation. En outre, le portefeuille de négociation s'avère essentiel dans les résultats. La suppression de ce portefeuille conduit à une diminution plus élevée du ratio de solvabilité. Toutefois, l'introduction du portefeuille de négociation apporte également plus de volatilité, à la fois sur la solvabilité et sur le taux d'endettement des banques. A cet égard, les résultats des petites banques semblent être plus volatils que ceux des plus grandes.

Après les premiers effets, quatre des six banques demeurent rentables dans plus de 99 % des cas. Parmi ces banques, trois ne font jamais faillite. Pour les banques ayant enregistré une forte probabilité de défaut, ceci s'explique principalement par la fragilité de leurs bilans lorsque le choc initial a été appliqué – vu que leurs ratios réglementaires étaient très faibles à cette époque. Dans certains cas, cependant, les résultats de ces banques aux cours des trimestres suivants ont également été insuffisants pour assurer une reprise significative. Cependant, il est intéressant de noter que ce sont plus les réserves en capitaux propres que la probabilité de rentabilité (i.e. le rendement espéré) qui fait qu'une banque soit susceptible ou pas de faire faillite. A cet effet, le risque de défaut est principalement accru chez deux banques du système considéré.

Le ratio de solvabilité des banques non rentables diminue après les effets de premier tour. Même si ce ratio est amélioré par les réponses privées de ces banques, il demeure nettement en dessous de son niveau du début de période. En outre, ce ratio reste relativement stable après la réponse, en raison de l'hypothèse selon laquelle, à court terme, les banques non rentables ne ciblent pas leurs ratios réglementaires. Il semble aussi que la réponse des banques non rentables n'affecte pas de manière significative leurs dettes, puisque leurs résultats négatifs ne sont pas nécessairement accompagnés d'une diminution sensible de la taille de leurs bilans.

En ce qui concerne les effets de contagion, deux principaux résultats sont obtenus pour les cas de faillites fondamentales. Tout d'abord, le nombre de défauts diminue à mesure que l'horizon se poursuit, même si leur probabilité est encore positive au terme de l'horizon. Ce résultat s'explique par le fait que le choc initial est progressivement absorbé vu qu'aucun choc supplémentaire n'est introduit. D'autre part, le taux maximal de défaut est atteint après deux à trois trimestres. C'est en raison de la fragilité des banques non rentables, suite aux effets du choc de stress sur les premières périodes. Ces banques sont, toutes choses égales par ailleurs, plus exposées aux effets de second tour. Avec le coût de la réputation reporté pour les périodes suivantes, les effets les plus lourds sont enregistrés tout de suite après le premier trimestre. A mesure que l'horizon avance, le choc s'estompe ; ce qui réduit également les effets inter-périodes.

La plupart des défauts de contagion sont concentrés sur la première année. Ceci peut être dû à différentes raisons. Premièrement, les banques non rentables sont plus vulnérables pendant les premières périodes, ce qui augmente leur probabilité de faire faillite. Deuxièmement, comme plusieurs banques font faillite, les effets de second tour augmentent également, d'où un impact sur les banques non rentables et profitables. Troisièmement, au cours des premières périodes, les défauts de contagion se prolongent à la suite de plusieurs itérations, jusqu'à ce que la plupart des banques vulnérables fassent faillite. Le nombre de défaut de contagion diminue ainsi dans les périodes suivantes – même si, dans certains scénarios, le processus de contagion provoque le défaut de l'ensemble du système. Enfin, il est à noter que les faillites de contagion sont relativement moins fréquentes que les faillites fondamentales, en raison de la présence de certaines banques plus vulnérables dans le système. Lorsque ces banques font faillite, le système restant devient relativement plus résistant.

La fraction des actifs liquidés est relativement faible par rapport à la taille du système bancaire. C'est ce qui explique la légère diminution du prix du marché financier. Deux conclusions peuvent être tirées en conséquence. D'un côté, la plupart des défaillances concernent les banques de petite et moyenne taille. D'autre part, le faible impact du portefeuille de négociation laisse entendre que la plupart des contagions sont entraînés par le marché monétaire et le canal des paiements interbancaires.

A la fin du processus de contagion, le ratio d'endettement des banques restantes augmente. Ceci peut être expliqué par le fait que les banques vulnérables ont été affectées par le processus de contagion et la plupart ont fait faillite. L'éviction de ces banques a également contribué à améliorer le ratio de solvabilité des banques restantes. Ce ratio est davantage renforcé par l'intervention de la banque centrale. En rachetant des titres aux banques commerciales, la banque centrale contribue à faire remonter le prix du marché financier; ce qui génère des gains supplémentaires. Ce résultat est valable pour les autres indicateurs réglementaires et de performance considérés. La réponse publique s'avère aussi être bien plus importante que la réponse privée des banques restantes. En effet, les rééquilibrages effectués par ces dernières ne semblent pas affecter de manière significative leurs ratios de solvabilité.

Résumé du chapitre 4
Identification Systématique des Scénarios de Stress
Une Approche Bayésienne

Dans un stress tests, seuls les scénarios crédibles doivent être conçus pour évaluer leur impact potentiel sur le portefeuille stressé. Pour être considéré comme crédible, le scénario doit engendrer des effets significatifs sur la valeur du portefeuille, tout en respectant un minimum de plausibilité. Choisir parmi tous les scénarios alternatifs est donc d'un intérêt central dans ces tests. Un compromis entre la plausibilité et la sévérité est nécessaire pour s'assurer que des scénarios sévères ne soient pas omis et que seuls des scénarios plausibles soient pris en compte. Dans la pratique, cependant, les stress tests réalisés s'appuient sur des méthodes subjectives pour identifier les scénarios de stress. Ces derniers sont souvent basés sur des épisodes de stress passés (scénarios historiques) ou sur des hypothèses concernant des événements futurs (scénarios hypothétiques). Pourtant, ce faisant, le critère de sévérité n'est pas toujours rempli. Le scénario retenu pourrait donc être moins néfaste que d'autres scénarios possibles, de niveaux de plausibilité égale ou supérieure. En d'autres termes, lors de la création du scénario en amont, ces tests se soucient plus du scénario lui-même que de son impact sur le portefeuille. En outre, les scénarios réalisés sont souvent non probabilisés, ce qui rend difficile d'évaluer leur plausibilité. Ces deux limites affectent, à plusieurs égards, l'utilité des résultats du test en question.

Tout d'abord, en omettant des scénarios sévères mais plausibles, les résultats conduisent à une mauvaise identification des « vrais » pertes extrêmes pour le portefeuille. Cela produit une fausse illusion de sécurité et peut conduire à de mauvaises décisions de gestion des risques. Un exemple concret est donné par les problèmes récents des systèmes bancaires grec et irlandais, contrastant avec les conclusions optimistes des stress tests auxquels ils ont été soumis. D'autre part, concevoir des scénarios invraisemblables peut conduire à ignorer tout ou partie des résultats de l'exercice. Ceci est particulièrement vrai lorsque l'impact révélé est jugé trop importants – et les mesures suggérées trop chères. Ici, le décideur ne prendrait pas de mesures significatives, sachant que le scénario considéré peut, à postériori, se révéler sans fondement. Les résultats sont, dans ce cas, simplement considérés comme une « fausse alerte ».

Deuxièmement, les résultats des scénarios non probabilisées (appelés aussi déterministes) ne peuvent pas être inclus dans tous les processus de prise de décision. En finance, la plupart des décisions en incertitude sont, en effet, fondées sur des programmes stochastiques d'optimisation sous contrainte. Ceci est particulièrement le cas en matière de gestion des risques, d'évaluation d'actifs financiers et d'allocation de portefeuille. Toutefois, lorsque la probabilité du scénario de stress n'est pas fixée, la pertinence des décisions prises en conséquence ne peut être évaluée. Ces décisions peuvent donc s'avérer inappropriée par rapport à la « vraie » probabilité du scénario considéré. Par conséquent, des pertes inattendues pourraient apparaître si les mesures de gestion se révèlent être insuffisantes. Lorsqu'en revanche celles-ci sont trop poussées, d'importants coûts d'opportunité peuvent être essuyés.

Enfin, les stress tests sont réalisés de façon occasionnelle. Or, la gestion des risques est une pratique effectuée sur une base quotidienne, en utilisant des outils plus conventionnels tels que la VaR. Lorsque les deux outils coïncident, les gestionnaires de risques sont souvent confrontés à un dilemme délicat : faut-il continuer de s'appuyer sur les résultats de la VaR, valables dans des conditions normales de marché, ou basculer vers les résultats de stress tests, valables uniquement en situation de stress ? Une réponse intuitive serait de dire : les deux à la fois. Mais là encore, le caractère déterministe des scénarios de stress empêche la combinaison de ces

deux approches. Par conséquent, les stress tests sont souvent considérés indépendamment des autres outils de gestion. Malgré les recommandations récurrentes des autorités de supervision et de régulation, la plupart des tests réalisés sont toujours basés sur des scénarios déterministes, laissant inexploité leur vaste potentiel.

L'objectif de ce chapitre est double :

(i) présenter une méthode flexible pour identifier les scénarios de stress sévères mais plausibles, pour un portefeuille non linéaire. Le théorème de Bayes, et en particulier la méthode dite *Nested Sampling* (ou échantillonnage imbriqué) est utilisé pour définir le domaine de plausibilité contenant tous les scénarios plausibles, pour un seuil de plausibilité prédéfini. Une recherche systématique est ensuite menée dans ce domaine pour identifier le(s) scénario(s) le(s) plus sévère(s) pour le portefeuille. Cette approche évite les considérations subjectives dans la sélection du scénario et assure le compromis nécessaire entre les critères de sévérité et de plausibilité.

(ii) présenter un cadre analytique pour les applications possibles de la méthode proposée à des fins de gestion des risques. Nous nous concentrons, en particulier, sur les méthodes permettant : (i) de combiner les résultats des stress tests et de la VaR, (ii) d'identifier les facteurs de risque clés pour un portefeuille donné, et (iii) d'effectuer une allocation optimale de portefeuille. Nous souhaitons que cela aide les gestionnaires de risques à surmonter les principaux problèmes causés par les scénarios de stress déterministes et leur permette de s'appuyer le plus souvent sur les résultats de ces tests.

Une sélection objective des scénarios de stress peut être réalisée à travers deux approches : (i) une recherche systématique des scénarios sévères dans un ensemble de scénarios plausibles, ou (ii) une recherche de scénarios plausibles parmi tous les scénarios sévères. Comme l'ensemble des scénarios sévères est souvent illimité, la deuxième approche présente donc un problème pratique. Ainsi, la littérature fait souvent référence à la première approche, comme nous le ferons dans cette étude. Cette approche systématique est basée sur deux éléments centraux : une définition quantitative des concepts de sévérité et de plausibilité, ainsi que le choix d'un seuil de plausibilité. Alors que ce dernier est laissé à la discrétion du décideur, qui le définit en fonction de sa culture du risque, la littérature offre quelques mesures statistiques pouvant être adaptées pour mesurer la sévérité et la plausibilité des scénarios de stress. Le choix de la mesure à adopter dépend souvent de la nature du scénario.

Un scénario de stress repose sur les changements dans les propriétés statistiques des facteurs de risque. Les variations en niveau (ex. rendement) constituent l'approche commune dans la pratique. Toutefois, bien qu'assez intuitives, ces hypothèses correspondent à une définition restreinte d'un scénario de stress. En effet, le scénario peut également supposer des changements dans le modèle de risque qui spécifie la distribution des facteurs de risque. Ainsi, en périodes d'agitation (euphorie ou stress), la volatilité et la structure de dépendance des facteurs de risque sont également susceptibles d'évoluer ; ce qui modifie la distribution sous-jacente. Cette modification peut consister en une modification dans les paramètres de la distribution, voire un changement radical de famille de distribution, en cas de ruptures structurelles (ex. changement de régime). Supposer fixe la distribution des facteurs de risque sur tout l'horizon de stress rend ainsi le test peu réaliste. Ci-après, les scénarios supposant des changements dans les valeurs des facteurs de risque sont appelés « scénarios purs », alors que le terme « scénarios généralisés » est utilisé pour qualifier les scénarios s'appuyant sur des changements dans le modèle de risque sous-jacent. Compte tenu de la différence dans les propriétés statistiques de ces deux approches,

des mesures statistiques spécifiques doivent être utilisées pour évaluer leurs niveaux de sévérité et de plausibilité respectifs. Cela est particulièrement vrai pour la plausibilité, comme nous le verrons ci-dessous.

La sévérité du scénario mesure son impact sur le portefeuille. Lorsque cet impact est donné par une valeur unique, comme le rendement du portefeuille, celle-ci peut être directement utilisée comme une mesure de sévérité. Dans la pratique, cependant, le profil de risque d'un portefeuille est généralement donné par sa fonction de distribution des profits et pertes (P&L). Pour comparer des P&L alternatives, les gestionnaires de risques ont recours à des modèles de sélection de portefeuille. Ces derniers incluent, entre autres, les modèles de rendement-risque (ex. variance, écart moyen absolu, moments partiels inférieurs, demi écarts-types centraux), les mesures de risque (ex. VaR, CVaR), la maximisation de l'utilité espérée, les critères de dominance stochastique, etc. Les gestionnaires de risques peuvent donc compter sur l'une de ces méthodes pour évaluer également la sévérité de scénarios de stress. Pour éviter toute confusion entre la distribution des pertes et profits du portefeuille et celle des facteurs de risque, nous référons à la première par P&L et garderons le terme distribution pour la seconde.

Le concept de plausibilité peut être assimilé à la divergence ou dissemblance entre les valeurs des facteurs de risque sous le scénario de stress et leurs valeurs sous un scénario de référence (ex. valeurs espérées). Les mesures statistiques de distance peuvent ainsi être utilisées à cet effet. Pour les scénarios purs, ces mesures seraient effectuées dans l'espace euclidien des valeurs de facteurs de risque. Pour les scénarios généralisés, ce serait fait dans un espace de dimensions supérieures, formé par les fonctions de probabilité des facteurs de risque. La distance est mesurée entre deux points, dans le premier cas, et entre deux ensembles, dans le second. La première série de mesures comprend, entre autres, les distances euclidienne, Minkowski, Mahalanobis, Manathan, Tchebychev, etc. Des mesures telles que la distance de Bhattacharyya, la divergence de Kullback-Leibler et la distance de Fisher-Rao ont été proposées pour les fonctions de distribution. Les deux approches sont largement utilisées en statistique, en économétrie et dans d'autres domaines pratiques, tels que le traitement et l'analyse d'image, l'imagerie médicale et la télédétection, entre autres. Cependant, seules quelques études ont adapté ces outils à des fins de stress tests. La distance de Mahalanobis (MD) et, plus récemment, la divergence de Kullback-Leibler (KLD) ont été particulièrement expérimentées. Leur adoption permet de faire un pas important vers la systématisation des scénarios de stress et l'intégration des résultats de l'exercice dans des procédures de gestion quantitative des risques. Néanmoins, plusieurs pistes d'extension sont laissées ouvertes, dues aux hypothèses sur lesquelles sont fondées ces mesures. Si nous prenons l'exemple de la divergence de la KLD, plusieurs pistes d'extension apparaissent. Tout d'abord, la KLD n'est pas une mesure de distance au sens propre sur l'espace des distributions de probabilité, car elle n'est pas symétrique et n'obéit pas à l'inégalité triangulaire. La KLD comme une mesure de plausibilité peut donc être débattue. Deuxièmement, le calcul de la KLD implique de résoudre une intégrale multidimensionnelle sur l'espace de données. Cependant, dans la littérature, il n'existe pas d'expression analytique explicite de la KLD pour les distributions elliptiques – sauf pour le cas gaussien dégénéré. Dans le cas général, la KLD est approximée par des méthodes numériques, assez lourdes. Enfin, cette méthode laisse ouvertes deux questions principales liées à sa mise en œuvre à des fins de stress test. D'un côté, il est difficile de définir l'ensemble de scénarios possibles sans introduire de jugements d'expert subjectifs. De l'autre côté, les gestionnaires de risques sont peu familiers avec la KLD, exprimée en termes géométriques. Ainsi, par exemple, la façon dont le rayon mesuré par la KLD peut être liée à la notion, plus opérationnelle, de probabilité du scénario, demeure peu évidente. Pour remédier à ces problèmes, nous proposons de remplacer les quelques mesures de

distance existantes par des méthodes bayésiennes pour sélectionner les scénarios de stress.

Les analyses basées sur l'approche bayésienne ont constitué un terrain de recherche fertile, qui s'est sensiblement enrichi au cours des dernières années. Un effort considérable a été fait pour développer des méthodes efficaces et robustes, utilisées principalement pour la sélection de modèles et l'estimation des paramètres. Dans ce dernier cas, l'idée consiste à définir une distribution a posteriori des paramètres d'un modèle donné, à partir d'une distribution a priori. Ainsi, l'inférence bayésienne peut être considérée comme un moyen de prendre en compte une partie de l'incertitude du modèle, à savoir « l'incertitude des paramètres ». Dit autrement, la méthode permet d'identifier le risque de modèle lié à l'évolution des valeurs de paramètres du modèle. Cette définition (étroite) du risque de modèle correspond assez à celle des scénarios généralisés présentée ci-dessus. En effet, pour un modèle donné, l'ensemble des scénarios possibles peut être défini par l'espace de valeurs des paramètres. Par conséquent, chaque scénario est représenté par un vecteur de paramètres. Des scénarios plausibles dans l'espace des paramètres peuvent être définis à l'aide de critères quantitatifs adaptés.

L'utilisation de l'inférence bayésienne pour définir des scénarios de stress présente plusieurs avantages par rapport aux mesures de distance. Tout d'abord, l'utilisation de méthodes non paramétriques évite toutes les critiques liées au choix spécifiques d'une mesure de distance particulière, parmi lesquels aucun critère objectif de classification n'est apporté par la littérature. D'autre part, l'ensemble des scénarios possibles est désormais donné par l'espace des valeurs des paramètres du modèle de risque considéré. Cela permet d'éviter tout jugement subjectif à cet effet. Enfin, la plausibilité d'un scénario est directement mesurée à partir de la distribution a posteriori des vecteurs correspondants aux valeurs des paramètres. Ceci permet d'éviter les problèmes liés à l'utilisation de mesures intermédiaires de plausibilité, telles que le rayon dans la KLD.

Cependant, les méthodes bayésiennes sont lentes en termes de calcul ; ce qui peut constituer une sérieuse gêne pour des modèles complexes. En effet, les méthodes MCMC peuvent être intensives, tout en rencontrant des difficultés dans l'inférence à partir de distributions a posteriori multimodales ou avec une grande dégénérescence entre les paramètres. Ceci est particulièrement le cas pour les modèles surdimensionnés. En réponse à ces difficultés, une méthode MC moins intensive, dite Nested Sampling, a été récemment proposée pour permettre de réduire les problèmes de calcul rencontrés par les méthodes classiques. Des algorithmes de calcul ont été proposés pour cette approche, qui a également montré une meilleure précision par rapport aux techniques MCMC existants. Dans cette étude, nous adoptons cette méthode pour inférer la distribution a posteriori des paramètres du modèle de risque utilisé pour le scénario de stress.

En considérant les vecteurs simulés de paramètres comme scénarios, nous proposons une méthode simple pour effectuer une sélection systématique des scénarios les plus sévères pour un portefeuille non linéaire donné. La méthode Nested Sampling permet de définir la distribution postérieure des paramètres du modèle. L'espace de paramètres simulés est considéré comme l'ensemble de tous les scénarios possibles. En d'autres termes, les scénarios sont assimilés à des paramétrisations alternatives du modèle de risque. Si l'on considère un portefeuille exposé à des facteurs de risque spécifiés par ce modèle, la recherche systématique du pire scénario vise à identifier les scénarios générant l'impact le plus sévère sur le portefeuille. Cette recherche est effectuée de la manière suivante. Tout d'abord, les scénarios sont classés en fonction de leur plausibilité. Cela peut être fait à l'aide d'un critère quantitatif adapté. Ici, nous utilisons les quartiles de la distribution a posteriori pour définir la probabilité de chaque scénario. Les scénarios dont les probabilités sont les plus élevées sont également les plus plausibles. Deuxièmement, nous fixons un seuil de plausibilité et définissons tous les scénarios présentant

une probabilité supérieure à ce seuil. Ces scénarios forment ainsi le domaine de plausibilité. Troisièmement, pour chaque scénario dans le domaine de plausibilité, nous procédons à une simulation du modèle de risque pour obtenir les valeurs espérées de facteurs de risque. Un ensemble de vecteurs de facteurs de risque est alors obtenu pour chaque scénario ou paramétrisation. Ceci peut se faire de deux façons : une simulation non conditionnelle en se fondant uniquement sur les valeurs des paramètres du modèle, ou une simulation conditionnelle qui comprend également des valeurs passées des facteurs de risque. La seconde approche est similaire à des exercices de prévision. Quatrièmement, à l'aide de la fonction de perte approprié du portefeuille, nous mesurons, pour chaque scénario, les valeurs du portefeuille résultant de l'ensemble des mouvements de facteurs de risque simulés. Un ensemble de valeurs de portefeuille est alors obtenu pour chaque scénario. Ces valeurs sont ensuite utilisées pour former la P&L empirique résultant de chaque scénario. Enfin, le pire scénario est déterminé en comparant les P&L obtenues à l'aide du critère *MaxLoss* adapté à l'esprit des stress tests.

Pour montrer les avantages concrets de l'utilisation de la méthode bayésienne présentée dans la conception des scénarios de stress, nous avons présenté un exemple expérimental de la méthode proposée, ainsi qu'un cadre méthodologique de ses principales applications en matière de gestion des risques. Celles-ci comprennent des méthodes permettant : de combiner les résultats de stress tests et de la VaR dans un cadre unifié, d'identifier les facteurs de risque clés pour le portefeuille, d'effectuer une allocation optimale de portefeuille, et de conduire des exercices de « reverse stress tests ». Nous espérons que cette démarche aidera les gestionnaires de risques à surmonter les principaux problèmes causés par les scénarios de stress déterministes, et leur permette de s'appuyer plus fréquemment sur des résultats les stress tests réalisés.

2. Financial Stress Tests Lessons from a Decade of Practice

Abstract

This chapter provides an analysis of financial stress tests, based on a qualitative and quantitative assessment of the main stages of the exercise. We compare, for each stage, the underlying conceptual and practical aspects to international regulatory standards in terms of risk management and stress testing. This method is applied to the main stress testing exercises carried out on major financial institutions and systems of developed countries over the period 2000-2014. The results allow assessing the strengths and weaknesses of current practices, the value to be placed on the related results, and the remaining challenges.

Keywords: Financial institution, risk management, risk model, scenario, stress testing, VaR

JEL Classification: G11, G17, G28

2.1 Introduction

In the narrow sense, stress testing exercises refer to a set of techniques allowing to estimate the likely impact of *severe yet plausible* events called "stress scenarios" on a portfolio of financial assets – or the "stressed portfolio". The nature of this portfolio determines the type of the corresponding stress test. We distinguish therefore:

- Individual (micro or portfolio) stress tests: related to portfolios held by departments or business units of a single entity (e.g. a financial institution).

- Systemic (macro or aggregate) stress tests: related to a set of entities in a systemic point of view (e.g. domestic or global banking systems).

In concrete terms, the stress test consists in reassessing – *ex ante* – the portfolio's value given a set of assumptions on the underlying risk factors.[1] These assumptions form the scenario. The more severe is the scenario – in terms of possible losses in the portfolio's value – the more it is assumed to reflect an adverse or a stress situation; hence the labels *stress scenario* and *stress test* (Jones et al., 2004). However, the stress test is not limited to the conception of the most severe scenarios, as a plausibility condition is also required (see below).

The above definition can be extended to include the integration of the stress testing results in the risk management process related to the stressed portfolio. This process may include an analysis of the results and the information therein, an assessment of possible risk management actions, and decision-making. Although this is more than a simple extension of the narrow definition of stress tests, empirical studies show that this aspect is missing in most realized tests, despite recurrent regulatory recommendations (see, e.g., BCBS, 2009; IIF, 2008; CRMPG III, 2008). The use of the stress testing result should, therefore, be at least as important as the result itself.

Combining the two previous definitions, three key elements can be retained to form what will be referred to as "the relevance criteria of a stress test" in the remainder of this thesis. These are, on one side, the severity and plausibility conditions required for the stress scenario, and, on the other, the usefulness of the stress testing results. The terms "test" and "exercise" will be used alternately in this chapter to refer to a stress testing exercise. Moreover, the term "process" will be retained in the sections related to the successive and iterative character of the exercise's stages.

The remainder of this chapter is structured as follows. Section 2.2 analyses the relevance criteria and emphasizes their importance in a stress test. Sections 2.3 and 2.4 present the conception issues of stress tests and their main applications and contributions to risk management. Section 2.5 defines the scope of the test, including the choice of portfolios, risk factors, and the test horizon. Section 2.6 discusses the issues related to the conception of stress scenarios. Section 2.7 covers the results and puts forward potential ways to capitalize on the outputs. Section 2.8 concludes.

[1] Risk factors may include financial, economic and/or political variables and indicators, for example. The related assumptions may relate to the risk factors' changes in level, in variance, in correlation, etc. (see section 2.6 for further details).

2.2 Relevance criteria of a stress test

The analysis of the triplet severity-plausibility-utility is the central point on which this study is focused. The present section defines the three concepts. A thorough analysis of their applications will be the subject of the subsequent chapters.

2.2.1 Severity

To be considered as severe a scenario should be designed to capture the most adverse events to the portfolio under review. The severity level is often likened to the range of changes applied to the underlying risk factors.[2] The concept of severity is therefore confused with that of extremity. Indeed, it is generally admitted in practice that: *the larger is the risk factor's change, the more it is of extreme nature, and the more severe is the corresponding scenario.*

However, this rule only applies on linear portfolios whose values are proportional to that of the underlying risk factors. It becomes inadequate to more complex portfolios (e.g. options, derivatives, swaps) for which the most severe losses do not necessarily stem from extreme levels of risk factors. Moreover, the extreme nature of a given risk factor value is a relative notion. It indeed varies over time, depending on the information set available when the scenario is designed. In other terms, a scenario considered as extreme during stable market conditions may become minor in periods of turmoil, when other even more extreme movements are conceivable. This fact can be noticed in the stress tests carried out before and during the 2007-09 financial crisis. While the related scenarios have been designed on the basis of the situation before the crisis, their results have suggested that potential losses would be less than a quarter of potential gains. This has been widely invalidated by the facts (see BCBS, 2009).

The scenario's severity varies then from one portfolio to another depending on their respective compositions, and from one period to another – for the same portfolio – depending on the range of potential risks. Neglecting one or both of these aspects is likely to cause bias in the test's results. To avoid any confusion in what follows, we abandon the term "extreme" in favour of the term "severe". This implies a customization of the scenario to the nature of the portfolio and to the context under study.[3] This leads us to the first challenge when implementing relevant stress tests.

Moreover, the analysis of a portfolio often leads to identify more than one potential severe scenario. Indeed, starting from a given scenario, it is often possible to design an even more severe one, and so on. Therefore, the next challenge relies on selecting one scenario among the – often unlimited – set of severe scenarios, or at least reducing this set with respect to some criteria. A first answer to this issue is provided by the introduction of the plausibility condition.

2.2.2 Plausibility

The scenario's plausibility indicates the confidence level associated with the underlying assumptions. To be considered as such the scenario should consist of likely changes in risk factors.

[2] This range is measured by the difference – in absolute value – of the risk factor's value (e.g. return) with respect to a reference value (e.g. average of past returns). The changes in disfavour of the portfolio's value are usually held in stress scenario – even though favourable movements could also be considered.

[3] When the impact of a scenario on the portfolio is captured for several periods, the resulting cumulative loss is considered to assess the severity level of this scenario.

When explicitly set, the plausibility level is usually assessed by the occurrence probability of these changes. Hence, *the higher is the probability of a risk factor's change, the more likely it is, and the more plausible is the corresponding scenario.*

The previous definition makes it possible to remove from the set of stress scenarios those presenting lower probabilities. Besides the severity condition, the scenario should also be plausible. However, this definition brings up two further issues.

- How to assess the probability of a scenario, as different values could be associated with the same risk factor's change depending on the assumed underlying distribution function? Which distribution function to retain so that we associate a single probability to each scenario?
- At what level of probability a stress scenario could be considered as plausible? How to set such a "plausibility threshold"?

The first question is widely discussed in the literature. The choice of the distribution function – also called "risk model" – is often made arbitrarily. In some recent studies, this choice is based on specification and Goodness-of-Fit tests instead.[4] The considered distribution function is then fitted to past data on risk factors' changes. Given the estimated parameters, a probability could be associated with each change – corresponding to the distribution quartiles.

The response to the second question allows defining the set of scenarios whose probability is greater than or equal to the plausibility threshold. However, scenarios with high probabilities – and therefore the most plausible – are sometimes less severe. This is especially the case for linear portfolios.[5] A trade-off between severity and plausibility is therefore needed for stress scenarios – hence the label *severe yet plausible* in the definition. This trade-off could be made through three alternative approaches:

- Set the risk factors' changes and measure their probability.
- Set the probability and derive the corresponding risk factors' changes.
- Set a plausibility threshold and search through the corresponding risk factors' changes that (or those) presenting the most severe impact on the stressed portfolio.

The first two approaches are ad-hoc methods while the last one is more systematic and frequently referred to as "the systematic research for the worst case scenario".

The major drawback of the first approach relies on the risk of omitting (more) severe scenarios. The second approach presents two main limitations: (i) the selection of a single risk factor's change among all changes corresponding to the defined probability, and (ii) the risk of omitting more severe changes with higher probabilities. The third approach allows avoiding most of these limitations as it is based on an endogenous trade-off between the severity and the plausibility criteria. However, additional issues are still associated with this approach. These

[4] In the next chapter, we will define objective criteria to select the distribution functions given the nature of the risk factors under review. We will also extend the notion of distribution function to that of risk model which also includes the risk factors' dynamics and dependence structure.

[5] For linear portfolios, any fall in the scenario's severity (i.e. minor changes in the risk factors' value) matches with an increase in its plausibility, and vice versa. This relation does not necessarily hold for non-linear portfolios, for which the most severe scenarios could also be the most plausible.

are mainly related to the choice of the portfolio's "valuation model" – sometimes called "loss function" – and that of the numerical algorithm allowing to search for the worst scenario(s) in the set of plausible scenarios. Chapters 3 and 4 are respectively dedicated to investigate these two issues.

The design of the worst case scenario is not imposed by national nor international regulations. These instead require in the scenario to be significant and reasonably conservative, without giving any quantitative definition of these concepts (BCBS, 2006). In what follows, we call "credible scenario" any stress scenario combining severity and plausibility criteria as defined above – whatever the approach retained to do so. This qualifier will also be retained for the stress tests based on such scenarios. Moreover, we call "starting assumptions" the choices related to the risk model, the valuation model, and the methods of setting credible scenarios.[6]

2.2.3 Usefulness

A stress testing exercise is assumed to provide additional informations allowing to improve the risk management process of the stressed portfolio. This third relevance criterion is of particular importance knowing the costs of the test (in terms of time, human, material, and financial resources) and the negative effects stemming from an unsuitable use of the test's results.

The usefulness of a stress test is assessed downstream. It mainly depends on how the obtained results are used to identify core vulnerabilities for the stressed portfolio, optimal steps through which these could be managed, as well as less risky investment opportunities (see BCBS, 2006). The decisions taken accordingly depend on two elements: (i) the credibility of the stress scenario, and (ii) the features of the decision-making process.

A decision based on a credible scenario ensures a sufficient risk management for the portfolio if such a scenario were to occur. Conversely, if more severe – but not less plausible – scenarios have been omitted, the decisions taken accordingly would be suboptimal, and the test pointless. Moreover, even when the scenario is credible enough, the responses could be inadequate or insufficient. In such cases, the benefits driven by the scenario will be offset by the negative effects of these responses. Unsuited actions could therefore form opportunity costs and even lead to important future losses. When the results of the test and the related responses are publicly displayed, additional negative externalities are to be expected (e.g. supervision sanctions, deposit withdrawals, funding constraints, lower ratings, etc.).

Unfortunately, neither the scenario's credibility nor the decision relevance could be assessed statistically using common tools such as backtesting procedures. This is partly due to the basic character of the stress testing scenarios carried out in practice. In most cases, these scenarios are not "probabilized", which compromises any statistical assessment of their credibility *ex post* (Aragonés et al., 2001). This issue still remains even when the stress scenario is probabilized. In this case, the complication is often due to data limitation during periods of stress, especially for new markets and assets. These issues require a conception of suitable methods adapted to stress testing exercises, as we shall see in chapter 2.

The next sections analyse conceptual and functional features related to the implementation of stress tests. We present the main stages of the exercise, focusing on the frequent issues, the proposed solutions, and the remaining challenges. The analysis is based on the major stress

[6] These issues form what is known in the literature as "the model risk".

tests carried out on the financial institutions and systems of developed countries over the period 2000-14.

2.3 Conceptual issues

The previous section emphasized the importance of the starting assumptions. Their impact goes beyond the corresponding formulas and numerical applications in the empirical stages of the test. Indeed, the design of objective assumptions ensures the scenario's credibility and the test's usefulness. However, the persons and structures responsible for these assumptions could also introduce subjective judgments due to the nature and the complexity of their missions.

After a brief presentation of the different stakeholders in a stress testing exercise, we discuss the possible considerations underlying the starting assumptions and the associated issues.

2.3.1 Stakeholders and prior arrangements

The nature of stakeholders and their responsibility in a stress testing exercise is related to the type of the test.[7] Hereafter, we refer to stakeholders as "the designer", whatever their nature and number.

An individual stress test is usually conducted by the institution to which it relates or by one of its departments or business units. The test is carried out either:

(i) In a voluntary action within the internal risk management process. It is therefore launched on the initiative of senior management, risk managers and/or heads of departments and business units, etc. (BCBS, 2006, §. 718 (LXXIII)).

(ii) In response to an external request from regulatory and supervisory authorities.[8] In this case, the test is performed in a context of micro-prudential supervision to assess the resilience of each institution to the considered scenario – hence the occasionally used label *resilience tests* (DBB, 2007; BCBS, 2006, §. 718 (LXXII)). These tests are typically used for spatial and temporal comparisons. They are therefore often designed in a uniform and standardized way, without considering the specificities of the stressed portfolios and institutions, which limits their usefulness. According to international regulators, these *one-size-fits-all* tests should be discarded in favour of *tailor-made* tests (BCBS, 2006, 2009; CGFS, 2001, 2005; CRMPG III, 2008; IIF, 2008). However, this task is often costly for supervisors; which presents an argument in favour of internal tests.

If the complementarity of internal and external tests is conceivable in a theoretical point of view, it is often difficult to integrate and/or to compare the results of both approaches. This is often due to a difference of perception between the respective designers, regarding the test objectives, scope (e.g. the choice of the database), and scenarios. Regarding the choice of the risk model, for example, different approaches could be used by risk managers on one side, and supervisors on the other. The first, frequently relates on

[7] The term stakeholders may include one or more of the following parties: Management, members of the Board of Directors, risk managers, trade managers, traders, experts and economists, regulators and supervisors, shareholders, creditors, suppliers and customers, etc.

[8] At the time of writing, we have faced no test performed after an explicit request from other external parties who are directly or indirectly related to financial institutions' activity.

analyses based on financial theory, mathematics and statistical models; while economists and supervisors prefer macroeconomic tools with a particular emphasis on systemic risk factors (Drehmann, 2008).

(iii) In a mixed framework, including internal and external stakeholders (BCBS, 2006, §. 718 (LXXI)). A common approach consists in setting the stress scenario by the supervisor, leaving it up to the institution to choose the appropriate models and data. The use of mixed tests is sometimes motivated by the problems faced by some (small) institutions to design stress scenarios incorporating systemic events – also called *big picture hypothetical scenarios*. Thus, these institutions often call on the supervisor' expertize, either directly by applying the designed scenarios, or indirectly using the information displayed in central banks and supervisors' publications, such as financial stability reports, surveys, etc. (CGFS, 2001, 2005; DBB, 2007; Drehmann, 2008).

Within a macro-prudential analysis, systemic stress tests can be carried out by supervisors on the basis of individual and/or aggregated data on the institutions that make up the system under review. Here again, it particularly consists of the so-called *resilience tests*. Beyond issues related to the design of credible scenarios, systemic stress tests involve several practical challenges. These include private data collection and harmonization, modelling the systemic network and the related linkages, managing the dynamic aspects in the stress events (e.g. contagion, second-round and feedback effects), aggregating individual results, etc.

Individual and systemic stress tests will be analysed in chapters 2 and 3, respectively.

2.3.2 Setting the starting assumptions

Whatever the stress testing type is, its success is closely linked to the content of the starting assumptions. However, given the involvement of different parties in this stage of the process, setting up consensual hypotheses could be long and complicated. In such a situation, the final word often rests with the dominant party in this *collective bargaining* (Haldane, 2009).

Stakeholders' viewpoints are influenced by internal and external aspects. Čihák (2007) and Haldane (2009), among others, show that risk managers in a financial institution are generally pessimistic and often tend to overestimate the risks facing the different portfolios. They thus confer more plausibility to the most severe scenarios. On the other hand, Management and the Board of Directors are more optimistic and deem implausible that kind of scenarios. Any use in a stress testing exercise would therefore be unnecessary. The risk manager should then convince Management of the accuracy of the considered assumptions. This task is often critical because of the subjective character of certain assumptions and the practical implications resulting from the choice of the most severe scenarios.

Having given his approval, Management shall be liable to take the necessary actions in response to the test's results. Depending on the scenario's severity level, these actions may present significant opportunity costs for the institution, which may result in a negative impact on its profits. When Management is convinced by the plausibility of the scenario, he can motivate his decisions advocating a trade-off profitability-risk according to the *risk culture* inside the institution. Otherwise, only scenarios of moderate severity, considered as reasonable by Management, will be admitted. This aspect is one of the main criticism levelled by researchers and market players against most tests carried out before and during the 2007-09 crisis (Haldane, 2009).

Even when the internal viewpoints come to converge, external considerations can lead the designer to reject severe scenarios. The implications of the results' publication and their notification to supervisors are the two major external barriers that may influence the choice of assumptions.

The publication of stress testing results is a practice that is being developed in recent years. This gives rise to uncontrolled comments and interpretations that may have adverse effects on the concerned institution(s). The assumed scenarios may therefore be understood by market players as the own expectations and the risk perception of the institution. In a situation of panic and confusion, the test's results may: (i) be a means of pressure on the decision-making process, (ii) affect the reputation of the institution, (iii) restrict its access to financial market, (iv) lower its ratings, and (v) generate new risk factors. To avoid these inconveniences, financial institutions tend to remain relatively silent regarding their stress testing practices. When they are forced to disclose the results' details (e.g. regulatory requirements, market pressures, or transparency concerns), they often opt for less severe scenarios. This practice, common in the business, ensures guarding against any alarmism caused – rightly or wrongly – by the publication of the results (ACP, 2013).

Meanwhile, supervisors may also ask financial institutions to take actions in response to potential losses revealed by the test's results.[9] This could consist of hedging and liquidating risky positions, rising capital and provisions, portfolios' restructuring and cleaning out, etc. The more severe is the scenario, the more expensive are these responses. At least in the short term, this means high opportunity costs and a loss in competitiveness. Therefore, the supervisory constraint indirectly leads to moderate the scenarios' severity.

There are no theoretical nor empirical studies in the literature that look into the trade-off between the advantages (additional efficiency) and disadvantages (destabilizing effects) of stress testing results. In this context, a comprehensive communication strategy on the whole process is vital. This should include sufficient information on the test's starting assumptions, objectives, scope, and results, as well as a detailed timetable of the intended responses. Before detailing these points in the next sections, we end the present section emphasizing a key issue related to the management of the starting assumptions.

The starting assumptions should be updated as much as possible. The designer is called to react to any news likely to maintain, enhance, or change these assumptions. In a stress testing exercise, this should at least be done twice: in the beginning of the test and after having observed the effects of the responses to the test's results. The process has to be rather iterative, as shown in Figure 2.1. Keeping an appropriate technical documentation at each stage of the process is also crucial to reduce the implementation costs and increase the test's efficiency. Yet, the facts show a pretty slowness of these practices, mainly due to some limits in the used techniques and the information systems in place in most institutions (see BCBS, 2009). The 2007-09 crisis has confirmed the problems of integrating the test's results in the management process. This is manifested by a lack of internal and external debates on risk factors, investment opportunities and costs, selling and hedging strategies regarding potential losses, etc. These common management issues are augmented by organizational aspects related to the perception of stress tests inside financial institutions. In most cases, these are regarded to as a routine – and even a mechanical – exercise, taken apart from the other management tools.

[9] According to the Second Pillar of Basel II Accord, the supervisor is required to control any stress test performed by the depending financial institutions (CGFS, 2005, §. 765 and 778 (iii)).

2.4 Applications

Having listed the main stress testing objectives and applications, we discuss the originality of this risk management tool, highlighting its contributions to the existing framework.

2.4.1 Objectives

A stress test is generally conducted in response to a well-defined goal. This stage is all the more crucial as it influences the rest of the exercise. Thus, the first decision of the designer is to define one or more purposes to the exercise. The latter often vary across tests, depending on their specific features. According to Blaschke et al. (2001):

- An individual stress test focuses on "*possible market breakdowns* [...] *from the perspective of its effect on the value of the portfolio*". It aims to "*assist in the process of managing risks within a firm, and ensuring the optimal allocation of capital across risk-taking activities*".
- A systemic stress test focuses on "*potential externalities and market failures, for example* [...] *an evaporation of liquidity or a flight to quality*". It aims to "*help regulators to identify structural vulnerabilities and overall risk exposures in the financial system that could lead to disruption of financial markets*".

The central objective of individual stress tests is therefore to improve extreme risk management. Having identified the core risk factors makes it easier to optimize investments with respect to the risk culture inside the institution. BCBS (2009) distinguishes between two main applications of individual stress tests: (i) qualitative applications, related to the identifying procedures of relevant scenarios, and (ii) quantitative applications, allowing to assess the portfolio's resilience to severe losses, identifying adequate strategies to reduce exposures, maximizing income, and preserving capital.

On the other side, a systemic stress test is primarily designed to assist regulators and supervisors. These exercises allow identifying both idiosyncratic and structural vulnerabilities. Drehmann (2008) outlines three main alternative objectives for systemic stress tests: assumptions' validation, decision-making, and communication. He shows that a trade-off is often necessary between these different objectives. That is, the criterion of accuracy is essential when the objective is to test the relevance of the starting assumptions. When it comes to help decision-making, the flexibility of the assumptions and the models' performances (e.g. forecast properties) are appreciated. All these criteria become of less importance when the test is performed to run a communication campaign on a specific aspect of the stressed portfolio. In this case, the clarity of the considered scenario in terms of storytelling would be required. Even though systemic stress tests offer different applications, they are usually designed to replicate past extreme events. They are therefore often considered as *resilience tests*, which is a rather reductive term.

2.4.2 Stress testing and risk management

The stock market crash of 1987 has revealed a pressing need for a more efficient risk management. This has been the starting point of several studies aiming to propose new management

tools able to ensure appropriate coverage against extreme events of the size of the 1987 crash. In 1996, in its first version of the RiskMetrics system, the U.S. bank J.P. Morgan has published the principles of a novel risk management tool, called "Value-at-Risk" or VaR. This tool has the particular advantage of combining two desirable properties for any risk manager: simplicity and flexibility. Initially reserved for large financial institutions, it has gradually been adopted by all market players to such an extent that many have referred to as "the VaR revolution". Under the leadership of most national and international regulatory authorities, it became a recognized official standard (see BCBS, 2006, 2009; CRMPG III, 2008).

However, after years of practice, it was found that frequent management errors have been induced by the use of the VaR, particularly in periods of market turmoil. This is where the first stress tests have been introduced and designed to manage extreme risks. The financial crisis of 2007-09 revealed even more significant losses, mostly due to an almost exclusive use of the VaR. These problems have renewed research interests around stress testing exercises to address the VaR drawbacks.

More specifically, the criticism directed at the use of the VaR in risk management relate to:

- Common limits for many other management tools – even for some stress tests – related to the underlying risk model used to estimate the VaR (see Appendix 2.B).
- Specific limits related to the conceptual and theoretical properties of the VaR.

Among the second series of limits, the most important are: (i) the subjective choice of VaR's parameters, such as the confidence level and the holding horizon, (ii) the lack of precision, as its performance decreases for extreme losses above the 99.9% confidence level (see Alexander and Sheedy, 2008), (iii) the lack of flexibility, as it gives no information about the size nor the nature of extreme losses beyond the considered confidence level, and (iv) the non-compliance with the *axioms of coherence* – especially the axiom of *subadditivity* – required for any risk measure (Artzner et al., 1999). The last point, revealing failures to account for diversification effects, is the source of most observed errors.

Conditional VaR (or CVaR) is a derivative measure of the VaR, allowing to reduce the last two limits.[10] It measures the average of losses beyond the confidence level, and provides therefore a new indication on the structure of extreme losses. This measure also satisfies all the axioms of coherence (Berkowitz et al., 2011). However, it provides only partial information about extreme losses while sharing the other common limits of risk measures.

In response, stress testing exercises brought two major innovations:

- Exploring in greater depth the tails of the distribution function of the portfolio's value beyond the confidence levels commonly used in the VaR and the CVaR.
- Connecting, in a flexible and explicit fashion, potential extreme losses in the portfolio's value to the causing events i.e. the stress scenarios. This allows deeply and explicitly identifying the sources of risk, as opposed to the pure statistical approach in the case of the VaR. Such a technique allows measuring the contribution of each risk factor, in each scenario, to the potential change in the portfolio's value.

[10] The CVaR is also known under alternative appellations, such as Conditional Loss (CL), Expected Tail Loss (ETL), Expected Shortfall (ES), Tail Conditional Expectation (TCE), and Worst Conditional Expectation (WCE). Yet, some minor discrepancies between these different concepts are sometimes encountered in the literature.

In what follows, we denote by "conventional management tools" all risk management instruments but stress tests.

2.5 Scope

The scope of a stress test can be understood through three main dimensions, namely the choice of the stressed portfolios, the underlying risk factors, and the test horizon.

2.5.1 Portfolios

A stress test is designed to measure the possible impact of a given scenario on the trading and non-trading positions that form the financial institutions' lending and trading books. These portfolios can be held by departments and business units of a same institution (individual stress tests) or by a group of institutions (systemic stress tests). Long and short balance sheet and off-balance sheet items may all be considered. Hereafter, we use the term "portfolio", regardless of its holder and composition.

More than three quarters of realized stress tests are related to trading portfolios (see CGFS, 2005; DBB, 2004, 2007). These exercises, essentially individual stress tests, have been subject to the earlier research works on the topic (see, e.g., Kupiec, 1998; Cherubini and Lunga, 1999; Berkowitz, 2000; Breuer and Krenn, 1999, 2001; Kim and Finger, 2000; Bee, 2001; Aragonés et al., 2001). The leading arguments in favour of this choice of portfolios consist of the use of relatively straightforward mark-to-market techniques in the valuation of the related risk profile as well as the availability of regular and high frequency data.

In contrast, testing for derivatives and off-balance sheet items are less common. This is mainly due to the private character of the required data, data limitations for most new products, and other practical issues related to the complexity of the exposures and the required valuation models (BCBS, 2009). Modelling profits, taxes, and dividends is also marginalized in practice, due to the granularity and the lower importance of these aspects in the payoff profile of the portfolio.

2.5.2 Risk factors

Each of the positions listed above is exposed to a set of risk factors of different natures (market risk, credit risk, liquidity risk, operational risk, etc.). The nature of the position often determines that of the underlying risk factors. However, given the large number of risk factors to which certain portfolios may be exposed, a prior selection is often needed before starting the test. Only "core" risk factors, presenting a significant impact on the portfolio's value, will be retained. Changes in core risk factors' value (in level or in return) are typically used in stress scenarios, while assumptions concerned with changes in volatility or correlation are quite rare.

The remaining risk factors, referred to as "peripheral", are discarded from the stress scenario. These factors are either:

- Completely excluded from the specification of the valuation model. In which case, only the core risk factors are considered.

- Included in the valuation model, but only through ad hoc values (e.g. last observed value, historical average, expected value, etc.). In other words, these factors are considered without any change that could cause severe losses in the portfolio's value.

Specifying particular interactions between core and peripheral risk factors depends on the starting assumptions (see Kupiec, 1998; Kim and Finger, 2000).

Most studies and surveys conducted on major financial institutions reveal a recurrent use of certain classifications of risk factors (CGFS, 2001, 2005; Fender and Gibson, 2001; DBB, 2004, 2007). For the asset-side of the balance sheet (i.e. long positions), market risk is often split into price-wise risk factors (stocks, commodities, derivatives, etc.) and rate-wise risk factors (interest rates, forex, risk premium, etc.). Some systemic stress tests in particular also include market liquidity risk factors. However, as for any dynamic phenomena, modelling market liquidity is subject to several practical issues, which explains the reduced number of tests of this nature.

On the liability side, risk factors are often related to funding liquidity. These risks mainly affect the portfolio's value through: (i) random behaviours of short-term liabilities (e.g. deposits, overnight debts, and interbank liabilities), (ii) creditworthiness of counterparties (rating downgrades, market liquidity), (iii) funding costs (interest rates, risk premia), (iv) collateral quality (required amounts and haircuts), (v) specific and generalized restrictions acceeding the money market, and (vi) regulatory and legal measures (deposit insurance), etc. These subclasses of funding liquidity risk are more documented than market liquidity risk, due to recent studies using relatively simple valuation models (CGFS, 2005; BCBS, 2009). Combinations of both liquidity risks (e.g. liquidity spirals phenomena) are still developing.

Finally, unlike trading credit risk (considered as a market risk factor), studying non-trading credit risk is still less common in practice, mainly due to data issues. The same applies for operational risk (BCBS, 2009).

2.5.3 Horizon

The horizon of a stress test consists of the following:

- The period during which the changes applied to risk factors persist. During this time, additional changes could be introduced.
- The period during which the impact of these changes on the value of the stressed portfolio are still recorded.

The two periods may be coincident, but not necessarily.

In what follows, we call "horizon", all the intermediate dates separating the initial changes in risk factors and the last period where their impact is measured on the portfolio's value.

The choice of the horizon depends on several criteria, such as the objectives of the test, the nature of the portfolio, and the underlying exposures. For an individual stress test, for example, it should be designed to take into account the time needed to hedge or to liquidate certain of the portfolio's positions. The speed at which these operations are completed is often determined by the market liquidity in times of stress, the size of the position, the extent of the generated externalities, the timing and the duration of any response from the concerned institution, the regulator and/or the government, etc. Even though the most severe impact is

not necessarily achieved at the end of the test horizon, losses in intermediate dates are often more severe than in the first period.

In the absence of any convention about the optimal stress testing horizon, it is frequent to assume:

- Shorter terms for individual stress tests, in the presence of market risk factors. The Basel Committee on Banking Supervision recommends a 10-day holding period for the most liquid positions. During this time, the portfolio is assumed to be non-transferable, and therefore accumulates the impacts of the risk factors' changes (see also DBB, 2004, 2007). For credit risk exposures, the horizon is usually more important, to allow accounting for the realization of the considered changes, their transmission to credit portfolios, and the decision-making process. For example, CEBS (2010) and FRB (2009, 2011) consider a 2-year horizon for the tests applied on the major EU-wide and U.S. banks, respectively.

- Longer terms for systemic stress tests, where they usually take place over several quarters or years. This is partly due to the frequency of the used data and models, allowing for slower responses. A long horizon also allows capturing the dynamics of systemic exposures and the transmission mechanism to individual portfolios, public and private responses, and negative externalities (e.g. contagion, second-round effects, and feedback effects on the real economy).

Having defined the objective and the scope of the test, the next steps in the process consist in:

- Setting the changes to be applied to some (or all) risk factors at the first period of the horizon, and the way these changes evolve through the intermediate periods up to the horizon term.

- Measuring, at each period, the corresponding impact on the portfolio's value.

These two steps could be seen as a practical implementation of the starting assumptions discussed earlier. The first one refers to the risk model, and includes the main components of the stress scenario. The second step refers to the valuation model, measuring the impact of the scenario on the portfolio's value. Remind that the two models operate simultaneously if the stress scenario is set according to a systematic research method.

2.6 Design of scenarios

The conception of a stress scenario requires a quantitative definition of: (i) the size of the changes to be applied – through simulation – to risk factors ("the initial shock"), (ii) the way these changes persist, evolve, and interact trough the test horizon ("the post-shock"), and (iii) the corresponding probabilities of occurrence. We refer to these three components as "the scenario parameters".

The initial shock is a common part for all stress scenarios, which is not the case for the remaining two parameters. In what follows, we qualify as "dynamic" (resp. "static"), a scenario that includes (resp. omits) post-shock events. We also qualify as "stochastic" (resp. "deterministic"), scenarios whose probabilities are explicitly set (resp. omitted).

2.6.1 Typology and selection criteria

The classification of stress scenarios relies on theoretical and operational criteria. Setting the scenario parameters can be based on: (i) observed past changes in risk factors (historical scenarios), (ii) theoretical assumptions (hypothetical scenarios), or (iii) a combination of both (hybrid scenarios).

However, as hypothetical scenarios could also be – directly or indirectly – based on past events, an alternative typology seems more appropriate to better discriminating these approaches. Here, we adopt a classification that distinguishes between purely hypothetical scenarios, also called "forward-looking" scenarios, and historical or "backward-looking" scenarios. In the remainder of this thesis, we qualify as "historical" (resp. "hypothetical"), scenarios exclusively based on a retrospective (resp. prospective) approach.

The choice between historical and hypothetical scenarios depends particularly on:

- The adequacy to the stressed portfolio and the context under review. This relies on the trade-off severity-plausibility.
- The compliance with the assigned test objectives. This allows improving the test's utility.
- The implementation costs, depending on the time allocated to the exercise, and to the human, material, and financial resources in place.

Considering historical scenarios, by replicating past extreme events, could reduce the conception costs and the scenario's subjectivity.[11] These scenarios may, however, mismatch with the nature of the stressed portfolio (e.g. including new assets with new exposures), the space-time context in which the exercise is realized (e.g. structural breaks, new regimes of risk factors' dependence structure), and the test's objectives. A hypothetical scenario, based on more flexible assumptions, offers a better consideration of these aspects. Unfortunately, the cost of such scenarios may present a material constraint, to which is added the human challenge consisting in demonstrating the objectivity of the considered assumptions. Hence, a trade-off flexibility-cost is often necessary. A first solution to this trade-off may come from hybrid scenarios. The latter are designed on the basis of past events, without producing a full replication of a specific episode.

Among researchers and market players, it is usual to consider that historical scenarios are necessarily more credible than hypothetical ones, as the former are based on manifest and observed events (see, e.g., Berkowitz, 2000; Breuer and Krenn, 1999, 2001). However, the credibility conferred to stress scenarios by historical facts could be more than offset by the inadequacy of these scenarios to the stressed portfolio and to the present and future contexts.

The distinction between historical and hypothetical stress scenario is the mostly cited in the literature. Four other classifications are still occasionally used, using alternative criteria. The first classification allows distinguishing between stress scenarios in the strict sense and "sensitivity tests". The former, considers changes applied – simultaneously and/or separately – to several risk factors, while the latter takes into account changes in a single risk factor. Beyond this difference, which may seem rather banal at first glance, other differentiation criteria are sometimes cited in the literature. These include, for example, a longer test horizon in a stress

[11] See CGFS (2005) for a comprehensive list of historical events considered as stress scenarios in practice.

scenario and its ability to identify the sources of the shock, while a sensitivity test is rather basic (CGFS, 2005).

The second classification distinguishes between "traditional" – also called "ad hoc" – scenarios and "modern" scenarios. The former include historical and hypothetical scenarios whose parameters are set subjectively. The scenarios could be deterministic or stochastic. In the latter, the scenario parameters are set following systematic methods allowing to search for the worst case scenarios for the stressed portfolio. Here, the scenario is stochastic by nature.

Another classification distinguishes between "pure" scenarios, assuming changes in risk factors' values, and "generalized" scenarios relying on changes in the underlying risk model. Hence, in turmoil periods, for example, the volatility and the dependence structure of risk factors are also likely to change. This modifies the underlying distribution (e.g. new parameter values, a new distribution family in case of structural breaks or regime switching, etc.). Note that the second category generalizes the first one, hence the denomination.

A final distinction allows highlighting the designer and/or the objectives of the test. We distinguish therefore "event-driven" scenarios from "portfolio-driven" scenarios. The former type is often carried out at the request of the risk manager, in response to recent informations and/or events (e.g. uncertainty in the marketplace). The latter responds to a Management request, without necessarily being motivated by a particular change in the risk environment.

In what follows, we use the term "scenario" to refer to the changes applied to risk factors, their evolution, and their probability, regardless of the starting assumptions, the designer, the objectives, and the scope of the test. In addition, we qualify as "univariate" (resp. "multivariate"), a scenario focused on a single (resp. several) risk factor(s).

2.6.2 Parameter setting

This section describes the strictly speaking setting procedure of a stress scenario. This stage comes after the choice of the objective, the scope, and the type of scenario, and consists in setting the values of the scenario's parameters. Considered as the most important stage of the process, this is the practical side of the conceptual starting assumptions

Shock and post-shock

The initial shock parameter is present in all stress scenarios. It allows setting the changes to be applied on each risk factor in the initial period. The post-shock parameter instead is only present in dynamic scenarios. It allows setting the dynamics of the risk factors' changes following the initial shock, up to the term of the test's horizon. These dynamics may include a shock persistence or a shock evolution, with or without interaction between risk factors – including peripheral factors that have not been affected by the initial shock – and possible new shocks at each period. For simplicity purposes, we focus the presentation on shocks consisting of changes in the risk factors' returns.

Let $h = 1, \ldots, H$ be the test horizon, where $h = 1$ is the initial date and $h = H$ is the horizon term. Consider a portfolio exposed to N risk factors whose past returns are given by the $N \times T$ matrix $R1$, where

$$R1 = (r_{i,t}), \qquad i = 1, \ldots, N; t = 1, \ldots, T$$

where $r_{i,t}$ is the return of the i-th risk factor at date t, and T is the sample size. This sample can be get using real data (i.e. historical observations of risk factors' returns) or by simulation.

Both the initial shock and the post-shock are random future values of the risk factors' returns. These values are obtained by combining past data, a risk model, and simulation methods. This procedure allows getting, at each future date, a series of returns or "realizations" for each risk factor. According to the starting assumptions, one realization per date and per risk factor will be considered to form the shock and the post-shock of the scenario.[12]

The initial shock is applied at the initial date of the test horizon, i.e. at $h = 1$, matching with $T + 1$. The post-shock is applied at the next periods up to the horizon term, i.e. at dates $h = 2, \ldots, H$. The changes applied, through the test horizon, to each risk factor's return can therefore be given by a $N \times H$ matrix $R2$, where

$$R2 = (r_{i,h}), \quad i = 1, \ldots, N;\ h = 1, \ldots, H$$

where $r_{i,h}$ denotes the return of the i-th risk factor at horizon date h.

Therefore, the initial shock $R2_1$ is given by the first column of $R2$, and the post-shock by the remaining vectors forming a sub-matrix $R2_h$. $R2_1$ and $R2_h$ could be written as follows

$$R2_1 = \begin{pmatrix} r_{1,1} \\ \vdots \\ r_{i,1} \\ \vdots \\ r_{N,1} \end{pmatrix}, \quad R2_h = \begin{pmatrix} r_{1,2} & \cdots & r_{1,h} & \cdots & r_{1,H} \\ \vdots & \ddots & & & \vdots \\ r_{i,2} & & r_{i,h} & & r_{i,H} \\ \vdots & & & \ddots & \vdots \\ r_{N,2} & \cdots & r_{N,h} & \cdots & r_{N,H} \end{pmatrix}$$

Having defined the risk factors' returns at each horizon date, the corresponding impact on the stressed portfolio's return r is measured through an appropriate valuation model.

Probability

The probability of a stress scenario measures the likelihood of the considered shock and post-shock. Depending on the approach used to perform the trade-off severity-plausibility, the probability can be set upstream or downstream from the setting of the shock and the post-shock. Note that the nature of this quantitative concept is different from that of plausibility, which is rather qualitative and subjective. Indeed, for a given probability of a scenario, the corresponding plausibility level may be perceived differently by different market players (e.g. according to their risk culture).

In a dynamic scenario, the probability often holds for the whole scenario, i.e. for both the shock and the post-shock. This is especially the case when no more shocks are introduced after the initial shock. In this case, the probability of the initial shock is identical to that of the related post-shocks up to the horizon term. These two quantities can differ if, for example, two different risk models are used to infer the probability of the initial shock and that of the post-shock. This could also be the case if new exogenous events are included during the intermediate dates

[12] The simulated realizations form the expected distributions of risk factors' returns, at each date. This distribution can then be used in different analyses, such as stress testing and forecasting. In a stress test, some points of the distribution causing a severe impact on the portfolio's value are considered. Meanwhile, a forecast exercise consists in considering the empirical mean or median of the expected distribution.

of the horizon (e.g. new risk factors, additional returns' changes, new dependence structure between risk factors, etc.). However, given their complexity, these assumptions are quite rare in practice.

The order of setting the three parameters of the scenario present its own importance. Remind that three alternative options are still open to the designer (see section 2.2.2). When the shock is set first, a subsequent modification may sometimes be necessary if, for example, the designer remarks that the corresponding probability is lower than the predefined plausibility threshold. In this case, an upward revision in the size of the shock can be considered. Conversely, when the plausibility or the plausibility threshold is fixed first, the corresponding shocks – obtained via the systematic research for the worst case method, for example – should not be subject to any revision. If such a decision has to be taken, it would be a carrier of subjective judgments.

2.7 Results

Before presenting the key fields where stress testing results could be used and their implications, two preliminary remarks are worthy of note:

- A stress scenario does not necessarily reflect a true image of future risks or a real perception of future events by the designer. The considered scenarios are certainly possible, but their probability is generally low. In most cases, the considered risk factors' changes match with extreme values in the underlying distribution function. In other words, a stress scenario is a simulation of future events and not a forecasting exercise. This is why the use of the test's result is much more important than the result itself.
- A stress test should not be seen as a *silver bullet* for all extreme risk management problems. Accompanying tools, such as VaR, CVaR, FSIs, and EWS, are also required to prevent model risk and any excessive dependence on this tool (see Berkowitz, 2000; BCBS, 2009).

2.7.1 Decision-making

The test's results provide a new information on potential risk sources for the stressed portfolio, as well as the nature of the relation between the expected returns and the risk profile of each position. This information can be used at different levels of the decision-making process, including:

- The trading level: to reveal the potential vulnerabilities of different markets and products.
- The managerial level: to compare risks between different asset classes and exposures, and to assess the need for potential investment limits or reallocations.
- The executive level: to compare between the risk profile of the stressed portfolios and the risk culture of the holding investors, ensuring an optimal capital allocation.

The next paragraphs discuss other utilization modes by different market players. These are mainly represented by supervisors, economic and business partners of financial institutions, and rating agencies.

(i) **Supervisors**

Whatever the type of the test (internal, external, or mixed; individual or systemic), the communication of the results can be used for different purposes by the supervision authorities, in particular, for:

- Individual supervision and regulation of financial institutions
  The results allow documenting the institution exposures and vulnerabilities. These results may reveal, for example, small margins regarding some regulatory ratios. In this case, the supervisor could encourage the concerned institution to take adequate measures to improve these indicators. The test's results can also be used to assess the financial soundness of the institution and its ability to meet its commitments. These criteria partly determine the refinancing conditions of financial institutions and the money market efficiency.
- Systemic risk monitoring and macro-prudential analysis
  Given the considered systemic and idiosyncratic risk factors, the test's results allow assessing the soundness of the financial system, the related vulnerabilities, and the feedback effects on the real economy.
- Policy issues
  The results also allow assessing the adequateness of public policies, management plans, and regulatory measures. Additional or alternative actions could therefore be considered accordingly.

(ii) **Economic partners**

Other market players can also use the results of individual and systemic stress tests. The provided message, including the results and the intended responses, is added to the public information in place. These two factors influence the decision-making of concerned economic agents (e.g. depositors, shareholders, lenders, insurers, etc.). Some investors may decide, for example, to reallocate their funds to lower risky portfolios, institutions, or systems. Others may refuse funding or insuring heavily exposed institutions that are not considering significant responses to the test's results, and so on.

(iii) **Rating agencies**

The test's results and the intended responses can be incorporated into the financial rating criteria of the stressed portfolio, the financial institution and its creditors, the financial system, and even the related State. Results revealing a high vulnerability towards a particular risk factor can therefore lower the rating of the institution for this exposure, as well as that of related creditors. This is particularly true when the intended responses are deemed insufficient by the rating agency. These intentions also play a major role in the rating of financial systems and governments, where factors like the tax regime, the deposit-insurance system, and official bailouts are usually looked into. Given the impact of ratings on market players' decision-making, the responses to the test's results should be as reassuring as possible.

2.7.2 Complementing other management tools

Beyond issues related to conceptual an operational aspects of a stress testing exercise, the use of the obtained results also causes several issues. One of the most important relies

on the treatment to be given to the stress test's results with respect to those of alternative risk management measures. The stress test is indeed a relatively recent tool added to the conventional framework, mainly based on the VaR. How then to use stress testing results to make it a complement rather than a substitute of the VaR's results provided in a more frequent basis? The risk manager often faces a practical issue: *by what means associating VaR's results, only valid under normal market conditions, and stress testing results, based on a few specific adverse scenarios?*

In practice, most stress tests are realized in a *stand-alone* basis, and the related results are used separately (Aragonés et al., 2001). Berkowitz (2000) is one of the first to emphasize the importance of this issue in real life. He has proposed a unified framework allowing to combine stress tests' and VaR's results to form a single *meta-distribution* of the expected portfolio's value (see Appendix 2.A). Aragonés et al. (2001) have presented an empirical illustration of this approach, for a portfolio of commodities, with one scenario and one risk factor. However, this approach imposes some strict assumptions that makes it difficult to carry out in practice. These are mainly related to: (i) the choice of risk models for the VaR and the stress scenarios, (ii) and the method of setting the probabilities of the stress scenarios. Despite the usefulness of this approach and the improving fields left open, Aragonés et al. (2001) is, to the best of our knowledge, the unique related empirical study. This idea will further be studied in chapter 5.

2.8 Conclusion

This chapter has analysed the conceptual and operational stages of financial stress tests and discussed the main related issues. The test is part of a complex and delicate process. It combines a large number of heterogeneous stakeholders, responsible for setting the test's objective and scope, the scenarios' assumptions, the results' analysis and decision-taking. Only a deep analysis of the whole process allows determining its relevance.

The study has revealed a relative slowness of stress testing practices, due to some limits in the used techniques and the information systems in place in most financial institutions. The recent crises have also confirmed the problems of integrating the tests' results in the risk management process, revealing a lack of internal and external debates on risk factors and investment opportunities, among others. These management issues are augmented with organizational aspects related to the perception of stress tests inside financial institutions, as they are, in most cases, regarded to as routine exercises, taken apart from the other management tools.

Moreover, this chapter revealed that most stress testing exercises are still experiencing several conceptual and practical challenges in all stages of the process. These mainly relate to the conception of risk models for the portfolio's underlying risk factors, selection procedures of credible stress scenarios, and comprehensive valuation models - especially for systemic stress tests.

References

Autorité de Contrôle Prudentiel (ACP) (2013): Analyses et Synthèses: Stress Tests sur le Système Bancaire et les Organismes d'Assurance en France, No. 11, January.

Alexander, C. and E. Sheedy (2008): Developing a Stress Testing Framework Based on Market Risk Models, *Journal of Banking and Finance*, 32(10), 2220-36.

Aragonés, J.R., C. Blanco, and K. Dowd (2001): Incorporating Stress Tests into Market Risk Modeling, *Derivative Quartely*, 7(3), 44–9.

Artzner, P., F. Delbaen, J.M. Eber, and D. Heath (1999): Coherent Measures of Risk, *Mathematical Finance*, 9(3), 203–28.

Basel Committee on Banking Supervision (BCBS) (2009): Principles for Sound Stress Testing Practices and Supervision, Consultative Document, May 20.

Basel Committee on Banking Supervision (BCBS) (2006): Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework – Comprehensive Version, *Bank for International Settlements*, June 26.

Bee, M. (2001): Mixture Models for VaR and Stress Testing, *ALEA Tech Reports*, Università di Trento, No. 12.

Berkowitz, J. (2000): A Coherent Framework for Stress-Testing, *Journal of Risk*, 2(2), 5-15.

Berkowitz, J., P. Christoffersen, and D. Pelletier (2011): Evaluating Value-at-Risk Models with Desk-Level Data, *Management Science*, 57(12), 2213–27.

Blaschke, W., M.T. Jones, G. Majnoni, and S. Martinez Peria (2001): Stress Testing of Financial Systems: an Overview of Issues, Methodologies, and FSAP Experiences, *International Monetary Fund*, Working Paper No. 01/88.

Board of Governors of the Federal Reserve (FRB) (2011): Comprehensive Capital Analysis and Review: Objectives and Overview, March 18.

Board of Governors of the Federal Reserve (FRB) (2009): The Supervisory Capital Assessment Program: Overview of Results, May 7.

Breuer T. and G. Krenn (2001): What is a Plausible Stress Scenario? in: L. Kuncheva et al. (Eds.), Computational Intelligence: Methods and Applications, ICSC Academic Press, Bangor, 215–21.

Breuer, T. and G. Krenn (1999): Stress Testing: Guidelines on Market Risk, Volume 5, Oesterreichische Nationalbank.

Cherubini, U. and G.D. Lunga (1999): Stress Testing Techniques and Value-at-Risk Measures: A Unified Approach, *Decisions in Economics and Finance*, 22(1–2), 77–99.

Čihák, M. (2007): Introduction to Applied Stress Testing, Technical Report, *International Monetary Fund*, Working Paper No. 07/59.

Committee of European Banking Supervisors (CEBS) (2010): Results of 2010 EU-Wide Stress Testing Exercise, July 23.

Committee on the Global Financial System (CGFS) (2005): Stress Testing at Major Financial Institutions: Survey Results and Practice, *Bank for International Settlements*, January 18.

Committee on the Global Financial System (CGFS) (2001): A Survey of Stress Tests and Current Practice at Major Financial Institutions, *Bank for International Settlements*, April 25.

Counterparty Risk Management Policy Group (CRMPG III) (2008): Containing Systemic Risk: The Road to Reform – The Report of the CRMPG III, August 6.

Deutsche Bundesbank (DBB) (2004): Stress Tests at German Banks: Methods and Results, *Monthly Report*, October, 75–84.

Deutsche Bundesbank (DBB) (2007): Stress Tests: Methods and Areas of Application. *Financial Stability Review*, November, 97–112.

Drehmann, M. (2008): Stress Tests: Objectives, Challenges and Modeling Choices, *Sveriges Riskbank Economic Review*, 2(2008), 60–92.

Fender, I. and M.S. Gibson (2001): Stress Testing in Practice: a Survey of 43 Major Financial Institutions, *Bank for International Settlements*, Quarterly Review, June 1.

Haldane, A. (2009): Why Banks Failed the Stress Test, Speech to Marcus-Evans Conference on Stress Testing, London, February 9-10.

Institute of International Finance (IIF) (2008): Final Report of the IIF Committee on Market Best Practices: Principles of Conduct and Best Practice Recommendations, Financial Services Industry Response to the Market Turmoil of 2007-2008, July 21.

Jones, M.T., P. Hilbers, and G. Slack (2004): Stress Testing Financial Systems: What to Do When the Governor Calls, *International Monetary Fund*, Working Paper No. 04/127.

Kim, J. and C.C. Finger (2000): A Stress Test to Incorporate Correlation Breakdown. *RiskMetrics Journal*, 1, 61-75.

Kupiec, P.H. (1998): Stress-Testing in a Value at Risk Framework, *Journal of Derivatives*, 6(1), 7-24.

Appendix

2.A Value-at-Risk (VaR)

The VaR of a risky portfolio is the maximum loss, over a holding time horizon h, not exceeded with a high probability or "confidence level". Thus, the probability that the loss exceeds the VaR is less than or equal to a low probability α called "coverage level".[13]

Statistically, the VaR is the α–quantile of the profits and losses (P&L) density function F of the portfolio. Thus

$$VaR(\alpha) = F^{-1}(\alpha) \Leftrightarrow \mathbb{P}\mathrm{r}[r < Var(\alpha)] = \alpha$$

where $r \in \mathbb{R}$ is the portfolio's value or return and $\alpha \in (0, 1)$.

Usually, the conditional density function is used to estimate the VaR, which is then given by

$$VaR(\alpha|\Omega_t) = VaR_{t+1|t}(\alpha) = F^{-1}_{t+1}(\alpha|\Omega_t)$$

where $F_{t+1}(\cdot)$ is the conditional density for date $t+1$, and Ω_t is the available information at t.[14]

The estimation of the VaR depends on three elements: the function F, the confidence level α, and the holding time h. The specification of F is the key factor in this approach. The main used methods include:

- Non-parametric methods: including historical simulations (e.g. simple, weighted, and filtered methods) and numerical simulations (e.g. Monte Carlo and Bootstrap procedures), etc.
- Parametric models: including models of type ARFIMA, ARCH-GARCH, stochastic volatility, etc.
- Semi-parametric models: including Extreme Value Theory (EVT), CAViaR models, etc.

Depending on the used method, also called "internal risk model", the Basel Committee defines the required amount of regulatory capital to hedge the market risk losses revealed by the VaR. This amount is given by the following formula

[13] For internal and regulatory considerations, α is usually set to 1% or 5%. For market risk, h is set to 1 or 10 days. Credit and operational risks instead require longer periods, to allow accounting for market liquidity especially.

[14] In practice, $-F^{-1}_{t+1}(\alpha|\Omega_t)$ is used instead, to allow dealing with positive VaR numbers.

$$CR(t) = \max\left(Var_{t+10|t}(0.01), \frac{k}{60 \cdot \Sigma_{i=1}^{60} VaR_{t+10|t-i+1}(0.01)} \right) + c$$

where $VaR_{t+10|j}(0.01)$ is the VaR for a 10-day holding period with 99% confidence level, estimated at date t given the available information at date j. The multiplier k, with $3 \leq k \leq 4$, refers to the *stress factor*, defined by the regulator depending on the internal risk model, and c captures the idiosyncratic risks but market risk.

The use of the VaR as a risk management tool has been criticized for different aspects. Alternative tools have been proposed. The most used in practice is the so-called Conditional VaR (CVaR). Allowing to measure the average of losses beyond the quantile α used in the VaR, the CVaR is given by

$$CVaR(\alpha) = \mathbb{E}[r|r < VaR(\alpha)] = \frac{1}{\alpha} \int_{\alpha}^{1} F^{-1}(p)\mathrm{d}p$$

2.B A unified framework VaR – stress tests

Berkowitz (2000) has been among the first to present a unified framework allowing to combine the results of stress tests and those of the VaR. This approach is carried out in the following steps:

- Set a probability to each stress scenario, which is also that of the related test's results. These probabilities may be fixed or dynamic, evolving according to a stochastic process depending on their past values, recent values of a set of economic and financial indicators, or simply based on expert judgments and forecasts. The probabilities of different scenarios can be independent or correlated. This is an interesting research topic to be explored when using stress tests including several scenarios.
- Using the appropriate valuation model, form the distributions of profits and losses (P&L) of the portfolio. This is done based on the risk models used to specify the risk factors in the VaR framework and in each of the stress scenarios.
- Mixed distributions methods allow forming a unified *meta-distribution* of the P&L, based on these P&L distributions. The corresponding probabilities of the VaR and the stress scenarios are used as mixing weights.
- Different risk measures (e.g. Stress-VaR, Stress-CVaR) can then be obtained from the unified distribution. These measures are therefore based on two management tools: stress tests and a traditional risk management tool.

Aragonés et al. (2001) have presented an illustrative application of the approach. To do so, they have used a portfolio of commodities, with a single risk factor and a single stress scenario. The Extreme Value Theory (EVT) is used as a risk model for both the underlying risk factor and the portfolio's P&L, while the mixing probabilities are set subjectively.

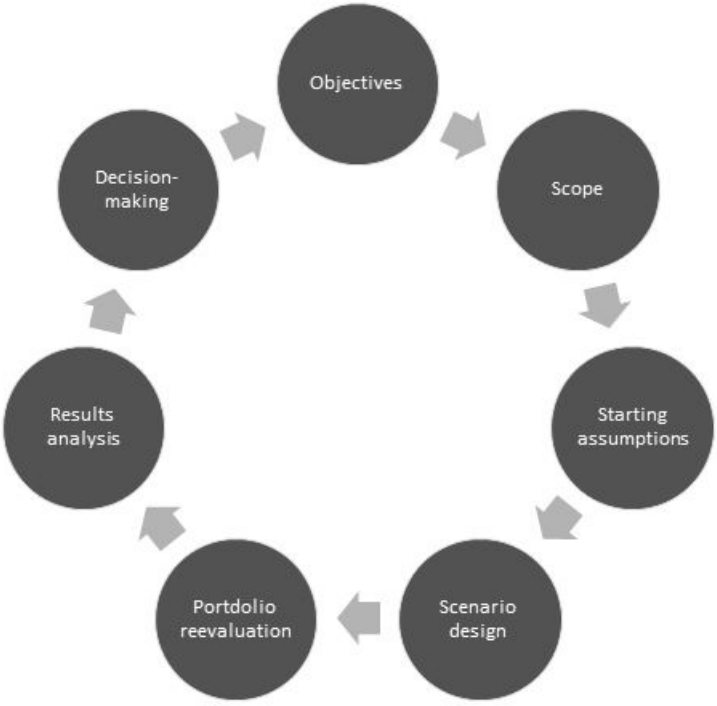

Figure 2.1: A simplified overview of a stress testing exercise

3. Extreme Risk Modelling A Sequential EVT – Pair-copulas Approach

Abstract

This chapter presents a semi-parametric EVT copula-GARCH (risk) model for financial return series. The marginal distributions of the returns are specified using the Extreme Value Theory (EVT), putting a specific emphasis on extreme returns. The joint distribution is then built using the pair-copulas theorem, based on the marginal distributions and the pair dependence structures. The model performance is assessed for three sets of assets, namely equity indices, exchange rates, and commodity prices. The empirical results support a better static and dynamic properties of the presented model compared to most common specifications used in practice. The proposed model and the alternative specifications are then carried out to perform stress testing exercises on analytical portfolios, where financial returns are considered as risk factors. The results show that the use of a wide range of risk models produce significantly different results, in terms of the corresponding stress scenario and the generated impact on the portfolios. Hence, considering flexible and consistent specifications, as in the proposed model, ensures a better credibility of the stress scenario and enhances the usefulness of the stress testing results.

Keywords: Financial time series, EVT, pair-copulas, R-vine, stress testing, scenario

JEL classification: C5, G11, G17

3.1 Introduction

The relevance of a stress testing exercise is highly related to the credibility of the considered scenarios. These are defined based on a risk model allowing to specify the marginal distributions and dynamics, as well as the dependence structure of the underlying risk factors. The estimated risk model is carried out to simulate a set of scenarios. Among these scenarios, those supposed as *severe yet plausible* are used for stress testing purposes. Hence, the stress testing result is conditioned by the prior choice of the risk model.

A risk model is made up of assumptions designed to specify the statistical properties of the risk factors facing a variable of interest. For a financial portfolio, risk factors are usually given by the returns, the volatilities and the dependence structure (e.g. correlation) of the underlying financial assets. Since the seminal works of Mandelbrot (1963a,b), the literature has highlighted a series of statistical properties (hereafter "stylized facts") shared by most financial assets' time series. It is shown that the empirical distributions of financial returns are left-skewed and leptokurtic. The return series also exhibit a weak autocorrelation, a strong heteroskedasticity, and volatility clusters. Moreover, past returns often have asymmetry and leverage effects on actual volatilities. In a multivariate framework, extreme returns are shown to be strongly dependent. This dependence is nonlinear, dynamic, and even stronger for bear markets (see Appendix 3.A).

To capture these stylized facts, various risk models have been proposed. Some of them have been used for simulation-based exercises, such as forecasting and stress testing. Over the recent period, however, significant errors have been recorded in this area. For example, several supposedly stress scenarios have proven to be harmless, while unexpected risk factors' moves and portfolios' losses have been observed. To explain this failure, analysts have put forward the inadequacy of the used risk models. Indeed, due to practical issues (e.g. data, expertise, and time requirements, communication issues, etc.), only simple model specifications have been considered. The latter often lack of flexibility and omit one or more stylized facts, which explains the gaps between the expected and the observed values.

The recurrent and increasing losses generated by these errors have motivated a crucial need for more reliable modelling frameworks. Recent advances in quantitative methods (e.g. econometrics, software programs, etc.) have favoured this kind of initiatives. In this chapter, we propose a multivariate risk model based on a flexible specification of financial returns, with a specific emphasis on extreme values. The model is structured in two parts. The first part specifies the conditional marginal distributions of the returns, associating the Extreme Value Theory (EVT) with an ARMA-EGARCH filter. The second part specifies the underlying dependence structure, based on the pair-copulas theorem.

The rest of this chapter is presented as follows. Section 3.2 reviews the main approaches to model financial returns. Section 3.3 presents the data used for the analysis and verifies the presence of the main stylized facts. Section 3.4 presents the model specification. Section 3.5 presents the estimation results and evaluates the model performance with respect to alternative specifications commonly used in practice. The different models are then carried out for stress testing exercises in section 3.6. The related results are compared using analytical financial portfolios. Section 3.7 summarizes and concludes.

3.2 Financial extreme returns modelling

3.2.1 The marginal distributions

In the financial literature, it is commonly admitted that normal-type distributions are unsuited to specify (i.e. capture and/or reproduce) financial assets' properties. However, there is still no agreement for the proposed alternatives. These specifications' performances often vary depending on the assets and the data samples under review. This is particularly true for the specification of extreme returns (i.e. the distribution tails). Student t (Praetz, 1972; Blattberg and Gonedes, 1974) and skewed Student t distributions (Hansen, 1994; Fernández and Steel, 1998; Branco and Dey, 2001; Azzalini and Capitanio, 2003; Jones and Faddy, 2003) are part of the first alternatives and the most used in practice. However, both Student t distributions are *two-fat-tails* specifications, which reduces their ability to specify strongly skewed distributions exhibited by some returns. This leads to an underestimation of the probability of extreme negative returns. To overcome this rigidity, Aas and Hobæk Haff (2006) have proposed a skewed t distribution, with a polynomial (fat) form on one tail and an exponential (semi-fat) form on the other. But yet again, this specification have proven to be unsuited for less asymmetric distributions (Banachewicz and van der Vaart, 2008; Jondeau, 2010).

Generalized hyperbolic distributions (Barndorff-Nielsen, 1977, 1978; Eberlein and Keller, 1995) offer another more flexible alternative to these specification issues. This distribution family is identified by five parameters, namely location, scale, skewness, kurtosis, and tails' forms parameters. It also presents interesting theoretical (closed under conditionality, marginalization and affine transformations, finite moments, etc.) and practical (fitting performance for several financial returns, generalizing most of the usual distributions) properties. Moreover, the estimation issues that have accompanied their introduction have been gradually reduced thanks to the progress made in programming softwares. In some cases, however, identification issues may complicate the distinction between the generalized distribution and some of its specific forms (see Appendix 3.B). This generates specification issues, especially for extreme returns.

Extreme Value Theory (EVT) offers a semi-parametric solution to these specification issues. This is done by focusing on the distribution's tails – of most interest for risk management. According to the Central Limit theorem (CLT), and under some regular conditions, random variables are proven to share common asymptotic behaviours, whatever are their underlying theoretical distributions (see Appendix 3.C). Accordingly, the EVT can be used to specify extreme returns beyond a high quartile called *extreme threshold*. Extreme value distributions (e.g. GPD or GEV) are used for this purpose. These are proven to perform better than most of the parametric distributions presented above. Moreover, focusing on the specification of distribution's tails prevents possible misspecifications of the whole distribution of returns. One possible challenge of this approach is the choice of the extreme threshold. A low quartile alters the asymptotic properties of the EVT, while a too high quartile creates estimation bias due to insufficient observations. Hence, a trade-off accuracy-efficiency is often required. However, the presence of a large sample may allow meeting both criteria. This is precisely why we have adapted this approach to model the marginal distributions of financial returns.

The EVT can only be applied to independent and identically distributed (i.i.d.) random variables. A prior filter of the return series is often necessary. To do so, ARMA-GARCH or stochastic volatility (SV) models are usually used in the literature. By filtering the data, these models also allow specifying the dynamic stylized facts of financial returns (i.e. autocorrela-

tion, heteroskedasticity, and volatility effects). We then specify the remaining distributional stylized facts (skewness and leptokurtosis) by applying the EVT to the filtered residuals of these models. We use a combination of an ARMA-EGARCH model and the EVT (hereafter "EVT-EGARCH") to perform a flexible specification of the individual return series.

3.2.2 The dependence structure

The joint dynamics of the series (hereafter "the multivariate system") is given by their joint distribution function. The latter can be specified in two ways (see appendices 3.D and 3.E). The first, called *direct approach*, consists in generalizing the analysis of univariate distributions presented above to the multivariate framework. A multivariate specification is then considered, and its parameters estimated. Among these parameters, the covariance matrix defines the linear correlation of the multivariate system. The second approach consists in factorizing the joint distribution into: (i) univariate (or marginal) distributions and (ii) a multivariate distribution called *copula* (Sklar, 1959). The marginal distributions specify the individual returns, while the copula specifies the dependence structure of the multivariate system. Therefore, this approach allows separating the shapes of the dependence structure from those of the marginal distributions. It also extends the concept of linear correlation coefficients to that of nonlinear dependence structure function (see Cherubini et al., 2004; McNeil et al., 2005; Jondeau et al., 2007; Patton, 2009).

While different specifications can be found in the literature for bivariate copulas (Joe, 1997; Nelsen, 2006), elliptical and Archimedean copulas are still the most used for the multivariate case (Dißmann et al., 2013). Unfortunately, these multivariate copulas (also called *meta-copulas*) are unsuited to reproduce the empirical dependence structure of extreme financial returns (high and asymmetric dependence). Indeed, elliptical copulas are relatively complex in terms of parameterization, but often inflexible when it comes to specify the dependence structure. Some Archimedean copulas allow specifying the asymmetry of the dependence structure. However, they are often based on few parameters (typically one or two) to specify all pair dependences within the multivariate system (Joe, 1997; McNeil and Nešlehová, 2009; Brechmann and Czado, 2011). Some extensions have been presented, by introducing parameter restrictions on the initial Archimedean copulas (Joe, 1993; Joe and Hu, 1996; Kotz and Nadarajah, 2004; Demarta and McNeil, 2005). But often, the provided extra flexibility is insufficient for oversized multivariate systems on one side, and generates estimation issues on the other (Kurowicka and Joe, 2011).[1]

Joe (1996) has proposed an interesting method to specify multivariate copulas while overcoming technical issues. This consists in factorizing the multivariate copula into conditional and unconditional bivariate copulas, the so-called *pair-copulas*. This technique, further explored by Bedford and Cooke (2001, 2002), offers more flexibility as it specifies separately each of the pair dependences within the multivariate system. It also simplifies the estimation procedure as it is based on bivariate distributions. Kurowicka and Cooke (2006) have presented a graphical representation, named "Regular-vine" or R-vine model, that summarizes all alternative factorization schemes for a multivariate copula (or R-vine copula) (see also Kurowicka and Joe, 2011). Each scheme is given by a sequence of trees starting from the marginal distributions, ending by the R-vine copula, through the pair-copulas (see Figure 3.2). Dißmann et al. (2013) provide a statistical inference of R-vine models.

[1] Note that these issues are also the main challenge facing the implementation of the direct method. This explains the use of more simple specifications in practice.

The choice of the decomposition scheme may become a tricky task in this approach, especially for oversized systems. Two particular schemes have been studied in the literature, namely the canonical-vine (C-vine) and the drawable-vine (D-vine) models. Aas et al. (2009) have proposed a sequential procedure allowing to perform simultaneously a selection and an estimation of the decomposition scheme and the pair-copulas of these models. For some data, the decomposition scheme is chosen among all alternative schemes of a C-vine or a D-vine model. The pair-copulas are selected from a predefined set of bivariate copulas (see Appendix 3.F). Brechmann and Czado (2011) have applied this procedure to the general R-vine model, for a multivariate system of 52 financial variables (including Euro Stoxx 50 securities). However, for oversized systems, this can be computationally intensive (Heinen and Valdesogo, 2009; Mendes et al., 2010; Czado, 2010; Joe et al., 2010; Kurowicka and Joe, 2011). To deal with, Brechmann et al. (2012) have proposed two parsimonious versions of the R-vine model, by restricting the pair-copulas used in the last trees of the model (using only independent and Gaussian copulas), or by simply truncating the last trees in the model. Their results show insignificant differences in terms of performance between the restricted and truncated versions on one side, and the original model on the other. These findings deserve to be confirmed by further studies.

Compared to meta-copulas, recent studies have shown that the extra-flexibility drawn by pair-copulas allows for a better specification of the dependence structure (Aas et al., 2009; Aas and Berg, 2009; Fischer et al., 2009; Min and Czado, 2010, 2011; Czado et al., 2011; Hua and Joe, 2011; Nikoloulopoulos et al., 2011). In this study, we rely on the R-vine model to specify the dependence structure of reasonably sized multivariate systems. Alongside with the EVT-EGARCH models presented above, we present a unified framework for financial returns. The association of the marginal models and the dependence structure model is performed within the "copula-GARCH approach" (Jondeau and Rockinger, 2006). To the best of our knowledge, this study is the first to associate the EVT with the pair-copulas approach. This would allow for more flexible modelling of extreme returns, in both the univariate and the multivariate cases. Knowing the importance of specifying extreme returns for stress testing exercises, we aim this study will reduce the errors observed by the use of simpler models over the recent period.

Before presenting the model (section 3.4) and analysing its performance (section 3.5) and implications for stress testing purposes (section 3.6), we present the data under review and verify the presence of the main stylized facts.

3.3 Data and descriptive statistics

3.3.1 Data

Our data sample consists of three sets of daily time series: two sets of financial series and a set of nonfinancial series. The first class includes three European and two North American equity indices. We consider the CAC40 (France), the DAX30 (Germany), the FTSE100 (UK), the SP500 (USA), and the SPTSX (Canada). To account for the differences in the closure times between the related markets, we consider the closing prices for the European indices, and the opening prices of the North American indices. The series are collected daily from 07/09/1987 to 07/18/2013 (6791 opening days). The second set includes the nominal exchange rates of the six most traded currencies in the Forex spot market (BIS, 2010). The value of six currencies with respect to the U.S. dollar (USD) are considered: the euro (USD/EUR), the Japanese yen (USD/JPY), the British pound (USD/GBP), the Australian dollar (USD/AUD),

the Swiss franc (USD/CHF), and the Canadian dollar (USD/CAD). We consider intermediate rates collected daily at 16:00 GMT on the London Forex market. The series cover the period 01/04/1999-07/18/2013 (3674 days). The third set includes the spot prices of five commodities, representing five distinct categories: the Brent (energy), gold (precious metal), copper (base metal), wheat (agricultural), and corn (agriculture-bioenergy). For each series, we consider daily closing prices – or intermediate prices for commodities traded OTC (i.e. gold and copper). The sample covers the period 07/05/1993-18/07/2013 (5229 days). All data are collected from Datastream. For all datasets, we have considered the maximum amplitude, given the data availability and synchronization issues related to the multivariate analysis. Last-observed-carried-forward (LOCF) method is considered for missing data. Table 3.1 summarizes the considered datasets.

We split each series into two sub-periods: the first one is used for the estimation procedure; the second, for forecasting and out-of-sample assessments. The first sub-period spans from 07/09/1987 to 08/02/2012 (6541 observations) for the first set, from 01/04/1999 to 07/23/2012 (3424 observations) for the second set, and from 7/5/1993 to 08/02/2012 (4979 observations) for the third set. The second sub-period consists of the last 250 observations for each series.

To ensure the stationarity of our series, prices are transformed into log-differences. That is, considering stock returns (1st dataset), changes in exchange rates (2nd dataset), and changes in commodity prices (3rd dataset). To simplify the presentation, we use the term "return" regardless of the nature of the datasets. Table 3.2 presents the results of three stationarity tests. We have reported the statistics of the augmented Dickey-Fuller (ADF), the Phillips-Perron (PP), and the Zivot-Andrews (ZA) tests. The latter assumes, for each series, a presence of one possible structural break over the entire sample period. For all series, and regardless of the test, the null hypothesis of the presence of a unit root is rejected with a confidence level of 99%. The ZA test detects a structural break for all series. For stock returns, one can note the proximity of the dates on which these breaks have been detected. They all occur within a 15-day window, following the stock market shocks of 10/19/1987 and 10/24/1987. Note that, because of their geographical proximity to the epicenter of these shocks, North American indices are affected first, followed by European indices. For exchange rates and commodities, no evident link between the structural breaks can be noted. This may augur a weak dependence among the return series of these two datasets (see below).

3.3.2 Statistical evidence of stylized facts

Univariate analysis

To set the specifications of the univariate EVT-EGARCH models, we have analysed the descriptive statistics of each return series. For each series, Table 3.3 reports the historical extreme values, the first empirical moments, and the results of normality, autocorrelation, and heteroskedasticity tests. All series have a non-zero conditional mean. Their variances vary depending on the dataset. Hence, commodities have the highest standard deviation (a maximum of 0.024 for wheat), followed by equities (between 0011 and 0015) – although the two datasets have comparable extreme values. Exchange rates are less volatile (between 0.06 and 0.09) and have lower extremes compared to the previous two sets. This could be one of the reasons that makes this market fairly attractive to investors.

In general, the series exhibit a negative skewness (except for USD/GBP, USD/AUD and USD/CHF) and a kurtosis excess (except for USD/EUR and Brent). This confirms the stylized

facts that financial returns' distributions are leptokurtic and left-skewed. That is, the frequency of extreme returns is higher than given by a normal distribution. This is particularly true for negative returns. To confirm this result, we have reported the results of two statistical normality tests, namely the Jarque-Bera (JB) and the Anderson-Darling (AD) tests. The first test is based on the values of the 3rd and 4th empirical moments of the return's distributions. The second, relies on the distance between the empirical and the normal distributions, with an overweighting of extreme quartiles. Both tests reject the null hypothesis of normality for all series, with a confidence level of 99%. Moreover, the tests' statistics show heterogeneous features between the empirical distributions on one side, and the normal distribution on the other. The distributions of stock returns are thus the most disconnected from the normal distribution, especially in the tails. To take into account the specificity of each return's distribution, which varies on the dataset and even between the returns of the same dataset, we will introduce more flexibility to specify the different marginal distributions. To do so, we will carry out a semi-parametric approach, in which the distribution tails are specified using the asymptotic EVT, while the centre of the distribution is given by a non-parametric specification.

To set the specification of the ARMA-EGARCH models – required as a prior filter to apply the EVT – we have checked the presence of autocorrelation and heteroskedasticity (i.e. the dynamic stylized facts) in the return series. These two phenomena are also kwon as linear and nonlinear serial dependence, respectively. They allow specifying the conditional mean (ARMA) and the conditional variance (EGARCH) processes of the series. The autocorrelation is assessed using the Ljung-Box (LB) and the Diebold (D) tests. The latter also allows separating autocorrelation and heteroskedasticity features. The statistics of the tests are reported for 25 and 50 lags. Except for USD/EUR and USD/JPY, the LB test rejects the null hypothesis of no autocorrelation, with confidence levels ranging from 95% to 99% for 25 lags, and 99% for 50 lags (except for wheat). Yet, the D test shows that only the USD/GBP is actually autocorrelated. This suggests that the series are rather heteroskedastic than autocorrelated. To check this aspect, we have reported the LB statistics for nonlinear transformations of the returns, namely squared returns and returns' absolute values. The LB test rejects the null hypothesis of no autocorrelation of transformed returns (i.e. of no heteroskedasticity of returns) with a confidence level of 99%.

The presence of heteroskedasticity is known to impact both the tails of the returns' distributions and their dependence structure (Jondeau, 2010). We will use ARMA-EGARCH models to isolate potential autocorrelation and heteroskedasticity effects (section 3.4.1). The filtered residuals of the ARMA-EGARCH models will then be used to estimate the marginal distributions of returns through the EVT (section 3.4.1). These distributions are used as an input in the R-vine model to specify and estimate the dependence structure of each dataset, using the adapted pair-copulas (section 3.4.2). For a better choice of these copulas, we have studied the empirical dependence structure of our three datasets, through a set of non-parametric statistics.

Analysis of the dependence structure

Table 3.4 analyses the empirical dependence structures of the return series. Within each dataset, and for each pair of returns, we have reported the (linear) correlation coefficient ρ and the (nonlinear) rank coefficient τ, as well as the extreme correlation $\rho^{\pm}$ and the tail dependence $\tau^{\pm}$ coefficients. The latter measure the correlation and the rank correlation of extreme positive and negative returns, respectively. Positive (resp. negative) extreme quartiles of 90% and 99% (resp. 10% and 1%) are considered to do so (see Appendix 3.F for more details on these

measures).

The results show that, overall, equities and exchange rates are more correlated than commodities. The correlation is higher for the European equities (from 0.71 to 0.80) compared to North Americans' (0.59) and cross correlations (0.27 to 0.46). The three European indices are also more correlated with the Canadian index than with the U.S. index. In contrast, the North American indices are more correlated with the FTSE100 than to the other two European indices. This may be due to historical, cultural, political, and economic proximity factors. European currencies are the most correlated in the system (between 0.55 and 0.84). However, the Australian dollar is positively correlated to all currencies (between 0.40 and 0.60) except the yen. The latter is also weakly – and even negatively – correlated to the rest of currencies. Geographical proximity and trade links largely explain these results. This is particularly the case between the euro and the Swiss franc (0.84) or the pound (0.67), or between the Canadian and the Australian dollars (0.60). However, trading volumes does not seem to significantly impact the correlations of exchange rates, given that the yen is the third most traded currency in the considered dataset. Finally, with the exception of the pair wheat-corn, which falls under the same category (agriculture), commodities are weakly correlated. This is mainly due to the heterogeneity of these products, especially in terms of supply and demand determinants in the corresponding markets. The analysis of the dependence through the rank coefficients confirms all the previous results.

Table 3.4 also provides a first indication of the dependence structure of extreme returns, through the extreme correlation $\rho^{\pm}$ and the extreme rank $\tau^{\pm}$ coefficients. The results show that, overall, the magnitude of the extreme coefficient follows that of the coefficient calculated on the whole distribution. Thus, highly correlated returns in normal times also exhibit high correlations in more volatile periods (stress or euphoria). Instead, returns with a (very) low correlation coefficient show a low – and even negative – extreme correlation. This is particularly the case of commodity prices. This result is still valid regardless of the considered extreme quartile.

Instead, the asymmetry of the extreme dependence structure is less obvious. Indeed, a comparison of the extreme correlation and tail dependence coefficients for positive and negative returns of the same quartile leads to different results, depending on the considered quartile and the dependence measure. Except for the Brent-gold and the Brent-wheat pairs, the correlation of negative returns below the 10% quartile are thus greater than that of positive returns beyond the 90% quartile. However, such a result is no longer valid when reducing the size of extreme quartiles to 1% and 99%, respectively (i.e. by considering the most extreme returns). In this case, negative returns are more correlated than positive returns for some pairs and less correlated for the remaining pairs. This ordering also varies with the considered dependence measure. The limited number of observations for the (most) extreme returns may justify this results instability.

To get a more precise idea on the extent of the (a-)symmetry of extreme returns' correlations, we have carried out the nonparametric statistical test proposed by Hong et al. (2007). It allows comparing all extreme positive and negative correlation coefficients, respectively, beyond an extreme quartile. We have reported the test's statistics and the corresponding p-values for 10%-90% and 1%-99% extreme quartiles. Regarding the results for the 10%-90% threshold, the test does not allow rejecting, in almost all cases, the null hypothesis of symmetric extreme correlation, with a confidence level of 95%. This may seem at odds with previous works (Longin and Solnik, 1995, 2001; Ang and Bekaert, 2002a,b; Ang and Chen, 2002; Patton, 2004, 2006a,b). However, as shown by the correlation coefficients for the 10%- 90% threshold, this asymmetry

does exist in most cases. Yet, it is often too weak to be significantly captured by the considered statistical test. The variability of extreme correlations according to the considered quartile may also explain this result. Figure 3.1 confirms this fact for stock returns, by reporting the extreme correlation coefficients for all empirical thresholds. To consider this result, we will introduce more flexibility in the specification of pair dependences between the return series. Pair-copulas with different dependence and extreme dependence shapes will be considered as an input in the selection/estimation procedure of the R-vine model (see Table 3.5).

3.4 The model

This section presents the two stages of the sequential model used for the analysis of financial returns. We first present the EVT-EGARCH model allowing for a semi-parametric specification of the marginal distributions. We then specify the dependence structure of the multivariate system, using an R-vine model associating the obtained marginal distributions and adapted pair-copulas.

3.4.1 The marginal model

The application of the EVT requires a prior filter of the return series. ARMA-EGARCH models with a – temporary – normally distributed errors is considered to capture the conditional mean and variance dynamics in the series. The EVT is then applied to the filtered residuals of the models. By doing so, we specify the tails of the marginal distributions. This specification holds for extreme negative and positive returns, beyond predefined negative and positive extreme thresholds. To complete the specification of the whole marginal distributions, we specify the returns falling between the two extreme thresholds using an empirical distribution. This is done by applying a linear interpolation to the corresponding returns. The marginal distribution is then constructed through a semi-parametric approach, associating an extreme value (EV) distribution in the tails and an empirical distribution in the centre.

Before presenting the different equations of the model, we highlight the main motivations of the considered specifications.

- The ARMA model captures the dynamics of the conditional mean in the series, by considering both autoregressive (AR) and moving average (MA) components.

- The GARCH model captures the conditional variance. The financial literature does not establish a strict order of preference between the alternative conditional variance models. For example, the performance of ARCH-GARCH models are more or less comparable to those of stochastic volatility (SV) models. The latter are often used for continuous time modelling relying on (very) high frequency data. The considered Exponential-GARCH (EGARCH) specification (Nelson, 1991) is part of the so-called *asymmetric volatility models*. It allows capturing the nonlinear dynamics in the conditional variance, based on a combination of ARCH and GARCH effects. ARCH effects are further split into an asymmetry (or sign) effect and a leverage (or amplitude) effect of past returns on actual volatilities. This provides an extra-flexibility of the EGARCH specification. The latter presents comparable properties to other asymmetric versions of GARCH models.

- The normal distribution assumed for the residuals in the ARMA-EGARCH models does not affect the quality of the estimation if the "true" distribution is part of the exponential distributions' family (Gourieroux, 1992). In this case, a Pseudo Maximum Likelihood (PML) estimation method is used.
- The generalized Pareto distribution (GPD) is adopted to specify the extreme returns under the EVT. This EV distribution is commonly used in practice. It is also proven to be more suitable if the EVT is performed through the *Peaks over Thresholds* (POT) method, as we shall do here.
- The choice of the two extreme thresholds is hand-picked, in line with the main practice in the literature.
- The use of a non-parametric method to specify the centre of the marginal distribution provides additional flexibility to the model. A linear interpolation is retained for this purpose. The use of a kernel density with a predefined smoothing (or *bandwidth*) parameter is an alternative approach. However, if the choice of the kernel density (e.g. normal, uniform, quadratic, triangular, Epanechnikov, etc.) has a minor impact on the results, the choice of the smoothing parameter is often crucial and can significantly impact the results.

The ARMA-EGARCH model

The process $\{r_{i,t}\}_{t=1}^{T}$ of return $i = 1, \ldots, N$ is defined by an $ARMA(p,q) - EGARCH(r,s)$ model, if

$$\begin{aligned} & r_{i,t} = \mu_i + \sum_{j=1}^{p} \phi_{ij} r_{i,t-j} + \sum_{j=1}^{q} \theta_{ij} \varepsilon_{i,t-j} + z_{i,t}\sqrt{\sigma_{i,t}^2} \\ & \ln(\sigma_{i,t}^2) = \omega_i + \sum_{j=1}^{s} \left(\alpha_{ij} z_{i,t-j} + \gamma_{ij} \left(|z_{i,t-j}| - \mathbb{E}|z_{i,t-j}|\right)\right) + \sum_{j=1}^{r} \beta_{ij} \ln(\sigma_{i,t-j}^2) \\ & z_{i,t} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1) \end{aligned} \tag{3.1}$$

where μ_{ij} is the unconditional mean; ϕ_{ij} and θ_{ij} capture, respectively, the AR and MA dynamics of the conditional mean; β_{ij} is the GARCH dynamics of the conditional variance; α_{ij} and γ_{ij} capture, respectively, the asymmetry and the leverage effects of past returns on volatility, and $\mathcal{N}(\cdot)$ is the standardized normal distribution of the filtered residuals $z_{i,t}$.

Given the filtered residuals of each return series, we specify the corresponding marginal distribution by the non-parametric approach presented below.

The semi-parametric marginal distribution

For an accurate modelling of the marginal distribution's tails, we apply the EVT to the filtered residuals $\hat{z}_{i,t}$ of the ARMA-EGARCH models. To do so, we use the POT method, according to which beyond of a sufficiently high threshold, the distribution of excess returns (i.e. return minus threshold) follows a GPD. The extreme lower and upper thresholds are set such as the proportion of exceedances on each side of the distribution represent 10% of the total observations. The remaining observations, falling between the two thresholds, are specified by a non-parametric function obtained by a linear interpolation. The semi-parametric marginal distribution of return i is thus given as follows

$$F_i(\hat{z}_i) = \begin{cases} \frac{T_{u_i^L}}{T}\left(1 + \xi_i^L \frac{u_i^L - \hat{z}_i}{\beta_i^L}\right) & \text{if} : \hat{z}_i < u_i^L \\ \varphi(\hat{z}_i) & \text{if} : u_i^R < \hat{z}_i < u_i^L \quad , i = 1, \ldots, N \\ 1 - \frac{T_{u_i^R}}{T}\left(1 + \xi_i^R \frac{\hat{z}_i - u_i^R}{\beta_i^R}\right)^{-\frac{1}{\xi_i^R}} & \text{if} : \hat{z}_i > u_i^R \end{cases} \tag{3.2}$$

where u_i^L (resp. u_i^R) is the lower (resp. upper) threshold, $T_{u_i^L}$ (resp. $T_{u_i^R}$) is the number of exceedances on the left-side (resp. right-side) of the distribution, $\varphi(\cdot)$ is the empirical distribution function, ξ_i^L and β_i^L (resp. ξ_i^R and β_i^R) are, respectively, the shape and the scale parameters of the GPD on the lower (resp. upper) tail.

3.4.2 The dependence structure model

Given the marginal distributions obtained by eq. (3.2), we model the dependence structure of the multivariate system of returns using the R-vine model (Bedford and Cooke, 2001, 2002; Kurowicka and Cooke, 2006; Kurowicka and Joe, 2011; Dißmann et al., 2013). This supposes a factorization of the joint distribution of the multivariate system into marginal distributions on one side, and an R-vine copula on the other. The latter specifies the conditional and the unconditional pair and joint dependence structures.

The joint density function $f(r_1, \ldots, r_N)$ of a multivariate system composed by N returns $r_1, \ldots, r_N$ can be factorized on successive conditional densities, as follows

$$f(r_1, \ldots, r_N) = f(r_1) \cdot f(r_2|r_1) \cdot f(r_3|r_1, r_2) \cdots f(r_N|r_1, \cdots, r_{N-1}) \tag{3.3}$$

Each but the first factor in the right-side product may, in turn, be decomposed into a bivariate density function (or pair-copula) and a conditional marginal density, by the formula

$$f(r|v) = c_{r,v_j|v_{-j}}\big(F(r|v_{-j}), F(v_j|v_{-j})\big) \cdot (r|v{-}j) \tag{3.4}$$

where v is a v-dimensionnal vector (with $v = 1, \ldots, N-1$), v_j is a component of vector v, and v_{-j} is the vector v but v_j. $c_{r,v_j|v_{-j}}$ is the density function of the bivariate copula, and $F(\cdot)$ is the marginal conditional distribution function, given by

$$F(r|v) = \frac{\partial C_{r,v_j|v_{-j}}\big(F(r|v_{-j}), F(v_j|v_{-j})\big)}{\partial F(v_j|v_{-j})} \tag{3.5}$$

where $C_{r,v_j|v_{-j}}$ is the distribution function of the bivariate copula.

Depending on how the joint density has been factorized in eq. (3.3), different decomposition schemes can be obtained. Kurowicka and Cooke (2006) have proposed a graphical structure, named Regular-vine or R-vine, representing the alternative schemes for a given multivariate density. The R-vine model is a sequence of trees $T_1, \ldots, T_{N-1}$. Each tree is given by nodes N_i, connected by edges E_i (for $i = 1, \ldots, N-1$). The nodes of the first tree correspond to the unconditional marginal densities. The related edges represent the unconditional bivariate copulas. For $i > 1$, the edges of a given tree become the nodes of the following one (i.e. $N_i = E_{i-1}$). Bedford and Cooke (2001) and Kurowicka and Cooke (2006) have shown that an edge $e \in E_i$ can be identified by only two nodes, namely conditioned nodes $j(e)$,$k(e)$ and a conditioning

set $D(e)$, such as $e = j(e), k(e)|D(e)$. Figure 3.2 shows an example of a decomposition scheme for an R-vine model with a five-dimensional multivariate system.

By associating to each edge $e = j(e), k(e)|D(e)$, the corresponding pair-copula $c_{e=j(e),k(e)|D(e)}$, Kurowicka and Cooke (2006) have shown that the density of the multivariate copula (or the R-vine copula) can be written as follows

$$c\big(F_1(r_1), \ldots, F_N(r_N)\big) = \prod_{i=1}^{N-1} \prod_{e \in E_i} c_{e=j(e),k(e)|D(e)}\big(F_1(r_{j(e)}|r_{D(e)}), \ldots, F_N(r_{k(e)}|r_{D(e)})\big) \quad (3.6)$$

where $r_{D(e)}$ is a given subset of $D(e)$.

To estimate this model with our three datasets, we have adopted the selection/estimation procedure proposed by Aas et al. (2009). This allows defining simultaneously the decomposition scheme and the adapted pair-copulas of the R-vine model (see Appendix 3.F). In this study, we have tested the suitability of different bivariate copulas exhibiting different features in terms of dependence and extreme dependence structures. To do so, twelve bivariate copulas, including elliptical and Archimedean specifications, have been considered. Table 3.5 summarizes these specifications and their statistical properties. Introducing more flexibility through non-parametric copulas will be explored in a future study.

3.5 Empirical results

3.5.1 The EVT-EGARCH models

To identify the optimal lags for the ARMA-EGARCH models, we have estimated all combinations for a maximum lag of $p = q = r = s = 3$. The Akaike (1973) criterion is considered to select the best specification among all combinations. For the selected models, Table 3.6 reports the estimation results and the related diagnostic tests. For all return series, most of the estimated parameters are significant at a 99% confidence level. Overall, the parameter estimates confirm the descriptive statistics presented earlier and, hence, justify the considered ARMA-EGARCH models. All unconditional means are null. The heteroskedasticity is confirmed, with a presence of ARCH effects in all the series, and the presence of GARCH effects in more than two thirds of cases. In section 3.3.2, we have shown that the autocorrelation is insignificant for most series, for 25 and 50 lags. It seems, however, that the autocorrelation is still present for lower lags. Indeed, the parameters of the second and third lags in the AR and MA processes are often significant. The specifications with the highest lags are even the best ones. Even though better specifications can still be found by relaxing the considered constraint of maximum lags in the ARMA processes, we have maintained our choice for two main raisons: a concern of parsimony, and because the considered specifications are proven to be sufficient to filter the return series, which is our main objective here. This aspect is shown by the last three columns of the table. The Ljung-Box (LB) test rejects the null hypotheses of both autocorrelation and heteroskedasticity (autocorrelation of squared returns) in the standardized residuals. The ARCH test (Engle, 1983) rejects the null hypothesis of the presence of ARCH effects. These results are valid with confidence levels ranging from 10% to 100%, and for lag orders from 1 to 50. Due to space limitation, however, only results for 20 lags are reported.

Before estimating the marginal distributions from the filtered residuals, we have analysed the impact of the presence of autocorrelation and heteroskedasticity on the returns' distributions. Specifically, we have compared, for each series, the empirical skewness and kurtosis coefficients in the distribution of filtered residuals (i.e. the unconditional distribution) on one side, with their corresponding values in the distribution of returns (i.e. the conditional distribution) on the other. Three main conclusions can be drawn, based on columns 3 and 4 of Table 3.7. First, the skewness and kurtosis coefficients vary significantly but converge to values zero and three, respectively. This result confirms the stylized fact that the conditional distributions are closer to the normal distributions than the corresponding unconditional distribution (see Appendix 3.A). Second, only equity returns, whose unconditional distributions showed the highest divergence with respect to the normal distribution (see section 3.3.2 and Table 3.3) conserve leptokurtic shapes for the conditional distributions. The other two datasets have now a mesokurtic conditional distribution – and even a platikurtic form for exchange rate returns. Third, the conditional distributions remain for the most asymmetric (in 12/16 cases, compared to 13/16 for unconditional distributions), even if the skewness coefficient is often close to zero for most series but stock returns. These results confirm the elusive character of financial returns' distributions. These may indeed vary between different datasets, different returns of a same dataset, and even for the same return series. This proves the unsuitability of any method assuming a (common) predefined parametric distribution for financial return series. To avoid this mistake, we have specified the return's distributions through the more flexible semi-parametric approach presented in section 3.4.1.

This approach is applied to the filtered residuals of the ARMA-EGARCH models. Columns 4-10 of Table 3.7 report the estimation results of the marginal distributions. The values of the extreme negative (u^L) and positive (u^U) thresholds, as well as those of the scale parameters of the GPD in the lower (β^L) and the upper (β^U) tail are comparable between the three datasets. However, the shape parameter varies substantially. GPD distributions of stock returns have a positive shape parameter on the lower tail (ξ^L), (i.e. a Pareto-type GPD). This value is higher than the corresponding value in the upper tail (ξ^L), which is close to zero or even negative (i.e. a GPD close to exponential or Pareto type II distribution). This result confirms the leptokurtic and skewed shape of all marginal distributions of stock returns. This is confirmed by the values of the scale parameter which has, for all series, a higher value on the lower tail, meaning a wider range of extreme negative returns. In four cases out of six, the exchange rate returns present a negative (resp. positive) value in the shape parameter on the lower (resp. upper) tail. Only in half the cases, this value is greater in the lower tail than its corresponding value in the upper tail. For the marginal distributions in this dataset, one can conclude to a non-fat lower tail and/or an unstable skewness. This remark also holds for commodities, even though these returns have at least one fat tail.

The suitability of the estimated marginal distributions to the stylized facts of the filtered residuals is evaluated through a battery of Goodness-of-Fit (GoF) tests. For each series, the test consists in: (i) obtaining a Probability Integral Transformation (PIT) of the filtered residuals using the corresponding marginal distribution, and (ii) comparing the distribution of the obtained PIT to a uniform distribution. The more the PIT distribution is close to the uniform distribution, the more the marginal distribution used in the transformation is deemed close to the "true" distribution of the filtered residuals – and to that of the corresponding returns. Columns 11-14 report the confidence levels of three versions of the Kolmogorov-Smirnov (KS) GoF test, considering three alternative hypotheses: a PIT distribution which is different from (KS), above (KS +), or bellow (-KS) the uniform distribution. As we have done earlier for

normality tests, we have also considered the Anderson-Darling (AD) test, which is a version of the KS test with an overweighting of the distribution tails. The results reject the alternative hypothesis with confidence levels ranging from 64% to 100%, regardless of the test and the return series. We can therefore conclude to a good performance of the considered semi-parametric specifications to model the financial returns' marginal distributions.

To confirm this result, we have compared the performance of this semi-parametric specification to that of five parametric distributions commonly used in practice. The KS and the AD GoF tests are used for this purpose. As parametric specifications, we have considered the following distributions: normal (Nor.), Student t (St.), skewed Student t (SSt.) of Fernández and Steel (1998), skewed t (GH-SSt) of Aas and Hobæk Haff (2006), and generalized hyperbolic (GH). Columns 15-24 show a clear advantage of the semi-parametric approach over all parameteric specifications, regardless of the considered GoF test and the return series. The normal and the GH-SSt distributions are adequate anywhere. The performance of the Student t distribution, the asymmetric Student t, and the GH distributions vary depending on the dataset. They are relatively well adapted to exchange rates whose empirical distributions are closer to normality. However, their performance drop significantly for stock and commodity returns whose empirical distributions have one or two fait tails.

The results of this subsection show the suitability of the semi-parametric distributions considered to specify the marginal distributions of financial returns. These marginal distributions are next used to model the dependence structure of the multivariate systems of financial returns, using the R-vine model. The following section presents the results of the estimation, as well as those of validation and robustness tests.

3.5.2 The R-vine model

The results of the specification/estimation procedure of the R-vine model are reported on Table 3.8. Figure 3.3 provides a graphical overview of the results. Three R-vine models are estimated for the three datasets. The results confirm the heterogeneity of the dependence structures, depending on the datasets and, within the same dataset, between the return series. Whatever the considered multivariate system, the first tree in the model – capturing the unconditional dependences between the return series – shows that extreme returns exhibit a strong dependence (Student t copulas). The dependence is often symmetric, except for two pairs of commodities (RG copulas). These results fairly confirm those of the non-parametric tests (see Table 3.4). The lower trees of the R-vine model capture the conditional dependences. Overall, these trees are consisting of symmetric copulas.

Except for the exchange rates (Student t copula), the extreme conditional dependence is often null (Gaussian, Frank, and BB8 copulas). In other words, only exchange rates are significantly correlated in turmoil periods (stress and/or euphoria). Stock and commodity returns are asymptotically independent. This result is confirmed by the last tree of each R-vine model, which captures the joint dependence structure of the multivariate system. This dependence is symmetric whatever the model. Overall, the joint dependence structure has: (i) a (high) joint extreme dependence for exchange rates (Student t copula), but (ii) a null joint extreme dependence for the remaining two datasets (Gaussian and Frank copulas). This findings indicate that, while presenting a low volatility in normal periods, a portfolio of foreign currencies becomes, in turmoil episodes, more risky than equity or commodity portfolios yet more volatile by nature.

3.5.3 Validation tests

To check the robustness of our results and the performance of the R-vine model to capture the empirical dependence structure of the dataset under review, we have simulated a sample of 10^6 bootstrap multivariate realizations for the three estimated R-vine models. For each sample, the extreme correlation coefficients have been calculated. These have then been compared to: (i) the empirical correlation coefficients calculated for the standardized residuals, and (ii) the correlation coefficients calculated for a 10^6 realizations resulting from a bootstrap simulation of two meta-copulas models (Student t and Gaussian) estimated for the standardized residuals. For the stock return series, Figure 3.4 shows a comparison of simulated and empirical extreme correlations for quartiles ranging from 1% to 10% (for negative returns) and from 90% to 99% (for positive returns).

Three main results can be drawn. First, the estimates obtained by the R-vine and the Student t models are quite similar, and are clearly distinguished from those of the Gaussian model. This result is due to the fact that, for the data under review, the R-vine model is essentially composed of Student t pair-copulas (see Figure 3.3). Second, the R-vine model seems to be more efficient to capture the empirical dependence of extreme negative returns. Indeed, it often outperforms the Student t model, and always does against the Gaussian model. Third, in most cases, the Gaussian model seems to dominate the other two with respect to the upper extreme quartiles. This discrepancy of the three models on the upper and lower extreme quartiles can be explained as follows. Some pairs of returns have a relative asymmetry in their dependence structure (i.e. the negative extreme returns are more dependent than positive returns of the same size). However, given that this asymmetry is often weak, it is rarely captured by statistical tests (see Table 3.4). Yet, these tests are part of the selection procedure of the pair-copulas used in the R-vine model. Thus, this procedure leads, in most cases, to select symmetric pair-copulas – note that Student t and Gaussian meta-copulas models are by nature symmetric. The use of symmetric pair-copulas may thus lead to an overestimation of the dependence for the upper extreme quartiles (R-vine and Student t models) and/or an underestimation for the lower extreme quartiles (Student t and Gaussian models). The dynamics of the dependence structure may also be a valid explanation of the latter result. In a future study, we will introduce an additional flexibility in the choice of pair-copulas, including non-parametric copulas and a dynamic specification for the dependence structure.

Despite these potential improvements, the considered R-vine model seems to outperform the other alternative copula-based models. Table 3.9 compares the static (in sample) performances of the R-vine model, the Student t and the Gaussian meta-copulas models, as well as the C-vine and the D-vine models – which are two restricted versions of the R-vine model (see Brechmann and Czado, 2011, for a similar analysis). Based on the estimation's log-likelihood and common information criteria, the results show a dominance of the R-vine specification, followed by the other two vine-copulas models. Meta-copulas models are clearly outperformed. To confirm these results, we also carried out the Vuong (1989) and Clarke (2007) tests, designed for non-nested models. They test the null hypothesis that an alternative model is equivalent or superior to a baseline model. Considering the R-vine as the baseline model, we have reported the test's statistics and the corresponding confidence levels. To account for the difference in the number of parameters between the models, we have also reported the corrected AIC-based and BIC-based adjusted versions. Except the C-vine model for stock returns and the Student t model for exchange rates, the Clarke test rejects the null hypothesis with confidence levels superior to 95%. That is, the R-vine outperforms the remaining models, with and without a correction by the number of parameters. The Vuong test only confirms this dominance with

respect to the Gaussian model, as the null hypothesis cannot be rejected for the remaining models.

Moreover, we have compared the dynamic (out-of-sample) performance of the R-vine model with the same models presented above. Due to the result of the previous paragrah, the meta-copula Gaussian model has been replaced by two DCC-GARCH models (Engle, 2002) currently used in practice (hereafter DCC). The comparison is based on common backtesting procedures. Table 3.10 reports the results for the stock returns, based on three groups of analytical portfolios. The first group contains two (long and short) equally weighted portfolios. The second group contains ten portfolios formed by a long (resp. short) position for one asset and a short (resp. long) position for the others. The third group includes twenty portfolios formed by a long (resp. short) position for two assets and a short (resp. long) position for the rest. The reported number of exceedances and the tests' p-values are the average values of each group. The table confirms the poor performance of the DCC-Normal model, regardless of the confidence level. The DCC-Student t and the Student t meta-copula models are fairly conservative above the 99% confidence level. Their use for risk management can then generate substantial opportunity costs related to the overestimated hedging costs of potential losses. Overall, copula-besed models seem more efficient than DCC models. However, the meta-copula model is less performant than the pair-copulas'. This is mainly due to the relatively lack of flexibility of the former, assuming a common dependence structure for all pairs of returns. Moreover, the R-vine model slightly dominates the C-vine model, which is relatively less flexible.

The confirmed performance and the flexlibility of the presented model framework justifies its use for extreme risk management, where stress testing exercises are a growing practice. In this respect, the presented marginal model (section 3.4.1) can be carried out to design univariate stress scenarios, while the multivariate model (section 3.4.2) can be used for multivariate scenarios. The next section provides an example of the latter, highlighting the main advantages of using the presented R-vine model.

3.6 An application to financial stress tests

Unlike historical scenarios, the design of hypothetical stress testing scenarios (also called *model-based scenarios*) is less common in the literature. The few exceptions include Kupiec (1998), Cherubini and Lunga (1999), Berkowitz (2000), Breuer and Krenn (1999), Breuer and Krenn (2001), Kim and Finger (2000), Bee (2001), and Aragonés et al. (2001). All these papers, however, consider one-period or static scenarios. Alexander and Sheedy (2008) have extended this literature to dynamic stress scenarios in a univariate framework. In this section, we further extend this approach to the multivariate framework, using the model framework of section 3.4.

We analyse the impact of hypothetical stress scenarios on analytical linear portfolios. An equally weighted portfolio is considered for each of our three datasets. The portfolio's return is given as a linear function of the individual returns, assumed as risk factors. The specification of the initial shock on the risk factors and the related probability is based on the marginal distributions of the return series. The evolution of the shock over an arbitrary stress horizon (i.e. the after-shock) is given by the EVT-EGARCH and the R-vine models. The result of the test is assessed in terms of losses in the portfolio's return, and in terms of the regulatory capital required to cover these losses. To show the importance of the risk model in the stress testing result, we have also reported the results obtained by alternative models. These are the same as already considered for validation purposes in the previous section.

Three scenarios are set. In each scenario, the initial shock is given by a joint extreme change in all returns of a same dataset. We have considered negative changes given by quartiles 0.05%, 0.02%, and 0.01% in the returns' marginal distributions.[2] These quartiles allow considering extreme risks not captured by the conventional VaR – where the coverage level is usually set at 5% or 1%. For each portfolio, Table 3.11 reports the resulting returns as given by four alternative marginal distributions: the empirical distribution, the normal and the Student t distributions, and the semi-parametric distribution of eq. (3.2).

The results show a significant difference in the size of the shock for the same quartile (or probability). The most severe changes are generated by the empirical distribution, followed by the semi-parametric distribution, then the Student t and the normal distributions. As we have noted earlier in chapter 1, one of the main challenges in stress testing exercises is to choose among all alternative specifications, the distribution that determines the plausibility level of the scenario – and thus the credibility of the stress test. Yet, in Table 3.7, we have shown the poor performance of most parametric specifications to capture the distribution of the return series. Hence, the use of these distributions in a stress testing exercise can generate an overestimation of the scenario's probability. Stress tests based on this kind of distributions may then suffer a lack of credibility. Instead, the semi-parametric specification has proven to be more flexible, less subjective when it comes to the choice of a specific distribution form, and also leads to a better performance for extreme returns. This latter specification seems then more suitable for stress testing exercises.

The model specification impacts the stress testing results through the size of the corresponding initial shock. For dynamic scenarios, this specification also impacts the results through the after-shock recorded over the test horizon. To illustrate this aspect, we have considered the initial shocks as estimated through the semi-parametric marginal distributions. We have then analysed the evolution of returns of our three portfolios, over a test horizon of $H = 20$ daily sub-periods. We assume, during this horizon, that the portfolio cannot be liquidated nor hedged. The initial shock occurs at date $h = 1$ (corresponding to date $T + 1$, where T is the size of the data sample). The post-shock is assessed for the next $H - 1$ periods. For a given period, the R-vine model captures the shock transmission across the different returns of the portfolio, given their extreme dependence structure. Moreover, the ARMA-EGARCH models allow measuring the shock transmission for the same return series across the horizon sub-periods. The combined effect of both models allows estimating the portfolio's return at each sub-period. The procedure is detailed in Appendix 3.G. For each initial shock, we have compared the resulting impact on the portfolio's return as given by three alternative multivariate specifications: the R-vine model, the Student t meta-copula model, and a DCC-Student t model with four degrees of freedom. The results are reported on Table 3.12.

The results show a significant difference across models. By using the Student t meta-copula model, with high extreme dependences, most of the initial shock is transmitted across the multivariate system. This generates significant losses on the portfolio's return as the horizon goes on. However, this loss is fairly overestimated as only few returns actually present a Student t-type pair dependence (see section 3.5.2). Moreover, the use of a DCC-Student t model can be confusing, as it does not allow distinguishing between the form of the dependence structure and that of the underlying marginal distributions (Garcia and Tsafack, 2011). Instead, the use of pair-copulas models allows avoiding these drawbacks, being based on an explicit and a customized modelling of all pair dependences that form the joint dependence structure. Here again, capitalizing on the extra-flexibility and the performance of R-vine models can bring more

[2] For simplicity reasons, the shocks are recorded on the marginal rather than on the joint distributions.

credibility to stress tests, leading to a better use of their results.

Given the impact of the shock and the after-shock on our analytical portfolios, we have measured the required regulatory capital to cover the generated losses. The results are reported on Figure 3.5. The required capital is indicated by a horizontal line. Whatever the considered shock, the estimated losses increase with the test horizon. Overall, this occurs in comparable proportions for the three portfolios. The losses remain below the regulatory capital up to say 15 (resp. 18) days after the initial shock of probability 0.02% (resp. 0.05%). In theory, the portfolios may be hedged or liquidated before these dates, to avoid any increase in the regulatory capital. In practice, however, the ability to perform these actions depends on several factors (e.g. market liquidity, portfolio size, strategic decisions, etc.). To account for these factors, regulators often recommend a 10-day non-transferability assumption.

3.7 Conclusion

In this chapter, we have presented a sequential model to capture the main stylized facts in financial return series. The model is used to perform extreme risk management analysis, namely a stress testing exercise. Flexible specifications of the conditional marginal distributions and the dependence structure of the return series is led by an EVT copula-GARCH model framework. The latter combines the EVT, for the univariate processes, and an R-vine model based on pair-copulas, for the multivariate system. The model is experimented for three datasets, including equity indices, bilateral exchange rates, and commodity prices. The results show fairly good performances of the model in reproducing the main stylized facts in the data and in forecasting extreme events.

Unlike several previous studies, our results suggest a symmetry of the dependence structure of the return series, and in two cases out of three, the extreme dependence is asymptotically null. To test the validity of these results, we have compared the static and dynamic performances of the presented model with those of the DCC and meta-copulas models commonly used in practice. The results showed a better performance of the R-vine model, especially for extreme returns.

The use of the R-vine model with severe stress testing scenarios has revealed an amplitude of losses proportional to the test horizon. These losses exceed the regulatory capital requirements after two weeks of the initial shock. After this period, an increase in capital becomes necessary for illiquid portfolios. The results also show a significant difference of losses stemming from alternative risk models. The choice of a more powerful and flexible specification ensures the credibility of the stress scenario and the usefulness of the related results. We have shown, in this respect, that the R-vine model has intersting properties.

Two main areas of future research could be considered as an extension of this work. First, introducing a dynamic specification in the dependence structure, distinguishing between normal and stressful periods in financial markets. Second, introducing market variables (e.g. volatility and liquidity) as risk factors to allow for more realism in the stress testing scenario and results. This would be performed, for example, by specifying the interaction between the shock and the reaction of market players.

References

Aas, K. and D. Berg (2009): Models for Construction of Multivariate Dependence: a Comparison Study, *The European Journal of Finance*, 5(7–8), 639–59.

Aas, K., C. Czado, A. Frigessi, and H. Bakken (2009): Pair-Copula Constructions of Multiple Dependence, *Insurance: Mathematics and Economics*, 44(2), 182–98.

Aas, K. and I. Hobæk Haff (2006): The Generalized Hyperbolic Skew Student's t-Distribution, *Journal of Financial Econometrics*, 4(2), 275–309.

Akaike, H. (1973): Information Theory and an Extension of the Maximum Likelihood Principle, in B.N. Petrov and F. Csaki (Eds.), Second International Symposium on Information Theory, *Academiai Kiado*, Budapest, 267-81.

Alexander, C. and E. Sheedy (2008): Developing a Stress Testing Framework Based on Market Risk Models, *Journal of Banking and Finance*, 32(10), 2220-36.

Ang, A. and G. Bekaert (2002a): International Asset Allocation with Regime Shifts, *The Review of Financial Studies*, 15(4), 1137–87.

Ang, A. and G. Bekaert (2002b): Regime Switches in Interest Rates, *Journal of Business and Economic Statistics*, 20(2), 163–82.

Ang, A. and J. Chen (2002): Asymmetric Correlations of Equity Portfolios, *Journal of Financial Economics*, 63(3), 443–94.

Aragonés, J.R., C. Blanco, and K. Dowd (2001): Incorporating Stress Tests into Market Risk Modeling. *Derivative Quartely*, 7(3), 44–9.

Azzalini, A. and A. Capitanio (2003): Distributions Generated by Pertubation of Symmetry with Emphasis on a Multivariate Skew t Distribution, *Journal of the Royal Statistical Society: Series B*, 65(2), 367–89.

Banachewicz, K. and A. van der Vaart (2008): Tail Dependence of Skewed Grouped t-Distributions, *Statistics and Probability Letters*, 78(15), 2388–99.

Barndorff-Nielsen, O.E. (1978): Hyperbolic Distributions and Distributions on Hyperbolae, *Scandinavian Journal of Statistics*, 5(3), 151–7.

Barndorff-Nielsen, O.E. (1977): Exponentially Decreasing Distributions for the Logarithm of Particle Size, *Proceedings of the Royal Society of London: Series A*, 353(1674), 401–19.

Bedford, T. and R.M. Cooke (2001): Probability Density Decomposition for Conditionally Dependent Random Variables Modeled by Vines, *Annals of Mathematics and Artificial Intelligence*, 32(1–4), 245–68.

Bedford, T. and R.M. Cooke (2002): Vines – A New Graphical Model for Dependent Random Variables, *The Annals of Statistics*, 30(4), 1031–68.

Bee, M. (2001): Mixture Models for VaR and Stress Testing, *ALEA Tech Reports*, Università di Trento, No. 12.

Berkowitz, J. (2000): A Coherent Framework for Stress-Testing, *Journal of Risk*, 2(2), 5-15.

Blattberg, R.C. and N.J. Gonedes (1974): A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices, *The Journal of Business*, 47(2), 244–80.

Branco, M.D. and D.K. Dey (2001): A General Class of Multivariate Skew-Elliptical Distributions, *Journal of Multivariate Analysis*, 79(1), 99–113.

Brechmann, E.C. and C. Czado (2011): Risk Management with High-Dimensional Vine Copulas: An Analysis of the Euro Stoxx 50, *Center for Mathematical Sciences*, Technische Universität München, Working Paper.

Brechmann, E.C., C. Czado, and K. Aas (2012): Truncated Regular Vines in High Dimensions with Application to Financial Data, *Canadian Journal of Statistics*, 40(1), 68–85.

Breuer T. and G. Krenn (2001): What is a Plausible Stress Scenario? in: L. Kuncheva et al. (Eds.), Computational Intelligence: Methods and Applications, *ICSC Academic Press*, Bangor, 215–21.

Breuer, T. and G. Krenn (1999): Stress Testing: Guidelines on Market Risk, Volume 5, Oesterreichische Nationalbank.

Candelon, B., G. Colletaz, C. Hurlin, and S. Tokpavi (2011): Backtesting Value-at-Risk: A GMM Duration-Based Test, *Journal of Financial Econometrics*, 9(2), 314-43.

Cherubini, U. and G.D. Lunga (1999): Stress Testing Techniques and Value-at-Risk Measures: A Unified Approach, *Decisions in Economics and Finance*, 22(1–2), 77–99.

Cherubini, U., E. Luciano, and W. Vecchiato (2004): Copula Methods in Finance, *Wiley*, Chichester.

Chollete, L., A. Heinen, and A. Valdesogo (2009): Modeling International Financial Returns with a Multivariate Regime-Switching Copula, *Journal of Financial Econometrics*, 7(4), 437–80.

Christoffersen, P.F. and D. Pelletier (2004): Backtesting Value-at-Risk: A Duration-Based Approach, *Journal of Financial Econometrics*, 2(1), 84-108.

Clarke, K. A. (2007): A Simple Distribution-Free Test for Nonnested Model Selection, *Political Analysis*, 15(3), 347-63.

Czado, C. (2010): Pair-Copula Constructions of Multivariate Copulas, in: P. Jaworski, F. Durante, W.K. Härdle, and T. Rychlik (Eds.), Copula Theory and its Applications, Lecture Notes in Statistics 198, *Springer*, 93–109.

Czado, C., U. Schepsmeier, and A. Min (2011): Maximum Likelihood Estimation of Mixed C-vines with Application to Exchange Rates, *Statistical Modelling*, 12(3), 229–55.

Demarta, S. and A.J. McNeil (2005): The t Copula and Related Copulas, *International Statistical Review*, 73(1), 111–29.

Dißmann, J., E.C. Brechmann, C. Czado, and D. Kurowicka (2013): Selecting and Estimating Regular Vine Copulae and Application to Financial Returns, *Computational Statistics and Data Analysis*, 59, 52–69.

Eberlein, E. and U. Keller (1995): Hyberbolic Distributions in Finance, *Bernoulli*, 1(3), 281-99.

Engle, R.F. (1983): Estimates of the Variance of U.S. Inflation Based on the ARCH Model, *Journal of Money Credit and Banking*, 15(3), 286-301.

Engle, R.F. (2002): Dynamic Conditional Correlation - A Simple Class of Multivariate GARCH Models, *Journal of Business and Economic Statistics*, 20(3), 339–50.

Fernández, C. and M.F.J. Steel (1998): On Bayesian Modelling of Fat Tails and Skewness, *Journal of the American Statistical Association*, 93(441), 359–71.

Fischer, M., C. Köck, S. Schlüter, and F. Weigert (2009): An Empirical Analysis of Multivariate Copula Models, *Quantitative Finance*, 9(7), 839–54.

Garcia, R. and G. Tsafack (2011): Dependence Structure and Extreme Comovements in International Equity and Bond Markets, *Journal of Banking and Finance*, 35(8), 1954–70.

Giacomini, E., W. Härdle, and V. Spokoiny (2009): Inhomogeneous Dependence Modeling with Time-Varying Copulae, *Journal of Business and Economic Statistics*, 27(2), 224–34.

Gourieroux, C. (1992): Modèles ARCH et Applications Financières, Collection ENSAE, *Economica*, Paris.

Hafner, C.M. and H. Manner (2012): Dynamic Stochastic Copula Models: Estimation, Inference and Applications, *Journal of Applied Econometrics*, 27(2), 269-95.

Hansen, B.E. (1994): Autoregressive Conditional Density Estimation, *International Economic Review*, 35(3), 705–30.

Heinen, A. and A. Valdesogo (2009): Asymmetric CAPM Dependence for Large Dimensions: The Canonical Vine Autoregressive Model, *Center for Operations Research and Econometrics*, Université Catholique de Louvain, CORE Discussion Papers No. 2009069.

Hong, Y., J. Tu, and G. Zhou (2007): Asymmetries in Stock Returns: Statistical Tests and Economic Evaluation, *The Review of Financial Studies*, 20(5), 1547–81.

Hua, L. and H. Joe (2011): Tail Order and Intermediate Tail Dependence of Multivariate Copulas, *Journal of Multivariate Analysis*, 102(10), 1454–71.

Joe, H. and T. Hu (1996): Multivariate Distributions from Mixtures of Max-Infinitely Divisible Distributions, *Journal of Multivariate Analysis*, 57(2), 240–65.

Joe, H. (1996): Families of m-Variate Distributions with Given Margins and m(m–1)/2 Biariate Dependence Parameters, in: L. Räuschendorf, B. Schweitzer, and M. Taylor (Eds.), Distributions with Fixed Marginals and Related Topics, *Lecture Notes-Monograph Series*, Volume 28, 120-41.

Joe, H. (1997): Multivariate Models and Dependence Concepts, *Chapman and Hall*, New York.

Joe, H. (1993): Parametric Families of Multivariate Distributions with Given Margins, *Journal of Multivariate Analysis*, 46(2), 262–82.

Joe, H., H. Li, and A.K. Nikoloulopoulos (2010): Tail Dependence Functions and Vine Copulas, *Journal of Multivariate Analysis*, 101(1), 252–70.

Jondeau, E. (2010): Asymmetry in Tail Dependence of Equity Portfolios, *FINRISK*, Working Paper No. 658.

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

Jondeau, E. and M. Rockinger (2006): The Copula-GARCH Model of Conditional Dependencies: An International Stock Market Application, *Journal of International Money and Finance*, 25(5), 827–53.

Jones, M.C and M.J. Faddy (2003): A Skew Extension of the t-Distribution, with Applications, *Journal of Royal Statistical Society: Series B*, 65(1), 159–74.

Kim, J. and C.C. Finger (2000): A Stress Test to Incorporate Correlation Breakdown, *RiskMetrics Journal*, 1, 61-75.

Kotz, S., and S. Nadarajah (2004): Multivariate t Distributions and their Applications, *Cambridge University Press*, Cambridge.

Kupiec, P.H. (1998): Stress-Testing in a Value at Risk Framework, *Journal of Derivatives*, 6(1), 7-24.

Kurowicka, D. and H. Joe (2011): Dependence Modeling: Vine Copula Handbook, *World Scientific*, London.

Kurowicka, D. and R. Cooke (2006): Uncertainty Analysis with High Dimensional Dependence Modelling, *Wiley*, Chichester.

Longin, F. and B. Solnik (2001): Extreme Correlation of International Equity Markets, *The Journal of Finance*, 56(2), 649–76.

Longin, F. and B. Solnik (1995): Is the Correlation in International Equity Eeturns Constant: 1960–1990? *Journal of International Money and Finance*, 14(1), 3–26.

Mandelbrot, B. (1963a): New Methods of Statistical Economics, *The Journal of Political Economy*, 71(5), 421–40.

Mandelbrot, B. (1963b): The Variation of Certain Speculative Prices, *Journal of Business*, 36(4), 394–419.

Manner, H. and O. Reznikova (2012): A Survey on Time-Varying Copulas: Specification, Simulations and Estimation, *Econometric Reviews*, 31(6), 654–87.

McNeil, A.J. and J. Nešlehová (2009): Multivariate Archimedean Copulas, d-Monotone Functions and l1–Norm Symmetric Distributions, *The Annals of Statistics*, 37(5B), 3059–97.

McNeil, A.J., R. Frey and P. Embrechts (2005): Quantitative Risk Management: Concepts, Techniques and Tools, *Princeton University Press*, Princeton, New Jersey.

Mendes, B.V.d.M., M.M. Semeraro, and R.P.C. Leal (2010): Pair-Copulas Modeling in Finance, *Financial Markets and Portfolio Management*, 24(2), 193–213.

Mercurio, D. and V. Spokoiny (2004): Statistical Inference for Time-Inhomogeneous Volatility Models, *Annals of Statistics*, 32 (2), 577-602.

Min, A. and C. Czado (2010): Bayesian Inference for Multivariate Copulas Using Pair-Copula Constructions, *Journal of Financial Econometrics*, 8(4), 511–46.

Min, A. and C. Czado (2011): Bayesian Model Selection for D-vine Pair-Copula Constructions, *Canadian Journal of Statistics*, 39(2), 239–58.

Morales-Nápoles, O., R. Cooke, and D. Kurowicka (2010): About the Number of Vines and Regular Vines on n Nodes, *Discrete Applied Mathematics*, forthcoming.

Nelsen, R.B. (2006): An Introduction to Copulas, *Springer*, New York.

Nelson, D.B. (1991): Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica*, 59 (2), 347–70.

Nikoloulopoulos, A.K., H. Joe, and H. Li (2011): Vine Copulas with Asymmetric Tail Dependence and Applications to Financial Return Data, *Computational Statistics and Data Analysis*, 56(11), 3659–73.

Okimoto, T. (2008): New Evidence of Asymmetric Dependence Structures in International Equity Markets, *Journal of Financial and Quantitative Analysis*, 43(3), 787–815.

Patton, A. (2009): Copula-Based Models for Financial Time Series, in: T. Andersen, R. Davies, J. Kreiss and T. Mikosch (Eds.), Handbook of Financial Time Series, *Springer*, New York, 767–85.

Patton, A. (2006a): Estimation of Multivariate Models for Time Series of Possibly Different Lengths? *Journal of Applied Econometrics*, 21(2), 147–73.

Patton, A. (2006b): Modelling Asymmetric Exchange Rate Dependence, *International Economic Review*, 47(2), 527–56.

Patton, A. (2004): On Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation, *Journal of Financial Econometrics*, 2(1), 130–68.

Praetz, P.D. (1972): The Distribution of Share Price Changes, *The Journal of Business*, 45(1), 49–55.

Rodriguez, J. (2007): Measuring Financial Contagion: a Copula Approach, *Journal of Empirical Finance*, 14(3), 401–23.

Stöber, J. and C. Czado (2012): Detecting Regime Switches in the Dependence Structure of High Dimensional Financial Data, mimeo.

Sklar, A. (1959): Fonctions de Répartition à n Dimensions et leurs Marges, *Publications de l'Institut de Statistique de l'Université de Paris*, 229–31.

Vuong, Q. H. (1989): Ratio Tests for Model Selection and Non-nested Hypotheses, *Econometrica*, 57(2), 307-33.

Zhao, X., C.J. Scarrott, L. Oxley, and M. (2011): GARCH Dependence in Extreme Value Models with Bayesian Inference, *Mathematics and Computers in Simulation*, 81(2011), 1430–40.

Appendix

3.A Stylized facts in financial time series

As proven by various empirical studies, financial time series share common individual and joint statistical properties.

(i) Univariate features

- Nonstationarity: the stochastic processes associated with financial assets' prices (i.e. levels) are generally nonstationary in the weak sense (a nonstationarity of order 2), while the assets' return processes (i.e. prices' first differences) are stationary in the weak sense.
- Seasonality: the return series generally exhibit several seasonal phenomena (e.g. Wednesday effect, weekend effect, January effect, etc.).
- The conditional mean of the return series is generally close to zero.
- Autocorrelation and heteroskedasticity: the squared and absolute values of returns generally exhibit high autocorrelation levels. The return series are then said to be "heteroskedastic" or having a "dynamic volatility". The return series exhibit very low autocorrelation levels, although they are not i.i.d. This is justified by the efficient-market hypothesis (EMH), according to which unexpected news may explain price movement between two successive periods.
- Volatility clustering: large variations in returns are generally followed by large variations in the opposite sign. Extreme values are thus concentrated or *clustered*.
- Asymmetry and leverage effects: the asymmetry effect means that the impact of positive returns on volatility are different from that of negative returns of the same absolute value. The leverage effect refers to a negative correlation between past returns and current volatilities. The leverage effect involves the asymmetry effect, but the opposite is not true.
- Leptokurtic distribution (or Kurtosis excess): the distribution of financial generally exhibit fatter tails than the normal distribution. In other words, extreme financial returns are more frequent than by the normal distribution.
- Skewed distribution (or negative Skewness): the distributions of financial returns are generally left-skewed. That is the probability of (extreme) negative returns is higher than that of positive returns of the same magnitude.
- Leptokurtic and asymmetric conditional distribution: the conditional distributions of the filtered returns (i.e. returns adjusted for the effects of autocorrelation, het-

eroskedasticity, and volatility) generally remain asymmetric and leptokurtic, even though at a lesser extent compared to the corresponding unconditional distributions.

(ii) Multivariate features

- Financial returns generally exhibit high correlation levels.
- Extreme returns correlation: extreme positive (resp. negative) values of a given financial return coincide with extreme positive (resp. negative) values of – most – other returns.
- Asymmetry of the dependence structure: for two return series, the correlation of extreme negative returns is often higher than that of extreme positive returns of the same magnitude.
- Nonlinearity and dynamics of the dependence structure: the correlation between financial returns – including their extreme values – is often nonlinear and changes over time.

3.B Generalized hyperbolic distributions (GH)

A GH distribution can be presented as a mixture of normal mean-variance distributions, with a mixing variable following a generalized inverse Gaussian distribution (GIG). A random vector $r \in \mathbb{R}^N$ is a mixture of normal mean-variance, if

$$r \overset{d}{=} \mu + W\gamma + \sqrt{W}AZ \quad (*)$$

where $Z \sim \mathcal{N}_k(0, I_k)$ is an i.i.d. random vector, $W \geq 0$ is an independent random scalar of Z, $A \in \mathbb{R}^{N\times k}$ is a matrix, and μ and γ in $\mathbb{R}^N$ are two vectors of parameters. The 1st and the 2nd moments are respectively given by

$$\mathbb{E}(r) = \mu + \mathbb{E}(W)\gamma$$

and

$$\mathbb{C}\text{ov}(r) = \mathbb{E}(W)\Sigma + \mathbb{V}\text{ar}(W)\gamma\gamma'$$

where $\Sigma = AA'$ and $\mathbb{V}\text{ar}(W) < \infty$.

Any change in the mixing variable W produces a change in the mean and/or variance of r. The nature of the latter is determined by that of W. If W follows a GIG distribution or $W \sim N^-(\lambda, \chi, \psi)$, then r follows a GH distribution i.e. $r \sim GH_N(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$ the joint density of which is given by

$$f(r) = c\frac{K_{\lambda-\frac{N}{2}}\sqrt{\Big(\chi + (r-\mu)'\Sigma^{-1}(r-\mu)\Big)\Big(\psi + \gamma'\Sigma^{-1}\gamma\Big)}\exp(r-\mu)'\Sigma^{-1}\gamma}{\Big(\sqrt{\Big(\chi + (r-\mu)'\Sigma^{-1}(r-\mu)\Big)\Big(\psi + \gamma'\Sigma^{-1}\gamma\Big)}\Big)^{\frac{N}{2}-\lambda}}$$

where K_λ denotes a 3rd-type modified Bessel function of index λ, c is a normalization constant given by

$$c = \frac{(\sqrt{\chi\psi})^{-\lambda}\psi^{\lambda}(\psi + \gamma'\Sigma^{-1}\gamma)^{\frac{N}{2}-\lambda}}{(2\pi)^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}K_{\lambda}(\sqrt{\chi\psi})}$$

and $|\cdot|$ is the matrix determinant operator. Σ is non-singular of rank N. μ is the location parameter, Σ is the scale parameter, γ is the parameter of asymmetry, and χ and ψ are the shape and the tail-shape parameters, respectively. These parameters satisfy: $\chi > 0$, $\psi \geq 0$ if $\lambda < 0$; $\chi > 0$, $\psi > 0$ if $\lambda = 0$; and $\chi \geq 0$ $\psi > 0$ if $\lambda < 0$.

The flexibility of the GH distribution allows deriving several specific distributions. A first family of distributions (more or less complex) are obtained from the limits and by restricting the parameters of the GIG distribution. A second family, containing the most commonly used (simple) distributions in practice, can be obtained from a simplification of equation $(*)$ and/or of the mixing variable W. The main specific forms of the GH distribution are:

(i) Normal mean-variance mixture distributions ($r = \mu + W\gamma + \sqrt{W}AZ$)

- The (multivariate) hyperbolic distribution (HYP): $\lambda = 1$
- The normal inverse Gaussian distribution (NIG): $\lambda = -1/2$
- The variance Gamma distribution (VG): $\lambda > 0$ and $\chi = 0$
- The asymmetric t distribution (ST): $\lambda = -\nu/2$, $\chi = \nu$ and $\psi = 0$

(ii) Normal variance mixture distributions ($r = \mu + \sqrt{W}AZ$)

- The Gaussian (or normal) distribution (N):
- The Student t distribution (T): $W \sim$ Inverse Gamma$(\frac{\nu}{2}, \frac{\nu}{2})$
- The normal mixture distribution (NM):

$$W = \begin{cases} k_1 & \text{with probability } \mathbb{P}\text{r} \\ k_2 & \text{with probability } 1 - \mathbb{P}\text{r} \end{cases}, \quad k_1, k_2 \in \mathbb{R}_+^*, \quad k_1 \neq k_2$$

The first family captures the kurtosis excess and the negative skewness of the return series, unlike the platokurtic and/or the symmetric distributions of the second family. However, if several studies have demonstrated the accuracy and the outperformance of the first distributions for modelling financial time series within a univariate framework, only few studies have been conducted on the multivariate case. This is particularly due to the encountered estimation issues (see section 3.2). The first empirical studies conclude to serious estimation problems beyond three-dimensional systems. It is therefore interesting to check whether the theoretical and the practical complexity of the GH distributions lead to superior performance over more classical distributions.

3.C The Extreme value theory (EVT)

The EVT is a statistical approach allowing to study the extreme values of a random variable (i.e. the tails of the corresponding distributions) within a univarate or a multivariate framework. The EVT can then be considered as an asymptotic specification of the distributions' tails, whose parameters are estimated given a data sample. In practice, two main approaches are used to model the extreme values of a random variable.

(i) Block Maxima (BM)

Assume N i.i.d. realizations of a finite-variance random variable r (e.g. observations of a financial return series). Let $M_n^{(1)}$ be the highest valued observation in a group (or block) consisting of the first n (with $n \ll N$ observations), $M_n^{(2)}$ the highest valued observation in the block of the following n observations, and so on. We can then note

$$M_n^{(i)} = \max(r_1^{(i)}, r_2^{(i)}, \ldots, r_n^{(i)}), \quad i = 1, \ldots, k$$

If k is the number of blocks of size n that can be formed from the data sample, then the set of block maximas (BM) is given by

$$M_n = (M_n^{(1)}, M_n^{(2)}, \ldots, M_n^{(k)}), \quad nk = N$$

According to the Central Limit theorem (CLT), it is possible to find a sequence of real numbers $a_n > 0$ and b_n, such that the distribution of standardized BM $(M_n - b_n)/a_n$ converge to a non-degenerate limit distribution H when the block size increases. The distribution of BM noted F^n is then said to be *in the maximum domain of attraction* of H or $F^n \in MDA(H)$, i.e.

$$\lim_{n\to\infty} \Pr\left\{\frac{M_n - b_n}{a_n} \leq x\right\} = \lim_{n\to\infty} F^n\left\{a_n x + b_n\right\} = H(x)$$

According to the Fisher-Tippett and Gnedenko theorems, if $F \in MDA(H)$ for a non-degenerate distribution H, then H is a Generalized Extreme Value (GEV) distribution. Moreover, this distribution is the only possible non-degenerate limit distribution for the standardized BM. The distribution function of the standardized GEV distribution is given by

$$H_\xi(x) = \begin{cases} \exp\left(-(1+\xi r)^{-\frac{1}{\xi}}\right), & \xi \neq 0 \\ \exp\left(-\exp(-x)\right), & \xi = 0 \end{cases}$$

with $1 + \xi x > 0$ and $x = (r - \mu)/\sigma$, for $\xi, \mu \in \mathbb{R}$, and $\sigma > 0$. μ, σ, and ξ are the location, scale, and shape parameters, respectively. This distribution is generelized in that it contains other forms, such as the Fréchet ($\xi > 0$), the Gumbel ($\xi = 0$), or the Weibull ($\xi < 0$) distributions. Given its leptokurtic form, the Fréchet distribution is particularly desired in the analysis of financial time series.

(ii) Peaks Over Threshold (POT)

For practical considerations, this method is more modern and useful than the previous, as it makes a more efficient use of the limited data available on extreme returns. It allows modelling the tails of the distribution from the observations beyond an extreme quartile or *threshold*. The POT approach can be conducted in two ways: a semi-parametric modelling with respect to the Hill estimator, for example, or a parametric modelling, typically based on the Generalized Pareto Distribution (GPD). Both methods are theoretically justified. The second, however, is more intuitive and easy to implement.

For a random variable r with a distribution function F, the distribution function of the excess returns over an extreme threshold u is given by

$$F_u(r) = \Pr\left\{\mathrm{r} - u \leq r | \mathrm{r} > r\right\} = \frac{F(r+u) - F(u)}{1 - F(u)}$$

for $0 \leq r < r_0 - u$, where $r_0 \leq \infty$ is the upper boundary (also called *the right endpoint*) of F. According to the Pickands-Balkema-de Hann theorem, one can find, for a large class of distributions, a function $\beta(u)$, such that

$$\lim_{u \to r_0} \sup_{0 \leq r < r_0 - u} |F_u(r) - G_{\xi,\beta(u)}(r)| = 0 \Leftrightarrow F \in MDA(H), \quad r \in \mathbb{R}$$

Hence, distributions for which the BM converge to a GEV distribution also exhibit excess values converging to a GPD. The distribution function of a standardized GPD is of the form

$$G_{\xi,\beta}(r) = \begin{cases} 1 - (1 + \frac{\xi r}{\beta})^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp(\frac{-r}{\beta}), & \xi = 0 \end{cases}$$

where $\beta > 0$ and $r \geq 0$ if $\xi \geq 0$, and $0 \leq r \leq -\beta/\xi$ if $\xi < 0$. ξ and β are the shape and the scale parameters, respectively. As for the GEV, the GPD distribution inclues several specifications, such as the Pareto distribution ($\xi > 0$), the exponential distribution ($\xi = 0$), and the Pareto type II distribution ($\xi < 0$). The former exhibits fat tails.

(iii) Block Maxima (BM) with heteroskedastic parameters

Zhao et al. (2011) have recently introduced a third approach to implement the EVT. Based on the Bayesian inference, this method does not require a prior filtering for the return series. The conditional variance is, in this case, captured by applying an ARMA-GARCH-type model directly to the parameters of the GEV distribution. The GEV may therefore be given by

$$\exp\left\{-\left[1 + \xi \frac{r_t - \mu_t}{\sigma_t}\right]^{-\frac{1}{\xi}}\right\}, \quad \sigma_t > 0, \quad 1 + \xi \frac{r_t - \mu_t}{\sigma_t} > 0$$

3.D The direct multivarite modelling approach

This approach is based on a generalization of ARMA-GARCH models to a multivariate framework. The dependence structure (assumed as linear) is derived from the covariance of the residuals associated with the individual processes. The latter are often assumed to follow a simple and homogeneous distribution for all the variables that make the multivariate system. The VAR model (Sims, 1980) is the baseline framework of this approach, with a linear dynamics of the conditional mean. Nonlinearity in the conditional mean is generally introduced by threshold (e.g. TAR, STAR) or Markov switching (MS) models. Tsay (1998) has proposed a vector threshold autoregressive model (VTAR), applied to the interest rates by Hansen and Seo (2002), Sander and Kleimeier (2004), and Rapach and Wohar (2006); to stock returns by Griffin, Nardari, and Stulz (2007); to exchange rates by Gouveia and Rodrigues (2004), Volckart and Wolf (2006), Kapetanios and Shin (2006), and Juvenal and Taylor (2008). Weise (1999), van Dijk (2001), Van dijk et al. (2002), and Camacho (2004) have introduced the vector smooth transition autoregressive model (VSTAR). Lekkos and Milas (2004), Osborn et al. (2005), Huang et al. (2005, 2008), Lekkos et al. (2007), Peng et al. (2011), and Auerbach and Gorodnichenko (2011) are the main applications in the financial literature. Lundbergh et al. (2003) have extended the VSTAR model to explicitly specify an AR time-varying dynamics in

the transition threshold (TV-VSTAR). Dueker et al. (2011) have proposed a smooth transition model, in which the transition is defined by the expected probabilities of the transition function.

To allow for a greater flexibility in the regime changes, Sola and Driffill (1994), Krolzig (1997), and Harris (1999) have adapted the Markov switching (MS) approach introduced by Hamilton (1989) to the multivariate framework – leading to MSVAR models. In these models, the transition process is defined by a state variable following a Markov chain, with fixed transition probabilities (FTP). Ang and Bekaert (2002a,b), Ang and Chen (2002), Patton (2006a, 2006b), and Guidolin and Timmermann (2006a,b, 2008) are the main applications. Filardo and Gordon (1998), then Kim (2008) and Henry (2009) in particular, have extended the Markov-switching approach to allow for a time-varying transition probabilities (TVTP). Yet, the main challenge of MSVAR models relies on the presence of nuisance parameters, preventing the use of specification tests (nonlinearity tests, in particular). Robustness tests are then usually conducted as an alternative.

By generalizing the ARCH-GARCH models, Bollerslev, Engle and Wooldridge (1988) have presented a first multivariate conditional variance model, the VEC model. However, the latter presents several practical challenges, mainly related to the large number of parameters and the difficulty of obtaining a positive definite covariance matrix. To remedy to this, some restrictions have been made to the VEC model, including the diagonal (D-)VEC (Bollerslev, Engle and Wooldridge, 1988; Bollerslev, Engle and Nelson, 1994), the BEKK (Engle and Kroner, 1995) and the D-BEKK (Engle and Kroner, 1995; Kroner and Ng, 1998) models. The initial issues are reduced but not completely resolved in all three versions. Bollerslev (1990) has proposed an alternative modelling approach, based on a decomposition of the conditional covariance matrix into a constant conditional correlation (CCC) matrix and a (diagonal) conditional standard deviations matrix. This specification ensures the positive definiteness of the covariance matrix, reduces the number of parameters, and provides, through a two-step estimation approach, the possibility to study multivariate systems of larger dimensions. Most of the subsequent models have therefore adopted this approach, and further extensions have been introduced. This is particularly the case of constant conditional correlations models, such as the Extended (E-)CCC (He and Teräsvirta, 2004) and the Unrestricted (U-)ECCC (Conrad and Karanasos, 2010) models; linear dynamic conditional correlations, such as the varying correlation (VC) (Tse and Tsui, 2002) and the dynamic conditional correlation (DCC) (Engle, 2002) models; and nonlinear dynamic conditional correlation models, such as smooth transition conditional correlation (STCC) (Silvennoinen and Teräsvirta, 2005), time-varying conditional correlation (TVCC), time-varying smooth transition conditional correlation (TVSTCC) (Berben and Jansen, 2005), dynamic smooth transition conditional correlation (DSTCC) (Silvennoinen and Teräsvirta, 2007), and regime-switching dynamic correlation (RSDC) (Pelletier, 2006). See Bauwens et al. (2006) and Silvennoinen and Teräsvirta (2008) for a detailed review.

Despite its reputation, the direct approach of modelling the conditional mean-variance dynamics admits several analytical and practical shortcomings. On a theoretical level, this method assumes linear the dependence structure (captured by a matrix of linear correlations). It also confuses the form of the dependence structure function with that of the underlying marginal distributions. Moreover, estimation issues usually appear for high dimensional multivariate systems, and often restrict the range of the used specifications for the marginal and the joint distributions. The latter are often reduced to the normal or a Student t distributions.

3.E Copulas

The theorem of copulas (or copula functions) is a statistical concept allowing to associate the marginal probability density functions (pdf) of a multivariate system and the pdf of the related dependence structure to build up the joint pdf of the system. That is, the concept allows separating, for of a multivariate distribution: (i) the dependence structure, captured by the copula function, and (ii) the individual features, described by the marginal distributions.

The copula C is a multivariate cumulative distribution function (cdf) defined on the N-dimensional unit hypercube $[0,1]^N$ in the unit interval

$$C : [0,1]^N \mapsto [0,1]$$

If C is a continuous function and N times partially differentiable, the corresponding density function $c : [0,1]^N \mapsto [0,1]$ is given by

$$c(u_1, u_2, \ldots, u_N) = \frac{\partial^N C(u_1, u_2, \ldots, u_N)}{\partial u_1 \cdots \partial u_N}$$

(i) The Sklar theorem

Let $F : \bar{\mathbb{R}}^N \mapsto [0,1]$ with $\bar{\mathbb{R}} = \mathbb{R} \cup [-\infty, \infty]$ be an N-dimensional joint cdf with univariate marginal cdfs $F_1, \ldots, F_N : \mathbb{R} \mapsto [0,1]$. The multivariate cdf is given by

$$F(r_1, \ldots, r_N) = \mathbb{P}\mathrm{r}(\mathrm{r}_1 \leq r_1, \ldots, \mathrm{r}_N \leq r_N)$$

The theorem of Sklar (1959) shows that there exist an N-dimensional copula $C : [0,1]^N \mapsto [0,1]$ such that for each $r_1, \ldots, r_N$ in $\bar{\mathbb{R}}^N$

$$F(r_1, \ldots, r_N) = C\left(F_1(r_1), \ldots, F_N(r_N)\right)$$

If $F_1, \ldots, F_N$ are continuous, then C is unique and given by

$$C(u_1, \ldots, u_N) = F(F^{-1}(u_1), \ldots, F^{-1}(u_N))$$

where $u_i = F_i(r_i), \quad i = 1, \ldots, N$. Conversely, if C is a copula and $F_1, \ldots, F_N$ are univariate distributions, then F is a joint distribution whose marginal distributions are $F_1, \ldots, F_N$. If F is N-times differentiable, then the joint pdf is deductible from the marginal distributions and the copula density, as follows

$$\begin{aligned} f(r_1, \ldots, r_N) &= \frac{\partial^N F(r_1, \ldots, r_N)}{\partial r_1 \cdots \partial r_N} \\ &= \frac{\partial^N C(F_1(r_1), \ldots, F_N(r_N))}{\partial u_1 \cdots \partial u_N} \\ &= c(F_1(r_1), \ldots, F_N(r_N)) \prod_{i=1}^{N} f_i(r_i) \end{aligned}$$

And the density of the corresponding copula is

$$c(u_1,\ldots,u_N)=\frac{f(F_1^{-1}(u_1),\ldots,F_N^{-1}(u_N))}{\prod_{i=1}^{N}f_i(F_i(u_i))}$$

where $f_i(i=1,\ldots,N)$ denote the marginal densities. Thus

$$f(r_1|r_2,\ldots,r_N)=\frac{f(r_1,\ldots,r_N)}{f_1(r_1)}=c(F_1(r_1),\ldots,F_N(r_N))\prod_{i=2}^{N}f_i(r_i)$$

(ii) Pair-copulas

The concept of pair-copulas is to specify the joint density function f only on the basis of the marginal distributions as well as conditional and unconditional bivariate copulas (or pair-copulas). The factorization of the joint density function is performed as follows

$$f(r_1,r_2,\ldots,r_N)=f(r_1)\cdot f(r_2|r_1)\cdot f(r3|r_1,r_2)\cdots f(r_N|r_1,\ldots,r_{N-1})(*)$$

Each factor of the left-side product of equation (*) is further factorized. For a D-vine model, for example, one possible decomposition scheme would be

$$\begin{aligned}f(r_1,r_2,\ldots,r_N) &= f(r_1)\cdot\prod_{i=2}^{N}f(r_i|r_1,\ldots,r_{i-1})\\ &= \prod_{i=1}^{N}f(r_i)\cdot\prod_{i=2}^{N}\prod_{j=1}^{i-1}c_{ij|(j+1),\ldots,(i-1)}\left(\mathcal{F}_i,\mathcal{F}_j\right)\end{aligned}$$

where $\mathcal{F}_i=F(r_i|r_{j+1},\ldots,r_{i-1})$ and $\mathcal{F}_j=F(r_j|r_{j+1},\ldots,r_{i-1})$.

(iii) Dynamic copulas

To capture the dynamics of the dependence structure represented by the multivariate copula, Patton (2006a,b) have extended the Copula-GARCH model, by introducing a dynamics (deterministic and AR-type) in the parameters of a bivariate Joe-Archimedean copula. Rodriguez (2007) has proposed indexing the parameters of a bivariate copula on the estimated parameters of a two-state Markov switching (MS-) SWARCH model. Okimoto (2008) has first estimated a MS-VAR model, then modelled the dynamics of the residuals using different bivariate copulas. Based on a GJR-GARCH model, Jondeau and Rockinger (2006) have presented a generalization of the previous approaches by applying: (i) a two-state MS model to the parameters of a four-dimension (Gaussian and Student t) copulas, and (ii) an AR model to the parameters of the marginal distributions. Garcia and Tsafack (2011) have associated a MGARCH model to a four-dimension copula, formed by a mixture of bivariate copulas. The weight of each copula in this mixture follows a two-state MS model. In Hafner and Manner (2012), the copula's parameters are defined by a latent variable. Finally, Giacomini et al. (2009) have adapted the Local Change Point (LCP) method, developed by Mercurio and Spokoiny (2004), to estimate the dependence structure over several sub-periods of the sample. See Manner and Reznikova (2012) for a detailed review.

Given the estimation issues related to multivariate copulas, any of the previous models exceeds four dimensions. The introduction of pair-copulas on the dynamic analysis has

contributed to relax this constraint. Chollete et al. (2009) have adopted a two-state MS model to capture the dynamics of pair-copulas within a C-vine framework. Their approach differs from the previous models as it allows for a direct change in the copula functions (from Gaussian to Student t copulas or vice versa) across regimes, and not a simple change in the parameters of a fixed copula. A similar approach is adopted by Stöber and Czado (2012).

3.F The dependence structure

(i) Measures of dependence and extreme dependence

The dependence structure between two random variables (rvs) is captured by a copula function. The dependence is often estimated using a scalar or a coefficient. The concepts of dependence and correlation are often assimilated in practice. Hence, the correlation coefficient (or a correlation matrix in a multivariate framework) is used to capture the dependence of the variables. However, this association is only valid if the latter follow an elliptical distribution. In this case, their joint distribution is defined by a mean vector, a covariance matrix, and a characteristic function. As means and variances are defined by the marginal distributions, the copula function only depend on the correlation matrix and the characteristic function. Only in this case, the correlation matrix matters in the parameterization of the copula, and hence captures the dependence structure. For any other model, using correlations to refer to the dependence structure would be wrong, as the marginal distributions and the correlation matrix of a multivariate system does not allow defining the corresponding joint distribution. Similarly, the joint distribution function, cannot be built on only by using the correlation matrix and the marginal distributions. In addition to these basic aspects, the correlation concept has several limitations.

The correlation coefficient ρ measures the linear dependence between two rvs, taking values in the interval $[-1, 1]$. When $\rho = 0$, the two variables are uncorrelated (or linearly independent). This does not mean, however, that they are necessarily independent. Moreover, when $|\rho| = 1$, the variables are perfectly correlated, but only linearly dependent. Also, the correlation is invariant by strictly increasing linear transformations, but not necessarily invariant for nonlinear transformations. Finally, the correlation is only defined for finite-variance rvs, and not if the underlying distributions are fat-tailed.

Appropriate measures based on copulas (copula-based coefficients) allow capturing the dependence structure between two rvs in a more general framework. These measures enter in the parameterization of the copula. They depend only on the copula and never on the marginal distributions, unlike correlation coefficients which depend on both. The dependence measures are often called *rank measures*, *rank correlation measures*, or *measures of concordance*. They measure the probability that the value of a random variable increases or decreases when that of the other variable moves in the same direction. They then focus on the rank (or the order) of the random variables rather than on their values. These measures are therefore more suited to the concept of copulas, compared to linear correlations. As they are derivable from the copula function, they can be estimated by fitting the copula to a given data sample. These measures are also defined in the interval $[-1, 1]$. They are invariant under strictly increasing transformations. Three main dependence measures are commonly used in practice: the Kendall's τ, the Spearman's ρ, and the Blomqvist's β coefficients. Their specifications are given by a copula function as

follows

$$\begin{aligned}\tau &= 4\int_0^1\int_0^1 C(u,v)\mathrm{d}C(u,v) - 1\\ \rho &= 12\int_0^1\int_0^1 (C(u,v) - uv)\mathrm{d}u\mathrm{d}v\\ \beta &= 4C\left(\frac{1}{2},\frac{1}{2}\right)\end{aligned}$$

where C is the distribution function of the bivariate copula, and u and v are the underlying marginal distributions.

There is no significant differences between these three measures. Although less used in practice, the Blomqvist's β coefficient is fairly straightforward. Unlike the correlation coefficient, these coefficients are insensitive to affine transformations of the random variables and the corresponding distributions. In this study, these measures are used in particular in: (i) the selection procedure of the R-vine model through the iterative selection/estimation method (see below), and (ii) the analysis of the dependence structure of multivariate systems, after the model's estimation.

The presented coefficients refer to all values (i.e. the whole distribution) of the random variables. To measure the particular dependence of extreme values, two main coefficients are used in practice: (i) the *Extreme Correlation* (EC) coefficient, which applies the Pearson ρ coefficient for extreme values, and (ii) the *Tail Dependence coefficient* (TDC) – also called *Quartile Dependence Coefficient* (QDC) – relying on the Kendall's τ coefficient. Nonparametric estimates exist for both types of measures. A parametric version of the TDC can also be obtained from a bivariate copula. Thus, the two TDC of a bivariate copula relate to its lower (i.e. extreme negative values) and upper (extreme positive values) tails. The two values are respectively given by

$$\begin{cases}\tau^L = \lim\limits_{\alpha\to 0}\frac{C(\alpha,\alpha)}{\alpha}\\ \tau^U = \lim\limits_{\alpha\to 0}\frac{2\alpha-1+C(1-\alpha,1-\alpha)}{\alpha}\end{cases}$$

If $\tau^L \in (0,1]$ (resp. $\tau^U \in (0,1]$), the considered copula is said et exhibit a negative (resp. positive) extreme dependence. If $\tau^L = 0$ (resp. $\tau^U = 0$), it exhibits independent extreme negative (resp. positive) values. If $\tau^L \neq \tau^U$, it has an asymmetric extreme dependence structure. For the R-vine model, Joe et al. (2010) have shown that a multivariate asymmetry does exist if and only if the pair-copulas of the first tree (i.e. the unconditional copulas) are asymmetric.

(ii) The sequential selection/estimation procedure of the R-vine model

Unlike the direct method (see Appendix 3.D) or the indirect method based on meta-copulas (see Appendix 3.E), where *only* a specification of the multivariate distribution is required to estimate the joint density, the estimation of pair-copula-based models requires the specification of all pair-copulas copula as well as that of the decomposition scheme.

For an N-dimensional density, the R-vine model has $\binom{N}{2} \times (N-2)! \times 2\binom{N-2}{2}$ possible decomposition schemes (Morales-Nápoles et al., 2010). Aas et al. (2009) have proposed an iterative selection method, based on a representation of the highest pair dependencies

on the first trees of the model. This allows for an explicit and a precise modelling of these dependences – which are the most relevant to analyse. Each spanning tree is formed by a set of nodes connected by edges measuring the bivariate dependence of each pair of nodes. Taking into account the set of nodes given by the previous tree, the selection procedure consists in retaining, in each stage, the tree that maximizes the sum of absolute dependencies. That is

$$\max \sum_{e=\{i,j\}} |\delta_{ij}|$$

where δ_{ij} is a measure of dependence, and $1 \leq i < j \leq N$.

At each stage, after selecting the corresponding tree, the pair-copulas are introduced to estimate (conditional) bivariate densities. Among a set of predefined bivariate copulas (see Table 3.5), those minimizing the AIC information criterion are selected. These copula functions form the nodes of the next step, according to which the corresponding spanning tree is selected, and so on. This iterative procedure (spanning tree – pair-copulas) ensures a better flexibility in the specification of the R-vine model and facilitates the related estimation.

3.G The R-vine model for risk management

Let a portfolio of M assets, with a vector of weights $w' = (w_1, \ldots, w_i, \ldots, w_N)$, with: $\sum_{i=1}^{N} w_i = 1$ and $w_i \in \mathbb{R}$. To estimate the VaR of the portfolio, we consider a moving window of size $T = 10^3$ (i.e. four civil years of daily return series) and a simulation sample size of $N = 10^6$. Based on the EVT-EGARCH and the R-vine models, the one-period ahead forecasting of the VaR is carried through the following steps (Brechmann and Czado, 2011):

1. Assuming a normal distribution of residuals, we specify and estimate the parameters of the following $ARMA(p,q) - EGARCH(r,s)$ models

$$r_{i,t} = \mu_i + \sum_{j=1}^{p} \phi_{ij} r_{i,t-j} + \sum_{j=1}^{q} \theta_{ij} \varepsilon_{i,t-j} + z_{i,t}\sqrt{\sigma_{i,t}^2}$$

$$\ln \sigma_{i,t}^2 = \omega_i + \sum_{j=1}^{s} \left(\alpha_{ij} z_{i,t-j} + \gamma_{ij} \left(|z_{i,t-j}| - \mathbb{E}|z_{i,t-j}|\right)\right) + \sum_{j=1}^{r} \beta_{ij} \ln \sigma_{i,t-j}^2$$

$$z_{i,t} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$$

The standardized residuals of the (marginal) models are obtained by:

$$\hat{z}_{i,t} = \frac{1}{\sigma_{i,t}} \left(r_{i,t} - \hat{\mu}_i - \sum_{j=1}^{p} \hat{\phi}_{ij} r_{i,t-j} - \sum_{j=1}^{p} \hat{\theta}_{ij} \hat{\sigma}_{i,t-j} \hat{z}_{i,t-j} \right)$$

2. Given these estimates, we measure, for each return series, the expected conditional variance for date $T+1$, as follows

$$\begin{aligned}\ln \hat{\sigma}^2_{i,T+1} &= \hat{\omega}_i + \sum_{j=1}^{r} \left(\hat{\alpha}_{ij} \hat{z}_{i,T+1-j} + \hat{\gamma}_{ij} \left(|\hat{z}_{i,T+1-j}| - \mathbb{E}|\hat{z}_{i,T+1-j}| \right) \right) \\ &+ \sum_{j=1}^{s} \hat{\beta}_{ij} \ln \hat{\sigma}^2_{i,T+1-j}\end{aligned}$$

3. Given the semi-parametric marginal distributions, fitted on the standardized residuals, we estimate the R-vine model.

4. For each $n = 1, \ldots, N$

   (i) Given the estimated R-vine model and the inverse of the marginal semi-parametric distributions, we simulate a sample of standardized residuals for date $T+1$: $(\hat{z}_{1,T+1}, \ldots, \hat{z}_{M,T+1})'$

   (ii) The samples, the parameter estimates and the conditional variances' forecasts are used to forecast the return $i = 1, \ldots, M$ at date $T+1$, as follows

$$\hat{r}_{i,T+1} = \hat{\mu}_i - \sum_{j=1}^{p} \hat{\phi}_{ij} r_{i,T+1-j} + \hat{\sigma}_{i,T+1} \hat{z}_{i,T+1} + \sum_{j=1}^{q} \hat{\theta}_{ij} \hat{\sigma}_{i,T+1-j} \hat{z}_{i,T+1-j}$$

   (iii) Based on the expected individual returns, the expected return of the portfolio is obtained by

$$\hat{r}_{T+1} = \sum_{i=1}^{M} w_i \hat{r}_{i,T+1}$$

5. The $VaR(\alpha)$ of the portfolio can then be estimated by considering the α-quartile of the expected distribution of the portfolio's returns $\hat{r}_{T+1}$ obtained by all simulations $m = 1, \ldots, M$.

6. The procedure is repeated daily. The obtained expected returns of the portfolio are compared to actual returns for backtesting purposes.

For the stress testing exercise, an ad hoc initial shock (a change in the standardized residuals) is introduced in step 4i. In this study, the shock is only introduced in the first period. The approach could, however, be generalized to introduce additional post-shocks. By introducing the stress shock, a sequence of stressed returns $(z^*_{1,T+1}, \ldots, z^*_{N,T+1})'$ replaces the normal sequence $(\hat{z}_{1,T+1}, \ldots, \hat{z}_{N,T+1})'$. The subsequent steps of the procedure remain identical for steps 4ii to 6 for the first period. The whole VaR procedure is repeated for the following periods, allowing to measure the impact of the initial shock on the portfolio $\hat{r}_{T+h}$ over the stress horizon $h = 1, \ldots, H$.

Data class	Series	Datastream Mnemonic	Market	Period	Obs.
Main domestic equity index	CAC40	FRCAC40	Paris	from 07/09/1987 to 07/18/2013	6791
	DAX30	DAXINDX	Frankfort	from 07/09/1987 to 07/18/2013	6791
	FTSE100	FTSE100	London	from 07/09/1987 to 07/18/2013	6791
	SP500	S&PCOMP	New York	from 07/09/1987 to 07/18/2013	6791
	SPTSX	TTOCOMP	Toronto	from 07/09/1987 to 07/18/2013	6791
Bilateral nominal exchange rate	USD/EUR	XUDLERD	London	from 01/04/1999 to 07/18/2013	3674
	USD/JPY	XUDLJYD	London	from 01/04/1999 to 07/18/2013	3674
	USD/GBP	XUDLGBD	London	from 01/04/1999 to 07/18/2013	3674
	USD/CAD	XUDLCDD	London	from 01/04/1999 to 07/18/2013	3674
	USD/CHF	XUDLSFD	London	from 01/04/1999 to 07/18/2013	3674
	USD/AUD	XUDLADD	London	from 01/04/1999 to 07/18/2013	3674
Commodity price	Brent	OILBREN	London (IPE)	from 07/05/1993 to 07/18/2013	5229
	Gold	GOLDBLN	London (LBM)	from 07/05/1993 to 07/18/2013	5229
	Copper	LCPCASH	London (LME)	from 07/05/1993 to 07/18/2013	5229
	Wheat	WHEATSF	Chicago (CBOT)	from 07/05/1993 to 07/18/2013	5229
	Corn	CORNUS2	Chicago (CBOT)	from 07/05/1993 to 07/18/2013	5229

Table 3.1: Data description

IPE: International Petroleum Exchange
LBM: London Bullion Market
LME: London Metal Exchange
CBOT: Chicago Board of Trade

				ZA		
	Obs.	ADF	PP	Statistic	Break point	Date
CAC40	6790	-18.75**	-6368.86**	-82.91**	89	11/11/1987
DAX30	6790	-18.78**	-6593.73**	-83.30**	88	11/10/1987
FTSE100	6790	-19.80**	-6455.04**	-82.83**	87	11/09/1987
SP500	6790	-19.39**	-6370.34**	-84.44**	78	10/27/1987
SPSTX	6790	-18.11**	-6912.94**	-85.23**	80	10/29/1987
USD/EUR	3673	-14.56**	-3638.50**	-59.99**	458	10/25/2000
USD/JPY	3673	-15.30**	-3653.02**	-62.34**	3470	09/27/2012
USD/GBP	3673	-14.13**	-3453.52**	-58.91**	2237	11/08/2007
USD/AUD	3673	-14.73**	-3685.81**	-62.63**	3605	04/11/2013
USD/CHF	3673	-15.45**	-3729.60**	-61.56**	3184	08/10/2011
USD/CAD	3673	-15.73**	-3523.31**	-60.71**	2236	11/07/2007
Brent	5228	-15.72**	-5240.08**	-72.90**	1418	12/10/1998
Gold	5228	-17.48**	-5201.25**	-72.07**	4740	09/05/2011
Copper	5228	-15.21**	-5385.71**	-74.92**	3353	05/11/2006
Wheat	5228	-18.19**	-5429.53**	-78.54**	12	07/21/1993
Corn	5228	-16.17**	-5240.34**	-71.87**	788	07/11/1996

Table 3.2: Stationarity tests

This table reports the statistics of the unit root tests for the daily return series of stock indices, exchange rates, and commodity prices. Three tests are considered: the augmented Dickey-Fuller (ADF), the Phillips-Perron (PP) test, and the Zivot-Andrews (ZA) test accounting for a possible structural break in the series. The ordering and the date of this break are in the last two columns of the table. For the three tests, the symbols * and ** denote a rejection of the null hypothesis of unit root at 5% and 1% confidence levels, respectively.

	Obs.	Min	Max	Mean	Std. dev.	Sk.	Ku.	JB	AD	LB(25)R	LB(50)R	D(25)R	LB(25)R^2	LB(50)R^2	LB(25)|R|
CAC40	6790	-0.101	0.106	0.000	0.014	-0.138	5.465	8501.74**	67.81**	55.20**	96.00**	22.00	5337.58**	6736.69**	6051.74**
DAX30	6790	-0.137	0.108	0.000	0.015	-0.319	6.430	11852.62**	90.23**	48.85**	91.32**	20.66	4167.89**	5422.23**	8156.23**
FTSE100	6790	-0.130	0.094	0.000	0.011	-0.502	9.958	28429.22**	84.22**	106.32**	165.07**	32.09	5268.39**	6293.67**	7844.60**
SP500	6790	-0.229	0.116	0.000	0.012	-1.375	30.528	266572.43**	142.19**	77.66**	151.21**	18.38	1262.50**	1547.70**	7828.64**
SPTSX	6790	-0.146	0.173	0.000	0.011	-0.770	27.770	219470.19**	165.97**	100.16**	214.08**	26.96	3189.15**	3804.26**	9893.97**
USD/EUR	3673	-0.046	0.038	0.000	0.007	-0.167	2.329	853.89**	11.52**	14.02	55.33	9.93	730.54**	1234.43**	647.88**
USD/JPY	3673	-0.046	0.037	0.000	0.007	-0.242	3.300	1714.23**	16.01**	31.28	60.35	24.77	315.10**	376.02**	481.81**
USD/GBP	3673	-0.044	0.039	0.000	0.006	0.038	3.992	2456.27**	15.16**	78.49**	163.17**	39.73*	2377.03**	4210.90**	2044.28**
USD/AUD	3673	-0.065	0.086	0.000	0.009	0.814	11.340	20199.39**	48.66**	43.38*	118.69**	12.26	3611.48**	4388.22**	3834.61**
USD/CHF	3673	-0.054	0.085	0.000	0.007	0.302	8.220	10459.77**	14.78**	40.64*	76.55**	27.16	301.24**	340.46**	504.66**
USD/CAD	3673	-0.050	0.043	0.000	0.006	-0.052	5.311	4345.37**	27.25**	50.86**	95.86**	23.91	2895.87**	4574.04**	3378.49**
Brent	5228	-0.136	0.135	0.000	0.022	-0.099	3.000	1980.09**	30.60**	43.48*	103.33**	29.04	1373.02**	2222.31**	1434.07**
Gold	5228	-0.101	0.074	0.000	0.010	-0.373	7.717	13149.18**	104.51**	65.66**	99.79**	32.51	1498.27**	2459.62**	3776.28**
Copper	5228	-0.105	0.117	0.000	0.017	-0.209	4.891	5274.89**	59.41**	58.25**	135.17**	25.98	4109.91**	5443.51**	4661.69**
Wheat	5228	-0.226	0.226	0.000	0.024	-0.176	8.308	15125.09**	64.06**	79.81**	112.02**	77.02	895.22**	1161.11**	2047.59**
Corn	5228	-0.121	0.109	0.000	0.018	-0.175	3.389	2541.35**	51.36**	40.46*	67.77*	27.13	1375.23**	2211.82**	1872.80**

Table 3.3: Descriptive statistics

This table presents the descriptive statistics (columns 3-8) and the results of normality (columns 9 to 10), autocorrelation (columns 11-13), and heteroskedasticity (columns 14-16) tests for the daily return series of stock indices, exchange rates, and commodity prices. For each series, the table reports: (i) the empirical extrema and the first four moments, (ii) the statistics of the Jarque-Bera (JB) and the Anderson-Darling (AD) normality tests, and (iii) the statistics of the Ljung-Box LB(p) and the Diebold D(p) autocorrelation/heteroskedasticity tests for p lags. The symbols * and ** denote a rejection of the null hypothesis of normality, no autocorrelation, or no heteroskedasticity (depending on the test) at the 5% and 1% confidence levels, respectively.

Series 1	Series 2	Obs.	ρ	τ	$\rho^-(1\%)$	$\rho^+(1\%)$	$J_\rho(\rho^-(1\%) = \rho^+(1\%))$	p-value	$\rho^-(10\%)$	$\rho^+(10\%)$	$J_\rho(\rho^-(10\%) = \rho^+(10\%))$	p-value	$\tau^-(1\%)$	$\tau^+(1\%)$
CAC40	DAX30	6790	0.80	0.57	0.62	0.70	0.06	0.81	0.79	0.71	6.94	0.73	0.84	0.73
CAC40	FTSE100	6790	0.79	0.56	0.85	0.76	0.05	0.83	0.79	0.72	4.63	0.91	0.82	0.78
CAC40	SP500	6790	0.28	0.17	0.33	0.45	0.12	0.73	0.39	0.43	6.24	0.79	0.43	0.43
CAC40	SPTSX	6790	0.43	0.26	0.19	0.13	0.04	0.83	0.42	0.28	8.17	0.61	0.51	0.47
DAX30	FTSE10	6790	0.71	0.48	0.37	0.60	0.32	0.57	0.66	0.63	8.06	0.62	0.61	0.75
DAX30	SP500	6790	0.27	0.18	0.28	0.35	0.06	0.81	0.43	0.32	5.59	0.85	0.35	0.39
DAX30	SPTSX	6790	0.41	0.25	0.22	0.13	0.12	0.73	0.48	0.28	8.36	0.59	0.34	0.47
FTSE100	SP500	6790	0.33	0.17	0.73	0.10	1.48	0.24	0.59	0.45	7.79	0.65	0.31	0.46
FTSE100	SPTSX	6790	0.46	0.27	0.65	-0.17	6.39	0.01	0.58	0.33	14.83	0.14	0.40	0.49
SP500	SPTSX	6790	0.59	0.38	0.71	0.57	0.00	0.94	0.63	0.56	4.97	0.89	0.46	0.59
USD/EUR	USD/JPY	3673	0.24	0.18	0.72	-0.91	21.20	0.00	0.53	0.30	32.33	0.00	0.35	0.37
USD/EUR	USD/GBP	3673	0.67	0.48	0.67	0.09	1.78	0.18	0.57	0.44	8.03	0.63	0.61	0.56
USD/EUR	USD/AUD	3673	0.55	0.39	0.48	-0.14	7.87	0.01	0.58	0.26	16.72	0.08	0.50	0.49
USD/EUR	USD/CHF	3673	0.84	0.71	0.79	0.92	0.06	0.80	0.81	0.72	7.27	0.70	0.80	0.61
USD/EUR	USD/CAD	3673	0.45	0.28	-0.12	-0.05	0.09	0.76	0.58	0.32	12.28	0.27	0.54	0.52
USD/JPY	USD/GBP	3673	0.13	0.13	-0.13	-0.05	0.00	0.00	0.40	0.12	7.25	0.61	0.23	0.54
USD/JPY	USD/AUD	3673	0.01	0.10	1.00	-	-0.00	0.00	0.55	0.42	5.76	0.76	0.22	0.13
USD/JPY	USD/CHF	3673	0.36	0.25	0.10	0.02	0.03	0.87	0.44	0.38	10.04	0.44	0.35	0.36
USD/JPY	USD/CAD	3673	-0.01	0.02	1.00	-1.00	0.00	0.00	0.48	0.30	6.95	0.64	0.10	0.13
USD/GBP	USD/AUD	3673	0.53	0.35	0.77	-0.02	4.94	0.03	0.57	0.33	16.62	0.08	0.23	0.49
USD/GBP	USD/CHF	3673	0.55	0.42	0.61	1.00	0.58	0.45	0.41	0.33	10.11	0.43	0.61	0.37
USD/GBP	USD/CAD	3673	0.45	0.26	0.60	0.19	1.39	0.24	0.61	0.48	5.39	0.86	0.58	0.59
USD/AUD	USD/CHF	3673	0.40	0.30	0.03	0.04	0.00	0.98	0.45	0.19	5.58	0.85	0.51	0.30
USD/AUD	USD/CAD	3673	0.60	0.39	0.85	0.53	0.64	0.42	0.73	0.50	9.44	0.49	0.71	0.65
USD/CHF	USD/CAD	3673	0.33	0.21	-0.42	0.30	3.41	0.06	0.39	0.26	10.63	0.39	0.42	0.49

Series 1	Series 2	Obs.	ρ	τ	$\rho^-(1\%)$	$\rho^+(1\%)$	$J_\rho(\rho^-(1\%) = \rho^+(1\%))$	p-value	$\rho^-(10\%)$	$\rho^+(10\%)$	$J_\rho(\rho^-(10\%) = \rho^+(10\%))$	p-value	$\tau^-(1\%)$	$\tau^+(1\%)$
Brent	Gold	5228	0.18	0.11	1.00	-0.34	0.00	0.00	-0.05	0.36	14.92	0.09	0.18	0.33
Brent	Copper	5228	0.23	0.12	0.10	0.21	0.11	0.74	0.34	0.29	11.74	0.30	0.41	0.36
Brent	Wheat	5228	0.11	0.06	1.00	1.00	0.00	0.00	0.10	0.24	16.75	0.05	0.21	0.25
Brent	Corn	5228	0.13	0.06	0.77	1.00	0.38	0.54	0.31	0.31	9.86	0.45	0.33	0.22
Gold	Copper	5228	0.28	0.17	-0.59	-0.07	2.85	0.09	0.26	0.20	9.64	0.47	0.36	0.39
Gold	Wheat	5228	0.10	0.06	-0.03	1.00	0.00	0.00	0.14	0.01	9.43	0.40	0.24	0.20
Gold	Corn	5228	0.11	0.06	-0.38	-1.00	0.00	0.00	0.26	0.15	15.21	0.09	0.29	0.28
Copper	Wheat	5228	0.14	0.07	0.17	0.52	0.86	0.35	0.28	0.22	7.16	0.71	0.18	0.08
Copper	Corn	5228	0.16	0.07	0.57	1.00	0.78	0.38	0.35	0.34	7.74	0.65	0.37	0.28
Wheat	Corn	5228	0.49	0.35	0.57	-0.15	7.52	0.01	0.41	0.28	22.28	0.01	0.57	0.54

Table 3.4: Nonparametric extreme dependence coefficients and tests of asymmetry

This table reports, for each pair of series: (i) the linear correlation (ρ) and rank (τ) coefficients, (ii) the extreme correlation ($\rho^\pm$) and tail dependence ($\tau^\pm$) coefficients for 1%, 10%, 90% and 99% quartiles, and (iii) the statistics and p-values of the Hong et al. (2007) test (J_ρ) for the asymmetry of extreme correlations. For the latter, a p-value close to 1 indicates a rejection of the null hypothesis of asymmetric extreme correlations.

	Dependence		Tail dependence			Parameter space	
Copula	Negative	Positive	Lower	Upper	Asymetry	1st parameter	2nd parameter
Gaussian	√	√				$(-1,1)$	–
Student t	√	√	√	√		$(-1,1)$	$(2,\infty)$
Clayton		√		√	√	$(0,\infty)$	–
Gumbel	√		√		√	$[1,\infty)$	–
Frank	√	√				$\mathbb{R}-\{0\}$	–
Joe	√		√		√	$(1,\infty)$	–
BB1 (Clayton-Gumbel)	√		√	√	√	$(0,\infty)$	$[1,\infty)$
BB6 (Joe-Gumbel)		√		√	√	$[1,\infty)$	$[1,\infty)$
BB7 (Joe-Clayton)	√		√	√	√	$[1,\infty)$	$(0,\infty)$
BB8 (Joe-Frank)	√	√				$[1,\infty)$	$(0,1]$
Rotated Clayton (RC)	√				√	$(-\infty,-1]$	$(-\infty,0)$
Rotated Gumbel (RG)	√				√	$(-\infty,-1]$	$(-\infty,-1]$

Table 3.5: Statistical features of the bivariate copulas

																LB(20)		ARCH(20)
	Obs.	Model	$\hat{\mu}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\omega}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	LLH	$\hat{z}$	$\hat{z}^2$	$\hat{z}$
CAC40	6540	ARMA(3,3)-EGARCH(3,2)[a]	0.00	-0.42	0.47	0.44	-0.47	-0.29	0.07	0.13	-0.09	-0.04	0.30	0.86	19630.77	0.85	0.86	0.93
			(0.00)	(0.33)	(0.42)	(0.33)	(0.42)	(0.04)	(0.02)	(0.02)	(0.01)	(0.01)	(0.11)	(0.06)				
DAX30	6540	ARMA(3,2)-EGARCH(1,2)[b]	0.00	0.03	-0.91	-0.03	0.89	-0.24	0.02	0.14	-0.20	0.11	0.97		19538.66	0.48	1.00	0.79
			(0.00)	(0.07)	(0.06)	(0.06)	(0.06)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.00)					
FTSE100	6540	ARMA(2,3)-EGARCH(3,2)[c]	0.00	0.14	0.32	-0.12	-0.33	0.02	0.15	-0.13	-0.14	0.13	1.71	-0.62	21224.07	0.19	0.63	0.31
			(0.00)	(0.46)	(0.42)	(0.45)	(0.42)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.08)	(0.14)				
SP500	6540	ARMA(2,3)-EGARCH(2,1)[d]	0.00	1.79	-0.91	-1.79	0.88	-0.20	0.14		-0.12		0.83	0.15	21328.63	0.80	0.98	0.52
			(0.00)	(0.03)	(0.04)	(0.04)	(0.05)	(0.02)	(0.01)		(0.01)		(0.07)	(0.07)				
SPTSX	6540	ARMA(3,3)-EGARCH(2,1)[e]	0.00	-0.85	-0.49	0.92	0.58	-0.34	0.27		-0.12		0.41	0.55	22104.60	0.49	0.98	0.95
			(0.00)	(0.15)	(0.18)	(0.15)	(0.19)	(0.02)	(0.01)		(0.01)		(0.02)	(0.02)				
USD/EUR	3423	ARMA(3,3)-EGARCH(1,2)[f]	-0.00	-0.74	-0.99	0.74	0.99	-0.05	-0.13	0.21	0.04	-0.03	0.99		12522.03	0.95	0.53	0.64
			(0.00)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)	(0.04)	(0.02)	(0.02)	(0.00)					
USD/JPY	3423	ARMA(2,3)-EGARCH(1,3)[g]	-0.00	-0.82	-0.85	0.80	0.83	-0.14	0.07	0.12	-0.08	-0.00	0.99		12421.88	0.64	0.87	074
			(0.00)	(0.11)	(0.11)	(0.11)	(0.11)	(0.03)	(0.03)	(0.04)	(0.02)	(0.03)	(0.00)					
USD/GBP	3423	ARMA(3,3)-EGARCH(1,3)[h]	-0.00	0.63	0.37	-0.62	-0.39	-0.09	0.02	0.04	0.09	-0.02	0.99		12939.67	0.86	0.98	0.93
			(0.00)	(0.17)	(0.25)	(0.16)	(0.24)	(0.02)	(0.04)	(0.05)	(0.02)	(0.03)	(0.00)					
USD/AUD	3423	ARMA(3,3)-EGARCH(3,1)[i]	-0.00	1.17	-1.18	-1.17	1.17	-0.20	0.19		0.08		0.61	0.00	11832.61	0.88	0.12	0.81
			(0.00)	(0.02)	(0.00)	(0.02)	(0.01)	(0.03)	(0.02)		(0.01)		(0.11)	(0.15)				
USD/CHF	3423	ARMA(2,3)-EGARCH(3,3)[j]	-0.00	-0.97	0.91	0.94	0.91	-0.20	0.01	0.15	0.00	-0.03	-0.06	0.18	12232.85	0.81	0.00	0.89
			(0.00)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.02)	(0.02)	(0.01)	(0.01)	(0.05)	(0.05)				
USD/CAD	3423	ARMA(3,2)-EGARCH(3,1)[k]	-0.00	-1.61	-1.02	1.60	1.00	0.18	0.18		0.03		0.42	0.47	13066.07	0.99	0.67	0.31
			(0.00)	(0.02)	(0.03)	(0.00)	(0.00)	(0.03)	(0.02)		(0.01)		(0.13)	(0.16)				

[a] $\hat{\phi}_3$= 0.17(0.30), $\hat{\theta}_3$= −0.20(0.30), $\hat{\beta}_3$= −0.19(0.09).
[b] $\hat{\phi}_3$= 0.01(0.01).
[c] $\hat{\theta}_3$= −0.03(0.01), $\hat{\beta}_3$= −0.09(0.08).
[d] $\hat{\theta}_3$= 0.02(0.01).
[e] $\hat{\phi}_3$= 0.40(0.15), $\hat{\theta}_3$= −0.33(0.15).
[f] $\hat{\phi}_3$= −0.01(0.02).
[g] $\hat{\theta}_3$= −0.04(0.02), $\hat{\alpha}_3$= −0.11(0.03), $\hat{\gamma}_3$= 0.06(0.02).
[h] $\hat{\phi}_3$= −0.88(0.17), $\hat{\theta}_3$= 0.89(0.16), $\hat{\alpha}_3$= 0.04(0.04), $\hat{\gamma}_3$= −0.06(0.02).
[i] $\hat{\phi}_3$= 0.99(0.02), $\hat{\theta}_3$= −0.98(0.02), $\hat{\beta}_3$= 0.36(0.09).
[j] $\hat{\theta}_3$= −0.03(0.02), $\hat{\alpha}_3$= 0.07(0.03), $\hat{\gamma}_3$= −0.04(0.01), $\hat{\beta}_3$= 0.09(0.04).
[k] $\hat{\phi}_3$= −0.02(0.02), $\hat{\beta}_3$= 0.10(0.13).

															LB(20)		ARCH(20)
	Obs.	Model	$\hat{\mu}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\omega}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	LLH	$\hat{z}$	$\hat{z}^2$	$\hat{z}$
Brent	4978	ARMA(3,2)-EGARCH(1,3)[a]	0.00	-1.12	-0.85	1.12	0.85	-0.07	0.10	0.00	-0.04	0.01	0.99		12267.83	0.54	0.68	0.10
			(0.00)	(0012)	(0.11)	(0.12)	(0.12)	(0.01)	(0.02)	(0.03)	(0.01)	(0.01)	(0.00)					
Gold	4978	ARMA(2,3)-EGARCH(3,2)[b]	0.00	1.21	-0.49	-1.20	0.50	-0.02	0.18	-0.16	0.05	-0.03	1.70	-0.74	16649.23	0.11	0.39	0.00
			(0.00)	(0.35)	(0.33)	(0.35)	(0.32)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.11)	(0.19)				
Copper	4978	ARMA(3,3)-EGARCH(2,2)[c]	0.00	1.80	-1.43	-1.80	1.40	-0.11	0.13		-0.01		0.99		1381761	0.96	0.27	0.91
			(0.00)	(0.19)	(0.33)	(0.20)	(0.34)	(0.02)	(0.01)		(0.00)		(0.00)					
Wheat	4978	ARMA(2,3)-EGARCH(1,1)[d]	0.00	1.45	-0.81	-1.19	0.84	-0.02	0.21	-0.14	0.03	-0.03	1.40	-0.41	12004.93	0.18	0.69	0.04
			(0.00)	(0.17)	(0.13)	(0.17)	(0.14)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.13)	(0.13)				
Corn	4978	ARMA(3,3)-EGARCH(2,2)[e]	0.00	0.20	-0.16	-0.19	0.16	-0.05	0.14	-0.06	-0.10	0.10	1.36	-0.37	13280.55	0.21	0.80	0.22
			(0.00)	(0.07)	(0.07)	(0.08)	(0.08)	(0.01)	(0.02)	(0.03)	(0.02)	(0.01)	(0.14)	(0.13)				

[a] $\hat{\phi}_3 = -0.02(0.02)$, $\hat{\alpha}_3 = -0.01(0.02)$, $\hat{\gamma}_3 = 0.00(0.01)$.
[b] $\hat{\theta}_3 = -0.02(0.02)$, $\hat{\beta}_3 = 0.04(0.10)$.
[c] $\hat{\phi}_3 = 0.62(0.17)$, $\hat{\theta}_3 = -0.58(0.18)$.
[d] $\hat{\theta}_3 = -0.03(0.02)$.
[e] $\hat{\phi}_3 = 0.89(0.07)$, $\hat{\theta}_3 = -0.87(0.08)$.

Table 3.6: Estimation and diagnostic tests of the ARMA-EGARCH models

This table presents the estimation results of the ARMA-EGARCH models for the daily return series of stock indices, exchange rates, and commodity prices. The table reports: the estimates of the ARMA (columns 4-8) and the EGARCH (columns 9 to 15) models. The corresponding standard deviations are given in brackets. LLH is the estimation log-likelihood. LB(p) and ARCH(p) refer, respectively, to the p-values of the Ljung-Box and the Engle (1983) tests for the absence of autocorrelation and the absence of ARCH effects in the models' residuals, for p lags.

										Semi-parametric distribution				KS					AD				
	Obs.	Sk.	Ku.	u^L	u^U	ξ^L	ξ^U	β^L	β^U	KS	KS+	KS-	AD	Nor	St	SSt	GH	GH-SSt	Nor	St	SSt	GH	GH-SSt
CAC40	6540	-0.41	2.93	-1.26	1.23	0.05	-0.04	0.56	0.49	0.98	0.64	0.87	1.00	0.00	0.03	0.01	0.01	0.00	0.00	0.04	0.10	0.11	0.00
DAX30	6540	-0.74	7.84	-1.24	1.20	0.12	0.51	0.52	0.45	1.00	0.86	0.79	1.00	0.00	0.02	0.15	0.20	0.00	0.00	0.03	0.20	0.23	0.00
FTSE100	6540	-0.41	3.80	-1.23	1.23	0.04	0.01	0.57	0.46	1.00	0.81	0.71	0.98	0.00	0.00	0.08	0.10	0.00	0.00	0.05	0.28	0.32	0.00
SP500	6540	-0.66	4.12	-1.22	1.20	0.10	-0.14	0.60	0.60	1.00	0.85	0.91	1.00	0.00	0.00	0.01	0.10	0.00	0.00	0.01	0.02	0.19	0.00
SPTSX	6540	-0.77	6.86	-1.20	1.16	0.15	0.05	0.58	0.53	1.00	0.89	0.91	1.00	0.00	0.12	0.04	0.03	0.00	0.00	0.12	0.22	0.21	0.00
USD/EUR	3423	-0.13	0.89	-1.26	1.24	0.01	-0.14	0.55	0.61	1.00	0.85	0.86	1.00	0.00	0.54	0.38	0.86	0.00	0.00	0.57	0.51	0.87	0.00
USD/JPY	3423	-0.22	2.05	-1.22	1.19	0.09	0.05	0.56	0.51	1.00	0.91	0.91	1.00	0.00	0.95	0.84	0.89	0.00	0.00	0.81	0.93	0.96	0.00
USD/GBP	3423	-0.02	0.69	-1.24	1.28	-0.01	-0.12	0.53	0.57	1.00	0.85	0.93	1.00	0.09	0.69	0.85	0.96	0.00	0.04	0.73	0.83	0.94	0.00
USD/AUD	3423	0.56	2.56	-1.18	1.21	-0.06	0.13	0.51	0.58	1.00	0.93	0.95	1.00	0.00	0.49	0.90	0.95	0.00	0.00	0.19	0.98	1.00	0.00
USD/CHF	3423	-0.12	1.81	-1.25	1.20	-0.02	0.06	0.61	0.49	1.00	0.85	0.84	1.00	0.00	0.47	0.76	0.49	0.00	0.00	0.25	0.87	0.76	0.00
USD/CAD	3423	0.10	0.98	-1.22	1.22	-0.10	0.03	0.58	0.55	1.00	0.90	0.89	1.00	0.10	0.88	0.94	0.93	0.39	0.05	0.89	0.91	0.85	0.48
Brent	4978	1.40	-0.17	-1.24	1.21	-0.02	0.04	0.63	0.50	1.00	0.92	0.91	1.00	0.00	0.11	0.06	0.13	0.00	0.00	0.09	0.22	0.40	0.00
Gold	4978	3.54	0.05	-1.18	1.14	-0.00	0.10	0.67	0.58	1.00	0.97	0.86	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.03	0.00
Copper	4978	3.11	-0.50	-1.17	1.20	0.15	-0.00	0.55	0.56	1.00	0.96	0.95	1.00	0.00	0.36	0.23	0.23	0.00	0.00	0.47	0.46	0.46	0.00
Wheat	4978	3.76	0.20	-1.14	1.20	0.04	0.08	0.60	0.57	1.00	0.93	0.88	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.05	0.05	0.00
Corn	4978	2.25	-0.20	-1.20	1.21	0.09	0.02	0.57	0.55	1.00	0.90	0.89	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.08	0.00

Table 3.7: Etimation and GoF tests for the marginal distributions

This table reports the parameter estimates of the marginal distributions, including extreme quartile values at 10% and 90% thresholds (columns 3 and 4), and the shape and the scale parameters of the GPD on the lower (columns 5 and 6) and upper (columns 7 and 8) tails. The adequacy of the estimated distributions is evaluated through Goodness of Fit tests applied to the probability integration transformation (PIT) of the filtered residuals of the ARMA-EGARCH models. The p-values of four tests have been reported. It consists in three versions of the Kolmogorov-Smirnov (KS) test, with as alternative hypotheses: a PIT distribution different (KS), above (KS+) or below (KS-) the uniform distribution, and the Anderson-Darling (AD) uniformity test. Moreover, the performance of the presented semi-parametric specification is compared to that of five parametric distributions commonly used in practice: the normal distribution (Nor), the Student t distribution (St), the asymmetric Student t distribution (SSt) of Fernández and Steel (1998), the generalized hyperbolic distribution (GH) and the asymmetric t distribution (GH-SSt) of Aas and Hobæk Haff (2006).

Tree	Obs.	Series 1	Series 2	Conditioning series	Pair-copula	1st parameter	2nd parameter	empirical τ	theoretical τ	LLH
1	6540	DAX30	CAC40		Student t	0.75(0.01)	3.11(0.17)	0.54	0.54	2825.78
	6540	CAC40	FTSE100		Student t	0.74(0.01)	4.10(0.26)	0.53	0.53	2623.70
	6540	FTSE100	SPTSX		BB1	0.31(0.02)	1.15(0.01)	0.26	0.25	577.28
	6540	SPTSX	SP500		Student t	0.54(0.01)	10.60(1.57)	0.37	0.36	1154.68
2	6540	DAX30	FTSE100	CAC40	Student t	0.21(0.01)	7.32(0.72)	0.14	0.14	208.35
	6540	CAC40	SPTSX	FTSE100	Gaussian	0.15(0.01)		0.10	0.10	79.05
	6540	FTSE100	SP500	SPTSX	Gaussian	0.05(0.01)		0.02	0.03	8.14
3	6540	DAX30	SPTSX	CAC40, FTSE100	Student t	0.09(0.01)	30.00(-)	0.06	0.06	31.53
	6540	CAC40	SP500	FTSE100, SPTSX	Gumbel	1.01(0.01)		0.02	0.01	4.13
4	6540	DAX30	SP500	CAC40, FTSE100, SPTSX	Gaussian	0.09(0.01)		0.05	0.06	26.40
										7539.04
1	3423	USD/GBP	USD/EUR		Student t	0.69(0.01)	9.74(1.58)	0.49	0.48	1106.95
	3423	USD/CAD	USD/AUD		Student t	0.55(0.01)	6.57(0.84)	0.37	0.37	631.87
	3423	USD/AUD	USD/EUR		Student t	0.57(0.01)	7.17(0.95)	0.39	0.39	682.03
	3423	USD/CHF	USD/EUR		Student t	0.90(0.00)	2.31(0.15)	0.71	0.71	2899.55
	3423	USD/CHF	USD/JPY		Student t	0.40(0.02)	6.00(0.72)	0.27	0.26	330.02
2	3423	USD/GBP	USD/AUD	USD/EUR	Student t	0.20(0.02)	19.17(5.90)	0.13	0.13	71.62
	3423	USD/CAD	USD/EUR	USD/AUD	Student t	0.14(0.02)	11.09(2.24)	0.09	0.09	44.64
	3423	USD/CHF	USD/AUD	USD/EUR	Student t	-0.11(0.02)	16.13(4.20)	-0.07	-0.07	32.37
	3423	USD/EUR	USD/JPY	USD/CHF	Student t	-0.10(0.02)	23.41(8.27)	-0.06	-0.07	24.92
3	3423	USD/GBP	USD/CHF	USD/AUD, USD/EUR	Student t	0.03(0.01)	20.44(6.79)	0.03	0.02	6.07
	3423	USD/CAD	USD/CHF	USD/EUR, USD/AUD	Frank	-0.47(0.10)		-0.05	-0.05	10.10
	3423	USD/AUD	USD/JPY	USD/CHF, USD/EUR	Student t	0.03(0.02)	8.00(1.24)	0.03	0.02	26.82
4	3423	USD/GBP	USD/CAD	USD/CHF, USD/AUD, USD/EUR	R.G.	1.03(0.01)		0.03	0.02	4.79
	3423	USD/CAD	USD/JPY	USD/CHF, USD/EUR, USD/AUD	Frank	-0.41(0.10)		-0.04	-0.05	7.68
5	3423	USD/GBP	USD/JPY	USD/CAD, USD/CHF, USD/AUD, USD/EUR	Student t	0.02(0.02)	18.56(5.63)	0.02	0.01	6.54
										5885.97

Tree	Obs.	Series 1	Series 2	Conditioning series	Pair-copula	1st parameter	2nd parameter	empirical τ	theoretical τ	LLH
1	4978	Gold	Copper		RG	1.16(0.01)		0.15	0.14	149.08
	4978	Brent	Copper		Student t	0.18(0.01)	12.63(2.61)	0.11	0.11	93.77
	4978	Copper	Corn		RG	1.07(0.01)		0.07	0.06	36.36
	4978	Corn	Wheat		Student t	0.51(0.01)	8.21(1.06)	0.35	0.34	768.19
2	4978	Gold	Brent	Copper	Frank	0.78(0.08)		0.09	0.09	41.10
	4978	Brent	Corn	Copper	Gaussian	0.08(0.01)		0.05	0.05	14.19
	4978	Copper	Wheat	Corn	BB8	1.21(0.13)	0.76(0.18)	0.04	0.04	10.15
3	4978	Gold	Corn	Brent, Copper	Gaussian	0.05(0.01)		0.03	0.03	6.39
	4978	Brent	Wheat	Corn, Copper	Student t	0.05(0.01)	21.15(6.05)	0.03	0.03	11.61
4	4978	Gold	Wheat	Corn, Brent, Copper	Frank	0.25(0.09)		0.03	0.03	4.23
										1135.08

Table 3.8: Sequential estimation of the R-vine model

This table reports the estimation results of the dependence structure specified by the R-vine model for the daily return series of stock indices, exchange rates, and commodity prices. The table reports: the tree number in the graphical model (column 1), the names of the conditional (columns 3 and 4) and conditioning (column 5) series, the name of the corresponding pair-copula (column 6), the corresponding parameter estimates (columns 7 and 8) and standard deviations (in brackets), as well as the empirical and the theoretical rank coefficients (columns 9 and 10) measuring the unconditional dependence on the first tree and the conditional dependence on the following trees. LLH denotes the log-likelihood of the sequential estimation of the pair-copulas (in light) and that of the multivariate system (in bold).

							Clarke			Vuong		
	Model	Obs.	LLH	No. of parameters	AIC	BIC	wo/correction	AIC	BIC	wo/correction	AIC	BIC
Equity indices	Rvine	6540	7539.04	16	-15046.08	-14937.51	-	-	-	-	-	-
	Cvine	6540	7528.27	17	-15022.54	-14907.18	3051	3056	3069	0.66	0.72	0.93
							(0.00)	(0.00)	(0.00)	(0.51)	(0.47)	
	Dvine	6540	7556.80	16	-15081.60	-14973.03	3031	3031	3031	-1.77(0.08)	-1.77	-1.77
							(0.00)	(0.00)	(0.00)	(0.00)	(0.08)	(0.08)
	Student t	6540	7234.67	11	-14447.34	-14372.7	3358	3349	3320	0.64	0.50	0.02
							(0.03)	(0.05)	(0.22)	(0.52)	(0.62)	(0.98)
	Gaussian	6540	6884.67	10	-13749.34	-13681.48	3894	3886	3864	13.55	13.43	13.00
							(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Exchange rates	Rvine	3423	5885.97	27	-11717.94	-11552.21	-	-	-	-	-	-
	Cvine	3423	5878.00	28	-11699.99	-11528.12	1676	1681	1697	0.98(0.33)	1.10	1.38
							(0.23)	(0.31)	(0.63)	(0.33)	(0.27)	(0.14)
	Dvine	3423	5760.69	27	-11467.38	-11301.65	1898	1898	1898	6.82	6.82	6.82
							(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	Student t	3423	5050.87	16	-10069.74	-9971.53	1901	1901	1901	0.94	0.94	0.94
							(0.00)	(0.00)	(0.00)	(0.35)	(0.35)	(0.35)
	Gaussian	3423	4616.02	15	-9202.04	-9109.97	2208	2200 2167	11.65	11.45	10.84	
							(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Commodity prices	Rvine	4978	1135.08	14	-2242.15	-2150.97	-	-	-	-	-	-
	Cvine	4978	1124.38	14	-2220.76	-2129.58	2325	2325	2325	1.34	1.34	1.34
							(0.00)	(0.00)	(0.00)	(0.18)	(0.18)	(0.18)
	Dvine	4978	1125.75	14	-2223.49	-2132.31	2407	2407	2407	1.38	1.38	1.38
							(0.02)	(0.02)	(0.02)	(0.17)	(0.17)	(0.17)
	Student t	4978	1102.79	11	-2111.93	-2183.58	2328	2340	2384	0.84	1.07	1.81
							(0.00)	(0.00)	(0.00)	(0.40)	(0.29)	(0.07)
	Gaussian	4978	1051.30	10	-2120.44	-2082.60	2816	2806	2763	5.44	5.19	4.40
							(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Table 3.9: Assessment of the static properties of the R-vine model

This table compares the static (in sample) properties of the R-vine model to restricted vine-copulas (C-vine D-vine vine) and meta-copulas (Gaussian, Student t) models. The comparison is based on the estimation log-likelihood (LLH), the number of parameters, and the AIC and BIC information criteria. For non-nested models, the Vuong (1989) and Clarke (2007) tests are also performed, with and without correction by the number of parameters – according to AIC and BIC criteria. Both tests consider, under the null hypothesis, that the considered model at least have a comparable performance with respect to the R-vine model. The tests' statistics are reported, while the corresponding p-values are given in brackets.

Portfolio	Model	VaR(%)	No. of exceedances	Weibull	GMM (2 orders)	GMM (5 orders)
Type A	R-vine	90	25	0.582	0.328	0.352
		95	14	0.605	0.352	0.532
		99	4	0.623	0.380	0.443
	C-vine	90	29	0.703	0.350	0.638
		95	16	0.713	0.366	0.701
		99	4	0.721	0.380	0.717
	Student t meta-copula	90	31	0.754	0.421	0.846
		95	17	0.636	0.383	0.449
		99	7	0.885	0.552	0.777
	DCC-Normal	90	45	0.954	0.721	0.891
		95	28	0.807	0.554	0.741
		99	11	0.982	0.782	0.992
	DCC-Student t (df=4)	90	22	0.655	0.402	0.667
		95	8	0.903	0.664	0.911
		99	0	-	-	-
	DCC-Student t (df=7)	90	24	0.437	0.474	0.811
		95	11	0.665	0.412	0.551
		99	1	0.992	0.872	0.997
Type B	R-vine	90	27	0.612	0.330	0.502
		95	16	0.665	0.382	0.522
		99	4	0.623	0.380	0.514
	C-vine	90	28	0.753	0.350	0.505
		95	17	0.703	0.366	0.510
		99	4	0.721	0.380	0.513
	Student t meta-copula	90	38	0.754	0.501	0.724
		95	15	0.756	0.363	0.511
		99	7	0.885	0.552	0.756
	DCC-Normal	90	41	0.784	0.841	0.882
		95	30	0.817	0.574	0.780
		99	12	0.954	0.792	0.865
	DCC-Student t (df=4)	90	20	0.655	0.482	0.738
		95	7	0.903	0.604	0.903
		99	0	-	-	-
	DCC-Student t (df=7)	90	26	0.732	0.587	0.801
		95	13	0.645	0.542	0.739
		99	1	0.992	0.872	0.965

Portfolio	Model	VaR(%)	No. of exceedances	Weibull	GMM (2 orders)	GMM (5 orders)
Type C	R-vine	90	28	0.653	0.346	0.511
		95	15	0.665	0.382	0.587
		99	4	0.623	0.380	0.441
	C-vine	90	29	0.782	0.389	0.338
		95	18	0.866	0.786	0.856
		99	4	0.721	0.380	0.614
	Student t meta-copula	90	42	0.804	0.598	0.888
		95	15	0.856	0.783	0.792
		99	7	0.885	0.552	0.838
	DCC-Normal	90	41	0.784	0.841	0.967
		95	26	0.671	0.446	0.641
		99	12	0.954	0.792	0.801
	DCC-Student t (df=4)	90	17	0.541	0.322	0.389
		95	8	0.983	0.664	0.547
		99	0	-	-	-
	DCC-Student t (df=7)	90	30	0.790	0.287	0.653
		95	13	0.695	0.542	0.722
		99	1	0.992	0.872	0.991

Table 3.10: Assessment of the dynamic properties of the R-vine model (Baktesting)

This table compares the results of the backtesting procedure for different multivariate specifications including: (i) three versions of the DCC model (DCC-Normal, and DCC-Student t with 4 and 7 degrees of freedom df), representing the direct approach, (ii) a Student t meta-copula model for the classical indirect approach, and (iii) two pair-copulas models (R-vine, C-Vine) for the modern indirect approach.
For each model and for different confidence levels of the VaR (column 3), the table reports the number of exceedances (column 4) and the p-values of two duration-based becktesting procedures (columns 5-7): the Weibull test of independence by Christoffersen and Pelletier (2004) and the GMM test of conditional coverage by Candelon et al. (2011) with 2 and 5 orders. Three distinct groups of analytical stock portfolios are considered (types A, B and C). The first, contains two (long and short) equally weighted portfolios. The second, contains ten portfolios formed by a long (resp. short) position for one asset and a short (resp. long) position for the others. The third group includes twenty portfolios formed by a long (resp. short) position for two assets and a short (resp. long) position for the rest. The reported number of exceedances and p-values are averaged per group.

		Marginal distribution			
	Probability (%)	Empirical	Semi-parametric	Normal	Student t
Equity indices	0.01	-11.65	-6.81	-3.76	-5.85
	0.02	-8.10	-6.06	-3.57	-5.20
	0.05	-5.33	-5.15	-3.30	-4.45
Exchange rates	0.01	-5.10	-5.02	-3.70	-5.47
	0.02	-4.99	-4.62	-3.55	-4.90
	0.05	-4.55	-4.11	-3.39	-4.28
Commodity prices	0.01	-6.64	-6.34	-3.72	-7.21
	0.02	-6.12	-6.72	-3.53	-6.23
	0.05	-4.99	-4.93	-3.29	-5.15

Table 3.11: Estimation of the initial shock in the stress scenarios

(in pourcentage of the long equally weighted portfolios)

			Multivariate model		
	Probability (%)	Horizon (days)	R-vine	DCC-Student t	Student t meta copula
Equity indices	0.02	3	8.88	8.12	9.07
	0.02	10	12.01	10.24	12.05
	0.05	3	8.44	7.92	8.72
	0.05	10	9.45	8.53	9.70
Exchange rates	0.02	3	8.78	8.27	10.12
	0.02	10	11.98	10.13	12.34
	0.05	3	8.84	7.87	9.40
	0.05	10	10.31	9.80	10.11
Commodity prices	0.02	3	8.35	7.72	8.55
	0.02	10	11.72	10.21	11.63
	0.05	3	7.62	6.53	8.02
	0.05	10	10.11	9.76	10.45

Table 3.12: Impact of the scenarios in terms of portfolios' return

This table presents the change (in percentage) in the portfolio's return in response to two stress scenarios with probabilities $\alpha = 0.02\%$ and $\alpha = 0.05\%$. The results are reported for two stress horizons $h = 3$ and $h = 10$ days. Equally weighted portfolios are considered for the three datasets.

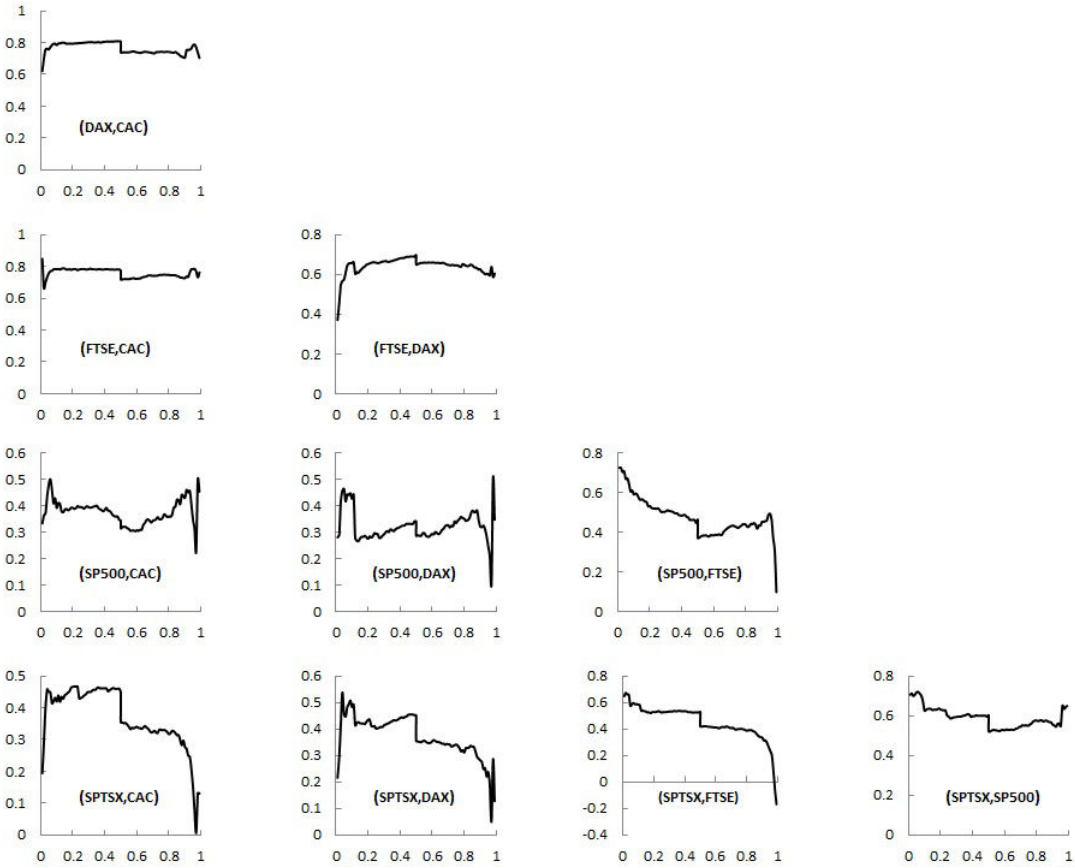

Figure 3.1: Empirical exceedance correlations (equity indices)

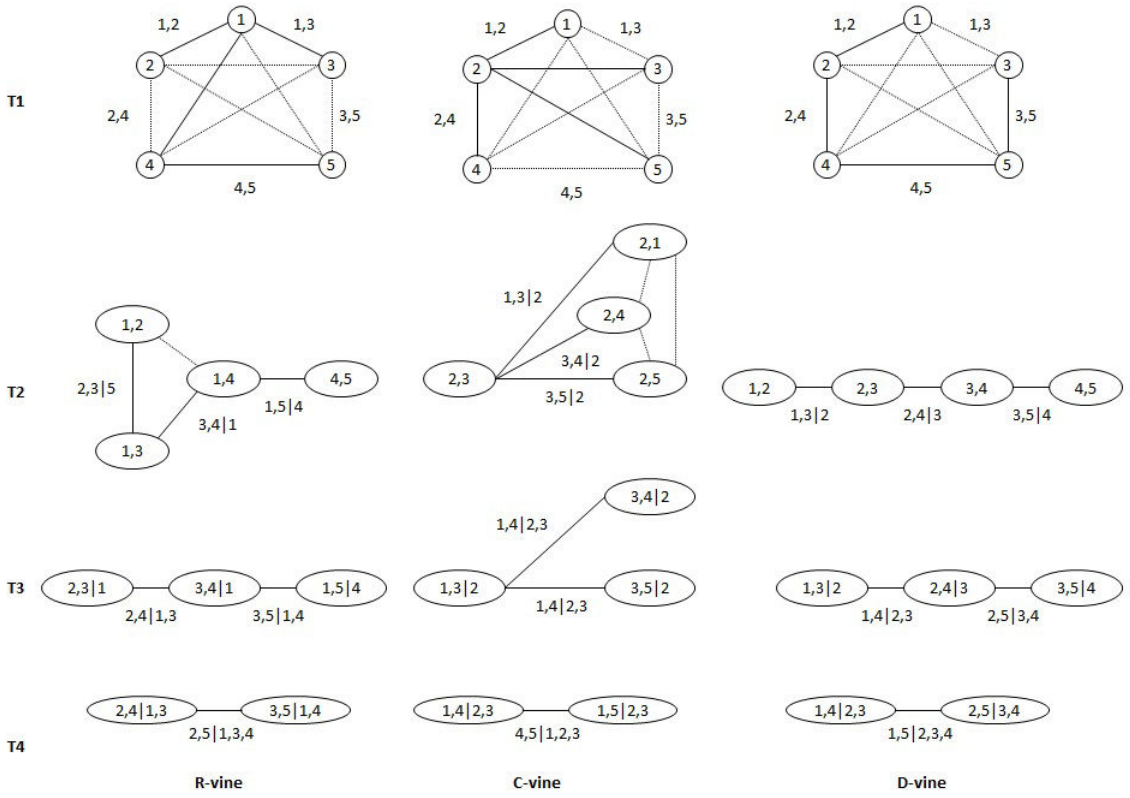

Figure 3.2: Vine-copulas models' structure

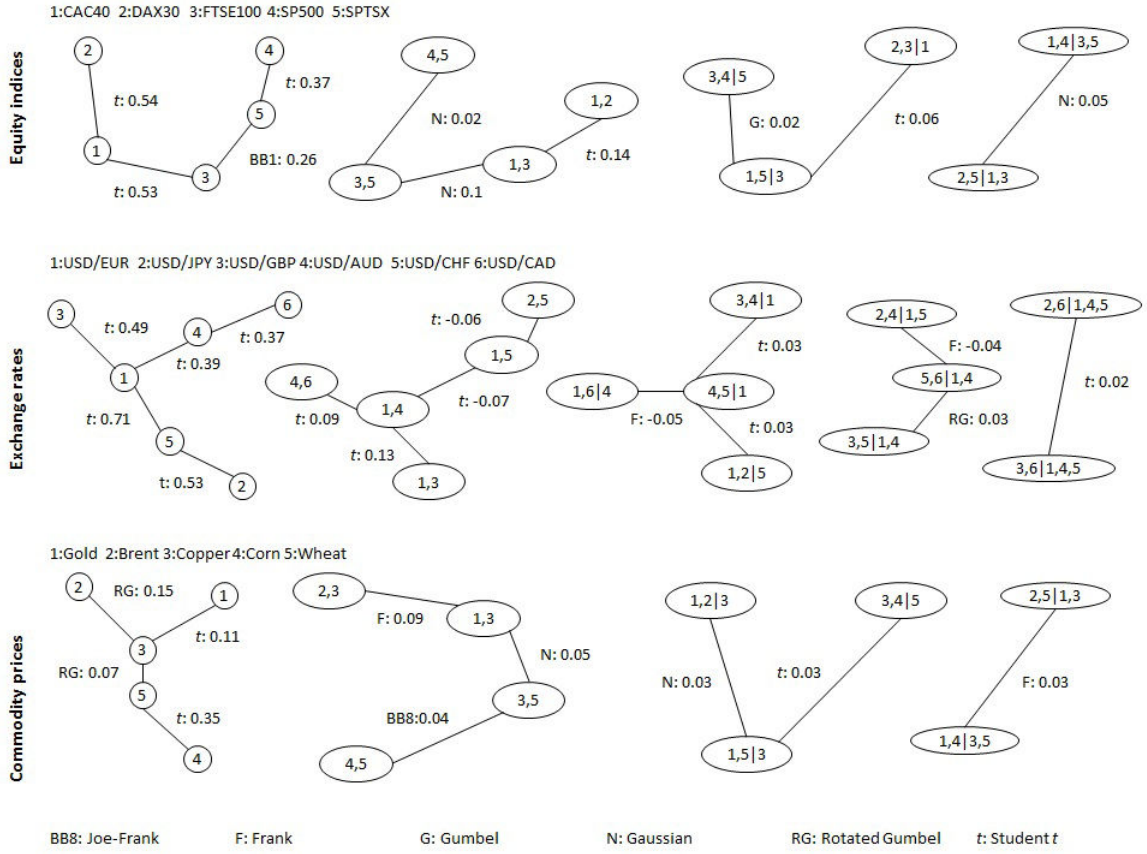

Figure 3.3: Graphical representation of the R-vine estimation results

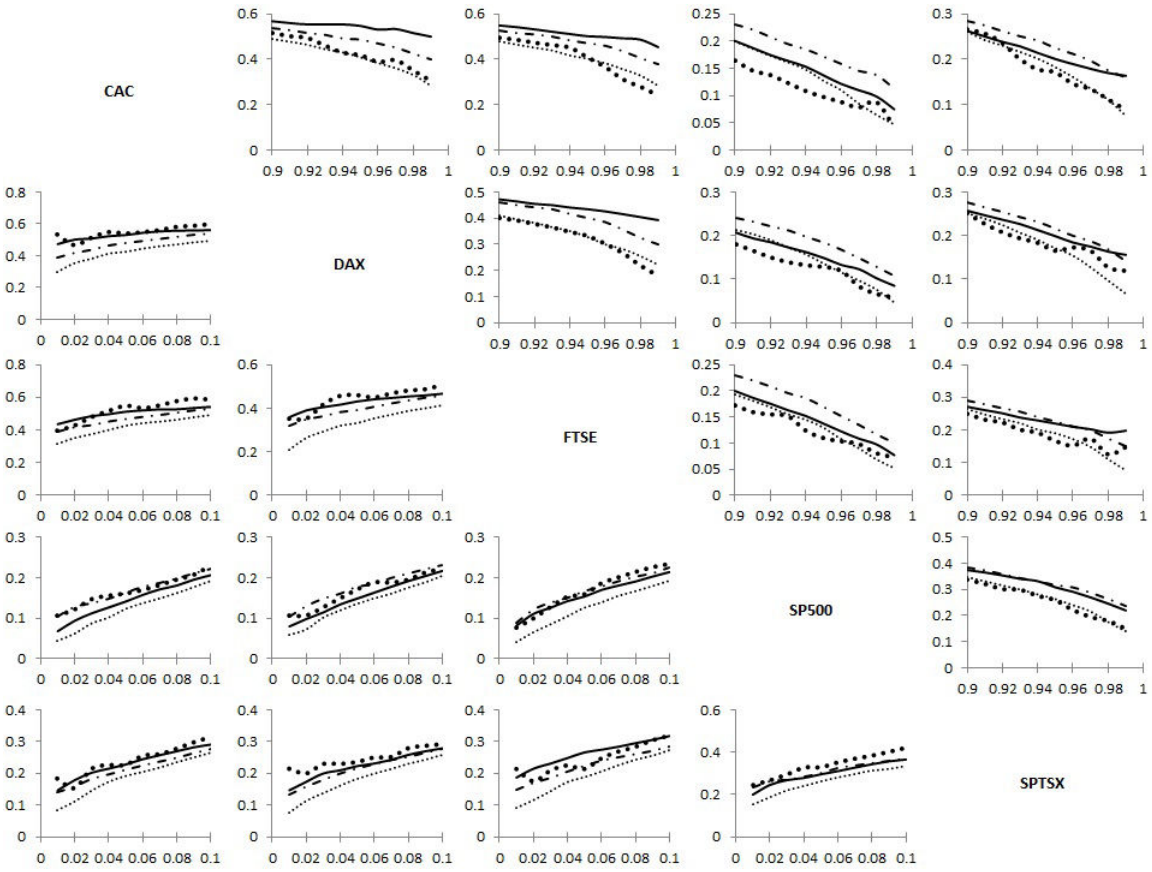

Figure 3.4: Out-of-sample performance of multivariate risk models (equity indices)

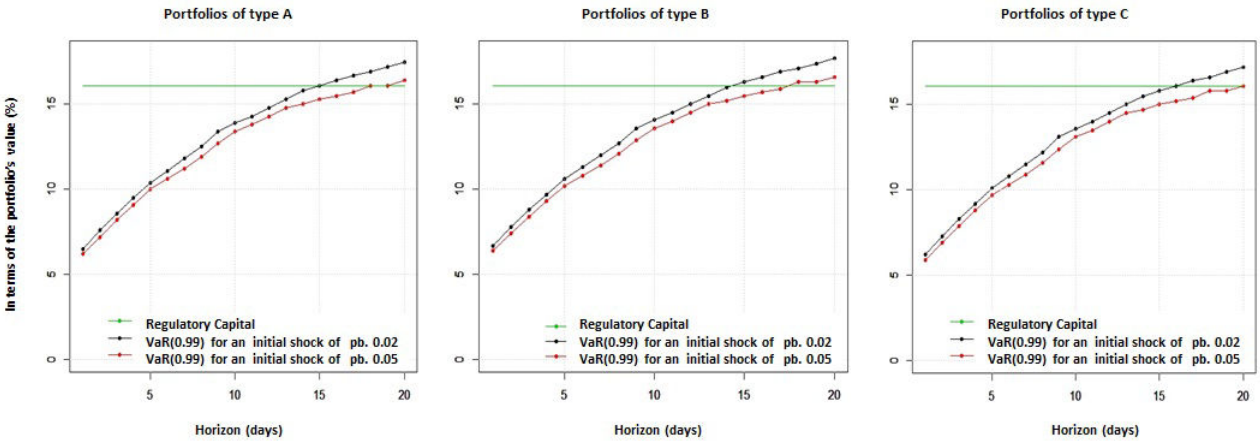

Figure 3.5: Impact of the stress scenarios in terms of regulatory capital

4. Systemic Risk Assessment
Stress Testing the French Banking System

Abstract

This chapter presents a valuation model adapted to systemic stress tests. The model allows measuring the impact of a stress scenario on a banking system within a bottom-up approach. A modular and sequential specification is used to capture the direct impact of the scenario on the banks' balance sheets, and the evolution of this impact over an arbitrary stress horizon. The direct impact is assessed on balance sheet items exposed to market, credit, and interest rate risks. A change in a regulatory solvency ratio is recorded accordingly. Given the new value of this ratio, a dynamic process is initiated, including one or more of the following: (i) individual bank reactions to the initial shock, (ii) second-round effects, (iii) the shock transmission in the banking system, through interbank networks, the financial market, and the money market contagion channels, and (iv) public response functions, represented by the central bank and the Treasury. The model is estimated and simulated quarterly for the French banking system. The results show a high vulnerability of the trading portfolio compared to other banking activities. Two out of six banks seem to be of particular vulnerability. Overall, contagion effects are relatively limited compared to first-round effects. Most losses stemming from both effects are observed in the first year of simulation. Finally, the central bank response turned to be crucial. Scenarios where this has been omitted result in a failure of a significant part of the system after only two quarters.

Key words: Bankruptcy, contagion, financial stability models, stress testing, systemic risk

JEL classification: C32, D85, G11, G17, G21, G28, G32, G33

4.1 Introduction

The subprime crisis stemming from the U.S. in the summer of 2007 suddenly converted into a financial global crisis of a serious impact on the real economic activity. These unexpected effects, both in quantitative and qualitative terms, have urged financial regulators and policy-makers towards a better understanding of the financial system functioning. These efforts are expected to allow preventing such events or, at least, to be able to manage them when they occur. Several analytical tools have been proposed to identify financial systems' vulnerabilities, to assess their expected impact on financial systems and the real activity, and to identify appropriate methods to prevent and to manage financial crises.

Macro or systemic stress tests are part of this management set. These exercises are designed to assess the potential impact of *harmful yet plausible* scenarios on the financial and economic stability of one or more systems (e.g. countries or regions). By doing so, they are assumed to capture the transmission channels of the supposed scenario and the evolution of the corresponding impact over a given horizon. On a conceptual level, stress tests respond exactly to what regulators and policy-makers are asking for. They then have been under the guidance of the main national and international regulations. However, the excitement generated by their introduction have been moderated by the limits identified in the first performed exercises. These limits are especially reflected by: (i) failures in identifying the sources of systemic vulnerability (the systemic risk factors), (ii) drawbacks of the performed models used to capture the transmission channels of these factors to the financial system and the real economy (the systemic risk), and (iii) the lack of credibility in the considered stress scenarios. These aspects have often led to hasty conclusions and to an illusion of strength of the stressed systems. As an example, one can hold up the case of most stress tests conducted within the FSAPs programs between 2005 and 2007, i.e. just before the outbreak of the crisis. More recently, Greek and Irish banks have been liquidated a few months after having successfully passed the stress tests to which they have been submitted.

The economic, social, and political costs of these errors have reinforced the need for a better control of stress tests. This study provides an answer to the second limit introduced above. It proposes a stress testing *valuation model* that assesses the systemic risk carried by a banking system. The model is based on a modular and sequential specification of the transmission channels through which changes in risk factors impact the considered system. Alongside with a *risk model* specifying the joint process of the systemic risk factors, the valuation model measures the systemic risk, based on the direct impact of a change in risk factors (hereinafter "the shock") on individual banks' balance sheets, and the dynamics of this impact over a given time horizon. The direct impact is measured in terms of changes in the value of balance sheet items exposed to market, credit, and interest rate risks. Depending on the direct impact, the model dynamics may include one or more of the following aspects: (i) private reactions to the shock, (ii) second-round effects resulting from these reactions, (iii) contagion effects across banks of the same system, through interbank networks, the financial market, and the money market channels, and (iv) public reactions represented by the central bank and the Treasury. The model is carried out within a bottom-up approach, and applied to the balance sheets of the six major French commercial banks.

The specification of the various modules is quite simple and often extensible. Reduced-form equations and rules of thumb are mainly considered in this study. This choice allows capturing, in operational terms, the evolution of the shock's impact and the contribution of the different modules in the final result. A detailed analysis and a straightforward interpretation of the

results are also strengthened through this approach. Besides providing an assessment of the potential systemic risk, the proposed model can also be regarded to as a quantitative tool allowing for a broad assessment of the systemic stress testing exercises conducted by financial institutions and supervisors.

The remainder of this chapter is structured as follows. Section 4.2 reviews the main approaches carried out in the literature to specify the systemic risk and systemic risk factors. A particular focus will be made on stress testing models. Section 4.3 presents the general framework of the proposed model and details its different modules. The dataset under review and the estimation results are given in section 4.4. Section 4.5 puts forward the practical implications of the model through an analytical stress testing exercise. Section 4.6 summarizes and concludes.

4.2 Systemic risk modelling

Until mid-2000s, systemic risk management for financial systems have been mainly performed within micro-prudential frameworks. In contrast, macro-prudential or financial stability models have been less stylized and mostly reduced to qualitative specifications. Those aspects have significantly reduced the practical utility of these models for risk management purposes, unlike monetary policy models for example (Borio et al., 2012). The criticism against the first macro-prudential models have been centered on their inability to – accurately – assess the potential impact of systemic risk factors on the financial system, i.e. the inability to quantify the systemic risk. This has produced awful consequences in recent times. On the one side, the prudential management measures taken before the crisis have proven to be suboptimal, which produced unexpected losses and several bankruptcies. On the other, when the crisis has erupted, public and private decision-makers were often confused about the reactions to adopt. This uncertainty was mainly due to a poor understanding of the financial system, and in particular, the risk transmission channels and the impacts of private portfolio management and public policy responses to the crisis. These factors explain inefficiencies in the undertaken measures, often inappropriate in qualitative and quantitative terms, as well as in terms of timing.

More generally, a macro-prudential model is based on three components: (i) the definition of the systemic risk, and (ii) the related risk factors, and (iii) the specification of the valuation model, which captures the transmission channels through which the risk factors are translated to the system, i.e. the concretization of the risk factors. A reliable assessment of the systemic risk requires an appropriate choice of the three components. In practice, however, this task faces several theoretical and practical challenges as we shall see below.

4.2.1 Systemic risk and systemic risk factors

The financial literature does not provide a conventional definition of systemic risk, which remains an elusive notion. In some papers, this definition is quite broad and often inaccurate. Systemic risk is, for example, assimilated to: distortions caused by financial frictions with respect to a critical threshold (Borio and Drehmann, 2009), circumstances threatening the stability or the confidence of the financial system (Billio et al., 2012), financial turmoil large enough to impact economic growth and welfare (ECB, 2009), etc. A second body of literature restricts the concept of systemic risk to some of its specific aspects. These include: common exposures to exogenous risk factors (Acharya et al., 2010), stock market bubbles (Rosengren, 2010), contagion (Moussa, 2011), endogeneity (Caballero, 2010), feedback effects (Kapadia et

al., 2013), information frictions (Mishkin, 2007), the impact on the real economy (G10, 2001), etc. Yet, broad definitions are often useless for operational purposes, while narrower definitions may shrink the scope of the analysis and mix up the notions of risk and the risk factors.

To circumvent these issues, we adopt a definition of systemic risk on the basis of the underlying risk factors. Three main categories of systemic risk (factors) will be considered: (i) the external risk, (ii) the internal risk, and (iii) the risk of contagion. A similar classification is adopted by ECB (2010). The external risk stems from external risk factors to the banking system. It impacts banks and financial markets in a direct fashion (e.g. macroeconomic, financial, geopolitical factors, etc.). The internal risk is generated by an accumulation of imbalances within the banking system. These factors include, among others, common positions (e.g. on securities and collaterals, with regard to the yield curve) and informational frictions (e.g. imitation, moral hazard due to a deposit-insurance system or the presence of a lender of last resort, etc.).

The contagion risk refers to the transmission of external and internal risks across different market players in the banking system. This transmission is carried by two main channels: the interbank settlement system, and the financial and the money markets. Thus, contagion risk factors often depend on the considered channel. A contagion generated by the settlement network can be initiated by an occurrence of a counterparty risk factor (Freixas and Parigi, 1998; Allen and Gale, 2000; Freixas et al., 2000; Kahn et al., 2003; Leitner, 2005; Brusco and Castiglionesi, 2007; ECB, 2010; Duffie and Zhu, 2011; Huang and Ratnovskib, 2011). It can also stem from liquidity risk factors. In this case, the concretization of risk factors often takes the form of a bank run (Chen, 1999; Diamond and Rajan, 2005; Acharya and Yorulmazer, 2007) or an uncertainty concerning future liquidity needs (Eisenschmidt and Tapking, 2009). A contagion through the financial and the money markets is often due to fire sales (Cifuentes et al., 2005), liquidity spirals, (Brunnermeier and Pedersen, 2009), currency crises, deposit withdrawals, etc. See Upper (2011) for a detailed review.

The internal and external risk factors affect in the same way banks of similar exposures. Instead, the impact of contagion risk factors differs with respect to the banks' idiosyncratic risk factors. In this study, we measure the impact of all risk factors on the exposures presented by banks' balance sheets. Exposures to market, credit, and interest rate risks will be considered. Hence, the next challenge of a macro-prudential framework is to implement a valuation model, which specifies – in a unified framework – the relation between banks' balance sheet items and the related risk factors.

4.2.2 Systemic risk valuation models

The literature identifies three main approaches to assess systemic risk. The first, measures systemic risk within a general equilibrium framework (Goodhart et al., 2013, 2006a,b, 2004). Being deeply stylized, these models present several practical challenges, particularly related to the estimation procedure (due to a lack of parsimony) and the interpretation of the results. Accordingly, some reduced versions have been proposed. These consist in restricting the number of variables, states, and time periods used in the model, and fixing some of the parameter values (Saade et al., 2007). However, the resulting specifications often lack of accuracy, which reduces their utility for management purposes. To prevent these issues, portfolio managers and policy-makers have switched on to a more operational approach. This consists in assessing the systemic risk through a set of quantitative tools such as the so-called *Financial Soundness Indicators* (FSIs). While the latter are fairly straightforward to implement and to interpret, compared

to the former approach, they still much more incomplete. Indeed, since different tools are designed to alternative purposes, they cannot be combined in a unified framework to assess systemic risks. Moreover, their construction is often based on simple reduced-form equations, which makes them unsuitable for economic storytelling. Finally, unlike structural approaches, FSIs do not allow specifying endogenous phenomena, such as risk transmission channels, the impact of private and public responses, etc.

Systemic stress testing models fall between the former two approaches. Conceptually speaking, these models are based on a sequential specification combining three main blocks. The first, said *risk model*, defines the joint dynamics of the exogenous risk factors (see chapter 1). The second block is called valuation model and captures the potential impact of a shock in risk factors on the banks' balance sheets. This model defines the individual and the systemic risk profiles. The last block allows drawing operational risk measures from the reassessed balance sheets. Tools such as VaR or FSIs can be used for this purpose.

Among the three blocks, the first two are of particular importance. Moreover, the valuation model is fundamental in systemic stress testing. Its specification is assumed to be based on a set of complementary modules allowing to specify: the direct impact of the shock on balance sheet items, private and public response functions to the shock, the impact of these responses (second-round effects), transmission channels of the shock – before and after reaction – to the rest of the system (contagion effects) and to the real economy (feedback effects), the dynamics of the shock's impact – with and without new shocks – in medium and long terms, etc. In practice, however, most existing models neglect one or more of these modules. In most cases, this is motivated by a lack of private data, estimation issues, model's implementation and management costs, concerns about internal and external communications related to the model, its outcomes, and the intended responses, etc. Even though the main issues are well identified, the existing stress testing models are still limited, which made them one of the most active research areas in the last period.

Elsinger et al. (2006a,b) have proposed a one-period quarterly model, adopted by the Austrian Central Bank (OeNB) to test the robustness of the domestic banking system. This model captures the direct impact of exogenous risk factors on balance sheet items exposed to market, credit, interest rate, and counterparty credit risks. Contagion is captured through the financial market channel, and its effects are recorded at the end of the period. Boss et al. (2006a,b) and Boss (2008) have extended this model (labelled SRM) to twelve periods. They have also introduced a private response function, which admits profit redistribution at the end of each period. Drehmann et al. (2008) and Wong et al. (2008) have presented a model designed to evaluate balance sheet exposures to credit and interest rate risks, and tested the impact of alternative response functions on the results of each period. Alessandri et al. (2009) extended the model of Elsinger et al. (2006a) by adding an interest rate risk valuation module as in Drehmann et al. (2008), a second contagion channel represented by counterparty credit risk, and a response function consisting in reinvesting the profits at the end of each period. This model, labelled RAMSI, is officially adopted by the Bank of England for its stress testing exercises. Aikman et al. (2009) have extended most of the RAMSI's modules and introduced a third contagion channel through funding liquidity – allowing to capture the liquidity spiral in a simplified fashion (see. Brunnermeier and Pedersen, 2009; Gauthier et al., 2012). For the Bank of the Netherlands (DNB), van den End (2010, 2012) have presented a stress testing model mainly focused on the liquidity risk. Compared to the previous models, this also includes an idiosyncratic reputation risk of individual banks and a response function from the central bank.

To date, the SRM and the RAMSI models are still the most advanced specifications in

practice. However, both of them are still a work-in-progress. The main remaining issues concern arbitrary selections of risk factors, an omission of crucial risk classes (e.g. off-balance sheet exposures, feedback effects on the real economy), and a simplicity in the specification of the different modules and the whole dynamics of the valuation model – often assumed as linear. These aspects show at what extent stress testing valuation models are still underexploited. This partly explains the recurrent active and prudential risk management errors recorded in the recent period, as well as the expressed cautions with respect to these exercises.

Our study is related to the existing literature, and especially to Alessandri et al. (2009). We extend this paper on several aspects. First, by introducing a new first-round class: an idiosyncratic reputation cost for a bank having suffered a loss after the direct impact. Second, we have explicitly considered public policy response functions, from the central bank and the Treasury. Third, a new contagion algorithm is introduced, including three interactive channels represented by the financial market, the interbank settlement network, and the money market. Four, the proposed model only relies on public information. This allows for a widespread implementation by financial market players facing a restriction to access private data. Finally, to the best of our knowledge, this study is the first to perform a bottom-up systemic stress testing analysis for the French banking system since the 2007-09 crisis.

4.3 The model

We model the systemic risk carried by a banking system within a dynamic bottom-up approach. This is based on a modular specification of the risk factors and their impact on the considered system. The model is estimated quarterly. Its flexibility allows using the model in simulation exercises for an arbitrary horizon. This will be illustrated through a hypothetical stress testing scenario. This section presents the general framework of the model and develops its different modules. The following sections report the estimation and the stress testing results, respectively.

4.3.1 Overview

For a typical period of one quarter, Figure 4.1 summarizes the role of each module and the sequence of events in the model. A risk model (4.3.2) defines the values of the (exogenous) risk factors at the start of the period. Given these values, four direct effects on individual banks' balance sheets are modelled: gains and losses on the trading portfolio, credit losses, net interest income, and reputation costs (4.3.3). After recording these direct effects, the profitability of each bank is assessed. Two situations can arise: all banks are still profitable, or at least one bank makes losses. In the first case, the balance sheets are "readjusted" according to four rules of thumb known to reflect a common behaviour of commercial banks in practice (4.3.4). This readjustment marks the end of the period and the start of the next period.

If one or more banks make a loss, their solvency is assessed. Regulatory ratios defined by the Basel III standards are used for this purpose. Two situations can arise: the solvency ratio is still above the regulatory minimum for all non-profitable banks, or this threshold is violated for at least one bank. In the first case, a loss of reputation cost is suffered by non-profitable banks. This cost is manifested by tightening their access to refinancing, represented by an increase in the interest rate on new loans. This cost will be recorded in the net interest income of the next period. This marks the end of the current period and a move to the next one, after

readjusting the balance sheets of profitable banks and "rebalancing" those of non-profitable banks to account for losses. These two actions generate second-round effects on profitable and non-profitable banks.

When a non-profitable bank records a solvency ratio under the regulatory threshold, it is assumed to default. It then follows a complex bankruptcy process defined in different stages (4.3.5). First, the assets of the bank lose part of their face value, corresponding to the so-called *bankruptcy costs*. To fulfil the bank's commitments, the remaining assets are sold in the financial markets. This action affects the remaining banks trough three complementary contagion channels. First, massive asset sales create a supply shock in the relevant market, leading to a price decrease and a devaluation of marked-to-market assets. The trading book of all banks is then impacted through the market liquidity channel. Second, the revenue of asset sales from the defaulting bank(s) may be insufficient to pay their full debt. In this case, creditor banks will suffer an additional loss related to their counterparty credit risk. This is the interbank network contagion channel. Finally, the failure of one or more banks undermines the market confidence towards the entire banking system. The remaining banks face tighter refinancing conditions making that their future borrowings are augmented by an additional risk premium. This is the funding liquidity channel. The sum of all contagion effects on the remaining banks is recorded. Their profitability and solvency are reassessed again, as explained above. If more failures occur, the procedure is repeated causing further effects on the remaining banks. This procedure continues until no new failure is recorded or – less likely – until the whole system goes bankrupt.

When the contagion chain ends, and before rebalancing the balance sheets of the remaining banks, a last module allows introducing public responses. Indeed, to make more realistic the model structure, we have considered two reaction functions each for the central bank and the public Treasury (4.3.6). The central bank's responses consist in restoring the asset prices (in the financial market) and the interbank interest rate (in the money market) to their start-of-the-period levels. For the remaining banks, this allows absorbing the losses resulting from the first and the last contagion channels. The Treasury's reactions consist in reducing the bankruptcy costs stemming from a bank failure, and a participation in the capital increasing of the remaining non-profitable banks. The period is ended by rebalancing the balance sheets to account for the gains and losses of the entire period

4.3.2 The risk model

In the previous chapter, we have introduced a multivariate risk model adapted to daily financial time series. This has been used to design individual stress testing scenarios. However, given the depth and the frequency of the data included in the current chapter, the application of such a framework could be a source of parsimony issues. Accordingly, a less sophisticated Bayesian VAR (BVAR) model is carried out here.

The $\{r_t\}_{t=1}^{T}$ process of the $N \times 1$ vector of exogenous risk factors is defined by a vector autoregressive model of order p, denoted VAR(p), and given by

$$A(L)r_t = \mu + \varepsilon_t \tag{4.1}$$

where $A(L)$ is a polynomial lag matrix, μ is an unconditional mean-vector, $\varepsilon_t \sim \mathcal{N}_N(0, \Sigma)$ a normally distributed standardized residuals, and $\mathcal{N}_N$ is an N-dimensional normal distribution and Σ is the related covariance matrix.

Given the parameter estimates, this model will be used to simulate the future paths of exogenous risk factors. These simulations can be carried out with (stress testing exercises) or without (forecasting exercises) an initial shock on risk factors. At each period of the simulation horizon, the impact of the simulated values of risk factors on the individual banks' balance sheets is assessed as follows.

4.3.3 First-round effects

The direct impact of risk factors is captured using a set of *balance sheet models*. Unlike the so-called *asset pricing models*, this approach allows taking into account a wide range of risks and provides a more detailed analysis of the transmission channels. Moreover, balance sheet models offer a clear understanding of the structure and the results of each sub-model. This makes it an attractive tool for supervisors in their systemic risk and financial stability analyses. In most cases however, this use is limited to credit risk models (Alessandri et al., 2009).

In this section, we consider four direct effects of the exogenous risk factors on individual banks' balance sheets: (i) trading book gains and losses, (ii) credit losses, (iii) net interest income, and (iv) reputation costs. To capture these effects, balance sheets have been split into fourteen asset classes: seven classes for each of the asset and the liability sides (see Table 4.1). The first class of each side captures gains and losses of the trading portfolio. The second class captures interbank exposures. The following four asset classes are exposed to credit risk, carried by households, administrations, large non-financial companies, and other financial institutions, respectively. The remaining classes are only exposed to interest rate risk. To measure the net interest income, we split non-interbank credits (classes 3 to 6) and liabilities (classes 3 to 6) into four buckets depending on their maturity: from zero to three month, from three to twelve months, from one to five years, and more than five years. The other classes are aggregated and analysed separately. The second and third columns of Table 4.1 summarize the considered methods to assess first-round effects. These are detailed below.

The trading book

Detailed data on the composition of the banks' trading portfolios are confidential. To estimate this first-round impact, we have then considered aggregated amounts for the trading assets and liabilities hold by each bank. Given their specificity, derivatives have been excluded from this model. With this respect, we have assumed that derivatives allows hedging a loss which is equivalent to their face value.[1] This allows introducing more flexibility in the model, as any additional private information can further be included to update this assumption. Anyway, the impact of this treatment on the result is fairly limited, given the amount of derivatives compared to the size of the trading portfolio.

Our model assumes that the value of trading assets evolves proportionally to stock market prices and to exchange rate, and inversely to interest rates and commodity prices. By doing so, we consider the main sources of market risk in the banking system. To prevent estimation inconsistencies stemming from highly volatile stock and commodity prices, these two risk factors have been demeaned with respect to their respective historical averages. For a given period, the trading book's return, given by the change in the net trading value (trading assets – trading

[1] Alessandri et al. (2009) include derivatives in the market portfolio, which overestimates the banks' exposure to market risk.

liabilities), is measured as follows

$$\Delta \text{NTV}_t = \alpha_0 + \alpha_1 \Delta \text{STOCK}_t + \alpha_2 \Delta \text{COM}_t + \alpha_3 \Delta \text{SR}_t + \alpha_4 \Delta \text{LR}_t + \alpha_5 \Delta \text{EX}_t + \varepsilon_t \tag{4.2}$$

where ΔSTOCK, ΔCOM, and ΔEX are quarterly log-differences for the stock market, the commodity market, and the exchange rate, respectively; ΔSR and ΔLR are quarterly differences for the short-term and the long-term interest rates, respectively; and $\varepsilon_t \sim \mathcal{N}(0,1)$ is a normal white noise.

Credit losses

The expected losses associated with the loans granted by the banks to the economy is calculated by multiplying the three following terms: the probability of default (PD) and the loss given default (LGD) of the borrower, and the exposition at default (EAD) of the loan. Apart for credit institutions, analysed separately in section 4.3.5, we distinguish four main categories of banks' borrowers: households, administrations, large non-financial companies, and other financial institutions. The more the bank has an information about its borrowers, the better it can estimate and manage the related credit risk. Yet, this information is still confidential, and sometimes even inexistent. Here, we only rely on public information and some hypotheses to estimate the impact of risk factors in terms of credit losses.

For simplicity, the default probability of administrations is assumed to be null. Given the values of the exogenous risk factors, the probability of default of the other categories of borrowers is estimated using a logit model. Since the PDs are not observable in practice, we have considered the historical default rates of each category as a proxy. Considering the real economic growth, equity and real estate indices, as well as the real interest rate as risk factors, the default rates are given by

$$\text{L}_t = \ln\left(\frac{\text{DR}_t}{1-\text{DR}_t}\right) = \beta_0 + \beta_1 \text{GDP}_t + \beta_2 \text{STOCK}_t + \beta_3 \text{REPI}_t + \beta_4 \text{RSR}_t + \eta_t \tag{4.3}$$

where L is the log-odds transformation of quarterly default rates DR; GDP, STOCK, REPI, and RSR are quarterly real GDP, stock and real estate indices, and real short-term interest rate, respectively; and $\eta \sim \mathcal{N}(0,1)$ is a normal white noise.

Hence, the expected PD for date $t+h$ (with $h>0$) is given by the following formula

$$\mathbb{E}(\text{PD}_{t+h}) = \mathbb{E}\left(\frac{e^{\text{L}_{t+h}}}{1+e^{\text{L}_{t+h}}}\right) \tag{4.4}$$

As is usual in the literature, LGDs are assumed to be constant over the simulation horizon. Following Alessandri et al. (2009), these are fixed to 50% for households, 80% for non-financial companies, and 10% for other financial institutions. Finally, in order to define the EADs amounts of each category, the last observed balance sheet decomposition for each bank has been replicated. This decomposition is carried forward for the entire simulation horizon (see section 4.5).

Credit losses of a given period are calculated by multiplying the estimated PDs of the same period, the fixed LGDs, and the EADs of the end of the previous period. By combining these

credit losses and the other credit losses due to contagion effects (section 4.3.5), the total amount of credits allowed by each bank at the end of the period is defined. This total is allocated as in the last observed balance sheet. This allows defining the EADs to be used in the next period. These EADs are multiplied by the LGDs and the estimated PDs to obtain the corresponding credit losses, and so on.

Net interest income

In order to estimate the net interest income, non-interbank loans and liabilities have been split into four maturity buckets: zero to three months, three to twelve months, one to five years, and more than five years. The remaining classes are treated separately, experiencing interest income (for assets) and charge (for liabilities) according to an ad hoc rate defined in Table 4.1.

For both the assets and the liabilities, we have considered the risk-neutral valuation model proposed by Drehmann et al. (2008) to determine the amount of paid (for debts) and received (for credits) coupons at each period.

Let A be an asset issued at date 0 with maturity at date T. The value of the coupon C paid by this asset for each period t (with $0 < t \leq T$) is defined at time 0. A "fair" value of C ensures an equality between the face value A_0 and the economic value V of asset A. More formally

$$A_0 = V = \sum_{t=1}^{T} D_t C A_0 + D_T A_0 \tag{4.5}$$

The coupon C is then obtained by

$$C = (1 - D_T) / \sum_{t=1}^{T} D_T \tag{4.6}$$

where D_t is a discount factor defined by

$$D_t = \prod_{l=1}^{t} (1 + R_{l-1,l})^{-1} \tag{4.7}$$

with

$$R_{l-1,l} = \frac{r_{l-1,l} + \mathrm{PD}_{l-1,l} \ \mathrm{LGD}}{1 - \mathrm{PD}_{l-1,l} \ \mathrm{LGD}} \tag{4.8}$$

where $r_{l-1,l}$ and $\mathrm{PD}_{l-1,l}$ denote, respectively, the expected values for the risk-free rate and the probability of default between dates $l-1$ and l. The yield curve is constructed using a linear interpolation of short-term and long-term interest rates. This curve provides the risk-free rate at each period of the simulation horizon. Default probabilities are obtained by eqs. (4.3) and (4.4) as described in the previous subsection.

The obtained coupon takes into account the interest rate risk and the credit risk associated with asset A. Mismatch maturities between the asset and liability sides of a balance sheet, as well as the probability of unanticipated shocks on the interest rates and/or the probabilities of

default are the main sources of interest rate risk. This is especially true for assets of a (very) long maturity. Indeed, the fundamental value of these assets may worsen significantly depending on changes in interest rates and/or the PDs, while their book value remains unchanged – since the coupon is set at the beginning of the period and holds until maturity.

4.3.4 Private response functions

Quite recently, the French supervision authority (ACP, 2013) argued that:

- Assuming institutional responses in stress testing exercises presents two opposing aspects. Namely, result comparability between the participating banks and the realism of the exercise.
- Assuming static balance sheets over the stress horizon guarantees a standardization and a comparability of results.
- Assuming an inability to respond to a stress shock may be the worst possible scenario.

The first viewpoint emphasizes the importance of private response functions for more realistic stress tests. However, the last two arguments prevent using this feature in most realized exercises, including those performed by the ACPR. The third argument is particularly questionable. Indeed, a stress testing exercise does not consist only in designing the worst possible scenarios, as a plausibility condition is also required. Yet a dynamic scenario includes an initial shock and an after-shock. Hence, assuming unrealistic after-shocks (e.g. no bank reaction over the entire stress horizon) affects the plausibility of the scenario, even if the assumed initial shock is plausible. This may lead to misleading results and to wrong active and prudential management decisions.

In our valuation model, we have explicitly specified commercial banks' private response functions. However, a lack of regular and filtered information on this topic prevents using econometric models as done for the first-round effects. Instead, we define private responses using a coherent set of the so-called *rules of thumb*. These have been proven by several empirical studies to reflect the common behaviour of commercial banks (see, e.g, Adrian and Shin, 2008; Alessandri et al., 2009). The nature of these reactions is defined by the situation of each bank after recording the first-round effects (i.e. profitable, non-profitable, or failure). Therefore, before presenting the reaction functions considered in this study, we will discuss the different potential outcomes after recording first-round effects.

Default rules

For each bank, the combination of eqs. (4.2) to (4.8) gives the first-round effects of the period. This is achieved by summing the values of the net trading value and the net interest income, minus the credit losses. If a bank has recorded a positive result, it is said to be "profitable". To reinvest the realized profit, the bank makes a readjustment of its balance sheet (section 4.3.4). If the result of a bank is negative, we use a solvability criterion to determine whether it is or not in default. The Common Equity Tier 1 (CET1) regulatory capital ratio is used in this respect. This is obtained by dividing equities by risk-weighted assets based on the Basel III standard approach (see Table 4.1). The choice of this ratio has been imposed by the high level

of aggregation in the available balance sheet data. Nevertheless, the flexibility of the method allows replacing, modifying, and/or adapting the default criterion depending on the detailed degree of the available information. Other regulatory ratios may therefore be considered.

Basel III standards require a minimum threshold value of 4.5% for the CET1 ratio. We rely on this value to distinguish: "non-profitable" banks, recording a negative result but meeting the regulatory criterion, and (ii) "failing" banks, recording a negative result and a solvency ratio below the regulatory threshold. After first-round effects, non-profitable banks rebalance their balance sheets to account for the losses (section 4.3.4). Failed banks liquidate their remaining assets to deal with their creditors.

Before being placed on the market, these assets support a discount in their face value. This is the so-called *bankruptcy cost*, widely documented in the literature (see, e.g., James, 1991; Bris et al., 2006). It corresponds, among others, to legal costs related to the liquidation process, a loss in human capital, and a loss of reputation. In practice, however, it is quite tricky to estimate the discount rate. Indeed, in addition to the cited determinants, one has also to consider aspects related the tax system and other structural constraints. It is then commonly admited to set the bankruptcy cost to 10% of the liquidated assets (Alessandri et al., 2009). In this study, we consider this parameter among the public response functions of the Treasury. We can therefore study the role of the Treasury's action by varying the value of the discount rate.

The failure of one or more banks can lead to those of the remaining banks, through the contagion channels (section 4.3.5). For more consistency in the used notation, we call "sound banks" those banks having survived to first-round effects, and "rescued banks" those who survived to contagion effects. Furthermore, we call "fundamental default" a default due to first-round effects, and "contagion default" a default produced by contagion effects. Before developing these processes, we end this section with a detailed presentation of the different private response functions to be considered after first-round effects.

Private mitigating response functions

After recording the first-round effects of the period, profitable banks "readjust" their balance sheets. To do so, they reinvest the generated positive income of the period targeting the following objectives:

(i) Maintaining the leverage ratio: in response to their capital increase, profitable banks subscribe to new debts in order to restore a predefined equity-to-debt ratio.

(ii) Maintaining the solvency ratio: profitable banks invest their new cash flows – caused by the positive result of the period and the new debts – aiming to restore a predefined CET1 ratio.

(iii) Maintaining the business model: to meet the two precedent criteria, debt increase and reinvestment are made in the same proportions as in a balance sheet of reference.

The predefined leverage and solvency ratios, as well as the balance sheet of reference are in line with the last observed balance sheet. That is the balance sheet used for the first period of simulation.

Non-profitable banks do not make these adjustments following first-round effects. Indeed, it is more intuitive to assume that they are more concerned about reacting to the suffered losses

than engaging a long-term targeting of leverage and solvency ratios. Given the difficulty of raising new equities in a too limited period of time, as well as the difficulty of liquidating assets at favourable prices given their situation, these banks will only subscribe to new debts at the extent of the recorded losses. In addition to the generated losses, the suffered first-round effects also cause a reputation cost for these banks (referred to as "reputation 1" in Table 4.1). We specify this cost by a degradation in the refinancing conditions, represented by a risk premium applied to all new debts contracted by non-profitable banks. The payment of this premium is made in the following period, hence impacting the corresponding net interest income.

After contracting new debts in response to the first-round effects, non-profitable banks "rebalance" their balance sheets. In this case, the first-round effects are reported for the following periods. That is, non-profitable banks accumulate losses across periods and see their situation even trickier, with serious risks of going bankrupt in case of further losses. Concretely, a readjusted balance sheet is a positive multiple of the balance sheet of reference, where the multiplier is the ratio of the profit in the current period to the size of the balance sheet in the previous period. Instead, a rebalanced balance sheet keeps the same size, with a different structure (marked by worst leverage and solvency ratios). The failure of one or more banks in the system impacts – at least through the market liquidity channel – the readjusted and the rebalanced balance sheets of the remaining banks. After recording contagion effects, rescued banks rebalance their balance sheets for the next simulation period. We assume that, given the short period of time allocated and the market stress generated by these effects, even profitable banks become unable to readjust their balance sheets. They shall, in this case, simply rebalance their balance sheets in the same way as non-profitable banks. Thus, when a failure occurs during a period, the related results are reported to the following period. After the first-round effects of the latter, profitable banks will again be able to readjust their balance sheets, which means a better visibility of their own, one quarter later. This assumption seems quite reasonable, given the recent developments in financial markets.

After recording the first-round effects and defining private responses, the simulation ends for the considered period if no default is recorded. If the first-round effects have led to one or more failures, these will create contagion effects. The latter can lead to further defaults, which may, in turn, generate other defaults, and so on. Whether it comes from first-round or contagion effects, we assume that a failure is transmitted to the rest of the system through three complementary channels, described in the next section.

4.3.5 Contagion effects

The transmission of a bank failure to the rest of the system can operate through three basic channels: the financial market, the interbank network, and the money market. These channels impact, respectively, the trading and the credit portfolios, and the net interest income. The first and the last channels are commonly referred to as *market liquidity* and *funding liquidity* channels, respectively. We first present the channels separately, before presenting a unified framework taking into account the dynamic interaction of their respective effects. To the best of our knowledge, this study is the first to introduce such a level of flexibility.

The financial market channel

After recording the first-round effects and applied the bankruptcy cost, the remaining trading assets of failed banks are sold in the financial market. Given the level of aggregation in the

available data on these assets, a one-asset/one-market model in considered here. This refers to the financial market in general terms. A massive sale of assets generates a supply shock in the market, leading to a price decrease. The extent of this decrease depends on the amount of sold assets. We have chosen to specify this relationship as in Cifuentes et al. (2005).

Let p be the financial market price and x the ratio of sold assets to the size of the market. The relation between these two variables is given by

$$p = e^{-\theta x}, \quad x > 0 \tag{4.9}$$

where $\theta > 0$ is the price elasticity with regard to the amount been traded. The price of reference $\bar{p} = 1$ prevails when no default is recorded (i.e. when $x = 0$).

A decrease in asset prices devaluates the trading book of sound banks, the value of which is indexed on the financial market price or marked-to-market. This is likely to worsen the solvency ratio of sound banks and lead to new failures. This is the first contagion effect.

The interbank network

Using the income from the sale of its trading book and credit portfolio, a failed bank repays its creditors, among which are sound banks. If the total income is superior to or equal the banks' total liabilities, the bank's creditors – excluding shareholders – are repaid at par, and the difference – if there is – is paid to shareholders. However, if the total income is less than total liabilities, the bank's creditors suffer a total loss equal to the amount of lost assets. This loss is assigned to creditors according to their respective seniority levels. Senior debts are paid off first. If the balance remains positive, it will be used to pay off the claims of creditors of lower priority, and so on, down to the least rated creditor. Unfortunately, an accurate information on the classification of banks' loans is not available. In such a situation, a one-class seniority is assumed here. The total of lost assets is then distributed over the creditors in proportion to the amount of their respective contributions.

A loss suffered by a sound bank in its interbank credit portfolio reduces its results and worsens its solvency ratio. This can generate a failure of this bank and further losses – or even failures – for its creditors. This is the second contagion effect.

The money market channel

The failure of one or more banks and the activation of the two previous channels create a climate of uncertainty, distrust, and a loss of market confidence towards the entire banking system (referred to as "reputation 2" in Table 4.1). This situation results in a degradation of sound banks' refinancing conditions: higher borrowing rates, limitation and denied access to credit lines, insurance costs, bank runs, etc. This multitude of reactions form the third contagion effect. In this study, we only focus on the rise in the money market interest rates, as it is the most observed in practice. This upward movement is applied to all new loans contracted after the first bank's failure is recorded. For simplicity, a same move of interest rate is assumed to hold for all banks, regardless of their respective conditions and the number of failures. This assumption seems fairly reasonable, given the market confusion in such circumstances. Indeed, there is often a limited discernibility about the effective proportion of the banking system actually affected by these failures.

Unlike the previous two effects, the impact of contagion effects through the money market is recorded in the following period. Hence, the charges related to a debt issued in a given period appear on the balance sheet of the following period. If new failures appear in the latter, the raised interest rate is maintained. Otherwise, it returns to its initial level. In other words, a one-quarter short memory is assumed to hold for the money market players, given the assumed central bank intervention at the end of each period (section 4.3.6).

The iterative process

This section presents an iterative algorithm through which, during the same period, bank failures are transmitted to the rest of the system. At each iteration, a dynamic interaction of the three previous contagion channels are considered.

The joint impact of the three channels is assessed on sound banks. If no further failure is recorded, the algorithm terminates at the first iteration. If one or more failures appear, a bankruptcy cost is applied to the bank(s) in question. This further reduces the value of their assets. The remaining trading assets are put on the financial market. The amount is added to that of assets put on the market following the first-round effects. The combined effect of these two supply shocks produces a decrease in the financial market price. This effect is measured by eq. (4.9). All trading portfolios are then reevaluated on the basis of this new price. The next step is to determine the ability of each sound bank to honour its commitments but shareholders. That is, to compare, for each bank, total assets and total liabilities. This task cannot be done analytically, since the payment of one bank defines – in part – the assets of the remaining banks (through interbank assets and liabilities) and vice versa. Hence, the inability of a bank to repay its loans reduces the revenue of its creditors. For some of them, the new total assets can become insufficient to repay liabilities. This generates losses for its creditors, which include the first failing bank. The latter therefore undergoes a further loss that affects the rest of the system, and so on.

The resolution of this iterative process is carried out using the so-called *clearing models*. The algorithms presented by Eisenberg and Noe (2001) and Furfine (2003) are the most popular in practice. They are used to define the vector of interbank payments and the number of failures in each iteration, allowing in particular to distinguish between fundamental and contagion failures. To take into account the flexibility introduced by our model framework, we have introduced two main extensions to the Eisenberg and Noe (2001) algorithm. These concern: (i) the default rule defined in section 4.3.4 and (ii) the interaction between the three contagion channels.

After having determined the interbank payments in the second iteration, the situation of each bank is assessed. If no further failures is recorded, the balance sheets are closed for the period after being rebalanced. If new failures occur, the procedure is repeated. The algorithm terminates, at the latest, after n iterations, where n is the number of commercial banks in the system.

When the chain of contagion ends, and to increase the realism of the model, a series of public response functions have been introduced over the different parts of the model. These functions allow mitigating the different effects of the shock on the banking system.

4.3.6 Public response functions

Even more than private response functions, public reactions are neglected by most existing models. Yet, their introduction allows increasing the plausibility of the scenarios, detailing the simulation results, and ensuring a better – public and private – understanding of the financial system. This is to considerably improve active and prudential risk management decisions. Accordingly, our model framework includes a series of actions, taken by the central bank and the public Treasury. These measures, frequently observed during the recent years, are defined as follows:

(i) When a fundamental failure is recorded, the central bank lightens the default rule described in section 4.3.4. Thus, to avoid – or at least to limit – contagion failures, a bank is declared bankrupt only if the amount of its liabilities exceeds that of its assets.

(ii) When a failure is recorded, the Treasury supports part of the bankruptcy cost.

(iii) For non-profitable banks, the Treasury participates in the capital increase after first-round and contagion effects. This allows banks overcoming the rise in the refinancing costs generated by the negative impact of first-round and contagion effects on the individual and the systemic reputations.

(iv) For non-profitable banks suffering a loss of reputation, a bilateral relationship allowing them to access central bank refinancing is established. This public response holds during periods when a failure is recorded. It is seen to as a complement of the previous Treasury response.

(v) At the end of each period, the central bank, through open market operations and the interest rate tool resets the financial market and the money market prices to their respective start-of-the-period levels. The location of this module is motivated by two main reasons. On one side, given the model frequency and the delay with which the central bank responds with respect to a shock occurrence, considering an intervention before recording the first contagion effects may seem implausible. On the other side, designing a reaction at the end of the period allows avoiding potential moral hazard biases due to an announced presence of a lender of last resort.

To assess the importance of these functions, their separate and joint impacts on the results of each period can be analysed by varying the value of the corresponding variable or parameter in the model's specification.

4.4 Data and estimation results

The estimation of the model is based on two datasets: macro-financial data to estimate the risk model, and balance sheet data to estimate the valuation model. The modules are assessed separately on a quarterly basis.

Eight indicators are used to characterize the exogenous risk factors in the French banking system: real gross domestic product (GDP), the consumer price index – excluding energy (CPI), a real estate price index (REPI), the French CAC 40 index (STOCK), the interbank overnight rate (SR), the rate of 30 years government bonds (LR), the nominal exchange rate

euro/U.S. dollars (EX) and the Brent crude oil price (COM). Data are collected at the end of each quarter, over the period 1992:Q1-2013:Q4 (88 observations). The series of GDP, CPI and REPI, obtained from the INSEE website, are adjusted for seasonal variations.[2] Exchange rate series are from the Bank of England website. The rest of the series are from Datastream. A summary description and prior data transformations are presented in Table 4.2.

These risk factors are used to obtain the yield curve and the PDs of the considered exposures (i.e. for households, administrations, large non-financial companies, and other financial institutions) in the French banking system. For simplicity, and in line with the existing literature, we have approximated the yield curve by linear interpolation of short-term and long-term interest rates. These rates are obtained by simulating the risk model for a given time horizon.

Balance sheet data consist of the balance sheet items of the six major French banking groups, with a net banking income superior to 10 billion euros at the end of 2012. The following groups are considered: BNP Paribas (BNPP), Société Générale (SG), Crédit Agricole (GCA), Banque Populaires Caisses d'Epargne (BPCE), Crédit Mutuel-CIC (GCM), and La Banque Postale (LBP). The total net banking income of the six groups accounts for 94% of the total French banking assets. Five of the six groups come from a series of mergers and acquisitions, carried out between 1990 and 2008 in order to consolidate the French banks' positions in a European and international environment. To circumvent the complexity of these arrangements and the lack of – detailed – data for the former entities, we have estimated some modules based on the data of the SG group.[3] All balance sheet data are obtained from quarterly financial communications published by the different banks. The estimated model is used to assess the impact of a hypothetical stress testing scenario. The scenario is applied on all banks on the basis of their 2013:Q4 balance sheets.[4] The latter are considered as the balance sheet of reference.

Equation (4.1) is estimated for the period 1992:Q4-2013:Q4. Based on conventional information criteria, the optimal lag order of the VAR model turned to be $p = 1$. Equation (4.2) is estimated for the period 2010:Q4-2013:Q4, for an aggregate portfolio composed by BNPP, SG and GCA assets – representing 70% of total assets of the considered banking system. Equation (4.3) is estimated over the period 2006:Q4-2013:Q4 for non-financial companies. In France, it is, indeed, the only category of economic agents for which data on default rates are public. The related data are available on the Banque de France website. Although the hypothesis may seem too strong, we have used estimates of PDs obtained for non-financial firms for the other two categories of borrowers. The results must then be treated with caution. Regarding the contagion channels, we have set to 0.8 the value of parameter θ in eq. (4.9). This is a commonly used assumption in the literature. Besides, an increase of 100 basis points in the borrowing interest rates is assumed for both the idiosyncratic and the systemic reputation costs (i.e. for reputation 1 and reputation 2 in Table 4.1). The estimation results of the different sub-models are summarized in Table 5.1. Due to space limitation and given the main focus of the current study, these are not commented here. Given the lack of data faced in the different stages of the estimation, there is necessarily room for improvement in this area.

[2] GDP is also working days adjusted.
[3] This is the only bank not to have suffered significant transformations in the considered period.
[4] Except for GCM, for which the 2013:Q2 balance sheet is considered, as the 2013:Q4 results were not published at the time of writing.

4.5 Simulation and stress testing results

This section presents the main results of a hypothetical stress scenario. The reported outcomes are in no case exhaustive. These are selected due to space limitation, while further results are still available from the author upon request. We also provide a Matlab code with functions to estimate the model, to carry out stochastic and deterministic scenarios, and to retrieve the results. This code is fairly straightforward and flexible. Moreover, alternative functions can be introduced to account for alternative data (e.g. private or more detailed information), to adjust the specification of the model, and to extract other outputs. The estimation procedure is fairly quick at the extent that they can easily be carried out on a daily basis. This feature is important to allow tacking account of any news to update the data and/or the model's equations.

We have carried out the model to assess the impact of a non-stochastic stress scenario on the French banking system. Even though the conception of the model also allows considering stochastic scenarios, we limit ourselves to non-stochastic scenarios, given the purpose of this study focusing on the impact rather than the conception of stress scenarios.[5] The scenario consists of a set of ad hoc changes in risk factors' values with respect to 2013-Q4 data. The considered shock is inspired by the systemic stress testing exercises commonly led by supervisors (see, e.g., CEBS, 2009, 2010; FRB, 2009, 2011; Alessandri et al., 2009). Here, we consider: (i) a fall of 1% in the French GDP, (ii) an increase of 10% in the REPI index, (iii) a fall of 15% in the real value of the CAC 40 index, (iv) an increase of 50 basis points of the interbank lending rate, (v) an increase of 30% in the PDs of households and non-financial companies, and (vi) a dry-up of market liquidity specified by setting to 0.9 – instead of 0.8 – the parameter θ in eq. (4.9). The plausibility of this scenario can be justified by an observation of similar events during the 2007-09 financial crisis or during the sample period used in this study.

The shock is applied at 2013-Q4, corresponding to date 0. The dynamics of the shock and its impact on the banking balance sheets is assessed on a quarterly basis over a three-year simulation horizon. Hence, the stress horizon corresponds to 2014-Q1 to 2016-Q4, i.e. twelve periods noted Q1 to Q12. The dynamics of the initial shock over the following periods is obtained using 10^6 Bootstrap simulations of the BVAR model. For each of the horizon periods, and for each simulated vector of risk factors, the corresponding impact on the individual balance sheets is measured using the presented valuation model. The balance sheets as of 2013-Q4 are considered to measure the impact of the scenario at Q1. Through the test horizon, only model-based endogenous changes affect these balance sheets. Exogenous modifications are thus excluded, for simplicity purposes. In the same spirit, no additional shocks are introduced during the stress horizon. Both assumptions are still easily adjustable in the provided code.

Hereafter, we report respectively: the simulated paths for risk factors, the first-round effects, the impact of banks' responses (i.e. second-round effects), and the contagion effects. To show the flexibility of the model and the role of each module in the final result, sensitivity analyses are also conducted. To keep this study as anonymous as possible, the reported results are aggregated for the six considered banking groups. Note that this does not mean a top-down approach, as individual results are obtained first before being aggregated. That is, our analysis of the impact of the shock is assessed on the balance sheet of each bank ant not on an aggregated balance-sheet.

Figure 4.2 shows the simulated paths of the exogenous risk factors over the stress horizon.

[5] See the next chapter for a detailed analysis of stress scenarios' selection.

Medians, 90% and 99% quartiles are reported. Overall, the simulation is fairly consistent with stable confidence intervals after Q1. Three main dynamics of risk factors are observed over the horizon. The first, includes stables changes, and holds for GDP, STOCK, and EX. The second, includes upward changes, and holds for CPI, REPI, and COM. The last risk factors record a continuous fall over the horizon. The monotonous dynamics of risk factors' values may be explained by the number of lags of the BVAR model and by the fact that no additional shocks are introduced during the horizon.

On each of the horizon's periods, the first-round impact of the simulated risk factors on the balance sheets is analysed using eight main indicators (Figure 4.3). These include: solvability (CET1) and leverage (LVG) ratios, net trading assets (NTA), credit losses (CL), net interest income (NII), profit (PROFIT), total liabilities (DEBT), and shareholder funds (SHARES). The first two indicators are Basel III regulatory ratios. The next three capture the impact of risk factors on the trading, the credit, and the net interest portfolios, respectively. The last three are the main indicators in the result and the balance sheet statements. Except for the leverage ratio, all other indicators do rise on the first period. LVG records a slight rise in the second period and remains at this level up to the horizon end. CET1 also remains stable during the remaining periods. One major result here is the strong relation between the dynamics of PROFIT and NTA. Indeed, even though NII is increasing during the remaining periods, and CL are fairly stable, PROFIT slightly decreases after the first period up to the end. This is due to a similar dynamics shown by the NTA. Despite of a decreasing PROFIT, the figure shows a continuously increasing balance sheet size (DEBT and SHARES) over the horizon. This confirms a common feature in banking systems, mainly observed during the last decade.

The reported results are aggregated for the whole banking system. Individual results are averaged using the banks' total assets as weighting factors. Table 4.4 compares these results with unweighted results for CET1 and LVG. The two ratios are clearly more dispersed in the unweighted version. For instance, the 5% and 95% quartiles of the CET1 ratio in the last period are respectively 7.5% and 14.31% for unweighted results. These values are more dispersed compared to the corresponding values in the weighted results (9.64% and 12.08%). This indicates that the aggregated results of Figure 4.3 are mainly influenced by those of big banks (over-weighted) whose ratios are less volatile. That is, results of small banks are more volatile than those of big ones. This feature also holds for the other indicators. However, due to space limitation and to the anonymous character of this study, we continue to report weighted aggregate results. One should just keep in mind that the reported results may present more outliers if banks are to be studied individually.

We now focus on the last period to analyse the distribution of the previous eight indicators, the weight of each portfolio on the results, and the banking population after first-round effects. Figure 4.4 presents the histogram of each indicator. The LVG distribution exhibits a bimodal form, due to the heterogeneity of the French banking system – especially between big and small banks. Overall, the leverage ratio is between 15% and 20%. However, for some scenarios (i.e. realizations of risk factors), the leverage ratio of big banks increases; causing an apparition of another cluster between 20% and 28% values. This bimodality is also observed in the DEBT and the SHARES figures, which is fairly intuitive as these indicators enter in the calculation of the leverage ratio. Regarding the solvability ratio, the result shows a right-skewed histogram. This can be explained as follows. In the last period only resilient banks are still alive, hence the good indicators in this period. Moreover, the NTA and the NII histograms are also right-skewed, impacting the form of the leverage ratio's histogram. Turning onto the PROFIT figure, we can see a right-skewed form with positive profits in most realizations. In some cases, however,

considerable losses (up to 50 millions) are recorded.[6] This aspect is confirmed by the left cluster in the SHARES histogram. This indicates near to zero values for shareholder funds, reflecting the considerable losses in such cases.

To study the impact of the different portfolios on the result of the period, we have decomposed the results of the last period by risk exposures. Table 4.5 compares the profit, the solvability ratio, and the leverage ratio for different portfolio's combinations. The result confirms the key role of the credit portfolio for the first two indicators. Hence, discarding credit exposures allows improving both regulatory ratios. However, the leverage ratio decreases due to an increase in its denominator. Discarding the trading or the net interest portfolio worsens all indicators. Using median values, the result seems to indicate that omitting the trading portfolio generates a higher decrease in the solvability ratio (9.9% vs. 10.43%), while omitting the net interest portfolio causes a higher decrease in the leverage ratio (21.13% vs. 22.32%). However, introducing the trading portfolio also brings more volatility in the ratio. Indeed, the 5%-95% quartile-range moves from 0.25% to 2.95% for the solvability ratio, and from 0% to 0.59% for the leverage ratio. Finally, given the specification of our different modules, the net interest income seems to be the main component of the profit.

Accounting for the first-round effects, we apply the default rules of section 4.3.4 to each balance sheet. Table 4.6 presents the resulting classification by bank and by scenario. The classification by bank refers to the percentage of cases in which the bank – the name of which is removed – goes bankrupt, records losses, or remains profitable. The classification by scenario is a classification of banks across all realizations. The classification by bank shows that four out of six banks remain profitable in more than 99% of cases. Among these four banks, three never go bankrupt – the last bank has a probability of failure of 0.3%. For banks 4 and 6, the probability of profitability is 70.9% and 80.2% respectively. This is mainly due to the fragility of these banks' balance sheets when the stress scenario is applied (i.e. at 2013:Q4). Their regulatory ratios have been quite low in that time. Moreover, in some cases, their results in the following quarters have been insufficient to ensure a significant recovery. A remarkable point in this result is that bank 4 is less exposed to failure than bank 6, even if it has a lower profitability probability. When this may be due to the considered default rules, this result also shows that bank 4 have more margin with respect to its regulatory ratios (i.e. more regulatory capital) compared to bank 6. Hence, in cases when the result of the period is negative, bank 4 is more likely to be sound (non-profitable) than bank 6 which has insufficient capital and liquidity buffers.

The classification by scenario shows that at least five banks remain profitable in 98.9% of cases. In the remaining cases, one or more banks are non-profitable or have to be liquidated. For example, the percentage of cases in which one failure is recorded in the banking system is 12.2%. This rate seems to be rather high. However, the discussed classification by bank helps to properly analyse this result. Indeed, the risk of failure is mainly presented by two banks, meaning that the exposure to failure is not homogeneous in the entire banking system. Yet, the failure of vulnerable banks may affect profitable ones through second-round and contagion effects. Before analysing this aspects, we present the impact of private decisions taken by the different banks in response to the recorded first-round effects.

Figure 4.5 shows the dynamics of three key indicators through the horizon (upper figures) as well as a focus on the last period (lower figures). These indicators include: new funds (profitable banks), new debts (non-profitable banks), and lost assets (failed banks). For profitable banks,

[6] All monetary results are displayed in thousand euros.

the results show a significant increase in funds – at least in the first periods – indicating a consolidation of their balance sheets. In cases where a failure is observed, the first period indicates a significant decrease in the total amount of assets in the banking system, due to the liquidation of failed banks. For non-profitable banks, the increase of liabilities seems to be insignificant. This is mainly explained by the fact that the negative result recorded by these banks is not necessarily accompanied by a shrinkage in their balance sheets – in which case no increase in liabilities is required. This result confirms that of Figure 4.3, in which this remark holds for the whole system. The lower panel of the figure focuses on the last period. The distribution of new funds by profitable banks is right-skewed with a significantly positive mean. The distribution of new debts by non-profitable banks shows a similar form. The observed cluster of negative values corresponds to liability reductions when the scenario leads to an expansion of the balance sheet. Finally, the distribution of lost assets is quite bimodal – and even trimodal. This is mainly due to the size of the liquidated bank, knowing that the studied banking system presents big, medium, and small banks.

The impact of these responses on the banks' balance sheets is presented by Table 4.7 – remind that failed banks have no response function. The latter compares the solvability and leverage ratios, the amount of liabilities, and the balance sheet's size for three stages: at the start of the period, after first-round effects, and after private response. For each period of the horizon, the results – the medians of the indicators – are presented, for profitable banks, non-profitable banks, and sound banks (i.e. profitable and non-profitable banks). The solvability ratio of profitable banks increases after first-round effects, which is due to an increase in their results. After reinvesting this profit, the ratio is reduced to converge towards its initial level. Overall, for non-profitable banks, the solvability ratio decreases after first-round effects. Even though the ratio is improved by the response of the bank, it now becomes clearly below its start-of-the-period level. The ratio of sound banks is fairly similar to that of profitable banks, as they are more influential in this stage (see Table 4.6). This remark holds for the other three indicators reported in the table. Regarding the leverage ratio, its value decreases for profitable banks and increases for non-profitable ones, given their positive and negative results, respectively. The reinvestment of the profit by profitable banks recovers their ratio, which converges to its start-of-the-period level. The ratio of non-profitable banks remains fairly stable after response. This is due to the assumption according to which these banks do not target regulatory ratios. Moreover, it seems that the response of non-profitable banks does not significantly impact their debts – which is also highlighted by Figure 4.5. As discussed above, this is because the negative results recorded by these banks is not necessarily accompanied by a significant decrease in the size of their balance sheet. This is exactly confirmed by the last three columns of the table. Indeed, for non-profitable banks, the balance sheet's size remains fairly stable in the three stages, while it increases considerably for profitable banks (due to their profit and reinvestment strategy).

We now turn towards the second-round effects caused by a failure of one or more banks in the system. These effects are transmitted to sound banks through the financial market (price decrease), the money market (refunding costs), and the interbank network (counterparty credit risk). Table 4.8 reports, for each period, the number of fundamental failures, the number of contagion failures, the number of iterations within the clearing algorithm, and the fraction of liquidated assets and its impact on the financial market price. Two main comments can be made for fundamental failures. On one side, the number of failures decreases as the horizon goes on. On the other, the maximum rate of failures is reached after two to three quarters. The first remark is explained by the fact that no additional shock is introduced during the horizon, as

the initial shock is progressively absorbed. Note, however, that the probability of fundamental failures is still positive at the end of the horizon. The second remark is motivated by the fact that first-round effects weakness non-profitable banks. Given the assumed response function for these banks, they become more exposed to second-round effects. Moreover, non-profitable banks suffer a reputation cost reflected by a tightening of their refunding costs for the following periods. All the results of a given period are then carried forward for the next one. Hence, the most severe impacts are recorded in the second and third periods – shortly after the initial shock. As the horizon goes on, the shock decreases as do the cross-period effects.

The features of fundamental failures are quite different from those of contagion failures. In the latter case, most failures are concentrated in the first year of simulation. It then considerably decreases during the remaining two years. These results can be explained by different facts. First, as discussed in the previous paragraph, non-profitable banks are more vulnerable during the first periods. This increases their probability of going bankrupt. Second, as more banks go bankrupt, second-round effects also increase. This is to impact other non-profitable banks, and even banks that have proven to be profitable after first-round effects. Third, in the first periods, contagion failures occur through several iterations (typically, four). This is because the contagion process terminates as most of vulnerable banks go bankrupt. These banks may be represented by the two banks discussed earlier or by a small bank. After these banks have been cleared from the system, the observed number of contagion failures clearly decreases in the following periods. Note, however, that in some scenarios, the contagion process causes the failure of the whole system. This occurs around the end of the horizon. Finally, contagion failures are relatively less frequent than fundamental failures. Again, this can be due to the presence of some more vulnerable banks in the system. When these banks make a failure after first-round effects, the remaining banks are relatively resilient and do not support contagion failures.

Regarding the fraction of liquidated trading assets and the generated impact on the market price, one can note that the results are quite similar over the stress horizon. The fraction of liquidated assets is relatively low compared to the size of the banking system. This explains a small decrease in the market price. Two conclusions may be drawn here. First, most failures concern small and medium banks – even though the size of the trading portfolio is not always proportional to the size of the bank. Second, the lower impact of the trading portfolio may suggest that most contagions are driven by the money market and the interbank network channels.

Finally, to study the weight of the contagion process in the outcomes, we have compared the results in different stages: at the start and at the end of the contagion process, after banks have rebalanced their balance sheets to account for contagion effects, after the central bank's intervention, and at the end of the period. The assumed central bank intervention consists in restoring the price in the financial market to its start-of-the-period level. Table 4.9 presents the results for rescued banks. Results for the whole system are also presented for the first and the last stages. Focusing on the solvability ratio, one can note that its level is too low for the whole system compared to that of rescued banks before the contagion took place. After the contagion terminates, the leverage ratio of rescued banks increases. This can be explained by the fact that vulnerable banks have been affected by the contagion process, and most of them go bankrupt. Discarding these banks has improved the solvability ratio of the remaining ones. The ratio is further enhanced by the intervention of the central bank. By increasing the financial market price, this reaction generates extra profits for the remaining banks. This has led to improve their solvability conditions. This comment also holds for the other indicators.

Finally, the rebalances made by rescued banks at the end of the period do not affect their solvability ratio. However, this is not the case for some other indicators (e.g. NTA) due to the considered rules of thumb in this study.

4.6 Conclusion

This chapter has presented a sequential model allowing to estimate the vulnerability of the French banking system to a set of systemic risk factors. The model is specified by a sequence of modules to capture the risk factors, their direct and indirect impact on the individual banks' balance sheets, as well as the nature and the size of the private and public response functions following the concretization of these risks.

The results show a strong relation between the dynamics of the profit and the trading portfolio. Yet, despite a decreasing profit, the banks' balance sheet are still continuously increasing over the simulation horizon. Moreover, the trading portfolio is shown to be key in the simulation results. Removing this portfolio from the analysis leads to a higher decrease in the solvability ratio. However, introducing the trading portfolio also brings more volatility in both the solvability and the leverage ratios. With this respect, the results of small banks seem to be more volatile than those of big ones.

After first round effects, four out of six banks remain profitable in more than 99% of cases. Among these banks, three never go bankrupt. For the banks having recorded a high failure probability, this is mainly explained by the fragility of their balance sheets when the initial shock has been applied – as their regulatory ratios were quite low in that time. In some cases, the results of these banks in the following quarters have also been insufficient to ensure a significant recovery. However, it is worth noting that it is the capital buffer rather than the profitability probability that makes a bank likely to go bankrupt. With this respect, the risk of failure is mainly presented by two banks in the system.

The solvability ratio of non-profitable banks decreases after first-round effects. Even though, this ratio is improved by these banks' responses, it becomes clearly below its start-of-the-period level. Moreover, this ratio remains fairly stable after response, due to the assumption that non-profitable banks do not target regulatory ratios. It seems also that the response of non-profitable banks does not significantly impact their debts, since their negative results are not necessarily accompanied by a significant decrease in the size of their balance sheets.

Regarding the contagion features, two main results are obtained for fundamental failures. First, the number of failures decreases as the horizon goes on, even though their probability is still positive at the horizon term. This is bacause the initial shock is progressively absorbed as no additional shock is introduced. Second, the maximum rate of failures is reached after two to three quarters. This is because first-round effects weakness non-profitable banks' situation, becoming more exposed to second-round effects. Together with the reputation cost carried forward for the following periods, the most severe impacts are recorded in the first quarters. As the horizon goes on, the shock decreases as do the cross-period effects.

Most contagion failures are concentrated in the first year. This can be explained by different facts. First, non-profitable banks are more vulnerable during the first periods, increasing their probability of going bankrupt. Second, as more banks go bankrupt, second-round effects also increase, hence impacting non-profitable and profitable banks. Third, in the first periods, contagion failures occur through several iterations until most vulnerable banks go bankrupt.

Afterwards, the number of contagion failures decreases in the following periods – even though, in some scenarios, the contagion process causes the failure of the entire system. Finally, contagion failures are relatively less frequent than fundamental failures, due to the presence of some more vulnerable banks in the system. When these banks make a failure, the remaining system becomes relatively resilient.

The fraction of liquidated assets is relatively low compared to the size of the banking system. This explains a small decrease in the market price. Two conclusions may be drawn accordingly. On one side, most failures concern small and medium banks. On the other, the lower impact of the trading portfolio suggests that most contagions are driven by the money market and the interbank network channels.

After the contagion terminates, the leverage ratio of rescued banks increases. This can be explained by the fact that vulnerable banks have been affected by the contagion process and most of them went bankrupt. Discarding these banks has also improved the solvability ratio of the remaining banks. This ratio is further enhanced by the intervention of the central bank which rises the financial market price, thus generating extra profits. A same result holds for the other considered regulatory and performance indicators. This public response is much more important than the private response of the rescued banks themselves. Indeed, the rebalances made by the latter do not seem to significantly affect their solvability ratio.

References

Acharya, V.V., L.H. Pedersen, T. Philippon, and M. Richardson (2010): Measuring Systemic Risk, *Federal Reserve Bank of Cleveland*, Working Paper No. 1002.

Acharya, V.V. and T. Yorulmazer (2007): Too Many to Fail – An Analysis of Time-Inconsistency in Bank Closure Policies, *Journal of Financial Intermediation*, 16(1), 1–31.

Adrian, T. and H.S. Shin (2008): Liquidity and Leverage, *Federal Reserve Bank of New York*, Staff Report No. 328.

Aikman, D., P. Alessandri, B. Eklund, P. Gai, S. Kapadia, E. Martin, N. Mora, G. Sterne, and M. Willison (2009): Funding Liquidity Risk in a Quantitative Model of Systemic Stability, *Bank of England*, Working Paper No. 372.

Alessandri, P., P. Gai, S. Kapadia, N. Mora, and C. Puhr (2009): Towards a Framework for Quantifying Systemic Stability, *International Journal of Central Banking*, 5(3), 47–81.

Allen, F. and D. Gale (2000): Financial Contagion, *Journal of Political Economy*, 108(1), 1–33.

Autorité de Contrôle Prudentiel (ACP) (2013): Analyses et Synthèses: Stress Tests sur le Système Bancaire et les Organismes d'Assurance en France, No. 11, January.

Billio, M., M. Getmansky, A.W. Lo, and L. Pellizzon (2012): Econometric Measures of Connectedness and Systemic Risk in the Finance and Insurance Sectors, *Journal of Financial Economics*, 104(3), 535–59.

Board of Governors of the Federal Reserve (FRB) (2011): Comprehensive Capital Analysis and Review: Objectives and Overview, March 18.

Board of Governors of the Federal Reserve (FRB) (2009): The Supervisory Capital Assessment Program: Overview of Results, May 7.

Borio, C., M. Drehmann, and K. Tsatsaronis (2012): Stress Testing Macro Stress Testing: Does it Live Up to Expectations? *Bank for International Settlements*, Working Paper No. 369.

Borio, C. and M. Drehmann (2009): Towards an Operational Framework for Financial Stability: "Fuzzy" Measurement and its Consequences, *Central Bank of Chile*, Working Paper No. 544.

Boss, M.G. (2008): Stresstests Für das Osterreichische FSAP-Update 2007: Methodik, Szenarien und Ergebnisse, Oesterreichische Nationalbank, *OeNB Finanzmarktsstabilitätsbericht*, 15, 72-99.

Boss, M., G. Krenn, C. Puhr, and M. Summer (2006): Systemic Risk Monitor: Risk Assessment and Stress Testing for the Austrian Banking System, Financial Stability Report 11, Oesterreichische Nationalbank, 83–95.

Boss, M., T. Breuer, H. Elsinger, M. Jandacka, G. Krenn, A. Lehar, C. Puhr, and M. Summer (2006): Systemic Risk Monitor: A Model for Systemic Risk Analysis and Stress Testing of Banking Systems, Technical Report, Oesterreichische Nationalbank.

Bris, A., I. Welch, and N. Zhu (2006): The Costs of Bankruptcy: Chapter 7 Liquidation versus Chapter 11 Reorganization, *Journal of Finance*, 61(3), 1253–1303.

Brunnermeier, M.K. and L.H. Pedersen (2009): Market Liquidity and Funding Liquidity, *The Review of Financial Studies*, 22(6), 2201–38.

Brusco, S. and F. Castiglionesi (2007): Liquidity Coinsurance, Moral Hazard, and Financial Contagion, *The Journal of Finance*, 62(5), 2275–302.

Caballero, R.J. (2010): The 'Other' Imbalance and the Financial Crisis, *NBER*, Working Paper No. 15636.

Chen, Y. (1999): Banking panics: The Role of the First-Come, First-Served Rule and Information, *Journal of Political Economy*, 107(5), 946–68.

Cifuentes, R., G. Ferrucci, and H.S. Shin (2005): Liquidity Risk and Contagion, *Journal of the European Economic Association*, 3(2–3), 556–66.

Committee of European Banking Supervisors (CEBS) (2010): Results of 2010 EU Wide Stress Testing Exercise, July 23.

Committee of European Banking Supervisors (CEBS) (2009): CEBS Press Release on the Results of the EU-Wide Stress Testing Exercise, October 1.

Diamond, D.W. and R.G. Rajan (2005): Liquidity Shortages and Banking Crises, *The Journal of Finance*, 60(2), 615–47.

Drehmann, M., S. Sorensen, and M. Stringa (2008): The Integrated Impact of Credit and Interest Rate Risk on Banks: An Economic Value and Capital Adequacy Perspective, *Bank of England*, Working Paper No. 339.

Duffie, D. and H. Zhu (2011): Does a Central Clearing Counterparty Reduce Counterparty Risk? *Review of Asset Pricing Studies*, 1(1), 74–95.

Eisenberg, L. and T.F. Noe (2001): Systemic Risk in Financial Systems, *Management Science*, 47(2), 236–49.

Eisenschmidt, J. and J. Tapking (2009): Liquidity Risk Premia in Unsecured Interbank Money Markets, *European Central Bank*, Working Paper No. 1025.

Elsinger, H., A. Lehar, and M. Summer (2006a): Risk Assessment for Banking Systems, *Management Science*, 52(9), 1301–14.

Elsinger, H., A. Lehar, and M. Summer (2006b): Using Market Information for Banking System Risk Assessment, *International Journal of Central Banking*, 2(1), 137–65.

European Central Bank (ECB) (2009): The Concept of Systemic Risk, *Financial Stability Review*, December 18, 134–42.

European Central Bank (ECB) (2010): Financial Networks and Financial Stability. *Financial Stability Review*, June, 155–60.

Freixas, X. and B.M. Parigi (1998): Contagion and Efficiency in Gross and Net Payment Systems, *Journal of Financial Intermediation*, 7(1), 3–31.

Freixas, X., B.M. Parigi, and J.-C. Rochet (2000): Systemic Risk, Interbank Relations, and Liquidity Provision by the Central Bank, *Journal of Money, Credit and Banking*, 32(3), 611–38.

Furfine, C.H. (2003): Interbank Exposures: Quantifying the Risk of Contagion, *Journal of Money, Credit and Banking*, 35(1), 111–28.

Gauthier, C., A. Lehar, and M. Souissi (2012): Macroprudential Capital Requirements and Systemic Risk, *Journal of Financial Intermediation*, 21(4), 594–618.

Goodhart, C.A.E., M.U. Peiris, and D.P. Tsomocos (2013): Global Imbalances and Taxing Capital Flows, *International Journal of Central Banking*, 9(2), 13–44.

Goodhart, C.A.E., P. Sunirand, and D.P. Tsomocos (2006a): A Model to Analyse Financial Fragility, *Economic Theory*, 27(1), 107–42.

Goodhart, C.A.E., P. Sunirand, and D.P. Tsomocos (2006b): A Time Series Analysis of Financial Fragility in the UK Banking System, *Annals of Finance*, 2(1), 1–21.

Goodhart, C.A.E., P. Sunirand, and D.P. Tsomocos (2004): A Model to Analyse Financial Fragility: Applications, *Journal of Financial Stability*, 1(1), 1–30.

Group of Ten (G10) (2001): Report on Consolidation in the Financial Sector: Chapter III. Effects of Consolidation on Financial Risk, *International Monetary Fund*, Working Paper.

Huang, R. and L. Ratnovskib (2011): The Dark Side of Bank Wholesale Funding, *Journal of Financial Intermediation*, 20(2), 248–63.

James, C. (1991): The Losses Realized in Bank Failures, *Journal of Finance*, 46(4), 1223–42.

Kahn, C.M., J. McAndrews, and W. Roberds (2003): Settlement Risk under Gross and Net Settlement, *Journal of Money, Credit and Banking*, 35(4), 591–608.

Kapadia, S., M. Drehmann, J. Elliott, and G. Sterne (2013): Liquidity Risk, Cash Flow Constraints, and Systemic Feedbacks, in: J.G. Haubrich and A.W. Lo (Eds.), Quantifying Systemic Risk, *NBER*, Chicago, 28–61.

Leitner, Y. (2005): Financial Networks: Contagion, Commitment, and Private Sector Bailouts, *The Journal of Finance*, 60(6), 2925–53.

Mishkin, F.S. (2007): Systemic Risk and the International Lender of Last Resort, Speech delivered at the Tenth Annual International Banking Conference, Federal Reserve Bank of Chicago, Chicago, September 28.

Moussa, A. (2011): Contagion and Systemic Risk in Financial Networks, Ph.D. Thesis, Columbia University, New York.

Rosengren, E.S. (2010): Asset Bubbles and Systemic Risk, Speech delivered at the Global Interdependence Center's Conference on 'Financial Interdependence in the World's Post-Crisis Capital Markets', Federal Reserve Bank of Boston, Philadelphia, March 3.

Saade A., D. Osorio, and S. Estrada (2007): An Equilibrium Approach to Financial Stability Analysis: The Columbian Case, *Annals of Finance*, 3(1), 75–105.

Upper, C. (2011): Simulation Methods to Assess the Danger of Contagion in Interbank Networks, *Journal of Financial Stability*, 7(3), 111–25.

van den End, J.W. (2012): Do Basel III and Unconventional Monetary Policy Work? *Applied Financial Economics*, 22(15), 1233–57.

van den End, J.W. (2010): Liquidity Stress-Tester: A Macro Model for Stress-Testing Banks' Liquidity Risk, *CESifo Economic Studies*, 56(1), 38–69.

Wong, J.H.-Y., K.-F. Choi, and P.-W. Fong (2008): A Framework for Stress-Testing Banks' Credit Risk, *The Journal of Risk Model Validation*, 2(1), 3–23.

Assets			
Class	Gains/losses modelling	Cash flows modelling	Risk weighting (%)
Trading assets	Trading book model	Risk-free rate + 15 basis points (bps)	50
Interbank assets	Interbank network matrix	Risk-free rate + 15 bps	35
Households	Credit risk model	Coupon given by the net interest income model + 50 bps	50
Administrations	Credit risk model	Risk-free rate	0
Non-financial companies	Credit risk model	Coupon given by the net interest income model + 50 bps	100
Other financial institutions	Credit risk model	Risk-free rate + 15 bps	40
Other assets		Risk-free rate	0
Liabilities			
Trading liabilities	Trading book model	Risk-free rate + 15 bps	
Interbank debts	Interbank network matrix	Risk-free rate + 15 bps + reputation 1 + reputation 2	
Household	Credit risk model	Risk-free rate + reputation 1 + reputation 2	
Administrations	Funding liquidity risk (reputation cost)	Risk-free rate + reputation 1 + reputation 2	
Non-financial companies	Funding liquidity risk (reputation cost)	Risk-free rate + reputation 1 + reputation 2	
Other financial institutions	Funding liquidity risk (reputation cost)	Risk-free rate + 15 bps + reputation 1 + reputation 2	
Other liabilities		Risk-free rate + 15 bps	

Table 4.1: Classification and valuation models for balance sheet items

Trading assets include financial instruments in their fair market value, financial assets available for sale, and derivatives.
Trading liabilities include financial instruments in their fair market value, debts represented by securities, and derivatives.
Households Include mortgage loans.

Series	Description	Source (Code)	Transformation	Sample	Obs.
GDP	Total nominal GDP (base 2005)	INSEE (001615898)	$\ln\left(\frac{\text{GDP}}{\text{CPI}}\right)$	1992:Q1-2013:Q4	88
CPI	Consumer price index – All items excluding energy (base 1998)	INSEE (000641193)	$\ln(\text{CPI})$	1992:Q1-2013:Q4	88
REPI	Price index of second-hand dwellings flats – Paris (base 2010)	INSEE (001587636)	$\ln(\text{REPI})$	1992:Q1-2013:Q4	88
STOCK	CAC 40 index (end-of-quarter closing prices)	Datastream (FRCAC40)	$\ln\left(\frac{\text{STOCK}}{\text{CPI}}\right)$	1992:Q1-2013:Q4	88
SR	Interbank overnight rate – EONIA (TMP before 01/01/1999)	Banque de France (QS.D.IEUEONIA)	$0.25 \cdot \ln\left(1+\frac{\text{SR}}{100}\right)$	1992:Q1-2013:Q4	88
LR	French government debt reference rate – 30 years (end-of-quarter rates)	Banque de France (QS.D.IFRPHF30)	$0.25 \cdot \ln\left(1+\frac{\text{LR}}{100}\right)$	1992:Q1-2013:Q4	88
EX	Nominal exchange rate euro/USD	Datastream (XUDLERD)	$\ln(\text{EX})$	1992:Q1-2013:Q4	88
COM	Crude oil dated Brent USD per barrel (end-of-quarter closing prices)	Datastream (OILBRN)	$\ln\left(\frac{\text{COM}\cdot\text{EX}}{\text{CPI}}\right)$	1992:Q1-2013:Q4	88

Table 4.2: Data description

Equation (4.1)

	GDP	CPI	REPI	STOCK	SR	LR	EX	COM
GDP_{l1}	0.86***	0.09***	0.12	1.30	0.06***	0.01**	-0.79*	-0.16
	(0.04)	(0.02)	(0.14)	(0.88)	(0.02)	(0.01)	(0.40)	(1.46)
CPI_{l1}	0.09	0.88***	-0.26	-2.67**	-0.17***	-0.04***	1.24**	-0.83
	(0.06)	(0.04)	(0.21)	(1.34)	(0.03)	(0.01)	(0.61)	(2.22)
$REPI_{l1}$	0.00	0.00	1.00***	-0.24*	0.01***	0.00**	-0.16**	0.82***
	(0.01)	(0.00)	(0.02)	(0.14)	(0.00)	(0.00)	(0.06)	(0.23)
$STOCK_{l1}$	0.02***	-0.01**	0.02	0.79***	0.00	0.00	0.05	0.27*
	(0.00)	(0.00)	(0.01)	(0.08)	(0.00)	(0.00)	(0.04)	(0.14)
SR_{l1}	-0.36*	-0.15	-0.51	-4.13	0.11	-0.12***	6.78***	-13.45*
	(0.20)	(0.12)	(0.71)	(4.32)	(0.10)	(0.04)	(1.97)	(7.19)
LR_{l1}	-0.00	-011	-3.77**	-29.76***	0.49**	0.81***	-9.12**	10.79
	(0.42)	(0.25)	(1.46)	(8.86)	(0.22)	(0.07)	(4.06)	(14.79)
EX_{l1}	0.00	0.00	0.00	-0.17	-0.00	-0.00*	0.87***	-0.39*
	(0.01)	(0.00)	(0.02)	(0.13)	(0.00)	(0.00)	(0.06)	(0.22)
COM_{l1}	0.00	-000	0.01	0.12*	0.00	-0.00	0.05*	0.46***
	(0.00)	(0.00)	(0.01)	(0.06)	(0.00)	(0.00)	(0.03)	(0.11)
Constant	0.71***	-011	0.19	4.00	0.21**	0.07**	1.21	0.23
	(0.18)	(0.11)	(0.62)	(3.79)	(0.09)	(0.03)	(1.73)	(6.30)
Obs.	83							
R^2	0.99	0.99	0.99	0.99	0.95	0.99	0.97	0.98
Adjusted R^2	0.99	0.99	0.99	0.99	0.95	0.99	0.96	0.98
LLH	2077.88							

Equation (4.2)

	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$
	3.07***	4.64	-1.12*	-0.88	0.01
	(0.88)	(3.22)	(0.60)	(0.48)	(0.01)
Obs.	12				
R^2	0.43				
Adjusted R^2	0.14				

Equation (4.3)

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
	10.68	-10.29	-5.68	-0.02	-9.65
	(24.10)	(27.13)	(4.56)	(0.59)	(4.39)
Obs.	25				
Adjusted R^2	-0.25				

Table 4.3: Estimation results

	Unweighted values								Weighted values							
	Quartile (%)								Quartile (%)							
Period	1	5	10	Mean	50	90	95	99	1	5	10	Mean	50	90	95	99
								CET1								
1	5.48	5.92	7.23	11.69	11.99	14.47	15.79	19.58	9.91	10.01	10.10	11.17	10.93	12.51	12.96	13.59
2	5.24	6.34	7.08	12.01	12.28	15.45	16.95	20.48	9.61	9.99	10.15	11.39	10.96	13.25	14.68	16.46
3	5.00	6.55	7.27	12.14	12.33	15.69	17.33	21.10	9.65	10.07	10.21	11.48	11.09	13.68	14.23	15.05
4	5.76	6.78	7.47	12.03	12.09	15.19	16.67	20.40	9.68	9.98	10.15	11.32	10.93	13.22	13.90	14.89
5	6.27	6.99	7.84	12.02	12.07	14.96	16.42	20.81	9.72	10.00	10.15	11.28	10.98	12.88	13.58	14.72
6	6.40	7.08	8.03	11.96	11.85	14.71	16.08	20.08	9.57	9.95	10.12	11.17	10.91	12.67	13.19	13.75
7	5.81	7.14	7.97	11.88	11.75	14.46	15.77	20.21	9.56	9.92	10.08	11.12	10.85	12.45	12.89	14.04
8	6.54	7.27	8.46	11.89	11.70	14.37	15.61	19.59	9.60	9.89	10.05	11.12	10.93	12.28	12.91	13.85
9	6.66	7.32	8.48	11.76	11.40	14.22	14.94	18.22	9.39	9.77	9.97	10.99	10.89	11.96	12.41	13.32
10	6.79	7.39	8.85	11.73	11.24	14.14	14.74	18.30	9.29	9.75	9.94	10.97	10.99	11.82	12.27	13.12
11	7.00	7.45	9.00	11.67	11.14	14.07	14.44	17.79	8.08	9.69	9.92	10.93	11.07	11.82	12.15	13.05
12	7.02	7.50	9.20	11.64	11.13	14.03	14.31	17.65	8.24	9.64	9.85 1	1.06	11.01	11.73	12.08	12.98
								LVG								
1	12.24	14.18	14.91	21.56	18.86	33.83	38.56	49.38	15.94	16.79	17.42	19.80	20.10	21.86	22.06	22.28
2	12.01	13.93	14.73	22.85	19.02	37.61	38.79	57.53	12.70	14.69	17.02	20.42	20.99	23.12	23.78	25.00
3	11.96	13.73	14.59	22.44	18.86	38.52	38.90	56.09	14.25	15.19	16.02	20.32	20.76	23.30	23.84	25.40
4	12.37	14.00	14.73	22.31	18.92	38.50	38.94	58.63	14.24	15.57	16.64	20.59	21.01	23.73	24.51	25.64
5	12.68	14.29	14.81	22.14	18.84	37.39	38.91	55.37	14.56	16.09	17.14	20.67	20.57	24.01	24.59	25.67
6	13.27	14.52	14.93	22.41	18.89	37.78	38.96	55.43	15.76	16.77	17.70	20.85	20.44	24.29	24.93	26.21
7	13.27	14.60	14.96	21.67	18.90	37.25	38.95	61.21	15.07	17.05	17.97	20.95	20.05	24.61	25.21	26.29
8	13.41	14.65	14.98	22.10	18.87	36.29	38.92	57.27	15.53	17.14	18.01	20.93	19.91	24.70	25.35	26.49
9	14.26	14.80	15.03	21.74	18.90	36.33	38.94	53.24	16.19	17.75	18.49	21.04	19.90	25.00	26.02	27.35
10	14.28	14.83	15.04	21.43	18.84	35.50	38.93	56.27	16.20	18.01	18.69	21.13	19.85	25.24	26.15	27.48
11	14.33	14.88	15.06	26.33	18.88	35.60	38.93	54.23	16.47	18.22	18.71	21.24	19.82	25.34	26.36	27.85
12	14.55	14.93	15.07	21.33	18.92	33.84	38.93	51.52	16.98	18.33	18.87	21.16	19.79	25.62	26.46	28.03

Table 4.4: Solvability and leverage ratios: unweighted vs. weighted values

	CET1			LVG			PROFIT		
	Quantile (%)								
Portfolio	5	50	95	5	50	95	5	50	95
All	10.01	10.93	12.96	23.07	23.07	23.07	13463.86	50147.82	106886.27
Market	9.59	9.59	9.59	17.24	20.66	22.77	3799.69	33116.09	96118.45
Credit	8.81	9.45	9.56	23.14	23.45	25.28	-24839.24	-4604.37	-862.24
Net Interest	9.87	10.05	10.63	20.69	21.96	22.37	9870.79	14709.81	33233.03
w/o market	9.75	9.90	10.00	22.09	22.32	22.68	6047.77	9880.79	12999.79
w/o credit	10.14	11.20	13.28	16.33	19.58	21.75	17754.59	51218.60	117876.72
w/o net interest	9.45	10.43	12.50	17.43	21.13	23.46	-4749.39	26087.31	91926.03

Table 4.5: First-round effects: A decomposition by portfolio exposures (last period)

Classification by bank

	Probability (%)					
Status	Bank 1	Bank 2	Bank 3	Bank 4	Bank 5	Bank 6
Profitable	99.3	99.7	99.7	70.9	99.6	80.2
Non-profitable	0.7	0.3	0.0	23.3	0.4	13.5
Failed	0.0	0.0	0.3	5.8	0.0	6.3
Total	100	100	100	100	100	100

Classification by scenario

	Probability (%)							
Status	0 case	1 case	2 cases	3 cases	4 cases	5 cases	6 cases	Total
Profitable	0.0	0.0	0.1	0.3	0.7	47.9	51	100
Non-profitable	62.3	37.2	0.5	0.0	0.0	0.0	0.0	100
Failed	87.7	12.2	0.1	0.0	0.0	0.0	0.0	100

Table 4.6: First-round effects: classification of the banking population (last period)

Profitable banks

	CET1 (%)			LVG (%)			DEBTS (millions)			BS (millions)		
Period	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
1	9.45	11.36	9.24	23.25	19.40	23.98	6179.45	6189.87	7710.72	6445.56	6510.66	8031.51
2	9.91	11.53	9.30	23.38	20.14	23.91	7371.81	7383.80	8842.99	7685.97	7751.03	9210.22
3	10.00	11.57	9.39	23.26	20.11	23.70	8237.09	8250.54	9827.37	8589.51	8661.29	10238.12
4	10.10	11.39	9.56	23.07	20.43	23.28	9210.01	9225.72	10604.01	9605.87	9675.95	11054.25
5	10.19	11.32	9.64	22.89	20.53	23.08	10227.13	10244.56	11628.94	10668.93	10741.07	12125.45
6	10.21	11.19	9.70	22.89	20.78	22.97	11359.01	11378.41	12705.26	11848.26	11921.35	13248.19
7	10.27	11.16	9.78	22.79	20.86	22.81	12344.98	12366.54	13671.02	12876.55	12952.48	14256.96
8	10.37	11.14	9.88	22.66	20.89	22.63	13418.21	13441.63	14763.28	13997.28	14076.40	15398.05
9	10.39	11.02	10.02	22.60	21.09	22.42	14398.67	14424.95	15530.09	15020.14	15096.44	16201.58
10	10.53	11.31	10.04	22.33	20.97	22.30	15212.67	15239.12	16274.98	15873.28	15947.57	16983.43
11	10.48	10.97	10.10	23.01	21.19	22.16	15887.41	15914.70	16758.67	16574.47	16648.75	17492.72
12	10.57	11.14	10.16	22.13	20.98	21.86	16644.90	16673.39	17548.22	17371.98	17444.53	18319.36

Non-profitable banks

	CET1 (%)			LVG (%)			DEBTS (millions)			BS (millions)		
Period	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
1	11.40	9.71	9.66	23.44	28.06	28.23	791.60	792.94	798.07	829.98	825.04	830.18
2	11.72	10.10	10.04	26.09	31.00	31.18	723.71	724.71	726.26	757.75	754.07	755.62
3	11.97	10.33	10.20	26.70	32.10	32.39	735.65	736.75	737.89	769.86	766.55	767.69
4	12.12	10.62	10.41	27.65	33.00	33.37	781.85	783.25	786.82	817.23	814.59	818.16
5	12.39	10.95	10.69	28.13	33.33	33.74	822.93	824.40	827.58	859.90	857.32	860.51
6	12.25	10.76	10.55	28.79	34.44	34.81	937.35	938.66	940.72	979.44	976.61	978.67
7	12.28	11.18	10.87	28.63	33.03	33.42	1407.96	1410.58	1418.84	1473.23	1470.82	1479.08
8	12.49	11.15	10.77	27.66	32.91	33.42	1557.27	1560.10	1570.91	1629.17	1626.47	1637.28
9	12.16	11.19	10.80	28.81	33.08	33.52	1737.76	1741.32	1750.57	1818.50	1813.15	1822.40
10	12.20	11.31	10.84	29.48	33.52	34.04	2203.86	2208.22	2221.14	2300.71	2293.62	2306.53
11	12.02	11.09	10.70	28.13	31.98	32.42	2738.72	2743.07	2761.57	2869.14	2861.17	2879.68
12	11.60	14.88	11.03	28.66	32.55	33.01	2806.01	2811.82	2802.51	2945.87	2933.77	2924.46

Sound banks

	CET1 (%)			LVG (%)			DEBTS (millions)			BS (millions)		
Period	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
1	9.57	11.26	9.28	23.09	19.65	23.97	6477.88	6488.80	8011.59	6758.46	6821.70	8344.49
2	9.97	11.47	9.33	23.28	20.28	23.92	7581.69	7593.97	9053.61	7905.72	7969.71	9429.35
3	10.06	11.53	9.42	23.17	20.21	23.69	8441.60	8455.36	10032.50	8803.53	8874.39	10451.54
4	10.13	11.36	9.57	23.02	20.51	23.29	9388.27	9404.30	10783.41	9792.20	9861.68	11240.79
5	10.22	11.30	9.65	22.85	20.58	23.09	10397.47	10415.21	11800.25	10846.93	10918.53	12303.57
6	10.24	11.18	9.71	22.84	20.82	22.98	11544.79	11564.46	12891.71	12042.38	12114.91	13442.16
7	10.29	11.14	9.79	22.73	20.89	22.80	12589.13	12611.14	13915.82	13132.06	13207.47	14512.15
8	10.38	11.13	9.88	22.60	20.91	22.62	13662.23	13686.09	15008.21	14252.61	14331.20	15653.31
9	10.42	11.00	10.00	22.51	20.99	22.29	14705.13	14732.04	15838.81	15340.84	15416.20	16522.97
10	10.50	10.98	10.00	22.19	21.06	22.01	15571.56	15598.73	16634.79	16247.96	16320.85	17356.91
11	10.49	10.93	10.09	22.90	21.23	22.20	16357.99	16386.02	17230.21	17067.75	17140.17	17984.35
12	10.55	11.07	10.11	22.20	21.14	22.04	17021.88	17051.19	17919.15	17768.38	17838.43	18706.38

Table 4.7: Second-round effects (last period)

This table compares the solvability and leverage ratios, the amount of liabilities, and the balance sheet size for profitable, non-profitable, and sound banks. The comparison is led in three stages: at the start of the period (T1), after the first-round effects of the shock (T2), and after the second-round effects of private responses (T3).

Period	Fondamental failures (%)						Contagion failures (%)						Iterations						Fraction	Price fall
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6		
1	8.8	0.1	0.0	0.0	0.0	0.0	0.0	2.0	1.8	5.1	0.0	0.0	91.1	2.3	6.6	0.0	0.0	0.0	0.97	0.04
2	11.2	0.0	0.1	0.0	0.0	0.0	0.0	5.1	0.4	5.8	0.0	0.0	88.8	5.4	5.7	0.1	0.0	0.0	0.97	0.04
3	10.8	0.0	0.0	0.0	0.0	0.0	0.0	5.6	0.1	5.1	0.0	0.0	89.2	5.7	4.9	0.2	0.0	0.0	0.97	0.04
4	7.9	0.2	0.0	0.0	0.0	0.0	0.0	6.1	0.2	1.8	0.0	0.0	92.1	6.0	1.9	0.0	0.0	0.0	0.97	0.04
5	5.6	0.0	0.0	0.0	0.0	0.0	0.0	4.4	0.2	0.9	0.0	0.0	94.4	4.6	0.9	0.0	0.0	0.1	0.96	0.03
6	5.0	0.0	0.0	0.0	0.0	0.0	0.0	4.2	0.0	0.8	0.0	0.0	95.0	4.2	0.8	0.0	0.0	0.0	0.97	0.04
7	6.4	0.0	0.0	0.0	0.0	0.0	0.1	6.1	0.0	0.1	0.0	0.0	93.7	6.1	0.1	0.0	0.0	0.1	0.96	0.03
8	4.7	0.0	0.0	0.0	0.0	0.0	0.0	4.6	0.0	0.1	0.0	0.0	95.3	4.6	0.1	0.0	0.0	0.0	0.97	0.04
9	3.7	0.1	0.0	0.0	0.0	0.0	0.0	3.6	0.0	0.0	0.0	0.1	96.2	3.6	0.0	0.0	0.1	0.1	0.94	0.04
10	3.8	0.1	0.0	0.0	0.0	0.0	0.0	3.6	0.1	0.0	0.0	0.0	96.0	3.7	0.0	0.0	0.0	0.2	0.92	0.04
11	2.6	0.1	0.0	0.0	0.0	0.0	0.0	2.7	0.0	0.0	0.0	0.0	97.3	2.6	0.0	0.0	0.0	0.0	0.97	0.04
12	3.0	0.0	0.0	0.0	0.0	0.0	0.0	2.3	0.2	0.0	0.1	0.0	96.8	2.3	0.3	0.0	0.0	0.4	0.84	0.04

Table 4.8: Contagion effects

	CET1 (%)								LVG (%)							
	Quartile (%)								Quartile (%)							
Period	1	5	10	Mean	50	90	95	99	1	5	10	Mean	50	90	95	99
At the start of the contagion process																
1	0.08	0.08	0.08	0.09	0.09	0.10	0.10	0.10	22.76	22.80	22.80	25.72	27.16	28.40	29.00	29.51
2	0.08	0.08	0.08	0.09	0.09	0.10	0.10	0.10	22.63	22.68	22.70	25.88	27.36	29.61	30.38	30.64
3	0.08	0.08	0.08	0.09	0.09	0.10	0.10	0.10	22.49	22.59	22.63	25.98	26.48	29.36	29.78	30.87
4	0.08	0.08	0.08	0.10	0.10	0.10	0.10	0.10	22.47	22.54	22.63	25.07	23.86	29.11	29.96	31.10
5	0.08	0.08	0.09	0.10	0.10	0.10	0.11	0.11	22.20	22.42	22.51	24.58	23.57	29.15	29.37	30.04
6	0.08	0.08	0.09	0.10	0.10	0.10	0.11	0.11	22.31	22.47	22.64	25.23	24.77	29.09	29.36	29.95
7	0.08	0.08	0.09	0.10	0.10	0.11	0.11	0.11	21.62	22.28	22.57	24.38	23.74	26.81	29.70	30.54
8	0.08	0.09	0.09	0.10	0.10	0.10	0.10	0.11	22.91	23.41	23.78	25.44	25.21	28.14	28.98	30.12
9	0.09	0.09	0.09	0.10	0.10	0.11	0.11	0.11	22.35	22.66	22.74	24.89	24.78	27.05	27.97	29.20
10	0.02	0.09	0.09	0.09	0.10	0.10	0.10	0.10	22.44	22.77	23.22	18.32	24.78	27.61	28.14	29.37
11	0.07	0.09	0.09	0.10	0.10	0.10	0.11	0.14	14.30	21.99	23.85	25.18	25.40	27.71	28.73	29.97
12	0.02	0.03	0.09	0.09	0.10	0.10	0.10	0.11	23.21	23.30	23.35	32.23	25.08	28.55	94.34	28.09
At the end of the contagion process																
1	0.10	0.10	0.10	0.11	0.10	0.12	0.12	0.12	19.68	19.68	19.68	20.18	20.33	20.35	20.35	20.62
2	0.10	0.10	0.10	0.11	0.10	0.12	0.12	0.12	19.68	19.68	19.68	20.10	20.35	20.38	20.39	21.07
3	0.09	0.10	0.10	0.11	0.11	0.12	0.12	0.12	19.67	19.68	19.68	20.32	20.36	20.39	20.73	27.97
4	0.09	0.10	0.10	0.11	0.12	0.12	0.12	0.12	19.68	19.68	19.69	20.33	19.79	20.42	21.95	28.38
5	0.09	0.10	0.10	0.11	0.12	0.12	0.12	0.12	19.67	19.68	19.68	20.24	19.74	20.42	21.47	28.24
6	0.09	0.09	0.10	0.11	0.12	0.12	0.12	0.12	19.67	19.69	19.69	20.64	19.81	20.52	28.50	29.11
7	0.09	0.09	0.11	0.11	0.12	0.12	0.12	0.12	19.67	19.69	19.70	20.58	19.79	20.82	28.73	29.02
8	0.09	0.09	0.11	0.11	0.12	0.12	0.12	0.13	19.54	19.74	19.78	20.63	19.88	21.47	28.61	28.93
9	0.10	0.11	0.11	0.11	0.12	0.12	0.12	0.12	19.68	19.71	19.71	21.11	19.87	20.81	21.51	22.13
10	0.03	0.10	0.11	0.11	0.12	0.12	0.12	0.13	19.63	19.71	19.74	23.00	19.90	21.59	22.89	23.56
11	0.11	0.11	0.11	0.11	0.12	0.12	0.14	0.15	13.39	18.22	19.75	21.22	19.94	21.23	21.41	21.47
12	0.07	0.07	0.10	0.11	0.12	0.12	0.12	0.12	19.75	19.77	19.79	21.98	20.00	22.10	42.97	43.17

	CET1 (%)								LVG (%)							
	Quartile (%)								Quartile (%)							
Period	1	5	10	Mean	50	90	95	99	1	5	10	Mean	50	90	95	99
After balance sheet rebalancing																
1	0.11	0.11	0.12	0.13	0.12	0.15	0.15	0.15	14.91	15.41	15.43	17.06	17.64	18.41	18.86	19.21
2	0.10	0.10	0.10	0.12	0.12	0.14	0.14	0.14	15.92	16.07	16.10	18.27	18.49	20.99	21.56	21.81
3	0.08	0.10	0.10	0.12	0.12	0.14	0.14	0.14	16.18	16.29	16.34	19.30	18.82	22.18	22.69	28.62
4	0.08	0.09	0.09	0.12	0.13	0.14	0.14	0.14	16.35	16.54	16.58	18.97	17.83	22.78	23.89	29.00
5	0.08	0.09	0.09	0.12	0.13	0.14	0.14	0.14	16.48	16.63	16.70	19.01	17.55	23.91	25.84	28.82
6	0.08	0.08	0.08	0.12	0.12	0.14	0.14	0.14	16.78	17.07	17.13	20.21	18.79	25.65	29.11	29.71
7	0.08	0.08	0.09	0.12	0.13	0.13	0.14	0.14	16.93	17.04	17.32	19.59	18.51	23.79	29.26	29.71
8	0.08	0.08	0.09	0.11	0.11	0.12	0.13	0.13	17.85	18.33	18.57	21.07	20.25	24.74	29.14	29.54
9	0.09	0.10	0.10	0.11	0.12	0.13	0.13	0.13	17.42	17.75	17.90	21.44	19.90	22.37	23.74	25.10
10	0.09	0.09	0.10	0.11	0.11	0.13	0.13	0.13	17.92	18.38	18.59	21.46	20.93	26.16	27.36	29.13
11	0.07	0.08	0.09	0.11	0.11	0.12	0.12	0.13	18.70	19.16	19.52	22.22	21.44	27.02	28.01	29.87
12	0.09	0.09	0.09	0.11	0.11	0.12	0.12	0.12	18.84	18.98	19.26	22.29	21.16	26.13	27.78	28.28
After the central bank intervention																
1	0.11	0.11	0.12	0.13	0.12	0.14	0.14	0.15	15.10	15.67	15.69	17.01	17.38	18.11	18.53	18.86
2	0.10	0.10	0.11	0.12	0.12	0.14	0.14	0.14	16.17	16.32	16.36	18.20	18.34	20.52	21.16	21.83
3	0.08	0.10	0.10	0.12	0.12	0.14	0.14	0.14	16.43	16.55	16.61	19.22	18.82	21.62	22.08	28.53
4	0.08	0.09	0.10	0.12	0.13	0.14	0.14	0.14	16.61	16.82	16.87	19.08	18.19	22.18	23.19	28.91
5	0.08	0.09	0.09	0.12	0.13	0.14	0.14	0.14	16.76	16.91	17.00	19.17	17.90	23.21	24.96	28.71
6	0.08	0.08	0.09	0.11	0.12	0.13	0.13	0.13	17.07	17.40	17.46	20.38	19.13	24.79	29.00	29.62
7	0.08	0.08	0.09	0.12	0.12	0.13	0.13	0.13	17.22	17.36	17.67	19.93	18.93	23.54	29.19	29.61
8	0.07	0.08	0.09	0.11	0.11	0.12	0.12	0.13	18.23	18.71	18.99	21.63	20.77	24.68	29.20	35.62
9	0.09	0.09	0.10	0.11	0.11	0.13	0.13	0.13	17.76	18.13	18.29	22.04	20.38	23.04	24.31	25.93
10	0.08	0.08	0.09	0.11	0.11	0.12	0.12	0.13	18.31	18.80	19.02	22.20	21.60	26.49	30.88	32.71
11	0.07	0.07	0.08	0.11	0.11	0.12	0.12	0.12	19.16	19.68	20.01	22.55	22.13	27.63	30.36	37.31
12	0.08	0.08	0.09	0.10	0.11	0.12	0.12	0.12	19.31	19.47	19.79	22.97	21.72	27.18	29.03	29.99

	CET1 (%)									LVG (%)							
	Quartile (%)									Quartile (%)							
Period	1	5	10	Mean	50	90	95	99		1	5	10	Mean	50	90	95	99
	At the end of the period																
1	0.11	0.11	0.12	0.13	0.12	0.14	0.14	0.15	15.10	15.67	15.69	17.02	17.38	18.11	18.53	18.86	
2	0.10	0.10	0.11	0.12	0.12	0.14	0.14	0.14	16.20	16.35	16.39	18.22	18.34	20.52	21.16	21.85	
3	0.08	0.10	0.10	0.12	0.12	0.14	0.14	0.14	16.47	16.59	16.65	19.24	18.82	21.62	22.08	28.75	
4	0.08	0.09	0.10	0.12	0.13	0.14	0.14	0.14	16.65	16.86	16.90	19.11	18.22	22.18	23.19	29.08	
5	0.08	0.09	0.09	0.12	0.13	0.14	0.14	0.14	16.80	16.94	17.03	19.20	17.93	23.21	24.96	28.90	
6	0.08	0.08	0.09	0.11	0.12	0.13	0.13	0.13	17.11	17.43	17.49	20.41	19.16	24.79	29.16	29.72	
7	0.08	0.08	0.09	0.12	0.12	0.13	0.13	0.13	17.25	17.39	17.70	19.96	18.96	23.59	29.32	29.72	
8	0.07	0.08	0.09	0.11	0.11	0.12	0.12	0.13	18.26	18.74	19.02	21.67	20.79	24.70	29.32	35.70	
9	0.09	0.09	0.10	0.11	0.11	0.13	0.13	0.13	17.78	18.16	18.31	22.10	20.41	23.07	24.33	25.95	
10	0.08	0.08	0.09	0.11	0.11	0.12	0.12	0.13	18.34	18.82	19.05	22.21	21.62	26.51	30.56	32.76	
11	0.07	0.07	0.08	0.11	0.11	0.12	0.12	0.12	19.18	19.71	20.04	22.33	22.15	27.65	31.21	37.41	
12	0.08	0.08	0.09	0.10	0.11	0.12	0.12	0.12	19.34	19.50	19.81	22.98	21.74	27.22 2	8.97	30.02	

Table 4.9: Outcomes evolution across stages and over the simulation horizon

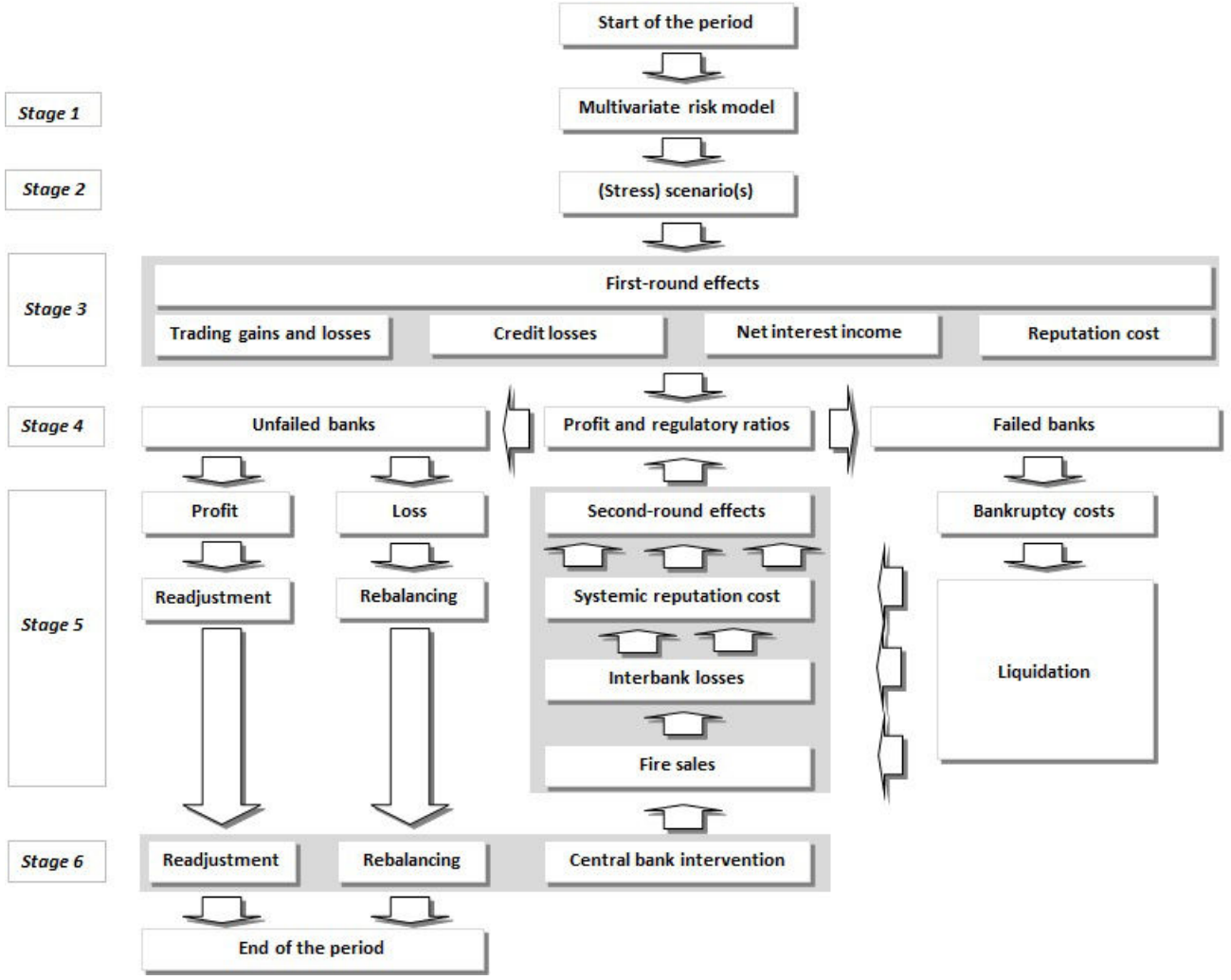

Figure 4.1: Overview of the valuation model's main stages

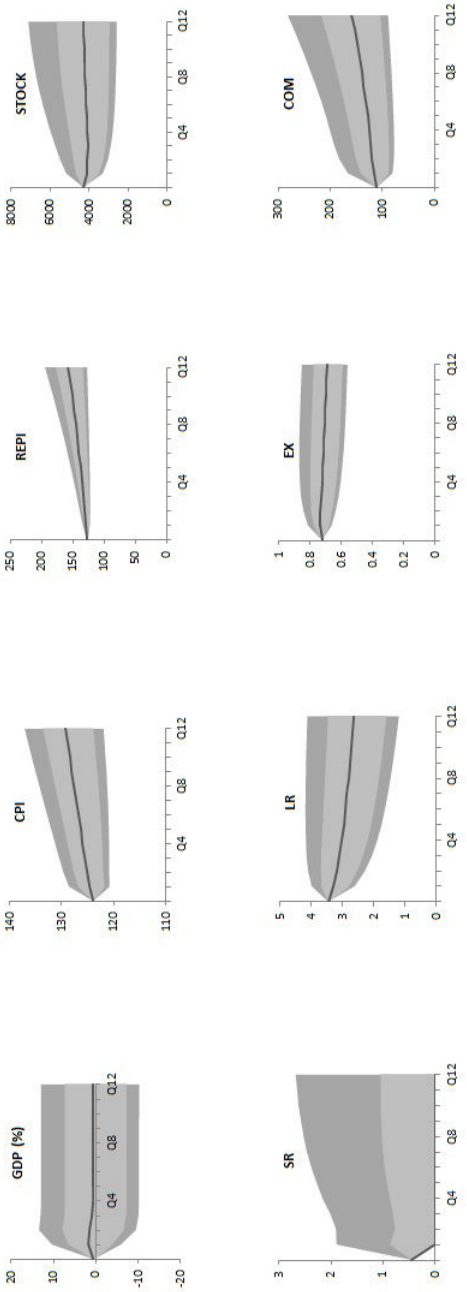

Figure 4.2: Simulated paths of the risk factors over the stress horizon

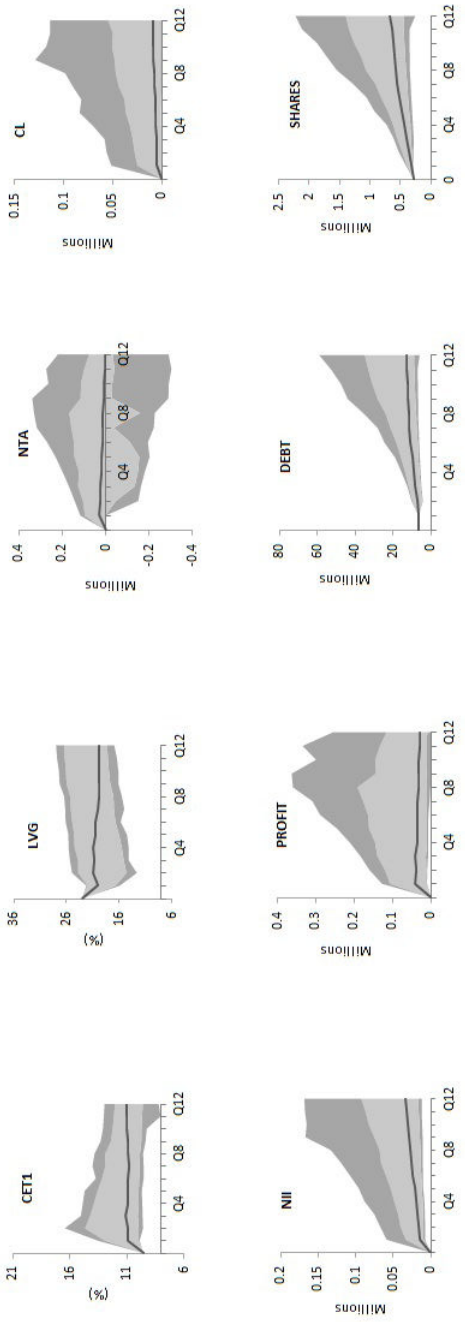

Figure 4.3: First-round effects over the stress horizon

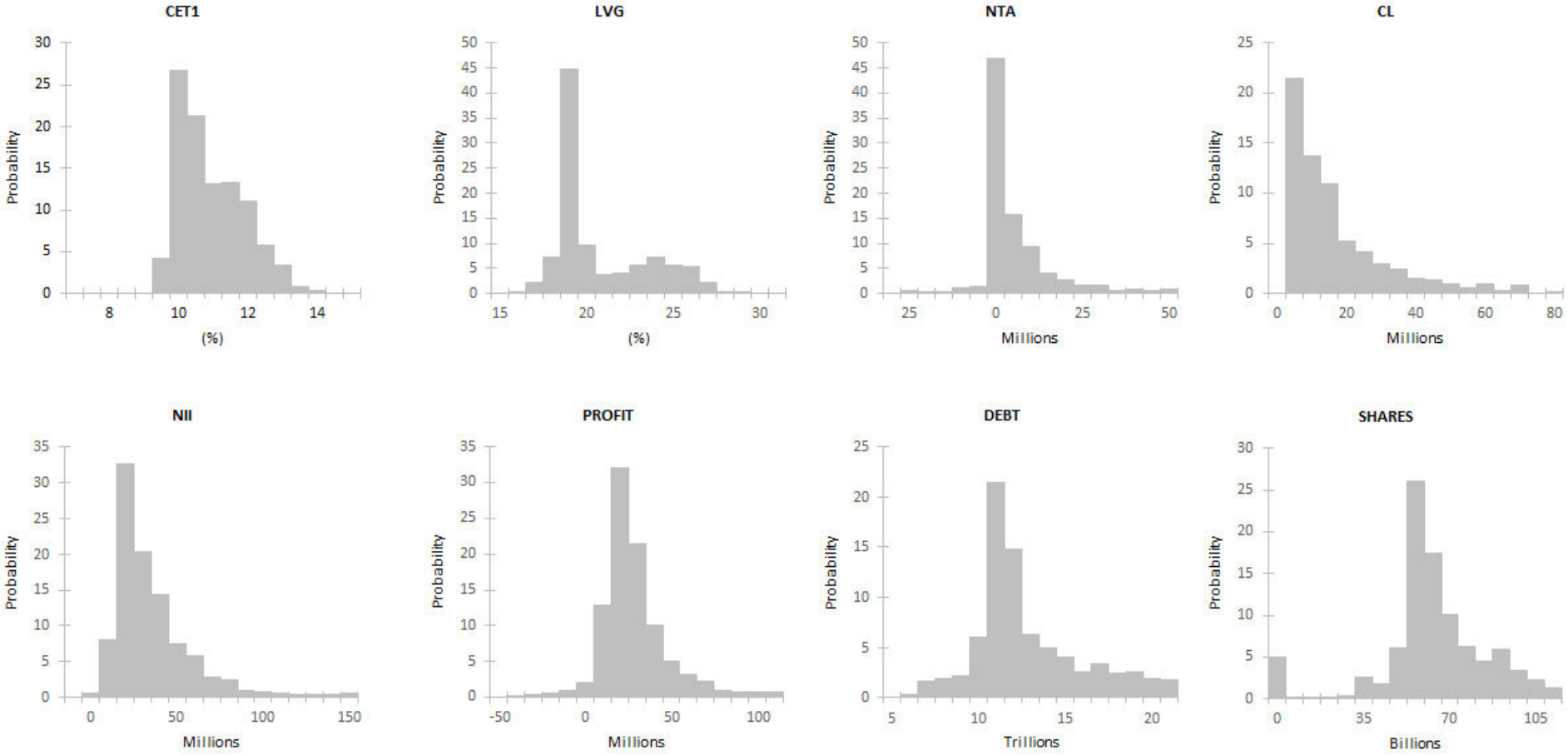

Figure 4.4: First-round effects: distribution of results (last period)

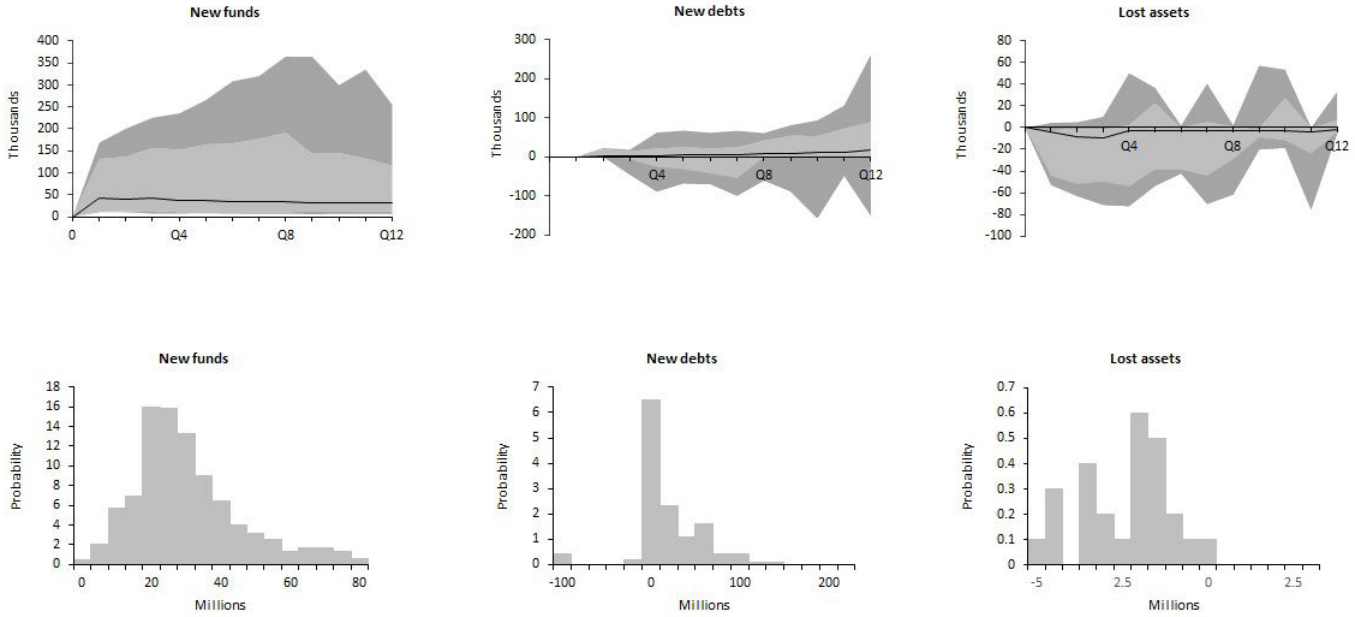

Figure 4.5: Private response functions

This figure shows the dynamics of new funds (profitable banks), new debts (non-profitable banks), and lost assets (failed banks). These indicators are reported for the whole simulation horizon (upper figures) and for the last period (lower figures).

5. Systematic Identification of Stress Scenarios A Bayesian Approach

Abstract

This chapter adapts the Nested Sampling Bayesian approach to carry out a systematic selection of stress testing scenarios. This approach supports scenarios consisting in changes in risk factors' values and uncertainties in the parameters of the underlying risk model. From the estimated posterior distribution of parameters, and given a predefined plausibility threshold, the set of plausible scenarios is bounded using extreme quartiles. The impact of each plausible scenario on the portfolio is then assessed using the Maximum Loss (MaxLoss) measure. This allows identifying the most severe scenario(s) for the considered portfolio and plausibility threshold. This method improves the commonly used subjective and ad hoc methods to set stress scenarios. Moreover, due to their flexibility, Bayesian methods are more suited to do so than alternative methods based on distance metrics. Finally, the stochastic framework of Bayesian methods allows for useful applications of stress testing exercises in risk management (e.g. core risk factors identification, portfolio allocation) and for a more complementarity between stress tests and more conventional risk measures.

Key words: Scenario, Bayesian methods, nested sampling, stress testing, risk management

JEL classification: C32, G11, G17, G21

5.1 Introduction

In a stress testing exercise, only credible scenarios should be designed to assess their potential impact on the stressed portfolio. To be considered as credible, the scenario has to be of severe impact on the portfolio's value while respecting a minimum of plausibility.[1] Choosing among all alternative scenarios is therefore of central interest in these tests. A trade-off between plausibility and severity is necessary to ensure that no severe scenarios have been omitted and only sufficiently plausible scenarios are considered. In practice, however, most realized stress tests rely on ad hoc methods to identify stress scenarios. The latter are set subjectively, based on past stress episodes (historical scenarios) or on some assumptions about future events (hypothetical scenarios). Yet, in doing so, the severity criteria is not always fulfilled. The selected scenario could therefore be less harmful than other possible scenarios of equal or higher plausibility levels. In other words, when setting the scenario upstream, these tests worry more about the scenario itself than about its impact on the portfolio. Moreover, realized scenarios are often not *probabilized*, which makes it difficult to assess their plausibility. These two limitations affect the utility of the stress testing results in several respects.

First, by omitting severe yet plausible scenarios, the results lead to a misidentification of the *true* extreme losses for the portfolio. This produces a false illusion of safety and may lead to wrong risk management decisions. A striking example is the recent problems of the Greek and the Irish banking systems, contrasting with the optimistic conclusions of the stress tests to which they have been submitted (see, e.g., Breuer et al., 2012). On the other hand, considering implausible scenarios may lead to ignore part or all of the test's results. This is particularly the case when the revealed impact is deemed too severe and the suggested measures too expensive. Here, the decision-maker would not assume such a decision, knowing that the considered scenario may later turn out to be unfounded. The results are then simply referred to as a *false alarm*.

Secondly, results of non-probabilized scenarios – also called deterministic scenarios – cannot be included in all decision-making processes. In finance, most decisions are based on stochastic programs performing *optimization under uncertainty* calculations. This is particularly the case in risk management, asset pricing, and investment decisions. However, when the probability of the stress scenario is not set, one cannot assess the relevance of the decisions taken accordingly. This may therefore be inadequate with respect to the *true* probability of the scenario. Hence, unexpected losses could appear if the management action turns to be insufficient – when it is too committed, instead, high opportunity costs are faced.

Finally, stress tests are realized in a casual fashion. Yet, risk management actions are carried out on a daily basis, using more conventional tools such as the VaR. When the two tools coincide, risk managers often face an awkward dilemma: should one continue to rely on VaR's results, valid under normal market conditions, or switch onto stress testing results, only valid for adverse situations? An intuitive answer would be: both. But yet again, the deterministic character of the stress scenarios prevent a combination of the two approaches. As a result, stress tests are often considered independently from the other risk management tools. Despite regulatory recommendations, most realized tests are still based on ad hoc scenarios, which left unexploited their vast potential.

The objective of this chapter is twofold:

[1] See chapter 1 for more details on these concepts.

(i) Presenting a flexible method to identify severe yet plausible stress testing scenarios for nonlinear financial portfolios. The Bayesian theorem, and specifically the Nested Sampling approach (Skilling, 2004, 2006) is used to define the "plausibility domain" containing all plausible scenarios with respect to some predefined "plausibility threshold". A systematic research is then conducted in this domain to identify the most severe scenario(s) for the portfolio. This approach prevents subjective issues in the selection of the scenario and ensures the required trade-off between the severity and the plausibility criteria.

(ii) Presenting a methodological framework of possible applications of the presented method in risk management. We specifically focus on methods allowing to combine stress testing and VaR's results in a unified framework, to identify core risk factors for a given portfolio, and to perform optimal portfolio allocation. We aim this would help risk managers to overcome the main problems caused by deterministic stress scenarios and allow them relying more frequently on stress testing results.

The rest of this chapter is presented as follows. Sections 5.2 and 5.3 present the concept of "the systematic research for stress testing scenarios under uncertainty" and discuss the few approaches encountered in the literature. Section 5.4 presents the Bayesian Nested Sampling approach and the way it can be adapted to systematically identify stress scenarios. Section 5.5 provides an experimental implementation of this approach and its main features. Section 5.6 highlights further applications in risk management purposes. Section 5.7 summarizes and concludes.

5.2 The systematic research of stress scenarios

Severe yet plausible scenarios are required for stress testing exercises. An objective selection of stress scenarios can then be conducted through two alternative approaches: (i) running a systematic research of severe scenarios within a set of plausible scenarios, or (ii) searching for plausible scenarios among all severe scenarios. As the set of severe scenarios is often unbounded, the second approach presents a major issue in practice.[2] The first approach is then usually referred to in the literature, as we shall do in this study. This systematic approach is based on two components, namely a quantitative definition of the severity and the plausibility concepts, and the choice of a "plausibility threshold". While the latter is left to the discretion of the decision-maker, who sets it according to his risk aversion, the literature provides some statistical measures that could be adapted for the severity and the plausibility concepts of stress scenarios. The choice of the measure to adopt often depends on the nature of the scenario.

A stress scenario rely on changes in the statistical patterns of risk factors. Changes in level (e.g. return) are a common approach in practice. However, where being fairly intuitive, these assumptions represent the narrowest definition of a stress scenario. The scenario can also assume changes in the *risk model* underlying the distribution of risk factors. Indeed, in turmoil periods, the volatility and the dependence structure of risk factors are also likely to change, which modifies the underlying distribution (Kim and Finger, 2000; Bee, 2001). This modification may consist in changes in the parameters of the initial distribution, and even changes in the distribution family in cases of structural breaks (i.e. regime switching). Hence,

[2] A related method to carry out the first approach is the so-called *reverse stress testing* exercises. Instead of defining the whole set of severe scenarios, this approach departs from a given impact on the portfolio's value, and searches for the most plausible scenario(s) likely to cause such an impact (see section 5.6.4).

assuming as fixed the distribution over the whole stress horizon is rather unrealistic.[3] Hereafter, scenarios assuming changes in risk factors values are referred to as "pure scenarios", while the term "generalized scenarios" is used to qualify scenarios relying on changes in the underlying risk model. Given the difference in the statistical properties of both approaches, specific statistical measures have to be used to assess their respective severity and plausibility levels. This is particularly true for plausibility, as we shall see below.

The scenario's severity measures its impact on the portfolio. When the impact is given by a single value, such as a loss in the portfolio's return, this value could directly be used as a severity measure. In practice, however, the risk profile of a portfolio is usually given by its distribution function of profits and losses (P&L). To compare alternative P&L distributions, risk managers often use the so-called *portfolio selection models*. These include return-risk models (variance, mean absolute deviation, lower partial moments, central semi-deviations, etc.), risk measures (VaR, CVaR, etc.), expected utility maximization, and stochastic dominance, among others.[4] Risk managers may therefore also rely on one of these methods to assess stress scenarios' severity. To avoid any confusion between the profits and losses distribution of the portfolio and the distribution of risk factors, we refer to the former as the P&L distribution and keep the term distribution for the latter.

The plausibility concept could be compared to that of divergence or dissimilarity between the risk factors' values under the stress scenario and their corresponding values in a baseline scenario (e.g. expected values). Measures of statistical distance could therefore be used for this purpose. For pure scenarios, these measures would be performed in the Euclidian space of risk factors' values. For generalized scenarios, this would be done in the space of higher-dimensional statistical manifolds formed by the risk factors' probability density functions. The distance is measured between two points in the first case, and between two sets or strings in the second. The first set of measures includes, for example, the Euclidian, the Minkowski, the Mahalanobis, the Manathan, and the Tchebychev distances. The concept of distance between two probability distributions has been first introduced by Mahalanobis (1936). Measures such as the Bhattacharyya distance, the Kullback-Leibler divergence, and the Fisher-Rao distance have been proposed afterwards (see Ullah, 1996). Both approaches are widely used in statistics, econometrics, and other practical domains such as image processing and analysis, medical imaging, and remote sensing, among others. However, only few studies have adapted these tools for stress testing purposes. The next section discusses the distance metrics as a method of identifying plausible scenarios, and highlights the main related difficulties. In the following section, we propose an alternative method based on the Bayesian inference and show its main advantages.

5.3 Distance metrics for stress scenario selection

Studer (1997, 1999) and Breuer and Krenn (1999) have presented a first theoretical framework of the systematic research for pure stress scenarios. They consider risk factors represented by the returns of financial assets specified using an elliptical distribution. Stress scenarios are represented by changes in the returns' values. The authors measure the plausibility of a given

[3] Note that a stress scenario may also be designed to consider changes in the loss function (or valuation model) used to assess the impact of risk factors' changes on the stressed portfolio. See chapter 3 for more details on this approach.

[4] See Roman and Mitra (2009) for a detailed review.

scenario by the Mahalanobis distance (MD) between the values of risk factors under the stress scenario and their expected values – corresponding to the centre of the distribution. That is

$$\text{Maha}(r) := \sqrt{[r - \mathbb{E}(r)]^T \Sigma^{-1} [r - \mathbb{E}(r)]}$$

where r is the vector of returns, and $\mathbb{E}(r)$ and Σ are the related mean vector and covariance matrix, respectively.

Scenarios with a MD lower than a threshold distance k are considered as plausible, and form "the plausibility domain" – also called "the trust region". The severity level of these scenarios is assessed by comparing the generated maximum loss (MaxLoss) on the portfolio. This single-value measure is given as follows

$$\text{MaxLoss}(L, k) := \max_{r: Maha(r) \leq k} L(r) - \mathbb{E}(L)$$

where L refers to both the loss function and the P&L distribution.

This optimization problem could be solved analytically if L is linear. Pistovčák and Breuer (2004) have presented a numerical algorithm for nonlinear functions. The solution of this problem is referred to as the "worst case scenario(s)" for the stressed portfolio at the considered threshold k.[5]

This approach presents an important step towards stochastic and more objective stress scenarios. However, the use of the MD makes it less flexible and the results are sometimes inconsistent. Indeed, MD is exclusively designed for elliptical distributions and cannot be used for other specifications. More general risk factors' classes, such as ratings and expert opinions, are also excluded from this framework. Moreover, while based on the first two moments of the elliptical distribution, the MD returns the same plausibility level for identical scenarios proceeding from alternative distributions. Yet, an extreme scenario is intuitively more plausible if it comes from a heavy-tailed distribution (Breuer and Csiszár, 2013). Breuer (2008) has shown that the solution of the previous problem is not low-invariant, as two portfolios might have the same P&L distribution without having the same maximum loss. This solution is also not coordinate-invariant. That is, for a given plausibility level, the maximum loss of a portfolio can vary by adding risk factors to the underlying joint distribution. This is even true when the extra risk factors are unreliable for the portfolio (i.e. with no additional information) or not highly correlated with the existing ones.[6]

McNeil and Smith (2012) have suggested to replace the MD by the half-space depth to form the set of plausible scenarios. This is then used to perform reverse stress testing exercises, consisting in a systematic research for the most plausible scenario(s) for a given value in the portfolio's loss. The proposed approach focuses on bivariate elliptical distributions. Aragonés et al. (2001) and Alexander and Sheedy (2008) measure the plausibility of stress scenarios

[5] Statistically speaking, scenarios of the same plausibility level form ellipsoids of radius equals to the corresponding MD. The greater is this distance, the less plausible is the scenario. Thus, the higher is k, the more the plausibility domain contains unlikely scenarios.

[6] Concretely, Breuer (2008) shows that the plausibility domain defined by Studer (1997) tends to contain more extreme scenarios as the dimension of the vector of risk factors increases. This is explained by the fact that the value of k increases to ensure the stability of the probability of the set $Ell_k := \{r : \text{Maha}(r) \leq k\}$. Breuer (2008) gets around this problem by fixing the value of k whatever is the number of risk factors, and allowing instead the probability to vary (McNeil and Smith, 2012).

through the corresponding probability in the underlying risk factors' distribution. However, both papers focus on univariate scenarios selected on a subjective way.

Breuer and Csiszár (2013) have extended the approach of Studer (1997) to generalized scenarios. They replace the MD by the Kullback-Leibler distance (KLD) – also called I-divergence and relative entropy. Hence, given a threshold distance k, the plausibility domain is formed by risk factors' distributions Q with a KLD lower than k with respect to a distribution of reference.[7] Every scenario's impact is measured in terms of the portfolio's P&L distribution. The most severe scenarios are identified by solving the following problem

$$\text{MaxLoss}(L,k) := \sup_{Q:D(Q||v)\leq k} \mathbb{E}_Q(L)$$

where $D(Q \,||\, v)$ is the KLD of distribution Q with respect to a distribution of reference v.

Breuer and Csiszár (2013) have presented an analytical method to solve this problem and showed that the solution is law-invariant and coordinate-invariant. In addition to the flexibility added by the use of generalized scenarios, the KLD also prevents the main shortcomings of the MD. The new method could then be carried out whatever the risk factors' distribution. Even though, new extension areas are left open due to the used assumptions. First, the KLD is not a metric in the proper sense on the space of probability distributions, as it is not symmetric and does not obey the triangle inequality.[8] The KLD as a plausibility measure can therefore be debated. Second, calculating the KLD involves solving a multidimensional integral over the data space. However, in the literature, there are no closed-form analytic expressions of the KLD for elliptical distributions – except for the degenerate Gaussian case. In the general case, the KLD is approximated through numerical methods, which are fairly cumbersome. This method also lefts open two main questions related to its implementation in stress testing purposes. On the one side, it is unclear from Breuer and Csiszár (2013) how the set Q of possible scenarios is defined. This suggests introducing subjective experts' judgments to do so. On the other side, using the KLD is somewhat unusual for risk managers. The way the KLD radius k can be linked to the more operational concept of probability of the scenario is also unspecified.

To remedy to these issues, we propose to replace distance metrics by less subjective Bayesian methods to select stress testing scenarios. The method and its applications are exposed in the next section.

5.4 Stress scenarios' selection: Bayesian methods

Analyses based on the Bayesian approach have been a growing area in the literature during the last decade. Considerable effort is made to develop efficient and robust methods, mainly used for model selection and parameter estimation. Within the last application, the idea is to define a posterior distribution of the model parameters given a prior distribution. In this case, Bayesian inference can be seen to as a way of considering part of model uncertainty,

[7] The family of this prior distribution (or risk model) is assumed by the risk manager. The related parameter values can be estimated by fitting the model to historical data on risk factors. We often refer to this method as the *maximum entropy* of Q when v is a uniform distribution and as the *relative entropy* of Q with respect to v, otherwise.

[8] A Distance $d : M \times M \to \mathbb{R}$, satisfies three conditions: (i) $d(x,y) \geq 0$ and $d(x,y) = 0$ if and only if $x = y$, (ii) $d(x,y) = d(y,x)$ (symmetry), and (iii) $d(x,z) \leq d(x,y) + d(y,z)$ (triangle inequality).

namely *parameter uncertainty*. That is, the method allows identifying the model risk related to changes in the model parameter values. This definition of model risk matches enough with that of generalized scenarios discussed above. Indeed, for a given model, the set of possible scenarios can be defined by the space of parameter values. Hence, each scenario is represented by a vector of parameters. Plausible scenarios within the parameter space can then be defined using adapted quantitative criteria – e.g. quartiles of the posterior distribution.

Using Bayesian inference to define stress scenarios presents several advantages over distance metrics measures. First, the use of general nonparametric methods prevents shortcomings related to specific or subjective choices of a particular distance measure, among which no classification criteria are provided in the literature. Second, the set of possible scenarios Q is now given by the space of parameter values of the assumed model. This allows preventing subjective considerations to do so. Finally, the plausibility of a given scenario is now directly measured from the posterior distribution of the corresponding vectors of parameter values. This prevents issues stemming from the use of intermediate plausibility measures, such as the MD radius.

However, Bayesian methods are time consuming; which can be a serious issue for complex models. Indeed, MCMC methods can be computationally intensive and often experience problems in sampling efficiently from a multimodal posterior distribution or one with large (curving) degeneracy between parameters. This is particularly the case for high dimensional models. Skilling (2004, 2006) have presented a less intensive MC method, called Nested Sampling, allowing to reduce computational issues of the most general methods. Successively, Feroz and Hobson (2008) and Feroz et al. (2009a,b, 2014) have proposed computational algorithms of this approach and proved its better accuracy over the existing MCMC techniques. In this study, we adopt this method to sample the posterior distribution of parameters for a stress testing risk model.

Assuming the sampled vectors of parameters as scenarios, we rely on a straightforward method to carry out a systematic selection of the most severe scenario(s) for a given portfolio. We successively present the Nested Sampling approach to form the posterior distribution of parameters and the research procedure for the worst case scenario(s).

5.4.1 Nested sampling parameter inference

Parameter inference consists in inferring the probability of parameters Θ, for risk model $\mathcal{M}$, given some data $\mathcal{D}$. To do so, Bayesian inference relies on the Bayes' theorem which states

$$\mathbb{P}\mathrm{r}(\Theta|\mathcal{D},\mathcal{M}) = \frac{\mathbb{P}\mathrm{r}(\mathcal{D}|\Theta|\mathcal{M})\mathbb{P}\mathrm{r}(\Theta|\mathcal{M})}{\mathbb{P}\mathrm{r}(\mathcal{D}|\mathcal{M})} \tag{5.1}$$

where $\mathbb{P}\mathrm{r}(\Theta|\mathcal{D},\mathcal{M}) = \mathcal{P}(\Theta)$ is the posterior probability, $\mathbb{P}\mathrm{r}(\Theta|\mathcal{D},\mathcal{M}) = \mathcal{L}(\Theta)$ is the likelihood, $\mathbb{P}\mathrm{r}(\Theta|\mathcal{M}) = \pi(\Theta)$ is the prior probability, and $\mathbb{P}\mathrm{r}(\mathcal{D}|\mathcal{M}) = \mathcal{Z}(\Theta)$ is the Bayesian evidence (Sivia and Skilling, 2006). The latter, is the factor required to normalize the posterior over parameter space. It can thus be computed by marginalizing the likelihood $\mathcal{L}(\Theta)$ over the parameter space,

$$\mathcal{Z} = \int \mathcal{L}(\Theta)\pi(\Theta)\mathrm{d}^D\Theta \tag{5.2}$$

where D is the dimensionality of the parameter space. For most applications, this quantity has to be estimated through MCMC methods. However, this can be computationally intensive as it needs high-dimensional numerical integration. To overcome this issue, Skilling (2006) showed that the Bayesian evidence can be calculated by transforming the multi-dimensional integral over the parameter space into a one-dimensional integral over the likelihood space. This can be done by considering a survival function $X(\lambda)$ – called *prior volume* or *prior mass*. This function is the proportion of the prior with likelihood greater than λ. That is

$$X(\lambda) = \int_{\mathcal{L}(\Theta)>\lambda} \pi(\Theta)\mathrm{d}^D\Theta \tag{5.3}$$

Performing an integration by parts, the evidence can then be given by

$$\mathcal{Z} = \int_0^\infty X(\lambda)\mathrm{d}\lambda \tag{5.4}$$

When $\mathcal{L}(\Theta)$, the inverse of $X(\lambda)$ exists, the evidence integral may thus be further expressed as

$$\mathcal{Z} = \int_0^1 \mathcal{L}(X)\mathrm{d}X \tag{5.5}$$

where $\mathcal{L}(X(\lambda)) \equiv \lambda$. The sampling algorithm proceeds as follows (Feroz et al., 2014):

1. Sample the prior n times to generate an active set of objects $\Theta_1, \ldots, \Theta_n$, and calculate each object's likelihood.
2. Sort the objects based on likelihood.
3. Remove the point with lowest likelihood ($\mathcal{L}^*$) from active set, leaving $n-1$ active samples.
4. Generate a new sample point from the prior, subject to the likelihood constraint $\mathcal{L}(\Theta) > \mathcal{L}^*$.
5. Add the new sample Θ_{new} to the active set to return the set to n objects.
6. Repeat steps 2 to 5 until termination.

   With the generated samples, the evidence can be approximated numerically using standard quadrature methods, as follows

$$\mathcal{Z} \approx \hat{\mathcal{Z}} = \sum_{i=1}^{N} \mathcal{L}_i w_i \tag{5.6}$$

where $w_i = 1/2(X_{i-1} - X_{i+1})$ is the width between successive sample points, and N is the total number of samples $0 < X_N < \cdots < X_2 < X_1 < X_0 = 1$ (including discarded objects from the active set and the remaining objects at termination). For a more detailed presentation on the Nested Sampling approach and the related computational algorithm, see Feroz et al. (2014) and the references therein.

5.4.2 Worst case scenario selection

The method presented above allows defining the posterior parameter distribution of model $\mathcal{M}$. The space of sampled parameters is assumed to be the set of all possible scenarios. In other words, scenarios consist in alternative parameterizations of model $\mathcal{M}$. If we consider a portfolio exposed to risk factors specified by $\mathcal{M}$, the systematic research for the worst case scenario(s) aims to identify scenario(s) causing the most severe impact on the portfolio. This research is carried out as follows. First, scenarios are ordered according to their plausibility. This can be done with respect to some adapted quantitative criterion. Here, we use the quartiles of the posterior distribution to set the probability of each scenario. Scenarios with higher probabilities are the most plausible. Second, we set a plausibility threshold and define all scenarios with a greater probability. These form the plausibility domain. Third, for each scenario in the plausibility domain, we perform a simulation of model $\mathcal{M}$ to get the corresponding expected values of risk factors. This can be done in two ways: an unconditional simulation relying only on the values of the model parameters, or a conditional simulation which also includes past values of risk factors. The second approach is similar to forecasting exercises. Fourth, using the appropriate loss function, measure, for each scenario, the portfolio's values resulting from all simulated risk factors. A set of portfolio's values is then obtained for each scenario. These could be used to form the empirical P&L distribution for each scenario. Finally, the worst case scenario is identified by comparing the obtained P&L distributions, using a portfolio selection model. To do so, we use the MaxLoss criteria which is well adapted to the spirit of stress tests.

Hereafter, we provide a formal description of the five stages of the systematic research procedure. To avoid duplication, the first two stages are put together.

(i) Plausibility domain definition

Set the plausibility threshold α and the corresponding values in the vector of parameters Θ. To do so, we discard the most extreme values of each parameter in the underlying multivariate empirical cumulative distribution function.

The plausibility domain $\mathcal{S}$ is therefore defined as follows

$$\mathcal{S} = \left\{ \Theta_d : \mathbb{P}\mathrm{r}[\Theta_d \leq \Theta_{d_min}] = \frac{\alpha}{2} \wedge \mathbb{P}\mathrm{r}[\Theta_d \geq \Theta_{d_max}] = \frac{\alpha}{2} \right\} \tag{5.7}$$

where Θ_{d_min} and Θ_{d_max} are, respectively, the $\frac{\alpha}{2}$ upper and $1 - \frac{\alpha}{2}$ lower quartiles in the marginal cumulative distribution function of parameter d (with $d = 1, \ldots, D$). Therefore, $\mathcal{S}$ contains all combinations of models whose parameter values fall inside the plausible parameter interval $[\Theta_{min}, \Theta_{max}]$.

(ii) Discretization

Discretize the continuous plausible interval for each parameter $[\Theta_{d_min}, \Theta_{d_max}]$, through varying the parameter value by a small increment ε_d. Let S be the resulting number of values for each parameter. This leads to S plausible parameterizations or generalized scenarios.

For each parameterization, get N simulations of model $\mathcal{M}$. By doing so, NS vectors of risk factors (or pure scenarios) are obtained. Note r_{ij} these vectors, with $i = 1, \ldots, S$ and $j = 1, \ldots, N$.

(iii) Portfolio's reevaluation

For each vector r_{ij}, measure the corresponding impact on the stressed portfolio, as follows

$$V_{ij} = wL(r_{ij}) \tag{5.8}$$

where V_{ij} is the portfolio's value resulting from the r_{ij} risk factor's move, L is the loss function, and w is the vector of investment weights.

Then, form the P&L distributions corresponding to the S generalized scenarios, as follows

$$\mathcal{L}_j(r, L) \equiv V_{ij} \tag{5.9}$$

(iv) Identification of the worst case scenario(s)

The worst case scenario is defined by comparing the P&L distributions obtained under all S plausible scenarios. Relying on the MaxLoss criteria to do so, the worst case scenario is finally given by

$$\text{MaxLoss}(L, \alpha) := \sup_{\Theta:\Theta\in[\Theta_{min},\Theta_{max}]} \mathbb{E}_\Theta(\mathcal{L}) \tag{5.10}$$

5.5 Experimental results

We present a simple numerical example to illustrate the implementation of the presented method in practice. The CCPFOrge website contains a set of software packages to run the Nested Sampling algorithm. In this section, we have used an augmented version of the Matlab MultiNest package designed by Pitkin and Romano.

Consider a nonlinear portfolio made of a currency-translated option. Precisely, the portfolio is a short position on a European call denominated in euros, and the underlying asset is the S&P500 index (SP) denominated in U.S. dollars (USD). If we note K the strike of the option and EX the nominal exchange rate euro/USD, the current value of the underlying asset is X=SP·EX euros. The portfolio is subject to two risk factors: the equity index and the exchange rate. The option is out-of-the money if K > X and is in-the-money if K < X.

The pay-off of an investor holding a long position on this option is: $c = \max(\text{X} - \text{K}, 0)$ euros. Hence, up to maturity, the portfolio's value V is given by

$$\text{V} = L(\text{SP}, \text{EX}|\text{K}) = -c = -\max(\text{SP} \cdot \text{EX} - \text{K}, 0)$$

Figure 5.1 shows the pay-off of the portfolio, as a function of the two underlying risk factors.

To estimate the (expected) values of the equity index and the exchange rate, we have modelled their joint dynamics using a bivariate Bayesian VAR (BVAR) model. We specifically assume a BVAR(1) model with normal homoscedastic errors. The resulting model has then $D = 6$ parameters.

The 2×1 vector of risk factors r_t is specified by

$$r_t = \mu + Ar_{t-1} + \varepsilon_t$$

where μ and A are, respectively, a 2×1 vector and 2×2 matrix of parameters, and $\varepsilon_t \sim \mathcal{N}(0,1)$ is a 2×1 vector of normal white noise.

We have estimated this model on daily data collected over the period $6/1/2013 - 5/31/2014$. Data are from Datastream. To estimate the posterior distributions of the model parameters, we use the Nested Sampling approach based on the core function of the MultiNest package. To adapt the package to our purpose, we have introduced two main modifications on the original code, replacing the model specification and the corresponding likelihood function. Figure 5.2 presents a matrix plot of the estimation results. The Model parameters' posterior distributions and pairwise dependence structures are reported. We note the relatively normal-shaped distributions of the estimated parameters. Indeed, the related skewness coefficients range from -0.042 to 0.055, while kurtosis range from 2.84 to 3.09. These values show the consistency of the Nested Sampling as an estimation method. The pairwise correlation are relatively low, except for two pairs of parameters where it is negative and below 0.7. The common feature of these correlations is that they rely, in both of the BVAR equations, on the constant and the lagged value of the non-dependent variable.

We now perform a systematic research of the worst case scenario(s) for the portfolio. Following the method presented in section 5.4.2, this research is carried out as follows. First, we consider all combinations of parameter values (given by the estimation results) as possible scenarios. We set the plausibility threshold at 5% and determine, for each parameter distribution, the 2.5% and 97.5% quartiles. This allows defining the plausibility domain at 95% confidence level. Table 5.1 reports the obtained results for each parameter.

We set the discretization increment to $\varepsilon_d = \frac{(\Theta_{d_max} - \Theta_{d_min})}{10^3}$. This leads to consider 10^3 plausible values for each parameter and $S = (10^3)^6$ multivariate scenarios or parameterizations. Note, however, that this optional step is proposed to improve the accuracy of the research procedure. Indeed, we can use only the sampled parameters as scenarios. This allows reducing the time of calculation for complex models.

Given the last observation of the two considered risk factors, we estimate their expected one-period ahead values for each scenario. Here, we set $N = 1$, meaning that a point forecast is performed. We do so for simplicity reasons and because of the use of MaxLoss to compare scenarios' severity levels.[9] Accordingly, $SN = (10^3)^6$ expected risk factor values are obtained. Figure 5.3 reports their corresponding empirical distributions.

Given the portfolio's loss function, the last step in the procedure consists in reassessing the portfolio's value for each scenario, and determining the worst case scenario. The obtained results are reported on Table 5.2.

The first six columns report the values of the parameters under the worst case scenario. These correspond to the worst parameterization of the BVAR model for the considered scenario. By comparing these values to those of the posterior distributions' quartiles (see Table 5.1), we can note that the worst case parameter values are not necessarily the most extreme values. This fact denotes the importance of considering a systematic research of stress scenarios relying on severe scenarios rather than considering extreme scenarios. The last concept is indeed irrelevant for nonlinear portfolios. This remark also holds for the risk factors' values under the worst case scenario.

[9] Indeed, the MaxLoss criteria requires a unique value in the portfolio's distribution (e.g. mean or median) under each scenario to perform the comparison. When the whole distributions is needed for the comparison (e.g. under the stochastic dominance criterion), one can run a numerical simulation of the model to get the desired number of expected realizations.

5.6 Stress tests in risk management

In this section, we propose four main applications in risk management of stress testing exercises based on stochastic scenarios as described above.

5.6.1 A unified framework for stress tests and VaR

Risk managers often rely on a limited set of statistical tools to set investment and hedging strategies related to financial portfolios. When these tools are proven to be consistent and supported by decision-makers, it becomes a tricky task to remove the risk management routine inside the institution. When additional tools are proposed, these are usually seen to as casual and independent exercises which should not disrupt the orderly functioning of the internal practices. If no guidelines are provided for the use of these tools, the related results are more likely to be ignored. When internal or external (e.g. regulatory) constraints require some reaction on the basis of these new tools, this is usually done in a suboptimal fashion. The results of stress tests fall in this area. Risk managers often wonder how to combine these results with those of more conventional tools, such as the VaR. Should they continue to rely on VaR's results, valid under normal market conditions, or switch onto stress testing results, only valid for some adverse situations? Berkowitz (2000) was among the first to present a unified framework allowing to combine the stress testing and the VaR's results. This approach is carried out in the following steps:

1. Set a probability to each stress scenario, which is also the probability of the related results.
2. Using the appropriate valuation model, form the P&L distributions of the portfolio under each stress scenario and under normal market conditions in the VaR framework
3. Set the unified P&L *meta-distribution* using distribution mixture methods, where the mixing weights are given by the probabilities of the VaR and the stress scenarios.
4. Deduce different risk measures (e.g. Stress-VaR, Stress-CVaR) from the unified distribution, which is based on two management tools: stress tests and a traditional risk management tool.

Aragonés et al. (2001) have presented an illustrative application of this approach. They have used a portfolio of commodities, with a single risk factor and a single stress scenario. The Extreme Value Theory (EVT) is used as a risk model, while the mixing probabilities are set subjectively. Despite the usefulness of this approach, Aragonés et al. (2001) is still the unique related study in the literature. The main reason of this is given by the subjective nature of the considered scenarios and the related probabilities. Relying on objective methods based on stochastic and systematic identification of stress scenarios, as shown in the previous section, favours a more adapted and frequent use of the stress testing results.

If we set $\mathcal{L}_0(r, L; \psi_0)$ and $\mathcal{L}_\alpha(r, L; \psi_\alpha)$ the P&L distributions, respectively, under the baseline scenario and the most severe scenario with a plausibility threshold α, we can form the aggregated P&L distribution $\mathcal{L}$ as follows

$$\mathcal{L}(r, L; \psi) = \begin{cases} \mathcal{L}_0(r, L; \psi_0) & \text{with probability } \alpha \\ \mathcal{L}_\alpha(r, L; \psi_\alpha) & \text{with probability } 1 - \alpha \end{cases} \tag{5.11}$$

This distribution combines therefore the risk model used for the stress scenario and the baseline risk model carried out in a daily basis by risk managers. Risk measures can be drawn accordingly. These have the advantage of taking account of both normal and stress market conditions.

5.6.2 Identification of core risk factors

Here, we present a practical approach to analyse the contribution of the risk factors in the most severe scenario. This approach allows: (i) identifying the *core* factors having the highest impact on the portfolio's value, (ii) analysing the evolution of this contribution over calm and stressful periods, and (iii) investigating whether the core factors in the most severe scenario are also those that contribute most to the portfolio's return. Such an analysis is the basis of asset allocation decisions, based on an optimization of the return-risk trade-off. It also allows removing *peripheral* risk factors from complex analyses (e.g. optimization and identification of stress scenarios) and from the reporting of results to ease communication (Pistovčák and Breuer, 2004). For presentation purposes, we simply use the term "contribution" to refer to the contribution of risk factors to the portfolio's value under the most severe scenario at a given plausibility threshold.

To measure the contribution of a small group of n risk factors among the N portfolio's underlying risk factors, we form partial scenarios. In each scenario, the n factors take their corresponding values in the most severe scenario while the remaining $N-n$ factors take their expected values. By doing so, n^2+1 partial scenarios of $i=1,\ldots,n$ risk factors could be formed. This approach is fairly close to the concepts of *marginal VaR* and *partial VaR*, commonly used in practice. The individual and joint contributions of the n risk factors are measured by comparing the portfolio's P&L distribution under the most severe scenario with the P&L distributions resulting from each partial scenario. This comparison can be made by dividing some statistical moment (e.g. mean, median, variance, etc.) of both P&L distributions. The resulting ratio is called "the contribution ratio" of the risk factor(s) included in the partial scenario. We can therefore identify core risk factors presenting the most individual and/or joint contribution to the portfolio's return, the portfolio's volatility, etc. This helps identifying the portfolio's main vulnerabilities during stress market conditions, and assists decision-making on asset allocation (see below).

The joint contribution of a group of risk factors can be higher, less or equal to the sum of their individual contributions, depending on whether these factors are positively, negatively or not correlated. Thus, a given loss cannot be explained by an addition of individual impacts of different factors. Moreover, the joint contribution of a group of factors can be higher than that of a second group, even when the individual contributions are higher in the latter. Thus, the factors of high individual contributions are not necessarily the core risk factors of the portfolio (Breuer et al., 2002).

5.6.3 Portfolio allocation

This section extends a previous work by Breuer (2006) who has presented an ad hoc approach for asset allocation, based on the MaxLoss measure discussed above. Breuer (2006) assumes financial asset returns as risk factors. For each asset, he measures the historical return and the individual contribution to the portfolio's MaxLoss under the worst case scenario. The ratios of

return to contribution are then deduced. Finally, the author performs a manual reallocation by reducing (resp. increasing) the weight of the assets that present a lower (resp. higher) return-to-contribution ratio.

This approach has three main drawbacks. First, the MaxLoss cannot be considered as a risk measure, as it represents the mean of the P&L distribution. Then, the proposed allocation is rather based on a return-return trade-off rather than a return-risk one. Second, the allocation procedure is carried out on a subjective and deterministic basis. When it improves the actual allocation, it does not ensure the optimal allocation as it lacks of a stochastic program (SP) to consider all reallocation alternatives. Third, by changing the asset weights in the portfolio, the resulting MaxLoss is also likely to evolve. Yet the allocation decision is based on a fixed MaxLoss corresponding to the actual weighting of the portfolio. By doing so, the new portfolio will present a new worst case scenario which may be more severe than that of the initial portfolio.

To overcome the first limit, we use an allocation decision rule based on the ratio of historical return means to a coherent risk measure. For the latter, we consider the CVaR drawn from the aggregated P&L distribution (section 5.6.1). By doing so, we also discard the third limit, as now the decision in not only based on the stress scenario but also on more likely events. We finally propose to search for the optimal asset allocation through a more general systematic framework. This is carried out within the set of plausible scenarios S. Using the approach presented in section 5.4 and the concepts of sections 5.6.1 and 5.6.2, the procedure is run in the following steps:

1. Consider a sufficiently high number of $N \times 1$ vectors of weights in the set of possible allocations $w \in [w_{min}, w_{max}]$.

2. For each vector in the set of possible allocations:

   (i) Form the P&L distributions for the S plausible scenarios, using the valuation model L.

   (ii) Given the baseline P&L distribution, form the S corresponding aggregated P&L distributions $\mathcal{L}_i$ (with $i = 1, \ldots, S$).

   (iii) Draw the CVaR measure at a predefined confidence level.

   (iv) Given the vector of weights, measure the portfolio's average return using the historical means of the underlying risk factors.

   (v) Deduce the average return-to-CVaR ratio.

3. Compare the resulting ratios and select the allocation leading to the highest ratio.

5.6.4 Reverse stress tests

Instead of searching the plausible scenario(s) leading to the most severe impact on the portfolio, a reverse stress test consists in searching the most plausible scenario(s) likely to cause a given impact on the portfolio. The use of reverse stress tests has been recommended by most international regulations with the same insistence than they have done for the more classical stress tests – referred to as *forward stress tests*, by comparison. However, only few applications have been proposed for this concept. McNeil and Smith (2012) is a rare example in

this sense. Assuming a given portfolio's value, they use the half-space depth technique to draw plausible corresponding scenarios. The *Least solvent likely event* (LSLE) and the *most likely ruin event* (MLRE) are mainly identified. The systematic method presented in section 5.4 can also be used to carry out reverse stress testing exercises.

Let $\mathcal{S}$ be the set of plausible scenarios with respect to the plausibility threshold α. The most plausible scenario x_{MP} for a given portfolio's value $\bar{V}$ is identified by solving the following problem

$$\mathrm{x}_{\mathrm{MP}} := \min_{r:r\in\mathcal{S}} |L(r) - \bar{V}| \tag{5.12}$$

A systematic research for the most plausible scenario can then be run over the generated NS pure scenarios. The larger is this set, the more accurate is the solution of this problem.

5.7 Conclusion

In this chapter, we have proposed a method allowing to adapt the Bayesian approach to carry out a systematic selection of stress testing scenarios. Given a model of risk factors, the scenarios are given by the space of sampled parameters. This method allows considering generalized scenarios, which also support parameter uncertainty. Given the space of sampled parameters, we identify plausible scenarios by discarding the most extreme values of the posterior distribution. This is done with respect to some predefined plausibility threshold, represented by the extreme (lower and upper) quartiles in this distribution. We then simulate the risk model for the remaining parameterizations. By assessing the portfolio's value for each simulated vector of risk factors, we define the profits and losses (P&L) distribution under each scenario. We compare these distributions using the Maximum Loss (MaxLoss) criteria to identify the worst case scenario for the portfolio, for the considered plausibility threshold.

Using Bayesian methods to set stress scenarios has several advantages over some distance metrics methods recently proposed in the literature. First, using nonparametric methods prevents the shortcomings related to any subjective choice of a distance measure. Second, generalized scenarios are given by the space of model parameters. This prevents any subjective considerations to set the scenarios. Finally, the plausibility levels of the scenarios are directly measured from the posterior probability distribution. This allows dealing with stochastic scenarios, which is of great interest for quantitative risk management.

To reduce the time consumption commonly associated with Bayesian methods, we have used the Nested Sampling approach first proposed by Skilling (2004, 2006). This approach also proved to be of a better accuracy compared to the existing MCMC techniques. This feature reduces the implementation costs of the proposed systematic research for stress scenarios and eases its repetition in a high frequency routine (e.g. daily). With this respect, we have also presented a methodological framework of the main applications of this method for risk management. This includes methods allowing to combine stress testing and VaR's results in a unified framework, identify core risk factors for a given portfolio, perform optimal portfolio allocation, and run reverse stress tests. We hope this would help risk managers to overcome the main problems caused by deterministic stress scenarios, and allow them relying more frequently on stress testing exercises.

References

Alexander, C. and E. Sheedy (2008): Developing a Stress Testing Framework Based on Market Risk Models, *Journal of Banking and Finance*, 32(10), 2220–36.

Aragonés, J.R., C. Blanco, and K. Dowd (2001): Incorporating Stress Tests into Market Risk Modeling, *Derivative Quartely*, 7(3), 44–9.

Bee, M. (2001): Mixture Models for VaR and Stress Testing, *ALEA Tech Reports*, Università di Trento, No. 12.

Berkowitz, J. (2000): A Coherent Framework for Stress-Testing, *Journal of Risk*, 2(2), 5–15.

Breuer, T. and I. Csiszár (2013): Systematic Stress Tests with Entropic Plausibility Constraints, *Journal of Banking and Finance*, 37(5), 1552–9.

Breuer, T., M. Jandačka, J. Mencía, and M. Summer (2012): A Systematic Approach to Multi-Period Stress Testing of Portfolio Credit Risk, *Journal of Banking and Finance*, 36(2), 332–40.

Breuer, T. (2008): Overcoming Dimensional Dependence of Worst Case Scenarios and Maximum Loss, *Journal of Risk*, 11(1), 79–92.

Breuer, T. (2006): Providing Against the Worst: Risk Capital for Worst Case Scenarios, *Managerial Finance*, 32(9), 716–30.

Breuer, T., Krenn, G., and F. Pistovčák (2002): Stress Tests, Maximum Loss, and Value at Risk, in: B. Britzelmaier, S. Geberl, M. Menichetti, and H.-R. Kaufmann (Eds.), Regulierung oder Deregulierung der Finanzmärkte: Strategieimplikationen für Finanzdienstleister, *Physica Verlag*, Berlin, 301–14.

Breuer, T. and G. Krenn (1999): Stress Testing: Guidelines on Market Risk, Volume 5, Oesterreichische Nationalbank.

Csiszár, I. (1963): Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, *8* (1963), 85–108.

Csiszár, I. (1967): Information-Type Measures of Difference of Probability Distributions and Indirect Observations. *Studia Scientiarum Mathematicarum Hungarica*, *2* (1967), 299–318.

Feroz, F., M.P. Hobson, E. Cameron, and A.N. Pettitt (2014): Importance Nested Sampling and the MULTINEST Algorithm, *MNRA*, forthcoming.

Feroz, F., M.P. Hobson, and M. Bridges (2009): MultiNest: An Efficient and Robust Bayesian Inference Tool for Cosmology and Particle Physics, *MNRAS*, 398, 1601–14.

Feroz, F., J.R. Gair, M.P. Hobson, and E.K. Porter (2009): Use of the MultiNest Algorithm for Gravitational Wave Data Analysis, *Classical and Quantum Gravity*, 26(21), 215003-19(17).

Feroz, F. and M.P. Hobson (2008): Multimodal Nested Sampling: An Efficient and Robust Alternative to Markov Chain Monte Carlo Methods for Astronomical Data Analyses, *MNRAS*, 384, 449–63.

Kim, J. and C.C. Finger (2000): A Stress Test to Incorporate Correlation Breakdown, *RiskMetrics Journal*, 1, 61–75.

Mahalanobis, P.C. (1936): On the Generalized Distance in Statistics, *Proceedings of the National Institute of Sciences*, India, 2(1), 49–55.

McNeil, A.J. and A.D. Smith (2012): Multivariate Stress Scenarios and Solvency, *Insurance: Mathematics and Economics*, 50(3), 299–308.

Pistovčák, F. and T. Breuer (2004): Using Quasi Monte Carlo Scenarios in Risk Management, in: H. Niederreiter (Eds.), Monte Carlo and Quasi-Monte Carlo-Methods 2002, *Springer*, Berlin, 379-92.

Roman, D. and G. Mitra (2009): Portfolio Selection Models: A Review and New Directions, *Wilmott Journal*, 1(2), 69–85.

Sivia, D. and J. Skilling (2006): Data Analysis: A Bayesian Tutorial, *Oxford University Press*, New York.

Skilling, J. (2006): Nested Sampling for General Bayesian Computation, *Bayesian Analysis*, 1(4), 833–60.

Skilling, J. (2004): Nested Sampling, in: R. Fischer, R. Preuss, U.V. Toussaint (Eds.), American Institute of Physics Conference Series, 395–405.

Studer, G. (1999): Market Risk Computation for Nonlinear Portfolios, *Journal of Risk*, 1(4), 33–53.

Studer, G. (1997): Maximum Loss for Measurement of Market Risk, Ph.D. Thesis, Swiss Federal Institute of Technology, Zürich.

Ullah, A. (1996): Entropy, Divergence and Distance Measures with Econometric Applications, *Journal of Statistical Planning and Inference*, 49 (1), 137–62.

		Quartiles (%)					
Parameter	Mean	1	2.5	5	95	97.5	99
μ_1	0.002	-0.011	-0.009	-0.007	0.011	0.013	0.015
μ_2	-0.003	-0.028	-0.024	-0.021	0.014	0.018	0.022
A_{11}	0.231	-1.931	-1.595	-1.274	1.745	2.083	2.338
A_{12}	0.094	-3.970	-3.283	-2.748	3.206	3.690	4.346
A_{21}	-0.168	-1.589	-1.340	-1.152	0.784	0.969	1.225
A_{22}	-0.107	-2.845	-2.484	-2.103	1.941	2.232	2.664

Table 5.1: Descriptive statistics of the marginal posterior distributions

Parameter values						Risk factor values		Portfolio's value
μ_1	μ_2	A_{11}	A_{12}	A_{21}	A_{22}	SP	EX	
0.009	0.018	-1.120	-2.297	0.370	1.374	1703.712	0.787	-70.260

Table 5.2: Outputs of the worst case scenario

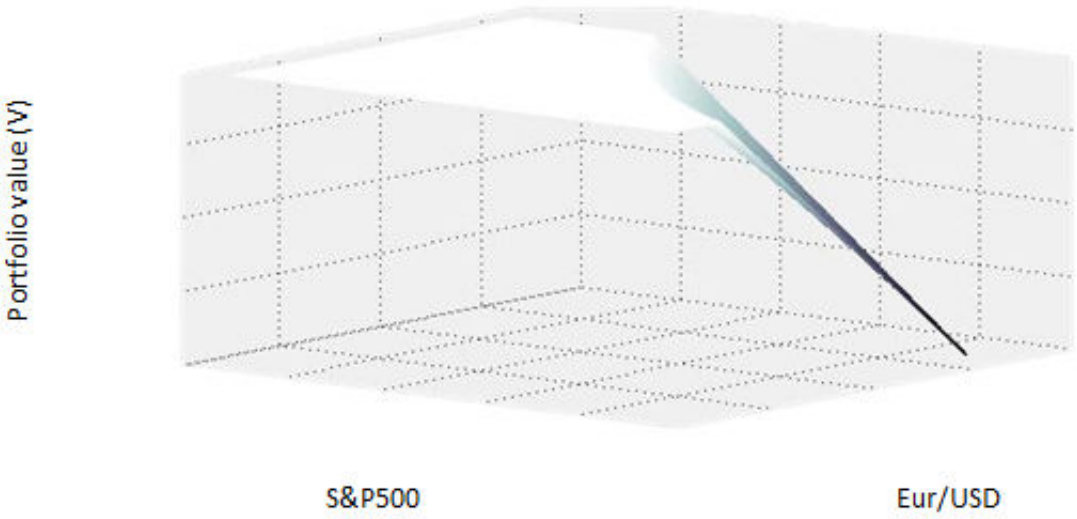

Figure 5.1: The portfolio's value profile

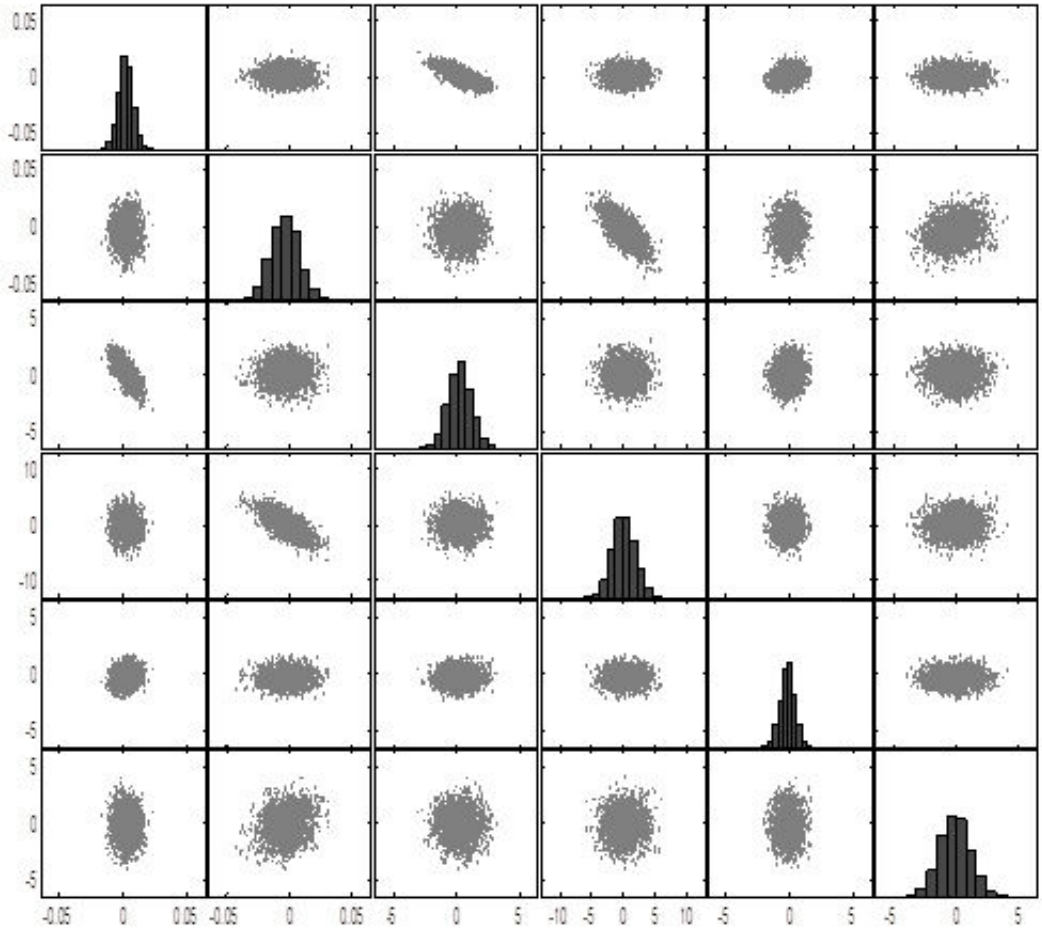

Figure 5.2: Nested Sampling estimation results for the BVAR model

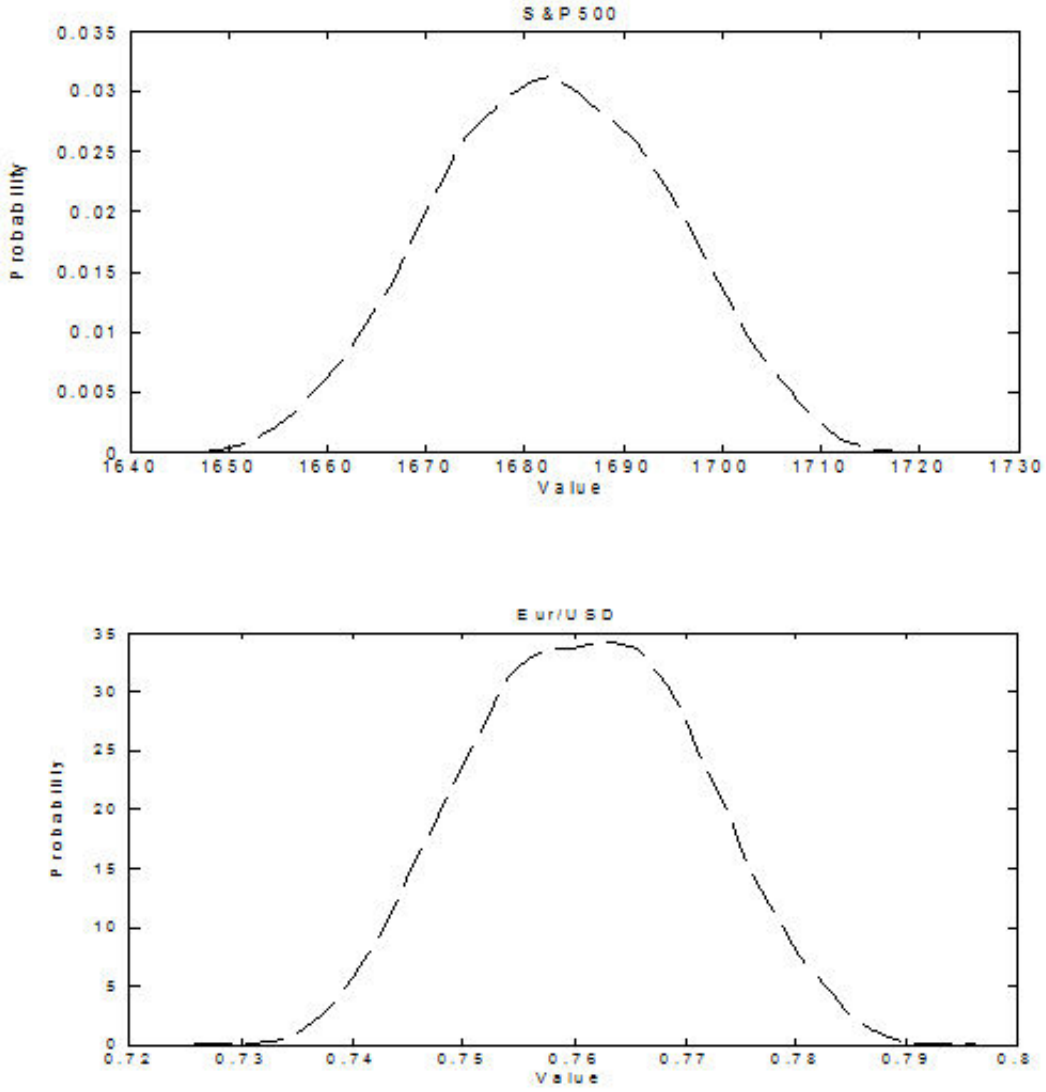

Figure 5.3: Expected empirical density functions for the risk factors

6. Conclusion

In this thesis, we have analysed the conceptual and operational stages of financial stress testing exercises and discussed the main related issues. The exercise is part of a complex and subtle process. It combines a large number of heterogeneous stakeholders, responsible for setting the test's objectives and scope, the stress scenarios, and the outcomes' analysis and use in the decision-making process. Only a deep analysis of the whole process allows determining its relevance. The latter is defined by the stress scenario's severity and plausibility levels (altogether defining the scenario's credibility) and by the utility of the outcomes. The key stage of this process is the definition of the starting assumptions used for the stress scenario. These assumptions relate to the choice of the risk model for the risk factors, the choice of an adapted valuation model to measure their impact on a stressed portfolio, and the definition of credible stress scenarios. These three key points are respectively analysed through the three last chapters of the thesis.

The first chapter has performes a deep analysis of existing stress tests. It has revealed a relative slowness of stress testing practices, due to some limits in the used techniques and the information systems in place in most financial institutions. The recent crises have also confirmed the problems of integrating the tests' results in the risk management process, revealing a lack of internal and external debates on risk factors and investment opportunities, among others. These management issues are augmented by organizational aspects related to the perception of stress tests inside financial institutions, as they are, in most cases, regarded to as routine exercises, taken apart from the other management tools. Moreover, this chapter revealed that most stress testing exercises are still experiencing several conceptual and practical challenges in all stages of the process. These mainly relate to the conception of risk models for the portfolio's underlying risk factors, the selection procedures of credible stress scenarios, and comprehensive valuation models – especially for systemic stress tests.

The second chapter has presented a sequential model to capture the main stylized facts in financial return series. The model is then carried out to specify stress scenarios where the financial returns are taken as risk factors. We have proposed flexible specifications for conditional marginal distributions and for the dependence structure of the return series. This is led using a copula-GARCH model, combining the Extreme Value Theory (EVT), for the univariate processes, and a pair-copulas-based R-vine model, for the multivariate system. The model performance is assessed for three datasets, including equity indices, nominal exchange rates, and commodity prices. The results have shown good in-sample and out-of-sample performances of the proposed model. To check the relevance of these results, we have also compared the static and the dynamic properties of the model with those of commonly used models in practice. The results have supported the better properties of the presented copula-GARCH specification, especially for extreme returns.

The use of the model for a stress testing exercise with hypothetical scenarios has revealed

increasing losses across the test horizon. These losses exceed the regulatory capital requirements two weeks after the initial stress shock. A capital increase becomes then necessary for illiquid portfolios (in stress market conditions). Moreover, the stress testing results have shown significant differences in terms of the corresponding shocks and after-shocks, and in terms of the generated losses on the portfolio, depending on the used risk model. With this respect, the choice of a more flexible and performant specification, such as the presented copula-GARCH model, ensures a better credibility of the stress scenario. The utility of the test's results are enhenced accordingly.

In the following chapter, we have proposed a valuation model allowing to estimate the vulnerability of the French banking system to a set of domestic and global systemic risk factors. The model is structured as a series of sequential modules capturing the risk factors, their direct and indirect impacts on the banks' balance sheets, as well as the nature and the size of private and public response functions after recording the impact of the shock. An ad hoc hypothetical stress scenario is considered, while the results are recorded quarterly over a three-year horizon within a bottom-up approach. The first main result has shown a high vulnerability of the (aggregate) trading portfolio compared to other banking activities. Hence, introducing the trading portfolio brings more volatility in both the solvability and the leverage ratios. With this respect, the results of small banks seem to be more volatile than big ones'. The result also shows that it is the capital buffer more than the profitability probability (i.e. the expected results) that makes a bank likely to go bankrupt. In the French banking system, the risk of failure is mainly presented by two banks in the time of study.

Second-round effects have proven to be relatively limited, even after having cumulated the results over the simulation horizon. However, the introduction of hypothetical stress scenarios has led to an important shift to the left of the distribution of results. The latter also exhibits a bimodal form, due to the concentration of negative results stemming from contagion spirals corresponding to the most extreme realizations of the scenario.

The second main result of this chapter concerns the important – and even vital – role of the central bank's intervention to stop negative contagion spirals. The cases where this non-conventional action has been omitted result in a failure of an important part of the banking system after only two quarters. This public response is much more important than the private response of rescued banks, which does not seem to significantly impact their situation. This result is in line with the reaction of most central banks observed in the wake of the first episodes of the recent financial crisis. An even more extensive modelling of public response functions need to be considered as an extension of this study. This would allow considerably reducing the reaction time and to optimize its content.

In the final chapter, we have adapted the Bayesian approach to perform a systematic selection of credible stress testing generalized scenarios. Given a model of risk factors, the scenarios are given by the space of sampled parameters. From this space, we have identified plausible scenarios by discarding the most extreme values of the posterior distribution of parameters, with respect to some plausibility threshold. We have then simulated the model for the remaining parameterizations, and assessed the portfolio's value for each simulated vector of risk factors. The profits and losses (P&L) distributions are defined according to each scenario. By comparing these distributions through the Maximum Loss (MaxLoss) criteria, we have been able to identify the worst case scenario for the portfolio.

Relying on Bayesian methods to run a systemic selection of stress scenarios has several advantages over some distance metrics also used in this context. First, through nonparametric

Bayesian methods, one avoids criticism related to subjective choices among alternative and performance-varying distance measures. Second, generalized scenarios are given by the space of model parameters, preventing any subjective considerations to set the scenarios. Finally, the plausibility level of each scenario is directly measured from the posterior probability distribution of parameters. This produces the stochastic scenarios required for quantitative risk management. To reduce the time consumption of the Bayesian inference, we have used the Nested Sampling approach, which has proven to perform better than existing MCMC techniques. This method also reduces the costs of implementation of our systematic selection of stress scenarios and allows repeating the procedure in a more frequent basis. To demonstrate this aspect, we have presented a methodological framework allowing to: (i) combine stress testing and VaR's results in a unified framework, (ii) identify core risk factors for a given portfolio, (iii) perform stochastic portfolio allocation, and (iii) run reverse stress testing exercises.

This thesis is among the first studies exclusively dedicated to financial stress tests. It has not the vocation of being neither complete nor exhaustive. Instead, several fields are still open on this topic. First, more flexible risk models are still needed to perform richer and more realistic scenarios, especially for systemic stress tests. However, this requires more detailed (private) balance sheet data, and longer samples with homogeneous accounting standards. Second, a further effort has also to be made to share estimation and calibration results among researchers and supervisors. Third, techniques such as Bayesian methods should be capitalized on to carry out more objective stress scenarios. The last advancements in numerical and software programs allow reducing the implementation costs of these exercises and easing their repetition in a more regular basis. Finally, with respect to the propositions made through this thesis, we hope this study would help risk managers to rely more frequently on systematic stress testing exercises. Given the stochastic framework of this method, we also encourage more complementarity between stress testing and other conventional risk measures.

Stress Testing and Financial Risks

Abstract: This thesis has set a comprehensive framework to assess the relevance of financial stress tests, identifying their main drawbacks. Three robust and flexible model frameworks have been proposed to improve current practices in each of the tests' stages. This is achieved through: (i) a semi-parametric EVT–Pair-copulas model for financial risk factors, with a specific focus on extreme values, (ii) a valuation model to assess the impact of risk factors on a financial system, through direct and indirect effects, contagion channels, and considering private and public response functions, and (iii) a Bayesian-based approach to run a systematic selection of stress scenarios for nonlinear portfolios. The presented risk model has proven to outperform commonly used specifications, hence increasing the test's credibility. Estimated for the French banking system, the valuation model revealed the related risk profile and the main vulnerabilities. Public responses turned to be of vital interest. Finally, the Bayesian approach allows replacing the traditional subjective scenarios and including the tests' results in quantitative risk management alongside with other conventional tools.

Key words: Stress testing, scenario, EVT, pair-copulas, systemic risk, risk management, Bayesian inference

JEL classification: C5, C32, D85, G11, G17, G21, G28, G32, G33

Stress Tests et Risques Financiers

Résumé : Cette thèse a établi un cadre d'évaluation des stress tests financiers, en identifiant leurs principales limites. Trois approches ont été proposées pour améliorer les pratiques actuelles à chaque étape du processus. Elles incluent : (i) un modèle semi-paramétrique TVE–copules-paires pour les facteurs de risque financiers, avec un accent particulier sur les valeurs extrêmes, (ii) un modèle d'évaluation pour estimer l'impact de ces facteurs sur un système financier, via des effets directs, indirects et de contagion, en considérant les réactions endogènes publiques et privées, et (iii) une approche bayésienne pour mener une sélection systématique des scénarios de stress pour des portefeuilles non linéaires. Le modèle de risque a montré de meilleures performances par rapport à la plupart des spécifications courantes ; ce qui augmente la crédibilité du test. Le modèle d'évaluation est estimé pour le système bancaire français, révélant ses principales sources de vulnérabilité et le rôle clé des réactions publiques. Enfin, l'approche bayésienne a permis de remplacer les scénarios subjectifs traditionnels et d'inclure les résultats de stress tests dans la gestion quantitative des risques aux côtés des autres outils conventionnels.

Mots clés : Stress tests, scénario, TVE, copules-paires, risque systémique, gestion des risques, inférence bayésienne

Codes JEL : C5, C32, D85, G11, G17, G21, G28, G32, G33

Vu
Le président
M.

Vu
Les suffragants

Vu et permis d'imprimer par :

Le Vice-Président du Conseil Scientifique Chargé de la Recherche de l'Université Paris-Dauphine

Stress Testing and Financial Risks

Abstract: This thesis has set a comprehensive framework to assess the relevance of financial stress tests, identifying their main drawbacks. Three robust and flexible model frameworks have been proposed to improve current practices in each of the tests' stages. This is achieved through: (i) a semi-parametric EVT–Pair-copulas model for financial risk factors, with a specific focus on extreme values, (ii) a valuation model to assess the impact of risk factors on a financial system, through direct and indirect effects, contagion channels, and considering private and public response functions, and (iii) a Bayesian-based approach to run a systematic selection of stress scenarios for nonlinear portfolios. The presented risk model has proven to outperform commonly used specifications, hence increasing the test's credibility. Estimated for the French banking system, the valuation model revealed the related risk profile and the main vulnerabilities. Public responses turned to be of vital interest. Finally, the Bayesian approach allows replacing the traditional subjective scenarios and including the tests' results in quantitative risk management alongside with other conventional tools.

Key words: Stress testing, scenario, EVT, pair-copulas, systemic risk, risk management, Bayesian inference

JEL classification: C5, C32, D85, G11, G17, G21, G28, G32, G33

Stress Tests et Risques Financiers

Résumé : Cette thèse a établi un cadre d'évaluation des stress tests financiers, en identifiant leurs principales limites. Trois approches ont été proposées pour améliorer les pratiques actuelles à chaque étape du processus. Elles incluent : (i) un modèle semi-paramétrique TVE–copules-paires pour les facteurs de risque financiers, avec un accent particulier sur les valeurs extrêmes, (ii) un modèle d'évaluation pour estimer l'impact de ces facteurs sur un système financier, via des effets directs, indirects et de contagion, en considérant les réactions endogènes publiques et privées, et (iii) une approche bayésienne pour mener une sélection systématique des scénarios de stress pour des portefeuilles non linéaires. Le modèle de risque a montré de meilleures performances par rapport à la plupart des spécifications courantes ; ce qui augmente la crédibilité du test. Le modèle d'évaluation est estimé pour le système bancaire français, révélant ses principales sources de vulnérabilité et le rôle clé des réactions publiques. Enfin, l'approche bayésienne a permis de remplacer les scénarios subjectifs traditionnels et d'inclure les résultats de stress tests dans la gestion quantitative des risques aux côtés des autres outils conventionnels.

Mots clés : Stress tests, scénario, TVE, copules-paires, risque systémique, gestion des risques, inférence bayésienne

Codes JEL : C5, C32, D85, G11, G17, G21, G28, G32, G33