Three essays in financial econometrics

Three Essays
in
Financial Econometrics

ISBN 978 94 6380 612 1
Production: Datawyse | Universitaire Pers Maastricht

Three Essays
in
Financial Econometrics

DISSERTATION

in candidature for the degree of

Docteur de l'Université d'Orléans en Sciences Economiques,

and

Doctor at Maastricht University,
on the authority of the Rector Magnificus Prof. dr. Rianne Letschert
in accordance with the decision of the Board of Deans,
to be defended in public in Maastricht, the Netherlands
on Thursday 5 December 2019, at 12.45 hours

by

Jérémy Leymarie

Supervisors:

Prof. Dr. A. Hecq (Maastricht University)
Prof. Dr. C. Hurlin (Université d'Orléans)

Assessment Committee:

Prof. Dr. F. Palm (Maastricht University, chairman)
Prof. Dr. S. Laurent (Aix-Marseille Université)
Prof. Dr. J-M. Zakoïan (CREST)
Dr. N. Basturk (Maastricht University)

Dissertation Defense Committee:

Prof. Dr. F. Palm (Maastricht University)
Prof. Dr. S. Laurent (Aix-Marseille Université)
Prof. Dr. J-M. Zakoïan (CREST)
Dr. N. Basturk (Maastricht University)
Prof. Dr. G. Colletaz (Université d'Orléans)

This research was partially supported by the ANR Grant MultiRisk (ANR-16-CE26-0015-01) ("Econometric Methods for the Modelling of Multiple Risks").

Je dédie cette thèse à ma grand-mère,

Acknowledgements

It is a pleasant task to express my thanks to all those who contributed in many ways to the successful completion of this thesis. First of all, I would like to express my deep and sincere gratitude to my supervising guides Prof. Christophe Hurlin and Prof. Alain Hecq. Thank you for your continuous support, your patience, motivation, and immense knowledge. Your guidances helped me to become a more rigorous, skilled, and self-motivated young researcher. Also, I would like to express my deep appreciation to Prof. Olivier Scaillet, a valuable colleague to me during my Ph.D., who has always been there to advise me in my doctoral research, future career choices, and for being part of the people who gave me the joy in the pursuit of knowledge and to awaken my taste for academic research. I am also grateful in particular to Prof. Denisa Banulescu-Radu, Prof. Sylvain Benoit, Prof. Sébastien Laurent, Prof. Yannick Lucotte, Prof. Christophe Pérignon, and Prof. Sessi Tokpavi for their advices, discussions about research and encouragement.

Next, I take this opportunity to express gratitude to my friends and colleagues from both universities. Warm thoughts to my colleagues from Orléans: Christian, Dylan, Florian, Hajare, Ismael, Jean-Charles, Jose, Laura, Léonard, Marie-Laure, Marie-Pierre, Maxime, Moktar, Olessia, Nada, Nicolas, Réda, Yunzie, Wassim, and from Maastricht: Aditya, Benoit, Caterina, Elisa, Luca, Niloofar, Rasmus, Sean, for the great moments together and the long discussions about everyday life. A very special thank to Sullivan my research partner who had become a close friend. Thanks also to my friend Maxime despite you turned to the "Dark Side of the Force", making your Ph.D. in mathematics instead of economics. I forgive you! Finally, many heartfelt thanks to Karine for your help with my administrative tasks, and even more importantly, for the good time spent together.

I would like now to state my gratitude to Ophélie even if words are short to express my thankfulness to her. Thank you for your patience, constant support and motivation with me. These years would had not been the same without your love. I also send a special thought to Ophelie's family and more particularly to Béatrice, Lucile, and Patrick. The link we have built is much more important than you can imagine. I am very fortunate to have you in my life. My childhood friend, Christian, also deserves special mention for first being my best friend, but also for being with me always like a brother. It is an

honor for me to see you building your life, and share important moments together. Who was to know that we are on the brink of defending our Ph.D. dissertation at the same time! Also, and despite the distance that separates us, I would like to sincerely thank my family, and specifically my parents, Brigitte and Alain, and my sister Angélique. My thoughts are also with my uncle, Christian, my aunt Françoise, and my cousins Sylvain and Quentin. I naturally conclude these thanks with a person who counts a lot in my life, my grand mother Marie-Angèle. Thank you grandma for your financial help during my studies, and thank you to be proud of me at every moment. Honestly, I would not even be able to accomplish so much in my professional career without your precious support.

Jérémy Leymarie
Orléans, July 2019

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Risk management is a central area of expertise for financial institutions including banks, insurance companies, investment funds and others. One of the most important lessons we have learned from the global economic and financial crisis is that measuring risk should become even more a matter of necessity. In today's financial environment, the constant increase in the size and complexity of financial institutions and in the pace of their financial transactions has undeniably introduced a new variable into the equation. In parallel, fortunately, technological advances in communication and data gathering have lowered the cost of acquiring, managing and analyzing data to monitor rapidly changing risk exposures and better reflect the more complex and fast-paced business environment. This context has triggered the development of sophisticated financial instruments and new risk management techniques. In particular, the academic research in financial econometrics has provided impulse and direction to (*i*) the development of new financial risk measures, (*ii*) the introduction of appropriate estimation and inference methods, and (*iii*) the implementation of validation techniques that should be devoted to those indicators.

1.1 Financial Risk Measures

Techniques for the measurement of financial risk are key components in the process of managing risk. Measuring financial risks can be performed in terms of probability distributions. However, it is not always clear how to extract the relevant information through the whole probability distribution, and it may be preferable to summarize it into a single figure that says something on a particular dimension of risk (asymmetry, expectile, central tendency, tail distribution, variance, etc.). Risk measures are typically dedicated to that purpose by quantifying financial risks with one number representing the future losses that could be potentially experienced on a risky position. According to Artzner et al. (1999), when positive, this number discloses the minimum extra cash that the agent needs to add to the risky position to fit the set of acceptable risks as

decided by a supervisor. For that reason, this number is generally interpreted as a required capital amount for that risky position to regulate the risk assumed by market participants, traders, or insurance underwriters. A wide range of risk measures, designed to cover various types of financial risks and reference instruments, has been proposed in the academic literature. In what follows, we give a description of the main financial risk measures according to the type of financial risks they aim to quantify.

Credit Risk Measures. Credit risk measures are necessary tools to model the so-called expected losses and unexpected losses issued from credit portfolios. The former can be described as the "usual" or average losses that an institution incurs in its natural course of business, while the latter correspond to large potential losses that are experienced in adverse conditions and may threaten financial stability (see Gouriéroux and Tiomo, 2007; Roncalli, 2009; Genest and Brie, 2013, for more details). Four key risk measures are relevant in analyzing these losses, namely, the probability of default (PD), the loss given default (LGD), the exposure at default (EAD), and the maturity (M). The risk measure PD provides an estimate of the likelihood of a default over a particular time horizon and hence represents the risk that the borrower will be unable or unwilling to repay his or her debt in full or on time. The LGD is the amount of debt that is lost by the bank in case of default of the obligor. As a consequence, PD and LGD are closely related. A positive LGD is likely to occur for a borrower who has a high level of PD. Finally, the EAD can be defined as the gross exposure under a facility upon default of an obligor. The scope of lending activities also encompasses the default risk in central counterparties (CCPs) issued from derivative portfolios. Risk measures or approaches commonly used in CCPs are the standard portfolio analysis of risk (SPAN) or the value-at-risk (VaR) approach to estimate collateral requirements based on a coverage level of potential losses for an individual contract or portfolio of contracts (Chicago Mercantile Exchange, 2012). Recently, more sophisticated techniques have emerged. For instance, Cruz Lopez et al. (2017) develop the CoMargin methodology, which estimates collateral requirements considering both the tail risk of a given market participant and its interdependence with other participants.

Market Risk Measures. Market risk measures are used to quantify the risk of facing losses in positions due to fluctuations in market prices. As with other forms of risk, the potential loss amount due to market risk may be measured in a number of ways or by using different conventions. Traditionally, one convention is to use VaR, which reports the maximum potential loss from holding an asset (or a portfolio) over a given period and for a given probability level. VaR has become a building block of internal risk management systems in financial institutions, following the success of the J.P. Morgan (1996) RiskMetrics system. The convention of using VaR is therefore historically well established and accepted in risk management practices. However, VaR is known to contain a number

of theoretical deficiencies. In particular, VaR is not a coherent risk measure as it is not always subadditive (Artzner et al., 1999). It has also been criticized for lacking sensitivity to capture tail risk during periods of significant financial market stress (BCBS, 2016). Another suggestion for measuring market risk is expected shortfall (ES), also known as conditional value-at-risk (CVaR) or tail value-at-risk (TVaR). ES is the conditional expected loss given exceedance of VaR at a given probability level. In comparison to VaR, ES is a coherent risk measure (Acerbi and Tasche, 2002) and is much more tail-sensitive, capturing both the size and the likelihood of incurred loss events. Finally, it is worth noting that the variance (or standard deviation) also plays a significant role as a measure of market risk. In particular, the variance has become a leading indicator in modern portfolio theory (Markowitz, 1952) or, more generally, in any risk-based investment strategy based on the second-order moment (see Roncalli, 2014).

Systemic Risk Measures. The recent financial crisis has fostered extensive research on systemic risk. Of particular interest is the identification of financial institutions that contribute the most to overall risk in the financial system – the so-called systemically important financial institutions (SIFIs). The Financial Stability Board (FSB, 2011) defines SIFIs as "*financial institutions whose distress or disorderly failure, because of their size, complexity and systemic interconnectedness, would cause significant disruption to the wider financial system and economic activity*". As SIFIs pose a major threat to the system, regulators and policy makers from around the world have called for tighter supervision, extra capital requirements, and liquidity buffers for them. To that end, many systemic risk measures have been proposed in the academic literature in recent years (see Benoit et al., 2017), the most well known being the marginal expected shortfall (MES) and the systemic expected shortfall (SES) of Acharya et al. (2017), the systemic risk measure (SRISK) of Acharya et al. (2012) and Brownlees and Engle (2017), and the delta conditional value-at-risk (ΔCoVaR) of Adrian and Brunnermeier (2016). As for any risk measure, they summarize the systemic risk contribution of each financial institution into a single figure in order to identify SIFIs whose failure might trigger a crisis in the whole financial system.

1.2 Estimation

Financial risk cannot be directly measured even after the actual event is realized, and this is why econometricians perceive it as a latent process. For that reason, financial risk measures are unobservable, and we need to estimate them. These estimates are generally delivered through the use of a risk model. As part of the regulatory framework, banks have the possibility to develop their own risk models to estimate credit risk measures (BCBS, 2001) and market risk measures (BCBS, 2016), which are ultimately used to compute their regulatory capital charge for credit risk and market risk. Banks have

an incentive to hold the lowest possible regulatory capital level because reducing capital frees up economic resources that can be directed to profitable investments. Consequently, the vast majority of banks develop their own risk models that allow them to reduce the required level of capital, compared to if they applied the standardized approach provided by supervisors. However, determining how to estimate these measures and which models should be applied is a tricky exercise. Thus, banks have an important role to play in the calculation of prudential requirements by means of responsible management practices.

Various risk models are available for modeling risk measures. The nature of the response variable provides important guidance on which kind of model is appropriate. For instance, because PD modeling can be viewed as a binary classification problem, the typical modeling techniques are based on classifiers. The convention in credit scoring practices is to use binary choice models such as logistic and probit regression models or to apply statistical approaches such as discriminant analysis. Commonly used approaches for LGD modeling include, among many others, simple look-up (contingency) tables and regression models (linear regression, survival analysis, fractional response regression, inflated beta regression, or Tobit models, for instance). Over the last decades, machine learning techniques have also grown in popularity for modeling credit risk measures due to their ability to achieve significant improvement in model performance. These techniques include regression and classification trees, support vector machine and support vector regression, random forest, gradient boosting, and artificial neural network methods, among others (see Baesens et al., 2003; Lessmann et al., 2015 for a benchmarking study of PD models, and Loterman et al., 2012 for the case of LGD models).

Due to their time-varying dynamics, market risk measures are typically expressed conditionally on an information set, and the forecasts are generally issued from a dynamic parametric or semiparametric model. For instance, univariate and multivariate GARCH models can be used to produce conditional VaR or ES forecasts (see Palm, 1996 for a survey of the univariate GARCH models, and Bauwens et al., 2006 for the case of multivariate GARCH models), or a dynamic conditional correlation (DCC) model can be used to estimate a dynamic conditional beta (Engle, 2002, 2016). Modeling approaches also include univariate and multivariate quantile regression models (Engle and Manganelli, 2004; White et al., 2008, 2015, etc.), realized volatility models (Andersen et al., 2003; Corsi, 2009; Cubadda et al., 2017, etc.), and nonparametric estimation approaches (e.g., Scaillet, 2004; Cai and Wang, 2008, etc.). Several advances in modeling approaches have been provided in recent years. For instance, Darolles et al. (2018) propose a new model with time-varying slope coefficients based on the Cholesky decomposition of the conditional variance matrix. This model outperforms a model with constant betas and the DCC model. Taylor (2019) introduces a method for predicting ES corresponding to VaR forecasts produced by quantile regression models. Patton et al. (2019) use recent results

from statistical decision theory by jointly modeling ES and VaR, and propose new dynamic models for these risk measures. These types of models are also of great interest for systemic risk measurement. They are typically used to forecast the systemic risk measures computed with financial market data. For instance, Adrian and Brunnermeier (2016) consider a simple quantile regression model to forecast ΔCoVaR, while Girardi and Ergün (2013) estimate it from a multivariate GARCH model. Brownlees and Engle (2017) implement a DCC model to estimate the long-run MES and SRISK. These similarities of the modeling approaches between market risk and systemic risk come from the fact that they both rely on financial market data and that the underlying risk models must be able to capture the same type of stylized facts on financial series, such as nonstationarity in levels, fat tails, and volatility clustering.

A common and important shortcoming shared by financial risk measures is the fact that their estimates are affected by two types of measurement error. The first source of measurement error, called the estimation risk, implies that replacing the parameters of the model by their estimates has an impact on the estimation accuracy of the risk measure itself. The need for inference hence becomes a crucial issue to quantify estimation errors and correct them. Various contributions have been provided to address this concern. Hurlin et al. (2017) propose a bootstrap procedure that accounts for estimation uncertainty to test for equality of conditional risk measures for different assets, portfolios or firms at a single point in time. Francq and Zakoïan (2015) explicitly take into account the effect of estimation risk under the class of GARCH-type model and non-Gaussian quasi-maximum likelihood and provide an asymptotic distribution and confidence intervals for the VaR. Gouriéroux and Zakoïan (2013) show that VaR is affected by an asymptotic bias in the coverage probabilities induced by estimation, and they derive a correction. The second source of measurement error, called the model risk, corresponds to the risk that the forecasting model may be incorrectly specified and may lead to inconsistent outcomes of risk forecasts. Boucher et al. (2014) propose a methodology to compute risk measures intended to be robust to model risk. They show that model bias is large in general and that it strongly depends upon the probability confidence level. Daníelsson et al. (2016) propose a general framework for quantifying model risk and show that the degree of model risk of such models is quite high. Their results indicate that risk measures are subject to significant model risk during periods of financial distress, which are, unfortunately, when they are most needed. This brief overview of the literature shows that financial risk measures are often affected by these two types of measurement errors, which may skew the evaluation of risk and regulatory capital levels held by banks.

1.3 Validation

The need for sound risk management and for effectiveness in assessing risk measures have never been more essential than in today's financial environment. Of great importance is the fact that the risk measure may be conceptually valid (see Artzner et al. 1999 and Chen et al. 2013 for desirable properties of market and systemic risk measures) but not conveniently estimated. Of crucial importance for regulators and risk managers is the ability to identify misspecified risk models as they lead to misrepresentation of actual risk exposures. This assessment is generally known as backtesting. Jorion (2007) defines backtesting as a formal statistical framework that consists in verifying if actual losses are in line with projected losses. This approach involves a comparison of the historical model-generated risk measure forecasts with actual losses.

The objective of PD models is to predict the default rate. To deliver valid predictions, the PD model should correctly classify credit applicants into "good" and "bad" risk classes (Hand and Henley, 1997). For this reason, backtesting techniques for PD generally rely on discriminatory power analysis (Gouriéroux, 1992) or simple binomial statistical tests (Brown et al., 2001). For LGD models, even if there are no particular guidelines to assess their estimates, Loterman et al. (2014) propose a backtesting framework for LGD models using statistical hypothesis tests. Kalotay and Altman (2017) show that variation in the composition of the defaulted debt pool at the time of default generates time variation in the LGD distribution. They quantify the importance of accounting for such time variation in out-of-sample comparisons of alternative LGD models. Finally, for standard credits and loans, the EAD is observed, and it is unnecessary to estimate this risk parameter. However, for off-balance-sheet EAD, this quantity becomes unknown, and the bank has to estimate a credit conversion factor (CCF). Gürtler et al. (2018) develop a theoretical and empirical assessment for EAD modeling. More generally, several approaches allow the evaluation of the credit risk model as a whole. For instance, Lopez and Saidenberg (2000) provide an approach to evaluate credit risk models based on simulated credit portfolios. Medema et al. (2009) implement a simple validation methodology that can be used by banks to validate their credit risk modeling exercise.

Financial market regulation applies strict control over the internal risk models used by banks to compute their market risk capital requirements. A key responsibility of banks is to backtest certain risk models. Over the past two decades, a number of contributions have been provided to gauge the ability of predictive models to provide valid risk forecasts. These techniques are typically based on tests over violations. VaR forecasts are valid when the violation process satisfies the unconditional coverage (UC) hypothesis. Another important hypothesis for VaR forecasts is the independence hypothesis (IND), which assumes that VaR violations observed at two different dates for the same coverage rate must be distributed independently. When the UC and IND hypotheses are simul-

taneously valid, VaR forecasts are said to have a correct conditional coverage (CC), and the VaR violation process is a martingale difference sequence (see Christoffersen, 1998, for a more detailed description of these hypotheses). Engle and Manganelli (2004) develop the so-called dynamic quantile approach, focusing directly on the correlation of the violations with the actual return series. Dumitrescu et al. (2012) propose a refinement of this approach substituting the linear regression model by a nonlinear dynamic binary regression model. Several backtesting tests have also been proposed to assess the validity of VaR at various coverage rates. Colletaz et al. (2013) develop a backtest of the UC hypothesis at two coverage rates to distinguish between a situation in which losses are below but close to the VaR and a situation in which losses are considerably below the VaR. Hurlin and Tokpavi (2006) use a multivariate portmanteau statistic to test for the IND hypothesis at several probability levels. Finally, it is important to mention the class of duration-based tests that take into account the time interval between two violations (Berkowitz et al., 2011; Candelon et al., 2011; Christoffersen and Pelletier, 2004; Pelletier and Wei, 2016).

Until recently, the Basel Committee has advocated the use of ES as the new regulatory risk measure, complementing, and in parts substituting, the more familiar VaR. This regulatory shift has encouraged the development of validation procedures dedicated to the ES predictive models. McNeil and Frey (2000) develop a nonparametric backtesting framework for ES based on exceedance residuals. Acerbi and Szekely (2014) develop three new ES backtests relying on Monte Carlo simulations. Nolde and Ziegel (2017) devise the so-called conditional calibration tests for assessing ES. More recently, Bayer and Dimitriadis (2018) provide a regression-based backtest exploiting the joint elicitability of the pair VaR-ES. Kratz et al. (2018) propose to generalize the popular binomial backtest of VaR exceptions at a single coverage level to a multinomial backtest of VaR exceptions at several coverage levels. Exploiting the relationship between VaR and ES, Kerkhof and Melenberg (2004) provide a backtesting framework based on PIT exceedances that encompasses VaR and ES as special cases. Costanzino and Curran (2015) derive a coverage backtest for spectral risk measures in the spirit of the traditional VaR coverage backtests, which nests ES as a spectral risk measure. Du and Escanciano (2017) develop a cumulative violation process for ES, generalizing the initial violation process for VaR. Costanzino and Curran (2018) provide a traffic light backtest for ES that extends the so-called traffic light backtest for VaR.

Existing procedures for evaluating the correctness of systemic risk measures are surprisingly underdeveloped. There are no formal statistical procedures to assess this class of risk measures. However, even if no formal techniques have been proposed, some attempts have been made to empirically assess the predictive contents in the systemic risk measures. Idier et al. (2014) study the firms with high systemic risk scores and their like-

lihood of suffering the highest financial losses in a financial crisis. Wu and Zhao (2018) investigate whether these firms are more likely to become insolvent. Brownlees and Engle (2017) show that banks with higher SRISK before the financial crisis were more likely to be bailed out by the government and to receive capital injections from the Federal Reserve. Engle et al. (2015) compare the ranking of European financial institutions obtained with the SRISK to the list of SIFIs produced by the FSB. Recently, Brownlees et al. (2018) propose a historical assessment of the SRISK and ΔCoVaR based on two dimensions. The first dimension, called the SIFI ranking challenge, consists in investigating whether ranking financial institutions by SRISK and ΔCoVaR allows the identification of institutions with notable deposit declines around panic events. The second dimension, called the financial crisis prediction challenge, investigates whether these systemic risk measures are significant predictors of system-wide deposit declines during panic events.

The systemic indicators listed above share the same feature of relying on publicly available data, e.g., stock, asset returns, option prices, and CDS spreads. Obviously, one can also question the validity of methods for systemic risk measurement that rely on proprietary data, e.g., balance sheet, cross-positions, size, leverage, liquidity, and interconnectedness. These proprietary-based methods have been embedded in the toolbox of banking supervisors because they include a more theoretical foundation than those based on financial market data. Even if access to the data is rendered more difficult, if not impossible, several attempts have been made to assess the quality of the systemic risk methodology issued from private data. Philippon et al. (2017) provide a first attempt to empirically assess the quality of the European Banking Authority (EBA) 2014 stress tests. They find that stress tests are informative and provide reliable information for regulators on the resilience of banks. Benoit et al. (2019) identify several shortcomings in the systemic-risk scoring methodology currently used to identify and regulate the SIFIs. They propose a new methodology that addresses these shortcomings and improve regulatory capital allocation among banks.

1.4 Contribution

In this renewed context, our research focuses on financial risk measures and the validation techniques dedicated to their predictive models. The broad goal of this dissertation is to provide advanced tools for the evaluation of risk measure estimates. Our methodological developments for the assessment of the risk measures cover three prominent classes of financial risks, namely, (*i*) credit risk, (*ii*) market risk, and (*iii*) systemic risk. For each of these categories, our overall objective remains the same: improving the soundness of the banking system by means of the development of formal validation methods for risk estimates. Regarding credit risk, our dissertation contributes to enhancing the reliability of estimates of future losses issued from loan portfolios and making regulatory capital al-

location more efficient. Regarding market risk, our work aims to make asset management practices sounder in order to properly cover investment companies for adverse shocks and potential losses. Finally, this thesis also contributes to the reinforcement of financial stability as a whole by improving bank's monitoring via a precise identification of the SIFIs through the systemic risk measures. This work has been concretized in three chapters (articles) that can be studied independently one from another.

The first chapter addresses issues related to credit risk assessment. We focus on LGD and propose an original comparison methodology that selects the LGD predictive model inducing the lowest estimation errors on regulatory capital and, as a result, improving banks' solvability. Chapters 2 and 3 investigate the validity of the market-based risk measures. Chapter 2 meets the requirements of regulators to provide more efficient validation tools for the risk measure ES. We develop a new approach to assess the validity of the ES predictive models based on multivariate quantile regressions. Because our underlying validation methodology makes use of the relationship between VaR and ES, this new class of statistical tests is consistent with the current Basel regulatory guidelines that recommend backtesting ES by verifying the validity of several VaRs in the tail of the portfolio loss distribution. Exploiting our evaluation procedure, we provide an original technique to adjust the imperfect ES forecasts, which are cleansed of estimation risk and model risk. In Chapter 3, we investigate market-based systemic risk measures and the quality of their predictions. We rely on standard VaR backtesting procedures and develop a first backtest for the UC hypothesis and a second for the IND hypothesis. To the best of our knowledge, ours is the first formal statistical procedure to backtest systemic risk measures. We then exploit our methodology to provide an early warning system that demonstrates remarkable ability to detect early signs of crisis. In the following, we synthesize the contents of each chapter.

Chapter 2: Loss functions for Loss Given Default model comparison

Chapter 2, "Loss functions for Loss Given Default model comparison", proposes an alternative comparison methodology for the LGD models that is based on expected loss functions expressed in terms of regulatory capital charge.[1] Within the credit risk regulatory framework, the level of regulatory capital is determined to cover bank's unexpected credit loss (BCBS, 2005). To derive this unexpected loss, the Basel Committee provides a theoretical framework based on the asymptotic single risk factor (ASRF) model, inspired by the seminal Merton-Vasicek "model of the firm" (Merton, 1974; Vasicek, 2002). It is then possible to compute the capital charge for credit risk based on the ASRF model and

[1]This chapter is based on Hurlin, Leymarie, and Patin (2018) published in the *European Journal of Operational Research*.

some external estimated risk parameters. LGD is one of the most important parameters involved in this formula. The LGD can be broadly defined as the ratio of losses (expressed as percentage of the EAD) that will never be recovered by the lender or equivalently as one minus the recovery rate. Because the LGD enters the regulatory capital formula in a linear way, any underestimation of this risk parameter may induce an underestimation of the regulatory capital and a lowest bank's solvency.

According to the advanced internal rating-based (AIRB) approach adopted by most major international banks, the LGD forecasts are issued from internal risk models. No particular guidelines have been proposed concerning how LGD models should be selected, compared, and evaluated. As a consequence, the model benchmarking method simply consists in evaluating LGD forecasts with standard statistical criteria such as the mean square error, or the mean absolute error, computed between the observed LGD and its forecast, as for any continuous variable. Therefore, the current LGD model comparison is done regardless of the other Basel risk parameters (exposure at default, probability of default, maturity) and by neglecting the impact of LGD forecast errors on regulatory capital. This approach may lead to the selection of a LGD model that has the smallest mean square error among all the competing models but that induces small errors on small exposures and large errors on large exposures.

This chapter aims to address these weaknesses by developing an alternative comparison methodology that improves the banks' solvability. Contrary to the existing approach, our model's comparison method more heavily penalizes LGD forecast errors made on credits associated with high exposure and long maturity and selects the model associated with the lowest estimation errors on regulatory capital. This is not the case with the comparison method currently used by banks and academics, which selects the model that minimizes the estimation errors on the LGD itself. We show theoretically that our approach ranks models differently compared to the traditional approach.

Using a sample of credit and leasing contracts provided by an international bank, we illustrate the interest of our method by comparing the rankings of six competing LGD models. Our empirical findings clearly show that model rankings based on capital charge losses differ substantially from those based on the LGD loss functions currently used by regulators, banks, and academics. The proposed method allows the identification of the best LGD models associated with the lowest estimation errors on regulatory capital. Beyond these traditional statistical criteria, we also introduce asymmetric criteria especially designed to improve financial stability. These loss functions only penalize LGD forecast errors that lead to underestimating regulatory capital. We find that the rankings based on symmetric criteria are drastically different from the model rankings obtained with asymmetric criteria, which highlights the usefulness of asymmetric functions to enhance the soundness and stability of the banking system.

Chapter 3: Backtesting Expected Shortfall via Multi-Quantile Regression

Chapter 3, "Backtesting Expected Shortfall via Multi-Quantile Regression", proposes an easy-to-use regression-based approach to backtesting ES based on quantile models.[2] Within the context of financial market regulation and banking supervision, the Third Basel Accord has drawn new attention on ES for the computation of market risk capital requirements, complementing, and in parts substituting the more familiar VaR measure (BCBS, 2010). As an alternative tail risk measure, ES offers a number of appealing properties that overcome the theoretical deficiencies of VaR. In particular, ES is coherent, meaning that this risk measure satisfies the properties of monotonicity, subadditivity, homogeneity, and translational invariance (see Artzner et al., 1999; Acerbi and Tasche, 2002). In its revised standards for market risk, the BCBS emphasizes the important role of ES in place of VaR "*to ensure a more prudent capture of 'tail risk' and capital adequacy during periods of significant financial market stress*" (BCBS, 2016, page 1).

As a result, the key challenge for risk managers and policymakers is to develop appropriate modeling methods for ES (see the recent work of Taylor, 2019; Patton et al., 2019, for instance) and to elaborate advanced tools to assess ES forecasts on a real-time basis. This chapter precisely focuses on the second point. Indeed, validation and backtesting procedures are key requirements for any financial risk measure to become an industry standard. Even more important, the validity of ES estimates is crucial given that the ES parameter is a key constituent of market risk regulatory capital. As a consequence, any underestimation of ES that has not been identified in time may threaten the bank's solvability.

In this chapter, we suggest a natural extension to standard VaR backtesting procedures that allow us to test VaR estimates at several probability levels jointly. Because ES can be broadly defined as a function of VaR at different probability levels along the tail distribution of the portfolio loss, our approach can be viewed as providing an implicit backtest for ES. Our testing strategy follows the general recommendation of financial supervisors. According to the BCBS guidelines on ES assessment "*Backtesting requirements are based on comparing each desk's 1-day static value-at-risk measure [...] at both the 97.5th percentile and the 99th percentile*" (BCBS, 2016, page 57). To implement our testing strategy, we develop a validation framework based on multivariate quantile regression. The procedure extends the test of Gaglianone et al. (2011) that allows the backtesting of VaR at a single probability level.

This approach has many advantages. First, our procedure is flexible since the user may choose the number and values of quantiles for the assessment and can easily focus

[2]This chapter is based on Couperier and Leymarie (2019) currently R&R in the *Journal of Business and Economic Statistics*.

on various aspects of the tail distribution of the forecasting model. Second, our methodology has the advantage of being consistent with the regulatory guidance to verify if the underlying ES model delivers correct quantiles at levels 0.975 and 0.990. Finally, the procedure is easy to implement for risk managers and regulatory supervisors as it is based on the well-established VaR. This validation tool is therefore more likely to be embraced by financial institutions as a new standard for financial risk management.

To illustrate the benefits of our method, we assess the ES estimates issued from the popular AR(1)-GARCH(1,1) model assuming that the portfolio returns of the investor are given by the S&P500 index over the period 2007-2012. In this period of financial turbulence, the procedure concludes that the forecasts of ES are misleading. Our results also suggest that one should be very cautious in using a single VaR to assess the tail distribution of the portfolio loss distribution. Furthermore, the use of two VaRs, as recommended by financial supervisors, is not always enough to identify improper risk forecasts and thus can lead to inaccurate levels of market risk capital requirements, threatening financial stability. As a general result, we show that four to six VaRs for the tail distribution assessment provide a better ability to identify an improper ES predictive model, which should be accordingly taken into account by financial regulators.

Chapter 4: Backtesting Marginal Expected Shortfall and Related Systemic Risk Measures

Chapter 4, "Backtesting Marginal Expected Shortfall and Related Systemic Risk Measures", develops the first statistical procedure for the assessment of market-based systemic risk measures.[3] The ultimate goal of a systemic risk measure is to better identify the vulnerabilities of the financial system, and in practice, there are three ways of measuring systemic risk (see Benoit et al., 2017, for a survey). A first approach, called the supervisory approach, relies on firm-specific proprietary data on size, leverage, liquidity, interconnectedness, complexity, and substitutability and a scoring methodology (Benoit et al., 2019). A second approach relies on structural models that identify specific sources of systemic risk, such as contagion, bank runs, or liquidity crises. A third approach aims to derive global measures of systemic risk based on publicly available market data, such as stock returns, option prices or CDS spreads, encompassing the so-called MES, SES, SRISK, and ΔCoVaR market-based systemic risk measures.

However, these methodological approaches may lead to opposite conclusions, which has caused a fierce debate in the academic and regulatory spheres in the last few years. It was typically the case in October 2014, when the EBA disclosed the effect of losses in a stress test on bank capital ratios and concluded that French banks were among

[3]This chapter is based on Banulescu-Radu, Hurlin, Leymarie and Scaillet (2018) and has been awarded a research grant sponsored by the Fondation Banque de France.

the safest banks in the Eurozone. These conclusions were immediately cast in doubt by Viral Acharya (Financial Times, October 27, 2014), who argued that French banks were the riskiest in Europe according to the results of the SRISK displayed on the Volatility Lab (NY University). These controversies raise the issue of validation of stress tests (Philippon et al., 2017) but also of systemic risk measures. However, to the best of our knowledge, no backtesting procedure based on sound econometrics has been proposed yet for the systemic risk measures.

In this context, this chapter proposes the first general framework for backtesting systemic risk measures. Our testing strategy follows the baselines of the standard backtests used for VaR that exploit the martingale difference sequence (mds) property of a violation process (Kupiec, 1995; Christoffersen, 1998, 2010; Berkowitz et al., 2011, among others). The main technical novelty of our approach is to exploit the mds property for what we call the cumulative joint violation process. This new class of violation is specifically devised for assessing the validity of the systemic risk measures and generalizes to the bivariate case the cumulative violation process introduced by Du and Escanciano (2017) for backtesting ES. In a first step, we focus our research on the MES, and then, we extend our backtesting procedure to the other systemic risk measures (SRISK, SES, ΔCoVaR). Exploiting the mds property of the cumulative joint violation process enables the implementation of various types of backtests for systemic risk indicators. Here, we propose two tests based on the so-called UC hypothesis and IND hypothesis (Christoffersen, 1998). These tests are designed to verify if the systemic risk measure forecasts are in line with the ex-post losses by controlling for the number and correlation of the cumulative joint violations. These statistical backtests are easy to implement and similar to those currently used by risk managers for assessing the market risk measures. They may thus be adopted in financial regulation without too much difficulty.

Our empirical results based on U.S. banks reveal that the one-day-ahead systemic risk forecasts are misleading for a large subset of banks before the systemic risk crisis occurs. However, when considering a longer forecasting horizon (one month), our tests conclude that the systemic risk predictions are valid and suggest that these indicators are more suited to capturing the long-run dynamics of systemic risk. Finally, we show that our procedure can also be used as an early warning system (EWS) for systemic risk crisis. To this end, we introduce an EWS indicator defined as the difference between the systemic risk forecast issued from a potentially misspecified risk model and its adjusted counterpart. The latter is derived from our backtesting procedure and represents the systemic risk prediction for which we do not reject the null hypothesis of UC. This adjustment is obtained by tailoring the probability level of the crisis event considered for the systemic risk measure. This technique has already been considered for market risk measures such as ES and VaR (see Boucher et al., 2014; Lazar and Zhang, 2019,

for instance) but to date, it was not still available for systemic risk measures because of the lack of theoretical background on backtesting this class of risk measure. Our EWS indicator reports a sharp increase before the early signs of the crisis and reaches its highest value during the historic collapse of Lehman Brothers. Therefore, it may give useful insights for monitoring the financial system on a real-time basis, complete the toolbox used by academics and regulators to capture the build-up of systemic risk in tranquil times, and improve allocation efficiency of regulatory capital among banks.

Finally, Chapter 5 summarizes the main findings of this thesis and puts forward several objectives for future research.

Chapter 2

Loss functions for Loss Given Default model comparison[1]

This chapter proposes a new approach for comparing Loss Given Default (LGD) models which is based on loss functions defined in terms of regulatory capital charge. Our comparison method improves the banks' ability to absorb their unexpected credit losses, by penalizing more heavily LGD forecast errors made on credits associated with high exposure and long maturity. We also introduce asymmetric loss functions that only penalize the LGD forecast errors that lead to underestimate the regulatory capital. We show theoretically that our approach ranks models differently compared to the traditional approach which only focuses on LGD forecast errors. We apply our methodology to six competing LGD models using a sample of almost 10,000 defaulted credit and leasing contracts provided by an international bank. Our empirical findings clearly show that models' rankings based on capital charge losses differ from those based on the LGD loss functions currently used by regulators, banks, and academics.

2.1 Introduction

Since the Basel II agreements, banks have the possibility to develop internal rating models to compute their regulatory capital charge for credit risk, through the internal rating-based approach (IRB). The IRB approach can be viewed as an external risk model based on the asymptotic single risk factor (ASRF) model. This risk model relies on four key risk parameters: the exposure at default (EAD), the probability of default (PD), the loss given default (LGD), and the effective maturity (M). The Basel Committee on Banking Supervision (BCBS) allows financial institutions to use one of the following two methods: (1) the Foundation IRB (FIRB), in which banks only estimate the PD, the

[1]This chapter is based on Hurlin, Leymarie, and Patin (2018) published in *European Journal of Operational Research*.

other parameters being arbitrarily set; (2) the Advanced IRB (AIRB), in which banks estimate both the PD and the LGD using their own internal risk models.[2]

In this chapter, we propose a new approach for comparing LGD models which is based on loss functions defined in terms of regulatory capital charge. Given the importance of the LGD parameter in the Basel risk weight function and the regulatory capital for credit risk, the LGD model comparison is a crucial problem for banks and regulators. Unlike PD, the LGD estimates enter the capital requirement formula in a linear way and, as a consequence, the estimation errors have a strong impact on required capital. Furthermore, there is no benchmark model emerging from the "zoo" of LGD models currently used by regulators, banks, and academics.[3] Indeed, the academic literature on LGD definition, measurement, and modelling is surprisingly underdeveloped and is particularly dwarfed by the one on PD models. The LGD can be broadly defined as the ratio (expressed as percentage of the EAD) of the loss that will never be recovered by the bank in case of default, or equivalently by one minus the recovery rate. While this definition is clear, the measurement and the modelling of LGD raise numerous issues in practice. Regarding the measurement, both the BCBS and the European Banking Authority (see for instance EBA, 2016) made tremendous efforts to clarify the notion of default and the scope of losses that should be considered by the banks to measure the *workout* LGD. On the contrary, no particular guidelines have been provided for the LGD models. This may explain why there is such a large heterogeneity in the modelling approaches used by AIRB banks and academics (see Section 2.2.3, for a survey). Commonly used approaches include among many others, simple look-up (contingency) tables, parametric regression models (linear regression, survival analysis, fractional response regression, inflated beta regression, or Tobit models, for instance), and non-parametric techniques (regression tree, random forest, gradient boosting, artificial neural network, support vector regression, etc.). Within an extensive benchmarking study based on six real-life datasets provided by major international banks, Loterman et al. (2012) evaluate 24 regression techniques. They found that the average prediction performance of the models in terms of R-square ranges from 4% to 43%. Similarly, Qi and Zhao (2011) compare six models that provide very different results.

How should LGD models be compared? The benchmarking method currently adopted by banks and academics simply consists in (1) considering a sample of defaulted credits split in a training set and a test set, (2) estimating the competing models on the training set and then, (3) evaluating the LGD forecasts on the test set with standard statistical

[2]In the FIRB approach, the LGD is fixed at 45% for senior claims on corporate, sovereigns, and banks not secured by recognized collateral, 75% for all subordinated claims on corporate, sovereigns, and banks. The effective maturity M is fixed at 2.5 years for corporate exposures except for repo-style transactions where the maturity is fixed at 6 months (Roncalli, 2009).

[3]By analogy with the "factor zoo" evoked by Cochrane (2011).

criteria such as the mean square error (MSE) or the mean absolute error (MAE). Thus, LGD model comparison is made independently from the other Basel risk parameters (EAD, PD, M). The first shortcoming of this approach lies with the lack of economic interpretability of the loss function applied to the LGD estimates. What do a MSE of 10% or a MAE of 27% exactly imply in terms of financial stability? These figures give no information whatsoever about the estimation error made on capital charge and bank's ability to face an unexpected credit loss. The second shortcoming is related to the two-step structure of the AIRB approach. The LGD forecasts produced by the bank's internal models are, in a second step, introduced in the regulatory formula to compute the capital charge. If LGD models are compared independently from this second step, the same weight is given to a LGD estimation error of 10% made on two contracts with an EAD of 1,000€ and 1,000,000€, respectively. Similarly, it gives the same weight to a LGD estimation error of 10% made on two contracts, one with a PD of 5% and another with a PD of 15%.

On the contrary, within our approach the LGD forecast errors are assessed in terms of regulatory capital and ultimately, in terms of bank's capacity to face unexpected losses on its credit portfolio. To do so, we define a set of expected loss functions for the LGD forecasts, which are expressed in terms of *regulatory capital charge* induced by these forecasts. Hence, these loss functions take into account the EAD, PD, and maturity of the loans. For instance, they penalize more heavily the LGD forecast errors made on credits associated to high exposure and long maturity. Furthermore, we propose asymmetric loss functions that only penalize the LGD forecast errors that lead to underestimating the regulatory capital. Such asymmetric functions may be preferred by the banking regulators in order to neutralize the impact of the LGD forecast errors on the required capital and ultimately, to enhance the soundness and stability of the banking system. We show theoretically that the models ranking determined by a LGD-based loss function (MSE, MAE, etc.) may differ from the ranking based on the corresponding capital charge loss function. In particular, we demonstrate the conditions under which both rankings are consistent and show that these conditions are likely to be violated in practice. This theoretical analysis confirms the relevance of our comparison framework for the LGD models and the usefulness of the regulatory capital estimation errors as comparison criteria.

We apply our methodology using a sample of almost 10,000 defaulted credit and leasing contracts provided by the bank of a worldwide leader automotive company. The originality of our dataset lies in the fact that the LGD observations incorporate all expenses (with an appropriate discount rate) arising during the workout process, to meet the Basel II requirements. Hartmann-Wendels et al. (2014) and Miller and Töws (2018) argue that workout costs are rarely considered in empirical studies, even if they are es-

sential for LGD modelling. Indeed, Gürtler and Hibbeln (2013) show that neglecting the workout costs leads to underestimate the LGD. Given this dataset, we compare six competing LGD models which are among the most often used in the empirical literature, namely (1) the fractional response regression, (2) the regression tree, (3) the random forest, (4) the gradient boosting, (5) the artificial neural network, and (6) the least squares support vector regression. We find that the models ranking based on the LGD loss function is generally different from the models ranking obtained with the capital charge loss function. Such a difference clearly illustrates that the consistency conditions previously mentioned are not fulfilled, at least in our sample. Our findings are robust to (1) the choice of the explanatory variables considered in the LGD models, (2) the inclusion (or not) of the EAD as a covariate, and (3) the use of the Basel PDs (collected one year before the default) in the capital charge loss function. We also find that the LGD forecast errors are generally right-skewed. In this context, the use of asymmetric loss functions provides a models ranking which is very different from the ranking obtained with symmetric loss functions.

The main contribution of this chapter is to propose a comparison method for LGD models which improves the banks' solvability. Within the BCBS framework, the level of regulatory capital is determined such as to cover unexpected credit losses. This level depends on estimated risk parameters, and in particular on the LGD. As a consequence, any underestimation of these risk parameters induces an underestimation of the regulatory capital and in fine, a lowest bank's solvency. In this context, when considering a set of competing LGD models that produce different LGD forecasts, an appropriate comparison method should select the model associated with the lowest estimation errors on the regulatory capital. This is not the case with the comparison method currently used by banks and academics which is only based on the LGD estimation errors. Conversely, our approach allows us to select the LGD model which induces the lowest estimation errors on the regulatory capital. Hence, we believe that adopting this new model comparison approach should be of general interest. Furthermore, our work complements the nascent literature on the LGD model validation. Loterman et al. (2014) propose a backtesting framework for LGD models using statistical hypothesis tests. Kalotay and Altman (2017) show that variation in the composition of the defaulted debt pool at the time of default generate time variation in the LGD distribution. They quantify the importance of accounting for such time variation in out-of-sample comparisons of alternative LGD models.

The rest of this chapter is structured as follows. We discuss in Section 2.2 the main features of the AIRB approach and the regulatory capital for credit risk portfolios. The discussion continues thereafter with a brief survey of LGD models and the method currently used to compare them. In Section 2.3, we present the capital charge loss function

that is at the heart of our comparison methodology. In Section 2.4, we describe the dataset as well as the six competing LGD models. In section 2.5, we conduct our empirical analysis and display our main takeaways. In Section 2.6, we discuss various robustness checks. We summarize and conclude this chapter in Section 2.7.

2.2 Capital charge for credit risk portfolios

In this section, we propose a brief overview of the importance of the LGD within the AIRB approach. Then, we present the main issues related to LGD measurement and we summarize the existing literature on LGD models. Finally, we discuss the method which is currently used to compare LGD models.

2.2.1 Capital requirement, individual risk contributions, and LGD

Let us consider a portfolio of n credits indexed by $i = 1, ..., n$. Each credit is characterized by (1) an EAD defined as the outstanding debt at the time of default, (2) a LGD defined as the percentage of exposure at default that is lost if the debtor defaults, (3) a PD that measures the likelihood of the default risk of the debtor over a horizon of one year, and (4) an effective maturity M, expressed in years. The credit portfolio loss is then equal to

$$L = \sum_{i=1}^{n} \mathrm{EAD}_i \times \mathrm{LGD}_i \times D_i,$$

where D_i is a binary random variable that takes a value 1 if there is a default before the residual maturity M_i and 0 otherwise.

In the AIRB approach, the regulatory capital (RC) charge is designed to cover the unexpected bank's credit loss. The unexpected loss is measured as the difference between the 99.9% value-at-risk of the portfolio loss and the expected loss $\mathbb{E}(L)$. In order to derive this unexpected credit loss, the Basel Committee proposes a framework based on the ASRF model. This model is based on the seminal Merton-Vasicek "model of the firm" (Merton, 1974; Vasicek, 2002) with additional assumptions such as the infinite granularity of considered portfolios, the normal distribution of the risk factor, and a time horizon of one year (BCBS, 2005). Under these assumptions, the unexpected loss, and hence the regulatory capital, can be decomposed as a sum of independent risk contributions (RC_i) which only depend on the characteristics of the i^{th} credit (cf. Appendix 2.8.1). The regulatory capital is then equal to

$$\mathrm{RC} = \sum_{i=1}^{n} \mathrm{RC}_i.$$

The risk contribution RC_i for the i^{th} credit is given by

$$\mathrm{RC}_i \equiv \mathrm{RC}_i\left(\mathrm{EAD}_i, \mathrm{PD}_i, \mathrm{LGD}_i, \mathrm{M}_i\right) = \mathrm{EAD}_i \times \mathrm{LGD}_i \times \delta\left(\mathrm{PD}_i\right) \times \gamma\left(\mathrm{M}_i\right), \tag{2.1}$$

with

$$\delta\left(\mathrm{PD}_i\right) = \Phi\left(\frac{\Phi^{-1}\left(\mathrm{PD}_i\right) + \sqrt{\rho\left(\mathrm{PD}_i\right)}\Phi^{-1}\left(99.9\%\right)}{\sqrt{1-\rho\left(\mathrm{PD}_i\right)}}\right) - \mathrm{PD}_i, \tag{2.2}$$

where $\Phi\left(.\right)$ denotes the cdf of a standard normal distribution, $\rho\left(\mathrm{PD}\right)$ a parametric decreasing function for the default correlation, and $\gamma\left(\mathrm{M}\right)$ a parametric function for the maturity adjustment. The maturity adjustment and the correlation functions suggested by the BCBS depend on the type of exposure: corporate, sovereign or bank exposures, versus residential mortgage, revolving, or other retail exposures (see Appendix 2.8.2, for more details).

These equations highlight the key role of LGD within the Basel II framework. Since LGD enters the capital requirement formula in a linear way, LGD forecast errors have necessarily a strong impact on the regulatory capital. Consequently, the LGD measurement and the choice of an efficient forecasting model are crucial for bank's solvability.

2.2.2 LGD measurement

The LGD measurement raises numerous practical issues. Schuermann (2004) identifies three ways of measuring LGD. The market LGD is calculated as one minus the ratio of the trading price of the asset some time after default to the trading price at the time of default. The implied market LGD is derived from risky (but not defaulted) bond prices using a theoretical asset pricing model. As they are based on trading prices, the market and implied market LGDs are generally available only for bonds issued by large firms. On the contrary, the workout LGD can be measured for any type of instrument. The workout LGD is based on an economic notion of loss including all the relevant costs tied to the collection process. The Basel II Accord identifies three types of costs: (1) the direct (external) costs associated to the loss of principal and the foregone interest income, (2) the indirect (internal) costs incurred by the bank for recovery in the form of workout costs (administrative costs, legal costs, etc.), and (3) the funding costs reflected by an appropriate discount rate tied to the time span between the emergence of default and the actual recovery. So, the scope of necessary data for proper LGD measurement is very broad.[4] However, the workout approach is clearly preferred by the regulators. For instance, in its guidelines on LGD estimation, the EBA states that "*the workout LGD*

[4]This may explain why most empirical academic studies neglect workout costs because of data limitations (cf. Miller and Töws, 2018, for a discussion), even if Khieu et al. (2012) found evidence that market LGDs are biased estimates of the workout LGD.

is considered to be the main, superior methodology that should be used by institutions." (EBA, 2016, page 11).

Whatever the measure considered, the LGD distribution across defaulted bank loans or bonds generally exhibits two main stylized facts. Firstly, the LGD theoretically ranges between 0 and 100% of the EAD, meaning that the bank cannot recover more than the outstanding amount and that the lender cannot lose more than the outstanding amount. However, several studies (Schmit, 2004; Gürtler and Hibbeln, 2013; Miller and Töws, 2018) show that when workout costs are incorporated, the LGD is sometimes larger than 100%. Secondly, many empirical studies show a bimodal LGD distribution (see Miller and Töws, 2018, for instance). Most of the LGD values of defaulted contracts are either concentrated around high values (typically 70-80%) or low values (typically 20-30%).

2.2.3 LGD models

The general purpose of the LGD (internal) models consists in providing an estimate of the LGD for the credits which are currently in the bank's portfolio and for which the bank does not observe the potential losses induced by a default of the borrower. These models are generally estimated on a sample of defaulted credits for which the ex-post workout LGD is observed. By identifying the main characteristics of these contracts and the key factors of the recovery rates, it is then possible to forecast the LGD for the non-defaulted credits.

Because of the specific nature of the LGD distribution, a large variety of LGD models are currently used by academics and practitioners. Within the empirical literature, we can distinguish parametric and non-parametric approaches. The simplest parametric approach consists in using linear regression models based on debt characteristics and macroeconomic variables (Gupton and Stein, 2002; Bastos, 2010; Khieu et al., 2012). However, the linear model generally yields poor out-of-sample predictive performances.[5] Consequently, many other parametric models have been considered for LGD forecasting. Since the LGD is theoretically defined over $[0, 1]$, these models are generally based on various transformations of LGD data which are done prior to the modelling stage or within the model itself. The most often used transformations are either based on beta (Credit Portfolio View of Gupton and Stein, 2002), exponential-gamma (Gouriéroux et al., 2006), or logistic-Gaussian distributions. In a similar way, the fractional response regression or log-log models, which keep the predicted values in the unit interval, have also been used for LGD modelling by Dermine and Neto De Carvalho (2006), Bastos (2010), Qi and Zhao (2011) or Bellotti and Crook (2012). Calabrese (2014b) uses an inflated beta regression model based on a mixture of a continuous beta distribution on $[0, 1]$ and a discrete

[5]Notice that Zhang and Thomas (2012) found that linear regression models yield better performance than survival analysis models.

Bernoulli distribution, in order to model the probability mass at the boundaries 0 and 1. Similarly, Calabrese (2014a) proposes a parametric mixture distribution approach for downturn LGD. More recently, Kalotay and Altman (2017) suggest conditional mixtures of distributions allowing time variation in the LGD distribution. Using a different approach, Tanoue et al. (2017) propose a parametric multi-step approach for the LGD of bank loans in Japan.

The main advantage of parametric models is their interpretability, but they usually have weak predictive performances compared to non-parametric methods that do not assume a specific distribution for LGD. Qi and Zhao (2011) compare fractional response regression to other parametric and non-parametric methods, such as regression trees and neural networks. They conclude that non-parametric methods perform better than parametric ones when overfitting is properly controlled for. A similar result is obtained by Bastos (2010) who recommends the use of non-parametric regression trees. If the predictive performance of non-parametric techniques is largely documented, it is difficult to identify the best models given the great heterogeneity of datasets and benchmarks considered. Using data from Moody's Ultimate Recovery Database (MURD), Bastos (2014) recommends a bagging algorithm. Hartmann-Wendels et al. (2014) use three datasets from German leasing companies to compare hybrid finite mixture models, model trees and regression trees. Their conclusions depend on the sample size and differ according to out-of-sample or in-sample performance criteria. Yao et al. (2015) compare the predictions of support vector regression techniques with thirteen other algorithms using data from MURD. They conclude that all support vector regression models substantially outperform other statistical models in terms of both model fit and out-of-sample predictive accuracy. The previously mentioned benchmarking study of Loterman et al. (2012) compares 24 parametric and non-parametric techniques, including ordinary least squares regression, beta regression, robust regression, ridge regression, regression splines, neural networks, support vector regressions, and regression trees. They conclude that non-linear techniques, and in particular support vector regressions and neural networks, perform significantly better than more traditional linear techniques.[6]

In addition to single-stage models, some studies implement two-stage models to forecast LGD. These methods have the advantage to model the extreme values concentrating on the boundaries at 0 and 1. Bellotti and Crook (2012) propose a two-stage model based on a decision tree algorithm (with two logistic regression sub-models) which is applied to

[6]Other studies aim to model the LGD distribution using non-parametric estimators. Renault and Scaillet (2004) or Hagmann et al. (2005) propose different kernel estimators of the LGD density for defaulted loans. Calabrese and Zenga (2010) consider a mixture of beta kernels estimator to model the LGD density of a large dataset of defaulted Italian loans. These approaches have the common advantage to reveal a number of bumps which can be larger than those obtained with parametric distributions. We can also mention Krüger and Rösch (2017) who consider quantile regressions for modelling downturn LGD.

split the whole sample into three groups according to the values of LGD (0, 1, or between 0 and 1). Then the values in $]0, 1[$ are fitted by an OLS regression model. Yao et al. (2017) improve this two-stage approach by considering a least squares support vector classifier rather than logistic regressions. They show that this two-stage model outperforms the single-stage support vector regression model in terms of out-of-sample R-square. Considering two datasets of home equity and corporate loans, Tobback et al. (2014) also find that a two-stage model (which combines linear regression and support vector regression) outperforms the other techniques when forecasting out-of-time. But, they observe that non-parametric regression tree has better performance when forecasting out-of-sample. Miller and Töws (2018) propose an original multi-step estimation approach based on an economic separation of the LGD determined by the workout process. Nazemi et al. (2017) implement a fuzzy fusion model which uses a function to combine the results of several base models. They show that the fuzzy fusion model has higher predictive accuracy compared to support vector regression models.

This brief overview of the literature shows that there is no benchmark model emerging from the "zoo" of LGD models. Consequently, for each new real-life database, academics and practitioners have to consider several LGD models and compare them according to appropriate comparison criteria.

2.2.4 LGD models comparison

In this section, we briefly present the method currently used both by academics and banks to compare the predictive performances of LGD models. Consider a set of $\mathcal{M}$ LGD models indexed by $m = 1, ..., \mathcal{M}$ and a sample of n_d defaulted credits which is randomly split into a training set including n_t credits and a test set including n_v credits, with $n_t + n_v = n_d$. In a first step, the models are estimated (for parametric models) or calibrated on the training set.[7] In a second step, the models are used to produce pseudo out-of-sample forecasts of the LGD for the credits of the test set. The test set is then used solely to assess the prediction performances of the models. Denote by LGD_i the true LGD value observed for the i^{th} credit of the test set, for $i = 1, ..., n_v$ and by $\widehat{\text{LGD}}_{i,m}$ the corresponding forecast issued from model m.

The assessment of the prediction performances of the LGD models is generally based on an expected loss $\mathcal{L}$ defined as

$$\mathcal{L}_m \equiv \mathcal{L}\left(\text{LGD}_i, \widehat{\text{LGD}}_{i,m}\right) = \mathbb{E}\left(L\left(\text{LGD}_i, \widehat{\text{LGD}}_{i,m}\right)\right),$$

[7] For machine learning methods (regression trees, neural networks, etc.), the training set is sometimes further split into training and validation subsets. The validation set is used to select the optimal tuning parameters that provide the best in-sample predictive performance.

where $L(.,.)$ is an integrable loss function, with $L : \Omega^2 \to \mathbb{R}^+$.[8] Since the LGD is a continuous variable defined over a subspace Ω of $\mathbb{R}^+$ (typically $[0,1]$ or $[0,\delta]$ with $\delta > 1$), the loss functions generally considered in academic literature are the quadratic loss function $L(x,\widehat{x}) = (x-\widehat{x})^2$ and the absolute loss function $L(x,\widehat{x}) = |x-\widehat{x}|$. Thus, the LGD models are compared through the empirical mean of their losses computed on the test set, defined as

$$\widehat{\mathcal{L}}_m = \frac{1}{n_v}\sum_{i=1}^{n_v} L\left(\mathrm{LGD}_i, \widehat{\mathrm{LGD}}_{i,m}\right).$$

Given the functional form of the loss function, the empirical mean $\widehat{\mathcal{L}}_m$ corresponds to a common measure of predictive accuracy such as the MSE, MAE, or RAE, with

$$\text{MSE: } \widehat{\mathcal{L}}_m = \frac{1}{n_v}\sum_{i=1}^{n_v}\left(\mathrm{LGD}_i - \widehat{\mathrm{LGD}}_{i,m}\right)^2,$$

$$\text{MAE: } \widehat{\mathcal{L}}_m = \frac{1}{n_v}\sum_{i=1}^{n_v}\left|\mathrm{LGD}_i - \widehat{\mathrm{LGD}}_{i,m}\right|,$$

$$\text{RAE: } \widehat{\mathcal{L}}_m = \sum_{i=1}^{n_v}\left|\mathrm{LGD}_i - \widehat{\mathrm{LGD}}_{i,m}\right| \Big/ \sum_{i=1}^{n_v}\left|\mathrm{LGD}_i - \overline{\mathrm{LGD}_i}\right|.$$

Other standard comparison criteria (deduced from these loss functions) can also be used for models comparison (see Yao et al., 2017; Nazemi et al., 2017, for instance), such as R-square, RMSE, etc. Whatever the criterion used, the LGD models are compared and ranked according to the realization of the statistic $\widehat{\mathcal{L}}_m$ on the test set. A model m is preferred to a model m' as soon as $\widehat{\mathcal{L}}_m < \widehat{\mathcal{L}}_{m'}$. Denote by $\widehat{m}^*$ the model associated to the minimum realization $\widehat{\mathcal{L}}_m$ for $m = 1, ..., \mathcal{M}$. Under some regularity conditions, $\widehat{\mathcal{L}}_m$ converges to $\mathcal{L}_m$, and the model $\widehat{m}^*$ corresponds to the optimal model m^* defined as

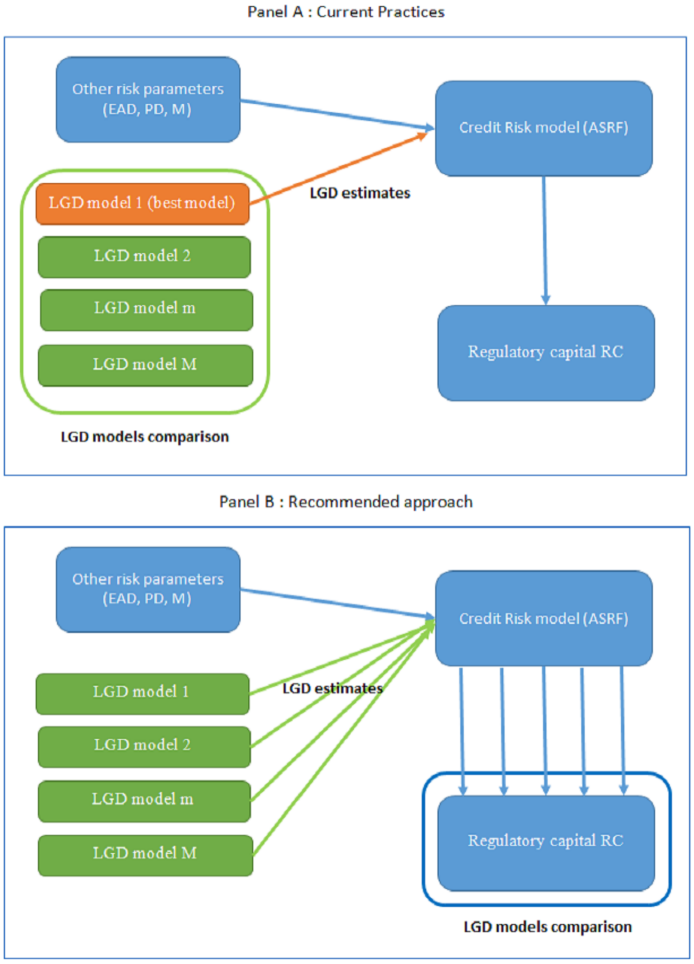
$$m^* = \underset{m=1,...,\mathcal{M}}{\arg\min}\, \mathbb{E}\left(L\left(\mathrm{LGD}_i, \widehat{\mathrm{LGD}}_{i,m}\right)\right).$$

This general approach has two main shortcomings. The first one is the lack of interpretability of the loss function. What do a MSE of 10% or a MAE of 27% exactly imply in terms of regulatory capital? These figures give no information about the estimation error made on the capital charge, and ultimately on the ability of the bank to absorb unexpected losses. The second pitfall is related to the two-step structure of the AIRB approach. The output of the bank's internal models, including the LGD models, are the Basel risk parameter estimates. These estimates are, in a second step, introduced in the ASRF model to compute the capital charge for each credit. As shown in the top panel

[8] If we denote by $e = x - \widehat{x}$ the error, the loss function is assumed to satisfy the following properties: (i) $L(0) = 0$, (ii) $\min L(e) = 0$ so that $L(e) \geq 0$, (iii) $L(e)$ is monotonically non-decreasing as e moves away from zero so that $L(e_1) \geq L(e_2)$ if $e_1 > e_2 > 0$, and if $e_1 < e_2 < 0$.

of Figure 2.1, the LGD model comparison is currently done independently of this second step and, as a consequence, of the ASRF model and the other risk parameters (EAD, PD, etc.).

Figure 2.1: Comparison of LGD models in the regulatory framework



2.3 Capital charge loss functions for LGD models

> "*Of great importance, and almost always ignored, is the fact that the economic loss associated with a forecast may be poorly assessed by the usual statistical metrics. That is, forecasts are used to guide decisions, and the loss associated with a forecast error of a particular sign and size is induced directly by the nature of the decision problem at hand.*". (Diebold and Mariano, 1995, page 2).

This quotation issued from the seminal paper of Diebold and Mariano (1995), perfectly illustrates the drawbacks of the current practices for LGD models comparison. In the BCBS framework, the LGD estimates are only *inputs* of the ASRF model which produces

the key estimate, namely the capital charge for credit risk. Consequently, the economic loss associated to the LGD models has to be assessed in terms of regulatory capital. The bottom panel of Figure 2.1 summarizes the alternative approach that we recommend for LGD model comparison. The LGD forecasts issued from the competing models and the other risk parameters (EAD, PD, etc.) are jointly used to compute the capital charges. Then, our approach consists in comparing the LGD models not in terms of forecasting abilities for the LGD itself, but in terms of forecasting abilities for the regulatory capital charges. The main advantage of this approach is that it favors the LGD model that leads to the lowest estimation errors associated to the loans with the highest EAD and PD.

2.3.1 Capital charge expected loss

The capital charge expected loss $\mathcal{L}_{CC,m}$ is simply defined as the expected loss defined in terms of regulatory capital charge, which is associated to a LGD model m. Formally, we have

$$\mathcal{L}_{CC,m} \equiv \mathcal{L}\left(\mathrm{RC}_i, \widehat{\mathrm{RC}}_{i,m}\right) = \mathbb{E}\left(L\left(\mathrm{RC}_i, \widehat{\mathrm{RC}}_{i,m}\right)\right),$$

where $L(.,.)$ is an integrable *capital charge* loss function with $L : \mathbb{R}^{+2} \to \mathbb{R}^{+}$, and

$$\mathrm{RC}_i = \mathrm{EAD}_i \times \mathrm{LGD}_i \times \delta(\mathrm{PD}) \times \gamma(\mathrm{M}_i),$$

$$\widehat{\mathrm{RC}}_{i,m} = \mathrm{EAD}_i \times \widehat{\mathrm{LGD}}_{i,m} \times \delta(\mathrm{PD}) \times \gamma(\mathrm{M}_i).$$

The variable RC_i denotes the risk contribution of the i^{th} credit, defined by the regulatory formula (Equation 2.1). This risk contribution depends on the risk parameters, namely EAD_i, LGD_i, and M_i. Notice that PD is not indexed by i, meaning that we consider the same default probability for all the credits. As we only consider defaulted credits in the test set, PD is fixed to an arbitrary value, typically close to 1. Similarly, $\widehat{\mathrm{RC}}_{i,m}$ denotes the estimated risk contribution for credit i, which is based on the individual risk parameters (EAD_i and M_i), the common value for the PD, and the LGD forecast issued from model m.

Given the functional form of $L(.,.)$, the empirical counterpart $\widehat{\mathcal{L}}_{CC,m}$ can be defined in terms of MSE, MAE, RAE, RMSE, R^2, or any usual criteria, with for instance

$$\text{Capital Charge MSE: } \widehat{\mathcal{L}}_{CC,m} = \frac{1}{n_v}\sum_{i=1}^{n_v}\left(\mathrm{RC}_i - \widehat{\mathrm{RC}}_{i,m}\right)^2,$$

$$\text{Capital Charge MAE: } \widehat{\mathcal{L}}_{CC,m} = \frac{1}{n_v}\sum_{i=1}^{n_v}\left|\mathrm{RC}_i - \widehat{\mathrm{RC}}_{i,m}\right|,$$

$$\text{Capital Charge RAE: } \widehat{\mathcal{L}}_{CC,m} = \sum_{i=1}^{n_v}\left|\mathrm{RC}_i - \widehat{\mathrm{RC}}_{i,m}\right| / \sum_{i=1}^{n_v}\left|\mathrm{RC}_i - \overline{\mathrm{RC}_i}\right|,$$

where n_v denotes the size of the test set of defaulted credits. Beyond these traditional statistical criteria, we also introduce asymmetric criteria especially designed to improve financial stability. These loss functions only penalize the capital charge underestimates and they do not take into account the overestimations. As the regulatory capital is designed to absorb the unexpected credit losses, any underestimate of this charge can threaten the bank's solvability. Thus, we propose asymmetric loss functions defined as

$$\text{Asymmetric MSE: } \widehat{\mathcal{L}}_{CC,m} = \frac{1}{n_v^+} \sum_{i=1}^{n_v} \left(\text{RC}_i - \widehat{\text{RC}}_{i,m}\right)^2 \times \mathbb{I}_{\left(\text{RC}_i > \widehat{\text{RC}}_{i,m}\right)},$$

$$\text{Asymmetric MAE: } \widehat{\mathcal{L}}_{CC,m} = \frac{1}{n_v^+} \sum_{i=1}^{n_v} \left|\text{RC}_i - \widehat{\text{RC}}_{i,m}\right| \times \mathbb{I}_{\left(\text{RC}_i > \widehat{\text{RC}}_{i,m}\right)},$$

where $\mathbb{I}_{(.)}$ denotes the indicator function that takes a value 1 when the event occurs and 0 otherwise, and n_v^+ is the number of defaulted credits for which we observe $\text{RC}_i > \widehat{\text{RC}}_{i,m}$. These loss functions are particularly suitable to compare LGD models which produce skewed LGD estimation errors (cf. Section 2.5).

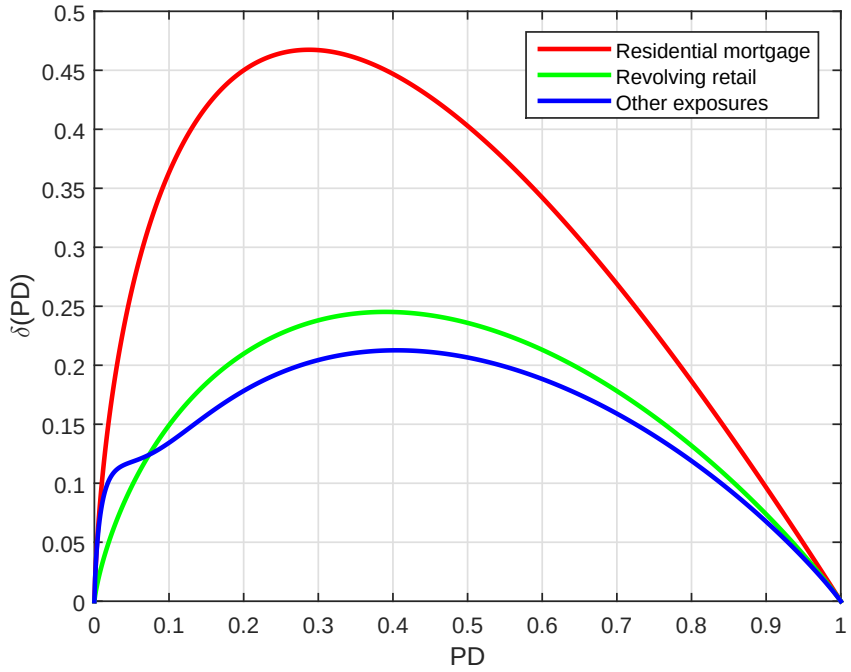
The expected loss $\mathcal{L}_{CC,m}$ has a direct economic interpretation. For instance, the capital charge MAE represents the average absolute estimation error observed between the capital charge estimates (associated to the LGD estimates issued from a given model) and the true ones (based on the observed LGD for the defaulted credit). Similarly, the asymmetric MSE corresponds to the variance of the capital charge underestimates produced by a given LGD model. Furthermore, these comparison criteria take into account the exposure and the maturity of the credits. Finally, the comparison rule for the LGD models is the same as before. A model m is preferred to a model m' as soon as $\widehat{\mathcal{L}}_{CC,m} < \widehat{\mathcal{L}}_{CC,m'}$. Denote by $\widehat{m}^*_{CC}$ the model associated to the minimum empirical mean $\widehat{\mathcal{L}}_{CC,m_{CC}}$ among the set of $\mathcal{M}$ models. Under some regularity conditions, $\widehat{\mathcal{L}}_{CC,m_{CC}}$ converges to $\mathcal{L}_{CC,m_{CC}}$, and allows to identify the optimal model in terms of capital charge expected loss.

As previously mentioned, the expected loss expressed in terms of capital charge depends on the value of PD chosen for the defaulted credits that belong to the test set. However, the *ranking* of the LGD models based on the capital charge expected loss, does not depend on the choice of the PD value. Indeed, since $\delta\,(\text{PD})$ is a constant term that does not depend on the contract i or the model m, the choice of PD does not affect the *relative* values of the expected losses observed for two alternative models, m and m'. This choice only affects the *absolute* value of the expected losses $\mathcal{L}_{CC,m}$ and $\mathcal{L}_{CC,m'}$.

Equation 2.2 implies that $\delta\,(1) = 0$ and $\delta\,(0) = 0$. As a consequence, the PD value has to be chosen on the interval $]0, 1[$. Here, we recommend to use the value PD^* that maximizes the value of $\delta\,(\text{PD})$ and hence, the regulatory capital since RC_i is an increasing function of $\delta\,(\text{PD})$. The profile of the capital charge coefficient $\delta\,(\text{PD})$ depends on the type of exposure (cf. Appendix 2.8.2) and is displayed on Figure 2.2. The capital charge

coefficient increases with PD until an inflexion point, and then decreases to 0 when the PD tends to 1. This profile is explained by the fact that once this inflexion point is reached, losses are no longer absorbed by the regulatory capital (which covers the unexpected bank's credit loss), but by the provisions done for the expected credit losses $\mathbb{E}(L)$. The maximum of the $\delta(.)$ function is reached for a PD value of 28.76% in the case of residential mortgage, 38.98% for revolving retail, and 40.45% for other exposures.

Figure 2.2: δ function for different types of retail exposure



2.3.2 Ranking consistency

The LGD models comparison can be based either on traditional LGD-based loss functions L_m or capital charge-based loss functions $L_{CC,m}$. Suppose that both approaches lead to the same models ranking (e.g, in the case of two models m and m', $L_m < L_{m'}$ and $L_{CC,m} < L_{CC,m'}$). Then, one should favor the simplest approach that only focuses on LGD errors. In this case, it is useless to collect additional data for other risk parameters (EAD, PD, maturity, etc.) and to compute the capital charges for each credit in order to compare LGD models. However, nothing guarantees ex-ante that both approaches will necessarily lead to consistent models' rankings.

The goal of this section is twofold. First, we determine the conditions under which the models ranking induced by a LGD-based loss function and the models ranking obtained with a capital charge-based loss function are consistent. Second, we show that these conditions are very particular and are likely to be violated in practice. Hence, this theoretical analysis illustrates the relevance of our comparison framework for LGD models, which is based on regulatory capital estimation errors.

Consider the following assumptions on the LGD loss functions.

Assumption A1: $L(x, \widehat{x}) = g(x - \widehat{x})$ *with* $g : R \to R^+$, *a continuous and integrable function.*

Assumption A2: *The function* $g(.)$ *is multiplicative:* $\forall k \in R$, $g(k(x - \widehat{x})) = g(k)\, g(x - \widehat{x})$.

Notice that assumptions A1 and A2 are satisfied by the usual loss functions considered in the LGD literature.[9]

Consider a set of $\mathcal{M}$ LGD models, indexed by $m = 1, \cdots, \mathcal{M}$. We refer to the ordering based on the expected loss as the true ranking and we assume that LGD-based expected losses are ranked as follows

$$\mathcal{L}_1 < \mathcal{L}_2 < ... < \mathcal{L}_{\mathcal{M}},$$

with $\mathcal{L}_m = \mathbb{E}(g(\varepsilon_{i,m}))$ and $\varepsilon_{i,m} = \text{LGD}_i - \widehat{\text{LGD}}_{i,m}$, $\forall m = 1, ..., \mathcal{M}$. Now, define the corresponding capital charge expected loss, $\mathcal{L}_{CC,m}$, for the model m as

$$\mathcal{L}_{CC,m} = \mathbb{E}(g(\eta_{i,m})),$$

with $\eta_{i,m} = \text{RC}_i - \widehat{\text{RC}}_{i,m}$. By definition of the regulatory capital charge, we have[10]

$$\eta_{i,m} = \text{EAD}_i \times \delta(\text{PD}) \times \gamma(\text{M}) \times \varepsilon_{i,m}.$$

Proposition 1. *The models' rankings produced by LGD-based and capital charge-based expected losses are consistent, i.e.* $\mathcal{L}_1 < \mathcal{L}_2 < ... < \mathcal{L}_{\mathcal{M}}$ *and* $\mathcal{L}_{CC,1} < \mathcal{L}_{CC,2} < ... < \mathcal{L}_{CC,\mathcal{M}}$, *as soon as,* $\forall m = 1, ..., \mathcal{M} - 1$

$$cov(g(EAD_i), g(\varepsilon_{i,m})) - cov(g(EAD_i), g(\varepsilon_{i,m+1})) < \mathbb{E}(g(EAD_i))(\mathcal{L}_{m+1} - \mathcal{L}_m).$$

The proof of Proposition 1 is reported in Appendix 2.8.3. Since $\mathcal{L}_m < \mathcal{L}_{m+1}$, the consistency condition of Proposition 1 is satisfied as soon as $cov(g(\text{EAD}_i), g(\varepsilon_{i,m})) < cov(g(\text{EAD}_i), g(\varepsilon_{i,m+1}))$. Thus, both rankings are consistent as soon as the covariances of the LGD forecast errors with the exposures are ranked in the same manner as the LGD models themselves. Consider a simple case with two LGD models A and B, where model A has a smaller LGD-based MSE than model B. Model A will have also a smaller MSE in terms of capital charge, if its squared LGD estimation errors are less correlated to the squared EAD than the errors of model B. For instance, if model B produces

[9] For instance, the quadratic loss function $L(x, \widehat{x}) = (x - \widehat{x})^2$ with $g(y) = y^2$ implies that $L(kx, k\widehat{x}) = L(g(k(x - \widehat{x}))) = k^2 (x - \widehat{x})^2 = g(k)\, g(x - \widehat{x})$. Regarding the absolute loss function $L(x, \widehat{x}) = |x - \widehat{x}|$ with $g(y) = |y|$, we have $L(kx, k\widehat{x}) = L(g(k(x - \widehat{x}))) = |k|\,|x - \widehat{x}| = g(k)\, g(x - \widehat{x})$.

[10] For simplicity, we assume that the credits have the same maturity M. In the general case, the consistency condition of the models' rankings can be easily deduced from the formula given in this benchmark case.

large LGD estimation errors for high exposures and low LGD errors for low exposures, whereas it is not the case for model A, both model comparison approaches will provide the same rankings. Obviously, this condition is very particular and in the general case, the two comparison approaches are likely to provide inconsistent LGD models' rankings. Proposition 1 has a direct interpretation in the special case where the exposures are independent from the estimation errors of the LGD models.

Corollary 1. *As soon as the EAD_i and the LGD estimation errors $\varepsilon_{i,m}$ are independent, the models' rankings based on the LGD and capital charge expected losses are consistent.*

The proof is provided in Appendix 2.8.4. This corollary implies that when credit exposures and LGD estimation errors $\varepsilon_{i,m}$ are independent, the current model comparison approach that consists to compare the MSE, MAE or RAE in terms of LGD estimation errors is sufficient. However, this independence assumption is likely to be violated in practice. First, even if the variables EAD_i and LGD_i are independent, it does not necessarily imply that EAD_i and $\text{LGD}_i - \widehat{\text{LGD}}_{i,m}$ are independent. Second, it is important to notice that the introduction of the EAD as an explanatory variable in the LGD model, does not necessarily guarantee that the EAD and estimation errors are independent. It depends on the model (linear or not) and the estimation method used. For instance, the independence assumption is satisfied for linear regression model estimated by OLS. Conversely, for nonlinear models or machine learning methods, such as regression tree, support vector regression, or random forest, the forecast errors may be correlated with the explanatory variables.

2.4 Comparison framework

We now propose an empirical application of our comparison approach for LGD models. In this section, we describe our dataset, the experimental set-up, and the six competing LGD models.

2.4.1 Data description

Our dataset consists in a portfolio of retail loans (credit and leasing contracts) provided by an international bank specialized in financing, insurance, and related activities for a worldwide leader automotive company. The initial sample includes 23,933 loans that defaulted between January 2011 and December 2016. For the more recent defaults, the recovery processes are not necessarily completed and we don't observe the bank's final loss. As a consequence, we exclude these contracts and limit our analysis to the 9,738 closed recovery processes for which we observe the final workout LGD. This approach has also been used by Gürtler and Hibbeln (2013) and Krüger and Rösch (2017) who recommend restricting the observation period of recovery cash flows to avoid the under-

representation of long workout processes, which might result in an underestimation of LGD and regulatory capital.

The final sample covers 6,946 credit and 2,792 leasing contracts granted to individual (6,521 contracts) and professional (3,217 contracts) Brazilian customers that defaulted between January 2011 and November 2014. For each contract, we observe the characteristics of the loan (e.g. type of contract, interest rate, original maturity, etc.) and the borrower (professional, individual, etc.), as well as the workout LGD and EAD. All the contracts are in default, so by definition their PD is equal to 1 (certain event). However, we collect for each contract the PD calculated by the internal bank's risk model one year before the default occurs. For the contracts that entered in default in less than one year, the PD is set to the value determined by the internal bank's risk model at the granting date. Hence, we have all the information to compute the regulatory capital charge for each credit. Finally, we complete the database with three macroeconomic variables, namely (1) the quarterly Brazilian GDP growth rate, (2) the monthly unemployment rate and (3) the monthly average of the daily interbank rates.[11] For each contract, the macroeconomic variables are considered at the date of default. Their introduction in LGD models aims to capture the influence of the business cycles on the recovery process, as suggested by Bellotti and Crook (2012) and Tobback et al. (2014). The description of the dataset variables is reported in Appendix 2.8.5.

Table 2.1 displays some descriptive statistics (mean, q25, and q75) about the LGD, PD, and EAD by year, exposure and customer type. The number of defaulted contracts per year ranges between $1,573$ and $2,946$. The mean of losses is equal to 33.12%, a similar value to that reported by Miller and Töws (2018) for a German leasing company, and tends to decrease between 2011 and 2014. The average PD is equal to 9.53%, but this figure hides a large heterogeneity since the PD values range from less than 1% to 71%, whereas 3/4 of the PD values are below 11.08%. Similarly, the EAD ranges from less than 1 BRL to $123,550$ BRL, with an average exposure equal to $20,830$ BRL. The credit and leasing contracts exhibit the same level of exposure, but the average PD is higher for leasing than for credit (10.24% against 9.25%). As often reported in the literature, the LGD for leasing contracts (32.01%) is slightly smaller than for credits (33.57%). We observe that the average PD is higher for the professional clients than for individuals, but their average LGD is smaller due to their highest collateral.

The empirical distribution of the 9,738 workout LGDs is displayed on the top panel of Figure 2.3. Three remarks should be made here. First, 10.58% of the defaulted contracts have a recovery rate that exceeds 100%, with a maximum value of 116.14%, due to the

[11]The data have been collected from OECD.Stat databases: Monthly Monetary and Financial Statistics (MEI), Quarterly National Accounts (QNA) and Labour Market Statistics (LMS).

Table 2.1: Descriptive statistics on LGD, PD, and EAD

	Nb of obs	LGD (%)			PD (%)			EAD (thous. BRL)		
Panel A. All loans										
	–	q25	mean	q75	q25	mean	q75	q25	mean	q75
	9,738	5.03	33.12	57.22	0.78	9.53	11.08	10.63	20.83	28.11
Panel B. By year										
	–	q25	mean	q75	q25	mean	q75	q25	mean	q75
2011	1,573	6.82	40.95	77.29	0.80	10.56	16.93	11.12	20.80	27.78
2012	2,430	8.09	39.14	67.31	0.71	8.36	8.98	11.67	22.20	29.57
2013	2,946	7.16	36.43	61.98	0.72	9.07	11.00	11.02	21.43	28.50
2014	2,789	2.85	19.97	22.92	1.05	10.46	17.42	9.14	19.02	26.18
Panel C. By exposure										
	–	q25	mean	q75	q25	mean	q75	q25	mean	q75
Credit	6,946	5.03	33.57	56.83	0.77	9.25	10.14	10.36	20.94	28.40
Leasing	2,792	5.02	32.01	58.15	0.84	10.24	14.79	11.28	20.57	27.39
Panel D. By customer type										
	–	q25	mean	q75	q25	mean	q75	q25	mean	q75
Individuals	6,521	5.15	33.80	59.34	0.70	8.75	9.89	11.08	20.39	27.61
Professionals	3,217	4.64	31.75	52.70	1.11	11.12	14.79	9.77	21.73	29.71

workout costs.[12] Second, we also confirm that the kernel density estimate of the LGD distribution is bimodal (bottom panel of Figure 2.3). Finally, the LGD distributions for the credit and leasing contracts are relatively close, except for the right part of the distribution. The outcome of a loss event is less severe for leasing than for credit. This difference illustrates the role of the collateral in the recovery processes (in the case of leasing, the vehicle belongs to the bank and plays the same role as a collateral).

Finally, Figure 2.4 displays the scatter plot of LGD versus the logarithm of EAD. We find a positive correlation between the LGD and the EAD. The correlation is relatively small (0.11), but significant. This observation justifies the introduction of the exposure as explanatory variable in our LGD models.

[12]Notice that this percentage is smaller than those generally observed in the literature. For instance, in the leasing industry, Schmit (2004) report that up to 59% of all defaulted contracts in their sample have a recovery rate that exceeds 100%.

Figure 2.3: Empirical distribution of the LGDs

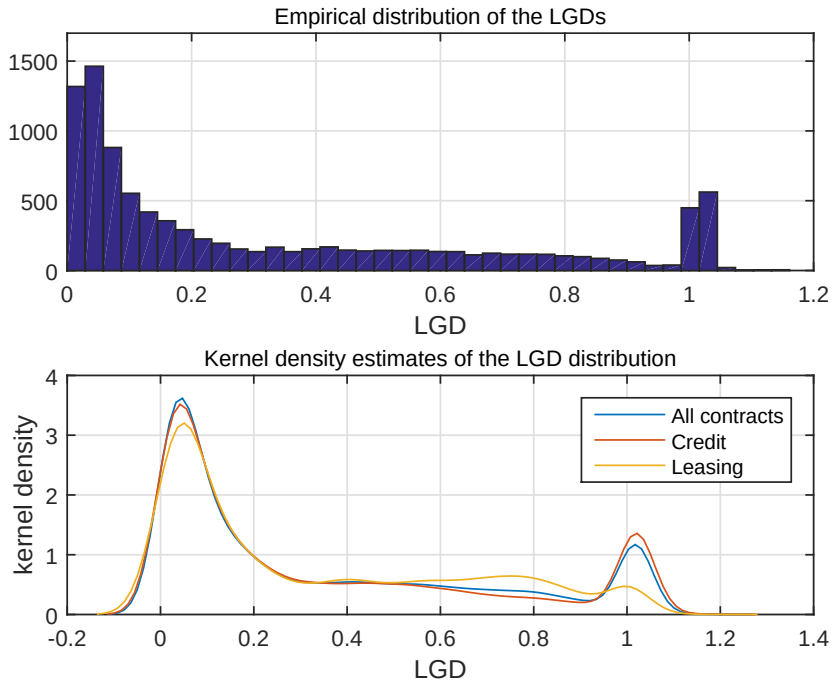
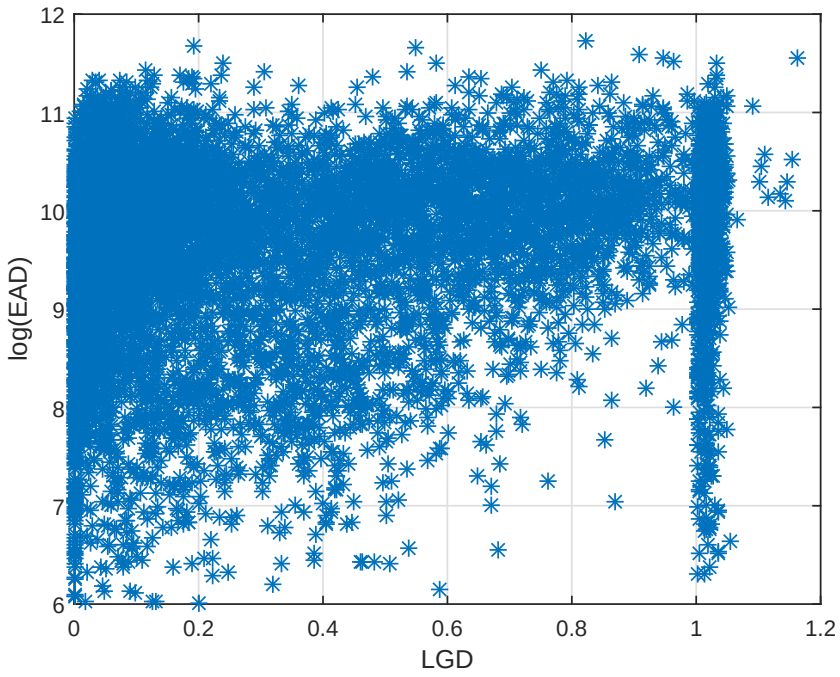


Figure 2.4: Scatter plot of LGD versus log(EAD)



2.4.2 Competing LGD models

For our comparison, we consider six competing LGD models which are commonly used in academic literature, namely (1) the fractional response regression (FRR) model, (2) the regression tree (TREE), (3) the random forest (RF), (4) the gradient boosting

(GB), (5) the artificial neural network (ANN), and (6) the least squares support vector regression (LS-SVR).[13]

The FRR model allows to estimate the conditional mean of a continuous variable defined over $[0, 1]$. It is often considered as a benchmark parametric model for LGD (see Bastos, 2010; Qi and Zhao, 2011, etc.). The TREE model consists in recursively partitioning the covariates space according to a prediction error and then, to fit a simple mean prediction within each partition. Here, we consider the CART algorithm which has been applied to LGD estimation by Matuszyk et al. (2010), Qi and Zhao (2011), Bastos (2010, 2014), and Loterman et al. (2012), among many others. The RF is a boostrap aggregation method of regression trees, trained on different parts of the same training set, with the goal of reducing overfitting. This model has been used for LGD modelling by Bastos (2014) and Miller and Töws (2018), among others. The ANNs are a class of flexible non-linear models. It produces an output value by feeding inputs through a network whose subsequent nodes apply some chosen activation function to a weighted sum of incoming values. Here, we consider a multilayer perceptron similar to that used by Qi and Zhao (2011) or Loterman et al. (2012) for the LGD forecasts. Finally, we consider the LS-SVR model. Compared to other support regression techniques, the LS-SVR has a low computational cost as it is equivalent to solving a linear system of equations. Loterman et al. (2012), Yao et al. (2015, 2017) and Nazemi et al. (2017) illustrate the good predictive performance of LS-SVR for LGD modelling. For more details and references about these models, see Appendix 2.8.6.

2.4.3 Experimental set-up

The six competing models are estimated on a training set of 7,791 loans (80% of the sample) and the out-of-sample LGD forecasts are evaluated on a test set of 1,947 loans. For each model, we consider the same set of explanatory variables including the exposure at default, the original maturity, the time to default, the relative duration (defined as the ratio between time to default and maturity), the interest or renting rate, the type of exposure (credit versus leasing), the customer type (individual or professional), the state of the car (new or second-hand), and the brand of the car.[14] Appendix 2.8.5 displays the description of these independent variables, as well as descriptive statistics. For each model and each contract within the test set, we compute the LGD forecast and the regulatory capital charge (based on the LGD forecast or the true LGD value), by using the other Basel risk parameters. The regulatory capital charges are computed with a PD of 40.45%, which corresponds to the maximal charge for the retail exposures.

[13] We thank an anonymous referee for the suggestion of the LS-SVR model.

[14] An extended set of information with macroeconomic variables will be considered in Section 2.6.

The hyperparameters of the machine learning algorithms are tuned using five-fold cross validation on the training set. They were all selected based on the MSE criterion. For the TREE model, the procedure leads to select the optimal depth of the tree. For the GB, the cross validation procedure determines the optimal number of iterative training cycles (candidates 10, 50, 100, 250, 500, 1,000 are considered). The same approach is applied for the RF for identifying the optimal number of trees in the forest. For the ANN, the five-fold cross validation procedure is used to select the number of hidden neurons (a value from 1 to 20 is considered). A logistic function is used as the activation function in hidden layer neurons. Finally, in order to implement the LS-SVR, we consider a radial basis function kernel. The radial basis function kernel parameter σ and the regularization parameter C, are tuned using five-fold cross validation on the training dataset. A grid search procedure firstly evaluates a large space of possible hyperparameter combinations to determine suitable starting candidates. The search limits are set to $[\exp(-10), \exp(10)]$. Then, given these starting values, the hyperparameters σ and C are optimized with a simplex routine so as to find the combination that minimizes the MSE.

2.5 Empirical results

2.5.1 LGD and RC estimation errors

Table 2.2 displays some figures about LGD and regulatory capital forecast errors, respectively defined by $\mathrm{LGD}_i - \widehat{\mathrm{LGD}}_{i,m}$ and $\mathrm{RC}_i - \widehat{\mathrm{RC}}_{i,m}$. Notice that, given this notation, a positive error implies an underestimation of the true value. We observe that the empirical means of the LGD and RC forecast errors are slightly positive, whereas the medians are generally negative. This feature is due to the positive skewness observed for the errors of all models. We also observe that the excess kurtosis for the regulatory capital are positive, indicating fat tails for the errors distribution. When one considers the LGD errors, the LS-SVR, ANN and the GB models have the smallest variance. However, it is no longer the case for the ANN when one considers the RC forecast errors. This result clearly illustrates the usefulness of our comparison approach.

The kernel density estimates of the forecast errors distributions displayed in Figure 2.5 confirm the positive skewness of the errors' distributions. This figure shows that one can frequently observe capital requirement underestimates larger than 4,000 BRL, whereas similar overestimates are much rarer. Such a feature is clearly problematic within a regulatory perspective, and justifies the use of asymmetric loss functions for comparing LGD models.

Figure 2.6 displays the scatter plot of the LGD forecast errors (x-axis) and the RC forecast errors (y-axis), obtained with the GB model. Each point represents a contract (credit or leasing). This plot shows the great heterogeneity that exists between both

Table 2.2: Descriptive statistics on the LGD and regulatory capital forecast errors

	FRR	ANN	TREE	LS-SVR	RF	GB
	LGD errors					
mean	0.011	0.011	0.010	0.011	0.009	0.010
median	-0.136	-0.122	-0.142	-0.127	-0.114	-0.138
variance	0.117	0.116	0.118	0.115	0.117	0.116
skewness	0.824	0.804	0.817	0.828	0.791	0.830
excess kurtosis	-0.618	-0.525	-0.605	-0.551	-0.473	-0.626
	Regulatory capital errors					
mean	60	44	66	56	38	66
median	-257	-231	-256	-233	-232	-257
variance	3,813,596	3,799,691	3,730,561	3,698,999	3,737,503	3,717,518
skewness	0.55	0.41	0.80	0.58	0.61	0.79
excess kurtosis	2.52	2.83	2.01	2.54	2.73	1.91

type of errors. Due to the differences in EAD across borrowers, the magnitudes of the RC errors can drastically differ for the same level of LGD forecast error. Consider the two credits represented by the symbols A and B, with an EAD equal to $61,271$ BRL and $2,034$ BRL, respectively. For the same level of LGD forecast error (64.3%), the GB slightly underestimates the capital requirement (278 BRL) in the case of the credit B, whereas the underestimation reaches $8,367$ BRL in the case of credit A. Obviously, from a regulatory perspective, the second LGD error should be more penalized than the first one, as its consequence on the RC estimates are more drastic. The dispersion of the observations within the y-axis fully justifies our comparison approach for LGD models, based on loss functions expressed in terms of capital charge. Furthermore, the scatter plot confirms the asymmetric pattern of the errors distribution associated to the GB model. This model leads to relatively small overestimates (negative errors), both for LGD and RC, while it leads to large underestimates (positive errors). Thus, any competing LGD model that leads to less severe underestimates than the GB should be preferred from a regulatory perspective. For this reason, we recommend the use of asymmetric loss functions to compare the LGD models. These features (heterogeneity and asymmetry) are not specific to the GB model, even if the skewness of the error distribution is more pronounced for this model compared to the other ones. The scatter plots of the LGD and RC errors are quite similar for the six competing models (cf. Appendix 2.8.7). This

Figure 2.5: Kernel density estimate of the estimation error

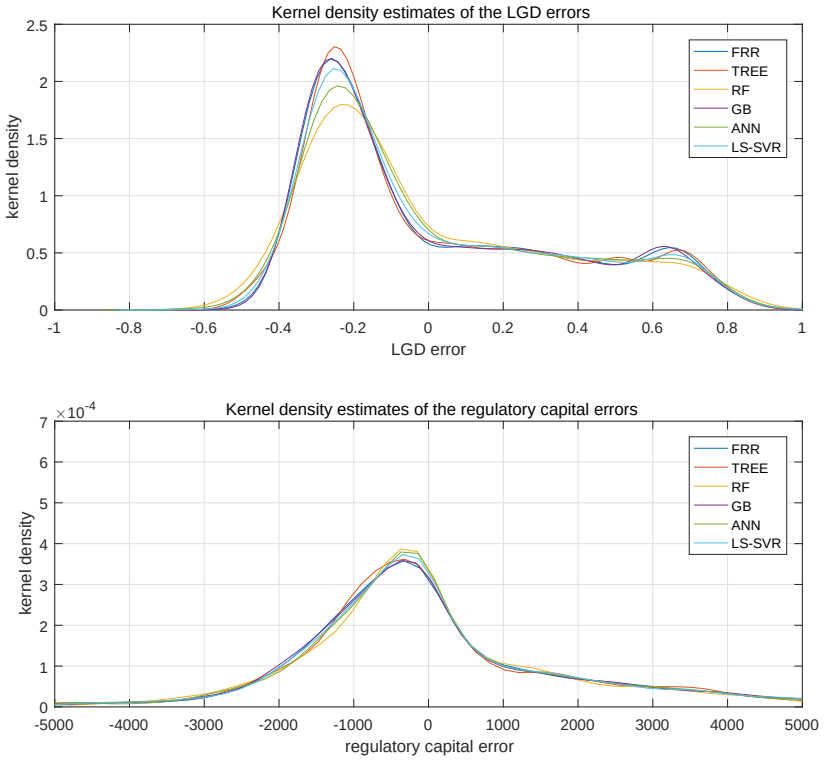
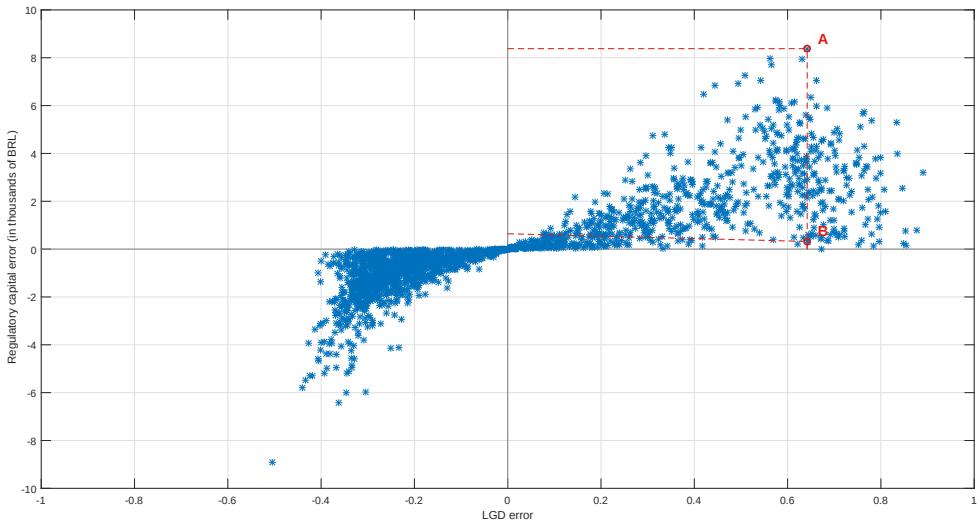


Figure 2.6: Scatter plot of LGD versus regulatory capital forecast errors for the GB model



similarity comes from two sources. First, it is due to the fact that we use the same set of covariates for all the models. Second, this similarity is also related to the definition

of the regulatory capital that leads to increase the RC errors dispersion and induces a similar appearance for the different scatter plots (i.e. for the different LGD models).

2.5.2 LGD models' rankings

Table 2.3 displays the models' rankings issued from two usual loss functions, namely the MSE and MAE, associated to the LGD (column 1) and regulatory capital (column 2) forecast errors. We also report the rankings based on the asymmetric expected losses (columns 3 and 4) that only penalize the LGD forecast errors which lead to underestimating the regulatory capital. The values of the losses (MSE, MAE) are displayed in Appendix 2.8.8, along with the corresponding R^2 and RMSE.

Table 2.3: Models' rankings based on LGD and capital charge expected loss functions

Ranking	LGD Loss	CC Loss	Asym. LGD Loss	Asym. CC Loss
		Mean squared error		
1.	LS-SVR	LS-SVR	RF	RF
2.	GB	GB	LS-SVR	ANN
3.	ANN	TREE	FRR	LS-SVR
4.	FRR	RF	ANN	FRR
5.	RF	ANN	TREE	TREE
6.	TREE	FRR	GB	GB
		Mean absolute error		
1.	LS-SVR	RF	RF	RF
2.	RF	LS-SVR	LS-SVR	LS-SVR
3.	ANN	ANN	FRR	ANN
4.	GB	TREE	ANN	FRR
5.	FRR	GB	TREE	TREE
6.	TREE	FRR	GB	GB

Note: The two columns LGD Loss and CC Loss correspond to the models' rankings obtained with loss functions (MSE or MAE) respectively defined in terms of LGD forecast errors and regulatory capital forecast errors. The columns Asym. LGD Loss and Asym. CC Loss display the rankings obtained with asymmetric loss functions either defined in terms of LGD or regulatory capital forecast errors.

The models' rankings that we obtain with the MSE or MAE criteria computed with LGD estimation errors, are similar to those generally obtained in the literature. As in Loterman et al. (2012), Yao et al. (2015, 2017), and Nazemi et al. (2017), we observe that the LS-SVR model outperforms the five competing LGD models. As in Bastos

(2010), Qi and Zhao (2011), Loterman et al. (2012), Hartmann-Wendels et al. (2014) or Miller and Töws (2018), we observe that non-parametric approaches such as LS-SVR, ANN, and RF generally yield better predictive performances than the FRR parametric model. Furthermore, we observe similar values for the RMSE of the LS-SVR model (see Appendix 2.8.8) as those reported in Yao et al. (2015) or Loterman et al. (2012).

However, considering the loss function based on RC estimation errors leads to different conclusions. Regarding the MSE criterion, the LS-SVR is still ranked as the best model, whatever the errors considered (LGD or RC). The GB is also consistently ranked as the second best model. But, the rest of the LGD models ranking is not consistent. For instance, ANN is identified as the third-best model with the LGD loss, while it holds the penultimate rank with the capital charge loss. Conversely, the TREE model is ranked third with the capital charge loss while it is ranked at the last position with the LGD-based loss. Similar results are obtained when one compares the rankings associated to the asymmetric LGD and RC-based loss functions. These inversions prove that the ranking consistency condition of Proposition 1 is not valid, at least in our sample, for some couples of models. Ranking the LGD models according to their LGD forecast errors or their RC forecast errors is not equivalent. The conclusions are similar for the MAE criteria. In this case, the LS-SVR is the best model when one considers LGD forecast errors, but the RF should be preferred when one considers RC forecast errors.

Furthermore, our results highlight the usefulness of asymmetric loss functions. These functions penalize more the models with the largest positive errors (underestimates), as the GB, for instance. Indeed, the GB is ranked as the worst model when the MSE is computed with asymmetric LGD or RC forecast errors. It also remains the worst model when one considers asymmetric MAE, due to the large skewness of its forecast errors. On the contrary, the RF exhibits the lowest asymmetric MSE and MAE whatever the type of errors considered. These findings clearly illustrate the fact that ranking the LGD models according to their estimation errors (whatever their signs) or to their underestimates, is not equivalent.

These conclusions are confirmed by rank correlation tests. Table 2.4 displays the Spearman's and Kendall's rank correlation coefficients ρ, with the p-values associated to the null hypothesis $\rho = 0$. These p-values are computed with the exact permutation distributions for small sample sizes. Our goal is to test if the models' rankings differ significantly. We compare the rankings obtained with (1) MSE (MAE) based on LGD errors, (2) MSE (MAE) based on RC errors, and (3) asymmetric MSE (MAE) based on LGD or RC errors. The conclusions are clear-cut: for a 5% significance level, the rank correlation coefficients are never statistically different from 0. These tests confirm that using a regulatory capital-based criterion (MSE or MAE) do not provide the same ranking as that obtained with a similar criterion only based on LGD estimation errors.

Furthermore, the tests show that the rankings are sensitive to the choice of a symmetric or asymmetric criterion.

Table 2.4: Spearman's and Kendall's rank correlation coefficients

	Spearman rank correlation		Kendall rank correlation	
		MSE		
	ρ	p-value	ρ	p-value
LGD vs. CC	0.4857	0.1778	0.3333	0.2347
LGD vs. Asym. LGD	-0.0286	0.5403	-0.0667	0.6403
LGD vs. Asym. CC	-0.0857	0.5986	-0.0667	0.6403
		MAE		
	ρ	p-value	ρ	p-value
LGD vs. CC	0.7714	0.0514	0.6000	0.0681
LGD vs. Asym. LGD	0.6571	0.0875	0.4667	0.1361
LGD vs. Asym. CC	0.7714	0.0514	0.6000	0.0681

Note: The ρ coefficients denote the Spearman's or Kendall's rank correlation coefficients. The p-values are computed under the null hypothesis $\rho = 0$ and are based on the exact permutation distributions for small sample sizes. LGD and CC respectively denote the rankings based on LGD or regulatory capital errors. Asym. LGD and Asym. CC respectively denote the rankings obtained with asymmetric loss functions based on LGD or regulatory capital errors.

Beyond the rank correlations, we also investigate if the choice of the loss function may affect the pairwise models comparison. In Appendix 2.8.9, we display the paired t-tests for comparisons of MSE and MAE, based on LGD estimation errors or regulatory capital estimation errors. The logic here is similar to Yao et al. (2015, 2017) or Nazemi et al. (2017). The main takeaway of these pairwise tests is the following. Considering out-of-sample criteria (MSE or MAE) based on regulatory capital estimation errors sometimes change the conclusions of the pairwise comparison of LGD models performance. For instance, if we consider the MSE criterion based on LGD errors, the LS-SVR model outperforms all other models (except the GB), as the differences between the corresponding MSEs are always positive for a 5% significance level. However, when considering the MSE based on regulatory capital errors, the MSE difference between the LS-SVR and the TREE model is not significant, meaning that both models lead to similar regulatory capital estimation errors. Similarly, the LS-SVR does not make significant improvements compared to the RF when one considers regulatory capital errors.

2.6 Robustness checks

Our empirical results are robust to a variety of robustness checks. Firstly, instead of considering a common PD for the computation of the capital charges, we use the individual PD calculated by the internal bank's risk model for each credit one year before the default occurs. The corresponding LGD models' rankings are reported in Table 2.5. The rankings based on the MSE are similar to those obtained with a common PD (cf. Table 2.3). For the symmetric capital charge MSE, the only change concerns the RF and the TREE models. For the asymmetric MSE, the ranking changes for the LS-SVR, the ANN, the GB and TREE models. But, we still observe ranking inversions compared to the ranking based on the LGD loss functions.

Table 2.5: Models' rankings based on Basel PDs

	LGD Loss	CC Loss	Asym. LGD Loss	Asym. CC Loss
		Mean squared error		
1.	LS-SVR	LS-SVR	RF	RF
2.	GB	GB	LS-SVR	LS-SVR
3.	ANN	RF	FRR	ANN
4.	FRR	TREE	ANN	FRR
5.	RF	ANN	TREE	GB
6.	TREE	FRR	GB	TREE
		Mean absolute error		
1.	LS-SVR	RF	RF	RF
2.	RF	LS-SVR	LS-SVR	LS-SVR
3.	ANN	ANN	FRR	ANN
4.	GB	TREE	ANN	FRR
5.	FRR	GB	TREE	TREE
6.	TREE	FRR	GB	GB

Note: The two columns LGD Loss and CC Loss correspond to the models' rankings obtained with loss functions (MSE or MAE) respectively defined in terms of LGD forecast errors and regulatory capital forecast errors (computed with Basel PD values). The columns Asym. LGD Loss and Asym. CC Loss display the rankings obtained with asymmetric loss functions either defined in terms of LGD or regulatory capital forecast errors.

Secondly, we also consider the same type of regressions by excluding the exposure at default from the set of explanatory variables. The qualitative results (not reported) remain the same: we observe a global inconsistency of the LGD models' rankings based on the LGD estimates or on the capital charge estimates. So, include (or exclude) the

EAD as explanatory variable in the LGD models, has no consequence on the validity of the condition of Proposition 1, since we only consider non-linear LGD models in our application.

Finally, we extend the set of explanatory variables by considering three macroeconomic variables in order to capture the influence of the business cycles on the recovery process, as suggested by Schuermann (2004), Bellotti and Crook (2012), and Tobback et al. (2014). These variables are the Brazilian GDP growth, the unemployment and interbank rates. Table 2.6 displays the corresponding LGD models' rankings. With the MSE criterion, the RF outperforms all competing models whatever the loss function considered. It is also the case for the MAE criterion. As in the previous cases, we observe a ranking inconsistency for other models, meaning that the condition of Proposition 1 is not valid for these couples of models. The values of the losses (MSE, MAE) are displayed in Appendix 2.8.8, along with the corresponding R^2 and RMSE. Our results are similar to those obtained in the literature. For instance, Hartmann-Wendels et al. (2014) who examine three leasing datasets, report a MAE that ranges from 0.2710 to 0.3370 for the TREE model (0.2768 in our case), while their RMSE takes values between 0.3462 and 0.3958 depending on the dataset (0.3343 in our case). We also get similar results for the RF as those reported in Miller and Töws (2018). Within their sample, the authors obtain a MAE of 0.3272 (0.2705 in our case) and a MSE of 0.1722 (0.1092 in our case). We obtain relatively low R^2 values (around 10%) when we consider LGD errors, but the R^2 reaches higher values (around 35%) when considering RC errors.

Beyond the rankings analysis, we report in Appendix 2.8.10 the marginal effects of the debt characteristics and macroeconomic variables on the LGD estimates obtained within the FRR model. Our qualitative results are similar to those obtained in the literature. As in Bastos (2010), the credit interest rate (fixed at the beginning of the credit contract) positively affects the LGD. This positive effect reflects the fact that the risky clients, who have the lowest collateral, have generally also the highest interest rates and in fine the lowest recovery rates. The original maturity has also a positive and significant effect on LGD. Indeed, for a given retail credit or leasing contract, longer maturities are generally negotiated by riskiest clients with the lowest collateral and revenues, and as a consequence the highest LGD. Contrary to Schuermann (2004), we observe a significant and positive impact of the EAD. The brand and the characteristics (new or second hand) of the car, the customer type (professional or individual), and the credit type (leasing versus standard credit) do not significantly impact the LGD. Finally, the time to default has a negative impact meaning that bank generally suffers limited losses for contracts that default close to their maturity. This result is similar to that obtained by Bellotti and Crook (2012) who found a negative and significant impact for the date of default.

Table 2.6: Models' rankings based on LGD and capital charge expected loss functions: LGD models with macroeconomic variables and common PD

	LGD Loss	CC Loss	Asym. LGD Loss	Asym. CC Loss
			Mean squared error	
1.	RF	RF	RF	RF
2.	LS-SVR	LS-SVR	ANN	ANN
3.	GB	TREE	TREE	LS-SVR
4.	ANN	GB	LS-SVR	TREE
5.	TREE	ANN	GB	GB
6.	FRR	FRR	FRR	FRR
			Mean absolute error	
1.	RF	RF	RF	RF
2.	LS-SVR	LS-SVR	ANN	ANN
3.	ANN	ANN	LS-SVR	LS-SVR
4.	GB	TREE	TREE	TREE
5.	TREE	GB	GB	GB
6.	FRR	FRR	FRR	FRR

Note: The two columns LGD Loss and CC Loss correspond to the models' rankings obtained with loss functions (MSE or MAE) respectively defined in terms of LGD forecast errors and regulatory capital forecast errors. The columns Asym. LGD Loss and Asym. CC Loss display the rankings obtained with asymmetric loss functions either defined in terms of LGD or regulatory capital forecast errors.

Concerning macroeconomic variables, we observe that the interbank interest rate (measured at the date of default) has a negative and significant coefficient. Tobback et al. (2014) also find a negative impact of the Federal Funds rate and explain it by the fact that a higher Federal Funds rate decreases the ability of the borrowers to pay off already defaulted loans. We observe that the unemployment rate has a significant positive impact on LGDs, as in Tobback et al. (2014). Finally, we observe that the GDP growth rate coefficient is negative, but non-significant at the 5% level. During a period of economic boom, banks are willing to issue loans to more risky borrowers against a high return. Therefore, the expansion and peak phases of the business cycle are accompanied by an accumulation of risks which result in greater losses once the growth starts slowing down. However, as pointed out by Tobback et al. (2014), one should be very cautious about interpreting the GDP growth effect, as the peak phase of the business cycle generally corresponds to a low growth percentage of GDP.

2.7 Conclusion

LGD is one of the key modelling components of the credit risk capital requirements. According to the AIRB approach adopted by most major international banks, the LGD forecasts are issued from internal risk models. While the practices seem to be well established for the PD modelling, no particular guideline has been proposed concerning how LGD models should be compared, selected, and evaluated. As a consequence, the model benchmarking method generally adopted by banks and academics simply consists in evaluating the LGD forecasts on a test set, with standard statistical criteria such as MSE, MAE, etc., as for any continuous variable. Thus, the LGD model comparison is done regardless of the other Basel risk parameters and by neglecting the impact of the LGD forecast errors on the regulatory capital. This approach may lead to select a LGD model that has the smallest MSE among all the competing models, but that induces small errors on small exposures, but large errors on large exposures.

We propose an alternative comparison methodology for the LGD models which is based on expected loss functions expressed in terms of regulatory capital charge. These loss functions penalize more heavily the LGD forecast errors associated to large exposure or to long credit maturity. We also define asymmetric loss functions that only penalize the LGD models which lead to underestimating the regulatory capital, since these underestimations weaken the bank's ability to absorb unexpected credit losses. Using a sample of credits provided by an international bank, we illustrate the interest of our method by comparing the rankings of six competing LGD models. Our approach allows to identify the best LGD models associated with the lowest estimation errors on the regulatory capital. Besides, the empirical results confirm that the ranking based on a naive LGD loss function are generally different from the models ranking obtained with the capital charge symmetric (or asymmetric) loss.

A natural extension of our work includes the identification of a Model Confidence Set (Hansen et al., 2011) that contains the "best" LGD models for a given level of confidence and a given criterion. This method of "models clustering", based on pairwise t-tests and an iterative algorithm, have been recently used to compare conditional risk measures (Hurlin et al., 2017) and could be adapted to compare LGD models.

2.8 Appendix

2.8.1 Appendix A: Asymptotic Single Risk Factor model

Here, we detail the sketch of the proof of the regulatory formula for the credit capital charge (for more details, see Gouriéroux and Tiomo, 2007; Roncalli, 2009; Genest and Brie, 2013). Let us consider a portfolio of n credits indexed by $i = 1, ..., n$. The portfolio loss is equal to

$$L = \sum_{i=1}^{n} \text{EAD}_i \times \text{LGD}_i \times D_i,$$

where EAD_i is the exposure at default for the i^{th} credit (assumed to be constant), LGD_i is the loss given default (random variable) and D_i is a binary random variable that takes a value 1 if there is a default before the residual maturity M_i and 0 otherwise. Formally, $D_i = 1_{(\tau_i \leq M_i)}$ where τ_i is the default time (random variable).

Assumption A1: *The default depends on a set of factors X and we denote by x the realization of X.*

Assumption A2: *The loss given default LGD_i is independent from the default time τ_i.*

Assumption A3: *The default times τ_i, $i = 1, ..., n$ are independent conditionally to the X factors.*

Assumption A4: *The portfolio is infinitely fine-grained, which means that there is no concentration, with*

$$\lim_{n \to \infty} \max \frac{\text{EAD}_j}{\sum_{i=1}^{n} \text{EAD}_i} = 0 \quad \forall j.$$

Under assumptions A1-A4, it is possible to show that the conditional distribution of L given X degenerates to the conditional expectation $\mathbb{E}_X(L) = \mathbb{E}(L|X = x)$ and we get

$$L|X \xrightarrow{p} \mathbb{E}_X(L) = \sum_{i=1}^{n} \text{EAD}_i \times \mathbb{E}(\text{LGD}_i) \times p_i(x),$$

where $p_i(x) = \mathbb{E}_X(D_i) = \mathbb{E}(D_i = 1|X = x)$ is the conditional default probability. Notice that under assumption A2, $\mathbb{E}_X(\text{LGD}_i) = \mathbb{E}(\text{LGD}_i)$. As a consequence, the portfolio loss has a marginal distribution given by

$$L \xrightarrow{d} g(X) = \sum_{i=1}^{n} \underbrace{\text{EAD}_i}_{\text{constant term}} \times \underbrace{\mathbb{E}(\text{LGD}_i)}_{\text{constant term}} \times \underbrace{p_i(X)}_{\text{random var.}}.$$

Denote by F_L the cdf of L such that $F_L(l) \equiv \Pr(L \leq l) = \Pr(g(X) \leq l)$.

Assumption A5: There is only one factor X, with a cdf $F_X(.)$ and $p_i(X)$ is a decreasing function of X.

Under assumption A5, the α-VaR of the portfolio loss L is defined as $VaR_L(\alpha) = F_L^{-1}(\alpha) = g\left(F_X^{-1}(1-\alpha)\right)$ or equivalently by

$$VaR_L(\alpha) = \sum_{i=1}^{n} \text{EAD}_i \times \mathbb{E}(\text{LGD}_i) \times p_i\left(F_X^{-1}(1-\alpha)\right) = \sum_{i=1}^{n} RC_i,$$

where RC_i denotes the risk contribution of the credit i. The VaR of an infinitely fine-grained portfolio can be decomposed as a sum of independent risk contributions, since RC_i only depends on the characteristics of the i^{th} credit (exposure at default, loss given default and probability of default). Similarly, the marginal loss expectation is defined as

$$\mathbb{E}(L) = \sum_{i=1}^{n} \text{EAD}_i \times \mathbb{E}(\text{LGD}_i) \times p_i,$$

where $p_i = \Pr(D_i = 1)$ corresponds to the unconditional probability of failure.

Assumption 6: *Let Z_i be the normalized asset value of the entity i. The default occurs when Z_i is below a given barrier B_i (level of debt), with*

$$D_i = 1 \quad \text{if} \quad Z_i \leq B_i.$$

Assumption 7: *The asset value Z_i depends on a common risk factor X and an idiosyncratic risk factor ε_i*, with

$$Z_i = \sqrt{\rho}X + \sqrt{1-\rho}\varepsilon_i,$$

where X and ε_i are two independent standard normal random variables, and ρ is the asset's correlation (or with the factor).

Under assumptions A6-A7, the conditional probability of default is equal to

$$p_i(x) = \Phi\left(\frac{B_i - \sqrt{\rho}x}{\sqrt{1-\rho}}\right),$$

where $\Phi(.)$ is the cdf of the standard normal distribution and the barrier B_i corresponds to the quantile associated to the unconditional probability of default, $B_i = \Phi^{-1}(p_i)$. Since $\Phi^{-1}(1-\alpha) = -\Phi^{-1}(\alpha)$, we get

$$VaR_L(\alpha) = \sum_{i=1}^{n} \text{EAD}_i \times \mathbb{E}(\text{LGD}_i) \times \Phi\left(\frac{\Phi^{-1}(p_i) + \sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right).$$

In order to determine the regulatory capital (RC), the BCBS considers the unexpected loss as the credit risk measure

$$\text{RC} = UL(\alpha) = VaR_L(\alpha) - \mathbb{E}(L).$$

Then, we get

$$\mathrm{RC} = \sum_{i=1}^{n} \mathrm{EAD}_i \times \mathbb{E}\left(\mathrm{LGD}_i\right) \times \left(\Phi\left(\frac{\Phi^{-1}\left(p_i\right) + \sqrt{\rho}\Phi^{-1}\left(\alpha\right)}{\sqrt{1-\rho}} \right) - p_i \right).$$

By considering a risk level $\alpha = 99.9\%$ and by denoting PD the unconditional probability of default, we get the IRB formula (without maturity adjustment).

2.8.2 Appendix B: Maturity adjustment and correlation functions

The maturity adjustment suggested by the BCBS depends on the type of exposure. For the corporate, sovereign, and bank exposures, it is defined as

$$\gamma\left(\mathrm{M}\right) = \frac{1 + \left(\mathrm{M} - 2.5\right) \times b\left(\mathrm{PD}\right)}{1 - 1.5 \times b\left(\mathrm{PD}\right)},$$

with the smoothed maturity adjustment equal to

$$b\left(\mathrm{PD}\right) = \left(0.11852 - 0.05478 \log\left(\mathrm{PD}\right)\right)^2.$$

For the retail exposures, there is no maturity adjustment, i.e. $\gamma\left(\mathrm{M}\right) = 1$. The correlation function $\rho\left(\mathrm{PD}\right)$ describes the dependence of the asset value of a borrower on the general state of the economy. Different asset classes show different degrees of dependency on the overall economy, so it's necessary to adapt the correlation coefficient to these classes. The correlation function $\rho\left(\mathrm{PD}\right)$ for corporate, sovereign, and bank exposures is defined as

$$\rho\left(\mathrm{PD}\right) = 0.12 \times \left(\frac{1 - e^{-50\ \mathrm{PD}}}{1 - e^{-50}}\right) + 0.24 \times \left(1 - \left(\frac{1 - e^{-50\ \mathrm{PD}}}{1 - e^{-50}}\right)\right).$$

For small and medium-sized enterprises (SME), a firm-size adjustment is introduced that depends on the sales. In the sequel, we neglect this adjustment for simplicity. For retail exposures, the correlation function $\rho\left(\mathrm{PD}\right)$ depends on the exposures. For the residential mortgage exposures, the BCBS recommends to fix the correlation at 0.15, for revolving retail exposures at 0.04 and for other retail exposures, to use the following formula

$$\rho\left(\mathrm{PD}\right) = 0.03 \times \left(\frac{1 - e^{-35\ \mathrm{PD}}}{1 - e^{-35}}\right) + 0.16 \times \left(1 - \left(\frac{1 - e^{-35\ \mathrm{PD}}}{1 - e^{-35}}\right)\right).$$

2.8.3 Appendix C: Proof of Proposition 1

Proof. Under assumptions A1-A2, the capital charge expected loss can be expressed as

$$\begin{aligned}\mathcal{L}_{CC,m} &= \mathbb{E}\left(g\left(\eta_{i,m}\right)\right) \\ &= \mathbb{E}\left(g\left(\text{EAD}_i \times \delta\left(\text{PD}\right) \times \gamma\left(\text{M}\right) \times \varepsilon_{i,m}\right)\right) \\ &= g\left(\delta\left(\text{PD}\right)\right) \times g\left(\gamma\left(\text{M}\right)\right) \times \mathbb{E}\left(g\left(\text{EAD}_i \times \varepsilon_{i,m}\right)\right),\end{aligned}$$

since $\delta\left(\text{PD}\right)$ and $\gamma\left(\text{M}\right)$ are positive constant terms. Rewrite $\mathcal{L}_{CC,m}$ as

$$\mathcal{L}_{CC,m} = \Delta \times cov\left(g\left(\text{EAD}_i\right), g\left(\varepsilon_{i,m}\right)\right) + \Delta \times \mathbb{E}\left(g\left(\text{EAD}_i\right)\right) \times \mathcal{L}_m,$$

with $\Delta = g\left(\delta\left(\text{PD}\right)\right) \times g\left(\gamma\left(\text{M}\right)\right)$ and $\mathcal{L}_m = \mathbb{E}\left(g\left(\varepsilon_{i,m}\right)\right)$. Consider two LGD models m and $m+1$. The rankings of the two models are consistent as soon as $\mathcal{L}_m < \mathcal{L}_{m+1}$ and $\mathcal{L}_{CC,m} < \mathcal{L}_{CC,m+1}$. Since $\Delta > 0$, these conditions can be expressed as

$$cov\left(g\left(\text{EAD}_i\right), g\left(\varepsilon_{i,m}\right)\right) + \mathbb{E}\left(g\left(\text{EAD}_i\right)\right) \times \mathcal{L}_m < cov\left(g\left(\text{EAD}_i\right), g\left(\varepsilon_{i,m+1}\right)\right) + \mathbb{E}\left(g\left(\text{EAD}_i\right)\right) \times \mathcal{L}_{m+1}.$$

Or equivalently as

$$cov\left(g\left(\text{EAD}_i\right), g\left(\varepsilon_{i,m}\right)\right) - cov\left(g\left(\text{EAD}_i\right), g\left(\varepsilon_{i,m+1}\right)\right) < \mathbb{E}\left(g\left(\text{EAD}_i\right)\right)\left(\mathcal{L}_{m+1} - \mathcal{L}_m\right),$$

with $\mathcal{L}_{m+1} - \mathcal{L}_m < 0$ and $\mathbb{E}\left(g\left(\text{EAD}_i\right)\right) > 0$. ■

2.8.4 Appendix D: Proof of Corollary 1

Proof. If the variables EAD_i and $\varepsilon_{i,m}$ are independent, the variables $g(\text{EAD}_i)$ and $g(\varepsilon_{i,m})$ are also independent. Then, the capital charge expected loss becomes

$$\mathcal{L}_{CC,m} = \mathbb{E}\left(g\left(\eta_{i,m}\right)\right) = g\left(\delta\left(\text{PD}\right)\right) \times g\left(\gamma\left(\text{M}\right)\right) \times \mathbb{E}\left(g\left(\text{EAD}_i\right)\right) \times \mathbb{E}\left(g\left(\varepsilon_{i,m}\right)\right).$$

Consider two LGD models m and $m+1$, $\forall m = 1, ..., \mathcal{M}-1$, for which $\mathcal{L}_m < \mathcal{L}_{m+1}$, then we have

$$\Delta_i \times \mathbb{E}\left(g\left(\varepsilon_{i,m}\right)\right) < \Delta_i \times \mathbb{E}\left(g\left(\varepsilon_{i,m+1}\right)\right),$$

with $\Delta_i = g\left(\delta\left(\text{PD}\right)\right) \times g\left(\gamma\left(\text{M}\right)\right) \times \mathbb{E}\left(g\left(\text{EAD}_i\right)\right) > 0$. The ranking of LGD models are necessarily consistent, i.e. $\mathcal{L}_{CC,m} < \mathcal{L}_{CC,m+1}$. ■

2.8.5 Appendix E: Dataset description

Table 2.7: List of the variables

Variables type	Variables name	Description
Contract	Original maturity	Original maturity of the contract (in months)
	Time to default	Number of months before default
	Relative duration	Time to default divided by maturity
	Interest rate	Interest (or renting) rate
	Exposition type	Credit or leasing
	Customer type	Individual, professional (natural or legal)
	Brand of the car	Brand name of the car
	State of the car	New or second-hand
Macroeconomic	GDP Growth rate	Brazil, quarterly
	Unemployment rate	Brazil, monthly
	Interbank interest rate	Brazil, monthly
Basel parameters	EAD	Exposure at default
	PD	Basel default probability estimated by the bank
	LGD	Loss Given Default

Table 2.8: Descriptive statistics of the variables

Credit characteristics			
	q25	Median	q75
Original maturity (month)	36	48	60
Time to default (month)	11	19	29
Relative duration	0.23	0.43	0.70
Interest rate	17.04	19.94	23.19
Exposure at default	10,631	19,035	28,112
	Percentage		
Exposition type			
Credit		71.33	
Leasing		28.67	
Customer type			
Individuals		66.96	
Professionals		33.04	
Brand of the car			
Brand A		76.62	
Brand B		17.86	
Other		5.52	
State of the car			
New hand		89.09	
Second hand		10.91	
Macroeconomic variables			
	q25	Median	q75
GDP growth rate	-0.04	0.24	0.99
Unemployment rate	5.09	5.45	5.78
Interbank interest rate	8.00	10.50	11.25

2.8.6 Appendix F: Competing LGD Models

For our comparison, we consider six competing LGD models which are commonly used in academic and practitioner literature (see for instance Bastos, 2010; Qi and Zhao, 2011; Loterman et al., 2012, etc.), namely (1) the fractional response regression model, (2) the regression tree, (3) the random forest, (4) the gradient boosting, (5) the artificial neural network, and (6) the least squares support vector regression. In the sequel, we briefly present these competing models and mention the main references for further details.

2.8.6.1 Fractional response regression

The fractional response regression (FRR) model, initially proposed by Papke and Wooldridge (1996), allows to estimate the conditional mean of a continuous variable defined over $[0, 1]$. The FRR specification is defined as

$$\mathbb{E}\left(\mathrm{LGD}_i|\, X_i\right) = G\left(X_i'\beta\right),$$

where X_i is a k-vector of explanatory variables for the i^{th} loan, β a k-vector of parameters and $G\left(.\right)$ a link function, with $G : \mathbb{R} \to [0, 1]$. A natural choice for the link function is the logistic function with

$$G\left(X_i'\beta\right) = \frac{1}{1 + \exp\left(-X_i'\beta\right)}.$$

The model parameters are estimated by quasi-maximum likelihood (QML), where the quasi likelihood is defined as a modified Bernouilli likelihood. If we denote by $\widehat{\beta}$ the QML estimator of β, the LGD estimator is then given by $\widehat{\mathrm{LGD}}_i = G(X_i'\widehat{\beta})$.

2.8.6.2 Regression tree

The regression tree (TREE), initially introduced by Breiman et al. (1984), is a machine-learning forecasting method. For a continuous variable, the tree is obtained by recursively partitioning the covariates space according to a prediction error (defined as the squared difference between the observed and predicted values) and then, by fitting a simple mean prediction within each partition.

The sketch of a regression tree algorithm is the following. The algorithm starts with a root node gathering all observations. For each covariate X, find the set R that minimizes the sum of the node impurities in the two child nodes and choose the split that gives the minimum overall X and R. The splitting procedure continues until no significant further reduction of the sum of squared deviations is possible. At the end of the procedure, we get a partition into K regions $R_1, \ldots, R_K$, also called terminal nodes or leaves. For each terminal node k, the LGD forecast is then given by the average LGD, denoted $\overline{\mathrm{LGD}}_k$,

estimated from all the contracts that belong to the region R_k, with

$$\widehat{\text{LGD}}_i = \sum_{k=1}^{K} \overline{\text{LGD}}_k \times \mathbb{I}_{(X_i \in R_k)}.$$

There exist many algorithms for regression trees. Here, we consider the CART algorithm (Breiman et al., 1984).

2.8.6.3 Random forest

Random forest (RF), introduced by Breiman (2001), is a bootstrap aggregation method of regression trees, trained on different parts of the same training set, with the goal of reducing overfitting (or, equivalently estimator variance). Random forest generally induces a small increase in the bias compared to regression trees and a loss of interpretability, but generally greatly boosts the performance of the model. In addition to constructing each tree using a different bootstrap sample of the data as in bagging approaches, random forests change how the regression trees are constructed. Indeed, each node is split using the best among a subset of covariates randomly chosen at that node. Assume that B regression trees are combined and denote by $\widehat{\text{LGD}}_{i,b}$ the prediction of the b^{th} tree, then the random forest prediction is defined as

$$\widehat{\text{LGD}}_i = \frac{1}{B} \sum_{b=1}^{B} \widehat{\text{LGD}}_{i,b}.$$

2.8.6.4 Gradient boosting

Gradient boosting (GB) is an iterative aggregation procedure that consecutively fits new models (typically regression trees) to provide a more accurate estimate of the dependent variable (Friedman, 2001). The general feature of this algorithm consists in constructing for each iteration, a new base-learner which is maximally correlated with the negative gradient of a loss function, evaluated at the previous iteration over the whole sample. In general, the choice of the loss function is up to the researcher, but most of the studies consider the quadratic loss function.[15]

The gradient boosting algorithm can be summarized as follows. A first regression tree is built on the LGD training set. Denote by $f_0(X_i)$ the prediction for the i^{th} loan and define the corresponding residuals $r_{i0} = \text{LGD}_i - f_0(X_i)$ for $i =, .., n_t$. At the first iteration, a new regression tree is applied to the residuals r_{i0}. The LGD predictions are then updated using the iterative formula $f_1(X_i) = f_0(X_i) + r_{i1}$, where r_{i1} denotes the adjusted residuals issued from the regression tree. After M iterations, algorithm stops

[15] Another possibility would consist to use our capital charge loss function for the gradient boosting algorithm. But, this new estimation method for LGD is beyond the scope of this chapter.

and the final LGD predictions are given by

$$\widehat{\text{LGD}}_i = f_0(X_i) + \sum_{m=1}^{M} r_{im}.$$

2.8.6.5 Artificial neural network

Artificial neural networks (ANN) are a class of flexible non-linear models, initially introduced by Bishop (1995). It produces an output value by feeding inputs through a network whose subsequent nodes apply some chosen activation function to a weighted sum of incoming values. The type of ANN considered in this study is a multilayer perceptron similar to that used by Qi and Zhao (2011) for the LGD forecasts. It consists in a three-layer network based on an input layer, a hidden layer, and an output layer. The central idea of the algorithm is (1) to extract linear combinations of the covariates from the input layer to the hidden layer, and (2) to apply nonlinear function on these derived features in the output layer to predict the dependent variable.

Let f be the unknown underlying function, through which a vector of input variables X explains LGD, i.e. $LGD_i = f(X_i)$. Derived features Z_m are created using linear combinations of the covariates such as

$$Z_{im} = G(\alpha'_m X_i), \quad \forall m = 1, \ldots, M,$$

where M is the number of neurons in the hidden layer, α_m a vector of coefficients (including a constant term) from the input layer to the hidden layer and $G(.)$ the logistic function, which is the common activation function used in neural network. The LGD are then modeled as a function of these linear combinations such that

$$f(X_i) = \beta_0 + \sum_{m=1}^{M} \beta_m Z_{im} + \varepsilon_i,$$

where β_m are coefficients from the hidden layer to the output layer. The LGD forecasts are then given by $\widehat{\text{LGD}}_i = f(X_i)$.

2.8.6.6 Least squares support vector regression

Initially introduced by Vapnik (1995), support vector machine (SVM) is a machine learning tool for classification and regression. The method has become popular for its ability to deal with large data, its small number of meta-parameters, and its good results in practice. The key principle of SVM is to map the covariates into a higher dimensional feature space through a mapping function which increases the learning capabilities of the algorithm. In the context of LGD modelling, we consider support vector regression (SVR) since the LGD variable is continuous. There exist various types of SVR. Here, we

consider the least squares support vector regression (LS-SVR) introduced by Suykens and Vandewalle (1999) and Suykens et al. (2002). This method has a low computational cost as it is equivalent to solving a linear system of equations instead of solving a quadratic programming problem. Loterman et al. (2012), Yao et al. (2015, 2017) and Nazemi et al. (2017) illustrate the good predictive performance of LS-SVR for LGD modelling.

Suppose a set of training data $\{y_i, X_i\}_{i=1}^N$ in which y_i is the observed response value (i.e. LGD_i in our case) and X_i the associated k-vector of explanatory variables for the i^{th} individual. Let us assume that y_i can be approximated by the following function such that

$$f(X_i) = \beta' \varphi(X_i) + b,$$

where β is the k-vector of unknown parameters, b is the intercept, and $\varphi(X_i)$ denotes the kernel function that maps the data from the original data space to a higher dimensional space. The LS-SVR is hence based on the following quadratic minimization problem

$$\begin{cases} \min\limits_{\beta,b,u_i} & J(\beta, b, u_i) = \frac{1}{2}\beta'\beta + \frac{C}{2}\sum_{i=1}^N u_i^2 \\ \text{s.t.} & y_i = \beta'\varphi(X_i) + b + u_i, \quad i = 1, \ldots, N \end{cases}$$

The minimization of the first part of the J objective allows to control appropriately for overfitting, while the second serves to reduce the training error. Notice that the error terms u_i^2 are scaled by a positive regularization parameter C that controls the penalty imposed on prediction errors. In others words, this parameter determines the trade-off between the model complexity (flatness) and the degree to which large deviations are tolerated. This optimization problem can be solved in a simpler way using its Lagrange dual formulation counterpart. The dual formula requires the introduction of Lagrangian multipliers denoted $\{\alpha_i\}_{i=1}^N$ in the optimization problem leading to the maximization of

$$L(\beta, b, u_i, \alpha_i) = J(\beta, b, u_i) - \sum_{i=1}^N \alpha_i \left(\beta'\varphi(X_i) + b + u_i - y_i\right).$$

The KKT condition allows to reformulate the dual form in terms of linear equation systems such as

$$\begin{pmatrix} 0 & 1_N' \\ 1_N & \overline{K} \end{pmatrix} \begin{pmatrix} b \\ \alpha \end{pmatrix} = \begin{pmatrix} 0 \\ Y \end{pmatrix},$$

with $Y = (y_1, \ldots, y_N)'$, $1_N = (1, \ldots, 1)'$, $\alpha = (\alpha_1, \ldots, \alpha_N)'$, and $\overline{K} = K + \frac{1}{C} I_N$. By using Mercer's condition, the uv^{th} element of K is given by

$$K_{uv} = \varphi(X_u)'.\varphi(X_v) = K(X_u, X_v) \qquad u, v = 1, \ldots, N.$$

In order to implement the LS-SVR model, we consider a radial basis function kernel defined as

$$K\left(X_u, X_v\right) = exp\left(-\frac{||X_u - X_v||^2}{2\sigma^2}\right),$$

where σ is the scale parameter of the kernel. The closed form solution for α and b is then given by

$$\begin{cases} \alpha^* = \overline{K}^{-1}\left(Y - b^* 1_N\right) \\ b^* = \frac{1'_N \overline{K}^{-1} Y}{1'_N \overline{K}^{-1} 1_N} \end{cases}$$

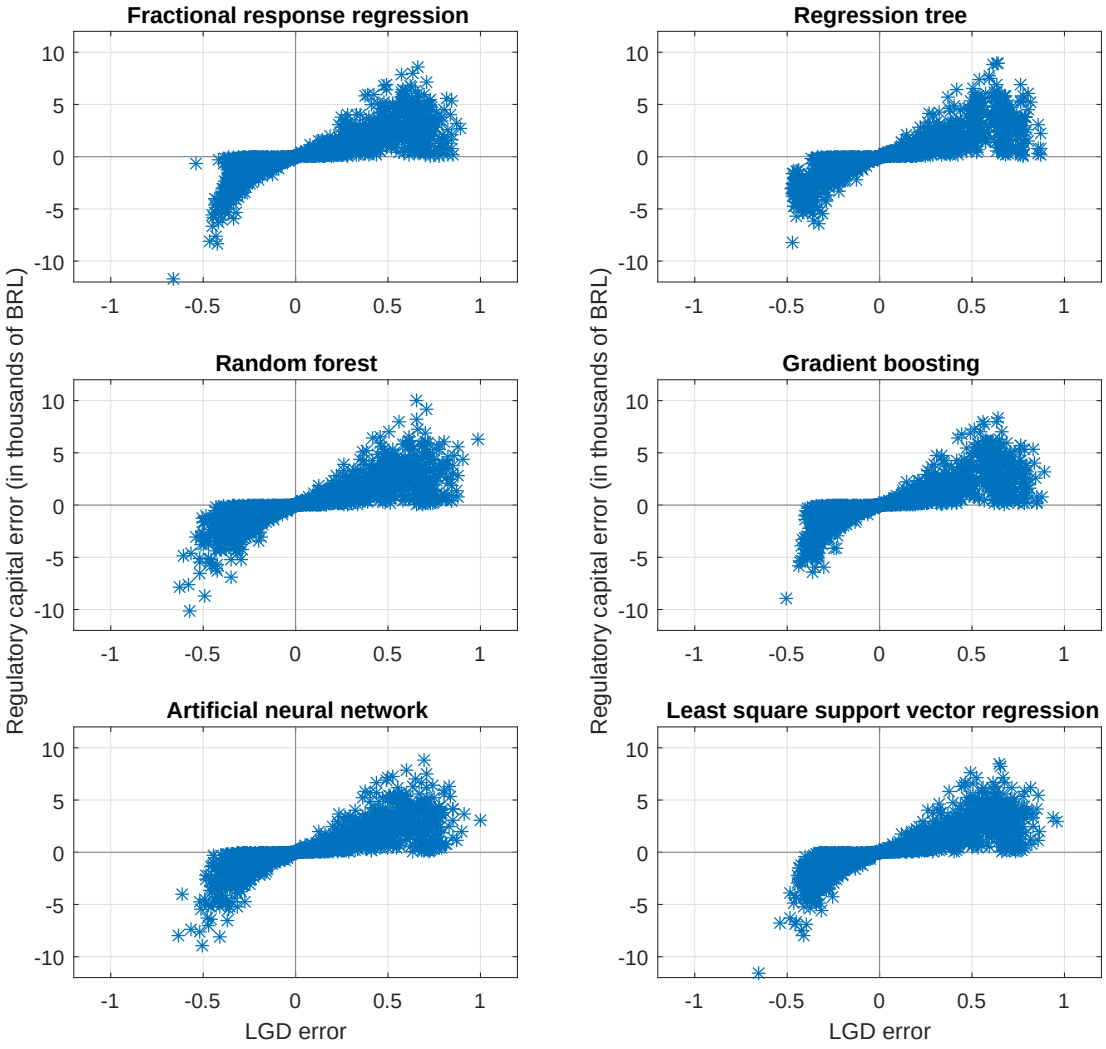
We finally get the LGD forecast for the j^{th} individual as

$$f\left(X_j\right) = \sum_{i=1}^{N} \alpha_i^* K\left(X_i, X_j\right) + b^*.$$

2.8.7 Appendix G: Scatter plot of the LGD and regulatory capital forecast errors

Figure 2.7 displays the scatter plots of the LGD versus regulatory capital (RC) forecast errors for the six competing models.

Figure 2.7: Scatter plot of the LGD and regulatory capital forecast errors (all models)



2.8.8 Appendix H: Out-of-sample criteria

Table 2.9: Out-of-sample criteria based on LGD and capital charge errors

		FRR	ANN	TREE	LS-SVR	RF	GB
				MSE			
Standard	LGD	0.1169	0.1160	0.1176	0.1148	0.1170	0.1160
	CC	3,815,235	3,799,711	3,732,987	3,700,248	3,737,030	3,720,010
Asymmetric	LGD	0.2010	0.2014	0.2018	0.1990	0.1982	0.2019
	CC	6,036,659	5,898,476	6,131,352	5,910,540	5,835,150	6,179,716
				MAE			
Standard	LGD	0.2906	0.2856	0.2908	0.2856	0.2856	0.2896
	CC	1,373.96	1,353.61	1,366.04	1,347.04	1,342.20	1,367.34
Asymmetric	LGD	0.3815	0.3817	0.3817	0.3787	0.3752	0.3843
	CC	1,815.28	1,800.22	1,819.83	1,792.60	1,758.73	1,836.43
				RMSE			
Standard	LGD	0.3419	0.3406	0.3430	0.3388	0.3420	0.3406
	CC	1,953.26	1,949.28	1,932.09	1,923.60	1,933.14	1,928.73
Asymmetric	LGD	0.4483	0.4488	0.4492	0.4461	0.4452	0.4493
	CC	2,456.96	2,428.68	2,476.16	2,431.16	2,415.61	2,485.90
				R^2			
Standard	LGD	0.0447	0.0521	0.0390	0.0622	0.0441	0.0522
	CC	0.3003	0.3032	0.3154	0.3214	0.3147	0.3178
Asymmetric	LGD	0.0209	0.0363	0.0200	0.0388	0.0398	0.0278
	CC	0.4566	0.4769	0.4504	0.4714	0.4761	0.4506

Table 2.10: Out-of-sample criteria (LGD models with macroeconomic variables)

		FRR	ANN	TREE	LS-SVR	RF	GB
				MSE			
Standard	LGD	0.1125	0.1108	0.1117	0.1094	0.1092	0.1101
	CC	3,589,815	3,530,643	3,470,465	3,445,874	3,441,599	3,476,828
Asymmetric	LGD	0.1994	0.1870	0.1904	0.1908	0.1843	0.1921
	CC	5,792,513	5,480,483	5,678,123	5,488,234	5,443,856	5,716,877
				MAE			
Standard	LGD	0.2805	0.2731	0.2768	0.2724	0.2705	0.2766
	CC	1,316.02	1,279.29	1,296.38	1,273.56	1,270.13	1,304.60
Asymmetric	LGD	0.3807	0.3613	0.3683	0.3678	0.3583	0.3721
	CC	1,784.87	1,691.39	1,741.11	1,714.11	1,686.07	1,758.48
				RMSE			
Standard	LGD	0.3355	0.3329	0.3343	0.3308	0.3305	0.3318
	CC	1,894.68	1,879.00	1,862.92	1,856.31	1,855.15	1,864.63
Asymmetric	LGD	0.4465	0.4324	0.4364	0.4368	0.4292	0.4382
	CC	2,406.76	2,341.04	2,382.88	2,342.70	2,333.21	2,391.00
				R^2			
Standard	LGD	0.0805	0.0947	0.0870	0.1061	0.1077	0.1003
	CC	0.3417	0.3525	0.3636	0.3681	0.3689	0.3624
Asymmetric	LGD	0.0571	0.1157	0.0986	0.0981	0.1260	0.0959
	CC	0.4925	0.5151	0.5013	0.5191	0.5206	0.4987

2.8.9 Appendix I: Paired t-test for comparisons of MSE and MAE

Table 2.11: Paired t-test for comparisons of MSE

Models	FRR	ANN	TREE	LS-SVR	RF	GB
Panel A. LGD Loss						
FRR	—					
ANN	1.0138	—				
TREE	−0.7712	−1.3892	—			
LS-SVR	3.1996**	2.1222*	2.9342**	—		
RF	−0.0588	−0.8185	0.4766	−2.0093*	—	
GB	1.9272	0.0052	2.0023*	−1.7521	0.7907	—
Panel B. CC Loss						
FRR	—					
ANN	0.2817	—				
TREE	1.1020	0.7267	—			
LS-SVR	3.2161**	2.5371*	0.4376	—		
RF	0.9583	0.8966	−0.0481	−0.5752	—	
GB	1.9318	1.1372	0.2803	−0.3983	0.2358	—

Note: Values are paired t statistics where a positive value means the accuracy statistic for the model on the vertical axis is better than that for the model on the horizontal axis, and vice versa. *: 5% Significance level. **: 1% Significance level.

Table 2.12: Paired t-test for comparisons of MAE

Models	FRR	ANN	TREE	LS-SVR	RF	GB
Panel A. LGD Loss						
FRR	—					
ANN	3.9300**	—				
TREE	−0.1094	−3.2109**	—			
LS-SVR	5.1559**	0.0518	3.7981**	—		
RF	2.7557**	0.0049	2.8167**	−0.0223	—	
GB	1.6076	−3.0733**	1.0514	−3.9863**	−2.2727*	—
Panel B. CC Loss						
FRR	—					
ANN	2.8610**	—				
TREE	0.8975	−1.1809	—			
LS-SVR	5.0206**	1.3452	2.1153*	—		
RF	2.9521**	1.2095	2.1702*	0.5432	—	
GB	1.4570	−1.8047	−0.1879	−3.5635**	−2.5179*	—

Note: Values are paired t statistics where a positive value means the accuracy statistic for the model on the vertical axis is better than that for the model on the horizontal axis, and vice versa. *: 5% Significance level. **: 1% Significance level.

2.8.10 Appendix J: Marginal effects in the FRR model

Table 2.13: Estimation results of the fractional response regression model

Independent variables	(1)	(2)
Interest rate	0.0282^{**}	0.0253^{**}
Original maturity	0.0168^{**}	0.0154^{**}
Time to default	-0.0158^{*}	-0.0133
Relative duration	1.3152^{**}	1.2437^{**}
Exposure at default	$1.7e^{-05**}$	$1.7e^{-05**}$
Customer type		
Individuals	0.0453	0.0378
Professionals	–	–
Brand of the car		
Brand A	-0.1143	-0.0960
Brand B	0.1084	0.0944
Other	–	–
Exposition type		
Credit	0.0125	0.0558
Leasing	–	–
State of the car		
New hand	-0.1771	-0.2302^{*}
Second hand	–	–
Macroeconomic Variables		
GDP growth rate	–	-0.0231
Unmeployment rate	–	0.7546^{**}
Interbank interest rate	–	-0.0814^{**}

Note: The first and second columns display the estimation results of the FRR model, with and without including macroeconomic variables, respectively. *: 5% Significance level. **: 1% Significance level.

Chapter 3

Backtesting Expected Shortfall via Multi-Quantile Regression[1]

In this chapter we propose a new approach to backtest Expected Shortfall (ES) exploiting the definition of ES as a function of Value-at-Risk (VaR). Our methodology examines jointly the validity of the VaR forecasts along the tail distribution of the risk model, and encompasses the Basel Committee recommendation of verifying quantiles at risk levels 97.5%, and 99%. We introduce four easy-to-use backtests in which we regress the ex-post losses on the VaR forecasts in a multi-quantile regression model, and test the resulting parameter estimates. Monte-Carlo simulations show that our tests are powerful to detect various model misspecifications. We apply our backtests on S&P500 returns over the period 2007-2012. Our tests clearly identify misleading ES forecasts in this period of financial turmoil. Empirical results also show that the detection abilities are higher when the evaluation procedure involves more than two quantiles, which should accordingly be taken into account in the current regulatory guidelines.

3.1 Introduction

In response to the market failures revealed by the global 2007-2008 financial crisis, the Basel Committee on Banking Supervision (BCBS) has adopted the Basel III accords to improve the banking sector's ability to absorb shocks arising from financial and economic stress (BCBS, 2010). Among the number of fundamental reforms that must be implemented until January 1st, 2019, the BCBS has substituted Value-at-Risk (VaR) by Expected Shortfall (ES) for the calculation of market risk capital requirements. Expected Shortfall, also referred to as Conditional VaR (CVaR) or Tail VaR (TVaR), measures the expected loss incurred on an asset portfolio given that the loss exceeds VaR. That is, if

[1]This chapter is based on Couperier and Leymarie (2019) currently R&R in *Journal of Business and Economic Statistics*.

L_t is the ex-post loss on a portfolio at time t, Ω_{t-1} is the information at time $t-1$, and $Q_{L_t}(.)$ is the quantile function of L_t, the τ-level ES and VaR are given by

$$ES_t(\tau) = \mathbb{E}\left[L_t \mid L_t \geq VaR_t(\tau); \Omega_{t-1}\right], \tag{3.1}$$

$$VaR_t(\tau) = Q_{L_t}(\tau; \Omega_{t-1}). \tag{3.2}$$

As an alternative tail risk measure, ES offers a number of appealing properties that overcomes the theoretical deficiencies of the more-familiar VaR. In particular, ES is *coherent* meaning that this risk measure satisfies the properties of monotonicity, sub-additivity, homogeneity, and translational invariance (see Artzner et al., 1999; Acerbi and Tasche, 2002). Furthermore, ES provides information about the expected size of the potential loss given that a loss bigger than VaR is experienced, while VaR only captures the likelihood of an incurred loss, and tells us nothing about tail sensitivity. In its revised standards for market risk, the BCBS emphasizes the important role of ES in place of VaR "*to ensure a more prudent capture of "tail risk" and capital adequacy during periods of significant financial market stress*" (BCBS, 2016, page 1).

Although ES is now considered as the new standard for risk management and regulatory requirements, there are still outstanding questions about the modeling of ES (see for instance Taylor, 2019; Patton et al., 2019), and the validation of the ES forecasts, or backtesting. Jorion (2006) defines backtesting as a formal statistical framework that consists in verifying if actual losses are in line with projected losses. Because ES is unobservable, its evaluation cannot be performed conventionally as a direct comparison of the observed value with its forecast, and thus generally relies on the elicitability property. A risk measure is said to be *elicitable* if there exists a loss function such that the solution of minimizing the expected loss is the risk measure itself. However, it has been established that, in contrast to VaR, ES does not meet the general property of elicitability (Gneiting, 2011), but satisfies narrower properties such as conditional elicitability (Emmer et al., 2015), or joint elicitability with VaR (Acerbi and Szekely, 2014; Fissler and Ziegel, 2016), making its evaluation trickier than VaR in practice. Several contributions are tied to these properties, and provide backtests by making explicit reference of the ES forecasts in the testing procedure (McNeil and Frey, 2000; Acerbi and Szekely, 2014; Nolde and Ziegel, 2017; Bayer and Dimitriadis, 2018).

To circumvent the lack of elicitability of ES, several alternative testing strategies have been proposed in the literature. Following the recent classification of Kratz et al. (2018), these backtests enter the category of *implicit* backtests, as they focus on the tail distribution characteristics of the model rather than directly on ES. They generally exploit the fact that ES can be expressed as a function of VaR, which itself is elicitable. Indeed, definition of a conditional probability and a change of variable yield a useful

representation of ES in terms of VaR

$$ES_t(\tau) = \frac{1}{1-\tau}\int_\tau^1 VaR_t(u)\,du. \tag{3.3}$$

Based on this analogy, Costanzino and Curran (2015) derive a coverage backtest for spectral risk measures such as ES in the spirit of the traditional VaR coverage backtests. Du and Escanciano (2017) define a cumulative violation process for ES as a generalization of the violation process for VaR. Costanzino and Curran (2018) provide a Trafic Light backtest for ES which extends the so-called Traffic Light backtest for VaR. More largely, several additional techniques have been proposed to assess the whole return distribution encompassing ES as a special case (Berkowitz, 2001; Kerkhof and Melenberg, 2004; Wong, 2008). For more details on risk forecast evaluation, see the survey of Argyropoulos and Panopoulou (2016).

In this chapter, we also propose to exploit the relationship that prevails between ES and VaR, but contrary to the existing literature, our procedure aims at focusing on a finite number of VaRs. Definition of a Riemann sum gives a handy approximation of ES.

$$ES_t(\tau) \approx \frac{1}{p}\sum_{j=1}^{p} VaR_t(u_j), \tag{3.4}$$

where the risk level u_j is defined by $u_j = \tau + (j-1)\frac{1-\tau}{p}$ for $j = 1, 2, \ldots, p$. This representation suggests that p quantiles with appropriate risk levels would be convenient to assess the performance of an ES model. In other words, an estimate/forecast of $ES_t(\tau)$ issued from a given model could be considered valid if the sequence of $VaR_t(u_j)$ estimates/forecasts issued from the same model is itself valid. This testing strategy is fully consistent with the general recommendation of financial supervisors, indicating that "*Backtesting requirements [for ES] are based on comparing each desk's 1-day static value-at-risk measure [...] at both the 97.5th percentile and the 99th percentile*" (BCBS, 2016, page 57).

The main contribution of this chapter is to propose an original backtesting methodology for ES based on the theory of multi-quantile regression. This approach has many advantages. First, our procedure is flexible since the user may choose the number and values of quantiles to be investigated, and can easily focus on various aspects of the tail distribution. Second, our methodology encompasses the regulatory standards which consist in verifying the validity of two given quantiles. Finally, our testing strategy enters the category of regression-based backtests, and complements the existing literature on regression-based risk forecast evaluation proposed by Engle and Manganelli (2004), Christoffersen (2011), Bayer and Dimitriadis (2018), among others. In addition, this

original approach represents an alternative to the multiple VaR exceptions backtests of Colletaz et al. (2013), and Kratz et al. (2018).

Our procedure extends the seminal idea of Gaglianone et al. (2011) to evaluate the validity of the VaR forecasts applying quantile regression. We develop a multivariate framework focusing on multi-quantile regression to jointly assess VaR at multiple levels in the tail distribution of the risk model. We show that the parameter estimates issued from the multi-quantile regression model satisfy specific properties under the hypothesis of correct ES forecasts. We propose four backtests which correspond to various linear restrictions on these parameters. Finally, we introduce a procedure to correct the imperfect predictions relying on our multi-quantile regression framework.

We provide several Monte Carlo experiments and an empirical application using the S&P500 series. Our backtests deliver good performances to detect misleading ES forecasts. We also find that the use of asymptotic critical values is prone to substantial size distortions, and address these deficiencies via the implementation of bootstrap critical values. The latter provide satisfactory size performances regardless of the sample size, and hence should be preferred when asymptotic theory does not apply conveniently. We also show that the BCBS recommendation of verifying quantiles at coverage levels 97.5% and 99% is not always sufficient to reject the validity of the ES forecasts issued from misspecified models. We hence recommend the use of additional risk levels to improve the soundness of the decision. Finally, we show numerically that our approximation of ES as a combination of several VaRs is close to its theoretical counterpart, which strongly supports its implementation in a risk management viewpoint.

The rest of this chapter is organized as follows. In Section 3.2, we introduce the multi-quantile regression framework. Section 3.3 describes the null hypotheses of our tests, the test statistics, their asymptotic properties, and the procedure to implement the bootstrap critical values. Section 3.4 examines the finite sample performance of the proposed backtests through a set of Monte Carlo experiments. In Section 3.5, we apply our backtesting methodology on the S&P500 index, and introduce the procedure to adjust the imperfect ES forecasts. Finally, we conclude this chapter in Section 3.6.

3.2 Multi-quantile regression framework

This section is devoted to the description of our proposed multi-quantile regression approach. In the first part, we discuss the usefulness of approximating ES via a finite sum of VaRs. In a second part, we describe the multi-quantile regression model that we employ in our testing strategy. Finally, the last part is devoted to the description of the estimation method, and the asymptotic theory.

3.2.1 ES as an approximation of VaRs

Our backtesting procedure exploits the relationship between VaR and ES. We suppose that ES can be appropriately approximated as a weighted sum of VaRs. This assertion stems from the representation of ES as the limit of a Riemann sum when the partition becomes infinitely fine.

Definition 1 (ES approximation)**.** *Let $\tau \in]0,1[$ denote the coverage level. The τ-level ES can be approximated by a finite Riemann sum involving p VaRs by*

$$ES_t(\tau) \approx \frac{1}{p} \sum_{j=1}^{p} VaR_t(u_j), \tag{3.5}$$

where risk levels u_j, $j = 1, 2, \ldots, p$, satisfy $u_j = \tau + (j-1)\frac{1-\tau}{p}$, and p denotes the number of subdivisions taken in the definite integral.

Our approximation of ES averages VaRs in the upper tail distribution of the risk model. The number of quantiles involved in the sum is given by p, and characterizes the approximation's accuracy. In particular, $p = 1$ involves a single VaR at coverage level τ, while increasing p to infinity leads the Riemann sum to converge to the theoretical ES. In practice, p may be chosen small as the interval of the definite integral is restricted to the extreme upper tail distribution, and therefore a few quantiles are generally enough to get good approximations. Finally, the risk levels u_j, $j = 1, 2, \ldots, p$, are determined so that the interval is equally partitioned between the two boundaries τ, and 1.

Our approximation is useful for several reasons. First, this simple formula is appealing in a regulatory and risk management viewpoint since the estimation of VaR is well-established compared to ES and pretty easier to compute. Secondly, and it is the purpose of this chapter, the above relationship greatly simplifies the assessment of ES in practice, by focusing on the validity of several VaRs, and is more intelligible in the context of banking regulation. This approach is indeed fully consistent with the BCBS guidelines on ES assessment indicating that "*Backtesting requirements [for ES] are based on comparing each desk's 1-day static value-at-risk measure [...] at both the 97.5th percentile and the 99th percentile*" (BCBS, 2016, page 11). Finally, our validation strategy offers a certain flexibility since the risk manager or the supervisor may select both the number of probability levels and their magnitude depending on the objective in mind (regulatory guidelines, ES statistical approximation, etc.).

3.2.2 Multi-quantile regression model

In the sequel, we consider an asset or a portfolio, and denote by L_t the corresponding loss observed at time t, for $t = 1, 2, \ldots, T$. In addition, we denote by Ω_{t-1} the information

set available at time $t-1$, with $(L_{t-1}, L_{t-2}, \ldots) \subseteq \Omega_{t-1}$. Formally, the Ω_{t-1} conditional VaR at level u_j of the L_t distribution is the quantity $VaR_t(u_j)$ such that

$$\Pr\left(L_t \geq VaR_t(u_j) \,|\Omega_{t-1}\right) = u_j. \tag{3.6}$$

A VaR model is said to be correctly specified (at coverage level u_j) as soon as Equation (3.6) holds for all t. In practice, VaR forecasts are assessed through the evaluation of this simple equality. Given the ES approximation introduced in Definition 1, this equality may arguably be adapted for the assessment of ES models. The chief insight is to evaluate Equation (3.6) for a number p of risk levels as set out in Definition 1. Accordingly, one should conclude to the appropriateness of a given ES model as soon as the sequence $VaR_t(u_j)$, $t = 1, 2, \ldots, T$, issued from the ES model satisfies Equation (3.6) jointly for $j = 1, 2, \ldots, p$.

We refer to the original idea of Gaglianone et al. (2011) who derive a backtest of VaR at a single coverage level, introducing VaR as a regressor of a quantile regression model. We generalize their approach for the assessment of multiple VaRs. To do so, we regress the ex-post losses $\{L_t, t = 1, 2, \ldots, T\}$ on the p VaR forecasts $\{VaR_t(u_j), t = 1, 2, \ldots, T\}_{j=1,2,\ldots,p}$ in a multi-quantile regression model.

$$L_t = \beta_0(u_j) + \beta_1(u_j) VaR_t(u_j) + \epsilon_{j,t} \quad \forall\ j = 1, 2, \ldots, p, \tag{3.7}$$

where $\beta_0(u_j)$, and $\beta_1(u_j)$, respectively, denote the intercept and the slope parameters at level u_j, and where $\epsilon_{j,t}$ is the error term at risk level u_j and time t, such that the u_j-th conditional quantile of $\epsilon_{j,t}$ satisfies $Q_{\epsilon_{j,t}}(u_j; \Omega_{t-1}) = 0$.

This specification could be interpreted as a multi-quantile regression version of Koenker and Xiao (2002). More specifically, the representation is tightly related to the multi-quantile CaViAR model (MQ-CaViAR) proposed by White et al. (2008, 2015) which allows a joint modeling of multiple conditional VaRs. Given the multi-quantile regression model of Equation (3.7), the u_j-th conditional quantile of L_t is defined as

$$Q_{L_t}(u_j; \Omega_{t-1}) = \beta_0(u_j) + \beta_1(u_j) VaR_t(u_j) \quad \forall\ j = 1, 2, \ldots, p. \tag{3.8}$$

This equation is central for our backtesting methodology as it establishes a direct link between the VaR forecasts (issued from the external ES model), with the true unknown conditional quantile (issued from the ex-post observed losses). Our procedure consists in verifying if there exists a perfect match between $VaR_t(u_j)$ and $Q_{L_t}(u_j; \Omega_{t-1})$. Consistently with Gaglianone et al. (2011), we rely on the regression parameters, and test if the intercept parameter $\beta_0(u_j)$, and the slope parameter $\beta_1(u_j)$, are respectively equal to zero, and one, for $j = 1, 2, \ldots, p$. For these parameter values, and given Definition 1,

the risk model is accepted as a valid proxy of the true unknown data generating process to deliver the ES forecasts.

3.2.3 Parameter estimation and asymptotic properties

Our backtesting procedure requires to consistently estimate the parameters $\beta_0(u_j)$, and $\beta_1(u_j)$, for $j = 1, 2, \ldots, p$. Under the hypothesis that a sequence of VaR is valid, coefficients satisfy $\beta_0(u_j) = 0$, and $\beta_1(u_j) = 1$, for $j = 1, 2, \ldots, p$. In what follows, we denote by $\beta(u_j) = (\beta_0(u_j),\ \beta_1(u_j))'$ the vector of parameters for the u_j-th quantile index, and we write $\beta = \left(\beta(u_1)',\ \beta(u_2)',\ \ldots, \beta(u_p)'\right)'$ the stacked vector of $2p$ coefficients. In addition, we assume that the sequence $\{u_j, j = 1, 2, \ldots, p\}$ is ordered in the sense that $u_1 < u_2 < \ldots < u_p < 1$.

In order to estimate β, we consider the multi-quantile regression approach recently proposed by White et al. (2008, 2015). A consistent QMLE estimator is given by

$$\widehat{\beta} = arg \min_{\beta \in \mathbb{R}^{2p}} T^{-1} \sum_{t=1}^{T} \left(\sum_{j=1}^{p} \rho_{u_j} \left(L_t - \beta_0(u_j) - \beta_1(u_j) VaR_t(u_j) \right) \right), \tag{3.9}$$

where $\rho_{u_j}(x) = x\psi_{u_j}(x)$ is the standard "check function", and $\psi_{u_j}(x) = u_j - \mathbb{1}(x \leq 0)$ is the usual quantile step function. Under suitable regularity conditions, White et al. (2008, 2015) show that this estimator is asymptotically normally distributed:

$$\sqrt{T}\left(\widehat{\beta} - \beta\right) \xrightarrow{d} \mathcal{N}(0, \Sigma), \tag{3.10}$$

where Σ denotes the asymptotic covariance matrix which takes the form of a Huber (1967) sandwich. Its expression is given by $\Sigma = A^{-1}VA^{-1}$, with:

$$V = \mathbb{E}\left[\eta_t \eta_t'\right], \tag{3.11}$$

$$\eta_t = \sum_{j=1}^{p} \nabla Q_{L_t}(u_j; \Omega_{t-1})\, \psi_{u_j}(\epsilon_{j,t}), \tag{3.12}$$

$$A = \sum_{j=1}^{p} \mathbb{E}\left[f_{j,t}(0)\, \nabla Q_{L_t}(u_j; \Omega_{t-1})\, \nabla' Q_{L_t}(u_j; \Omega_{t-1}) \right], \tag{3.13}$$

where $\nabla Q_{L_t}(u_j; \Omega_{t-1})$ denotes the $2p$ gradient vector differentiated with respect to β, and $\epsilon_{j,t} = L_t - Q_{L_t}(u_j; \Omega_{t-1})$, and where $f_{j,t}(0)$ denotes the pdf of $\epsilon_{j,t}$ evaluated at zero. In Appendix 3.7.1, we provide a consistent estimator $\widehat{\Sigma}$ of the true unknown quantity Σ that will be used for the computation of our test statistics.

3.3 Backtesting ES

In this section, we present our backtests for ES. Our procedures assess whether the parameters $\beta_0(u_j)$ and $\beta_1(u_j)$ coincide with their expected values for the risk levels u_j, $j = 1, 2, \ldots, p$. To this end, we propose four backtests that analyze various settings on the regression coefficients. In the sequel, we introduce the null hypotheses, the test statistics, and establish their asymptotic properties. Finally, we discuss the use of finite sample critical values and provide a bootstrap algorithm when the asymptotic theory does not apply conveniently.

3.3.1 The backtests

Formally, our goal is to test $\beta_0(u_j) = 0$, and $\beta_1(u_j) = 1$, for $j = 1, 2, \ldots, p$. We propose to test various implications of these coefficient restrictions by taking into consideration four distinct null hypotheses based on a reduced number of constraints. Many backtests test implications of a more general hypothesis. In this context, Du and Escanciano (2017) assess two implications for the martingale difference sequence of their cumulative violation process. McNeil and Frey (2000) and Nolde and Ziegel (2017) propose to test the zero mean hypothesis of their residuals which more largely behave as white noise.

Definition 2 (Null hypotheses)**.** *Denote by J_1, J_2, I, and S, the four backtests. The corresponding null hypotheses H_{0,J_1}, H_{0,J_2}, $H_{0,I}$, $H_{0,S}$, are defined as follows:*

$$H_{0,J_1} : \sum_{j=1}^{p} \left(\beta_0(u_j) + \beta_1(u_j)\right) = p, \tag{3.14}$$

$$H_{0,J_2} : \sum_{j=1}^{p} \beta_0(u_j) = 0, \text{ and, } \sum_{j=1}^{p} \beta_1(u_j) = p, \tag{3.15}$$

$$H_{0,I} : \sum_{j=1}^{p} \beta_0(u_j) = 0, \tag{3.16}$$

$$H_{0,S} : \sum_{j=1}^{p} \beta_1(u_j) = p, \tag{3.17}$$

where notations J_1 and J_2 indicate the "joint" backtests, and where I and S refer to the "intercept" backtest and to the "slope" backtest, respectively.

Definition 2 gives the null hypotheses H_{0,J_1}, H_{0,J_2}, $H_{0,I}$, $H_{0,S}$. They are devised to assess various implications that the regression coefficients should satisfy when the ES forecasts are valid. Notice that the coefficients are summed across risk levels u_j, $j = 1, 2, \ldots, p$. This coefficients' aggregation substantially reduces the number of constraints

to be tested. The structure of H_{0,J_2} is hence characterized by two constraints, and those of H_{0,J_1}, $H_{0,I}$, $H_{0,S}$ involve a single constraint.

Our null hypotheses analyze various settings on the regression coefficients. The null of the joint backtests, H_{0,J_1} and H_{0,J_2}, look at the expected value of both the intercept and slope parameters $\beta_0(u_j)$ and $\beta_1(u_j)$ for $j = 1, 2, \ldots, p$. H_{0,J_1} sums the two types of coefficient together, while H_{0,J_2} sums the coefficients separately depending on whether they are slope parameters or intercept parameters. Finally, the null hypotheses of the intercept backtest and the slope backtest, $H_{0,I}$ and $H_{0,S}$, focus solely on one of the two parameter components. $H_{0,I}$ is built to examine the intercept parameters $\beta_0(u_j)$, $j = 1, 2, \ldots, p$, and $H_{0,S}$ is devoted to the analysis of the slope parameters $\beta_1(u_j)$, $j = 1, 2, \ldots, p$. These additional null hypotheses complement the joint backtests to identify the nature of the misspecification. When $H_{0,I}$ is rejected, this indicates that the forecasting errors are constant across time. In contrast, the rejection of $H_{0,S}$ suggests that the errors are time-varying since they change with respect to the VaR predictions.

Definition 3 (Wald-test statistics)**.** *Let us denote by $W \in \{J_1, J_2, I, S\}$ the generic notation for the test statistic, and consider the classical formulation of a Wald-type test such as $H_{0,W}$: $R_W\beta = q_W$. The general expression of the test statistics is given by*

$$W = T\left(R_W\widehat{\beta} - q_W\right)'\left(R_W\widehat{\Sigma}R_W'\right)^{-1}\left(R_W\widehat{\beta} - q_W\right), \tag{3.18}$$

where T is the out-of-sample size, and $\widehat{\Sigma}$ denotes a consistent estimator of the asymptotic covariance matrix.

To assess our null hypotheses we consider Wald-type inference. Definition 3 provides the general expression of the test statistics. In accordance with our notations, substituting W by J_1, J_2, I, and S, yields the four test statistics. For ease of presentation, the null hypotheses are now presented in a classical formulation, such that $H_{0,W} : R_W\beta = q_W$. Given the null hypotheses of Definition 2, the quantities R_W and q_W are as follows: $R_{J_1} = \iota_p \otimes \begin{pmatrix} 1 & 1 \end{pmatrix}$, $q_{J_1} = p$, $R_{J_2} = \iota_p \otimes I_2$, $q_{J_2} = \begin{pmatrix} 0 & p \end{pmatrix}'$, $R_I = \iota_p \otimes \begin{pmatrix} 1 & 0 \end{pmatrix}$, $q_I = 0$, $R_S = \iota_p \otimes \begin{pmatrix} 0 & 1 \end{pmatrix}$, $q_S = p$, where ι_p is a p-row unit vector, and I_2 denotes the identity matrix of size 2.

Proposition 1 (Chi-squared distribution)**.** *Suppose the covariance matrix Σ is non singular. Under the normality condition of Equation (3.10), and the null hypotheses of Definition 2, the test statistics J_1, I, and S, converge to a chi-squared distribution with 1 degree of freedom, and the test statistic J_2 converges to a chi-squared distribution with 2 degrees of freedom.*

Proposition 1 gives the asymptotic distribution of the Wald statistics J_1, J_2, I, S under their respective null hypotheses H_{0,J_1}, H_{0,J_2}, $H_{0,I}$, $H_{0,S}$. As a result of coefficients' aggregation, the asymptotic distributions are based on a small and fixed number of degrees of freedom no matter how p is chosen. Thus, the four backtests have unchanged critical values whatever the number of quantiles considered in the ES approximation. Finally, we provide in Appendix 3.7.2 the proof for consistency of the tests under fixed untrue hypothesis.

3.3.2 Finite sample inference

Our four backtests are asymptotically chi-squared distributed, and it is thus possible to employ them if the asymptotic conditions are fulfill for realistic sample sizes. However, in the specific context of ES assessment, the focus is on the extreme tail distribution, i.e. for risk levels above the regulatory coverage level, namely $\tau = 0.975$. In practice, this may induce scarce information, and affect the inference when the sample size is not sufficiently large. To overcome these typical deficiencies, we implement a bootstrap procedure to adjust the critical values of our test statistics in small samples.

The backtests of tail risk measures such as VaR and ES are typically affected by small sample size distortions. For instance, Gaglianone et al. (2011) report size distortions for moderate sample sizes. In the same vein, the regression procedure to assess ES proposed by Bayer and Dimitriadis (2018) induces size distortions even for large sample sizes, e.g. $T = 1000$. The authors also show that the conditional calibration test of Nolde and Ziegel (2017), and the exceedance residual test of McNeil and Frey (2000) display poor results for realistic samples. They propose a bootstrap algorithm to correct these biases. Finally, it should be noticed that Hurlin et al. (2017) introduce a bootstrap procedure dedicated to the inference of the risk measures themselves that is valid in finite samples.

In the following, we propose a pairs bootstrap algorithm (Freedman, 1981) in order to correct the finite sample size distortions of our backtests. This is a fully non-parametric procedure that can be applied to a very wide range of models, including quantile regression model (Koenker et al., 2018). This approach consists in resampling the data, keeping the dependent and independent variables together in pairs. The procedure is valid for any sample sizes T, and large levels u_j, $j = 1, 2, \ldots, p$, and ideally applies in our case when the constraints of the null hypothesis are linear in the parameters. The algorithm is as follows:

1. Estimate β and Σ on the original data $\{L_t, VaR_t(u_j)\}_{j=1,2,\ldots,p}$, $t = 1, 2, \ldots, T$, to obtain $\widehat{\beta}$ and $\widehat{\Sigma}$, and compute the unconstrained test statistic W given by

$$W = T\left(R_W\widehat{\beta} - q_W\right)'\left(R_W\widehat{\Sigma}R_W'\right)^{-1}\left(R_W\widehat{\beta} - q_W\right).$$

2. Build a bootstrap sample by drawing with replacement T pairs of observations from the original data $\{L_t, VaR_t(u_j)\}_{j=1,2,\ldots,p}$, $t = 1, 2, \ldots, T$.

3. Estimate the model on the bootstrap sample, to obtain $\widehat{\beta}^b$ and $\widehat{\Sigma}^b$, and compute the bootstrapped test statistic W^b under the null hypothesis as follows:

$$W^b = T\left(R_W\widehat{\beta}^b - R_W\widehat{\beta}\right)'\left(R_W\widehat{\Sigma}^b R_W'\right)^{-1}\left(R_W\widehat{\beta}^b - R_W\widehat{\beta}\right).$$

4. Repeat $B-1$ times steps 2 and 3, to obtain the bootstrap statistics W^b, $b = 1, 2, \ldots, B$.

Two remarks should be made about the algorithm. First, when we use the pairs bootstrap we cannot impose the null hypothesis on the bootstrap data generating process since imposing restrictions on β is unfeasible. To overcome this issue, we calculate the bootstrap statistics by considering the difference $R_W\beta - R_W\widehat{\beta}$ rather than $R_W\beta - q$. Since the estimate of β from the bootstrap samples should, on average, be equal to $\widehat{\beta}$, at least asymptotically, the null hypothesis tested by W^b becomes "true" for the pairs bootstrap data generating process. Second, the critical value c_α is obtained as the α-quantile of the bootstrap statistics W^b, $b = 1, 2, \ldots, B$. The decision rule is as follows. If the original test statistic W is greater than the α-level bootstrapped critical value c_α, we conclude to the rejection of the null hypothesis. In addition, we compute the p-value of the test as $P = B^{-1}\sum_{b=1}^{B}\mathbb{1}\left(W^b > W\right)$.

3.4 Simulation Study

In this section, we provide Monte Carlo simulations to illustrate the finite sample properties (empirical size and power) of our four backtests. The simulation study is performed on 5000 replications, and we consider two sample sizes $T = 500, 2500$. The results are reported using both the asymptotic critical values (based on a χ^2 distribution), and the bootstrap critical values. We calculate the latter with $B = 1000$ bootstrap samples. Finally, the backtests are computed with $\tau = 0.975$, which is the coverage level applied in the context of the current banking regulation.

Beyond the traditional size and power analysis, a second important objective of this section is to characterize the influence of the number p of quantiles used to assess the ES forecasts. We aim at examining whether an ES backtest based on a large number of quantiles may provide better performances than a backtest based on a small number of quantiles, as it is recommended by the current BCBS guidelines. For that, we consider different choices for the number of risk levels, namely $p = 1, 2, 4, 6$. The p risk levels $u_1, u_2, \ldots, u_p$ are computed in accordance with Definition 1. Notice that $p = 1$ coincides with the VaR backtest at level τ of Gaglianone et al. (2011). With $p = 2$ risk levels, our

backtests are in accordance with the number of quantiles of the regulatory guidances. Finally, the case $p = 4$ corresponds to the framework considered by Emmer et al. (2015).

The correct data generating process is given by the popular AR(1)-GARCH(1,1) specification with Student innovations. Accordingly, we define the ex-post portfolio loss L_t, $t = 1, 2, \dots T$ as

$$\begin{aligned} L_t &= \delta_0 + \delta_1 L_{t-1} + \epsilon_t, \\ \epsilon_t &= \sigma_t \eta_t, \ \eta_t \sim t_v, \\ \sigma_t^2 &= \gamma_0 + \gamma_1 \epsilon_{t-1}^2 + \gamma_2 \sigma_{t-1}^2, \end{aligned} \tag{3.19}$$

where t_v denotes the Student's t distribution with v degrees of freedom. Given the model in Equation (3.19), the true ES and VaR at coverage level τ are given by

$$ES_t(\tau) = \delta_0 + \delta_1 L_{t-1} + \sigma_t m(\tau), \tag{3.20}$$

$$VaR_t(\tau) = \delta_0 + \delta_1 L_{t-1} + \sigma_t F_v^{-1}(\tau), \tag{3.21}$$

with $m(\tau) = \mathbb{E}[\eta_t | \eta_t \geq F_v^{-1}(\tau)]$, and where $F_v^{-1}(\tau)$ denotes the τ-quantile of the Student distribution with v degrees of freedom. Model parameters (δ_0, δ_1, γ_0, γ_1, γ_2, v) are calibrated using the S&P500 series over the period 2013-2017, which leads us to consider the following numeric values in the simulation study (-0.085, -0.093, 0.034, 0.214, 0.748, 5). Finally to investigate the power, we consider several misspecified alternatives for L_t:

A_1 : AR(1)-GARCH(1,1) model with underestimated conditional variances: L_t is as Equation (3.19), with $\sigma_t^2 = \left(\gamma_0 + \gamma_1 \epsilon_{t-1}^2 + \gamma_2 \sigma_{t-1}^2\right) \times (1 - \kappa)$, where $\kappa = 0.25, 0.50, 0.75,$ respectively.

A_2 : GARCH in mean model: $L_t = \kappa \times \sigma_t^2 + \epsilon_t, \ \epsilon_t = \sigma_t \eta_t, \ \sigma_t^2 = \gamma_0 + \gamma_1 \epsilon_{t-1}^2 + \gamma_2 \sigma_{t-1}^2, \ \eta_t \sim t_v,$ where $\kappa = +2.5, -2.5$, respectively.

A_3 : AR(1)-GARCH(1,1) model with mixed normal innovations: L_t satifies Equation (3.19), with $\eta_t \sim \left(0.5X^+ + 0.5X^-\right)/\sqrt{10}$, where $X^+ \sim \mathcal{N}(3,1)$ and $X^- \sim \mathcal{N}(-3,1)$.

A_4 : 12-month historical simulation model : VaR and ES are given by their empirical counterparts from the 250 previous trading days such that $VaR_t(\tau) = \text{percentile}\left(\{L_{t-i}\}_{i=1}^{250}, 100\tau\right)$, and $ES_t(\tau) = \dfrac{1}{\sum_{i=1}^{250} \mathbb{1}_{(L_{t-i} \geq VaR_{t-i}(\tau))}} \sum_{i=1}^{250} L_{t-i} \times \mathbb{1}_{(L_{t-i} \geq VaR_{t-i}(\tau))}.$

In A_1, the conditional variance of the series σ_t is alternately underestimated of 25%, 50%, and 75% to examine whether our tests are able to detect an underestimation of ES stemming from a misleading appreciation of volatility. In A_2, the misspecification occurs in the conditional mean by assuming a GARCH in mean model. In A_3, the distribution

of the innovations η_t is incorrect, and should imply misleading ES predictions compared to the t-distribution. Finally in scenario A_4, the time-varying dynamics is incorrectly captured by the historical simulation method. It should be noticed that our alternatives are in line with the existing literature on risk assessment. Bayer and Dimitriadis (2018) look at an alternative close to A_1 by varying the coefficients related to the GARCH component. A_2, and A_3 were applied by Du and Escanciano (2017) to illustrate the performance of their unconditional and conditional ES backtests. Finally, scenario A_4 was extensively studied by Kratz et al. (2018), Bayer and Dimitriadis (2018), Gaglianone et al. (2011), among many others.

Tables 3.1 and 3.2 report the rejection frequencies of the tests at 5% significance level for sample sizes $T = 500$, and $T = 2500$, respectively. The first four columns report the results of the asymptotic backtests, and the last four columns embed the bootstrap based tests. As previously discussed, the use of asymptotic critical values (based on a χ^2 distribution) induces important size distortions. For instance, with sample size $T = 500$, and $p = 6$, the four test statistics J_1, J_2, I, and S, display empirical sizes equal to 0.126, 0.273, 0.165, 0.216, respectively. These distortions are caused by poor inference made on regression parameters in the extreme upper tail when the sample size is not sufficiently large. It is worth noting, however, that the tests become well-sized when applying more central coverage levels (cf. Appendix 3.7.3). In contrast, the backtests based on bootstrap critical values give satisfactory size performances. Empirical sizes are close to the nominal size of 5% for all reported sample sizes and risk levels. For large coverage levels and moderate sample sizes, we thus recommend the use of bootstrap critical values rather than asymptotic ones.

Our backtests display good power performances. The results are discussed in details hereinafter. First, we find that the tests generally detect well the misspecified alternatives A_1, A_2, A_3, and A_4, and we verify that there is a general improvement of powers as the sample size T increases, suggesting that these tests are consistent for these alternatives. For instance, with $T = 500$, and $p = 4$, the test statistic J_1 identifies the misleading scenario A_3 in 49.3% of times, while it reaches 98.1% of times with $T = 2500$. Second, the joint test statistics, J_1 and J_2, generally deliver higher power performances compared to the intercept and slope test statistics I, and S. This finding comes from the definition of the joint null hypotheses that focus on both intercept and slope coefficients, and are thus more conservative than the null of the intercept and slope backtests. In details for the two joint tests, we find that J_1 performs generally better to detect A_1 and A_4, while J_2 more often identifies A_2 and A_3, which suggests complementarity between the two joint backtests.

Third, although the intercept and slope backtests exhibit lower power performances, they provide useful informations about the type of misspecification. We observe that the

Table 3.1: Empirical rejection rates of the backtests at 5% significance level, $T = 500$

		J_1	J_2	I	S	$J_1^{(b)}$	$J_2^{(b)}$	$I^{(b)}$	$S^{(b)}$
					$p = 1$				
H_0		0.130	0.303	0.186	0.241	0.057	0.058	0.058	0.061
A_1	$\kappa = 0.25$	0.068	0.053	0.054	0.055	0.070	0.059	0.054	0.047
	$\kappa = 0.50$	0.591	0.055	0.053	0.061	0.434	0.070	0.061	0.044
	$\kappa = 0.75$	0.665	0.811	0.061	0.079	0.575	0.639	0.068	0.062
A_2	$\kappa = +2.5$	0.141	0.995	0.391	0.833	0.091	0.994	0.342	0.846
	$\kappa = -2.5$	0.104	0.997	0.989	0.763	0.065	0.996	0.988	0.773
A_3		0.623	0.994	0.247	0.106	0.523	0.992	0.202	0.117
A_4		0.120	0.145	0.208	0.120	0.079	0.128	0.165	0.128
					$p = 2$				
H_0		0.116	0.278	0.166	0.223	0.055	0.058	0.057	0.059
A_1	$\kappa = 0.25$	0.059	0.054	0.052	0.058	0.071	0.061	0.051	0.042
	$\kappa = 0.50$	0.637	0.055	0.049	0.062	0.355	0.062	0.049	0.044
	$\kappa = 0.75$	0.802	0.693	0.089	0.072	0.643	0.553	0.053	0.059
A_2	$\kappa = +2.5$	0.078	0.987	0.353	0.840	0.041	0.977	0.341	0.854
	$\kappa = -2.5$	0.060	0.976	0.968	0.692	0.042	0.967	0.967	0.709
A_3		0.640	0.971	0.220	0.145	0.487	0.955	0.209	0.152
A_4		0.172	0.149	0.211	0.144	0.069	0.128	0.206	0.164
					$p = 4$				
H_0		0.150	0.277	0.165	0.199	0.054	0.057	0.058	0.058
A_1	$k = 0.25$	0.057	0.054	0.050	0.051	0.068	0.061	0.053	0.049
	$\kappa = 0.50$	0.582	0.071	0.069	0.053	0.281	0.061	0.040	0.045
	$\kappa = 0.75$	0.816	0.602	0.073	0.054	0.659	0.430	0.072	0.065
A_2	$\kappa = +2.5$	0.060	0.971	0.318	0.776	0.053	0.941	0.317	0.821
	$\kappa = -2.5$	0.088	0.924	0.932	0.627	0.045	0.891	0.930	0.691
A_3		0.646	0.975	0.203	0.121	0.493	0.933	0.202	0.161
A_4		0.211	0.151	0.239	0.155	0.098	0.141	0.237	0.172
					$p = 6$				
H_0		0.126	0.273	0.165	0.216	0.054	0.059	0.059	0.062
A_1	$\kappa = 0.25$	0.058	0.052	0.054	0.054	0.070	0.047	0.050	0.048
	$\kappa = 0.50$	0.552	0.054	0.048	0.055	0.286	0.069	0.053	0.043
	$\kappa = 0.75$	0.841	0.471	0.047	0.058	0.715	0.404	0.050	0.063
A_2	$\kappa = +2.5$	0.045	0.958	0.319	0.786	0.034	0.949	0.334	0.841
	$\kappa = -2.5$	0.111	0.878	0.927	0.571	0.045	0.862	0.930	0.636
A_3		0.651	0.955	0.180	0.110	0.502	0.938	0.192	0.148
A_4		0.236	0.162	0.272	0.181	0.148	0.241	0.261	0.194

Note: The results based on asymptotic critical values are reported in the first four columns. The results using bootstrap critical values are displayed in the last four columns, and indicated by (b) in the table. Reported powers are size-corrected.

slope backtest performs better in alternatives A_1 and A_3, while the intercept backtest is superior for alternative A_4. In line with these rejections, A_1 and A_3 mainly affect the expected value of the slope parameters indicating that the error is essentially multiplicative with respect to the true quantiles. On the contrary, alternative A_4 induces distortions in the expected value of the intercept coefficients, suggesting that the VaRs issued from the ES model are affected additively. In the latter case, the error is hence more global and not directly related to the true quantiles.

Table 3.2: Empirical rejection rates of the backtests at 5% significance level, $T = 2500$

		J_1	J_2	I	S	$J_1^{(b)}$	$J_2^{(b)}$	$I^{(b)}$	$S^{(b)}$
					$p = 1$				
H_0		0.090	0.186	0.141	0.172	0.045	0.051	0.054	0.054
A_1	$\kappa = 0.25$	0.935	0.503	0.063	0.046	0.897	0.222	0.052	0.055
	$\kappa = 0.50$	0.998	1.000	0.074	0.168	0.995	1.000	0.071	0.112
	$\kappa = 0.75$	0.994	1.000	0.108	0.472	0.985	1.000	0.080	0.364
A_2	$\kappa = +2.5$	0.360	1.000	0.914	1.000	0.308	1.000	0.872	0.997
	$\kappa = -2.5$	0.303	1.000	1.000	0.987	0.249	1.000	1.000	0.983
A_3		0.990	1.000	0.749	0.807	0.984	1.000	0.679	0.711
A_4		0.855	0.518	0.718	0.552	0.761	0.361	0.614	0.448
					$p = 2$				
H_0		0.103	0.184	0.160	0.178	0.045	0.052	0.052	0.056
A_1	$\kappa = 0.25$	0.849	0.308	0.058	0.069	0.745	0.208	0.055	0.026
	$\kappa = 0.50$	0.998	1.000	0.065	0.153	0.997	0.999	0.051	0.113
	$\kappa = 0.75$	0.995	1.000	0.107	0.485	0.992	1.000	0.086	0.407
A_2	$\kappa = +2.5$	0.141	1.000	0.893	0.997	0.102	1.000	0.865	0.996
	$\kappa = -2.5$	0.231	1.000	0.998	0.988	0.192	1.000	0.997	0.985
A_3		0.988	1.000	0.700	0.879	0.983	1.000	0.638	0.827
A_4		0.879	0.507	0.794	0.640	0.803	0.426	0.725	0.545
					$p = 4$				
H_0		0.090	0.163	0.126	0.137	0.049	0.054	0.049	0.055
A_1	$\kappa = 0.25$	0.632	0.271	0.058	0.069	0.503	0.209	0.050	0.052
	$\kappa = 0.50$	0.998	1.000	0.079	0.085	0.998	1.000	0.054	0.090
	$\kappa = 0.75$	0.997	1.000	0.085	0.415	0.996	1.000	0.076	0.428
A_2	$\kappa = +2.5$	0.104	1.000	0.893	0.997	0.080	1.000	0.871	0.997
	$\kappa = -2.5$	0.384	1.000	0.995	0.980	0.330	1.000	0.994	0.981
A_3		0.984	1.000	0.577	0.826	0.981	1.000	0.549	0.840
A_4		0.894	0.530	0.767	0.651	0.851	0.474	0.741	0.589
					$p = 6$				
H_0		0.108	0.172	0.125	0.139	0.051	0.053	0.055	0.057
A_1	$\kappa = 0.25$	0.470	0.318	0.055	0.060	0.416	0.219	0.047	0.052
	$\kappa = 0.50$	0.997	0.997	0.058	0.114	0.997	0.997	0.057	0.114
	$\kappa = 0.75$	0.999	1.000	0.069	0.497	0.998	1.000	0.080	0.502
A_2	$\kappa = +2.5$	0.095	1.000	0.854	1.000	0.082	1.000	0.865	1.000
	$\kappa = -2.5$	0.573	0.998	0.995	0.985	0.548	0.998	0.995	0.985
A_3		0.996	1.000	0.549	0.878	0.992	1.000	0.580	0.879
A_4		0.913	0.609	0.812	0.664	0.897	0.521	0.798	0.646

Note: The results based on asymptotic critical values are reported in the first four columns. The results using bootstrap critical values are displayed in the last four columns, and indicated by (b) in the table. Reported powers are size-corrected.

Finally, we show that the selection of the number p of risk levels is not crucial for detecting alternatives A_1, A_2, and A_3 since reported powers are weakly affected by p. This finding may be explained by the nature of these alternatives for which the misspecification is relatively uniform along the tail, and does not require many levels. In contrast, for the last alternative A_4, we conclude that an increase of p is valuable for detecting the misleading one-year historical simulation method since we observe a general improvement of powers for larger values of p. This is due to the fact that, for this alternative, the error made along the tail is more irregular and requires the use of additional levels.

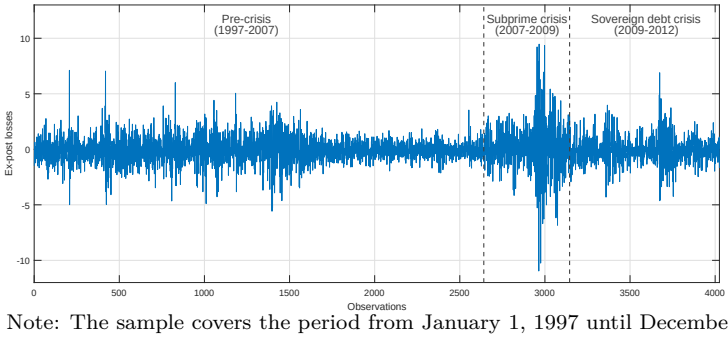
3.5 Empirical application

In this section, we apply our backtests to the daily returns of the S&P500 index, and illustrate their capabilities to identify a misspecified ES model. In addition, we provide a method for the adjustment of imperfect forecasts relying on our backtesting framework. In the sequel, we set $\tau = 0.975$ to coincide with the regulatory ES coverage level. The probability levels u_j, $j = 1, 2, \ldots, p$, are calculated accordingly with Definition 1. In addition, we consider the risk levels suggested by the BCBS, i.e. $u_1 = 0.975$, and $u_2 = 0.990$, respectively. Finally, for comparison purposes and to provide useful backtesting recommendations, we consider several values $p = 1, 2, 4$, and 6.

3.5.1 Data

We consider the daily adjusted closing prices of the S&P500 index over the period January 1, 1997 - December 31, 2012. The in-sample period spans from January 1, 1997 until June 30, 2007, and we consider two out-of-sample periods (1) from July 1, 2007 to June 30, 2009, corresponding to the subprime mortgage crisis, and (2) from July 1, 2007 to December 31, 2012, which pools together the subprime mortgage crisis and the European sovereign debt crisis, two major episodes of economic and financial instability. We compute the daily log-returns and denote by L_t the opposite returns. In line with our notations, a positive value indicates a loss.

Figure 3.1: S&P500 daily losses (%), and descriptive statistics



	1997-2007	2007-2009	2007-2012
Nb of obs	2641	504	1384
Mean	-0.028	0.105	0.005
Median	-0.059	-0.032	-0.067
Variance	1.280	4.932	2.643
Skewness	0.090	0.059	0.242
Kurtosis	3.171	4.028	6.943
Minimum	-5.574	-10.957	-10.957
Maximum	7.113	9.470	9.470

Note: The sample covers the period from January 1, 1997 until December 31, 2012. Source: *finance.yahoo.com* website.

The S&P500 series is displayed in Figure 3.1 with the three aforementioned sub-periods. The in-sample period (1997-2007) is weakly volatile, while the out-of-sample crisis periods (2007-2009, and 2007-2012) are characterized by more severe levels of volatility, with several extreme events. Figure 3.1 also provides some descriptive statistics. The variance and the average ex-post losses are higher in the out-of-sample periods than in the in-sample period, especially for the period 2007-2009. In addition, we observe that the series is right-skewed and has an excess kurtosis.

To predict the ES risk measure, we fit an AR(1)-GARCH(1,1) model with Student innovations, as defined in (3.19). The ES and VaR forecasts are defined as in Equations

(3.20) and (3.21), respectively. The set of unknown parameters is estimated by maximum likelihood over the in-sample period. We obtain the following coefficient estimates $\left\{\widehat{\delta}_0, \widehat{\delta}_1, \widehat{\gamma}_0, \widehat{\gamma}_1, \widehat{\gamma}_2, \widehat{\upsilon}\right\} = \{-0.0568, -0.0321, 0.0067, 0.0603, 0.9356, 9\}$.

3.5.2 Empirical results

We start by evaluating the relevancy of the ES approximation of Definition 1, consisting in averaging several quantiles in the tail of the risk model. To do so, we compare the approximation considering $p = 1, 2, 4, 6$ quantiles, with what we refer to as "exact ES". The latter corresponds to an ES which is computed via an exact method of calculation. The technique relies on simulations, and is described in Appendix 3.7.4.

Figure 3.2: In-sample ES estimates issued from the approximation, and the exact calculation method

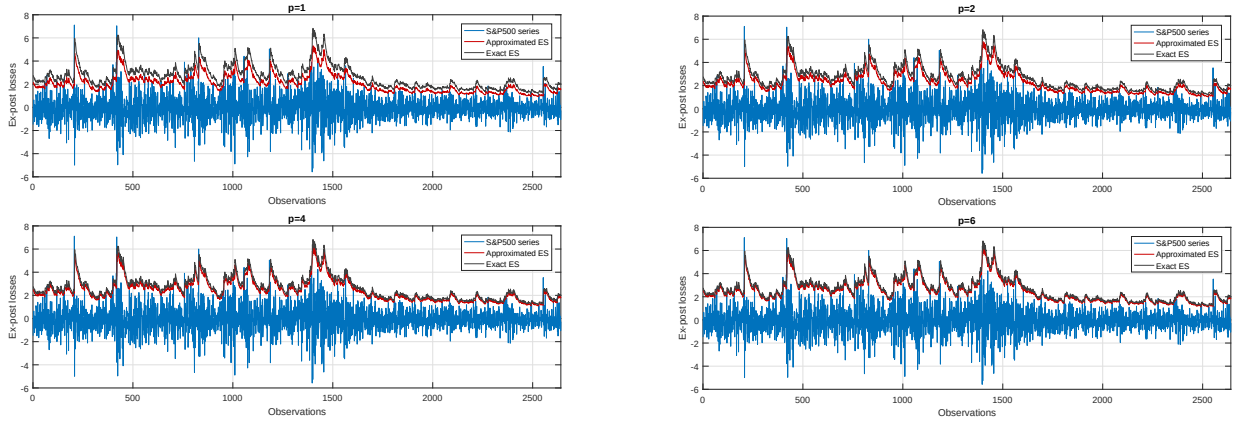


Figure 3.2 reports the in-sample ES estimates obtained using the approximation and the exact calculation method. Two remarks should be made here. First, the ES forecasts issued from the approximation and the exact method strongly correlate regardless of the value p, indicating that the approximation performs very well to capture the ex-post losses information. Because the approximation is obtained by combining VaRs, our finding is in accordance with several papers. Gouriéroux and Liu (2012) study the relationship between VaR and ES and show that they are related through their risk levels by some link function. Daníelsson et al. (2016) argue that the two measures of risk are related by a small constant, and hence are conceptually equally informative. Second, we observe that the approximation is substantially improved when p slightly increases. As an illustration, we find that the ES estimates issued from the two competing approaches coincide almost completely using six quantiles. For its ease of implementation and its accuracy, the approximation hence appears highly appealing to compute ES and evaluate the performance of its forecasts.

Table 3.3 reports the p-values of the backtests. For a sake of clarity, we only report the p-values obtained with the bootstrap critical values. Panel A provides the results

Table 3.3: p-values of the backtests for several number p of quantiles

p	$J_1^{(b)}$	$J_2^{(b)}$	$I^{(b)}$	$S^{(b)}$
		Panel A. 2007-2009		
1	0.035	0.051	0.125	0.949
2	0.014	0.041	0.038	0.200
4	0.009	0.040	0.023	0.103
6	0.009	0.038	0.021	0.123
2 *(regulatory levels)*	0.024	0.047	0.053	0.351
		Panel B. 2007-2012		
1	0.056	0.040	0.176	0.554
2	0.004	0.013	0.014	0.215
4	0.002	0.004	0.003	0.096
6	0.004	0.005	0.009	0.196
2 *(regulatory levels)*	0.006	0.012	0.032	0.448

Note: p-values of the four backtests computed with $p = 1, 2, 4, 6$ risk levels successively, and the two regulatory levels $u_1 = 0.975$, $u_2 = 0.990$. Reported p-values are obtained using bootstrap critical values. Panel A gives the results for the period 2007-2009, and Panel B provides results for the period 2007-2012.

over the sample 2007-2009. The test statistic J_1 leads to reject the validity of the ES predictions regardless of the number of quantiles p. We also observe that the larger p, the smaller the p-value, indicating that the rejections are more severe when the number of risk levels increases. The test statistic J_2 displays higher p-values. At 5% significance level, the backtest based on a single VaR no longer rejects the validity of the ES predictions, and the p-value based on the regulatory levels of the BCBS is close to 5%, making the decision rule more unclear for those number of risk levels. Finally, given the p-values of the test statistic I for $p = 2, 4, 6$, we reject the expected value of the intercept coefficients, leading to the rejection of the ES forecasts in an additive viewpoint. On the contrary, the test statistic S leads to the conclusion that the slope parameters are as expected under the null hypothesis. Panel B contains the p-values for the period 2007-2012. Overall, we obtain similar results, but the rejections are found more severe in this enlarged sample due to the consistency of our backtesting methodology when applying it to larger sample sizes.

Given these results, it emerges that we should be very cautious in using a single quantile to assess the tail distribution of the risk model. Such procedures may lead market practitioners to select a model that generates mistaken ex-post forecasts. In addition, the results issued from the regulatory guidelines are contrasted. Two risk levels are not always enough to provide a sound conclusion about the correctness of the ES forecasts. We recommend the use of additional risk levels beyond the regulatory coverage level $\tau = 0.975$ to improve the reliability of the decision. Typically, our tests give satisfactory results with four to six risk levels.

Table 3.4 displays the coefficient estimates of the multi-quantile regression of Equation (3.8) for $p = 6$ risk levels, to help understand the reasons that explain the rejections of the

Table 3.4: Coefficient estimates issued from the multi-quantile regression, $p = 6$

	u_1	u_2	u_3	u_4	u_5	u_6
			Panel A. 2007-2009			
β_0	0.661	0.696	0.808**	0.846**	0.965*	1.076*
	(0.295)	(0.296)	(0.227)	(0.240)	(0.429)	(0.265)
β_1	1.005	0.953	0.911*	0.847**	0.804	0.689**
	(0.093)	(0.088)	(0.056)	(0.053)	(0.142)	(0.042)
joint	*	*	**	**		**
			Panel B. 2007-2012			
β_0	0.376	0.510*	0.692***	0.808***	0.777**	0.784
	(0.200)	(0.182)	(0.195)	(0.186)	(0.284)	(0.611)
β_1	1.031	0.974	0.902	0.851**	0.826	0.787
	(0.073)	(0.067)	(0.065)	(0.050)	(0.107)	(0.232)
joint	**	**	**	**	**	

Note: Standard errors are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5% and 1% level, respectively, and are obtained using bootstrap critical values. Panel A gives estimation results for the period 2007-2009, and Panel B provides estimation results for the period 2007-2012.

ES forecasts. Panel A, and B, respectively provide the results for periods 2007-2009, and 2007-2012. It must be recalled that, if the risk model is correctly specified, the intercept coefficient β_0 and the slope coefficient β_1 take values zero and one, respectively. With these values in mind, we observe in both panels that the coefficients β_0 are overestimated for all the risk levels $u_1, u_2, \ldots, u_6$, while the coefficient β_1 is overestimated for the first level u_1, and it becomes underestimated for all the remaining risk levels $u_2, u_3, \ldots, u_6$. The average errors of β_0 and β_1 are respectively equal to 0.84 and -0.13 in panel A, and 0.66 and -0.10 in panel B, indicating that the magnitude of errors is more important in panel A than in panel B, and that the intercept coefficients are more affected than the slope coefficients. Finally, we observe that the distortion of the regression coefficients with respect to their expected values is more pronounced for the highest risk levels suggesting that the errors are more severe far in the tail.

Table 3.4 also provides inference on the regression parameters obtained for each risk level. Hereinafter, we discuss the results at 5% significance level. We observe that the intercept parameters are statistically not equal to zero for the intermediary risk levels u_3 and u_4 in panel A, and the additional u_5 risk level is also significantly different from zero in panel B. For the slope coefficients, the u_4 and u_6 order quantiles are statistically different from one in panel A, and only the level u_4 is misspecified in panel B. Finally, we also report the joint inference, i.e. looking at both the intercept and slope coefficients. The results are provided in the row labeled as "joint" (bottom of the panels). Similarly to the previous findings, we find that the intermediary, and highest order quantiles u_3, u_4 and u_6 are misleading in panel A, whereas in panel B, all the quantiles are misspecified (except for the highest, presumably because the coefficients have large standard errors), meaning that the entire tail distribution is incorrectly estimated.

3.5.3 Adjusted ES forecasts

In what follows, we exploit our testing strategy to provide adjusted ES forecasts. Our routine is designed to take into account both misspecification and estimation uncertainty, without having to change the misspecified risk model. Furthermore, the procedure may serve to identify whether the model overestimates, or underestimates the true unknown ES, by comparing the initial forecast with its adjusted counterpart, which appears useful in a risk management and regulatory viewpoint.

The correction of imperfect risk forecasts is not a novel concept in the financial literature. Gouriéroux and Zakoïan (2013) propose to adjust the VaR forecasts affected by estimation uncertainty. Similarly, Boucher et al. (2014) adjust imperfect VaR forecasts based on backtesting frameworks, and recently Lazar and Zhang (2019) apply the same strategy to adjust imperfect ES forecasts. The method typically consists in modifying the coverage level τ of the risk measure so as to meet the null hypothesis of valid risk forecasts. The originality of our technique stems from the fact that we employ a regression-based framework to correct the ex-ante forecasts, while available techniques are generally based on the concept of violation. This allows us to directly adjust the risk forecasts by application of a regression model, without having to rescale the coverage level τ.

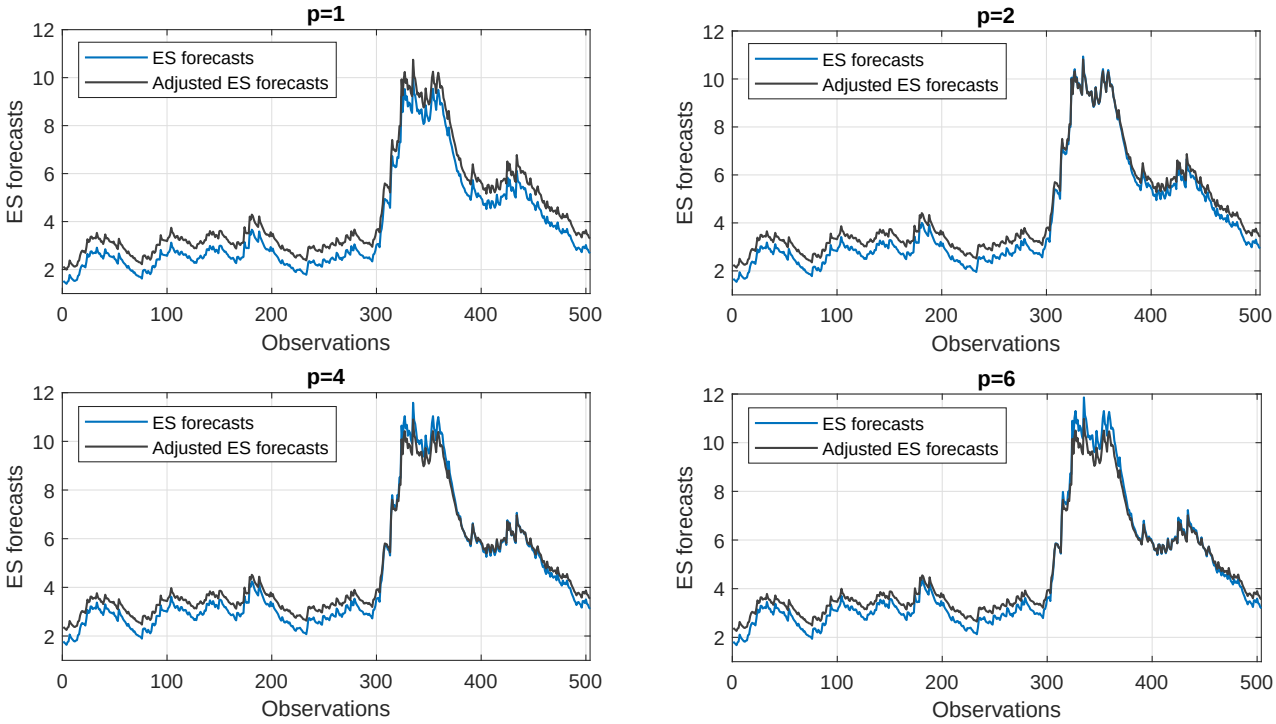
For ease of notation, we assume the parameters of the multi-quantile regression to be known. Formally, the adjusted VaR forecast at level u_j, and time t, is defined as the ex-ante prediction of the multi-quantile regression model, namely $Q_{L_t}(u_j; \Omega_{t-1})$. In view of representation (3.8), the initial imperfect VaR forecast is subsequently weighted by the regression parameters $\beta_0(u_j)$ and $\beta_1(u_j)$, which provides an adjustment corresponding to the global bias caused by misspecification and estimation uncertainty. The adjusted ES forecast at coverage level τ, and time t, is derived from the approximation of ES in terms of several adjusted VaRs as follows:

$$ES_t^*(\tau) = \frac{1}{p}\sum_{j=1}^{p} Q_{L_t}(u_j; \Omega_{t-1}). \tag{3.22}$$

The adjusted ES forecasts are robust to model risk, as they meet the desirable properties on the regression coefficients. Indeed, if we compute the backtesting procedure with the sequence $\{Q_{L_t}(u_j; \Omega_{t-1})\}_{j=1}^{p}$ instead of the initial misleading $\{VaR_t(u_j)\}_{j=1}^{p}$, the parameters would exactly coincide with the expected values under the null hypothesis, i.e. $\beta_0(u_j) = 0$, and $\beta_1(u_j) = 1$, for the risk levels $u_1, u_2, \ldots, u_p$.

Figure 3.3 reports the ES predictions, and adjusted ES predictions for the period 2007-2009. The risk forecasts are built using the approximation with $p = 1, 2, 4, 6$. We observe that the AR(1)-GARCH(1,1) structure generally provides underestimated risk forecasts compared to the adjusted predictions. We note that the underestimation is

Figure 3.3: ES forecasts and adjusted ES forecasts over the period 2007-2009



more pronounced for the smallest predictions, the error being more severe when the risk forecasts are originally small. Our procedure hence serves at identifying whether the model generates overestimates, or underestimates, the latter case being more worrisome in a financial stability perspective. Finally, we observe that the ES forecasts may be slightly overestimated when the variance of the series is larger, suggesting that the risk model tends to overestimate the true volatility in turbulent financial times. This is caused by the volatility persistence in the GARCH component. We observe similar results for the period 2007-2012, as well as when applying the risk levels of the BCBS. The corresponding Figures 3.4, and 3.5 are reported in Appendix 3.7.5.

3.6 Conclusion

The financial crisis of 2007-2008 and its aftermath has led to a reassessment of risk-management practices and financial market regulation through the Basel III accords (BCBS, 2010). Among the number of fundamental reforms for the market risk, the BCBS has adopted ES in place of VaR as the new standard for risk management and regulatory requirements. One of the major obstacle to its implementation was undoubtedly the deficit of simple tools for the evaluation of its forecasts. In this chapter, we have introduced four easy-to-use regression-based backtests of ES. Our methodology explores jointly the validity of the VaR forecasts along the tail distribution of the risk model. This approach has the advantage of being consistent with the regulatory guidances to verify

if the underlying ES model delivers correct quantiles at levels 0.975 and 0.990 (BCBS, 2016). To do so, we generalize the testing procedure introduced by Gaglianone et al. (2011) to the multivariate quantile framework. This econometric approach consists in regressing the ex-post losses on the VaRs forecasts in a multi-quantile regression model, and then, testing the resulting parameter estimates using Wald-type inference.

Several simulation studies are provided. We find that the use of asymptotic critical values may lead to important size distortions if the sample size is not large enough. We propose a pairs bootstrap algorithm to correct these small-sample biases (Freedman, 1981), and show that our regression-based tests are reasonably sized within this bootstrap framework. We consider several misleading alternatives in line with the existing literature on risk assessment (Gaglianone et al., 2011; Du and Escanciano, 2017; Bayer and Dimitriadis, 2018; Kratz et al., 2018, etc.). Our methodology detect misspecifications in all considered simulation experiments. In particular, they identify the most frequent inaccuracies in risk modeling, namely mean, variance, tail, and dynamic misspecifications.

We apply our tests on the S&P500 index over the period 2007-2012. Our backtests clearly reject the validity of the ES forecasts based on a simple AR(1)-GARCH(1,1) model during this period of financial turmoil. Beyond this result, we highlight the importance of choosing a sufficient number of quantiles to assess ES. The use of one or two risk levels is especially inadvisable as they are not always enough to identify improper risk forecasts. On the contrary, four to six risk levels deliver much more sound decisions, suggesting an update of the current regulatory guidelines in favor of the evaluation of more than two quantiles.

Within the current debate on risk assessment, the proposed regression-based backtests are valuable diagnostic tools in line with the expertise and skills of financial supervisors. They have the advantages to be easy to implement, and to complete the toolbox commonly used by market practitioners. They are therefore more likely to be embraced by financial institutions as new standards for financial risk management.

3.7 Appendix

3.7.1 Appendix A - Consistent variance-covariance matrix estimation

In what follows, we provide a consistent estimator of the variance-covariance matrix Σ. The methodology is derived from White et al. (2008, 2015) who provide asymptotic theory in a multi-quantile regression framework. A consistent estimate of Σ can be obtained from the decomposition of the Huber (1967) sandwich form and is thus given by $\widehat{\Sigma} = \widehat{A}^{-1}\widehat{V}\widehat{A}^{-1}$. In the sequel, we provide consistent estimators $\widehat{A}$, and $\widehat{V}$. To obtain $\widehat{V}$, we apply a simple plug-in estimator as follows:

$$\widehat{V} = T^{-1}\sum_{t=1}^{T}\widehat{\eta}_t\widehat{\eta}_t', \tag{3.23}$$

where $\widehat{\eta}_t$ is given by its estimated counterpart $\widehat{\eta}_t = \sum_{j=1}^{p}\nabla\widehat{Q}_{L_t}(u_j,\Omega_{t-1})\,\psi_{u_j}(\widehat{\epsilon}_{j,t})$, with $\widehat{Q}_{L_t}(u_j,\Omega_{t-1}) = \widehat{\beta}_0(u_j) + \widehat{\beta}_1(u_j)\,VaR_t(u_j)$, and $\widehat{\epsilon}_{j,t} = L_t - \widehat{Q}_{L_t}(u_j,\Omega_{t-1})$.

The estimation of A is trickier because it requires to consistently estimate $f_{j,t}(0)$, namely the density of the error term $\epsilon_{j,t}$ given Ω_{t-1} evaluated at zero. Because the function is unknown, we follow Powell (1984) and use a non parametric estimator. The method was also implemented by Engle and Manganelli (2004) to estimate the variance-covariance matrix of a set of coefficients issued from the so-called CaViaR model. $\widehat{A}$ is then given by

$$\widehat{A} = (2\widehat{c}_T T)^{-1}\sum_{t=1}^{T}\sum_{j=1}^{p}\mathbb{1}\left(|\widehat{\epsilon}_{j,t}| \leq \widehat{c}_T\right)\nabla\widehat{Q}_{L_t}(u_j,\Omega_{t-1})\,\nabla'\widehat{Q}_{L_t}(u_j,\Omega_{t-1}), \tag{3.24}$$

where $\widehat{c}_T$ is a bandwidth parameter that must verify $\widehat{c}_T/c_T \xrightarrow{p} 1$, with c_T a nonstochastic positive sequence satisfying $c_T = o(1)$, and $c_T^{-1} = o(T^{1/2})$. Throughout this chapter, we select a bandwidth parameter $\widehat{c}_T = T^{-1/7}$ which verifies the above properties.

3.7.2 Appendix B - Proof of consistency under fixed untrue hypothesis

Proof. In line with our previous notations, we term W the generic notation of the test statistic such that $W \in \{J_1, J_2, I, S\}$. The test statistic is given by

$$W = T(R_W\widehat{\beta} - q_W)'(R_W\widehat{\Sigma}R_W')^{-1}(R_W\widehat{\beta} - q_W). \tag{3.25}$$

The null hypothesis of the proposed Wald-type test can be written as $H_{0,W} : R_W\beta - q_W = 0$, against the two-sided alternative $H_{1,W} : R_W\beta - q_W \neq 0$. The continuous mapping theorem implies under $H_{1,W}$ that

$$R_W\widehat{\beta} - q_W \xrightarrow{p} R_W\beta - q_W \neq 0. \tag{3.26}$$

Rearranging the term T in the test statistic and using the continuous mapping theorem leads

$$WT^{-1} \xrightarrow{p} (R_W\beta - q_W)'(R_W\Sigma R_W')^{-1}(R_W\beta - q_W). \tag{3.27}$$

Because $(R_W\Sigma R_W')^{-1}$ is positive definite, we get under $H_{1,W}$: $(R_W\beta - q_W)'(R_W\Sigma R_W')^{-1}(R_W\beta - q_W) > 0$. Multiplying $(R_W\beta - q_W)'(R_W\Sigma R_W')^{-1}(R_W\beta - q_W)$ by T under $H_{1,W}$ hence gives

$$\lim_{T\to+\infty} W = +\infty, \tag{3.28}$$

and therefore we get

$$\lim_{T\to+\infty} \mathbb{P}(W > \chi^2_{1-\alpha}(d_W) \,|H_{1,W}) = 1, \tag{3.29}$$

where $\chi^2_{1-\alpha}(d_W)$ is the fractile of order $1-\alpha$ of the chi-square distribution with d_W degrees of freedom, and where α is the significance level of the test. ■

3.7.3 Appendix C - Empirical sizes for more central coverage levels

Table 3.5: Empirical sizes of the asymptotic backtests at a 5% significance level, T=500, T=2500

	T=500				T=2500			
τ	J_1	J_2	I	S	J_1	J_2	I	S
				$p = 1$				
0.5	0.045	0.034	0.042	0.048	0.048	0.044	0.048	0.045
0.6	0.053	0.045	0.028	0.051	0.064	0.059	0.048	0.061
0.7	0.081	0.063	0.042	0.075	0.074	0.055	0.055	0.072
0.8	0.085	0.082	0.072	0.087	0.070	0.073	0.065	0.069
0.9	0.107	0.149	0.117	0.135	0.071	0.105	0.090	0.087
0.975	0.130	0.303	0.186	0.241	0.090	0.186	0.141	0.172
				$p = 2$				
0.5	0.049	0.041	0.042	0.049	0.046	0.050	0.042	0.044
0.6	0.058	0.077	0.062	0.056	0.059	0.065	0.066	0.058
0.7	0.072	0.079	0.062	0.072	0.060	0.069	0.060	0.070
0.8	0.082	0.115	0.086	0.089	0.067	0.085	0.083	0.081
0.9	0.118	0.160	0.138	0.142	0.081	0.121	0.106	0.107
0.975	0.116	0.278	0.166	0.223	0.103	0.184	0.160	0.178
				$p = 4$				
0.5	0.049	0.057	0.054	0.050	0.043	0.052	0.053	0.046
0.6	0.055	0.072	0.066	0.053	0.061	0.066	0.073	0.060
0.7	0.073	0.097	0.076	0.073	0.069	0.089	0.079	0.076
0.8	0.081	0.110	0.088	0.083	0.064	0.104	0.083	0.085
0.9	0.092	0.158	0.114	0.128	0.067	0.104	0.084	0.086
0.975	0.150	0.277	0.165	0.199	0.090	0.163	0.126	0.137
				$p = 6$				
0.5	0.055	0.058	0.056	0.057	0.043	0.055	0.055	0.046
0.6	0.061	0.079	0.079	0.067	0.047	0.068	0.058	0.047
0.7	0.079	0.107	0.088	0.080	0.052	0.070	0.057	0.055
0.8	0.079	0.118	0.100	0.093	0.061	0.089	0.073	0.074
0.9	0.098	0.168	0.122	0.137	0.060	0.105	0.086	0.086
0.975	0.126	0.273	0.165	0.216	0.108	0.172	0.125	0.139

Note: Empirical sizes of the four backtests using asymptotic critical values (based on the χ^2 distribution) for sample sizes $T = 500$ and $T = 2500$. The results are reported for more central coverage levels $\tau = 0.5, 0.6, 0.7, 0.8, 0.9, 0.975$, and number of risk levels $p = 1, 2, 4, 6$.

3.7.4 Appendix D - Exact calculation method of ES

In this section, we describe the methodology used for the computation of exact ES forecasts at a given coverage level τ. Several techniques are available in practice. In the following, because the distribution of the innovations is parametric, we rely on Monte Carlo simulations. For ease of notation, we assume parameters to be known while in practice we use estimated parameters. The algorithm is as follows:

1. Randomly draw S pseudo standardized innovations $\{\eta_t^s\}_{s=1}^S$ from the Student distribution, with degrees of freedom v. We set the number $S = 100000$ in the empirical application.

2. Compute the ES at time t of the standardized innovation η_t as the Monte Carlo average of the simulated innovations:

$$m(\tau) = \frac{1}{\sum_{s=1}^S \mathbb{1}\left(\eta_t^s \geq F_v^{-1}(\tau)\right)} \sum_{s=1}^S \eta_t^s \times \mathbb{1}\left(\eta_t^s \geq F_v^{-1}(\tau)\right),$$

   where $F_v^{-1}(\tau)$ is the τ-quantile of the innovation distribution and is obtained as follows

$$F_v^{-1}(\tau) = \text{percentile}\left(\{\eta_t^s\}_{s=1}^S, 100\tau\right).$$

3. Compute the ES at time t as follows:

$$ES_t(\tau) = \delta_0 + \delta_1 L_{t-1} + \sigma_t m(\tau).$$

3.7.5 Appendix E - Adjusted ES forecasts

Figure 3.4: ES forecasts and adjusted ES forecasts over the period 2007-2012

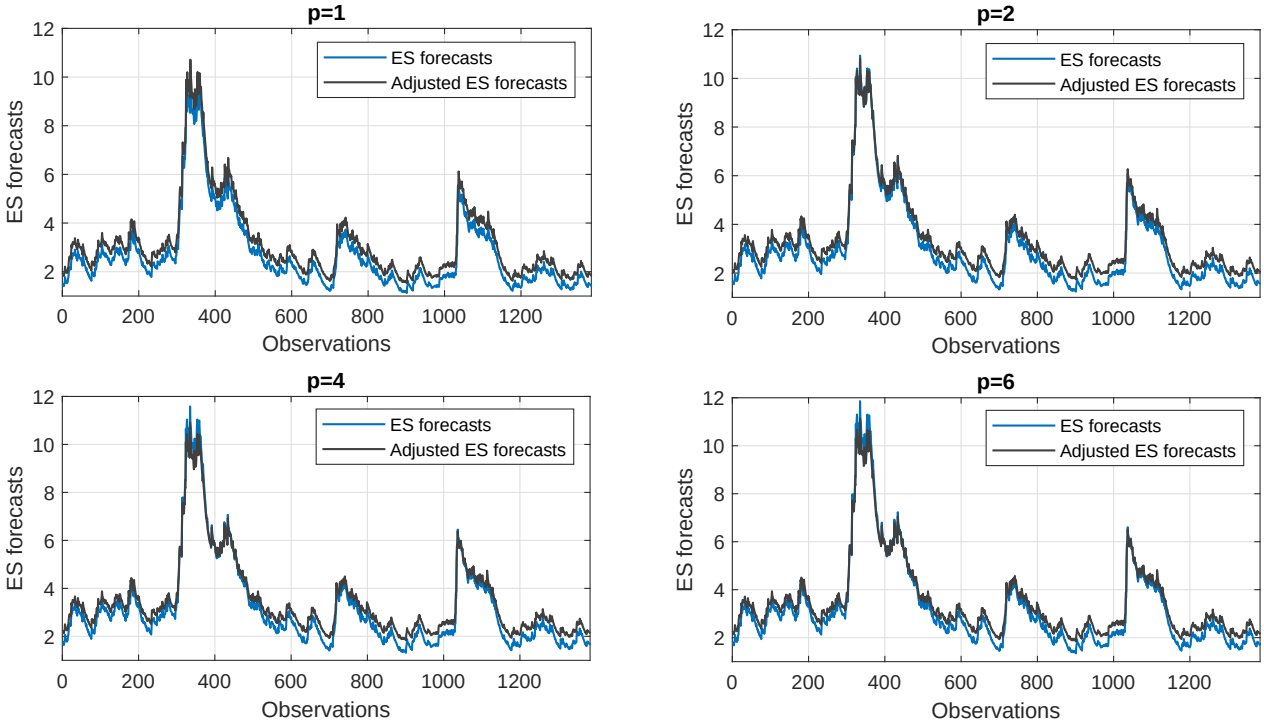
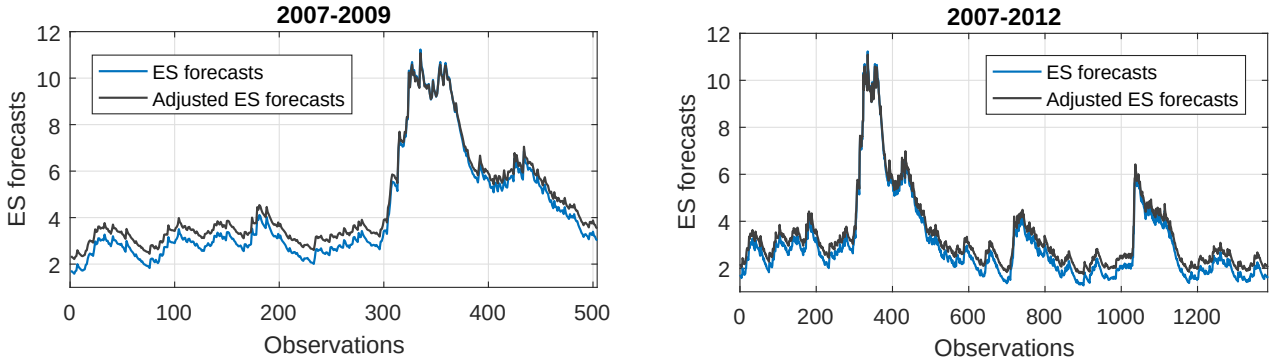


Figure 3.5: ES forecasts and adjusted ES forecasts over the periods 2007-2009 (on the left) and 2007-2012 (on the right) with the two regulatory risk levels



Chapter 4

Backtesting Marginal Expected Shortfall and Related Systemic Risk Measures[1]

This chapter studies validation tools in backtesting systemic risk measures. They allow assessing systemic risk measure forecasts used to identify the financial institutions contributing the most to the whole risk of the financial system. They are similar to those assessing the market risk measures such as the value-at-risk or the expected shortfall. We introduce a concept of violation associated to the marginal expected shortfall (MES), and we define unconditional coverage and independence tests for these violations. We can generalize these tests to any MES-based systemic risk measures such as SES, SRISK, or ΔCoVaR. We study their asymptotic properties in presence of estimation risk and investigate their finite sample performances via Monte Carlo simulations. An empirical application is then carried out in order to check the validity of the MES, SRISK, and ΔCoVaR forecasts issued from a GARCH-DCC model for a panel of U.S. financial institutions. We conclude that this risk model is able to produce valid forecasts for the MES and SRISK when considering a medium term horizon. Finally, we show how to derive an early warning system for financial crisis from these backtesting tests, by defining an adjusted MES risk measure.

4.1 Introduction

October 26, 2014: the European Banking Authority (EBA) publishes the results of the 2014 EU-wide stress test, which involves 123 banks from 22 countries, covering broadly 70% of the total EU banking sector assets (EBA, 2014). The main conclusion is that only

[1]This chapter is based on Banulescu-Radu, Hurlin, Leymarie and Scaillet (2018), and has been awarded a research grant sponsored by the Fondation Banque de France

24 participating banks fall below the defined capital threshold, leading to a maximum aggregate capital shortfall of €24.6bn. There is no French bank among these banks, and Christian Noyer, governor of the Banque de France at that time, salutes immediately the success of the "*French banks which are in the best positions in the Eurozone*" (Noyer, October 26, 2014). A day after these announcements, Viral Acharya (NYU) declares for Financial Times in an article entitled "French banks are the weakest in Europe", that in case of a crisis, the French financial institutions would face a capital shortfall of almost $400bn, i.e., around 15 times more than the shortfall identified by EBA for the weakest European banks (Financial Times, 2014). These results are based on SRISK, an alternative measure of capital shortfall. Looking just at the French banks tested in the ECB stress tests, for which no capital shortfall was identified, the SRISK reaches $189bn, i.e., approximately 9% of the 2014 French GDP.

This debate illustrates the difficulty of measuring systemic risk and the need for validation tools of its measures.[2] In this chapter, we propose a general framework for assessing the validity of systemic risk measures which are expressed as functions of expected equity loss conditional to a financial crisis, as it the case for the SRISK for instance.

The ultimate goal of a systemic risk measure is to better identifying the vulnerabilities of the financial system (see De Bandt and Hartmann, 2002; Benoit et al., 2017, for a survey). Ideally, regulators need measures that are timely, capture well-identified economic mechanisms, and can consist in inputs for regulatory instruments. However, in practice, measuring the systemic risk is challenging, and we can identify two main families of systemic risk measures. The first set aggregates low-frequency regulatory data. A typical example is the systemic risk score currently implemented by the Basel Committee on Banking Supervision (BCBS) and the Financial Stability Board (FSB) in order to identify the so-called systemically important financial institutions (SIFIs), i.e., the firms whose failure might trigger a crisis in the whole financial system.[3] Another example is the bank capital shortfall computed from the regulatory banking stress tests previously mentioned. Until 2014, we could not assess empirically the performance of these measures because the necessary data were not in the public domain. Recently, Philippon et al. (2017) propose the first evaluation of the quality of the EU banking stress tests, and Benoit et al. (2019) identify some pitfalls in the systemic risk scoring. The second set relies on high-frequency market data such as stock or asset returns, option prices, or CDS spreads. These measures have the advantage of being easily implemented with public data and standard econometric models. Many global risk measures have been proposed

[2]For similar discussions see also Acharya et al. (2014), Tavolaro and Visnovsky (2014), Acharya et al. (2016a,b), and Pierret and Steffen (2018).

[3]The systemic risk score aggregates information about five broad categories of systemic importance: size, interconnectedness, substitutability, complexity, and cross-jurisdictional activity. In order not to favor any particular facet of systemic risk, the BCBS computes an equally weighted average score of these categories. For more details, see Basel Committee on Banking Supervision (2014).

in the academic literature.[4] The most prominent examples are the Marginal Expected Shortfall (MES) and the Systemic Expected Shortfall (SES) of Acharya et al. (2017), the Systemic Risk Measure (SRISK) of Acharya et al. (2012) and Brownlees and Engle (2017), and the Delta Conditional Value-at-Risk (ΔCoVaR) of Adrian and Brunnermeier (2016).[5] The MES of a financial firm is defined as its short-run expected equity loss conditional on the market taking a loss greater than its value-at-risk (VaR). It constitutes one of the key elements, together with the leverage and the firm market value, of the SRISK and the SES. These two systemic risk measures represent the expected amount a bank is undercapitalized in a future systemic event in which the overall financial system is undercapitalized. Finally, the CoVaR is defined as the VaR of the financial system conditional on institutions being under distress. Then, the ΔCoVaR corresponds to the difference between the CoVaR and the unconditional VaR of the financial system, and captures the marginal contribution of a particular institution to the overall systemic risk. Thus, all these global measures are designed to summarize the systemic risk contribution of a given financial institution into a single figure, and could be used to rank the financial institutions according to their systemic importance. However, to the best of our knowledge, no backtesting procedure has been proposed to evaluate their ex-post validity. This issue is of crucial importance as the validation is a key requirement for any systemic risk measure to become an industry standard. The current proposal by the BCBS strongly encourages backtesting risk measures.

We propose here a general framework for backtesting the MES, and by extension the related systemic risk measures especially the SES, SRISK, and ΔCoVaR. As defined by Jorion (2007), the backtesting is a formal statistical framework that consists in verifying if actual losses are in line with projected losses. This involves a systemic comparison of the historical model-generated risk measure forecasts with actual losses. Since the true value of the risk measure is unobservable, this comparison generally relies on violations. For instance, in the case of VaR, a violation is said to occur when the ex-post portfolio return is lower than the VaR forecast. The goal here consists in defining an appropriate concept of violation for the MES forecasts and for the other systemic risk measures. Then, it is possible to adopt the same tests as those currently used for backtesting other market and credit risk measures, such as VaR (see Christoffersen, 2010, for a survey) or expected shortfall (ES). We proceed as follows. First, we introduce a concept of conditional-VaR, inspired by the systemic risk measure proposed by Adrian and Brunnermeier (2016). A (β, α)-CoVaR is defined as the β-quantile of the truncated distribution of the firm

[4] Bisias et al. (2012) surveyed 31 quantitative measures, whereas Giglio et al. (2016) propose systemic risk indexes computed from 19 alternative systemic risk measures.

[5] These articles are among the most influential in the academic literature on systemic risk. For instance, the two articles of Acharya et al. (2012) and Brownlees and Engle (2017) that defined the MES and the SRISK, have been cited 1,000 times since their publication (source: Google Scholar). For online computation of some of these systemic risk measures, see the Stern-NYU V-Lab initiative website.

returns given that the financial system takes a loss greater than its α-VaR. We express the MES as an integral of the CoVaRs for all coverage rate β between 0 and 1. To the best of our knowledge, it is the first time that a relationship is established between the CoVaR and the MES, and extended then to the SES and SRISK. Second, we define a concept of joint violation of the (β, α)-CoVaR of the firm returns and α-VaR of the market returns. This extends the concept of cumulative violation recently proposed by Du and Escanciano (2017) for the ES backtests, to a bivariate case. We define a cumulative joint violation process defined as the integral of the joint violation processes for all coverage rate β between 0 and 1. We show that if the risk model used to forecast the MES is well specified, this cumulative violation process is a martingale difference sequence (mds). Exploiting this mds property, we propose two backtests for the MES: an unconditional coverage (UC, thereafter) test and an independence (IND, thereafter) test, as those generally considered for the VaR (Christoffersen, 1998). The UC test refers to the fact that the violations frequency should be in line with the theoretical probability to observe a violation. Failure of UC means that the MES forecast does not measure the risk accurately. Besides UC, the violations should satisfy the independence property, which implies that past violations should not be informative about current and future violations, if the dynamics of the risk model used to forecast the MES is well-specified.

Our backtesting procedure has many advantages. It allows backtesting either conditional (with respect to the past information set) MES or unconditional MES. Furthermore, we derive the asymptotic distribution of our test statistics while taking into account the estimation risk (Escanciano and Olmo, 2010). Indeed, the MES forecasts are generally issued from a parametric risk model for which the parameters have to be estimated. Then, the use of standard backtesting procedures to assess the MES forecasts in an out-of-sample framework can be misleading, because these procedures do not consider the impact of estimation risk. That is why, we also propose a robust version of our test statistics. Monte Carlo simulations show that these robust test statistics have good finite sample properties for realistic sample sizes.

Another advantage of our backtesting procedure is its applicability for any MES-based systemic risk measures, and typically for the SES and SRISK. As we can express these measures as a linear deterministic function of the (long run) MES, we show that testing their validity is equivalent to testing the validity of the MES forecast itself. This result is due to the assumptions made on the other constituents of the SES and SRISK, i.e., a constant level of liabilities and a constant initial market value of the financial institution. We can further extend our test to testing the validity of ΔCoVaR forecasts. To do so, we propose a simple approach based on a vector of two joint violations associated to two conditional VaRs for the financial system: one for the situation in which an institution is in distress and a second one with respect to a median state of the institution. The

intuition is similar here to the backtests proposed for a multi-level VaR, i.e., VaR defined for a finite set of coverage rates.

We apply our backtesting tests to assess the empirical validity of the MES, SRISK, and ΔCoVaR issued from a GARCH-DCC model for a panel of large U.S. financial institutions over the period January 3, 2000 to December 30, 2016. First, we observe that the one-day-ahead forecasts of these systemic risk measures are generally misspecified in periods of financial instability. The UC tests reject the validity of the risk forecasts for most of the financial institutions in our panel during the 2008-2009 global financial crisis. Second, when considering a longer forecasting horizon, say one month, the periods of significant rejection of the UC hypothesis for MES or SRISK are no longer there. Results suggest that for longer horizons, the GARCH-DCC model is able to provide much more trustworthy forecasts of systemic risk. Finally, we apply the UC test to the ΔCoVAR daily forecasts. The results are fully in line with those observed for the short-term MES forecasts. The UC hypothesis is rejected for most of the financial institutions during the global 2008-2009 financial crisis, and to a lesser extent during the 2011-2012 European debt crisis. In addition, our tests show that the stressed CoVaR is generally more affected by model misspecification than the median CoVaR.

Some attempts of validation procedures for the systemic risk measures have been proposed in the literature. Following the coherent risk approach of Artzner et al. (1999), Chen et al. (2013) define an axiomatic framework for systemic risk measures.[6] However, most of the validation techniques are empirical. They generally consist in testing whether firms with high systemic risk scores are more likely to become insolvent (Wu and Zhao, 2018), or to suffer the highest financial losses (Idier et al., 2014) in a financial crisis, etc. Brownlees and Engle (2017) show that banks with higher SRISK before the financial crisis were more likely to be bailed out by the government and to receive capital injections from the Federal Reserve. Engle et al. (2015) compare the ranking of European financial institutions obtained with the SRISK, to the list of SIFIs produced by the FSB. Recently, Brownlees et al. (2018) propose an historical assessment of the SRISK and ΔCoVaR based on two dimensions. The first one, called SIFI ranking challenge, consists in investigating whether ranking financial institutions by SRISK and ΔCoVaR allows identifying the institutions with notable deposit declines around panic events. The second one, called the financial crisis prediction challenge, investigates whether these systemic risk measures are significant predictors of system-wide deposit declines during panic events. Based on an original historical dataset for the New York banking system between 1866 and 1933,

[6] A systemic risk measure must satisfy the main conditions that define any coherent risk measure, namely the monotonicity, positive homogeneity, and outcome convexity axioms. However, it must also satisfy an additional preference consistency axiom. This axiom states that the risk measure has to reflect the preference of the regulator on the cross-sectional profile of losses across firms and the distribution of the aggregate outcomes across states.

their results show that both measures identify well the SIFIs, especially in periods of distress. However, the SRISK and ΔCoVaR exhibit poor performances as early warning signal of distress in the financial system as a whole.

Our validation approach is different as it relies on backtesting tests in the same spirit of those used for other market or credit risk measures (Jorion, 2007). We propose test statistics which are similar to those generally used by the regulator or the risk manager to backtest the VaR (Kupiec, 1995; Christoffersen, 1998) or the ES (Du and Escanciano, 2017; Kratz et al., 2018). Even if the validation approaches are different, they show some similarities in their main lines. As in Brownlees et al. (2018), albeit with a different empirical approach, we propose an Early Warning System (EWS) indicator which aims at identifying the periods of forecast breakdown that depict significant changes in market conditions. Our indicator is defined as the difference between the original MES forecast issued from the risk model, and an adjusted MES forecast. The latter represents the forecast that meets the null hypothesis of unconditional coverage. The adjustment of risk forecasts issued from a misspecified or mis-estimated risk model has already been considered in the risk management literature. For instance, Gouriéroux and Zakoïan (2013) propose a method to adjust the VaR forecasts affected by estimation risk. Boucher et al. (2014) empirically adjust imperfect VaR forecasts by outcomes from backtesting frameworks, considering desirable qualities of VaR models such as the frequency, independence and magnitude of violations. A similar approach has been recently applied to adjust imperfect ES forecasts by Couperier and Leymarie (2019) and Lazar and Zhang (2019). In the same spirit, we propose to adjust imperfect MES-based forecasts considering the frequency property of the cumulative joint violation process. Formally, it consists in determining an adjusted coverage level for the MES such that the null hypothesis of the UC test is not rejected. Our empirical results show that the adjusted MES diverges from the unadjusted MES at the beginning of the 2007-2009 financial crisis, and to a lesser extent, during the European debt crisis. Such a divergence should warn regulators about the severity of the global financial crisis, since the risk models initially applied by the financial institutions are no longer appropriate to deliver valid systemic risk forecasts.

We organize this chapter as follows. In Section 4.2, we define the MES, and we introduce the concept of cumulative joint violation process. We present the backtesting tests for MES in Section 4.3 and we illustrate their finite sample properties via Monte Carlo simulations. Section 4.4 extends our backtests to any MES-based systemic risk measure, and especially to the SES, the SRISK and to the ΔCoVaR. An empirical application is proposed in Section 4.5. Section 4.6 illustrates how to use these backtests as early warning systems to detect financial crisis episodes. Finally, we conclude this chapter in Section 4.7.

4.2 MES and cumulative joint violation process

In this section, we define the MES and introduce an associated concept of cumulative joint violation. We consider the following notations. Let $Y_t = (Y_{1t}, Y_{2t})'$ denotes the vector of stock returns of two assets at time t. In the specific context of systemic risk, Y_{1t} corresponds to the stock return of a financial institution, whereas Y_{2t} corresponds to the market return. Denote by Ω_{t-1} the information set available at time $t-1$, with $(Y_{t-1}, Y_{t-2}, ...) \subseteq \Omega_{t-1}$ and $F(.;\Omega_{t-1})$ the joint cumulative distribution function (cdf) of Y_t given Ω_{t-1}, such that $F(y_t;\Omega_{t-1}) \equiv \Pr(Y_{1t} < y_1, Y_{2t} < y_2 | \Omega_{t-1})$ for any $y = (y_1, y_2)' \in \mathbb{R}^2$. Assume that $F(.;\Omega_{t-1})$ is continuous.

4.2.1 Marginal expected shortfall

Following Acharya et al. (2012), we define the MES of a financial firm as its short-run expected equity loss conditional on the market taking a loss greater than its VaR.[7] Formally, the α-level MES of the financial institution at time t given Ω_{t-1} is defined as

$$MES_{1t}(\alpha) = \mathbb{E}(Y_{1t} | Y_{2t} \leq VaR_{2t}(\alpha); \Omega_{t-1}), \tag{4.1}$$

where $VaR_{2t}(\alpha)$ is the α-level VaR of the marginal distribution of Y_{2t}, denoted $F_{Y_2}(.;\Omega_{t-1})$, with $VaR_{2t}(\alpha) = F^{-1}_{Y_2}(\alpha;\Omega_{t-1})$, and $\alpha \in [0,1]$.[8] If we define the market return Y_{2t} as the value-weighted average of the firm returns (for all the firms that belong to the financial system), then the MES of one firm corresponds to the derivative of the market ES with respect to the firm market share (Scaillet, 2004; Acharya et al., 2017), hence the term "marginal". Thus, MES measures how the financial institution adds to the financial system overall risk.

From Equation (4.1), it follows that MES is a conditional expectation, and as such, we can express the risk measure as a function of the quantiles of the conditional distribution of Y_{1t} given $Y_{2t} \leq VaR_{2t}(\alpha)$. For that purpose, we introduce a concept of Conditional-VaR (CoVaR) inspired from the systemic risk measure proposed by Adrian and Brunnermeier (2016). For any coverage level $\beta \in [0,1]$, the CoVaR for the firm 1 at time t is the quantity $CoVaR_{1t}(\beta,\alpha)$ such that

$$\Pr(Y_{1t} \leq CoVaR_{1t}(\beta,\alpha) | Y_{2t} \leq VaR_{2t}(\alpha); \Omega_{t-1}) = \beta. \tag{4.2}$$

There are two main differences between $CoVaR_{1t}(\beta,\alpha)$ and the CoVaR introduced by Adrian and Brunnermeier (2016). First, the conditioning event is based on an inequal-

[7] The definition of the MES was extended to a Ω_{t-1}-conditional version by Brownlees and Engle (2017).

[8] For ease of notations, we do not use the usual convention that defines the VaR as the opposite of the α-quantile of the returns distribution.

ity, i.e., $Y_{2t} \leq VaR_{2t}(\alpha)$ as in Girardi and Ergün (2013), rather than on the equality $Y_{2t} = VaR_{2t}(\alpha)$. Second, we introduce a distinction between the coverage level β of the CoVaR and the coverage level α of the VaR, which is used to define the conditioning event. We also define the (β, α)-level CoVaR as $CoVaR_{1t}(\beta, \alpha) = F^{-1}_{Y_1|Y_2 \leq VaR_{2t}(\alpha)}(\beta; \Omega_{t-1})$ where $F_{Y_1|Y_2 \leq VaR_{2t}(\alpha)}(.; \Omega_{t-1})$ is the cdf of the conditional distribution of Y_{1t} given $Y_{2t} \leq VaR_{2t}(\alpha)$ and Ω_{t-1}. Definition of a conditional probability and a change of variables yield a meaningful representation of MES in terms of CoVaR.

$$MES_{1t}(\alpha) = \int_0^1 CoVaR_{1t}(\beta, \alpha)\mathrm{d}\beta. \tag{4.3}$$

The above Equation (4.3) is key for our backtesting approach. It gives a simple relationship between two risk measures, i.e., the MES and the CoVaR. To the best of our knowledge, this is the first attempt to establish a link between these systemic measures that are both broadly used in the systemic risk literature (see Acharya et al., 2017, for a comparison). We can use this link either for the Ω_{t-1}-conditional $MES_{1t}(\alpha)$, as in Brownlees and Engle (2017), or for the unconditional $MES_1(\alpha)$, as in Acharya et al. (2012). Furthermore, this definition of the MES is valid for any bivariate distribution.

For some particular distributions, the conditional cdf $F_{Y_1|Y_2 \leq VaR_{2t}(\alpha)}(.; \Omega_{t-1})$ that defines the CoVaR has a closed-form expression. For instance, Arnold et al. (1993) calculate the marginal of a bivariate normal distribution with double truncation over one variable. Horrace (2005) formalizes analytical results on the truncated multivariate normal distribution, where the truncation is one-sided and at an arbitrary point. Ho et al. (2012) focus on the truncated multivariate t-distribution. Whatever the distribution considered, it is also possible to express the cdf of the truncated distribution of Y_{1t} given $Y_{2t} \leq VaR_{2t}(\alpha)$ as a simple function of the cdf of the joint distribution of Y_t, with $F_{Y_1|Y_2 \leq VaR_{2t}(\alpha)}(y_1; \Omega_{t-1}) = \frac{1}{\alpha} F(\widetilde{y}; \Omega_{t-1})$, where the vector $\widetilde{y}$ is defined as $\widetilde{y} = (y_1, VaR_{2t}(\alpha))'$.

In general, the MES forecasts are issued from a parametric model specified by the researcher, the risk manager, or the regulation authority. For instance, Brownlees and Engle (2017) and Acharya et al. (2012) consider a bivariate dynamic conditional correlation (DCC) model to compute the MES and the SRISK. In practice, the cdf $F(.; \Omega_{t-1}, \theta_0)$ of the joint distribution of Y_t, the cdf $F_{Y_2}(.; \Omega_{t-1}, \theta_0)$ of the marginal distribution of Y_{2t} and the cdf $F_{Y_1|Y_2 \leq VaR_{2t}(\alpha;\theta_0)}(.; \Omega_{t-1}, \theta_0)$ of the truncated distribution of Y_{1t} given $Y_{2t} \leq VaR_{2t}(\alpha, \theta_0)$ depend on $\theta_0 \in \Theta \subseteq \mathbb{R}^p$, a vector of unknown parameters. It is therefore necessary to estimate these parameters before delivering the MES forecasts.

4.2.2 Cumulative joint violation process

In order to backtest the CoVaR and the MES, we define a joint violation of the (β, α)-CoVaR of Y_{1t} and the α-VaR of Y_{2t} at time t. We represent this violation process by the following binary variable

$$h_t(\alpha, \beta, \theta_0) = \mathbb{1}\left(\left(Y_{1t} \leq CoVaR_{1t}\left(\beta, \alpha, \theta_0\right)\right) \cap \left(Y_{2t} \leq VaR_{2t}\left(\alpha, \theta_0\right)\right)\right), \tag{4.4}$$

where $\mathbb{1}(.)$ denotes an indicator function. The violation (4.4) takes the value one if the loss of the firm exceeds its CoVaR and the loss of the market exceeds its VaR, and zero otherwise.

The VaR backtesting tests (Kupiec, 1995; Christoffersen, 1998; Berkowitz et al., 2011, among others) generally exploit the mds property of the violation process (see Christoffersen, 2010, for a survey). Here, we adopt a similar approach to backtest the CoVaR, and the MES. Notice that Bayes theorem implies that $\Pr\left(h_t(\alpha, \beta, \theta_0) = 1 | \, \Omega_{t-1}\right) = \alpha\beta$. Then, it follows from Equation (4.2) that the violations are Bernouilli distributed with mean $\alpha\beta$, and that the centered violation $\{h_t(\alpha, \beta, \theta_0) - \alpha\beta\}_{t=1}^{\infty}$ is a mds for risk levels $(\alpha, \beta) \in [0, 1]^2$. Formally, if the (β, α)-CoVaR of Y_{1t} and the α-VaR of Y_{2t} are correctly specified, we have

$$\mathbb{E}\left(h_t(\alpha, \beta, \theta_0) - \alpha\beta | \, \Omega_{t-1}\right) = 0.$$

In order to test for the validity of the MES, we consider a cumulative joint violation process which we can view as a kind of violations "counterpart" relying on the MES definition in Equation (4.3). We define this cumulative joint violation process as the integral of the violations $h_t(\alpha, \beta, \theta_0)$ for all the risk levels β between 0 and 1, with

$$H_t\left(\alpha, \theta_0\right) = \int_0^1 h_t(\alpha, \beta, \theta_0) \mathrm{d}\beta.$$

We can view this cumulative joint violation process as an extension to the bivariate case of the cumulative violation process recently introduced by Du and Escanciano (2017) to backtest the ES. The random variable $H_t\left(\alpha, \theta_0\right)$ has a Ω_{t-1}-conditional distribution defined as the product of a Bernouilli(α) random variable times a continuous uniform distribution over $[0, 1]$. To ease its numerical computation, we derive a closed-form solution for $H_t(\alpha, \theta_0)$ based on two generalized errors in Appendix 4.8.2. Du and Escanciano (2017) derive a closed-form solution for a single generalized error. The mean and variance of $H_t(\alpha, \theta_0)$ are equal to $\alpha/2$ and $\alpha\left(1/3 - \alpha/4\right)$ (see Appendix 4.8.2 for details). Furthermore, the Fubini theorem implies that the mds property of the sequence $\{h_t(\alpha, \beta, \theta_0) - \alpha\beta\}_{t=1}^{\infty}, \, \forall\left(\alpha, \beta\right) \in [0, 1]^2$ is preserved by integration. As a consequence,

the sequence $\{H_t(\alpha, \theta_0) - \alpha/2\}_{t=1}^{\infty}$ is also a mds for any $\alpha \in [0, 1]$, with

$$\mathbb{E}\left(H_t\left(\alpha, \theta_0\right) - \alpha/2 | \, \Omega_{t-1}\right) = 0.$$

4.3 Backtesting MES

Exploiting the mds property of the cumulative joint violation process enables the implementation of various types of backtests for the MES (see Nieto and Ruiz, 2016, for a survey). Here, we propose two tests which are similar to those generally used by regulators or risk managers for VaR backtesting (Christoffersen, 1998). The unconditional coverage (hereafter, UC) test relies on the null hypothesis

$$\mathrm{H}_{0,UC} : \mathbb{E}\left(H_t\left(\alpha, \theta_0\right)\right) = \alpha/2.$$

Since $\mathbb{E}\left(H_t\left(\alpha, \theta_0\right)\right) = \int_0^1 \mathbb{E}\left(h_t(\alpha; \beta, \theta_0)\right) \mathrm{d}\beta$, the null $\mathrm{H}_{0,UC}$ is equivalent to

$$\mathrm{H}_{0,UC} : \Pr\left(h_t(\alpha, \beta, \theta_0) = 1\right) = \alpha\beta, \quad \forall \beta \in [0, 1].$$

The null hypothesis UC means that for any risk levels β, the joint probability to observe an ex-post return Y_{1t} exceeding its (β, α)-CoVaR and an ex-post return Y_{2t} exceeding its α-VaR must be equal to $\alpha\beta$. Note that in the case $\mathrm{H}_{0,UC}$ is violated, there are two scenarios. Case 1 corresponds to $\mathbb{E}\left(H_t\left(\alpha, \theta_0\right)\right) > \alpha/2$ and indicates that the number of joint exceedances is higher than expected. This scenario is associated to an underestimation of the MES. In the opposite case 2, $\mathbb{E}\left(H_t\left(\alpha, \theta_0\right)\right) < \alpha/2$ reveals that not enough exceedances are experienced in average, and induces an overestimation of MES.

The second backtest is based on the independence (hereafter, IND) property of the cumulative violation process $\{H_t\left(\alpha, \theta_0\right) - \alpha/2\}_{t=1}^{\infty}$: the cumulative violations observed at two different dates for the same coverage rate α must be distributed independently. Consistently with Christoffersen (1998), we propose a simple Box-Pierce test (Box and Pierce, 1970) in order to test for the nullity of the first K autocorrelations of $H_t\left(\alpha, \theta_0\right)$, denoted ρ_k. We define the null hypothesis of the IND test as

$$\mathrm{H}_{0,IND} : \rho_1 = ... = \rho_K = 0,$$

with $\rho_k = corr\left(H_t\left(\alpha, \theta_0\right), H_{t-k}\left(\alpha, \theta_0\right)\right)$. In theory, a rejection of $\mathrm{H}_{0,IND}$ reveals misspecification of MES coming from its time-varying dynamics, and thus should complete well with UC to detect an incorrect systemic risk model.

Implementing the tests requires to estimate the parameters $\theta_0 \in \Theta$ of the model used by the risk manager to forecast the MES. For simplicity, we adopt a fixed forecasting estimation scheme. We use an in-sample period from $t = 1$ to $t = T$ to estimate θ_0.

Denote by Ω_T the information set available at the end of the in-sample period, with $\{Y_1, ..., Y_T\} \subseteq \Omega_T$ and $\widehat{\theta}_T$ a consistent estimator of θ_0. Based on the ex-post returns observed from $t = T + 1$ to $t = T + n$, we compute the backtesting test statistics from the out-of-sample forecasts of the cumulative violation process given by

$$H_t(\alpha, \widehat{\theta}_T) = \left(1 - u_{12t}(\widehat{\theta}_T)\right) \mathbb{1}\left(u_{2t}(\widehat{\theta}_T) \leq \alpha\right), \quad \forall t = T + 1, ..., T + n,$$

with $u_{2t}(\widehat{\theta}_T) \equiv F_{Y_2}(Y_{2t}; \Omega_{t-1}, \widehat{\theta}_T)$ and $u_{12t}(\widehat{\theta}_T) \equiv F_{Y_1|Y_2 \leq VaR_{2t}(\alpha, \widehat{\theta}_T)}(Y_{1t}; \Omega_{t-1}, \widehat{\theta}_T)$. See Appendix 4.8.2 for more details.

4.3.1 Unconditional coverage test

By analogy with the backtest proposed by Du and Escanciano (2017) for ES and the well-known VaR backtest proposed by Kupiec (1995), we consider a standard t-test for the null hypothesis of unconditional coverage $\mathrm{H}_{0,UC}$ for the MES. We define the test statistic, denoted UC_{MES}, as

$$UC_{MES} = \frac{\sqrt{n}\left(\bar{H}(\alpha, \widehat{\theta}_T) - \alpha/2\right)}{\sqrt{\alpha\left(1/3 - \alpha/4\right)}}, \tag{4.5}$$

with $\bar{H}(\alpha, \widehat{\theta}_T)$ the out-of-sample mean of $H_t(\alpha, \widehat{\theta}_T)$

$$\bar{H}(\alpha, \widehat{\theta}_T) = \frac{1}{n} \sum_{t=T+1}^{T+n} H_t(\alpha, \widehat{\theta}_T).$$

In order to give the intuition of the asymptotic properties of the statistic UC_{MES}, let us define a similar statistic $UC_{MES}(\alpha, \theta_0)$ based on the true value of the parameters θ_0 instead of its estimator $\widehat{\theta}_T$. Under the null hypothesis, the sequence $\{H_t(\alpha, \theta_0) - \alpha/2\}_{t=T+1}^{T+n}$ is a mds with variance equal to $\alpha(1/3 - \alpha/4)$. As a consequence, the Lindeberg-Levy central limit theorem implies that $UC_{MES}(\alpha, \theta_0)$ has an asymptotic standard normal distribution. A similar result holds for the feasible statistic $UC_{MES} \equiv UC_{MES}(\alpha, \widehat{\theta}_T)$ when $T \to \infty$ and $n \to \infty$, whereas $\lambda = n/T \to 0$, i.e., when there is no estimation risk.

However, in the general case $T \to \infty$, $n \to \infty$ and $n/T \to \lambda < \infty$, and there is an estimation risk as soon as $\lambda \neq 0$. Escanciano and Olmo (2010) show that the use of standard backtesting procedures can be misleading if they do not take into account the uncertainty associated with parameter estimation. In presence of estimation error, the asymptotic distribution of UC_{MES} is not standard and depends on the ratio of the in-sample size T to the out-of-sample size n. Theorem 2 gives the corresponding asymptotic distribution of UC_{MES} when $T \to \infty$, $n \to \infty$ and $n/T \to \lambda$ with $0 < \lambda < \infty$.

Theorem 2. *Under Assumptions A1-A4 in the Appendix, we have:*

$$UC_{MES} \xrightarrow{d} \mathcal{N}\left(0, \sigma_\lambda^2\right),$$

where $\xrightarrow{d}$ *denotes the convergence in distribution and where the asymptotic variance* σ_λ^2 *is*

$$\sigma_\lambda^2 = 1 + \lambda \frac{R'_{MES} \Sigma_0 R_{MES}}{\alpha\left(1/3 - \alpha/4\right)},$$

with $R_{MES} = \mathbb{E}_0\left(\partial H_t\left(\alpha, \theta_0\right)/\partial\theta\right)$ *and* $\mathbb{V}_{as}(\widehat{\theta}_T) = \Sigma_0/T$.

The proof of Theorem 2 is reported in Appendix 4.8.3. The vector R_{MES} quantifies the parameter estimation effect on the test statistic UC_{MES} due to the difference between the estimate $\widehat{\theta}_T$ and the true value of the parameter θ_0. We can characterize the impact of the estimation risk on the UC_{MES} test statistic as follows

$$\begin{aligned} UC_{MES} &= \underbrace{\frac{1}{\sigma_H \sqrt{n}} \sum_{t=T+1}^{T+n} \left(H_t\left(\alpha, \theta_0\right) - \alpha/2\right)}_{UC_{MES}(\alpha, \theta_0)} \\ &\quad + \underbrace{\frac{\sqrt{\lambda}}{\sigma_H} \frac{1}{n} \sum_{t=T+1}^{T+n} \mathbb{E}\left(\left.\frac{\partial H_t\left(\alpha, \widetilde{\theta}\right)'}{\partial\theta}\right| \Omega_{t-1}\right) \sqrt{T}(\widehat{\theta}_T - \theta_0)}_{\text{Estimation risk}} + o_p\left(1\right). \end{aligned}$$

This formula is similar to that obtained by Du and Escanciano (2017) to backtest ES, but its constituents are different because the testing procedure now relies on a bivariate distribution. Whatever the dynamic model considered for the returns, we can deduce the vector R_{MES} from the cdf of the joint distribution Y_t given Ω_{t-1} with

$$R_{MES} = -\frac{1}{\alpha}\mathbb{E}_0\left(\frac{\partial F(\widetilde{y}_t; \Omega_{t-1}, \theta_0)}{\partial\theta} \mathbb{1}\left(u_{2t}\left(\theta_0\right) \leq \alpha\right)\right) + \mathbb{E}_0\left(\left(1 - u_{12t}\left(\theta_0\right)\right) \frac{\partial \mathbb{1}\left(u_{2t}\left(\theta_0\right) \leq \alpha\right)}{\partial\theta}\right),$$

where $\widetilde{y}_t = (y_{1t}, VaR_{2t}(\alpha, \theta_0))'$ and

$$\begin{aligned} \frac{\partial F(\widetilde{y}_t; \Omega_{t-1}, \theta_0)}{\partial\theta} &= \underbrace{\int_{-\infty}^{y_{1t}} f\left(u, VaR_{2t}\left(\alpha, \theta_0\right); \Omega_{t-1}, \theta_0\right) \mathrm{d}u \times \frac{\partial VaR_{2t}\left(\alpha, \theta_0\right)}{\partial\theta}}_{\text{Impact on the truncation}} \\ &\quad + \underbrace{\int_{-\infty}^{y_{1t}} \int_{-\infty}^{VaR_{2t}(\alpha, \theta_0)} \frac{\partial f\left(u, v; \Omega_{t-1}, \theta_0\right)}{\partial\theta} \mathrm{d}u\mathrm{d}v}_{\text{impact on the pdf of the joint distribution}}. \end{aligned}$$

In the bivariate case, the estimation error affects the cdf of the truncated distribution of Y_{1t} given $Y_{2t} \leq VaR_{2t}\left(\alpha, \theta_0\right)$ not only through its effect on the joint distribution, but

also through the truncation parameter, i.e., the VaR of the market return. There is no analytical expression for the derivative $\partial F(\widetilde{y}_t; \Omega_{t-1}, \theta_0)/\partial\theta$ except for some particular bivariate distributions (see Appendix 4.8.4 for the case of a bivariate normal distribution). In the general case, we have to evaluate that expression via numerical differentiation.

Corollary 3. *When there is no estimation risk, i.e., when* $\lambda = 0$, *under Assumptions A1-A4 in the Appendix, we have:*

$$UC_{MES} \xrightarrow{d} \mathcal{N}(0,1).$$

When the estimation period T is much larger than the evaluation period n, the unconditional coverage test is simplified since it does not require to evaluate R_{MES} and Σ_0. Given these results, it is possible to define a test statistic UC^C_{MES}, which takes into account explicitly the estimation risk, while having a standard limit distribution for any λ with $0 \leq \lambda < \infty$, when T and n tend to infinity. The feasible robust UC backtest statistic is

$$UC^C_{MES} = \frac{\sqrt{n}\left(\bar{H}(\alpha, \widehat{\theta}_T) - \alpha/2\right)}{\sqrt{\alpha\left(1/3 - \alpha/4\right) + n\widehat{R}'_{MES}\widehat{\mathbb{V}}_{as}(\widehat{\theta}_T)\widehat{R}_{MES}}},$$

with $\widehat{\mathbb{V}}_{as}(\widehat{\theta}_T) = \widehat{\Sigma}_0/T$ a consistent estimator of the asymptotic variance-covariance matrix of $\widehat{\theta}_T$ and $\widehat{R}_{MES}$, a consistent estimator of R_{MES}. In Appendix 4.8.5, we propose an estimator for the vector R_{MES} that we can easily implement.

4.3.2 Independence test

To test the independence hypothesis $\mathrm{H}_{0,IND} : \rho_1 = ... = \rho_m = 0$, we use a Portmanteau Box-Pierce test applied to the sequence of cumulative joint violation forecasts. We define the Box-Pierce test statistic as follows

$$IND_{MES} = n\sum_{j=1}^{m}\widehat{\rho}^2_{nj},$$

with $\widehat{\rho}_{nj}$ the sample autocorrelation of order j of the estimated cumulative joint violation $H_t(\alpha, \widehat{\theta}_T)$ given by

$$\widehat{\rho}_{nj} = \frac{\widehat{\gamma}_{nj}}{\widehat{\gamma}_{n0}} \quad \text{and} \quad \widehat{\gamma}_{nj} = \frac{1}{n-j}\sum_{t=T+1+j}^{T+n}\left(H_t(\alpha, \widehat{\theta}_T) - \alpha/2\right)\left(H_{t-j}(\alpha, \widehat{\theta}_T) - \alpha/2\right),$$

where $\widehat{\gamma}_{nj}$ denotes a consistent estimator of the j-lag autocovariance of $H_t(\alpha, \widehat{\theta}_T)$. Theorem 4 gives the asymptotic distribution of the statistic IND_{MES} when $T \to \infty$, $n \to \infty$ and $n/T \to \lambda < \infty$.

Theorem 4. *Under Assumptions A1-A4 in the Appendix, we have:*

$$IND_{MES} \xrightarrow{d} \sum_{j=1}^{m} \pi_j Z_j^2,$$

where $\{\pi_j\}_{j=1}^m$ *are the eigenvalues of the matrix* Δ *with the* ij*-th element given by*

$$\Delta_{ij} = \delta_{ij} + \lambda R_i' \Sigma_0 R_j,$$

$$R_j = \frac{1}{\alpha\,(1/3 - \alpha/4)} \mathbb{E}_0 \left((H_{t-j}(\alpha, \theta_0) - \alpha/2) \frac{\partial H_t(\alpha, \theta_0)}{\partial \theta} \right),$$

where δ_{ij} *is a dummy variable that takes the value 1 if* $i = j$ *and 0 otherwise,* $\{Z_j\}_{j=1}^m$ *are independent standard normal variables, and* $\mathbb{V}_{as}(\widehat{\theta}_T) = \Sigma_0/T$*.*

We give the proof of Theorem 4 in Appendix 4.8.6. The test statistic IND_{MES} has an asymptotic distribution which is a weighted sum of chi-squared variables. The weights depend on the asymptotic variance-covariance matrix of the estimator $\widehat{\theta}_T$, on the cumulative joint violation process and on its derivative with respect to the model parameter θ, as for the UC test. However, this limit distribution becomes standard when $\lambda = 0$, i.e., when there is no estimation risk.

Corollary 5. *When there is no estimation risk, i.e., when* $\lambda = 0$*, under Assumptions A1-A4 in the Appendix, we have:*

$$IND_{MES} \xrightarrow{d} \chi^2(m).$$

From the previous results, we can deduce a robust test statistic for the independence hypothesis which has standard distribution for any λ with $0 \leq \lambda < \infty$, when T and n tend to infinity. Denote $\widehat{\rho}_n^{(m)}$ the vector $(\widehat{\rho}_{n1}, ..., \widehat{\rho}_{nm})'$. The feasible robust IND backtest statistic is defined as

$$IND_{MES}^C = n \widehat{\rho}_n^{(m)\prime} \widehat{\Delta}^{-1} \widehat{\rho}_n^{(m)},$$

where $\widehat{\Delta}$ is a consistent estimator for Δ, such that

$$\widehat{\Delta}_{ij} = \delta_{ij} + n \widehat{R}_i' \widehat{\mathbb{V}}_{as}(\widehat{\theta}_T) \widehat{R}_j,$$

with $\widehat{\mathbb{V}}_{as}(\widehat{\theta}_T)$ a consistent estimator of the asymptotic variance-covariance matrix of $\widehat{\theta}_T$, and $\widehat{R}_j$ a consistent estimator of R_j. In Appendix 4.8.5, we provide an easy to implement estimator of R_j. When T and n tend to infinity, the robust statistic IND_{MES}^C converges to a chi-squared distribution with m degrees of freedom whatever the relative value of n and T.

4.3.3 Monte carlo simulations

This section assesses the finite sample properties of our two backtest statistics computed with and without taking into account the estimation risk. We consider two definitions of the MES in order to study the properties of our tests in various settings. Firstly, we look at the particular case of a marginal (time-invariant) MES. This configuration allows to easily control for the degree of misspecification of the firm risk and the financial interdependencies, and thus to assess the power of our tests in relevant cases. Secondly, we consider a conditional (time-varying) MES based on a multivariate GARCH-DCC model, as in Brownlees and Engle (2017). In the sequel, we briefly present the two data generating processes (DGP) and the results of our Monte Carlo simulations.

Backtesting marginal MES. In order to evaluate the empirical size and power of our tests in the case of a marginal MES, we consider a bivariate normal distribution for the daily demeaned returns $Y_t = (Y_{1t}, Y_{2t})'$, such that $Y_t = \Sigma^{1/2} z_t$, with Y_{1t} the firm return, Y_{2t} the market return, and where z_t denotes an i.i.d. Gaussian vector error process with $\mathbb{E}(z_t) = 0$ and $\mathbb{E}(z_t z_t') = I_2$. Under the null, the variance-covariance matrix Σ is defined as

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

where σ_1 and σ_2 respectively denote the volatility of firm and market returns, and ρ is the correlation between both returns. We calibrate the parameters $\theta_0 = (\sigma_1^2, \sigma_2^2, \rho)'$ using the daily log-returns of Bank of America and CRSP market value-weighted index over the period January 1, 2012 to December 30, 2016. Under the normality assumption, the MES of the bank does not depend on the market volatility.[9]

For each Monte Carlo replication b, we simulate the series $\{y_{1t}^{(b)}, y_{2t}^{(b)}\}_{t=1}^{T+n}$ and we estimate the variance-covariance parameters θ_0 using the first T simulated observations $\{y_{1t}^{(b)}, y_{2t}^{(b)}\}_{t=1}^{T}$. Then, using the estimated parameters $\widehat{\theta}_T^{(b)}$, we compute the cumulative violation process $H_t(\alpha, \widehat{\theta}_T^{(b)})$ for the out-of-sample periods $t = T+1, \ldots, T+n$. For a finite sample size T, the violations are affected by the estimation risk due to the difference between the estimate $\widehat{\theta}_T^{(b)}$ and the true parameter value θ_0. Finally, using the series $\{H_t(\alpha, \widehat{\theta}_T^{(b)})\}_{t=T+1}^{T+n}$, we compute the test statistics UC_{MES} and IND_{MES}, without estimation risk correction, and the corresponding robust test statistics UC_{MES}^C and IND_{MES}^C.

[9] Under these assumptions, the MES for the bank return has a closed-form expression (Brownlees and Engle, 2017) given by $MES_1(\alpha) = -\rho\sigma_1\phi\left(\Phi^{-1}(\alpha)\right)/\alpha$, where $\phi(.)$ and $\Phi(.)$ denote the pdf and the cdf of a standard normal distribution. We can also express the MES as the product of the beta of the firm by the ES of the market return (Benoit et al., 2017) such as $MES_1(\alpha) = \beta_1 ES_2(\alpha)$, with $\beta_1 = \rho\sigma_1/\sigma_2$, and $ES_2(\alpha) = \mathbb{E}(Y_{2t} | Y_{2t} \leq VaR_2(\alpha))$.

The simulation study is based on 10,000 replications. In addition, we consider various in-sample sizes ($T = 250$, 500, and 2,500) and out-of-sample sizes ($n = 250$ and 500) to illustrate the impact of T and n on the estimation error and the small-sample properties of the backtests. Finally, we set the coverage rate α for the MES to 5% and we compute the empirical size as the rejection frequency at a nominal level of 5%.

Because the MES depends on risk levels (volatilities) and dependencies (correlation), we propose various experiments designed to assess the capabilities of our tests to detect misspecification on these two quantities. We consider three misspecified models given by $Y_t = \widetilde{\Sigma}^{1/2} z_t$ under the alternative hypothesis:

$$\widetilde{\Sigma} = \begin{pmatrix} \widetilde{\sigma}_1^2 & \widetilde{\rho}\widetilde{\sigma}_1\widetilde{\sigma}_2 \\ \widetilde{\rho}\widetilde{\sigma}_1\widetilde{\sigma}_2 & \widetilde{\sigma}_2^2 \end{pmatrix},$$

- $\mathrm{H}_1(A_1)$: Undervalued variance of firm returns, $\widetilde{\sigma}_1^2 = \tau\sigma_1^2$, with $\tau < 1$.
- $\mathrm{H}_1(A_2)$: Undervalued variance of market returns, $\widetilde{\sigma}_2^2 = \tau\sigma_2^2$, with $\tau < 1$.
- $\mathrm{H}_1(A_3)$: Undervalued correlation between firm and market returns, $\widetilde{\rho} = \tau\rho$, with $\tau < 1$.

Under these alternatives, the variance-covariance matrix is misspecified and underestimates the variance of return (A_1 and A_2) or the correlation (A_3) by a factor τ, equal to either 25%, 50%, or 75%. Such undervaluations of risk or dependence induce an underestimated MES and, in fine, an undervaluation of the systemic risk. In the case of alternatives, we use the first T simulated observations to estimate the variance-covariance parameters θ_0 under the constraint H_1.[10] We denote the vector of estimated and constrained parameters obtained in the b^{th} replication by $\widehat{\theta}_T^{(b)}$. We compute the test statistics from the misspecified cumulative violation process $H_t(\alpha, \widehat{\theta}_T^{(b)})$ for $t = T+1, ..., T+n$, and we compute the empirical power as the rejection frequency at 5% nominal level. In order to take into account the potential size distortions for small T, our reported powers are all size-corrected.

Table 4.1 displays the empirical sizes and powers for the UC_{MES}, UC_{MES}^{C}, IND_{MES}, and IND_{MES}^{C} tests. Four main results stand out. First, the empirical sizes of the UC and IND tests, with or without correction for estimation risk, converge to the

[10] For each alternative, the reduction factor τ is applied to the true value of the parameter, instead of its estimated counterpart, so as to avoid variations of the constrained parameters across replications. In order to ensure positive semi-definiteness of the estimate of $\widetilde{\Sigma}$ for any τ, the other parameters are estimated by using the constrained maximum likelihood. This framework allows illustrating the power of our tests whereas taking into account the estimation risk. Notice that the data generating process considered under $\mathrm{H}_1(A_2)$ leads to a misspecified MES: even if MES does not depend on the market volatility σ_2, the other estimated parameters (firm return volatility and correlation) do not converge to their true value.

Table 4.1: Empirical rejection rates for backtesting MES(5%) at 5% nominal level (marginal case)

		$T = 250$, $n = 250$				$T = 250$, $n = 500$			
		UC_{MES}	UC^C_{MES}	IND_{MES}	IND^C_{MES}	UC_{MES}	UC^C_{MES}	IND_{MES}	IND^C_{MES}
H_0		0.0809	0.0499	0.0895	0.0728	0.1199	0.0553	0.0795	0.0626
$H_1(A_1)$	$\tau = 25\%$	0.0909	0.0906	0.0479	0.0467	0.1037	0.1042	0.0446	0.0484
	$\tau = 50\%$	0.2010	0.1886	0.0413	0.0408	0.2460	0.2250	0.0482	0.0462
	$\tau = 75\%$	0.3824	0.3516	0.0386	0.0344	0.5021	0.4436	0.0460	0.0418
$H_1(A_2)$	$\tau = 25\%$	0.2129	0.2359	0.0364	0.0353	0.2794	0.3333	0.0417	0.0384
	$\tau = 50\%$	0.7161	0.7174	0.0437	0.0220	0.8862	0.8912	0.0808	0.0346
	$\tau = 75\%$	0.9877	0.9830	0.1848	0.0165	0.9993	0.9984	0.4400	0.0296
$H_1(A_3)$	$\tau = 25\%$	0.1351	0.1291	0.0407	0.0391	0.1533	0.1487	0.0455	0.0440
	$\tau = 50\%$	0.2676	0.2462	0.0365	0.0337	0.3497	0.3095	0.0460	0.0424
	$\tau = 75\%$	0.4032	0.3590	0.0368	0.0340	0.5218	0.4409	0.0456	0.0388
		$T = 500$, $n = 250$				$T = 500$, $n = 500$			
		UC_{MES}	UC^C_{MES}	IND_{MES}	IND^C_{MES}	UC_{MES}	UC^C_{MES}	IND_{MES}	IND^C_{MES}
H_0		0.0648	0.0473	0.0891	0.0774	0.0898	0.0509	0.0755	0.0652
$H_1(A_1)$	$\tau = 25\%$	0.1014	0.1006	0.0437	0.0450	0.1111	0.1113	0.0437	0.0441
	$\tau = 50\%$	0.2161	0.2079	0.0391	0.0389	0.2751	0.2617	0.0446	0.0397
	$\tau = 75\%$	0.4115	0.3904	0.0394	0.0358	0.5572	0.5146	0.0452	0.0393
$H_1(A_2)$	$\tau = 25\%$	0.2372	0.2526	0.0327	0.0352	0.3481	0.3846	0.0460	0.0464
	$\tau = 50\%$	0.7524	0.7556	0.0438	0.0319	0.9365	0.9411	0.0826	0.0550
	$\tau = 75\%$	0.9942	0.9932	0.1774	0.0497	0.9998	0.9998	0.4554	0.1408
$H_1(A_3)$	$\tau = 25\%$	0.1379	0.1355	0.0375	0.0383	0.1814	0.1740	0.0393	0.0393
	$\tau = 50\%$	0.2813	0.2648	0.0347	0.0311	0.3905	0.3637	0.0420	0.0373
	$\tau = 75\%$	0.4231	0.3963	0.0367	0.0314	0.5765	0.5310	0.0447	0.0399
		$T = 2500$, $n = 250$				$T = 2500$, $n = 500$			
		UC_{MES}	UC^C_{MES}	IND_{MES}	IND^C_{MES}	UC_{MES}	UC^C_{MES}	IND_{MES}	IND^C_{MES}
H_0		0.0540	0.0503	0.0883	0.0851	0.0581	0.0498	0.0773	0.0724
$H_1(A_1)$	$\tau = 25\%$	0.0971	0.0971	0.0420	0.0402	0.1204	0.1189	0.0446	0.0447
	$\tau = 50\%$	0.2169	0.2144	0.0422	0.0389	0.3122	0.3079	0.0417	0.0406
	$\tau = 75\%$	0.4287	0.4239	0.0373	0.0325	0.6130	0.6018	0.0432	0.0402
$H_1(A_2)$	$\tau = 25\%$	0.2515	0.2544	0.0344	0.0346	0.4008	0.4125	0.0411	0.0443
	$\tau = 50\%$	0.7961	0.7968	0.0444	0.0414	0.9656	0.9669	0.0796	0.0775
	$\tau = 75\%$	0.9974	0.9973	0.1666	0.1336	1.0000	1.0000	0.4319	0.3710
$H_1(A_3)$	$\tau = 25\%$	0.1421	0.1409	0.0416	0.0402	0.1893	0.1875	0.0417	0.0427
	$\tau = 50\%$	0.2874	0.2824	0.0354	0.0324	0.4357	0.4251	0.0439	0.0430
	$\tau = 75\%$	0.4331	0.4281	0.0329	0.0301	0.6487	0.6367	0.0423	0.0401

Note: This table displays the Monte Carlo results associated to the marginal MES setting. UC_{MES} and IND_{MES} denote the unconditional coverage test and independence test, respectively. The test statistics robust to estimation risk are superscripted by C. For the independence test, we consider a maximum lag order $m = 5$. Reported powers are sized-corrected.

nominal level when both T and n increase. However, for small in-sample sizes T, the UC_{MES} test exhibits severe size distortions due to estimation errors. For instance, for $T = 250$ and $n = 500$, the UC_{MES} test is largely oversized as its empirical size (11.99%)

is twice the nominal one. These size distortions increase with the out-of-sample size n, as the test statistic UC_{MES} diverges when T is fixed and small, and n tends to infinity. Second, for small T samples, we recommend the use of the robust test statistics that properly control for estimation risk. Empirical sizes of the robust test statistic UC^C_{MES} are close to the nominal size of 5% for all reported samples. For instance, for $T = 250$ and $n = 500$, its empirical size is 5.53%. However, the robust IND^C_{MES} backtest exhibits a slower rate of convergence and displays slight size distortions in small samples.[11] Third, our tests demonstrate good capacity for detecting various misspecifications in the variance-covariance matrix used to compute the MES. The empirical power of UC increases with the misspecification factor τ and the out-of-sample size n. The simulations for very large n (not reported) confirm that our UC tests are consistent, as the rejection frequencies tend to 1. Interestingly, we obtain the highest empirical powers for the market volatility misspecification (A_2), indicating that a poor assessment of the market distress may have severe consequences on the MES. Finally, for all alternatives, the empirical power of the IND test is still very low whatever n and T. This result is consistent with the theory, as the underestimation of volatilities and correlation of returns does not have any consequences on the autocorrelations of the cumulative violation process H_t. As a consequence, the IND backtest is non-sensitive to the alternatives A_1, A_2, and A_3.

Backtesting conditional MES. For the conditional case, we consider a dynamic conditional correlation (DCC) model as in Brownlees and Engle (2017). This model is widely used in the context of systemic risk since it is able to well reproduce most of the stylized facts on financial data, and to capture the interdependencies observed between firm and market returns. Formally, we assume that the vector process of demeaned returns is now defined as $Y_t = \Sigma_t^{1/2} z_t$, where Σ_t denotes the conditional variance-covariance matrix and z_t is as defined previously. Under the null, we define the conditional variance-covariance matrix Σ_t as follows:

$$\Sigma_t = D_t R_t D_t,$$

where $D_t = diag\{\sigma_{1t}; \sigma_{2t}\}$ is a diagonal matrix which contains conditional volatilities, and R_t is the conditional correlation matrix of Y_t. We consider a GJR-GARCH specification for the conditional variances of the firm and market returns (Glosten et al., 1993; Rabemananjara and Zakoian, 1993),

$$\sigma^2_{it} = \alpha_{i0} + \alpha_{i1} Y^2_{i,t-1} + \alpha_{i2} Y^2_{it-1} \mathbb{1}\left(Y_{i,t-1} < 0\right) + \alpha_{i3} \sigma^2_{it-1} \quad \forall i = 1, 2.$$

[11] Using an independence test for backtesting expected shortfall, Du and Escanciano (2017) obtain similar size distorsions for small sample sizes.

The DCC specification imposes a time-varying correlation structure on the standardized returns $\epsilon_{1t} = Y_{1t}/\sigma_{1,t}$ and $\epsilon_{2t} = Y_{2t}/\sigma_{2,t}$ through the pseudo-correlation matrix Q_t,

$$Q_t = (1 - a - b)\overline{Q} + a\epsilon'_{t-1}\epsilon_{t-1} + bQ_{t-1},$$

where $\overline{Q}$ is the unconditional correlation matrix, and a, b are two non-negative parameters such that $a+b < 1$. The conditional correlation matrix is obtained by rescaling Q_t, such as $R_t = (diag\ Q_t)^{-1/2} Q_t (diag\ Q_t)^{-1/2}$. In the sequel, we refer to this conditional specification as GARCH-DCC model. We calibrate the parameters $\theta_0 = (\alpha_{i0}, \alpha_{i1}, \alpha_{i2}, \alpha_{i3}, a, b)$ at their maximum likelihood parameter estimates, obtained using the same dataset as for the marginal MES. The Monte Carlo study design is similar to that previously described, except that we consider now $1,000$ replications. We consider four misspecified models given by $Y_t = \widetilde{\Sigma}_t^{1/2} z_t$ under the alternative hypothesis:

- $\mathrm{H}_1(B_1)$: Undervalued conditional variance of firm returns, $\widetilde{\sigma}_{1t}^2 = \tau\sigma_{1t}^2$, with $\tau < 1$.
- $\mathrm{H}_1(B_2)$: Undervalued conditional variance of market returns, $\widetilde{\sigma}_{2t}^2 = \tau\sigma_{2t}^2$, with $\tau < 1$.
- $\mathrm{H}_1(B_3)$: Undervalued conditional correlation, $\widetilde{\rho}_t = \tau\rho_t$, with $\tau < 1$.
- $\mathrm{H}_1(B_4)$: Misspecification of the dynamics of firm and market returns, with $Y_t = \Sigma^{1/2} z_t$.

Under the alternatives B_1-B_3, the GARCH-DCC model is misspecified in the sense that the conditional variances or the conditional correlation are misleading by a factor τ, set to either $25\%, 50\%$, or 75%. In contrast, the alternative B_4 corresponds to a misspecification of the dynamics of returns: the cumulative violation process is computed using an estimated variance-covariance matrix which is assumed to be constant over time, whereas the true one is time-varying and exhibits a leverage effect. We expect that this setting creates autocorrelated violations, which should be detected by our independence test.

Table 4.2 provides the empirical size and size-corrected power at 5% nominal level of the conditional MES backtests. Our results are similar to those obtained for the unconditional setting and the key takeaways are the following. First, we observe size distortions for the UC backtests when there is no correction for the estimation risk and the estimation sample size T is small. For a fixed size T, these distortions increase with the out-of-sample size n. We also observe that the impact of estimation risk is amplified compared to the unconditional case, due to the increase of the number of estimated parameters (11 in the case of the GARCH-DCC model). As expected, the estimated parameters converge to the true parameter values and the impact of estimation error vanishes asymptotically. Second, our correction for estimation risk is efficient as it leads

Table 4.2: Empirical rejection rates for backtesting MES(5%) at 5% nominal level (conditional case)

		$T = 250$, $n = 250$				$T = 250$, $n = 500$			
		UC_{MES}	UC^C_{MES}	IND_{MES}	IND^C_{MES}	UC_{MES}	UC^C_{MES}	IND_{MES}	IND^C_{MES}
H_0		0.1080	0.0460	0.1000	0.0590	0.1500	0.0400	0.1100	0.0530
$H_1(B_1)$	$\tau = 25\%$	0.0780	0.0670	0.0400	0.0470	0.0820	0.0800	0.0400	0.0550
	$\tau = 50\%$	0.1550	0.1540	0.0540	0.0480	0.1970	0.1980	0.0610	0.0630
	$\tau = 75\%$	0.3090	0.2750	0.0430	0.0440	0.4470	0.3630	0.0480	0.0600
$H_1(B_2)$	$\tau = 25\%$	0.1520	0.1850	0.0320	0.0250	0.2210	0.2730	0.0270	0.0310
	$\tau = 50\%$	0.6490	0.6290	0.0310	0.0110	0.8500	0.7920	0.0530	0.0120
	$\tau = 75\%$	0.9750	0.9100	0.1710	0.0020	0.9970	0.9540	0.3610	0.0040
$H_1(B_3)$	$\tau = 25\%$	0.0970	0.1050	0.0410	0.0420	0.1180	0.1320	0.0390	0.0470
	$\tau = 50\%$	0.2240	0.2060	0.0450	0.0490	0.3050	0.2820	0.0600	0.0620
	$\tau = 75\%$	0.3190	0.2890	0.0450	0.0420	0.4600	0.3800	0.0490	0.0430
$H_1(B_4)$		0.2140	0.3170	0.4470	0.4690	0.2490	0.3710	0.5920	0.6460
		$T = 500$, $n = 250$				$T = 500$, $n = 500$			
		UC_{MES}	UC^C_{MES}	IND_{MES}	IND^C_{MES}	UC_{MES}	UC^C_{MES}	IND_{MES}	IND^C_{MES}
H_0		0.0760	0.0430	0.0940	0.0600	0.0920	0.0360	0.0900	0.0610
$H_1(B_1)$	$\tau = 25\%$	0.0810	0.0760	0.0470	0.0550	0.1050	0.1140	0.0470	0.0490
	$\tau = 50\%$	0.1800	0.1680	0.0420	0.0450	0.2760	0.2790	0.0430	0.0420
	$\tau = 75\%$	0.3810	0.3600	0.0530	0.0610	0.5360	0.5050	0.0570	0.0580
$H_1(B_2)$	$\tau = 25\%$	0.2250	0.2460	0.0260	0.0330	0.3260	0.3940	0.0430	0.0380
	$\tau = 50\%$	0.7340	0.7160	0.0420	0.0290	0.9310	0.9290	0.0680	0.0290
	$\tau = 75\%$	0.9900	0.9770	0.1750	0.0140	1.0000	0.9960	0.3640	0.0140
$H_1(B_3)$	$\tau = 25\%$	0.1090	0.1100	0.0450	0.0540	0.1630	0.1960	0.0420	0.0430
	$\tau = 50\%$	0.2790	0.2670	0.0450	0.0540	0.3960	0.3880	0.0470	0.0400
	$\tau = 75\%$	0.3830	0.3460	0.0320	0.0380	0.5790	0.5400	0.0500	0.0520
$H_1(B_4)$		0.2200	0.2880	0.4640	0.5030	0.2860	0.3850	0.5970	0.6390
		$T = 2500$, $n = 250$				$T = 2500$, $n = 500$			
		UC_{MES}	UC^C_{MES}	IND_{MES}	IND^C_{MES}	UC_{MES}	UC^C_{MES}	IND_{MES}	IND^C_{MES}
H_0		0.0500	0.0410	0.0890	0.0810	0.0650	0.0380	0.0680	0.0590
$H_1(B_1)$	$\tau = 25\%$	0.0980	0.1010	0.0540	0.0540	0.1000	0.0980	0.0480	0.0370
	$\tau = 50\%$	0.2080	0.2080	0.0450	0.0440	0.3050	0.2990	0.0560	0.0510
	$\tau = 75\%$	0.4120	0.4100	0.0330	0.0330	0.6190	0.6070	0.0590	0.0540
$H_1(B_2)$	$\tau = 25\%$	0.2700	0.2810	0.0300	0.0280	0.4020	0.4220	0.0570	0.0580
	$\tau = 50\%$	0.8200	0.8220	0.0420	0.0290	0.9680	0.9690	0.0870	0.0600
	$\tau = 75\%$	0.9990	0.9980	0.1620	0.0640	1.0000	0.9990	0.4380	0.2270
$H_1(B_3)$	$\tau = 25\%$	0.1380	0.1350	0.0280	0.0330	0.1940	0.2030	0.0560	0.0560
	$\tau = 50\%$	0.2940	0.2910	0.0360	0.0340	0.4470	0.4460	0.0500	0.0430
	$\tau = 75\%$	0.4630	0.4570	0.0380	0.0340	0.6160	0.6020	0.0520	0.0400
$H_1(B_4)$		0.2770	0.2960	0.4040	0.4200	0.2860	0.3210	0.6580	0.6660

Note: This table displays the Monte Carlo results associated to the conditional MES setting. UC_{MES} and IND_{MES} denote the unconditional coverage test and independence test, respectively. The test statistics robust to estimation risk are superscripted by C. For the independence test, we consider a maximum lag order $m = 5$. Reported powers are sized-corrected.

to precise inference, even for small T and n. The robust UC test statistic has an empirical size which is generally close to the nominal one, typically about 4% for $T = 250$ and $n = 500$. The robust IND test statistic is slightly over-sized for finite sample sizes. Third, our backtests display good power performances. The UC tests generally detect well the misspecified alternatives B_1, B_2, and B_3, and we verify that there is a general improvement of the power as the sample size n increases, suggesting that these tests are consistent for these alternatives. Fourth, the results of the empirical power suggest complementarity between the UC and the IND tests. As in the unconditional case, the UC test displays good power performances against alternatives B_1, B_2, and B_3, while the empirical power of the IND test is still very low, except for large risk undervaluations. By definition, the IND backtest demonstrates better abilities to detect misspecification errors in model dynamics. Under the alternative B_4, the IND test has an empirical power that is almost twice that of the UC test. As observed in the VaR backtesting literature (Christoffersen, 1998), the combination of UC and IND tests allows to detect a larger spectrum of potential misspecifications in the dynamic models used to forecast the MES.

4.4 Backtesting other systemic risk measures

A great advantage of our backtesting approach lies in its simple extension to any systemic risk measures defined as a function of the MES or the CoVaR. In the sequel, we focus our analysis on the SES of Acharya et al. (2017), the SRISK of Acharya et al. (2012) and Brownlees and Engle (2017), and the ΔCoVaR of Adrian and Brunnermeier (2016), since they are the most prominent in the literature.[12]

4.4.1 Backtesting SES and SRISK

Brownlees and Engle (2017) define the SRISK as the expected capital shortfall of a financial institution, conditional on a crisis affecting the whole financial system. The capital shortfall, denoted CS_{1t}, is taken as the capital reserves the firm needs to hold for regulation and/or prudential management, minus the firm equity. Formally, we define the capital shortfall of the firm indexed by 1 on day t as

$$CS_{1t} = k\left(L_{1t} + W_{1t}\right) - W_{1t},$$

with L_{1t} the book value of debt, W_{1t} the market value of the firm equity, and k a prudential ratio. We define a systemic event as a market decline below a given threshold over a time

[12]Note that our backtests apply to many other measures, such as the ΔCoVaR proposed by Girardi and Ergün (2013), the Component Expected Shortfall (CES) of Banulescu and Dumitrescu (2015), or the Δ-Conditional Expected Shortfall (ΔCoES) of Ferreiro (2018), among others. However, they cannot be applied to the class of network systemic risk measures such as proposed by Billio et al. (2012) or Hué et al. (2019).

horizon h. Denote by $\widetilde{Y}_{t+h} = (\widetilde{Y}_{1t+h}, \widetilde{Y}_{2t+h})'$, or equally by $Y_{t+1:t+h} = (Y_{1t+1:t+h}, Y_{2t+1:t+h})'$, the vector of multi-period arithmetic firm and market returns between $t+1$ and $t+h$, and $\widetilde{VaR}_{2t+h} \equiv VaR_{2t+1:t+h}(\alpha)$ the corresponding α-VaR of $\widetilde{Y}_{2t+h}$ used as market threshold. We define the SRISK of the firm at day t for an horizon h as

$$\begin{aligned} SRISK_{1t}(h) &= \mathbb{E}_t(CS_{1t+h} | \widetilde{Y}_{2t+h} < \widetilde{VaR}_{2t+h}(\alpha)) \\ &= k\mathbb{E}_t(L_{1t+h} | \widetilde{Y}_{2t+h} < \widetilde{VaR}_{2t+h}(\alpha)) - (1-k)\,\mathbb{E}_t(W_{1t+h} | \widetilde{Y}_{2t+h} < \widetilde{VaR}_{2t+h}(\alpha)), \end{aligned}$$

where $\mathbb{E}_t(.)$ denotes the conditional expectation with respect to Ω_t. In order to compute this expectation, Brownlees and Engle assume that the debt is constant during a systemic event, i.e., $\mathbb{E}_t(L_{1t+h} | \widetilde{Y}_{2t+h} < \widetilde{VaR}_{2t+h}(\alpha)) = L_{1t}$. Furthermore, they introduce the concept of Long Run Marginal Expected Shortfall (LRMES) in order to compute the expected market value of the firm. The LRMES is simply a MES defined in terms of cumulative returns over h periods.[13]

$$MES_{1t}(\alpha; h) = \mathbb{E}_t(\widetilde{Y}_{1t+h} | \widetilde{Y}_{2t+h} < \widetilde{VaR}_{2t+h}(\alpha)),$$

Thus, we get $\mathbb{E}_t(W_{1t+h} | \widetilde{Y}_{2t+h} < \widetilde{VaR}_{2t+h}(\alpha)) = W_{it}(1 + \mathbb{E}_t(\widetilde{Y}_{1t+h} | \widetilde{Y}_{2t+h} < \widetilde{VaR}_{2t+h}(\alpha)))$ and finally, the SRISK is defined as[14]

$$SRISK_{1t}(h) = k\ L_{1t} - (1-k)\,W_{1t}(1 + MES_{1t}(\alpha; h)). \tag{4.6}$$

A similar systemic risk measure, the SES, has been proposed by Acharya et al. (2017). The SES represents the amount a bank equity drops below its target level (defined as a fraction k of assets) in case of a systemic crisis when the aggregate capital is less than k times the aggregate assets. Acharya et al. (2017) provide theoretical justification on how SES relates to MES and show that:

$$SES_{1t}(h) = (k\ LV_{1t} - 1 - \Pi\ MES_{1t}(\alpha; h) + \Delta)\ W_{1t},$$

where Π and Δ are two constant terms with $\Pi > 0$, and $LV_{1t} = (L_{1t} + W_{1t})/W_{1t}$ denotes the quasi-leverage ratio.

These two formulas show how the SRISK and the SES extend the LRMES in order to take into account both the liabilities and the size of the financial institution. Assuming that the level of debt cannot be negotiated in the case of a systemic event, implies that the level of debt is known at time t. Thus, the need for inference for these risk measures

[13] For $h = 1$, the LRMES boils down to the MES given by Equation (4.1).

[14] Equation (4.6) is similar to the definition reported by Brownlees and Engle (2017) in their Equation (1) (page 52), except that we do not adopt the same sign convention for the MES. Here, we have defined the MES as a negative quantity accordingly to our Equation (4.1).

comes only from the expected value of the firm at horizon h, which can break down into an initial value known at time t and the long run MES. More generally, we define a MES-based systemic risk measure, as a risk measure for which the need for inference comes only from its MES component.

Definition 4. *A MES-based systemic risk measure $RM_{1t}(h)$ for the firm 1, at time t, and for an horizon $h \geq 1$, is defined as a deterministic function of the (long run) MES with*

$$RM_{1t}(h) = g_t(MES_{1t}(\alpha; h), X_t),$$

where $g_t(.)$ is a monotonic (in MES) function and X_t a set of variables that belong to Ω_t.

By definition, testing the validity of any MES-based systemic risk measure is equivalent to testing the validity of the LRMES itself. The intuition is as follows. We can rewrite any MES-based risk measure as a function of the CoVaR (defined in terms of cumulative returns), with

$$RM_{1t}(h) = \int_0^1 g_t(\widetilde{CoVaR}_{1t+h}(\beta, \alpha, \theta_0), X_t)\mathrm{d}\beta.$$

This expression is similar to that get for the MES in Equation (4.3). Furthermore, the violation process $h_t(\alpha, \beta, \theta_0, X_t)$ associated to the quantity $g_t(\widetilde{CoVaR}_{1t+h}(\beta, \alpha, \theta_0), X_t)$ is equivalent to the violation process $h_t(\alpha, \beta, \theta_0)$ associated to the CoVaR, $\widetilde{CoVaR}_{1t+h}(\beta, \alpha, \theta_0)$: both binary processes will take the value 1 at the same dates. Thus, the cumulative joint violation process used to backtest the SRISK, the SES, or any MES-based measure, is equivalent to the cumulative joint violation process used to backtest the MES itself (see Appendix 4.8.7). As a consequence, the UC and IND tests are identical to those reported in Section 4.3.

Implementing the aforementioned test statistics requires the computation of the cumulative violation process $H_t(\alpha, \theta_0)$ associated to the joint distribution of the multi-period returns $(\widetilde{Y}_{1t+h}, \widetilde{Y}_{2t+h})$ over h periods. For $h > 1$, it is in general not available in closed form for the class of standard dynamic models (e.g., GARCH-type models) typically used for daily returns. The problem here is similar for the estimation of the LRMES itself (see Brownlees and Engle, 2017, for more details). However, it is straightforward to implement a simulation-based procedure to obtain an estimate of the cdf of the joint distribution for any horizon h. In the empirical application, we simulate a large number of paths of returns for the periods $t+1$ to $t+h$, conditional on the information available at time t, and we compute the corresponding cumulative returns. Then, we apply a kernel estimator to the simulated cumulative returns in order to estimate the cdf of their joint distribution.

4.4.2 Backtesting ΔCoVaR

Our testing procedure can be extended in order to asses the validity of the ΔCoVaR (Adrian and Brunnermeier, 2016). In the sequel, we define this indicator as the difference between the conditional VaR (CoVaR) of an institution conditional on the financial system being in distress and the CoVaR conditional on the financial system being in its median state.[15] Adrian and Brunnermeier define the stress for the financial system as a situation in which the market return Y_{2t} is equal to its $VaR_{2t}(\alpha)$, and consider a quantile regression model for the estimation of the CoVaR. A more general approach consists in defining the financial stress as a situation in which $Y_{2t} \leq VaR_{2t}(\alpha)$, as in Girardi and Ergün (2013). In both cases, we can estimate the CoVaR with M-GARCH type models and some usual results about truncated distributions. Similarly, we can represent the normal state of the system by a situation in which $VaR_{2t}(\beta_{\text{inf}}) \leq Y_{2t} \leq VaR_{2t}(\beta_{\text{sup}})$, with $\alpha < \beta_{\text{inf}} < \beta_{\text{sup}}$, for instance $\beta_{\text{inf}} = 25\%$ and $\beta_{\text{sup}} = 75\%$. Formally, if we denote by Y_{2t} the return of a portfolio of financial institutions (financial system), then the ΔCoVaR of the financial firm is given by[16]

$$\Delta\text{CoVaR}_t(\alpha) = CoVaR_{1t}(\alpha, \alpha, \theta_0) - CoVaR_{1t}(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0), \tag{4.7}$$

where $CoVaR_{1t}(\alpha, \alpha, \theta_0)$ and $CoVaR_{1t}(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0)$ satisfy respectively

$$\Pr\left(Y_{1t} \leq CoVaR_{1t}(\alpha, \alpha, \theta_0) | Y_{2t} \leq VaR_{2t}(\alpha, \theta_0); \Omega_{t-1}\right) = \alpha,$$

$$\Pr(Y_{1t} \leq CoVaR_{1t}(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0) | VaR_{2t}(\beta_{\text{inf}}, \theta_0) \leq Y_{2t} \leq VaR_{2t}(\beta_{\text{sup}}, \theta_0); \Omega_{t-1}) = \alpha.$$

In order to backtest the two CoVaRs, we define two violations through the following binary processes

$$h_t(\alpha, \alpha, \theta_0) = \mathbb{1}\left(\left(Y_{1t} \leq CoVaR_{1t}(\alpha, \alpha, \theta_0)\right) \cap \left(Y_{2t} \leq VaR_{2t}(\alpha, \theta_0)\right)\right),$$

$$h_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0) = \mathbb{1}((Y_{1t} \leq CoVaR_{1t}(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0)) \cap (VaR_{2t}(\beta_{\text{inf}}, \theta_0) \leq Y_{2t} \leq VaR_{2t}(\beta_{\text{sup}}, \theta_0))).$$

The logic of the test is then similar to that used for backtesting the MES. If the risk model is correctly specified, the two violation processes satisfy the mds property with $\mathbb{E}\left(h_t(\alpha, \alpha, \theta_0) - \alpha^2 | \Omega_{t-1}\right) = 0$ and $\mathbb{E}(h_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0) - \alpha(\beta_{\text{sup}} - \beta_{\text{inf}}) | \Omega_{t-1}) = 0$. By exploiting the mds property, we can propose various types of backtests for the ΔCoVaR

[15] Adrian and Brunnermeier provide various definitions of CoVaR depending on the direction of the conditioning. In order to maintain consistency with the conditioning event used in the MES definition, we consider hereinafter the indicator referred as *Exposure*-ΔCoVaR which is a measure of an individual institution exposure to system wide distress.

[16] Adrian and Brunnermeier deals with the special case $\beta_{\text{inf}} = \beta_{\text{sup}} = 0.5$, implying that the $CoVaR_{1t}(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0)$ is null. For ease of presentation, we only present the ΔCoVaR at horizon $h = 1$.

or for each of its constituents. As for the MES, we consider an unconditional coverage test with the following joint null hypothesis

$$\mathrm{H}_{0,UC} : \mathbb{E}\left(\mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0)\right) = \mu,$$

where $\mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0) = \left(h_t(\alpha, \alpha, \theta_0), h_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0)\right)'$ denotes the vector of violations and where $\mu = \left(\alpha^2, \alpha\left(\beta_{\text{sup}} - \beta_{\text{inf}}\right)\right)'$. Here, the intuition is similar to the backtests proposed for the multi-level VaR, i.e., the VaR defined for a finite set of coverage rates (see Francq and Zakoïan, 2016, for estimation issues). We can cite the tests proposed by Pérignon and Smith (2008), Colletaz et al. (2013), Leccadito et al. (2014), Wied et al. (2016), among others. These multi-level VaR backtesting procedures have been recently adapted in order to test the validity of ES forecasts by Kratz et al. (2018) and Couperier and Leymarie (2019). Given this joint null hypothesis, we consider a Wald test statistic with

$$UC_{\Delta\text{CoVaR}} = n\left(\overline{\mathbf{h}}(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \widehat{\theta}_T) - \mu\right)' \gamma^{-1} \left(\overline{\mathbf{h}}(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \widehat{\theta}_T) - \mu\right),$$

where $\overline{\mathbf{h}}(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \widehat{\theta}_T)$ denotes the out-of-sample mean of $\mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \widehat{\theta}_T)$ and γ is the conditional variance-covariance matrix of $\mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0)$ such that

$$\gamma = \begin{pmatrix} \alpha^2(1-\alpha^2) & -\alpha^3(\beta_{\text{sup}} - \beta_{\text{inf}}) \\ -\alpha^3(\beta_{\text{sup}} - \beta_{\text{inf}}) & \alpha(\beta_{\text{sup}} - \beta_{\text{inf}})(1 - \alpha(\beta_{\text{sup}} - \beta_{\text{inf}})) \end{pmatrix}.$$

Under the null hypothesis, the sequence $\{\mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0) - \mu\}_{t=T+1}^{T+n}$ is a mds with variance equal to γ. Thus, the Lindeberg-Levy central limit theorem implies that $UC_{\Delta\text{CoVaR}}$ converges to a chi-squared distribution with two degrees of freedom as soon as we evaluate the test statistic with the true value of the parameters θ_0 instead of $\widehat{\theta}_T$. This asymptotic distribution remains valid for the feasible statistic $UC_{\Delta\text{CoVaR}} \equiv UC_{\Delta\text{CoVaR}}(\widehat{\theta}_T)$ as soon as $T \to \infty$ and $n \to \infty$, with $\lambda = n/T \to 0$, i.e., without estimation risk. The proof of this result is reported in Appendix 4.8.8. In the general case $n/T \to \lambda < \infty$, the test statistic $UC_{\Delta\text{CoVaR}}$ is no longer chi-squared distributed. However, by considering the same reasoning as in Section 4.3, it is possible to derive a robust test statistic $UC^C_{\Delta\text{CoVaR}}$ which has an asymptotic chi-squared distribution whatever the relative values of n and T. Monte Carlo simulations show that the robust test statistic provides satisfactory size performances regardless of the sample size, and hence should be preferred when asymptotic theory does not apply conveniently (see also Appendix 4.8.8).

Finally, in case the ΔCoVaR risk model fails the unconditional coverage test, it is worth examining if this rejection is due to the CoVaR associated to the distress and/or to the median state of the financial system. To identify the origin of the error, we propose two UC sub-tests defined as $\mathrm{H}^{\text{distress}}_{0,UC} : \mathbb{E}\left(h_t(\alpha, \alpha, \theta_0)\right) = \alpha^2$ and

$\mathrm{H}_{0,UC}^{\text{median}} : \mathbb{E}\left(h_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0)\right) = \alpha\left(\beta_{\text{sup}} - \beta_{\text{inf}}\right)$. The testing procedure is then similar to that previously described.

4.5 Empirical application

We now turn to the empirical analysis with the goal of assessing the MES, SRISK, and ΔCoVaR predictions for a panel of large U.S. financial institutions. In a first part, we describe the data and the empirical setup, with a special focus on the multiple testing problem. Then in the following sub-divisions, we apply our UC and IND backtests and present our empirical results for short-term (one day) and mid-term (1 month) forecasting horizons.

4.5.1 Data description and empirical setup

Our empirical analysis focuses on the same panel of large U.S. financial institutions as that considered by Brownlees and Engle (2017). This dataset contains all U.S. financial firms with a market capitalization greater than 5 billion dollars as of the end of June 2007, and covers the period from January 3, 2000 to December 30, 2016. The panel is unbalanced in that some companies have not been traded continuously along the sample period (for instance, Lehman Brothers after its bankruptcy on September 15, 2008). The corresponding list of tickers and company names is reported in Appendix 4.8.9. For each firm in the panel, we compute the daily logarithmic returns and market capitalization from CRSP data, and we consider the daily CRSP market value-weighted index return as the market return. For the SRISK, we collect the quarterly book value of total liabilities from Compustat.

We calculate the systemic risk forecasts using a GARCH-DCC model as in Brownlees and Engle (2017), Engle et al. (2015), Idier et al. (2014), Acharya et al. (2012), among others. We estimate the parameters by maximum likelihood over the in-sample period assuming conditional normal joint distribution for the bivariate daily returns. We consider two types of estimation schemes, namely the rolling-window and recursive estimation schemes.[17] In the rolling-window scheme, we estimate the parameters using the most recent T daily observations up to the end of each month. Here, we consider a 2-year rolling window with $T = 500$. In the recursive scheme, we estimate the parameters using all available information from January 3, 2000 up to the end of each month, and thus, the sample size T increases as time goes by. For each companies, we compute the MES, SRISK, and ΔCoVaR forecasts at the end of each month from January 2005 to December 2016, so that, the forecasts depict the changes over time in systemic risk. As in Brownlees and Engle (2017), we only consider positive estimates for the SRISK as it represents a

[17]The Stern-NYU V-Lab website uses the recursive estimation scheme to compute the SRISK for a large set of U.S. and European financial firms.

capital shortfall, and the negative values are set to zero. Finally, we compute the systemic risk forecasts for a coverage level α equal to 5%, and the prudential capital ratio k is set to 8%.

Our empirical analysis involves testing several hypotheses simultaneously. Each month, we apply (a maximum of) 95 backtests representing the maximum number of firms in our panel. This framework causes a multiple testing problem, and in that regard we need to control for the number of false rejections. This issue is addressed by using a controlling method based on the Family Wise Error rate (FWE) that is the probability of rejecting at least one of the true null hypotheses. Within our empirical analysis, we can interpret it as the probability to get one or more false rejections of the systemic risk forecasts correctness among the total of firms. In the sequel, we will consider the Bonferroni procedure which allows a strong control of the FWE (Romano et al., 2008). The method is applied as follows. Denote by γ the significance level used for a single backtest, and denote by M the number of backtests to be applied simultaneously, with typically $M \leq 95$ in our case. For each individual backtest, i.e., each firm $s = 1, \ldots, M$, we compute an individual p-value p_s, and we reject the null hypothesis at level γ if $p_s < \gamma/M$ for all backtests.

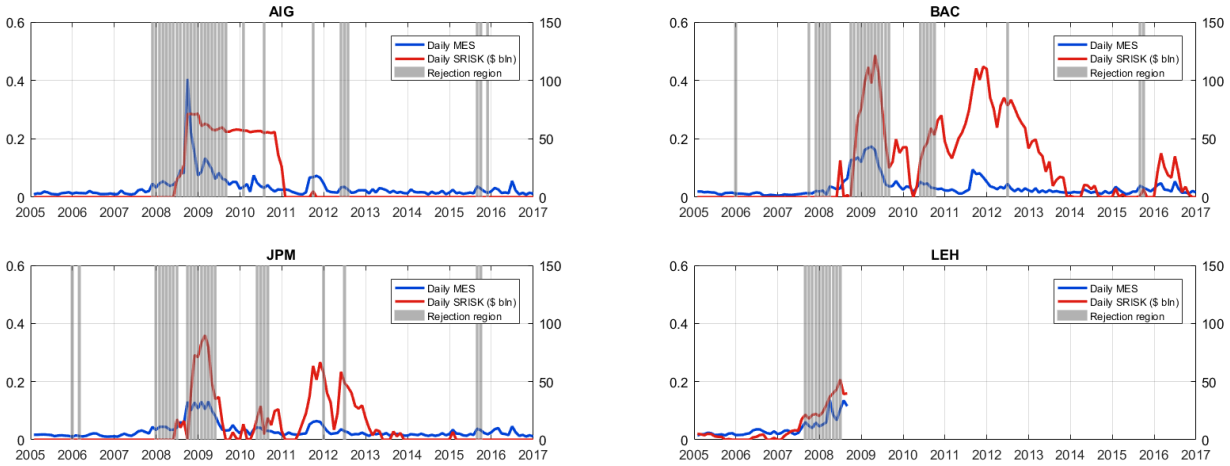
4.5.2 Empirical results for short-term MES and SRISK forecasts

We start by applying our backtests to MES and SRISK short-term forecasts by considering a time horizon of one day, i.e., $h = 1$. Figures 4.1 and 4.2 display the results of the UC and IND tests for four major financial institutions, namely Bank of America (BAC), JP Morgan (JPM), American International Group (AIG), and Lehman Brothers (LEH). For each firm, we display the one-day-step-ahead forecasts of the MES (blue line) and SRISK (red line) obtained with a recursive estimation window. We represent the rejection of the null hypothesis at a 5% significance level by shaded areas. These rejection regions rely on $n = 250$ out-of-sample observations.[18]

We observe that both MES and SRISK increase during the periods of financial instability, in particular the subprimes crisis (2008-2009) and the European debt crisis (2011-2012). At a first glance, these measures seem to capture well the impact of the financial crisis on the capital shortfall of these four financial institutions. However, the diagnosis is less obvious when one considers the backtests. Figure 4.1 highlights a significant number of rejections of the UC hypothesis during the 2008-2009 financial crisis, indicating that the MES and SRISK forecasts do not satisfy the unconditional coverage property during this period. The GARCH-DCC model does not fully capture the

[18]For ease of presentation, we only report the backtests that are robust to estimation risk. Overall, the unadjusted backtests display more rejections, especially for the rolling estimation scheme, as the ratio $\lambda = n/T$ increases. The corresponding results are available upon request.

Figure 4.1: Unconditional Coverage (UC) backtests for one-day risk forecast horizon (recursive estimation, $n = 250$)



increase in systemic risk, as it produces short-term MES forecasts which are associated to cumulative violations that are not observed with the right out-of-sample frequency. Interestingly, we note for Lehman Brothers severe rejections of the unconditional coverage hypothesis just before its bankruptcy, indicating for that bank a sharp change in its market conditions. For the rest of the sample, the unconditional coverage hypothesis is generally not rejected, indicating that MES and SRISK forecasts do a reasonably good job of depicting the true level of systemic risk.

Figure 4.2: Independence (IND) backtests for one-day risk forecast horizon (recursive estimation, $n = 250$, and $m = 5$)

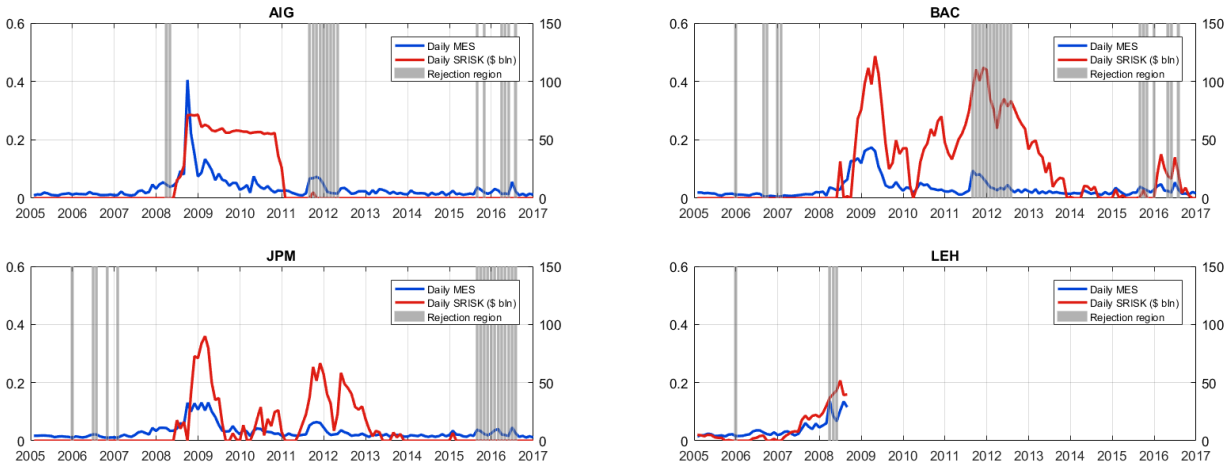
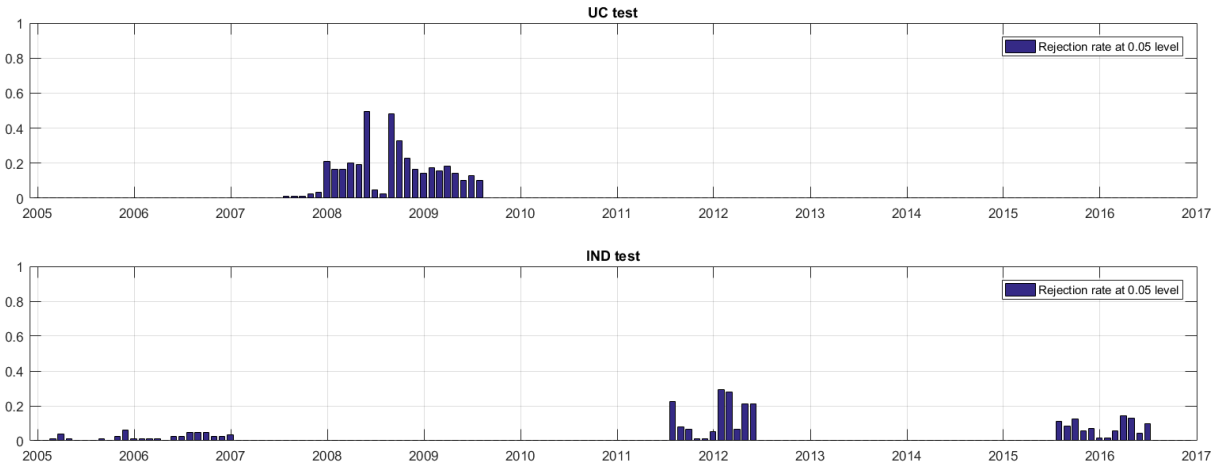


Figure 4.2 displays the results of the IND backtest, obtained for a lag-order $m = 5$. In contrast to the UC backtest, the rejection regions are not located during the global financial crisis, and differ between individual firms. For AIG and BAC, we observe a cluster of rejections of the independence hypothesis during the European debt crisis,

suggesting that our UC and IND statistics complement each other for detecting the two most prominent episodes of financial turmoil over the last two decades. In addition, JPM and to a lesser extent BAC and AIG, have experienced several rejections of the IND hypothesis in 2016-2017, which should warn regulators of the difficulties encountered by the MES and SRISK to capture the correct dynamics of the firms and market returns over these latest years. In Appendix 4.8.10, we provide various robustness check exercises by considering (i) a rolling window estimation scheme, and (2) a larger out-of-sample size, with $n = 500$. We observe overall the same findings.

To complete our analysis, we compute the UC and IND backtests for the full list of tickers, and report the rejection frequencies of the tests. As mentioned previously, we apply both tests for up to 95 firms simultaneously, and we implement a Bonferonni controlling method to address the multiple testing problem. Figure 4.3 displays the rejection rates of the UC test (top panel) and of the IND test (bottom panel) with a recursive estimation scheme and $n = 250$ out-of-sample observations. We reject the unconditional coverage hypothesis from 20% to 40% of the firms in the panel during the 2008-2009 subprime crisis, indicating that the GARCH-DCC model fails to capture the average level of systemic risk in times of crisis. In contrast to the UC backtest, we observe less pronounced rejections of the IND hypothesis, and the 2008-2009 financial crisis is no longer identified suggesting that the non-autocorrelation property of the violations is more easily satisfied than the unconditional coverage in global market distress situations. However, we observe up to 20% of rejections during the European debt crisis. Finally, we also highlight a cluster of rejections of the independence assumption at the end of the sample period, i.e., 2015-2017.

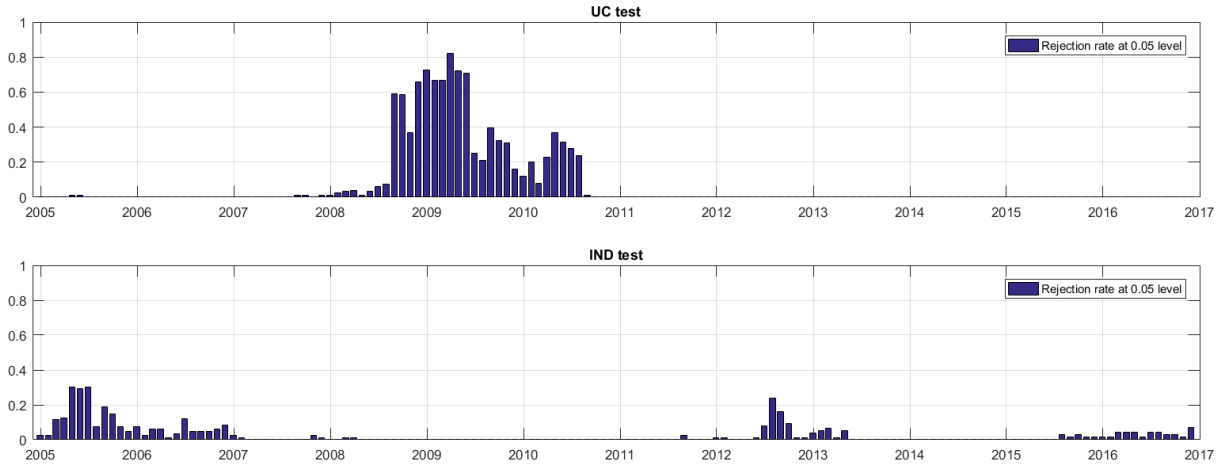
Figure 4.3: Rejection rates of the UC and IND backtests (recursive estimation, $n = 250$, and $m = 5$)



As further evidence of these measurement errors, Figure 4.4 displays the results obtained with $n = 500$, i.e., twice the previous out-of-sample size. The UC and IND

backtests identify the same periods of invalid risk forecasts, but the rejections are more frequent and more important due to the large size n that increases the power of the tests in detecting systemic risk measurement errors. In this situation, we observe up to 80% of rejections of the unconditional coverage hypothesis during the global financial crisis, providing evidence of the failure of the GARCH-DCC model to deliver valid MES and SRISK forecasts for most of the financial institutions for that period. We obtain similar results when considering a rolling estimation scheme (see Appendix 4.8.10).

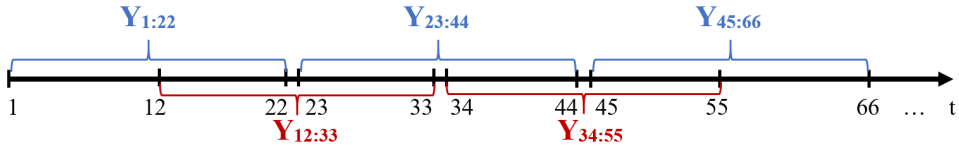
Figure 4.4: Rejection rates of the UC and IND backtests (recursive estimation, $n = 500$, and $m = 5$)



4.5.3 Empirical results for mid-term forecasts

Due to the difficulty of banks to adjust immediately their capital structure, it is generally preferable to forecast systemic risk at a longer horizon than a single day. Here, we follow Brownlees and Engle (2017) and consider a forecasting horizon of one month, i.e., $h = 22$ days, for LRMES and SRISK. Appendix 4.8.11 gives a detailed description of the method used to compute both measures over a time horizon $h > 1$. However, the use of a longer forecasting horizon implies to drastically increase the size of the out-of-sample dataset used to assess the unconditional coverage property of the violations. Indeed, as we consider cumulative firm and market returns $\widetilde{Y}_{t+h} \equiv Y_{t+1:t+h}$, the computation of only one observation of the cumulative violation process $H_t(\alpha, \widehat{\theta}_T)$ requires h observations of daily returns. Because the UC test statistic defined in Equation (4.5) depends on the out-of-sample mean of the violations $H_t(\alpha, \widehat{\theta}_T)$, its computation requires $n \times h$ daily observations. For instance, one year of out-of-sample observations (250 daily observations) and a forecasting horizon of 22 days, allows computing only 11 observations of $H_t(\alpha, \widehat{\theta}_T)$. In order to get 100 observations of $H_t(\alpha, \widehat{\theta}_T)$, we need about 9 years of daily returns. As the UC test statistic converges with n, this is clearly problematic.

Figure 4.5: Overlapping procedure



In order to address this issue, we propose a backtesting framework with overlapping blocks of data. Figure 4.5 summarizes the procedure for the case $h = 22$. Instead of considering the sequence of cumulative returns $\{Y_{1:22}, Y_{23:44}, \dots\}$ (in blue) to compute the violations, we consider the sequence $\{Y_{1:22}, Y_{12:33}, Y_{23:44}, \dots\}$ for which two subsequent cumulative returns share a common component of 11 daily returns (in blue and red). This method can be interpreted as a rolling window procedure applied to the out-of-sample observations, which allows increasing the number of observations available for the out-of-sample assessment, and thus the power of our test. Obviously, it also creates dependence across the violations $H_t(\alpha, \widehat{\theta}_T)$.[19] As a consequence, the property of independence of the cumulative joint violation process is no longer satisfied, and thus we cannot apply the IND backtest in this case.

Figure 4.6: Rejection rates of the UC backtest ($n = 100$, $h = 22$, overlap of 11 days)

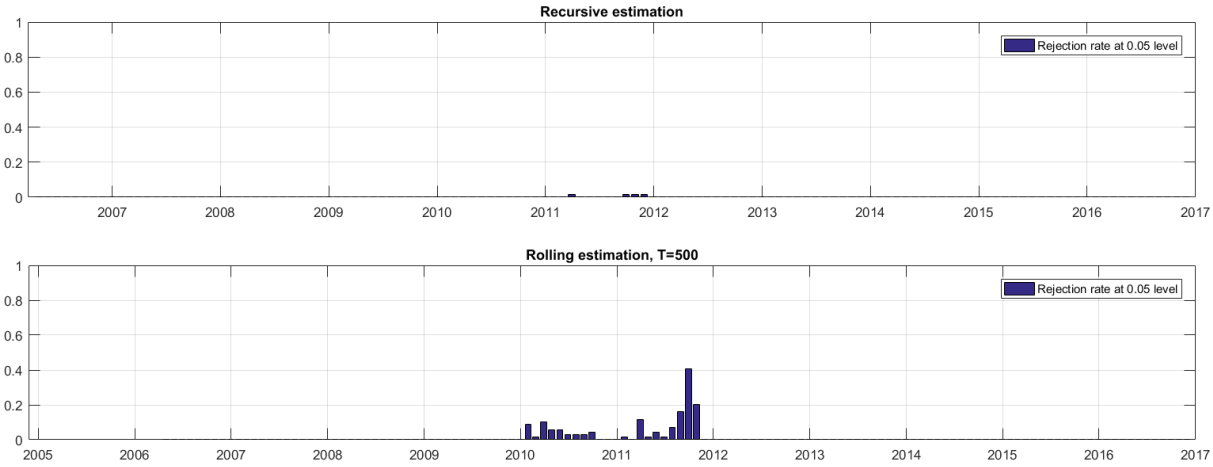


Figure 4.6 reports the rejection frequencies of the UC backtest for LRMES and SRISK calculated for the whole panel of firms . We estimate the GARCH-DCC parameters either with a recursive estimation (top panel) or with a rolling-window scheme (bottom panel). For each case, we backtest the SRISK over a time horizon $h = 22$ with $n = 100$ out-of-sample observations, requiring hence 1111 daily log returns given the 11 days

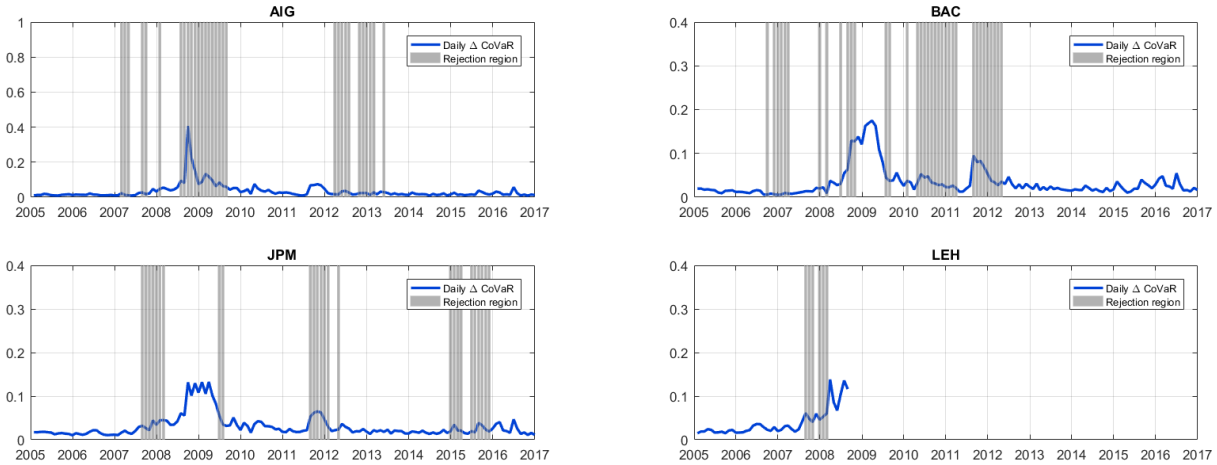
[19] If the overlapping is limited and generates only weak dependence across violations, the central limit theorem still applies and thus the UC test statistic remains normally distributed under the null. In order to correct the long run variance estimate, we consider a HAC estimator (Newey and West, 1987) for estimating the variance of the violation process. Monte Carlo experiments (not reported) show that this adjusted test is well-sized as soon as the overlap do not exceed 11 trading days for $h = 22$.

overlap window.[20] As mentioned earlier, the significance level is adjusted via Bonferonni controlling method. Interestingly, we find that the one-month-step ahead SRISK forecasts issued from the GARCH-DCC model are generally valid. For both estimation schemes, there are very few rejections of the UC hypothesis. Our findings differ drastically from those obtained for the daily forecasts where we found large rejection frequencies. In sum, if the risk model is unable to capture short term systemic risk, it succeeds in providing valid systemic risk forecasts for longer horizons. Our results are consistent with those of Brownlees and Engle (2017) who show that the predictive ability of SRISK is superior over longer horizons. Finally, we note that the rolling estimation setup is characterized by more rejections than the recursive estimation setup, highlighting that a widened sample period is more suitable to capture the long run dynamics.

4.5.4 Empirical results for ΔCoVaR forecasts

This section is devoted to the evaluation of the ΔCoVaR daily forecasts. We compute the ΔCoVaR with probability levels $\alpha = 0.05$, $\beta_{\text{inf}} = 0.25$, and $\beta_{\text{sup}} = 0.75$. As emphasized by Equation (4.7), the ΔCoVaR is defined as the difference between the CoVaR of the distress state and the CoVaR of the median state. In case the ΔCoVaR does not successfully pass the test, it is interesting to know whether this rejection stems from the CoVaR associated to the distress state or to the median state. Thus, we also report the two UC sub-tests associated to each type of CoVaR to complete the analysis, and identify which component of the ΔCoVaR is misspecified.

Figure 4.7: Unconditional Coverage (UC) backtests for ΔCoVaR (recursive estimation, $n = 250$)



[20]The joint distribution of the cumulative returns $\widetilde{Y}_{t+h} \equiv (Y_{1t+1:t+h}, Y_{2t+1:t+h})'$ is not available in closed form, and thus the derivative of their cdf is trickier to compute numerically. For this reason, we do not compute the robust statistics and alternatively, we set T to 500 in order to limit the impact of estimation risk.

Figure 4.7 reports the forecasts of ΔCoVaR for AIG, BAC, JPM, and LEH, and the rejection regions of the UC backtest in shaded areas. In all cases, we observe that the ΔCoVaR predictions increase in 2008-2009 and in 2012 suggesting that this indicator is sensitive to systemic events and thus may be helpful for systemic risk monitoring. However, at the same time, we observe several rejections of the unconditional coverage hypothesis for the ΔCoVaR forecasts, and thus we get empirical results associated to the ΔCoVaR and to the MES that coincide fairly well. Such similarities are not surprising given the strong relationship between MES and CoVaR as highlighted in Equation (4.3).[21] We also find additional rejections which are experienced at different periods depending on the firm. For instance, the ΔCoVaR forecasts of the institution JPM are generally misleading in 2015-2016.

In order to characterize the nature of these rejections, and identify which component of the ΔCoVaR is misspecified, Figures 4.8 and 4.9 display the results of the backtest for the CoVaR of the distress state and for the CoVaR of the median state, respectively. If we compare both CoVaRs, we find that the CoVaR calculated in distress conditions generally increases more markedly during financial crisis than the CoVaR evaluated in normal conditions. The results of the two UC sub-tests are clear-cut. The rejection periods of the ΔCoVaR identified in Figure 4.7 coincide with those of the CoVaR associated to the distress state, while the CoVaR of the median state is overall not affected by misspecification (except for AIG). To summarize, during global financial market turbulence, a common risk model such as the GARCH-DCC is not able to produce sound estimates of the CoVaR associated to distress situations simply because the frequency of the associated out-of-sample violations is not correct.

Finally, we can extend our ΔCoVaR assessment to our 95 U.S. financial institutions. Figure 4.10 reports the rejection rates for the full list of tickers for the ΔCoVaR (top panel), the CoVaR of the distress state (middle panel), and the CoVaR of the median state (bottom panel). We obtain these rejections rates with a recursive estimation scheme, $n = 250$ out-of-sample observations, and a Bonferroni correction. Our results confirm that the rejections of the risk model are generally observed for the ΔCoVaR and for the CoVaR of the distress state, while the validity of the CoVaR of the median state is generally not rejected. Furthermore, these rejections mainly occur during the 2008-2009 financial crisis indicating that the ΔCoVaR and the stressed CoVaR may be severely affected by forecasting errors during crisis periods. It hence appears that the tail sensitivity of the conditioning event is a key aspect in the modeling of the CoVaR. We obtain similar results for $n = 500$ with more pronounced rejections (not reported).

[21] In the same spirit, Benoit et al. (2013) show theoretically that the higher the correlation between the returns of the SIFIs and the market, the more likely it is that MES and CoVaR will lead to a convergent diagnostic.

Figure 4.8: Unconditional Coverage (UC) backtests for stressed CoVaR (recursive estimation, $n = 250$)

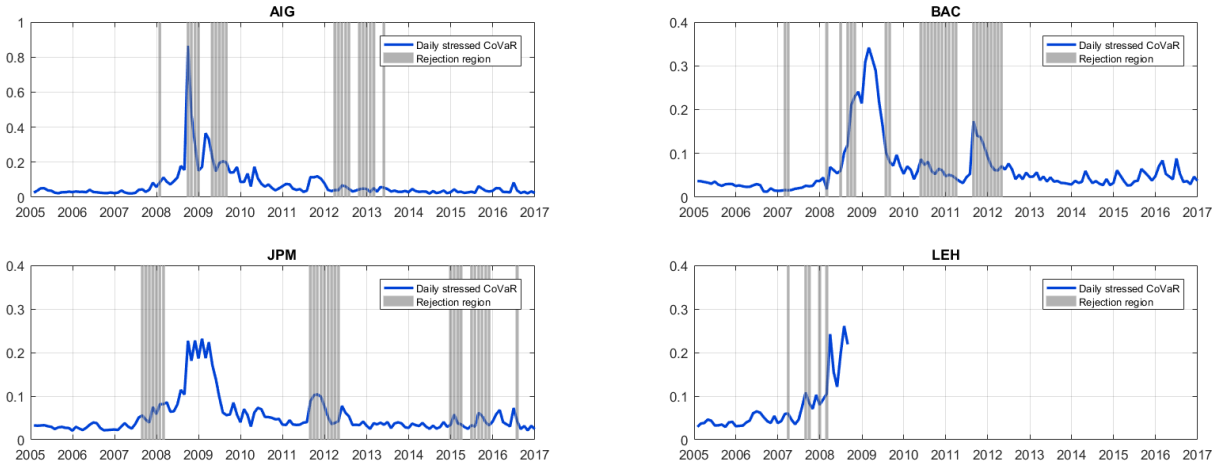
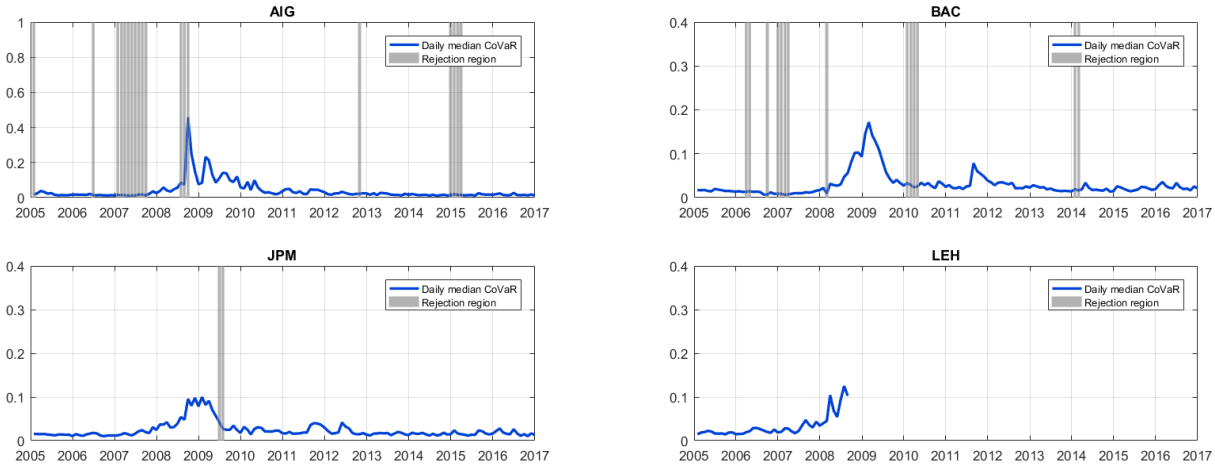


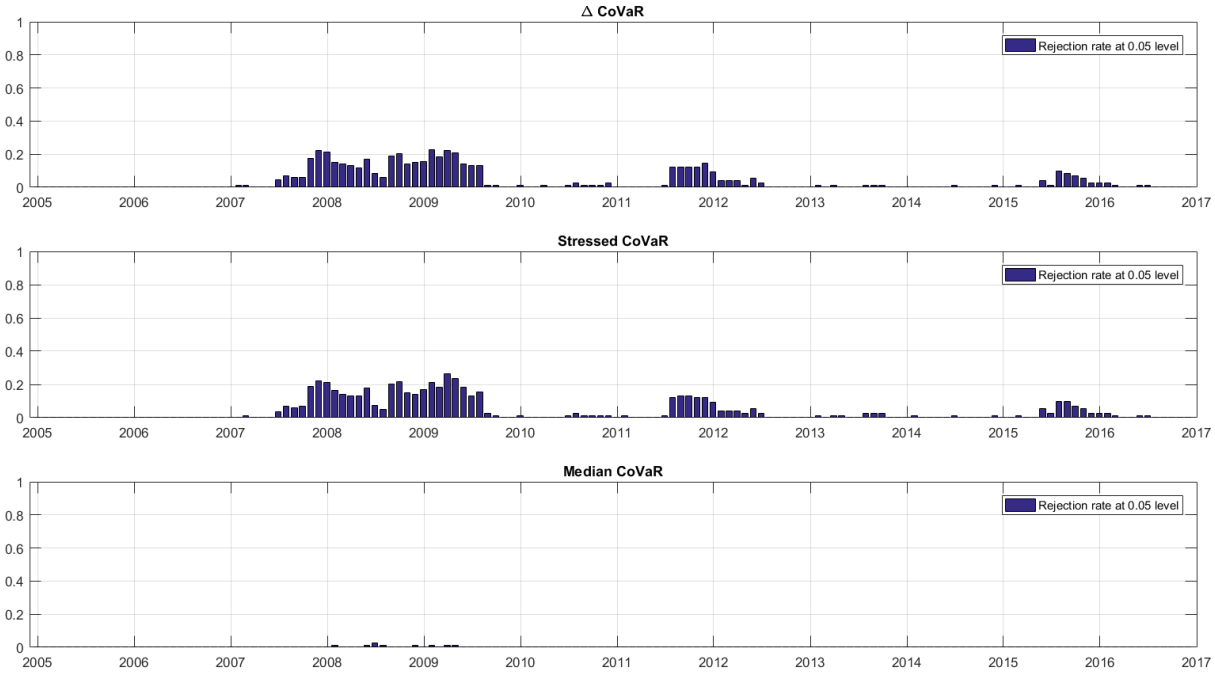
Figure 4.9: Unconditional Coverage (UC) backtests for median CoVaR (recursive estimation, $n = 250$)



4.6 Early warning system

A by-product of our backtesting procedure is to provide an efficient Early Warning System (EWS) indicator that allows to depict significant changes in the financial system stability. The general idea is the following: a given risk model, say a GARCH-DCC, is used to produce a systemic risk measure, say MES or SRISK. If the model is well-specified, we will observe the cumulative violations with the right frequency, and we will not reject the null hypothesis of unconditional coverage. On the contrary, if the model is misspecified and/or if the market conditions are likely to change, the model produces invalid systemic risk measures. Let us define an adjusted systemic risk measure as a measure produced by a potentially misspecified model, but for which we do not reject the null hypothesis of unconditional coverage. Thus, we can use the difference observed

Figure 4.10: Rejection rates for the ΔCoVaR, stressed CoVaR, and median CoVaR (recursive estimation, and $n = 250$)



between adjusted and unadjusted risk measures as a predictor of the financial distress. This gives useful insights for monitoring the financial system on a real-time basis.

In the same spirit as it was done for VaR or ES by Gouriéroux and Zakoïan (2013), Boucher et al. (2014) or Lazar and Zhang (2019), we propose to adjust imperfect MES-based forecasts by considering the mean property of the cumulative joint violation process. Formally, we define an adjusted coverage level $\widetilde{\alpha}$ for the MES-based forecast so that the null hypothesis of unconditional coverage is valid, i.e., $\mathbb{E}(H_t(\widetilde{\alpha}, \theta_0)) = \alpha/2$. If the risk model is well-specified then $\widetilde{\alpha} = \alpha$. We can obtain a feasible adjusted coverage level $\widetilde{\alpha}$ as the solution of the program

$$\widetilde{\alpha} = \underset{\alpha \in]0,1[}{\arg\min} \left(\bar{H}(\alpha, \widehat{\theta}_T) - \alpha/2 \right)^2,$$

where $\bar{H}(\alpha, \widehat{\theta}_T) = (1/n) \sum_{t=T+1}^{T+n} H_t(\alpha, \widehat{\theta}_T)$ denotes the mean of the cumulative joint violation process, and $\alpha/2$ represents the expected value of the cumulative joint violation process under the null hypothesis of correct unconditional coverage. Given the definition of $\widetilde{\alpha}$, the adjusted forecast $MES_{1t}(\widetilde{\alpha}, \widehat{\theta}_T)$ becomes valid because we observe the corresponding cumulative joint violations with the right frequency. For the sake of presentation, we only consider the MES, but we can easily generalize this adjustment to any MES-based systemic risk measure (e.g., SES and SRISK).

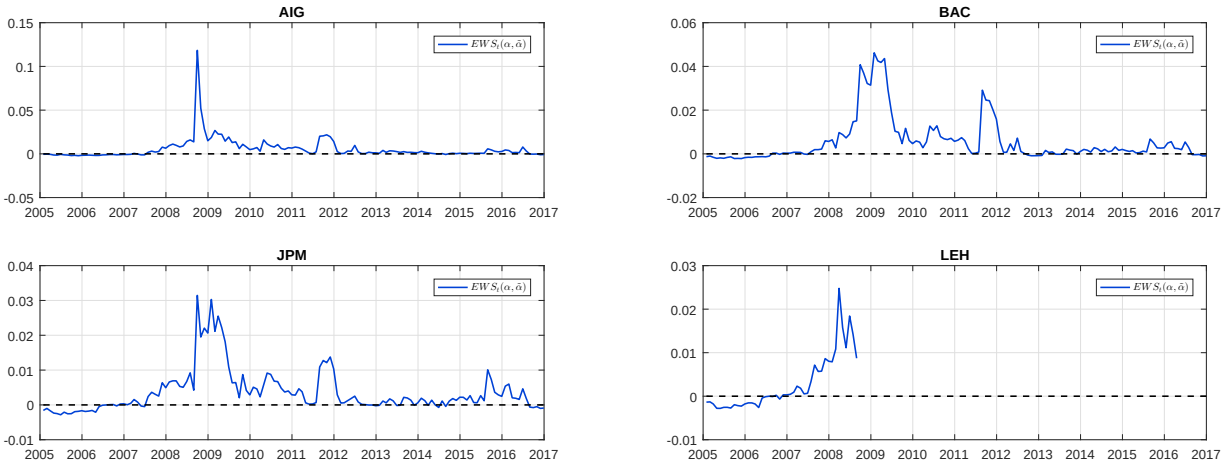
By considering the difference between the adjusted and unadjusted MES, it is then possible to build an indicator which depicts the unexpected changes in market conditions, and provides early warning signals of distress in the financial system. Formally, this indicator denoted $EWS_t(\alpha, \widetilde{\alpha})$ is defined as

$$EWS_t(\alpha, \widetilde{\alpha}) = MES_{1t}(\widetilde{\alpha}, \widehat{\theta}_T) - MES_{1t}(\alpha, \widehat{\theta}_T).$$

This quantity depicts the magnitude of model misspecification at time t expressed in terms of systemic risk, and represents the unexplained systemic risk component non-anticipated by market practitioners. In line with our notations, a positive value of $EWS_t(\alpha, \widetilde{\alpha})$ induces an unexpected growth in systemic risk, while a negative value indicates an unexpected decline in systemic risk.

Figure 4.11 reports the indicator $EWS_t(\alpha, \widetilde{\alpha})$ for AIG, BAC, JPM, and LEH. The results are reported using a recursive window estimation scheme and $n = 250$ out-of-sample observations. In line with our previous results, we observe a large increase of $EWS_t(\alpha, \widetilde{\alpha})$ during the 2007-2009 financial crisis, and to a lesser extent for the European debt crisis. During these periods, the GARCH-DCC model is unable to provide valid systemic risk forecasts and an adjustment on the coverage rate is required to get right measures. Interestingly, we also observe a sharp increase of this indicator for Lehman Brothers just at the beginning of 2007, largely before its bankruptcy.

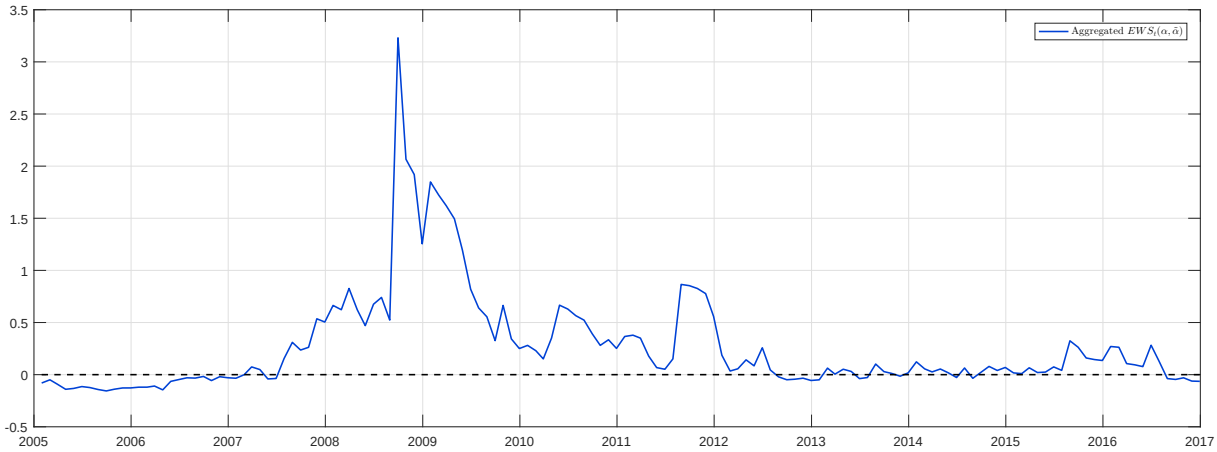
Figure 4.11: $EWS_t(\alpha, \widetilde{\alpha})$ for a panel of four firms, recursive estimation, $n = 250$



Finally, we carry out the same exercise at an aggregated level by averaging our indicator over the 95 U.S. firms of our panel. By doing so, we aim to evaluate its ability to be used as an EWS for the financial system as a whole. Figure 4.12 displays the aggregated $EWS_t(\alpha, \widetilde{\alpha})$. Our results are overall the same compared to what we observe at the firm level. We find that there is a large increase of the indicator during the global financial crisis. This increase starts at the mid of 2007 confirming the ability of our indicator

to early detect reversals. The highest value of our aggregated early warning indicator, 3.23, is observed in September 2008 and coincides with the collapse of Lehman Brothers. Similar results are obtained when considering a rolling windows estimation scheme, and using $n = 500$ for the out-of-sample period.

Figure 4.12: Aggregated $EWS_t(\alpha, \widetilde{\alpha})$, recursive estimation, $n = 250$



In their recent work, Brownlees et al. (2018) evaluate the ability of CoVaR and SRISK to provide early warning signals of distress in the financial system. They conclude that even if CoVaR and SRISK may be helpful to identify systemic institutions in periods of distress, both indicators would be non-efficient to predict when the next crisis is likely to occur. Our approach here is complementary: looking at the number of violations experienced by a systemic risk indicator, we aim to detect misspecification and quantify its level. The importance in systemic risk measurement error is indicative of the change in market conditions, and detecting these reversals may be an efficient way to forecast financial crisis.

4.7 Conclusion

This chapter develops the first statistical procedure to backtest systemic risk measures. The tests are built-up in analogy with the recent testing strategy for ES proposed by Du and Escanciano (2017). This approach has many advantages. First, the procedure is fully consistent with the testing strategies applied by risk managers to assess the validity of market risk measures such as VaR and ES and its implementation is no more difficult as it only requires for the user to evaluate the cdf of the bivariate returns. Second, it allows to perform a separate test for unconditional coverage and independence hypothesis (Christoffersen, 1998). Third, Monte-Carlo simulations show that for realistic sample sizes, the tests have good finite sample properties. Finally, we pay a particular attention to the consequences of estimation risk and derive a robust version of the test statistics.

In an empirical application, we apply our tests on a panel of large U.S. financial institutions considering one of the most popular specification for modeling systemic risk, namely the GARCH-DCC model. Three key findings emerge from this empirical analysis. Firstly, it appears that the short-term forecasts of systemic risk (one-day ahead forecast) do not satisfy the unconditional coverage hypothesis during the 2007-2009 global financial crisis. On a very short-term basis, the forecasting model is hence unable to capture the true level of systemic risk. In such situation, the risk model provides inaccurate systemic risk predictions, and leads to a misleading identification of the SIFIs. Secondly, and highly contrasted with the previous finding, we observe for longer forecasting horizons (one-month ahead forecast) that the unconditional coverage hypothesis is no longer rejected, suggesting that the GARCH-DCC model correctly anticipates the firm risk contribution on the mid and long run. As a general recommendation, academics and regulators should be more cautious in computing MES-based predictions over short horizons than over mid and long horizons. Finally, the empirical application clearly highlights that the unconditional coverage test and independence test complement each other well. They generally identify different periods of misspecification. Interestingly, the independence test identifies misleading forecasts during the 2011-2012 European debt crisis while the unconditional coverage test is clearly more sensitive to the global 2007-2009 financial crisis.

To address the issue of systemic risk measurement errors, we develop an original procedure which adjusts the misleading systemic risk predictions. The technique consists in modifying the coverage level of the MES-based risk measure, i.e., the severity of the market distress event, so as to meet the unconditional coverage hypothesis. As a by product of this procedure, we introduce an EWS indicator defined as the difference between the misleading forecast and its adjusted counterpart. Our indicator reports a sharp increase before the early signs of the crisis, and takes its highest value during the historic collapse of Lehman Brother. As a forward looking measure, it may complete the toolbox used by academics and regulators to capture the build-up of systemic risk in tranquil times, and improve the allocation of regulatory capital among banks.

4.8 Appendix

4.8.1 Appendix A: Assumptions

To derive the asymptotic properties of the test statistics and the robust test statistics, we introduce the following assumptions.

A1: The vectorial process $Y_t = (Y_{1t}, Y_{2t})'$ is strictly stationary and ergodic.

A2: The marginal distribution of Y_{2t} is given by $F_{Y_2}(Y_{2t}; \Omega_{t-1}, \theta_0)$ and the truncated distribution of Y_{1t} given $Y_{2t} \leq VaR_{2t}(\alpha, \theta_0)$ is given by $F_{Y_1|Y_2 \leq VaR_{2t}(\alpha,\theta_0)}(Y_{1t}; \Omega_{t-1}, \theta_0)$.

A3: $\theta_0 \in \Theta$, with Θ a compact subspace of $\mathbb{R}^p$.

A4: The estimator $\widehat{\theta}_T$ is consistent for θ_0 and is asymptotically normally distributed such that:

$$\sqrt{T}\left(\widehat{\theta}_T - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma_0\right),$$

with Σ_0 a positive definite $p \times p$ matrix. Denote $\mathbb{V}_{as}(\widehat{\theta}_T) = \Sigma_0/T$.

4.8.2 Appendix B: Cumulative joint violation process

Proof. First, let us rewrite $H_t(\alpha, \theta_0)$ in a more convenient way, through the Probability Integral Transformation (PIT). Notice that the cumulative joint violation process $H_t(\alpha, \theta_0)$ depends on the distribution of Y_t as follows:

$$\begin{aligned} H_t(\alpha, \theta_0) &= \mathbb{1}(Y_{2t} \leq VaR_{2t}(\alpha, \theta_0)) \times \int_0^1 \mathbb{1}(Y_{1t} \leq CoVaR_{1t}(\beta, \alpha, \theta_0)) \, \mathrm{d}\beta \\ &= \mathbb{1}(F_{Y_2}(Y_{2t}; \Omega_{t-1}, \theta_0) \leq \alpha) \times \int_0^1 \mathbb{1}\left(F_{Y_1|Y_2 \leq VaR_{2t}(\alpha;\theta_0)}(Y_{1t}; \Omega_{t-1}, \theta_0) \leq \beta\right) \mathrm{d}\beta. \end{aligned}$$

Let us introduce two terms that we can interpret as "generalized" errors, namely $u_{2t}(\theta_0) \equiv u_{2t} = F_{Y_2}(Y_{2t}; \Omega_{t-1}, \theta_0)$ and $u_{12t}(\theta_0) \equiv u_{12t} = F_{Y_1|Y_2 \leq VaR_{2t}(\alpha,\theta_0)}(Y_{1t}; \Omega_{t-1}, \theta_0)$. Then, the cumulative joint violation process becomes

$$H_t(\alpha, \theta_0) = \mathbb{1}(u_{2t} \leq \alpha) \int_0^1 \mathbb{1}(u_{12t} \leq \beta) \, \mathrm{d}\beta = \mathbb{1}(u_{2t} \leq \alpha) \int_{u_{12t}}^1 1 \mathrm{d}\beta.$$

Thus, we can express the process $H_t(\alpha, \theta_0)$ as a simple function of the transformed i.i.d. variables u_{2t} and u_{12t} defined over $[0, 1]$, such as

$$H_t(\alpha, \theta_0) = (1 - u_{12t}(\theta_0)) \times \mathbb{1}(u_{2t}(\theta_0) \leq \alpha).$$

The PIT implies that the variable u_{2t} has a uniform $U_{[0,1]}$ distribution. The binary variable $\mathbb{1}(u_{2t}(\theta_0) \leq \alpha)$ has a Bernouilli distribution with a success probability equal to α. The variable u_{12t} has also a $U_{[0,1]}$ distribution as soon as the PIT transformation $F_{Y_1|Y_2 \leq VaR_{2t}(\alpha,\theta_0)}(.; \Omega_{t-1}, \theta_0)$ is applied to observations Y_{1t} for which $Y_{2t} \leq VaR_{2t}(\alpha, \theta_0)$.[22]

$$\mathbb{1}(u_{2t}(\theta_0) \leq \alpha) | \, \Omega_{t-1} \sim \text{Bernouilli}(\alpha),$$

$$(1 - u_{12t}(\theta_0)) | \, \{\Omega_{t-1}, u_{2t}(\theta_0) \leq \alpha\} \sim U_{[0,1]}.$$

The two first conditional moments of the cumulative joint process $H_t(\alpha, \theta_0)$ are then given by

$$\begin{aligned} \mathbb{E}(H_t(\alpha, \theta_0) | \, \Omega_{t-1}) &= \Pr(u_{2t}(\theta_0) \leq \alpha | \, \Omega_{t-1}) \times \mathbb{E}(H_t(\alpha, \theta_0) | \, u_{2t}(\theta_0) \leq \alpha, \Omega_{t-1}) \\ &= \alpha - \alpha \, \mathbb{E}(u_{12t}(\theta_0) | \, u_{2t}(\theta_0) \leq \alpha, \Omega_{t-1}), \end{aligned}$$

$$\mathbb{E}\left(H_t^2(\alpha, \theta_0) \middle| \, \Omega_{t-1}\right) = \alpha \, \mathbb{E}\left(1 - 2u_{12t}(\theta_0) + u_{12t}^2(\theta_0) \middle| \, u_{2t}(\theta_0) \leq \alpha, \Omega_{t-1}\right).$$

Since the conditional distribution of $u_{12t}(\theta_0)$ given Ω_{t-1} is $U_{[0,1]}$ with $\mathbb{E}(u_{12t}(\theta_0) | \, u_{2t}(\theta_0) \leq \alpha, \Omega_{t-1}) = 1/2$ and $\mathbb{E}(u_{12t}^2(\theta_0) | \, u_{2t}(\theta_0) \leq \alpha, \Omega_{t-1}) = 1/3$,

[22] We thank Andrew Patton for this suggestion.

then we get

$$\mathbb{E}\left(H_t\left(\alpha,\theta_0\right)|\,\Omega_{t-1}\right)=\frac{\alpha}{2},$$

$$\mathbb{V}\left(H_t\left(\alpha,\theta_0\right)|\,\Omega_{t-1}\right)=\alpha\left(\frac{1}{3}-\frac{\alpha}{4}\right).$$

■

4.8.3 Appendix C: Proof of Theorem 2

Proof. Denote $H_t(\alpha,\theta) = (1-u_{12t}(\theta))\,1(u_{2t}(\theta)\leq\alpha)$ the cumulative violation process, with $u_{2t}(\theta) = F_{Y_2}(Y_{2t};\Omega_{t-1},\theta)$ and $u_{12t}(\theta) = F_{Y_1|Y_2\leq VaR_{2t}(\alpha,\theta)}(Y_{1t};\Omega_{t-1},\theta)$, $\forall t = T+1,...,T+n$ and $\forall\theta\in\Theta$. Under the null hypothesis $H_{0,UC}$, the sequence $\{H_t(\alpha,\theta_0)-\alpha/2\}_{t=T+1}^{T+n}$ is a mds with $\sigma_H^2 = \mathbb{V}(H_t(\alpha,\theta_0)) = \alpha(1/3-\alpha/4)$. For simplicity, we assume that Ω_{t-1} only includes a finite number of lagged values of Y_t, i.e., there is no information truncation. We can rewrite the test statistic UC_{MES} as

$$UC_{MES} = \frac{1}{\sigma_H\sqrt{n}}\sum_{t=T+1}^{T+n}\left(H_t(\alpha,\widehat{\theta}_T)-\alpha/2\right).$$

Under Assumptions A1-A4, the continuous mapping theorem implies that

$$\frac{1}{\sqrt{n}}\sum_{t=T+1}^{T+n}\left(H_t(\alpha,\widehat{\theta}_T)-\mathbb{E}\left(H_t(\alpha,\widehat{\theta}_T)\middle|\,\Omega_{t-1}\right)\right) = \frac{1}{\sqrt{n}}\sum_{t=T+1}^{T+n}\left(H_t(\alpha,\theta_0)-\mathbb{E}\left(H_t(\alpha,\theta_0)\middle|\,\Omega_{t-1}\right)\right)+o_p(1).$$

Rearranging these terms gives

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=T+1}^{T+n}\left(H_t(\alpha,\widehat{\theta}_T)-\mathbb{E}\left(H_t(\alpha,\theta_0)\middle|\,\Omega_{t-1}\right)\right) &= \frac{1}{\sqrt{n}}\sum_{t=T+1}^{T+n}\left(H_t(\alpha,\theta_0)-\mathbb{E}\left(H_t(\alpha,\theta_0)\middle|\,\Omega_{t-1}\right)\right) \\ &+\frac{1}{\sqrt{n}}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\left(H_t(\alpha,\widehat{\theta}_T)-H_t(\alpha,\theta_0)\right)\middle|\,\Omega_{t-1}\right)+o_p(1).\end{aligned} \tag{4.8}$$

The mean value theorem implies that

$$H_t(\alpha,\widehat{\theta}_T) = H_t(\alpha,\theta_0)+\left(\widehat{\theta}_T-\theta_0\right)'\frac{\partial H_t(\alpha,\widetilde{\theta})}{\partial\theta},$$

where $\widetilde{\theta}$ is an intermediate point between θ_0 and $\widehat{\theta}_T$. Equation (4.8) becomes

$$\begin{aligned}\frac{1}{\sigma_H\sqrt{n}}\sum_{t=T+1}^{T+n}\left(H_t(\alpha,\widehat{\theta}_T)-\alpha/2\right) &= \frac{1}{\sigma_H\sqrt{n}}\sum_{t=T+1}^{T+n}\left(H_t(\alpha,\theta_0)-\alpha/2\right) \\ &+\frac{\sqrt{\lambda}}{\sigma_H}\sqrt{T}(\widehat{\theta}_T-\theta_0)'\frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\frac{\partial H_t(\alpha,\widetilde{\theta})}{\partial\theta}\middle|\,\Omega_{t-1}\right)+o_p(1).\end{aligned}$$

Assume that $T\to\infty$, $n\to\infty$ and $n/T\to\lambda$ with $0\leq\lambda<\infty$. Under the null hypothesis $H_{0,UC}$, the first term on the right hand converges in distribution to a standard normal distribution. The covariance between the first term and $\sqrt{T}(\widehat{\theta}_T-\theta_0)$ is 0 as $\widehat{\theta}_T$ depends on the in-sample observations and the summand in the first term is for out-of-sample observations. Under Assumption A4, $\widetilde{\theta}\xrightarrow{p}\theta_0$ and since $\partial H_t(\alpha,\theta_0)/\partial\theta$ is also a mds, we

have

$$\frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\frac{\partial H_t(\alpha,\widetilde{\theta})}{\partial\theta}\middle|\,\Omega_{t-1}\right)\xrightarrow{p}R_{MES}=\mathbb{E}_0\left(\frac{\partial H_t\left(\alpha,\theta_0\right)}{\partial\theta}\right),$$

where $\mathbb{E}_0\left(.\right)$ denotes the expectation with respect to the true distribution of $H_t\left(\alpha,\theta_0\right)$. So, we get

$$\frac{\sqrt{\lambda}}{\sigma_H}\frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\frac{\partial H_t(\alpha,\widetilde{\theta})}{\partial\theta}\middle|\,\Omega_{t-1}\right)'\sqrt{T}(\widehat{\theta}_T-\theta_0)\xrightarrow{d}\mathcal{N}\left(0,\frac{\lambda}{\sigma_H^2}R'_{MES}\Sigma_0R_{MES}\right),$$

and finally

$$UC_{MES}\xrightarrow{d}\mathcal{N}\left(0,1+\lambda\frac{R'_{MES}\Sigma_0R_{MES}}{\alpha\left(1/3-\alpha/4\right)}\right).$$

■

4.8.4 Appendix D: Bivariate normal case

Proof. Let us assume that $Y_t = (Y_{1t}, Y_{2t})'$ such that

$$Y_t = \Sigma_t^{1/2} \varepsilon_t$$

where $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ are i.i.d. $\mathcal{N}(0, I_2)$, where I_2 denotes the 2×2 identity matrix and $\Sigma_t = \Sigma_t(\theta_0)$ is the conditional variance-covariance matrix of Y_t given Ω_{t-1}. Denote by $f(y, \Sigma_t) \equiv f(y_1, y_2, \Sigma_t)$ the pdf and by $F(y, \Sigma_t) \equiv F(y_1, y_2, \Sigma_t)$ the cdf of the joint distribution of Y_t such that

$$f(y, \Sigma_t) = \frac{1}{2\pi} |\Sigma_t|^{-\frac{1}{2}} \exp\left(-\frac{y' \Sigma_t y}{2}\right),$$

$$F(y, \Sigma_t) = \Pr((Y_1 \leq y_1) \cap (Y_2 \leq y_2)) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f(a, b, \Sigma_t) \, \mathrm{d}a\mathrm{d}b.$$

Using the Dwyer formula (1967), we know that

$$\frac{\partial f(y, \Sigma_t)}{\partial \Sigma_t} = \frac{\partial \ln f(y, \Sigma_t)}{\partial \Sigma_t} \times f(y, \Sigma_t) = -\frac{f(y, \Sigma_t)}{2} \left(\Sigma_t^{-1} - \Sigma_t^{-1} y y' \Sigma_t^{-1}\right),$$

and we get

$$\begin{aligned} \frac{\partial F(y, \Sigma_t)}{\partial \Sigma_t} &= \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} \frac{\partial f(a, b, \Sigma_t)}{\partial \Sigma_t} \mathrm{d}a\mathrm{d}b \\ &= -\frac{\Sigma_t^{-1}}{2} F(y, \Sigma_t) + \frac{1}{2} \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f(a, b, \Sigma_t) \, \Sigma_t^{-1} \Delta(a, b) \, \Sigma_t^{-1} \mathrm{d}a\mathrm{d}b, \end{aligned} \tag{4.9}$$

with $\Delta(a, b)$ a 2×2 matrix equal to $(a, b)' \times (a, b)$. If we define the conditional variance-covariance matrix as

$$\Sigma_t = \begin{pmatrix} \sigma_{1t}^2 & \sigma_{12t} \\ \sigma_{12t} & \sigma_{2t}^2 \end{pmatrix},$$

we can decompose the matrix expression given in Equation (4.9) as follows

$$\frac{\partial F(y, \Sigma_t)}{\partial \sigma_{1t}^2} = -\frac{\sigma_{2t}^2}{2\Delta_t} F(y, \Sigma_t) + \frac{1}{2\Delta_t^2} \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} \left(\sigma_{2t}^2 a^2 - 2\sigma_{12t}\sigma_{2t}^2 ab + \sigma_{12t}^2 b^2\right) f(a, b, \Sigma_t) \, \mathrm{d}a\mathrm{d}b,$$

$$\frac{\partial F(y, \Sigma_t)}{\partial \sigma_{2t}^2} = -\frac{\sigma_1^2}{2\Delta_t} F(y, \Sigma_t) + \frac{1}{2\Delta_t^2} \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} \left(\sigma_{12t}^2 a^2 - 2\sigma_{12t}\sigma_1^2 ab + \sigma_{1t}^2 b^2\right) f(a, b, \Sigma_t) \, \mathrm{d}a\mathrm{d}b,$$

$$\begin{aligned} \frac{\partial F(y, \Sigma_t)}{\partial \sigma_{12t}} &= -\frac{\sigma_{12t}}{2\Delta_t} F(y, \Sigma_t) \\ &\quad + \frac{1}{2\Delta_t^2} \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} \left(-\sigma_{2t}^2 \sigma_{12t} a^2 + \left(\sigma_{12t}^2 + \sigma_{1t}^2 \sigma_{2t}^2\right) ab - \sigma_{1t}^2 \sigma_{12t} b^2\right) f(a, b, \Sigma_t) \, \mathrm{d}a\mathrm{d}b, \end{aligned}$$

with $\Delta_t = \sigma_{1t}^2\sigma_{2t}^2 - \sigma_{12t}^2$. If θ denotes the parameters vector of the conditional variance-covariance model (CCC, DCC, etc.), then for any $\theta \in \Theta$, we have

$$\frac{\partial F\left(y, \Sigma_t\right)}{\partial \theta} = \left(\text{vec}\left(\frac{\partial F\left(y, \Sigma_t\right)}{\partial \Sigma_t}\right)\right)' \text{vec}\left(\frac{\partial \Sigma_t}{\partial \theta}\right),$$

where vec(.) denotes the vectorization operator of a matrix. ■

4.8.5 Appendix E: Consistent estimates of R_{MES}, R_j, and γ_λ

We provide consistent estimates of R_{MES}, R_j, and γ_λ involved in the computation of the robust test statistics (see Theorem 2 and 4, and Appendix 4.8.8).

Estimation of $\mathbf{R_{MES}}$. As a starting point, we consider the approach that consists in evaluating the convergence of

$$\frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\frac{\partial H_t\left(\alpha,\widetilde{\theta}\right)}{\partial\theta}\middle|\Omega_{t-1}\right).$$

Deriving the cumulative joint violation process with respect to θ yields

$$\frac{\partial H_t\left(\alpha,\widetilde{\theta}\right)}{\partial\theta}=-\frac{\partial u_{12t}(\widetilde{\theta})}{\partial\theta}\mathbb{1}\left(u_{2t}\left(\widetilde{\theta}\right)\leq\alpha\right)+\left(1-u_{12t}\left(\widetilde{\theta}\right)\right)\frac{\partial\mathbb{1}\left(u_{2t}\left(\widetilde{\theta}\right)\leq\alpha\right)}{\partial\theta}.$$

Since $F_{Y_2}\left(VaR_{2t}\left(\alpha,\theta\right);\Omega_{t-1},\theta\right)=\alpha$ for any $\theta\in\Theta$, we have

$$-\frac{\partial u_{12t}\left(\theta\right)}{\partial\theta}=-\frac{1}{\alpha}\frac{\partial F(\widetilde{y}_t\left(\theta\right);\Omega_{t-1},\theta)}{\partial\theta},$$

with the vector $\widetilde{y}_t\left(\theta\right)$ defined as $\widetilde{y}_t\left(\theta\right)=\left(y_{1t},VaR_{2t}\left(\alpha,\theta\right)\right)'$. Taking sums and conditional expectations yield

$$\begin{aligned}\frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\frac{\partial H_t\left(\alpha,\widetilde{\theta}\right)}{\partial\theta}\middle|\Omega_{t-1}\right)&=-\frac{1}{\alpha n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\frac{\partial F(\widetilde{y}_t(\widetilde{\theta});\Omega_{t-1},\widetilde{\theta})}{\partial\theta}\mathbb{1}\left(u_{2t}\left(\widetilde{\theta}\right)\leq\alpha\right)\middle|\Omega_{t-1}\right)\\&\quad+\frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\left(1-u_{12t}(\widetilde{\theta})\right)\frac{\partial\mathbb{1}\left(u_{2t}\left(\widetilde{\theta}\right)\leq\alpha\right)}{\partial\theta}\middle|\Omega_{t-1}\right).\end{aligned}$$

Now assume that $T\to\infty$, under assumption A4, this quantity converges to

$$\begin{aligned}\frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\frac{\partial H_t\left(\alpha,\theta_0\right)}{\partial\theta}\middle|\Omega_{t-1}\right)&=-\frac{1}{\alpha n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\frac{\partial F(\widetilde{y}_t(\theta_0);\Omega_{t-1},\theta_0)}{\partial\theta}\mathbb{1}\left(u_{2t}(\theta_0)\leq\alpha\right)\middle|\Omega_{t-1}\right)\\&\quad+\frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\left(1-u_{12t}(\theta_0)\right)\frac{\partial\mathbb{1}\left(u_{2t}(\theta_0)\leq\alpha\right)}{\partial\theta}\middle|\Omega_{t-1}\right).\end{aligned}\tag{4.10}$$

Assume that $n\to\infty$, this quantity converges to

$$\frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\frac{\partial H_t\left(\alpha,\theta_0\right)}{\partial\theta}\middle|\Omega_{t-1}\right)\xrightarrow{p}R_{MES}=\mathbb{E}_0\left(\frac{\partial H_t\left(\alpha,\theta_0\right)}{\partial\theta}\right).\tag{4.11}$$

We deduce from Equations (4.10) and (4.11) that

$$R_{MES} = -\frac{1}{\alpha}\mathbb{E}_0\left(\frac{\partial F(\widetilde{y}_t(\theta_0);\Omega_{t-1},\theta_0)}{\partial\theta}\mathbb{1}\left(u_{2t}(\theta_0)\leq\alpha\right)\right)+\mathbb{E}_0\left((1-u_{12t}(\theta_0))\frac{\partial\mathbb{1}\left(u_{2t}(\theta_0)\leq\alpha\right)}{\partial\theta}\right), \tag{4.12}$$

Note that explicit formulas for $\partial F(\widetilde{y}_t(\theta_0);\Omega_{t-1},\theta_0)/\partial\theta$ in the left hand side of Equation (4.12) are available for some particular bivariate distributions (see Appendix 4.8.4 for the normal distribution). In any case, we can evaluate the derivative numerically. The right hand side of Equation (4.12) is trickier to implement due to the presence of the indicator function $\mathbb{1}\left(u_{2t}(\theta_0)\leq\alpha\right)$ that cannot be continuously derived (see Engle and Manganelli, 2004; Lambert et al., 2012, for similar issues). We first approximate the indicator function with a continuously differentiable function. Denote by $\mathbb{1}^{\oplus}(u\leq\tau)$ a continuous approximation function of $\mathbb{1}(u\leq\tau)$, where u is a random variable, and τ is a probability level,

$$\mathbb{1}^{\oplus}(u\leq\tau)=\int_0^{\tau}\frac{1}{h}\phi\left(\frac{u-v}{h}\right)\mathrm{d}v=\Phi\left(\frac{u}{h}\right)-\Phi\left(\frac{u-\tau}{h}\right).$$

with $\phi(.)$ a Gaussian kernel function, and $h>0$ a bandwidth parameter. Under a suitable form of the law of large number, we can replace the expectation operator by the sample mean, and a consistent estimator of R_{MES} of Equation (4.12) is given by

$$\widehat{R}_{MES}=-\frac{1}{\alpha n}\sum_{t=T+1}^{T+n}\frac{\partial F(\widetilde{y}_t(\widehat{\theta}_T);\Omega_{t-1},\widehat{\theta}_T)}{\partial\theta}\mathbb{1}(u_{2t}(\widehat{\theta}_T)\leq\alpha)+\frac{1}{n}\sum_{t=T+1}^{T+n}(1-u_{12t}(\widehat{\theta}_T))\frac{\partial\mathbb{1}^{\oplus}(u_{2t}(\widehat{\theta}_T)\leq\alpha)}{\partial\theta},$$

with

$$\begin{aligned}\frac{\partial F(\widetilde{y}_t(\widehat{\theta}_T);\Omega_{t-1},\widehat{\theta}_T)}{\partial\theta} &= \int_{-\infty}^{y_{1t}} f(u,VaR_{2t}(\alpha,\widehat{\theta}_T);\Omega_{t-1},\widehat{\theta}_T)\mathrm{d}u\times\frac{\partial VaR_{2t}(\alpha,\widehat{\theta}_T)}{\partial\theta}\\ &\quad+\int_{-\infty}^{y_{1t}}\int_{-\infty}^{VaR_{2t}(\alpha,\widehat{\theta}_T)}\frac{\partial f(u,v;\Omega_{t-1},\widehat{\theta}_T)}{\partial\theta}\mathrm{d}u\mathrm{d}v,\end{aligned}$$

$$\frac{\partial\mathbb{1}^{\oplus}(u_{2t}(\widehat{\theta}_T)\leq\alpha)}{\partial\theta}=\frac{1}{h}\left(\phi\left(\frac{u_{2t}(\widehat{\theta}_T)}{h}\right)-\phi\left(\frac{u_{2t}(\widehat{\theta}_T)-\alpha}{h}\right)\right)\times\frac{\partial F_{Y_{2t}}(y_{2t},\widehat{\theta}_T)}{\partial\theta}.$$

This approach requires to choose an appropriate value for the bandwidth parameter h for the kernel smoothing. Throughout this work, we select a bandwidth parameter $h=n^{-1}$ in line with the suggestion in Engle and Manganelli (2004).

Estimation of $\mathbf{R_j}$. The sketch of the proof is similar to that of the quantity R_{MES}. We use a continuously differentiable function in place of the indicator function, and we compute a consistent estimator of R_j by the law of large number. The quantity R_j is

given by

$$R_j = \frac{1}{\alpha\,(1/3-\alpha/4)}\mathbb{E}_0\left(\left(H_{t-j}(\alpha,\theta_0)-\alpha/2\right)\frac{\partial H_t(\alpha,\theta_0)}{\partial\theta}\right).$$

Given the above equation, we need a consistent estimator of $\partial H_t(\alpha,\theta_0)/\partial\theta$ to estimate R_j. As it has been done for R_{MES}, we substitute the indicator function by the kernel approximation, and estimate $\partial H_t(\alpha,\theta_0)/\partial\theta$ as follows

$$\frac{\partial H_t^{\oplus}(\alpha,\widehat{\theta}_T)}{\partial\theta} = -\frac{1}{\alpha}\frac{\partial F(\widetilde{y}_t(\widehat{\theta}_T);\Omega_{t-1},\widehat{\theta}_T)}{\partial\theta}\mathbb{1}(u_{2t}(\widehat{\theta}_T)\leq\alpha) + (1-u_{12t}(\widehat{\theta}_T))\frac{\partial\mathbb{1}^{\oplus}(u_{2t}(\widehat{\theta}_T)\leq\alpha)}{\partial\theta}.$$

Finally, by the law of large number, we get

$$\widehat{R}_j = \frac{1}{\alpha\,(1/3-\alpha/4)}\frac{1}{n-j}\sum_{t=T+1+j}^{T+n}\left(\left(H_{t-j}(\alpha,\theta_0)-\alpha/2\right)\frac{\partial H_t^{\oplus}(\alpha,\widehat{\theta}_T)}{\partial\theta}\right).$$

Estimation of $\boldsymbol{\gamma_\lambda}$. Let us rewrite $h_t(\alpha,\alpha,\widehat{\theta}_T)$ and $h_t(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\widehat{\theta}_T)$ in terms of their generalized errors,

$$h_t(\alpha,\alpha,\widehat{\theta}_T) = \mathbb{1}(u_{12t}(\alpha,\widehat{\theta}_T)\leq\alpha)\times\mathbb{1}(u_{2t}(\widehat{\theta}_T)\leq\alpha),$$

$$h_t(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\widehat{\theta}_T) = \mathbb{1}(u_{12t}(\beta,\widehat{\theta}_T)\leq\alpha)\times(\mathbb{1}(u_{2t}(\widehat{\theta}_T)\leq\beta_{\text{sup}})-\mathbb{1}(u_{2t}(\widehat{\theta}_T)\leq\beta_{\text{inf}})),$$

with

$$u_{12t}(\alpha,\widehat{\theta}_T) = F_{Y_1|Y_2\leq VaR_{2t}(\alpha,\widehat{\theta}_T)}(Y_{1t};\Omega_{t-1},\widehat{\theta}_T),$$

$$u_{12t}(\beta,\widehat{\theta}_T) = F_{Y_1|VaR_{2t}(\beta_{\text{inf}},\widehat{\theta}_T)\leq Y_2\leq VaR_{2t}(\beta_{\text{sup}},\widehat{\theta}_T)}(Y_{1t};\Omega_{t-1},\widehat{\theta}_T).$$

As a key point, we need estimates of R_C^α and R_C^β (see Appendix 4.8.8 for a definition). As for R_{MES} and R_j, the quantities R_C^α and R_C^β depend on indicator functions, and we propose to approximate them with kernel smoother. A consistent estimator of R_C^α is given by

$$\widehat{R}_C^\alpha = \frac{1}{n}\sum_{t=T+1}^{T+n}\left(\frac{\partial}{\partial\theta}\mathbb{1}^{\oplus}(u_{12t}(\alpha,\widehat{\theta}_T)\leq\alpha)\times\mathbb{1}(u_{2t}(\widehat{\theta}_T)\leq\alpha)+\frac{\partial}{\partial\theta}\mathbb{1}^{\oplus}(u_{2t}(\widehat{\theta}_T)\leq\alpha)\times\mathbb{1}(u_{12t}(\alpha,\widehat{\theta}_T)\leq\alpha)\right),$$

with $\frac{\partial}{\partial\theta}\mathbb{1}^{\oplus}(u_{12t}(\alpha,\widehat{\theta}_T)\leq\alpha) = \frac{1}{h}\left(\phi\left(\frac{u_{12t}(\alpha,\widehat{\theta}_T)}{h}\right)-\phi\left(\frac{u_{12t}(\alpha,\widehat{\theta}_T)-\alpha}{h}\right)\right)\frac{\partial}{\partial\theta}u_{12t}(\alpha,\widehat{\theta}_T)$, and $\frac{\partial}{\partial\theta}\mathbb{1}^{\oplus}(u_{2t}(\widehat{\theta}_T)\leq\alpha) = \frac{1}{h}\left(\phi\left(\frac{u_{2t}(\widehat{\theta}_T)}{h}\right)-\phi\left(\frac{u_{2t}(\widehat{\theta}_T)-\alpha}{h}\right)\right)\frac{\partial}{\partial\theta}u_{2t}(\widehat{\theta}_T)$. Similarly, a consistent estimator of R_C^β is given by

$$\widehat{R}_C^{\beta} = \frac{1}{n}\sum_{t=T+1}^{T+n}\Bigg[\frac{\partial}{\partial\theta}\mathbb{1}^{\oplus}(u_{12t}(\beta,\widehat{\theta}_T)\leq\alpha)\times\Big(\mathbb{1}(u_{2t}(\widehat{\theta}_T)\leq\beta_{\text{sup}})-\mathbb{1}(u_{2t}(\widehat{\theta}_T)\leq\beta_{\text{inf}})\Big)$$
$$+\,\mathbb{1}(u_{12t}(\beta,\widehat{\theta}_T)\leq\alpha)\times\left(\frac{\partial}{\partial\theta}\mathbb{1}^{\oplus}(u_{2t}(\widehat{\theta}_T)\leq\beta_{\text{sup}})-\frac{\partial}{\partial\theta}\mathbb{1}^{\oplus}(u_{2t}(\widehat{\theta}_T)\leq\beta_{\text{inf}})\right)\Bigg],$$

with $\frac{\partial}{\partial\theta}\mathbb{1}^{\oplus}(u_{12t}(\beta,\widehat{\theta}_T)\leq\alpha) = \frac{1}{h}\left(\phi\left(\frac{u_{12t}(\beta,\widehat{\theta}_T)}{h}\right)-\phi\left(\frac{u_{12t}(\beta,\widehat{\theta}_T)-\alpha}{h}\right)\right)\frac{\partial}{\partial\theta}u_{12t}(\beta,\widehat{\theta}_T)$, and $\frac{\partial}{\partial\theta}\mathbb{1}^{\oplus}(u_{2t}(\widehat{\theta}_T)\leq\beta_{\text{sup}})-\frac{\partial}{\partial\theta}\mathbb{1}^{\oplus}(u_{2t}(\widehat{\theta}_T)\leq\beta_{\text{inf}}) = \frac{1}{h}\left(\phi\left(\frac{u_{2t}(\widehat{\theta}_T)-\beta_{\text{inf}}}{h}\right)-\phi\left(\frac{u_{2t}(\widehat{\theta}_T)-\beta_{\text{sup}}}{h}\right)\right)\frac{\partial}{\partial\theta}u_{2t}(\widehat{\theta}_T)$. Given the results in Appendix 4.8.8, we get

$$\widehat{\gamma}_\lambda = \begin{pmatrix}\lambda\widehat{R_C^{\alpha}}'\widehat{\Sigma}_0\widehat{R_C^{\alpha}} & \lambda\widehat{R_C^{\alpha}}'\widehat{\Sigma}_0\widehat{R_C^{\beta}}\\ \lambda\widehat{R_C^{\alpha}}'\widehat{\Sigma}_0\widehat{R_C^{\beta}} & \lambda\widehat{R_C^{\beta}}'\widehat{\Sigma}_0\widehat{R_C^{\beta}}\end{pmatrix}.$$

4.8.6 Appendix F: Proof of Theorem 4

Proof. Define the lag-j autocovariance and autocorrelation of the cumulative joint violation $H_t(\alpha, \theta_0)$ for $j \geq 0$ by

$$\rho_j = \frac{\gamma_j}{\gamma_0} \quad \text{and} \quad \gamma_j = Cov\left(H_t(\alpha, \theta_0), H_{t-j}(\alpha, \theta_0)\right).$$

For ease of notations, we drop the dependence of γ_j and ρ_j on α and θ_0. The sample counterparts of γ_j and ρ_j based on the sample $\{H_t(\alpha, \theta_0)\}_{t=T+1}^{T+n}$ are

$$\rho_{nj} = \frac{\gamma_{nj}}{\gamma_{n0}} \quad \text{and} \quad \gamma_{nj} = \frac{1}{n-j} \sum_{t=T+1+j}^{T+n} \left(H_t(\alpha, \theta_0) - \alpha/2\right)\left(H_{t-j}(\alpha, \theta_0) - \alpha/2\right),$$

respectively. Similarly, define $\widehat{\rho}_{nj}$ and $\widehat{\gamma}_{nj}$, the sample counterparts of γ_j and ρ_j based on the sample $\{H_t(\alpha, \widehat{\theta}_T)\}_{t=T+1}^{T+n}$ with

$$\widehat{\rho}_{nj} = \frac{\widehat{\gamma}_{nj}}{\widehat{\gamma}_{n0}} \quad \text{and} \quad \widehat{\gamma}_{nj} = \frac{1}{n-j} \sum_{t=T+1+j}^{T+n} \left(H_t(\alpha, \widehat{\theta}_T) - \alpha/2\right)\left(H_{t-j}(\alpha, \widehat{\theta}_T) - \alpha/2\right).$$

The sketch of the proof is similar to that used for Theorem 2. Under Assumptions A1-A4, the continuous mapping theorem implies that

$$\sqrt{n-j}\left(\widehat{\rho}_{nj} - \mathbb{E}\left(\widehat{\rho}_{nj} \middle| \Omega_{t-1}\right)\right) = \sqrt{n-j}\left(\rho_{nj} - \mathbb{E}\left(\rho_{nj} \middle| \Omega_{t-1}\right)\right) + o_p(1).$$

Rearranging these terms gives

$$\sqrt{n-j}\left(\widehat{\rho}_{nj} - \rho_{nj}\right) = \sqrt{n-j}\mathbb{E}\left(\widehat{\rho}_{nj} - \rho_{nj} \middle| \Omega_{t-1}\right) + o_p(1). \tag{4.13}$$

The mean value theorem implies that

$$\widehat{\rho}_{nj} = \rho_{nj} + \left(\widehat{\theta}_T - \theta_0\right)' \frac{\partial \widetilde{\rho}_{nj}}{\partial \theta},$$

with $\widetilde{\rho}_{nj}$ the lag-j autocorrelation of the process $H_t(\alpha, \widetilde{\theta})$, where $\widetilde{\theta}$ is an intermediate point between θ_0 and $\widehat{\theta}_T$. Define $\lambda = (n-j)/T$, Equation (4.13) becomes

$$\sqrt{n-j}\left(\widehat{\rho}_{nj} - \rho_j\right) = \sqrt{n-j}\left(\rho_{nj} - \rho_j\right) + \sqrt{\lambda}\sqrt{T}(\widehat{\theta}_T - \theta_0)'\mathbb{E}\left(\frac{\partial \widetilde{\rho}_{nj}}{\partial \theta}\middle| \Omega_{t-1}\right) + o_p(1).$$

Under Assumption A4, when $T \to \infty$ we have $\widetilde{\theta} \xrightarrow{p} \theta_0$. Then, we get for $j \neq 0$

$$\mathbb{E}\left(\frac{\partial \widetilde{\rho}_{nj}}{\partial \theta}\middle| \Omega_{t-1}\right) \xrightarrow{p} \mathbb{E}\left(\frac{1}{\gamma_{n0}} \frac{\partial \gamma_{nj}}{\partial \theta}\middle| \Omega_{t-1}\right) - \mathbb{E}\left(\frac{\rho_{nj}}{\gamma_{n0}} \frac{\partial \gamma_{n0}}{\partial \theta}\middle| \Omega_{t-1}\right).$$

When $n \to \infty$, $\gamma_{n0} \xrightarrow{p} \gamma_0$ and $\rho_{nj} \xrightarrow{p} \rho_j$. Since $\mathbb{E}(\,(H_t\,(\alpha,\theta_0) - \alpha/2)\partial H_{t-j}(\alpha,\theta_0)/\partial\theta|\,\Omega_{t-1}) = \partial H_{t-j}(\alpha,\theta_0)/\partial\theta\,\mathbb{E}(\,(H_t\,(\alpha,\theta_0) - \alpha/2)|\,\Omega_{t-1}) = 0$ for $j > 0$, we get

$$\mathbb{E}\left(\frac{\partial\widetilde{\rho}_{nj}}{\partial\theta}\middle|\,\Omega_{t-1}\right) \xrightarrow{p} R_j - 2\rho_j R_0,$$

with

$$R_j = \frac{1}{\gamma_0}\mathbb{E}_0\left(\frac{\partial\gamma_{nj}}{\partial\theta}\right) = \frac{1}{\gamma_0}\mathbb{E}_0\left(\left(H_{t-j}(\alpha,\theta_0) - \alpha/2\right)\frac{\partial H_t(\alpha,\theta_0)}{\partial\theta}\right),$$

and $\gamma_0 = \alpha\,(1/3 - \alpha/4)$. Under the null $\rho_j = 0$ for $j = 1, ..., m$. Therefore

$$\sqrt{n}\widehat{\rho}_{nj} = \sqrt{n}\rho_{nj} + \sqrt{\lambda}\sqrt{T}R_j'(\widehat{\theta}_T - \theta_0) + o_p\,(1)\,.$$

Notice that $\sqrt{n}\,(\rho_{n1}, ..., \rho_{nm})' \xrightarrow{d} \mathcal{N}\,(0, I_m)$ and the covariance between the first term and $\sqrt{T}(\widehat{\theta}_T - \theta_0)$ is 0 as $\widehat{\theta}_T$ depends on the in-sample observations and the correlation ρ_{nj} depends on the out-of-sample observations. Denote $\widehat{\rho}_n^{(m)}$ the vector $(\widehat{\rho}_{n1}, ..., \widehat{\rho}_{nm})'$. Under Assumptions A1-A4, we have

$$\sqrt{n}\widehat{\rho}_n^{(m)} \xrightarrow{d} \mathcal{N}\,(0, \Delta)\,,$$

with the ij-th element of Δ given by

$$\Delta_{ij} = \delta_{ij} + \lambda R_i'\Sigma_0 R_j,$$

$$R_j = \frac{1}{\alpha\,(1/3 - \alpha/4)}\mathbb{E}_0\left(\left(H_{t-j}(\alpha,\theta_0) - \alpha/2\right)\frac{\partial H_t(\alpha,\theta_0)}{\partial\theta}\right),$$

$\forall\,(i,j) \in \{1, ..., m\}^2$, where δ_{ij} is a dummy variable that takes a value 1 if $i = j$ and 0 otherwise. We can write $\Delta = Q\Lambda Q'$, where Q is an orthogonal matrix, and Λ is a diagonal matrix with elements $\{\pi_j\}_{j=1}^m$. So, we have

$$Q'\sqrt{n}\widehat{\rho}^{(m)} \xrightarrow{d} \mathcal{N}\,(0, \Lambda)\,.$$

Finally

$$IND_{MES} = n\sum_{j=1}^{m}\widehat{\rho}_{nj}^2 \xrightarrow{d} \sum_{j=1}^{m}\pi_j Z_j^2,$$

where $\{\pi_j\}_{j=1}^m$ are the eigenvalues of the matrix Δ and $\{Z_j\}_{j=1}^m$ are independent standard normal variables. ■

4.8.7 Appendix G: Backtesting MES-based risk measure

Proof. Consider a MES-based risk measure expressed as

$$RM_{1t}(h) = \int_0^1 g_t(Co\widetilde{V}aR_{1t+h}(\beta,\alpha,\theta_0), X_t)\mathrm{d}\beta,$$

with

$$\Pr\left(\widetilde{Y}_{1t+h} \leq Co\widetilde{V}aR_{1t+h}(\beta,\alpha,\theta_0)|\widetilde{Y}_{2t+h} \leq \widetilde{V}aR_{2t+h}(\alpha,\theta_0);\Omega_t\right) = \beta.$$

In order to backtest the CoVaR and the MES, we define a joint violation of the (β,α)-CoVaR of $\widetilde{Y}_{1t+h}$ and the α-VaR of $\widetilde{Y}_{2t+h}$. This violation process is represented by the following binary variable

$$h_t(\alpha,\beta,\theta_0) = \mathbb{1}(\widetilde{Y}_{1t+h} \leq Co\widetilde{V}aR_{1t}(\beta,\alpha,\theta_0)) \times \mathbb{1}(\widetilde{Y}_{2t+h} \leq \widetilde{V}aR_{2t+h}(\alpha,\theta_0)).$$

Denote by $h_t(\alpha,\beta,\theta_0,X_t)$ the violation process associated to $g_t(Co\widetilde{V}aR_{1t+h}(\beta,\alpha,\theta_0), X_t)$ and used to backtest the MES-based risk measure, for instance the SRISK. If the function $g_t(.)$ is monotonic decreasing with the MES (as it is the case for the SRISK given our sign convention for the MES), we have

$$h_t(\alpha,\beta,\theta_0,X_t) = \mathbb{1}(g_t(\widetilde{Y}_{1t+h},X_t) \geq g_t(Co\widetilde{V}aR_{1t+h}(\beta,\alpha,\theta_0),X_t)) \times \mathbb{1}(\widetilde{Y}_{2t+h} \leq \widetilde{V}aR_{2t+h}(\alpha,\theta_0)).$$

The violation processes $h_t(\alpha,\beta,\theta_0)$ and $h_t(\alpha,\beta,\theta_0,X_t)$ are identical, in the sense that they take a value 1 at the same date. As a consequence, the cumulative joint violation processes used for backtesting the MES and a MES-based risk measure are identical.

$$H_t(\alpha,\theta_0,X_t) = \int_0^1 h_t(\alpha,\beta,\theta_0,X_t)\mathrm{d}\beta = \int_0^1 h_t(\alpha,\beta,\theta_0)\mathrm{d}\beta = H_t(\alpha,\theta_0).$$

The unconditional coverage test for a MES-based risk measure, say SRISK, corresponds to the null hypothesis

$$\mathrm{H}_{0,UC}^{\mathrm{SRISK}} : \mathbb{E}(H_t(\alpha,\theta_0,X_t)) = \alpha/2.$$

The corresponding test statistic UC_{SRISK} is given by

$$UC_{\mathrm{SRISK}} = \frac{\sqrt{n}\left(\bar{H}(\alpha,\widehat{\theta}_T,X) - \alpha/2\right)}{\sqrt{\alpha(1/3-\alpha/4)}},$$

with $\bar{H}(\alpha, \widehat{\theta}_T, X)$ the out-of-sample mean of $H_t(\alpha, \widehat{\theta}_T, X_t)$. As $H_t(\alpha, \widehat{\theta}_T, X_t) = H_t(\alpha, \widehat{\theta}_T)$, $\forall t$, this statistic is equivalent to that used for backtesting the MES

$$UC_{\text{MES}} = \frac{\sqrt{n}\left(\bar{H}(\alpha, \widehat{\theta}_T) - \alpha/2\right)}{\sqrt{\alpha\left(1/3 - \alpha/4\right)}}.$$

■

4.8.8 Appendix H: Backtesting ΔCoVaR

Proof. The first step of the proof consists in evaluating the two first conditional moments of $\mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0)$. The Bayes theorem implies that

$$\mathbb{E}\left(h_t(\alpha, \alpha, \theta_0)\,|\Omega_{t-1}\right) = \alpha^2,$$

$$\mathbb{E}\left(h_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0)\,|\Omega_{t-1}\right) = \alpha(\beta_{\text{sup}} - \beta_{\text{inf}}),$$

leading to

$$\mathbb{E}\left(\mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0)\,|\Omega_{t-1}\right) \equiv \mu = \begin{pmatrix} \alpha^2 \\ \alpha(\beta_{\text{sup}} - \beta_{\text{inf}}) \end{pmatrix}. \tag{4.14}$$

According to Equation (4.14), $h_t(\alpha, \alpha, \theta_0)$ and $h_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0)$ are Bernouilli distributed with conditional success probabilities α^2 and $\alpha(\beta_{\text{sup}} - \beta_{\text{inf}})$, and conditional variances given by

$$\mathbb{V}\left(h_t(\alpha, \alpha, \theta_0)\,|\Omega_{t-1}\right) = \alpha^2\left(1 - \alpha^2\right),$$

$$\mathbb{V}\left(h_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0)\,|\Omega_{t-1}\right) = \alpha\left(\left(\beta_{\text{sup}} - \beta_{\text{inf}}\right)\left(1 - \alpha\left(\beta_{\text{sup}} - \beta_{\text{inf}}\right)\right)\right).$$

In order to compute the statistic $UC_{\Delta CoVaR}$, we have to determine the general expression of the conditional variance-covariance matrix of $\mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0)$, and in particular, we need the conditional covariance between the two violations,

$$\begin{aligned}\text{Cov}\left(h_t(\alpha, \alpha, \theta_0), h_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0)|\Omega_{t-1}\right) = {} & \mathbb{E}\left(h_t(\alpha, \alpha, \theta_0)\, h_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0)|\Omega_{t-1}\right) \\ & - \mathbb{E}\left(h_t(\alpha, \alpha, \theta_0)|\Omega_{t-1}\right)\mathbb{E}\left(h_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0)|\Omega_{t-1}\right).\end{aligned} \tag{4.15}$$

Since $\alpha < \beta_{\text{inf}}$, $Y_{2t} \leq VaR_{2t}(\alpha, \theta_0)$ and $VaR_{2t}(\beta_{\text{inf}}, \theta_0) \leq Y_{2t} \leq VaR_{2t}(\beta_{\text{sup}}, \theta_0)$ are incompatible events, we get $h_t(\alpha, \alpha, \theta_0) \times h_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0) = 0$ implying that the first expectation in Equation (4.15) is 0. We thus have

$$\text{Cov}\left(h_t(\alpha, \alpha, \theta_0), h_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0)|\Omega_{t-1}\right) = -\alpha^3(\beta_{\text{sup}} - \beta_{\text{inf}}),$$

and the conditional variance-covariance matrix of $\mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0)$ is given by

$$\mathbb{V}\left(\mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0)\,|\Omega_{t-1}\right) \equiv \gamma = \begin{pmatrix} \alpha^2(1-\alpha^2) & -\alpha^3(\beta_{\text{sup}} - \beta_{\text{inf}}) \\ -\alpha^3(\beta_{\text{sup}} - \beta_{\text{inf}}) & \alpha(\beta_{\text{sup}} - \beta_{\text{inf}})(1 - \alpha(\beta_{\text{sup}} - \beta_{\text{inf}})) \end{pmatrix}.$$

With the two first conditional moments of $\mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0)$ in hand, we now turn to the identification of its asymptotic distribution. Equation (4.14) implies that the sequence

$\{\mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0) - \mu\}_{t=1}^{\infty}$ is a mds for any $(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}) \in [0,1]^3$ with

$$\mathbb{E}\left(\mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0) - \mu | \Omega_{t-1}\right) = 0.$$

As a consequence, the Lindeberg-Levy central limit theorem implies that

$$\sqrt{n}\left(\bar{\mathbf{h}}(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0) - \mu\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \gamma), \tag{4.16}$$

with $\bar{\mathbf{h}}(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0) = 1/n \sum_{t=T+1}^{T+n} \mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0)$. From Equation (4.16), we get immediately that

$$UC_{\Delta CoVaR} = n\left(\bar{\mathbf{h}}(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0) - \mu\right)' \gamma^{-1} (\bar{\mathbf{h}}(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0) - \mu) \xrightarrow{d} \chi(2).$$

We now turn to the case of the feasible test statistic $UC_{\Delta CoVaR} \equiv UC_{\Delta CoVaR}(\widehat{\theta}_T)$ and its asymptotic distribution. The sketch of the proof is similar to that of Theorem 2. First, rewrite

$$\sqrt{n}(\bar{\mathbf{h}}(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \widehat{\theta}_T) - \mu) = \frac{1}{\sqrt{n}} \sum_{t=T+1}^{T+n} (\mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \widehat{\theta}_T) - \mathbb{E}(\mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0) | \Omega_{t-1})).$$

The continuous mapping theorem implies that

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{t=T+1}^{T+n} (\mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \widehat{\theta}_T) - \mathbb{E}(\mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \widehat{\theta}_T)) | \Omega_{t-1}) = \\
&\qquad \frac{1}{\sqrt{n}} \sum_{t=T+1}^{T+n} (\mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0) - \mathbb{E}(\mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0) | \Omega_{t-1})) + o_p(1).
\end{aligned}$$

Rearranging these terms gives

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{t=T+1}^{T+n} (\mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \widehat{\theta}_T) - \mathbb{E}(\mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0) | \Omega_{t-1})) = \\
&\qquad \frac{1}{\sqrt{n}} \sum_{t=T+1}^{T+n} (\mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0) - \mathbb{E}(\mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0) | \Omega_{t-1})) \\
&\qquad + \frac{1}{\sqrt{n}} \sum_{t=T+1}^{T+n} \mathbb{E}(\mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \widehat{\theta}_T) - \mathbf{h}_t(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \theta_0) | \Omega_{t-1}) + o_p(1).
\end{aligned} \tag{4.17}$$

The mean value theorem implies that

$$\begin{aligned}\mathbf{h}_t(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\widehat{\theta}_T) &\equiv \begin{pmatrix} h_t(\alpha,\alpha,\widehat{\theta}_T) \\ h_t(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\widehat{\theta}_T)\end{pmatrix} \\ &= \begin{pmatrix} h_t(\alpha,\alpha,\theta_0) \\ h_t(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\theta_0)\end{pmatrix} + \begin{pmatrix} \left(\widehat{\theta}_T-\theta_0\right)' \frac{\partial h_t(\alpha,\alpha,\tilde{\theta})}{\partial\theta} \\ \left(\widehat{\theta}_T-\theta_0\right)' \frac{\partial h_t(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\tilde{\theta})}{\partial\theta}\end{pmatrix},\end{aligned}$$

where $\tilde{\theta}$ is an intermediate point between θ_0 and $\widehat{\theta}_T$. Equation (4.17) becomes

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=T+1}^{T+n}(\mathbf{h}_t(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\widehat{\theta}_T)-\mathbb{E}(\mathbf{h}_t(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\theta_0)|\Omega_{t-1})) &= \frac{1}{\sqrt{n}}\sum_{t=T+1}^{T+n}(\mathbf{h}_t(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\theta_0)-\mu) \\ &+ \begin{pmatrix} \sqrt{\lambda}\sqrt{T}\left(\widehat{\theta}_T-\theta_0\right)' \frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\frac{\partial h_t(\alpha,\alpha,\tilde{\theta})}{\partial\theta}|\Omega_{t-1}\right) \\ \sqrt{\lambda}\sqrt{T}\left(\widehat{\theta}_T-\theta_0\right)' \frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\frac{\partial h_t(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\tilde{\theta})}{\partial\theta}|\Omega_{t-1}\right)\end{pmatrix} + o_p(1).\end{aligned} \tag{4.18}$$

Assume that $T\to\infty$, $n\to\infty$ and $n/T\to\lambda$ with $0\le\lambda<\infty$. Under the null hypothesis $\text{H}_{0,UC}$, the first term of Equation (4.18) converges in distribution to a bivariate normal distribution with mean $\mathbf{0}$ and variance γ. The covariance between the first term and $\sqrt{T}\left(\widehat{\theta}_T-\theta_0\right)$is 0 as $\widehat{\theta}_T$ depends on the in-sample observations and the summand in the first term is for out-of-sample observations. Under Assumption A4, $\tilde{\theta}\to\theta_0$, and since the two derivatives $\partial h_t(\alpha,\alpha,\theta_0)/\partial\theta$ and $\partial h_t(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\theta_0)/\partial\theta$ are also mds, we have

$$\frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\frac{\partial h_t(\alpha,\alpha,\tilde{\theta})}{\partial\theta}|\Omega_{t-1}\right)\xrightarrow{p} R_C^{\alpha}=\mathbb{E}_0\left(\frac{\partial h_t(\alpha,\alpha,\theta_0)}{\partial\theta}\right),$$

$$\frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\frac{\partial h_t(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\tilde{\theta})}{\partial\theta}|\Omega_{t-1}\right)\xrightarrow{p} R_C^{\beta}=\mathbb{E}_0\left(\frac{\partial h_t(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\theta_0)}{\partial\theta}\right).$$

where $\mathbb{E}_0(.)$ denotes the expectation with respect to the true distribution of $h_t(\alpha,\alpha,\theta_0)$ and $h_t(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\theta_0)$. As a consequence, we have

$$\begin{pmatrix} \sqrt{\lambda}\sqrt{T}\left(\widehat{\theta}_T-\theta_0\right)' \frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\frac{\partial h_t(\alpha,\alpha,\tilde{\theta})}{\partial\theta}|\Omega_{t-1}\right) \\ \sqrt{\lambda}\sqrt{T}\left(\widehat{\theta}_T-\theta_0\right)' \frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}\left(\frac{\partial h_t(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\tilde{\theta})}{\partial\theta}|\Omega_{t-1}\right)\end{pmatrix}\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\gamma_\lambda\right),$$

where $\gamma_\lambda=\begin{pmatrix}\lambda R_C^{\alpha\prime}\Sigma_0 R_C^{\alpha} & \lambda R_C^{\alpha\prime}\Sigma_0 R_C^{\beta} \\ \lambda R_C^{\alpha\prime}\Sigma_0 R_C^{\beta} & \lambda R_C^{\beta\prime}\Sigma_0 R_C^{\beta}\end{pmatrix}$, and we can deduce that

$$\sqrt{n}\left(\bar{\mathbf{h}}\left(\alpha,\beta_{\text{inf}},\beta_{\text{sup}},\widehat{\theta}_T\right)-\mu\right)\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\gamma+\gamma_\lambda\right).$$

The feasible test statistic $UC^{C}_{\Delta CoVaR}$ that takes into account the presence of estimation risk is then given by

$$UC^{C}_{\Delta CoVaR} = n\left(\bar{\mathbf{h}}\left(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \widehat{\theta}_T\right) - \mu\right)' (\gamma + \widehat{\gamma}_\lambda)^{-1} (\bar{\mathbf{h}}\left(\alpha, \beta_{\text{inf}}, \beta_{\text{sup}}, \widehat{\theta}_T\right) - \mu),$$

where $\widehat{\gamma}_\lambda$ denotes a consistent estimator of γ_λ (see Appendix 4.8.5). The robust statistic $UC^{C}_{\Delta CoVaR}$ converges to a chi-square distribution with two degrees of freedom whatever the value of λ, and is thus free of estimation risk. ■

Table 4.3: Empirical rejection rates for backtesting ΔCoVaR at 5% nominal level (marginal case)

		$T = 250$, $n = 250$		$T = 250$, $n = 500$	
		$UC_{\Delta CoVaR}$	$UC^{C}_{\Delta CoVaR}$	$UC_{\Delta CoVaR}$	$UC^{C}_{\Delta CoVaR}$
H_0		0.0766	0.0458	0.1051	0.0572
$H_1(A_1)$	$\tau = 25\%$	0.3100	0.3637	0.4881	0.4966
	$\tau = 50\%$	0.9002	0.9300	0.9941	0.9928
	$\tau = 75\%$	1.0000	0.9999	1.0000	1.0000
$H_1(A_2)$	$\tau = 25\%$	0.0585	0.0599	0.0555	0.0686
	$\tau = 50\%$	0.0681	0.0748	0.1001	0.1292
	$\tau = 75\%$	0.0938	0.1492	0.2513	0.3323
$H_1(A_3)$	$\tau = 25\%$	0.1027	0.1141	0.1449	0.1799
	$\tau = 50\%$	0.2669	0.3026	0.4957	0.5299
	$\tau = 75\%$	0.5644	0.5917	0.8471	0.8253
		$T = 500$, $n = 250$		$T = 500$, $n = 500$	
		$UC_{\Delta CoVaR}$	$UC^{C}_{\Delta CoVaR}$	$UC_{\Delta CoVaR}$	$UC^{C}_{\Delta CoVaR}$
H_0		0.0601	0.0396	0.0806	0.0525
$H_1(A_1)$	$\tau = 25\%$	0.3534	0.3922	0.5325	0.5341
	$\tau = 50\%$	0.9336	0.9467	0.9974	0.9969
	$\tau = 75\%$	1.0000	1.0000	1.0000	1.0000
$H_1(A_2)$	$\tau = 25\%$	0.0493	0.0501	0.0541	0.0634
	$\tau = 50\%$	0.0735	0.0798	0.1180	0.1278
	$\tau = 75\%$	0.1313	0.1635	0.3234	0.3444
$H_1(A_3)$	$\tau = 25\%$	0.1109	0.1210	0.1732	0.1886
	$\tau = 50\%$	0.3085	0.3453	0.5299	0.5525
	$\tau = 75\%$	0.5945	0.6273	0.8629	0.8613
		$T = 2500$, $n = 250$		$T = 2500$, $n = 500$	
		$UC_{\Delta CoVaR}$	$UC^{C}_{\Delta CoVaR}$	$UC_{\Delta CoVaR}$	$UC^{C}_{\Delta CoVaR}$
H_0		0.0547	0.0468	0.0642	0.0522
$H_1(A_1)$	$\tau = 25\%$	0.3769	0.3870	0.5723	0.5760
	$\tau = 50\%$	0.9414	0.9421	0.9974	0.9973
	$\tau = 75\%$	1.0000	1.0000	1.0000	1.0000
$H_1(A_2)$	$\tau = 25\%$	0.0493	0.0529	0.0612	0.0677
	$\tau = 50\%$	0.0715	0.0799	0.1234	0.1319
	$\tau = 75\%$	0.1336	0.1512	0.3164	0.3457
$H_1(A_3)$	$\tau = 25\%$	0.1100	0.1196	0.1864	0.1977
	$\tau = 50\%$	0.3074	0.3349	0.5794	0.5933
	$\tau = 75\%$	0.6112	0.6361	0.8943	0.9001

Note: This table displays the Monte Carlo results for the ΔCoVaR associated to a time-invariant setting as defined in Subsection "Backtesting marginal MES". $UC_{\Delta CoVaR}$ denotes the unconditional coverage test for ΔCoVaR. The probability levels are set to $\alpha = 0.05$, $\beta_{\text{inf}} = 0.25$, and $\beta_{\text{sup}} = 0.75$. The test statistic robust to estimation risk is superscripted by C. Reported powers are sized-corrected.

4.8.9 Appendix I: List of tickers

Tickers and company names

AET	Aetna	HBAN	Huntington Bancshares
AFL	Aflac	ICE	Intercontinental Exchange
ALL	Allstate Corp	JNS	Janus Capital
ABK	Ambac Financial Group	JPM	JP Morgan Chase
ACAS	American Capital	KEY	Keycorp
AXP	American Express	LM	Legg Mason
AIG	American International Group	LEH	Lehman Brothers
AMP	Ameriprise financial	LNC	Lincoln National
AMTD	TD Ameritrade	L	Loews
ANTM	Anthem	MI	Marshall & Ilsley
AON	Aon Corp	MMC	Marsh & McLennan
AIZ	Assurant	MA	Mastercard
MTB	M & T Bank Corp	MBI	MBIA
BAC	Bank of America	MER	Merrill Lynch
BK	Bank of New York Mellon	MET	Metlife
BBT	BB&T	MS	Morgan Stanley
BSC	Bear Stearns	NCC	National City Corp
WRB	W.R. Berkley Corp	NYCB	New York Community Bancorp
BRK	Berkshire Hathaway	NTRS	Northern Trust
BLK	Blackrock	NMX	Nymex Holdings
COF	Capital One Financial	NYX	NYSE Euronext
BOT	CBOT Holdings	PBCT	Peoples United Financial
CBG	CBRE Group	PNC	PNC Financial Services
CB	Chubb Corp	PFG	Principal Financial Group
CI	CIGNA Corp	PGR	Progressive
CINF	Cincinnati Financial Corp	PRU	Prudential Financial
CIT	CIT Group	RF	Regions Financial
C	Citigroup	SAF	Safeco
CME	CME Group	SCHW	Schwab Charles
CNA	CNA Financial Corp	SEIC	SEI Investments Company
CMA	Comerica	SLM	SLM Corp
CBH	Commerce Bancorp	SOV	Sovereign Bancorp
CBSS	Compass Bancshares	STT	State Street
CFC	Countrywide Financial	STI	Suntrust Banks
CVH	Coventry Health Care	SNV	Synovus Financial
AGE	A.G. Edwards	TMK	Torchmark
ETFC	E-Trade Financial	TRV	Travelers
FNM	Fannie Mae	TROW	T. Rowe Price
FNF	Fidelity National Financial	UB	Unionbancal Corp
FITB	Fifth Third Bancorp	UNH	Unitedhealth Group
BEN	Franklin Resources	UNM	Unum Group
FRE	Freddie Mac	USB	US Bancorp
GNW	Genworth Financial	WB	Wachovia
GS	Goldman Sachs	WM	Washington Mutual
HIG	Hartford Financial Group	WFC	Wells Fargo & Co
HNT	Health Net	WU	Western Union
HCBK	Hudson City Bancorp	ZION	Zion
HUM	Humana		

4.8.10 Appendix J: Robustness checks for backtesting short-term risk measures

Figure 4.13: UC backtest (recursive estimation, and $n = 500$)

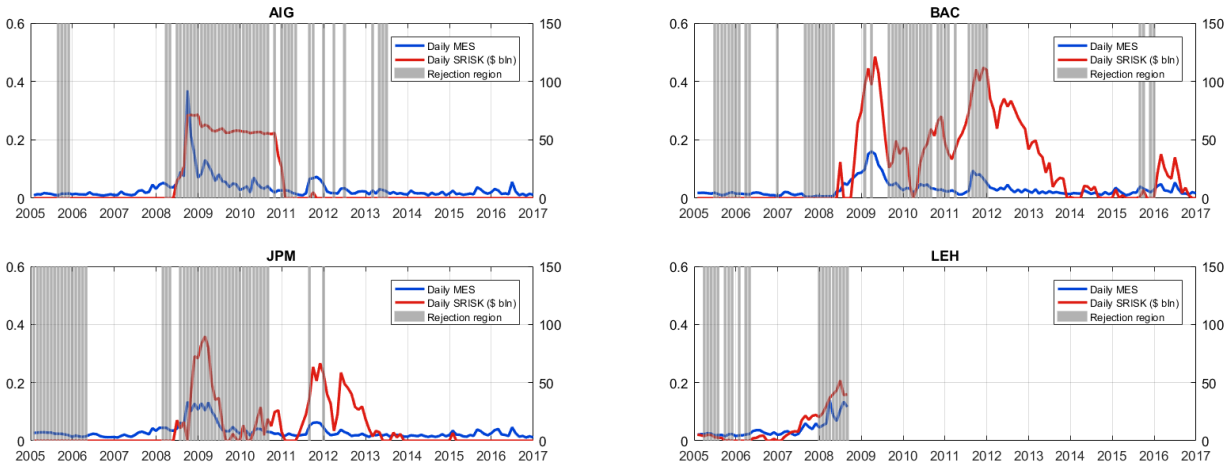


Figure 4.14: UC backtest (rolling estimation, $T = 500$, and $n = 250$)

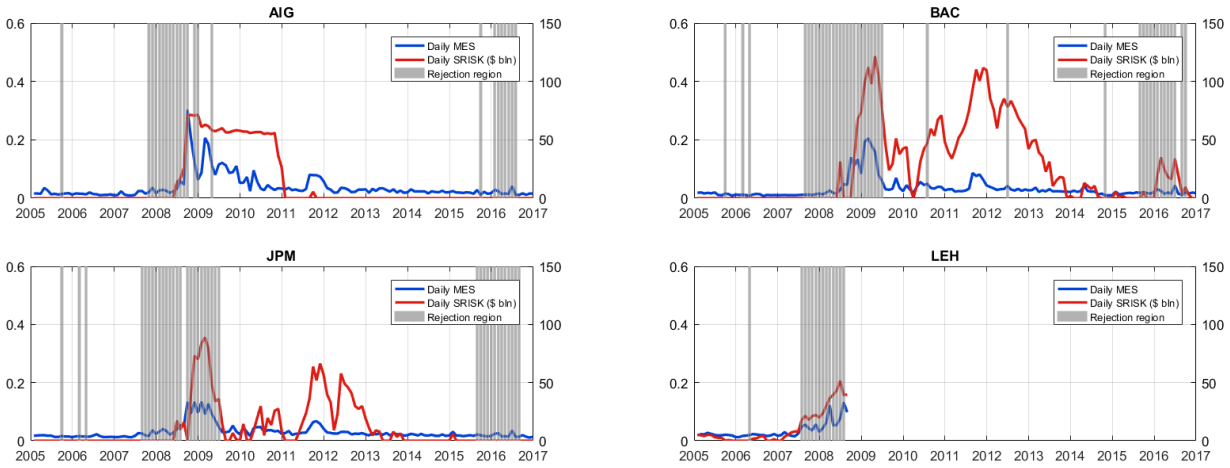


Figure 4.15: UC backtest (rolling estimation, $T = 500$, and $n = 500$)

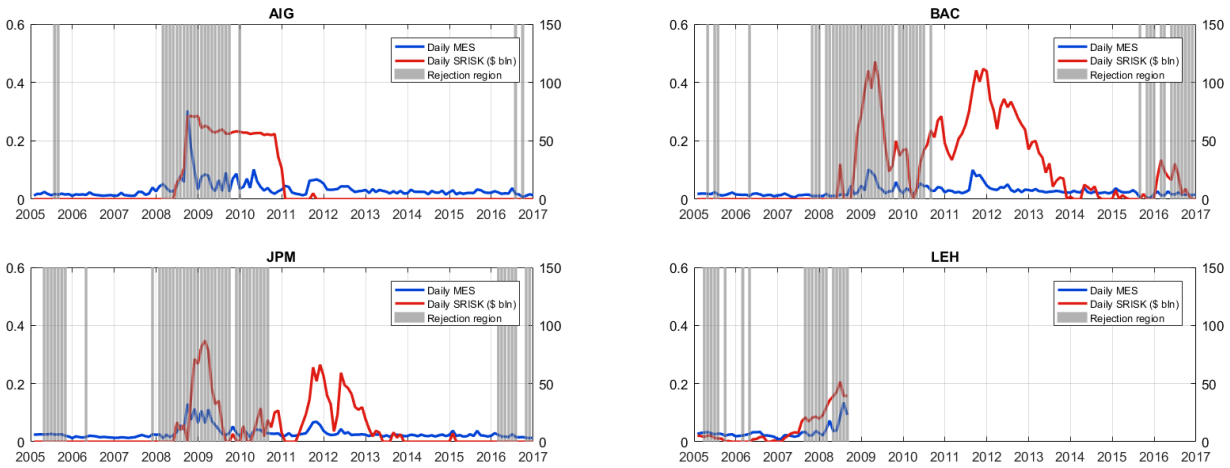


Figure 4.16: IND backtest (recursive estimation, $n = 500$, and $m = 5$)

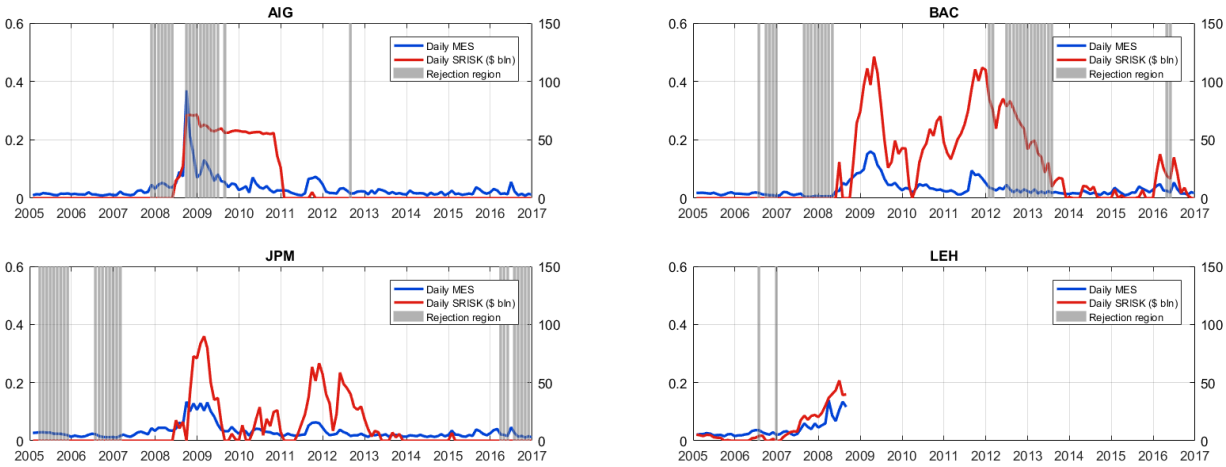


Figure 4.17: IND backtest (rolling estimation, $T = 500$, $n = 250$, and $m = 5$)

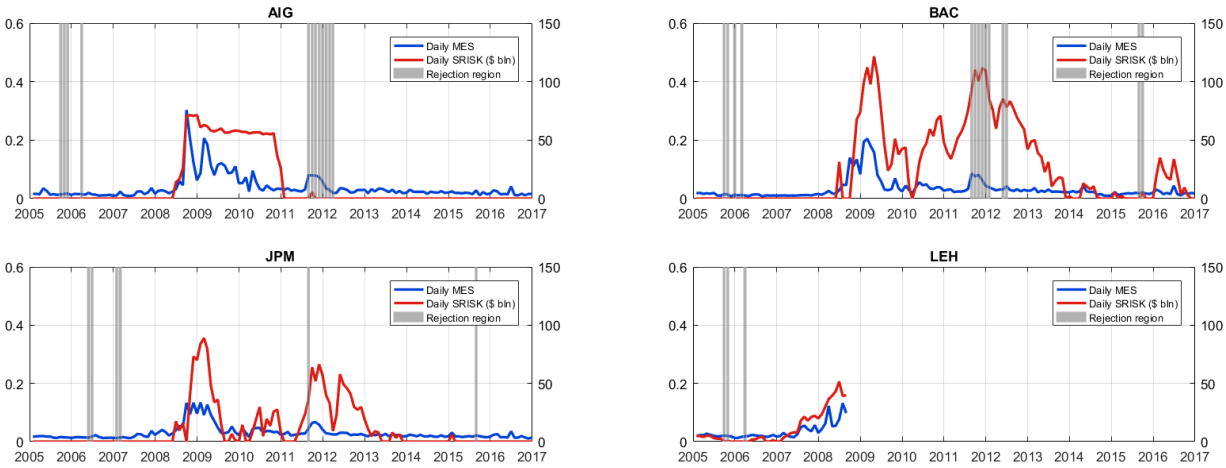


Figure 4.18: IND backtest (rolling estimation, $T = 500$, $n = 500$, and $m = 5$)

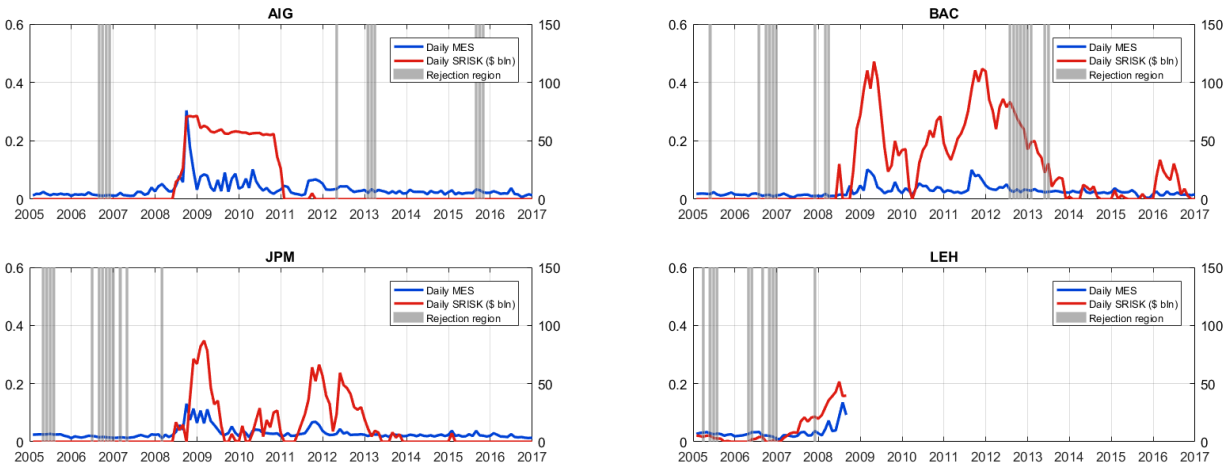


Figure 4.19: Rejection rates of the UC and IND backtests (rolling estimation, $T = 500$, $n = 250$, and $m = 5$)

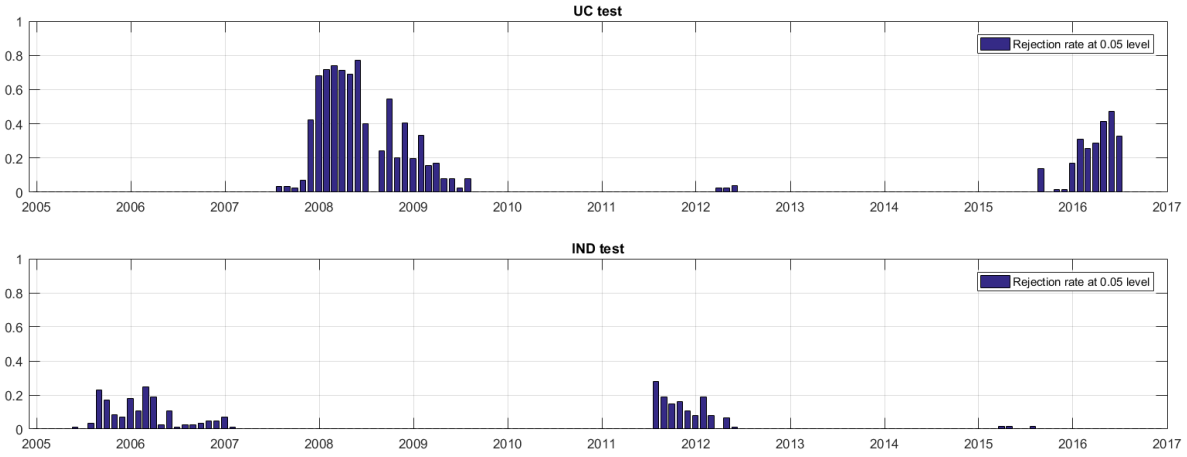
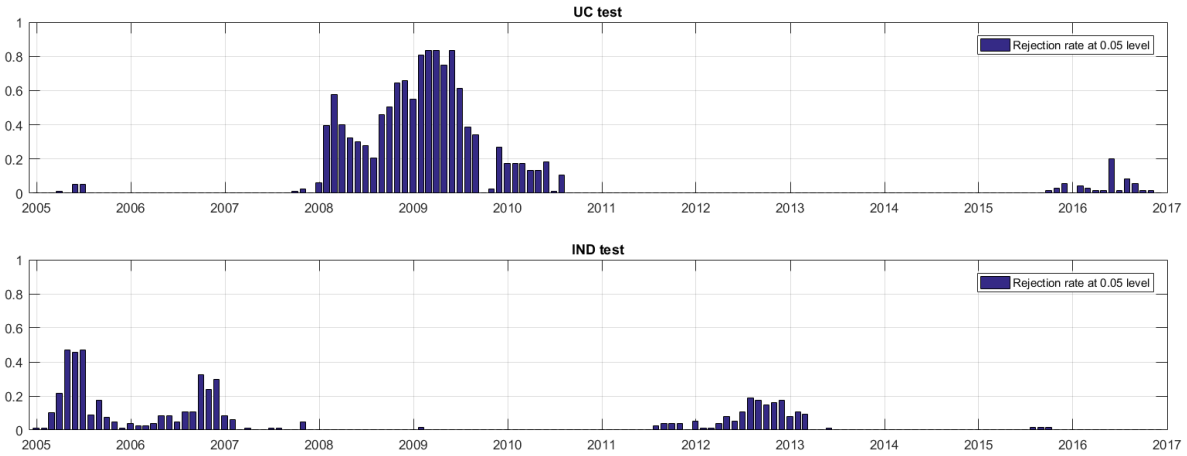


Figure 4.20: Rejection rates of the UC and IND backtests (rolling estimation, $T = 500$, $n = 500$, and $m = 5$)



4.8.11 Appendix K: Computation of LRMES

In this section, we describe the methodology used to compute the MES, hereafter Long Run MES (LRMES), and SRISK over an horizon $h > 1$. The computation approach differ from that used for the daily predictions, since the capital shortfall is calculated from the multi-period arithmetic returns instead of the arithmetic daily returns. The technique is thus closely related to that proposed by Brownlees and Engle (2017). Because we do not have closed form distribution for the multi-period arithmetic returns, we simulate daily log returns for which the distribution is known, and then aggregate these pseudo daily realizations to obtain its multi-period counterpart. The algorithm is as follows.

1. Randomly draw $S \times h$ pairs $\{z^s_{1T+t}, z^s_{2T+t}\}_{t=T+1}^{T+h}$ for $s = 1, \ldots, S$ of daily firm and market returns standardized innovations from the normal distribution. The number of pair is set to $S = 500000$ in the empirical application.

2. Build S paths of the couple $\{Y^s_{1T+t}, Y^s_{2T+t}\}_{t=T+1}^{T+h}$ given the information known at time T using the simulated standardized innovations.

3. Apply the formula $exp\left(\sum_{t=T+1}^{T+h} Y^s_{1T+t}\right) - 1$, and $exp\left(\sum_{t=T+1}^{T+h} Y^s_{2T+t}\right) - 1$, for $s = 1, \ldots, S$ to obtain the pseudo multi-period arithmetic returns $\{R^s_{1T+1:T+h}, R^s_{2T+1:T+h}\}_{s=1}^{S}$.

4. Compute the LRMES as the Monte Carlo average of the multi-period arithmetic returns as follows

$$LRMES_{1T+1:T+h}(\alpha, \widehat{\theta_T}) = -\frac{\sum_{s=1}^{S} R^s_{1T+1:T+h} \mathbb{1}\left(R^s_{2T+1:T+h} \leq VaR_{R_{2T+1:T+h}}\left(\alpha, \widehat{\theta_T}\right)\right)}{\sum_{s=1}^{S} \mathbb{1}\left(R^s_{2T+1:T+h} \leq VaR_{R_{2T+1:T+h}}\left(\alpha, \widehat{\theta_T}\right)\right)},$$

   where $VaR_{R_{2T+1:T+h}}\left(\alpha, \widehat{\theta_T}\right)$ is the multi-period market decline, and is defined as

$$VaR_{R_{2T+1:T+h}}\left(\alpha, \widehat{\theta_T}\right) = \text{percentile}\left(\{R^s_{2T+1:T+h}\}_{s=1}^{S}, 100\alpha\right).$$

5. Build the h step ahead SRISK forecasts as follows

$$SRISK_{1T+1:T+h}(\alpha, \widehat{\theta_T}) = kL_{1T} - (1-k)W_{1T}\left(1 + LRMES_{1T+1:T+h}(\alpha, \widehat{\theta_T})\right),$$

   where k is the value of the prudential capital ratio, L_{1T} is the book value of debt of the firm at time T, and W_{1T} is the market value of the firm at time T. As for the daily case, we set α to 5%.

Overall, our simulated SRISK forecasts (not reported) are similar to those reported by Brownlees and Engle (2017) even if we consider a conditional multivariate normal distribution, and not a semi-parametric GARH-DCC model estimated by QML.

Chapter 5

Conclusion

This dissertation proposes three essays that contribute to the financial econometrics literature in several ways. Our research focuses on financial risk measures and the development of financial tools dedicated to the evaluation of risk measure estimates. Three prominent classes of financial risks are investigated in this dissertation, namely, (*i*) credit risk, (*ii*) market risk, and (*iii*) systemic risk. Chapter 1 addresses issues related to the comparison and validation of methods devoted to credit risk measures, while Chapter 2 and Chapter 3 focus on the evaluation of the market-based risk measures used for the allocation of market risk capital requirements and within the context of financial regulations for the identification of SIFIs.

To compute the minimum capital requirements for credit risk, banks have the possibility to use their own internal risk models. These models are used to determine the PD, the LGD, the EAD and the maturity of the contract. This approach allows banks to be more effective in managing credit risk. However, it also raises the question of how these internal risk models should be selected, compared, and evaluated. Chapter 1 provides a solution to this question for the risk parameter LGD. The academic literature on the definition, measurement, and modeling of LGD is surprisingly underdeveloped compared to that of other risk parameters such as PD. LGD estimates enter the capital requirement formula in a linear way, and as a consequence, the estimation errors have a strong impact on required capital. Therefore, any underestimation of the LGD induces an underestimation of the regulatory capital. Chapter 1 develops a comparison methodology for LGD models that improves the banks' solvability. Contrary to the existing approach, which consists in evaluating the LGD forecasts with standard statistical criteria such as the mean square error or the mean absolute error computed between the observed and predicted LGD, our model's comparison method selects the model associated with the lowest estimation errors on regulatory capital. We show theoretically that our approach ranks models differently compared to the traditional approach, which only focuses on LGD forecast errors. To illustrate the interest of this methodology, we provide an empirical

application using a sample of credit and leasing contracts provided by an international bank. Six competing LGD models are implemented and compared based on our new methodology. Our method allows us to identify the best LGD models associated with the lowest estimation errors on the regulatory capital. We also introduce asymmetric criteria designed to improve financial stability. Rankings based on symmetric criteria are found to be substantially different from the model rankings obtained with asymmetric criteria. This result highlights the importance of penalizing more heavily the LGD models that lead to underestimation of the capital charge. As a significant step forward, it would be of particular interest to assess credit risk capital requirements as a whole. The BCBS provides a general framework to compute regulatory capital, but one can question the validity of this scope to deliver appropriate estimates. To the best of our knowledge, no formal statistical procedure has been proposed to assess these estimates, while it is a necessary condition to ensure that banks properly cover unexpected credit loss incurred on credit portfolios. Of great interest is the ASRF model providing the capital charge formula. This model is based on strong assumptions such as the infinite granularity of considered portfolios, the normal distribution of the risk factor, and a time horizon of one year (BCBS 2005), and we can question if these assumptions hold in practice.

In recent years, ES has received extensive attention from regulatory institutions and market practitioners. Recently, ES has replaced VaR for the calculation of market risk capital requirements to overcome its main deficiencies. As a relevant alternative, ES has the advantage of being a coherent risk measure (Artzner et al., 1999). In this renewed context, the modeling of ES and the validation of ES forecasts play a key role. Chapter 2 concentrates on this second point and introduces an econometric methodology to test for the validity of ES forecasts issued from market portfolios. Our testing strategy exploits the existing relationship that prevails between VaR and ES. We provide an implicit backtest of ES by verifying the validity of several VaRs in the tail distribution of the ES model. This strategy is fully consistent with the general recommendation of financial regulators to verify if the underlying ES model delivers correct quantiles at levels 0.975 and 0.990 (BCBS, 2016). The procedure generalizes the seminal idea of Gaglianone et al. (2011) of evaluating the VaR by applying quantile regression. We develop a multivariate quantile regression framework that allows the assessment of VaR at different probability levels in the tail distribution of the risk model. In an empirical application, we demonstrate the ability of our tests to reject a misspecified ES model. Furthermore, when ES forecasts are misleading, we provide an analytical correction that exploits our regression framework to adjust the imperfect forecasts. A natural step forward of this work would consist in considering the same approach but for assessing the whole return distribution (interval and density forecasts) in the same spirit as in Berkowitz (2001) and Kerkhof and Melenberg (2004).

The recent global financial crisis has led to the renewal of the financial regulation debate in order to prevent financial system-wide distress. Of particular interest is the identification of SIFIs. As they pose a major threat to the system, regulators and policy makers from around the world have called for tighter supervision, extra capital requirements, and liquidity buffers for SIFIs. Three prominent market-based systemic risk measures have been proposed for this purpose: the MES of Acharya et al. (2017), the SRISK of Acharya et al. (2012) and Brownlees and Engle (2017), and the ΔCoVaR of Adrian and Brunnermeier (2016). Very few crisis-related papers have had a higher impact both in academia and among regulators than this series of papers. Over the past four years, dozens of research papers have discussed, implemented, and sometimes generalized these systemic risk measures. Furthermore, these measures are currently used by several central banks and banking regulatory agencies. Within this context, Chapter 3 proposes the first general framework for backtesting systemic risk measures. In a first step, we focus our research on the MES, and then, we extend our backtesting procedure to other systemic risk measures (SRISK, SES, ΔCoVaR). The backtests are easy to implement and similar to those currently used for backtesting market risk measures (VaR, ES). We then apply our backtesting tests to assess the empirical validity of the MES, SRISK, and ΔCoVaR for a panel of large U.S. financial institutions over the period January 3, 2000, to December 30, 2016. Our results reveal that the one-day-ahead systemic risk forecasts are generally misspecified in periods of financial instability. However, when considering a longer forecasting horizon (one month), the periods of validity rejection are no longer there. This finding indicates that systemic risk indicators are more suited to capturing the long-run dynamics of systemic risk. Finally, we exploit our validation procedure to define an early warning system for the detection of advanced signs of crisis in the financial system. A promising avenue of research to extend the present work would be to identify whether systemic risk measures satisfy the property of elicitability. A statistical functional is deemed elicitable if there exists a loss function such that the correct forecast of the functional is the unique minimizer of the expected loss. This property opens the way to the possibility of modeling and estimating the statistical functional by means of regression and comparing its competing forecasts. As a possible starting point, Fissler and Ziegel (2016) show that ES is elicitable if combined with VaR. A natural extension should be to identify such a property for systemic risk measures that enter the class of shortfall measures including the MES, SES, and SRISK.

References

Acerbi, C., Szekely, B., 2014. Backtesting expected shortfall. Risk 27 (11), 76–81.

Acerbi, C., Tasche, D., 2002. On the coherence of expected shortfall. Journal of Banking & Finance 26 (7), 1487–1503.

Acharya, V., Engle, R., Pierret, D., 2014. Testing macroprudential stress tests: The risk of regulatory risk weights. Journal of Monetary Economics 65, 36–53.

Acharya, V., Engle, R., Richardson, M., 2012. Capital shortfall: A new approach to ranking and regulating systemic risks. American Economic Review 102 (3), 59–64.

Acharya, V., Pedersen, L., Philippon, T., Richardson, M., 2017. Measuring systemic risk. Review of Financial Studies 30 (1), 2–47.

Acharya, V., Pierret, D., Steffen, S., 2016a. Capital shortfalls of european banks since the start of the banking union. New York: Stern School of Business, New York University.

Acharya, V., Pierret, D., Steffen, S., 2016b. High time to tell european banks: No dividends. New York: Stern School of Business, New York University.

Adrian, T., Brunnermeier, M., 2016. Covar. American Economic Review 106 (7), 1705–1741.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Paul, L., 2003. Modeling and forecasting realized volatility. Journal of Applied Econometrics 71 (2), 579–625.

Argyropoulos, C., Panopoulou, E., 2016. A survey on risk forecast evaluation. Working paper.

Arnold, B., Beaver, R., Groeneveld, R., Meeker, W., 1993. The nontruncated marginal of a truncated bivariate normal distribution. Psychometrika 58 (3), 471–488.

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., 1999. Coherent measures of risk. Mathematical Finance 9 (3), 203–228.

Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., Vanthienen, J., 2003. Benchmarking state-of-the-art classification algorithms for credit scoring. Journal of the Operational Research Society 54, 627–635.

Banulescu, G.-D., Dumitrescu, E.-I., 2015. Which are the sifis? a component expected shortfall approach to systemic risk. Journal of Banking & Finance 50, 575–588.

Basel Committee on Banking Supervision, 2001. The internalratings-based approach. Consultation paper, January.

Basel Committee on Banking Supervision, 2005. An explanatory note on the basel ii irb risk weight functions. Consultation paper, July.

Basel Committee on Banking Supervision, 2010. Basel iii: A global regulatory framework for more resilient banks and banking systems. Consultation paper, December.

Basel Committee on Banking Supervision, 2014. The g-sib assessment methodology - score calculation. Report, November 6.

Basel Committee on Banking Supervision, 2016. Minimum capital requirements for market risk. Consultation paper, January.

Bastos, J. A., 2010. Forecasting bank loans loss-given-default. Journal of Banking & Finance 34 (10), 2510–2517.

Bastos, J. A., 2014. Ensemble predictions of recovery rates. Journal of Financial Services Research 46 (2), 177–193.

Bauwens, L., Laurent, S., Rombouts, J. V., 2006. Multivariate garch models: a survey. Journal of applied econometrics 21 (1), 79–109.

Bayer, S., Dimitriadis, T., 2018. Regression based expected shortfall backtesting. Working paper.

Bellotti, T., Crook, J., 2012. Loss given default models incorporating macroeconomic variables for credit cards. International Journal of Forecasting 28 (1), 171–182.

Benoit, S., Colletaz, G., Hurlin, C., Pérignon, C., 2013. A theoretical and empirical comparison of systemic risk measures. SSRN working paper.

Benoit, S., Colliard, J.-E., Hurlin, C., Pérignon, C., 2017. Where the risks lie: A survey on systemic risk. Review of Finance 21 (1), 109–152.

Benoit, S., Hurlin, C., Pérignon, C., 2019. Pitfalls in systemic-risk scoring. Forthcoming in Journal of Financial Intermediation.

Berkowitz, J., 2001. Testing density forecasts, with applications to risk management. Journal of Business & Economic Statistics 19 (4), 465–474.

Berkowitz, J., Christoffersen, P., Pelletier, D., 2011. Evaluating value at risk models with desk level data. Management Science 57, 2213–2272.

Billio, M., Getmansky, M., Lo, A., Pelizzon, L., 2012. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. Journal of Financial Economics 104 (3), 535–559.

Bishop, C. M., 1995. Neural networks for pattern recognition. Oxford University Press.

Bisias, D., Flood, M., Lo, A., Valavanis, S., 2012. A survey of systemic risk analytics. Annual Review of Financial Economics 4 (1), 255–296.

Boucher, C., Daníelsson, J., Kouontchou, P., Maillet, B., 2014. Risk models-at-risk. Journal of Banking & Finance 44, 72–92.

Box, G., Pierce, D., 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. Journal of the American statistical Association 65 (332), 1509–1526.

Breiman, L., 2001. Random forests. Machine learning 45 (1), 5–32.

Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and regression trees. Chapman and Hall/CRC.

Brown, L. D., Cai, T. T., DasGupta, A., 2001. Interval estimation for a binomial proportion. Statistical science, 101–117.

Brownlees, C., Chabot, B., Ghysels, E., Kurz, C., 2018. Back to the future: backtesting systemic risk measures during historical bank runs and the great depression. SSRN working paper.

Brownlees, C., Engle, R., 2017. Srisk: A conditional capital shortfall measure of systemic risk. Review of Financial Studies 30 (1), 48–79.

Cai, Z., Wang, X., 2008. Nonparametric estimation of conditional var and expected shortfall. Journal of Econometrics 147 (1), 120–130.

Calabrese, R., 2014a. Downturn loss given default: Mixture distribution estimation. European Journal of Operational Research 237 (1), 271–277.

Calabrese, R., 2014b. Predicting bank loan recovery rates with a mixed continuous-discrete model. Applied stochastic models in business and industry 30 (2), 99–114.

Calabrese, R., Zenga, M., 2010. Bank loan recovery rates: Measuring and nonparametric density estimation. Journal of Banking & Finance 34 (5), 903–911.

Candelon, B., Colletaz, G., Hurlin, C., Tokpavi, S., 2011. ÂŞbacktesting value-at-risk: A gmm duration-based test. Journal of Financial Econometrics 9, 314–343.

Chen, C., Iyengar, G., Moallemi, C., 2013. An axiomatic approach to systemic risk. Management Science 59 (6), 1373–1388.

Chicago Mercantile Exchange, 2012. Cme clearing risk management and financial safeguards. Report.

Christoffersen, P., 2010. Backtesting. Encyclopedia of Quantitative Finance R. Cont (ed). John Wiley and Sons.

Christoffersen, P., 2011. Elements of financial risk management. Academic Press, Second edition.

Christoffersen, P., Pelletier, D., 2004. Backtesting value-at-risk: A duration-based approach. Journal of Financial Econometrics 2 (1), 84–108.

Christoffersen, P. F., 1998. Evaluating interval forecasts. International Economic Review, 841–862.

Cochrane, J. H., 2011. Presidential address: Discount rates. Journal of Finance 66 (4), 1047–1108.

Colletaz, G., Hurlin, C., Pérignon, C., 2013. The risk map: A new tool for validating risk models. Journal of Banking & Finance 37 (10), 3843–3854.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics 7 (2), 174–196.

Costanzino, N., Curran, M., 2015. Backtesting general spectral risk measures with application to expected shortfall. Journal of Risk Model Validation 9 (1), 21–31.

Costanzino, N., Curran, M., 2018. A simple traffic light approach to backtesting expected shortfall. Risks 6 (1), 1–7.

Couperier, O., Leymarie, J., 2019. Backtesting expected shortfall via multi-quantile regression. Working paper.

Cruz Lopez, J. A., Harris, J. H., Hurlin, C., Pérignon, C., 2017. Comargin. Journal of Financial and Quantitative Analysis 52 (5), 2183–2215.

Cubadda, G., Guardabascio, B., Hecq, A., 2017. A vector heterogeneous autoregressive index model for realized volatility measures. International Journal of Forecasting 33 (2), 337 – 344.

Daníelsson, J., James, K., Valenzuela, M., Zer, I., 2016. Model risk of risk models. Journal of Financial Stability 23, 79–91.

Darolles, S., Francq, C., Laurent, S., 2018. Asymptotics of cholesky garch models and time-varying conditional betas. Journal of Econometrics 204 (2), 223 – 247.

De Bandt, O., Hartmann, P., 2002. Systemic risk: A survey, in financial crisis, contagion and the lender of last resort: A book of readings. ed. by C. Goodhart, and G. Illing. Oxford University Press.

Dermine, J., Neto De Carvalho, C., 2006. Bank loan losses-given-default: A case study. Journal of Banking & Finance 30 (4), 1219–1243.

Diebold, F. X., Mariano, R. S., 1995. Comparing predictive accuracy. Journal of Business & Economic Statistics 13 (3), 253–263.

Du, Z., Escanciano, J. C., 2017. Backtesting expected shortfall: Accounting for tail risk. Management Science 63 (4), 901–1269.

Dumitrescu, E.-I., Hurlin, C., Pham, V., 2012. Backtesting value-at-risk: From dynamic quantile to dynamic binary tests. Finance 33, 79–111.

Dwyer, P., 1967. Some applications of matrix derivatives in multivariate analysis. Journal of the American Statistical Association 62 (318), 607–625.

Emmer, S., Kratz, M., Tasche, D., 2015. What is the best risk measure in practice? a comparison of standard measures. Journal of Risk 18 (2), 31–60.

Engle, R., 2002. Dynamic conditional correlation. Journal of Business & Economic Statistics 20 (3), 339–350.

Engle, R., 2016. Dynamic conditional beta. Journal of Financial Econometrics 14 (4), 643–667.

Engle, R., Jondeau, E., Rockinger, M., 2015. Systemic risk in europe. Review of Finance 19 (1), 145–190.

Engle, R. F., Manganelli, S., 2004. Caviar: Conditional autoregressive value at risk by regression quantiles. Journal of Business & Economic Statistics 22 (4), 367–381.

Escanciano, J. C., Olmo, J., 2010. Backtesting parametric value-at-risk with estimation risk. Journal of Business & Economic Statistics 28 (1), 36–51.

European Banking Authority, 2014. Results of 2014 eu-wide stress test. Report, October 26.

European Banking Authority, 2016. Guidelines on pd estimation, lgd estimation and the treatment of defaulted exposures. Consultation paper, November.

Ferreiro, J., 2018. A component delta conditional expected shortfall measure in the financial systemic risk framework. Working paper.

Financial Stability Board, 2011. Policy measures to address systemically important financial institutions. November.

Financial Times, 2014. Alternative stress tests find french banks are weakest in europe. October 27.

Fissler, T., Ziegel, J., 2016. Higher order elicitability and osband's principle. Annals of Statistics 44 (4), 1680–1707.

Francq, C., Zakoïan, J.-M., 2015. Risk-parameter estimation in volatility models. Journal of Econometrics 184, 158–173.

Francq, C., Zakoïan, J.-M., 2016. Looking for efficient qml estimation of conditional vars at multiple risk levels. Annals of Economics and Statistics (123-124), 9–28.

Freedman, D., 1981. Bootstrapping regression models. Annals of Statistics 9 (6), 1218–1228.

Friedman, J. H., 2001. Greedy function approximation: a gradient boosting machine. Annals of Statistics 29 (5), 1189–1232.

Gaglianone, W. P., Lima, L. R., Linton, O., Smith, D. R., 2011. Evaluating value-at-risk models via quantile regression. Journal of Business & Economic Statistics 29 (1), 150–160.

Genest, B., Brie, L., 2013. Basel ii irb risk weight functions: Demonstration and analysis. SSRN working paper.

Giglio, S., Kelly, B., Pruitt, S., 2016. Systemic risk and the macroeconomy: An empirical evaluation. Journal of Financial Economics 119 (3), 457–471.

Girardi, G., Ergün, T., 2013. Systemic risk measurement: Multivariate garch estimation of covar. Journal of Banking & Finance 37 (8), 3169–3180.

Glosten, L., Jagannathan, R., Runkle, D., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48 (5), 1779–1801.

Gneiting, T., 2011. Making and evaluating point forecasts. Journal of the American Statistical Association 106 (494), 746–762.

Gouriéroux, C., 1992. Courbes de performance, de sélection et de discrimination. Annales d'Économie et de Statistique 28, 107–123.

Gouriéroux, C., Liu, W., 2012. Converting tail-var to var: An econometric study. Journal of Financial Econometrics 10 (2), 233–264.

Gouriéroux, C., Monfort, A., Polimenis, V., 2006. Affine models for credit risk analysis. Journal of Financial Econometrics 4 (3), 494–530.

Gouriéroux, C., Tiomo, A., 2007. Risque de crédit: une approche avancée. Economica.

Gouriéroux, C., Zakoïan, J.-M., 2013. Estimation-adjusted var. Econometric Theory 29 (4), 735–770.

Gupton, G., Stein, R., 2002. Losscalc: Moody's model for predicting loss given default. Moody's technical report.

Gürtler, M., Hibbeln, M., 2013. Improvements in loss given default forecasts for bank loans. Journal of Banking & Finance 37 (7), 2354 – 2366.

Gürtler, M., Hibbeln, M. T., Usselmann, P., 2018. Exposure at default modeling–a theoretical and empirical assessment of estimation approaches and parameter choice. Journal of Banking & Finance 91, 176–188.

Hagmann, M., Renault, O., Scaillet, O., 2005. Estimation of recovery rate densities: non-parametric and semi-parametric approaches versus industry practice. Recovery Risk: The Next Challenge in Credit Risk Management Altman, Resti, Sironi (ed.), 323–346.

Hand, D. J., Henley, W. E., 1997. Statistical classification methods in consumer credit scoring: a review. Journal of the Royal Statistical Society: Series A (Statistics in Society) 160 (3), 523–541.

Hansen, P. R., Lunde, A., Nason, J. M., 2011. The model confidence set. Econometrica 79 (2), 453–497.

Hartmann-Wendels, T., Miller, P., Töws, E., 2014. Loss given default for leasing: Parametric and non-parametric estimations. Journal of Banking & Finance 40, 364–375.

Ho, H., Lin, T.-I., Chen, H.-Y., Wang, W.-L., 2012. Some results on the truncated multivariate t distribution. Journal of Statistical Planning and Inference 142 (1), 25–40.

Horrace, W., 2005. Some results on the multivariate truncated normal distribution. Journal of Multivariate Analysis 94 (1), 209–221.

Huber, P. J., 1967. The behavior of maximum likelihood estimates under nonstandard conditions. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1, 221–233, University of California Press, Berkeley.

Hué, S., Lucotte, Y., Tokpavi, S., 2019. Measuring network systemic risk contributions: A leave-one-out approach. Journal of Economic Dynamics and Control 100, 86–114.

Hurlin, C., Laurent, S., Quaedvlieg, R., Smeekes, S., 2017. Risk measure inference. Journal of Business & Economic Statistics 35 (4), 499–512.

Hurlin, C., Tokpavi, S., 2006. Backtesting var accuracy: a new simple test. Journal of Risk 9 (2), 19–37.

Idier, J., Lamé, G., Mésonnier, J.-S., 2014. How useful is the marginal expected shortfall for the measurement of systemic exposure? a practical assessment. Journal of Banking & Finance 47, 134–146.

Jorion, P., 2006. Value at risk: the new benchmark for managing financial risk. McGraw-Hill, Third edition.

Jorion, P., 2007. Value-at-risk. McGraw-Hill, Third edition.

Kalotay, E. A., Altman, E. I., 2017. Intertemporal forecasts of defaulted bond recoveries and portfolio losses. Review of Finance 21 (1), 433–463.

Kerkhof, J., Melenberg, B., 2004. Backtesting for risk-based regulatory capital. Journal of Banking & Finance 28 (4), 1845–1865.

Khieu, H. D., Mullineaux, D. J., Yi, H.-C., 2012. The determinants of bank loan recovery rates. Journal of Banking & Finance 36 (4), 923–933.

Koenker, R., Chernozhukov, V., He, X., Peng, L., 2018. Handbook of quantile regression. Chapman and Hall/CRC Handbooks of Modern Statistical Methods.

Koenker, R., Xiao, Z., 2002. Inference on the quantile regression process. Econometrica 70 (4), 1583–1612.

Kratz, M., Lok, Y. H., McNeil, A. J., 2018. Multinomial var backtests: A simple implicit approach to backtesting expected shortfall. Journal of Banking & Finance 88, 393–407.

Krüger, S., Rösch, D., 2017. Downturn lgd modeling using quantile regression. Journal of Banking & Finance 79, 42–56.

Kupiec, P., 1995. Techniques for verifying the accuracy of risk measurement models. Journal of Derivatives 3 (2), 73–84.

Lambert, P., Laurent, S., Veredas, D., 2012. Testing conditional asymmetry: A residual-based approach. Journal of Economic Dynamics and Control 36 (8), 1229–1247.

Lazar, E., Zhang, N., 2019. Model risk of expected shortfall. Journal of Banking & Finance 105, 74–93.

Leccadito, A., Boffelli, S., Urga, G., 2014. Evaluating the accuracy of value-at-risk forecasts: New multilevel tests. International Journal of Forecasting 30 (2), 206–216.

Lessmann, S., Baesens, B., Seow, H.-V., Thomas, L., 2015. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research 247, 124–136.

Lopez, J. A., Saidenberg, M. R., 2000. Evaluating credit risk models. Journal of Banking & Finance 24 (1-2), 151–165.

Loterman, G., Brown, I., Martens, D., Mues, C., Baesens, B., 2012. Benchmarking regression algorithms for loss given default modeling. International Journal of Forecasting 28 (1), 161–170.

Loterman, G., Debruyne, M., Vanden Branden, K., Van Gestel, T., Mues, C., 2014. A proposed framework for backtesting loss given default models. Journal of Risk Model Validation 8 (1), 69–90.

Markowitz, H., 1952. Portfolio selection. Journal of Finance 7 (1), 77–91.

Matuszyk, A., Mues, C., Thomas, L. C., 2010. Modelling lgd for unsecured personal loans: Decision tree approach. Journal of the Operational Research Society 61 (3), 393–398.

McNeil, A. J., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. Journal of Empirical Finance 7 (3-4), 271–300.

Medema, L., Koning, R. H., Lensink, R., 2009. A practical approach to validating a pd model. Journal of Banking & Finance 33 (4), 701–708.

Merton, R. C., 1974. On the pricing of corporate debt: The risk structure of interest rates. Journal of Finance 29 (2), 449–470.

Miller, P., Töws, E., 2018. Loss given default adjusted workout processes for leases. Journal of Banking & Finance 91, 189–201.

Morgan, J., 1996. Riskmetrics. Technical Document (4th ed.), New York: Morgan Guaranty Trust Company of New York.

Nazemi, A., Fatemi Pour, F., Heidenreich, K., Fabozzi, F. J., 2017. Fuzzy decision fusion approach for loss-given-default modeling. European Journal of Operational Research 262 (2), 780–791.

Newey, W., West, K., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–708.

Nieto, M., Ruiz, E., 2016. Frontiers in var forecasting and backtesting. International Journal of Forecasting 32 (2), 475–501.

Nolde, N., Ziegel, J. F., 2017. Elicitability and backtesting: Perspectives for banking regulation. Annals of Applied Statistics 11 (4), 1833–1874.

Noyer, C., 2014. Présentation des résultats de l'evaluation complète des bilans bancaires. ACPR Banque de France, October 26.

Palm, F., 1996. 7 garch models of volatility. In: Statistical Methods in Finance. Vol. 14 of Handbook of Statistics. Elsevier, pp. 209–240.

Papke, L. E., Wooldridge, J. M., 1996. Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. Journal of Applied Econometrics 11 (6), 619–632.

Patton, A. J., Ziegel, J. F., Chen, R., 2019. Dynamic semiparametric models for expected shortfall (and value-at-risk). Forthcoming in Journal of Econometrics.

Pelletier, D., Wei, W., 2016. The geometric-var backtesting method. Journal of Financial Econometrics 14, 725–745.

Pérignon, C., Smith, D., 2008. A new approach to comparing var estimation methods. Journal of Derivatives 16 (2), 54–66.

Philippon, T., Pessarossi, P., Camara, B., 2017. Backtesting european stress tests. NBER working paper.

Pierret, D., Steffen, S., 2018. Capital shortfalls at european banks after the 2018 stress test. AllNews November 7.

Powell, J. L., 1984. Least absolute deviations estimation for the censored regression model. Journal of Econometrics 25 (3), 303–325.

Qi, M., Zhao, X., 2011. Comparison of modeling methods for loss given default. Journal of Banking & Finance 35 (11), 2842–2855.

Rabemananjara, R., Zakoian, J.-M., 1993. Threshold arch models and asymmetries in volatility. Journal of Applied Econometrics 8 (1), 31–49.

Renault, O., Scaillet, O., 2004. On the way to recovery: A nonparametric bias free estimation of recovery rate densities. Journal of Banking & Finance 28 (12), 2915–2931.

Romano, J., Shaikh, A., Wolf, M., 2008. Formalized data snooping based on generalized error rates. Econometric Theory 24 (2), 404–447.

Roncalli, T., 2009. La gestion des risques financiers. Economica, Seconde édition.

Roncalli, T., 2014. Introduction to risk parity and budgeting. Chapman and Hall/CRC Financial Mathematics Series.

Scaillet, O., 2004. Nonparametric estimation and sensitivity analysis of expected shortfall. Mathematical Finance 14 (1), 115–129.

Schmit, M., 2004. Credit risk in the leasing industry. Journal of Banking & Finance 28 (4), 811–833.

Schuermann, T., 2004. What do we know about loss given default? Credit Risk Models and Management Shimko (ed.).

Suykens, J., Van Gestel, T., De Brabanter, J., De Moor, B., Vandewalle, J., 2002. Least squares support vector machines. Singapore: World Scientific.

Suykens, J. A., Vandewalle, J., 1999. Least squares support vector machine classifiers. Neural Processing Letters 9 (3), 293–300.

Tanoue, Y., Kawada, A., Yamashita, S., 2017. Forecasting loss given default of bank loans with multi-stage model. International Journal of Forecasting 33 (2), 513–522.

Tavolaro, S., Visnovsky, F., 2014. What is the information content of the srisk measure as a supervisory tool? ACPR Banque de France, Economics and Financial Debates.

Taylor, J. W., 2019. Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric laplace distribution. Journal of Business & Economic Statistics 37 (1), 121–133.

Tobback, E., Martens, D., Van Gestel, T., Baesens, B., 2014. Forecasting loss given default models: impact of account characteristics and the macroeconomic state. Journal of the Operational Research Society 65 (3), 376–392.

Vapnik, V. N., 1995. The nature of statistical learning theory. Springer, First edition.

Vasicek, O., 2002. The distribution of loan portfolio value. Risk 15 (12), 160–162.

White, H., Kim, T. H., Manganelli, S., 2008. Modeling autoregressive conditional skewness and kurtosis with multi-quantile caviar. Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle In Bollerslev T., Russell, J., and Watson, M. editors.

White, H., Kim, T.-H., Manganelli, S., 2015. Var for var: Measuring tail dependence using multivariate regression quantiles. Journal of Econometrics 187 (1), 169–188.

Wied, D., Weiß, G., Ziggel, D., 2016. Evaluating value-at-risk forecasts: A new set of multivariate backtests. Journal of Banking & Finance 72, 121–132.

Wong, W. K., 2008. Backtesting trading risk of commercial banks using expected shortfall. Journal of Banking & Finance 32 (7), 1404–1415.

Wu, D., Zhao, X., 2018. Systemic risk and bank failure. SSRN working paper.

Yao, X., Crook, J., Andreeva, G., 2015. Support vector regression for loss given default modelling. European Journal of Operational Research 240 (2), 528–538.

Yao, X., Crook, J., Andreeva, G., 2017. Enhancing two-stage modelling methodology for loss given default with support vector machines. European Journal of Operational Research 263 (2), 679–689.

Zhang, J., Thomas, L. C., 2012. Comparisons of linear regression and survival analysis using single and mixture distributions approaches in modelling lgd. International Journal of Forecasting 28 (1), 204–215.

Valorization

The research presented herein is about financial econometrics with applications to financial risk management and banking regulation. I have pursued a focus on risk evaluation, model comparison, and estimation risk. These topics are fundamental within the context of academic financial research, but they are also relevant for professionals in the financial sector as well as for regulators and supervisory authorities. In this work, I have studied three prominent classes of financial risk: *(i)* credit risk, *(ii)* market risk, and *(iii)* systemic risk. They originate from different sources but are unified by a common will of the economic and business worlds to properly quantify them. As with any task involving predictions of the future, risk forecasting is afflicted with estimation errors. The econometric content of my work aspires to the identification of such misestimation under an acceptable margin of error decided by a financial entity. In this addendum, I will attempt to summarize the main contributions of my work from both a methodological standpoint and the perspectives they open up for banking regulation and financial stability.

Risk forecast evaluation and estimation uncertainty

Measuring financial risks is central in the process of managing risk. The set of underlying management problems is generally addressed via the estimation of financial risk measures whose purpose is to quantify financial risks with one number representing the future losses that could be potentially experienced on a risky position. Unhappily, modeling and estimating these measures is not sufficient in itself. Econometric methods to evaluate their ex-post validity are additionally needed. The relevance of this thesis with respect to the econometrics field can be recognized through two contributions. The first is through the alternative forecasting evaluation procedures and model comparison methods developed for financial risk measures. Because the financial risk measures are unobservable their evaluation cannot be conventionally performed as a direct comparison of the observed value with its forecast. One of the main contribution of this dissertation is to provide alternative methodologies that apply to the case of unobserved target functionals.

A second contribution of this thesis is an attempt to account for the problem of estimation uncertainty. Risk measures have to be estimated and their estimation counterparts are subject to estimation uncertainty. Replacing, in the theoretical formulas,

the true parameter value by an estimator induces uncertainty and errors on the subsequent forecast. This drawback is common to any financial risk measure and any facet of risk. Importantly, the question goes to the market activities (asset management and derivatives, portfolio allocation, etc.), the lending activities (secured and unsecured loans, bonds, etc.), and even to the systemic risk monitoring activities (contagion measurement, stress test exercises, etc.). The methods I offer in chapter 3 and chapter 4 are devoted to this question. I develop adjustment techniques associated with a given financial risk measure which is priorly affected by estimation uncertainty. In chapter 3, I have pursued a focus on expected shortfall estimation uncertainty, which can be easily extended to value-at-risk, while the purpose of chapter 4 is on systemic risk measurement. In addition, I formally establish in chapter 4 that the validation procedure itself is affected by estimation uncertainty. In presence of estimation uncertainty in the systemic risk indicators, I show that the standard inferential procedures no longer apply.

Banking regulation and financial stability

Beyond the methodological contribution, my work also benefits both banking regulation and financial stability. The primary mission of regulatory authorities and supervisory agencies is to continuously monitor the financial system risk exposure. The methods presented in chapter 2 and chapter 3 are well suited to this objective. Both chapters offer new insights on how to evaluate and compare risk measure estimates. Financial risk measures are involved in the calculation of banks' capital charge and are thus of great importance. Consequently, any underestimation of these parameters may induce an underestimation of the regulatory capital and a lower banks' solvency. In chapter 2, I propose a model comparison method for the loss given default which induces the lowest estimation errors on the banks' capital charge. I show theoretically and empirically that the proposed approach improves banks' solvency compared to the current method used by academics, banks, and regulators. This work constitutes a further step in the ongoing process of embedding a more economic content to risk evaluation. In chapter 3, I suggest a relationship between value-at-risk and expected shortfall that considerably simplifies the estimation and assessment of expected shortfall from a regulatory viewpoint. Using the proposed relationship allows the implementation of easy-to-use validation tests of the expected shortfall estimates. These tests promote a more intelligible evaluation of risks and also come in response to the market failures revealed by the global 2007-2008 financial crisis. The methodological developments of this work may thus stand a better chance of gaining acceptance from banks and their regulators while enabling them to push forward the current legislation and guidelines in banking regulation.

Lastly, a relevant contribution to financial stability I would highlight is the approach proposed to assess market-based systemic risk indicators, which has been the subject

of a considerable debate between the academic and regulatory spheres. It was the case when the European banking authority publishes the results of the 2014 EU-wide stress tests indicating that French banks are among the safest in Europe. These conclusions were immediately casted in doubt using systemic risk measures while revealing the exact opposite conclusions that French financial institutions would face an aggregate capital shortfall of almost $400bn in case of crisis (according to the SRISK definition). My work attempts to explain these inconsistencies. I show that systemic risk measures are affected by large estimation errors and more importantly in times of crisis. My results suggest that the systemic risk measures are not always able to accurately conclude which institution is systemically riskier than another, or to determine the right level of regulatory capital for the systemically important financial institutions. This provides a first answer to the question of the conflicting outcomes observed from both methodologies.

Résumé en Français

La gestion des risques est un domaine d'expertise central pour les institutions financières telles que les banques, les compagnies d'assurance, ou encore les fonds d'investissement. L'un des enseignements les plus importants que nous ayons tiré de la crise économique et financière mondiale est la nécessité de mesurer fidèlement les risques financiers. Dans l'environnement financier actuel, la croissance constante de la taille et de la complexité des institutions financières et du rythme de leurs transactions a indéniablement introduit une nouvelle variable dans l'équation. Parallèlement, et heureusement, les progrès technologiques en matière de communication et de collecte de données ont permis de réduire les coûts d'acquisition, de gestion et d'analyse des données, de suivre de près l'évolution des risques et de mieux refléter l'environnement économique plus complexe et au rythme rapide. Ce contexte a favorisé le développement d'instruments financiers sophistiqués et de nouvelles techniques de gestion des risques. En particulier, la recherche universitaire en économétrie financière s'est concentrée sur (i) l'élaboration de nouvelles mesures du risque financier, (ii) le dévelopement de méthodes d'estimation et d'inférence appropriées, et (iii) la mise en œuvre des techniques de validation consacrées à ces mesures financières.

I. Les Mesures du Risque Financier

Les techniques de mesure du risque financier sont au cœur du processus de gestion des risques. En théorie, la mesure du risque financier pourrait être abordée à partir de la notion de distribution de probabilité. Cependant, en pratique, il peut être délicat d'extraire les informations pertinentes à travers toute la distribution de probabilité, et il est préférable de les résumer en un seul chiffre qui donne de l'information sur une dimension particulière du risque (asymétrie, espérance, tendance centrale, queue de distribution, variance, etc.). Les mesures du risque financier sont généralement dédiées à cet objectif en quantifiant les risques financiers à partir d'une valeur numérique représentant les pertes futures qui seront probablement subies sur une position risquée. Pour Artzner et al. (1999), ce nombre, lorsqu'il est positif, indique le montant minimum de trésorerie supplémentaire que l'agent doit ajouter à la position risquée, pour satisfaire à l'ensemble des risques acceptables définis par une autorité de supervision. De ce fait, ce nombre est généralement interprété comme l'exigence en capitaux propres découlant de la détention

de cette position risquée, et qui permet de réguler le risque supporté par les acteurs du marché, les opérateurs financiers, ou les assureurs. Un large éventail de mesures de risque a été proposé dans la littérature scientifique, conçues pour couvrir divers types de risques financiers et d'instruments de référence. Dans la suite, nous décrivons les principales mesures du risque financier et les regroupons en fonction du type de risque financier qu'elles visent à quantifier.

Mesures de risque de crédit. Les mesures de risque de crédit sont des outils nécessaires pour modéliser les pertes dites attendues, et les pertes dites inattendues, enregistrées dans les portefeuilles de crédit des institutions financières. Les premières désignent les pertes "habituelles" ou moyennes qu'une institution subit dans le cours naturel de ses affaires, tandis que les secondes correspondent à des pertes potentielles importantes qui sont enregistrées dans des conditions défavorables et peuvent menacer la stabilité financière (voir Gouriéroux and Tiomo, 2007; Roncalli, 2009; Genest and Brie, 2013, pour plus de détails). Quatre mesures de risque de crédit sont dédiées à l'analyse de ces pertes: la *probability of defaut* (PD), la *loss given default* (LGD), la *exposure at default* (EAD) et la *maturity* (M). La mesure de risque PD fournit une estimation de la probabilité de défaut sur un horizon temporel donné et représente le risque que l'emprunteur ne puisse pas ou ne veuille pas rembourser sa dette intégralement ou à temps. La LGD représente le montant de créance perdue par la banque en cas de défaillance du débiteur. Par conséquent, PD et LGD sont étroitement liées. Il est plus probable d'observer une LGD positive pour un emprunteur dont la PD est initialement élevée. Enfin, l'EAD correspond à l'exposition encourue par un créancier en cas de défaut de son débiteur. Le champ des activités de prêt comprend également le risque de défaillance dans les chambres de compensation (CCPs) associé aux portefeuilles de produits dérivés. Les mesures ou approches de risque couramment utilisées dans les CCPs sont le *standard portfolio analysis of risk* (SPAN) ou l'approche de la *value-at-risk* (VaR) pour estimer les exigences de collatéral en fonction du niveau de couverture des pertes potentielles pour un contrat individuel ou un portefeuille de contrats (Chicago Mercantile Exchange, 2012). Récemment, des techniques plus sophistiquées ont émergé. Par exemple, Cruz Lopez et al. (2017) ont développé la méthodologie CoMargin qui estime les exigences de collatéral en tenant compte à la fois du risque extrême d'un acteur du marché et de son interdépendance avec les autres acteurs du marché.

Mesures de risque de marché. Les mesures de risque de marché servent à quantifier les risques de pertes résultant des fluctuations des prix des instruments financiers. Comme pour les autres formes de risque, le montant des pertes potentielles issues du risque de marché peut être mesuré de différentes manières ou conventions. Traditionnellement, une convention consiste à utiliser la VaR définie comme la perte maximale liée à la détention d'un actif (ou d'un portefeuille) sur une période donnée et pour un niveau

de probabilité donné. La VaR est devenue un élément constitutif des systèmes internes de gestion des risques dans les institutions financières, à la suite du succès du système RiskMetrics de J.P. Morgan (1996). Les conventions d'utilisation de la VaR sont bien établies et acceptées dans les départements de gestion des risques. Cependant, la VaR présente un certain nombre de faiblesses. En particulier, la VaR n'est pas une mesure de risque cohérente puisqu'elle n'est pas systématiquement sous-additive (Artzner et al., 1999). Par conséquent, de nouvelles propositions pour mesurer le risque de marché ont été suggérées telles que la mesure de risque *expected shortfall* (ES), également connue sous le nom de *conditional value-at-risk* (CVaR), ou *tail value-at-risk* (TVaR). Par définition, l'ES désigne la perte espérée lorsque cette dernière est plus extrême que la VaR pour un niveau de probabilité donné. Enfin, il convient de noter que la variance (ou l'écart type) joue également un rôle important en tant que mesure du risque de marché. En particulier, elle est devenue un indicateur clé de la théorie moderne du portefeuille (Markowitz, 1952), ou plus généralement de toute stratégie d'investissement basée sur le moment d'ordre deux (voir Roncalli, 2014).

Mesures de risque systémique. La crise financière globale a favorisé l'émergence de nombreuses recherches en lien avec le risque systémique. Un des objectifs principaux de cette littérature est d'identifier les institutions financières d'importance systémique (SIFIs) qui contribuent le plus au risque global du système financier. Le Conseil de Stabilité Financière (FSB, 2011) définit les SIFIs comme étant "*des institutions financières dont les difficultés, en raison de leur taille, de leur complexité, et de leur interconnexion systémique, perturberaient considérablement le système financier et l'activité économique au sens large*". Comme ces firmes constituent une menace majeure pour le système, les régulateurs et les décideurs du monde entier ont appelé à une surveillance plus étroite, et à des exigences de fonds propres et de réserves de liquidité supplémentaires pour les SIFIs. Dans cette perspective, de nombreuses mesures de risque systémique ont été proposées dans la littérature scientifique au cours de ces dernières années (voir Benoit et al., 2017, pour un résumé), les plus connues étant la *marginal expected shortfall* (MES) et la *systemic expected shortfall* (SES) de Acharya et al. (2017), la *systemic risk measure* (SRISK) de Acharya et al. (2012) et Brownlees and Engle (2017), et la *delta conditional value-at-risk* (ΔCoVaR) de Adrian and Brunnermeier (2016). Ces indicateurs rendent compte en un seul chiffre de la contribution du risque systémique de chaque institution financière dans le but d'identifier les SIFIs dont la défaillance pourrait déclencher une crise du système financier dans son ensemble.

II. Estimation

Le risque financier ne peut être mesurer directement même une fois l'événement réel observé. C'est pourquoi il est qualifié de processus latent en économétrie. De ce fait,

les mesures de risque sont des quantités non observables, et nous devons les estimer. Ces estimations sont généralement fournies par un modèle de risque. Dans le contexte réglementaire actuel, les banques ont la possibilité de développer leurs propres modèles de risque pour estimer les mesures de risque de crédit (BCBS, 2001) et les mesures de risque de marché (BCBS, 2016) utilisées dans le calcul des exigences en fonds propres réglementaires. Les banques sont incitées à maintenir le niveau de capital réglementaire le plus faible possible car la réduction de ce capital libère des ressources économiques pouvant être affectées à des investissements rentables. Ainsi, la plupart des banques développent leurs propres modèles de risque car ils conduisent généralement à un niveau de capital requis moins élevè que l'application de l'approche standard fournie par les autorités de contrôle. Cependant, l'estimation de ces mesures et le choix des modèles ne sont pas chose facile. Les banques ont donc un rôle déterminant dans le calcul de ces exigences en fonds propres réglementaires et dans la mise en place de pratiques de gestion des risques responsables.

De nombreux modèles sont disponibles pour modéliser les mesures de risque financières. La nature de la variable de réponse fournit des indications importantes sur le choix d'un modèle pertinent. Par exemple, la mesure de risque PD répond typiquement à un problème de classification binaire, et les techniques de modélisation reposent alors sur des classifieurs. En pratique, le choix se portera sur les modèles de choix binaires tels que les modèles de régression logistique et probit, ou sur les approches statistiques telles que l'analyse discriminante. Concernant la modélisation de la LGD, on utilise couramment de simples tableaux de contingences, ou des modèles de régression (régression linéaire, analyse de survie, régression à réponse fractionnelle, régression à partir de loi mixte, modèles Tobit, etc.). Au cours de ces dernières décennies, les techniques d'apprentissage automatique ont également gagné en popularité pour la modélisation des mesures de risque de crédit en raison de leur capacité à améliorer de manière significative les performances des modèles. Nous pouvons par exemple citer les arbres de régression et de classification, les machines et régressions à vecteurs de support, les forêts aléatoires, les approches de type gradient boosting, et les réseaux de neurones (voir Baesens et al., 2003; Lessmann et al., 2015 pour une étude comparative des modèles de PD, et Loterman et al., 2012 pour le cas des modèles LGD).

En raison de la présence d'une dynamique temporelle dans les prix de marché, les mesures de risque de marché sont typiquement exprimées conditionnellement à un ensemble d'information, et les prévisions sont généralement établies à partir de modèles dynamiques paramétriques ou semi-paramétriques. Par exemple, les modèles GARCH univariés et multivariés peuvent être utilisés pour produire des prévisions de VaR ou d'ES conditionnelles (voir Palm, 1996, pour une revue de littérature des processus GARCH univariés, et Bauwens et al., 2006, pour le cas des modèles GARCH multivariés), ou un

modèle de corrélation conditionnelle dynamique (DCC) peut être utilisé pour estimer des bêta dynamiques (Engle, 2002, 2016). Les approches de modélisation incluent, par ailleurs, les modèles de régression quantile univariés et multivariés (Engle and Manganelli, 2004; White et al., 2008, 2015, etc.), les modèles de volatilité réalisés (Andersen et al., 2003; Corsi, 2009; Cubadda et al., 2017, etc.) et les approches d'estimation non-paramétriques (Scaillet, 2004; Cai and Wang, 2008, etc.). Par ailleurs, de nombreuses avancées techniques de modélisation des mesures de risque de marché ont vu le jour ces dernières années. Darolles et al. (2018) proposent un nouveau modèle avec des coefficients de pente variant dans le temps basés sur la décomposition de Cholesky de la matrice de variance conditionnelle. Leur modèle affiche de meilleurs capacités prédictives qu'un modèle avec des bêtas constants ou que le modèle DCC. Taylor (2019) introduit une méthode pour prédire l'ES à partir de prévisions de VaR produites par des modèles de régression quantile. Patton et al. (2019) utilisent les résultats récents de la théorie statistique en modélisant conjointement l'ES et la VaR, et proposent de nouveaux modèles dynamiques pour ces mesures de risque. Par ailleurs, ces modèles ont l'avantage certain de s'appliquer aussi aux mesures de risque systémique calculées à partir de données de marché. Par exemple, Adrian and Brunnermeier (2016) considèrent un simple modèle de régression quantile pour prévoir la ΔCoVaR, tandis que Girardi and Ergün (2013) estiment cette dernière à partir d'un modèle GARCH multivarié. Brownlees and Engle (2017) implémentent un modèle DCC pour estimer la MES et la SRISK de long terme. Ces similitudes des approches de modélisation du risque de marché et du risque systémique s'expliquent par le fait qu'elles sont toutes deux fondées sur des données de marché et que les modèles de risque sous-jacents doivent être à même de capturer les faits stylisés des séries financières, tels que la non-stationnarité en niveau, les queues de distribution épaisses, et la persistance de la volatilité.

Les mesures du risque financier cachent une faiblesse commune : leurs estimations sont affectées par deux types d'erreur de mesure. La première source d'erreur de mesure, appelée risque d'estimation, vient du fait que les paramètres du modèle sont remplacés par leurs estimations. Cela a un impact sur la précision de l'estimation de la mesure de risque elle-même. L'inférence devient alors un outil puissant pour quantifier les erreurs d'estimation et les corriger. Diverses contributions ont été proposées pour répondre à cette lacune. Hurlin et al. (2017) proposent une procédure bootstrap qui tient compte de l'incertitude de l'estimation afin de tester l'égalité des mesures de risque conditionnelles pour différents actifs, portefeuilles, ou firmes, à une date prédéfinie. Francq and Zakoïan (2015) quantifient l'effet du risque d'estimation dans la classe des modèles GARCH estimés par quasi-maximum de vraisemblance non Gaussien, et dérivent une distribution asymptotique et des intervalles de confiance pour la VaR. Gouriéroux and Zakoïan (2013) montrent que la VaR est affectée par un biais asymptotique dans les probabilités

de couverture induit par l'estimation et en déduisent une correction. La deuxième source d'erreur de mesure, appelée risque de modèle, correspond au risque que le modèle de prévision soit mal spécifié et puisse conduire à des résultats incohérents de prévisions du risque. Boucher et al. (2014) proposent une méthodologie de construction de mesures de risque robustes au risque de modèle. Ils montrent que le biais de modèle est important et dépend fortement du niveau de couverture considéré. Daníelsson et al. (2016) proposent un cadre général pour quantifier le risque de modèle, et montrent que le degré de risque de modèle est assez élevé. Leurs résultats indiquent que les mesures de risque sont affectées par un risque de modèle significatif pendant les périodes de difficultés financières. Ce bref aperçu de la littérature montre que ces deux types d'erreurs de mesure ont souvent une incidence sur la mesure des risques financiers, altérant l'évaluation du risque et biaisant les niveaux de fonds propres réglementaires détenus par les banques.

III. Validation

La nécessité d'une gestion saine des risques financiers et le besoin de validation des mesures de risque n'ont jamais été aussi essentiels que dans l'environnement financier actuel. En particulier, les mesures de risque peuvent être valides du point de vue conceptuel (voir Artzner et al. 1999 et Chen et al. 2013, pour les propriétés désirables des mesures de risque de marché et systémique) mais ne pas être correctement estimées pour autant. De ce fait, la capacité à identifier des modèles de risque mal spécifiés, qui conduisent à une représentation erronée des expositions réelles au risque, également appelée backtesting, revêt une importance cruciale pour les régulateurs et les gestionnaires de risque. Jorion (2007) définit le backtesting comme un ensemble de procédures statistiques consistant à vérifier si les pertes réelles sont en adéquation avec les pertes prédites. Ces procédures visent alors à comparer les prévisions de mesures de risque générées par le modèle historique avec les pertes réelles.

L'objectif des modèles de PD est de prédire le taux de défaut. Pour fournir des prévisions valides, le modèle de PD doit séparer convenablement les demandeurs de crédit en classes de "bon" et "mauvais" risques (Hand and Henley, 1997). Pour cette raison, les techniques de backtesting dédiées à la PD s'appuient généralement sur des outils de discrimination (Gouriéroux, 1992) ou sur de simples tests statistiques binomiaux (Brown et al., 2001). Pour les modèles de LGD, même s'il n'existe pas de directives particulières pour évaluer leurs estimations, Loterman et al. (2014) proposent un cadre de backtesting qui repose sur des tests d'hypothèses statistiques. Kalotay and Altman (2017) montrent que les variations dans la composition des dettes en défaut au moment du défaut génèrent des variations temporelles dans la distribution de la LGD. Ils quantifient l'importance de la prise en compte de cette variation temporelle dans les comparaisons out-of-sample des modèles de LGD. Enfin, pour les crédits standard et les prêts, les EAD sont observées

et il n'est pas nécessaire d'estimer ce paramètre de risque. Cependant, pour les EAD hors bilan, cette quantité devient inconnue et la banque doit estimer un facteur de conversion de crédit (CCF). Gürtler et al. (2018) développent une évaluation théorique et empirique pour la modélisation de l'EAD. Plus généralement, plusieurs approches permettent d'évaluer le modèle de risque de crédit dans son ensemble. Par exemple, Lopez and Saidenberg (2000) développent une méthode d'évaluation des modèles de risque de crédit basée sur des portefeuilles de crédit simulés. Medema et al. (2009) implémentent une méthodologie de validation simplifiée que les banques peuvent utiliser pour valider leur exercice de modélisation du risque de crédit.

La réglementation des marchés financiers applique un contrôle strict des modèles de risque internes utilisés pour le calcul des exigences en fonds propres au titre du risque de marché. L'une des responsabilités principales des banques consiste à réaliser des exercices de backtesting pour certains modèles de risque. Au cours des deux dernières décennies, un certain nombre de contributions a été proposé pour évaluer la capacité des modèles prédictifs à fournir des prévisions de risque acceptables. Ces techniques reposent généralement sur des tests dit de violations. Les prévisions de VaR sont valides lorsque le processus de violation satisfait à l'hypothèse de couverture non conditionnelle (UC). Une autre hypothèse importante pour les prévisions de VaR est l'hypothèse d'indépendance (IND), qui suppose que les violations de VaR observées à deux dates différentes pour le même taux de couverture doivent être indépendamment distribuées. Lorsque les hypothèses UC et IND sont simultanément valides, les prévisions de VaR ont une couverture conditionnelle (CC) correcte et le processus de violation devient une séquence de différence de martingale (voir Christoffersen, 1998, pour une description plus détaillée de ces hypothèses). Engle and Manganelli (2004) développent le test dit *dynamic quantile*, qui se concentre directement sur la corrélation des violations avec la série de rendements observés. Dumitrescu et al. (2012) proposent de raffiner cette approche en remplaçant le modèle de régression linéaire par un modèle de régression dichotomique non linéaire et dynamique. Plusieurs tests de backtesting ont également été proposés pour évaluer la validité de la VaR pour différents niveaux de probabilité. Colletaz et al. (2013) développent un backtest de l'hypothèse UC à deux taux de couverture afin de faire la distinction entre une situation dans laquelle les pertes sont inférieures mais proches de la VaR et une situation dans laquelle les pertes sont considérablement inférieures à la VaR. Hurlin and Tokpavi (2006) utilisent une statistique de portmanteau multivariée pour tester l'hypothèse IND pour plusieurs niveaux de probabilité. Enfin, il est important de mentionner la classe des tests de durée qui tient compte de l'intervalle de temps entre deux violations (Berkowitz et al., 2011; Candelon et al., 2011; Christoffersen and Pelletier, 2004; Pelletier and Wei, 2016).

Depuis peu, le Comité de Bâle préconise l'utilisation de l'ES comme nouvelle mesure de risque réglementaire complétant et remplaçant partiellement la VaR. Ce changement de réglementation a encouragé la communeauté académique au développement de procédures de validation dédiées aux modèles de prévision de l'ES. McNeil and Frey (2000) développent un cadre de backtesting non paramétrique de l'ES basé sur des résidus en excés. Acerbi and Szekely (2014) développent trois nouveaux backtests d'ES basés sur des simulations de Monte-Carlo. Nolde and Ziegel (2017) conçoivent des tests de calibration conditionnels pour évaluer l'ES. Plus récemment, Bayer and Dimitriadis (2018) proposent un backtest fondé sur la régression et qui exploite l'élicitabilité conjointe du couple VaR-ES. Kratz et al. (2018) proposent de généraliser le backtest binomial populaire des exceptions de VaR à un seul niveau de couverture, à un backtest multinomial d'exceptions de VaR à plusieurs niveaux de couverture. En exploitant la relation entre la VaR et l'ES, Kerkhof and Melenberg (2004) fournissent un cadre de backtesting basé sur les dépassements en PIT, qui englobe la VaR et l'ES en tant que cas particuliers. Costanzino and Curran (2015) dérivent un backtest de couverture pour les mesures de risque spectrales telles que l'ES dans l'esprit des backtests traditionnels de couverture de VaR. Du and Escanciano (2017) définissent un processus de violation cumulé pour l'ES, qui généralise le processus de violation de la VaR. Costanzino and Curran (2018) implémentent un backtest de type *traffic light* pour l'ES, qui étend le backtest *traffic light* utilisé pour la VaR.

De façon étonnante, les procédures pour évaluer l'exactitude des mesures de risque systémique sont trés peu développées. Il n'existe aucune procédure statistique formelle permettant d'évaluer cette classe de mesures de risque. Cependant, même si aucune technique formelle n'a été proposée, des tentatives ont permis d'évaluer de manière empirique le contenu prédictif des mesures de risque systémique. Idier et al. (2014) étudient les firmes ayant un score de risque systémique élevé et leur probabilité de subir les pertes financières les plus importantes en cas de crise financière. Wu and Zhao (2018) cherchent à savoir si ces firmes risquent davantage de devenir insolvables. Brownlees and Engle (2017) montrent que les banques dont la SRISK était élevée avant la crise financière risquaient davantage d'être sauvées par le gouvernement et de recevoir des injections de capital de la Réserve Fédérale. Engle et al. (2015) comparent le classement des institutions financières européennes obtenu avec la SRISK à la liste des SIFIs produites par le FSB. Récemment, Brownlees et al. (2018) ont proposé une évaluation historique de la SRISK et de la ΔCoVaR basée sur deux dimensions. La première, appelée *SIFI ranking challenge*, consiste à déterminer si le classement des institutions financières construit au moyen de la SRISK et de la ΔCoVaR permet d'identifier les institutions dont les dépôts ont sensiblement baissé autour d'événements de panique. La seconde, intitulée *the financial crisis prediction challenge*, cherche à établir si ces mesures de risque systémique

sont des prédicteurs significatifs de la baisse des dépôts à l'échelle du système pendant les événements de panique.

Les indicateurs systémiques énumérés ci-dessus sont tous construits à partir de données accessibles au public, telles que les actions, les rendements d'actifs, les prix d'options, ou les spreads de CDS. Bien entendu, on peut également s'interroger sur la validité des méthodes de mesure du risque systémique reposant sur des données propriétaires, telles que les données de bilan, les données de positions croisées, la taille, l'effet de levier, la liquidité, ou les interconnexions. Ces méthodes propriétaires ont été intégrées dans la boite à outils des autorités de contrôle bancaire car elles reposent sur davantage de bases théoriques que celles fondées sur les données des marchés financiers. Même si l'accès aux données est rendu plus difficile, voire impossible, plusieurs tentatives ont été faites pour évaluer la validité des méthodes de mesure du risque systémique issues de données privées. Philippon et al. (2017) fournissent une première tentative d'évaluation empirique de la qualité des tests de résistance du secteur bancaire organisé par l'Autorité Bancaire Européenne (EBA) en 2014. Ils constatent que les tests de résistance sont informatifs et fournissent aux régulateurs des indications fiables sur la résilience des banques. Benoit et al. (2019) identifient plusieurs faiblesses dans la méthodologie de notation du risque systémique actuellement utilisée pour identifier et réglementer les SIFIs. Ils proposent une nouvelle méthodologie pour pallier à ces lacunes qui améliore l'allocation du capital réglementaire entre les banques.

IV. Contribution

Dans ce nouveau contexte, notre recherche porte sur les mesures de risque financier et les techniques de validation dédiées à leurs modèles prédictifs. L'objectif général de cette thèse est de fournir des outils avancés pour l'évaluation des estimations de mesure du risque. Nos développements méthodologiques pour l'évaluation des mesures de risque couvrent trois grandes catégories de risques financiers : (*i*) le risque de crédit, (*ii*) le risque de marché, et (*iii*) le risque systémique. Pour chacune de ces catégories, notre objectif reste globalement le même: améliorer la solidité du système bancaire grâce au développement de méthodes de validation performantes des estimations de risque. Dans le cadre du risque de crédit, cette thèse contribue à améliorer la fiabilité des prévisions des pertes futures générées par les portefeuilles de prêts et permet une allocation plus efficace du capital réglementaire. En ce qui concerne le risque de marché, nos travaux visent à améliorer la qualité des pratiques de gestion d'actifs afin de couvrir de manière adéquate les fonds d'investissement en cas de chocs défavorables et de pertes potentielles. Enfin, cette thèse contribue également au renforcement de la stabilité financière dans son ensemble en améliorant le suivi des banques via une identification précise des SIFIs par le

biais des mesures de risque systémique. Ce travail se compose en trois chapitres (articles) qui peuvent être étudiés indépendamment les uns des autres.

Le premier chapitre traite des questions liées à l'évaluation du risque de crédit. Nous nous concentrons sur la mesure de risque LGD et proposons une méthode de comparaison de modèle originale qui sélectionne le modèle prédictif de la LGD induisant les erreurs d'estimation les plus faibles sur le capital réglementaire, et qui de ce fait, améliore la solvabilité des banques. Les chapitres 2 et 3 examinent la validité des mesures de risque fondées sur des données de marché. Le chapitre 2 répond aux exigences des régulateurs de fournir des outils de validation plus efficaces pour la mesure de risque ES. Nous développons une nouvelle approche pour évaluer la validité des modèles prédictifs d'ES basée sur des régressions quantiles multivariées. Puisque notre méthodologie repose sur la relation entre la VaR et l'ES, cette nouvelle classe de tests statistiques est conforme aux directives réglementaires de Bâle en vigueur qui recommandent d'effectuer un backtest de l'ES en vérifiant la validité de deux VaRs de la distribution des pertes du portefeuille. De plus, l'exploitation de notre procédure d'évaluation permet de proposer une technique d'ajustement des prévisions d'ES imparfaites, qui sont alors débarrassées du risque d'estimation et du risque de modèle. Dans le chapitre 3, nous examinons les mesures de risque systémique issues de données de marché et la qualité de leurs prévisions. Nous nous appuyons sur les procédures standard de backtesting de la VaR et développons un premier backtest pour l'hypothèse UC et un second pour l'hypothèse IND. À notre connaissance, il s'agit de la première procédure statistique de backtesting dédiée aux mesures de risque systémique. Nous exploitons alors notre méthodologie pour fournir un *early warning system* (EWS) qui démontre une capacité remarquable à détecter les premiers signes de la crise. Dans ce qui suit, nous synthétisons le contenu de chaque chapitre.

Chapitre 2: Loss functions for Loss Given Default model comparison

Le chapitre 2, "Loss functions for Loss Given Default model comparison", propose une méthode originale de comparaison des modèles de *loss given default* (LGD), basée sur des fonctions de pertes espérées exprimées en termes d'exigences de fonds propres réglementaires.[1] Dans le cadre réglementaire de couverture du risque de crédit, le niveau de capital réglementaire est déterminé de manière à couvrir la perte de crédit inattendue de la banque (BCBS, 2005). Pour déduire cette perte inattendue, le Comité de Bâle fournit un cadre théorique basé sur le modèle *Asymptotic Single Risk Factor* (ASRF), inspiré du modèle fondateur de Merton (Merton, 1974; Vasicek, 2002). Ce dernier, à partir de l'estimation de paramètres de risque additionnels, permet de calculer l'exigence en capital règlementaire de la banque. Un des paramètres essentiel dans ce calcul est la LGD.

[1] D'après Hurlin, Leymarie et Patin (2018) publié dans *European Journal of Operational Reasearch*.

La LGD peut être définie au sens large comme le ratio de pertes (exprimé en pourcentage de l'exposition au défaut) qui ne sera jamais recouvré par le prêteur, ou de manière équivalente à un moins le taux de recouvrement. La LGD entre dans la formule du capital réglementaire de manière linéaire, et à ce titre, toute sous-estimation de ce paramètre induit une sous-estimation du capital réglementaire entraînant un amoindrissement de la solvabilité bancaire.

Conformément à l'approche *advanced internal rating-based* adoptée par la plupart des grandes banques internationales, les prévisions de LGD sont issues de modèles de risque internes. Aucune directive particulière n'a été proposée concernant la manière dont les modèles de LGD devraient être évalués, comparés, puis sélectionnés. En conséquence, la méthode de référence consiste simplement à évaluer les prévisions de LGD avec des critères statistiques standard tels que l'erreur quadratique moyenne ou l'erreur absolue moyenne calculées entre les LGD observées et les prévisions, comme pour toute variable continue. Par conséquent, la comparaison actuelle des modèles de LGD est effectuée indépendamment des autres paramètres de risque Bâlois (EAD, PD, maturité, etc.) et en négligeant l'impact des erreurs de prévision de la LGD sur le capital réglementaire. Cette approche peut conduire à sélectionner un modèle de LGD présentant la plus petite erreur quadratique moyenne parmi tous les modèles concurrents, mais induisant de petites erreurs sur les petites expositions, et de grandes erreurs sur les grandes expositions.

Ce chapitre vise à remédier à ces faiblesses en développant une méthode de comparaison alternative qui renforce la solvabilité des banques. Contrairement à l'approche existante qui sélectionne le modèle minimisant les erreurs d'estimation sur la LGD elle-même, notre méthode de comparaison sélectionne le modèle associé aux erreurs d'estimation les plus faibles sur le capital réglementaire. Nous montrons théoriquement que notre approche classe les modèles différemment par rapport à l'approche traditionnelle qui se concentre uniquement sur les erreurs de prévision de la LGD.

À l'aide d'un échantillon de contrats de crédit et de leasing fournis par une banque internationale, nous illustrons l'intérêt de notre méthode en comparant les classements de six modèles de LGD concurrents. Nos résultats empiriques montrent clairement que les classements de modèles basés sur les fonction de pertes de charge en capital diffèrent considérablement de ceux basés sur les fonctions de perte de LGD actuellement considérées par les régulateurs, les banques, et la communeauté académique. La méthode proposée permet d'identifier les meilleurs modèles de LGD associés aux erreurs d'estimation les plus faibles sur le capital réglementaire. Au-delà de ces critères statistiques traditionnels, nous introduisons également des critères asymétriques spécialement conçus pour améliorer la stabilité financière. Ces fonctions de perte pénalisent les erreurs de prévision de la LGD qui conduisent à une sous-estimation du capital réglementaire. Nous constatons que le classement basé sur des critères symétriques est radicalement différent

du classement des modèles obtenu avec des critères asymétriques, ce qui met en évidence l'utilité des fonctions asymétriques pour améliorer la solidité et la stabilité du système bancaire.

Chapitre 3: Backtesting Expected Shortfall via Multi-Quantile Regression

Le chapitre 3, "Backtesting Expected Shortfall via Multi-Quantile Regression", propose une nouvelle approche d'évaluation de la qualité des prévisions de la mesure de risque *expected shortfall* (ES) fondée sur la régression quantile.[2] Dans le contexte de la régulation des marchés financiers et de la supervision bancaire, les accords de Bâle III ont accordé une place importante à l'ES dans le calcul des exigences de fonds propres pour le risque de marché, complétant, et se substituant pour partie à la mesure de risque plus familière connue sous le nom de *value-at-risk* (VaR) (BCBS, 2010). En tant que mesure de risque alternative, l'ES offre un certain nombre de propriétés intéressantes remédiant aux insuffisances théoriques de la VaR. En particulier, l'ES est une mesure de risque cohérente, ce qui signifie qu'elle satisfait les propriétés de monotonicité, de sous-additivité, d'homogénéité, et d'invariance par translation (voir Artzner et al., 1999; Acerbi and Tasche, 2002). Le BCBS souligne le rôle important que joue l'ES à la place de la VaR "*pour assurer une capture plus prudente du 'risque extrême' et de l'importance de la mobilisation d'un capital adéquat en période de fortes tensions sur les marchés financiers*" (BCBS, 2016, page 1).

Compte tenu du changement en faveur de l'ES, le principal défi consiste à développer des méthodes de modélisation appropriées pour l'ES (voir les travaux récents de Taylor, 2019; Patton et al., 2019, entre autres), et à construire des outils avancés pour en évaluer les prévisions. Ce chapitre se concentre précisément sur le second point. En effet, les procédures de validation et de backtest sont des critères essentiels pour qu'une mesure de risque puisse obtenir le statut de standard dans l'industrie. Plus important encore, la validité des estimations d'ES est cruciale étant donné que ce paramètre entre dans le calcul du capital réglementaire pour le risque de marché. Par conséquent, toute sous-estimation de l'ES qui n'a pas été identifiée à temps peut menacer la solvabilité des banques.

Dans ce chapitre, nous proposons une extension naturelle des backtests standard de VaR, qui nous permet de tester les estimations de VaR à plusieurs niveaux de probabilité conjointement. Étant donné que l'ES peut être définie comme une fonction de VaR pour différents niveaux de probabilité appartenant à la queue de distribution des pertes du portefeuille, notre approche peut être considérée comme délivrant un backtest implicite

[2]D'après Couperier et Leymarie (2019), actuellement R&R dans *Journal of Business and Economic Statistics*.

pour l'ES. Notre stratégie de test est conforme aux recommandations générales des superviseurs financiers. Selon les directives du BCBS sur l'évaluation de l'ES, "*les exigences en matière de backtesting reposent sur la comparaison de la mesure de VaR à un jour [...] à la fois au percentile d'ordre 97,5 et au percentile d'ordre 99*" (BCBS, 2016, page 57). Pour mettre en oeuvre notre stratégie de test, nous développons un cadre de validation basé sur la régression quantile multivariée. La procédure étend le test de Gaglianone et al. (2011) qui permet de valider la VaR pour un niveau de probabilité unique.

Notre approche présente de nombreux avantages. Premièrement, notre procédure est flexible car l'utilisateur peut choisir le nombre et les valeurs de quantiles pour l'évaluation de l'ES et peut facilement se concentrer sur divers aspects de la queue de distribution du modèle de prévision. Deuxièmement, notre méthodologie présente l'avantage d'être conforme aux directives réglementaires qui consistent à vérifier si le modèle d'ES sous-jacent fournit des quantiles adaptés aux niveaux de probabilité 0,975 et 0,990. Enfin, la procédure est facile à mettre en oeuvre pour les gestionnaires de risque et les autorités de contrôle, car elle repose sur la validation de la VaR qui est la mesure de référence historique. Pour ces raisons, cet outil de validation est susceptible d'être adopté par les institutions financières en tant que nouvelle norme de gestion des risques financiers.

Pour illustrer les avantages de notre méthode, nous évaluons les estimations d'ES calculées à partir d'un modèle AR(1)-GARCH(1,1) en supposant que les rendements du portefeuille de l'investisseur sont construits à partir de l'indice S&P500 sur la période 2007-2012. Pendant cette période d'instabilité financière, la procédure conclut que les prévisions d'ES sont incorrectes. Nos résultats suggèrent également que nous devrions être très prudents lorsque nous utilisons une VaR unique pour évaluer la distribution de perte du portefeuille. En outre, deux VaRs comme le recommande les autorités de contrôle financières ne permettent pas toujours d'identifier des prévisions de risque erronées et peuvent conduire à un niveau inexact d'exigences de fonds propres menaçant la stabilité financière. De manière générale, nous trouvons que l'évaluation de quatre à six VaRs dans la distribution des pertes du portfeuille améliore la capacité à rejeter un modèle de prévision d'ES incorrect, ce que le superviseur devrait prendre en considération dans ses futures directives règlementaires.

Chapitre 4: Backtesting Marginal Expected Shortfall and Related Systemic Risk Measures

Le chapitre 4, "Backtesting Marginal Expected Shortfall and Related Systemic Risk Measures", propose la toute première procédure statistique d'évaluation des mesures de risque systémique qui sont construites à partir de données de marché.[3] L'objectif général

[3]D'après Banulescu-Radu, Hurlin, Leymarie et Scaillet (2018). Ce chapitre a reçu une subvention de recherche de la Fondation Banque de France.

d'une mesure de risque systémique est l'identification des vulnérabilités du système financier. Dans la pratique, il existe trois façons de mesurer le risque systémique (voir Benoit et al., 2017, pour une revue de littérature). Une première approche, appelée approche prudentielle, repose sur des données propriétaires relatives à la firme que sont la taille, l'effet de levier, la liquidité, l'interconnexion, la complexité et la substituabilité, ainsi que sur une méthode de notation (Benoit et al., 2019). Une deuxième approche repose sur des modèles structurels qui identifient des sources spécifiques de risque systémique, telles que la contagion, les *bank runs*, ou des crises de liquidité. Une troisième approche vise à dériver des mesures globales de risque systémique sur la base de données de marché accessibles au public, telles que les rendements d'actions, les prix d'options ou les spreads de CDS. Trois exemples remarquables de mesures de risque systémique basées sur des données de marché sont la *marginal expected shortfall* (MES) de Acharya et al. (2017), la *systemic risk measure* (SRISK) de Acharya et al. (2012) et de Acharya et al. (2017), et la *delta conditional value-at-risk* (ΔCoVaR) de Adrian and Brunnermeier (2016).

Cependant, ces approches méthodologiques peuvent aboutir à des conclusions opposées, ce qui a provoqué un vif débat dans les sphères universitaires et réglementaires au cours de ces dernières années. Cela s'est par exemple produit en octobre 2014, lorsque l'EBA a révélé l'effet des pertes dans un test de résistance sur les ratios de fonds propres bancaires et a conclu que les banques françaises étaient parmi les banques les plus sûres de l'eurozone. Viral Acharya (Financial Times, 27 octobre 2014) a immédiatement mis en doute ces conclusions, affirmant que les banques françaises étaient les plus risquées en Europe, selon les résultats de la SRISK affichés sur le *Volatility Lab* (Université de New York). Ces controverses soulèvent la question de la validation des tests de résistance (Philippon et al., 2017), mais aussi des mesures de risque systémique. Néanmoins, à notre connaissance, aucune procédure de backtesting basée sur une approche économétrique solide n'a encore été proposée pour les mesures de risque systémique.

Dans ce contexte, ce chapitre propose le premier cadre général de backtesting des mesures de risque systémique. Notre stratégie de test est similaire à celle des backtests standard utilisés pour la VaR qui exploitent la propriété de séquence de différence de martingale (mds) suivie par les processus de violation (Kupiec, 1995; Christoffersen, 1998, 2010; Berkowitz et al., 2011, entre autres). La principale nouveauté technique de notre approche consiste à exploiter la propriété mds pour aboutir à ce que nous appelons le processus de violation joint cumulé. Cette nouvelle catégorie de violation est spécialement conçue pour évaluer la validité des mesures de risque systémique et généralise au cas bivarié le processus de violation cumulé de Du and Escanciano (2017) utilisé pour le backtesting de l'ES. Dans un premier temps, nous concentrons nos recherches sur la MES, puis nous étendons notre procédure de backtesting aux autres mesures de risque

systémique (SRISK, SES, ΔCoVaR). L'exploitation de la propriété mds du processus de violation joint cumulé permet la mise en oeuvre de divers types de backtests pour les indicateurs de risque systémique. Nous proposons ici deux tests basés sur les hypothèses UC et IND (Christoffersen, 1998). Ces procédures statistiques sont faciles à mettre en oeuvre et similaires à celles actuellement utilisées par les gestionnaires de risque pour évaluer les mesures de risque de marché. Ces backtests peuvent donc être aisément adoptés dans la règlementation financière.

Nos résultats empiriques, obtenus à partir de banques américaines, révèlent que les prévisions de risque systémique journalières sont non valides pour un grand nombre de banques avant la crise financière des subprimes. Cependant, lorsque nous considérons un horizon de prévision plus long (un mois), nos tests concluent que les prévisions de risque systémique sont valides, ce qui suggère que ces indicateurs sont plus aptes à rendre compte de la dynamique de long terme du risque systémique. Enfin, nous montrons que notre procédure peut également être utilisée en tant que *early warning system* (EWS) en cas de crise systémique. Dans cette perspective, nous introduisons un indicateur EWS correspondant à la différence entre la prévision de risque systémique issue d'un modèle de risque potentiellement mal spécifié et sa contrepartie ajustée. Cette dernière est obtenue à partir de notre procédure de backtesting et représente la prévision de risque systémique pour laquelle nous ne rejetons pas l'hypothèse nulle UC. Cet ajustement est obtenu en adaptant le niveau de probabilité de sévérité de la crise de la mesure de risque systémique considérée. Cette technique a déjà été proposée pour les mesures de risque de marché telles que l'ES et la VaR (voir Boucher et al., 2014; Lazar and Zhang, 2019, par exemple), mais elle n'était pas encore disponible pour les mesures de risque systémique en raison de l'absence de bases théoriques solides pour le backtesting de cette classe de mesure de risque. Notre indicateur EWS affiche une forte hausse avant les premiers signes de la crise et atteint son maximum lors de l'effondrement historique de Lehman Brothers. Par conséquent, il peut fournir des informations utiles en temps réel pour la surveillance du système financier, compléter la boite à outils utilisée par les universitaires et les régulateurs pour saisir l'accumulation du risque systémique par temps calme, et améliorer l'efficacité de l'allocation du capital réglementaire entre les banques.

Enfin, le chapitre 5 résume les principaux résultats de cette thèse et expose plusieurs objectifs pour des recherches futures.

Nederlandse Samenvatting

Risicomanagement is een belangrijk vakgebied voor financiële instellingen, zoals banken, verzekeringsmaatschappijen en beleggingsfondsen. Een van de belangrijkste lessen die we hebben geleerd van de wereldwijde economische en financiële crisis is dat het meten van risico nog noodzakelijker moet worden. De almaar toenemende omvang en de complexiteit van financiële instellingen, in het tempo van hun financiële transacties, hebben in het huidige financiële klimaat ontegenzeggelijk een nieuwe variabele geïntroduceerd wat betreft het risicomanagement. Tegelijkertijd heeft de technologische vooruitgang in communicatie en dataverzameling geleid tot lagere kosten voor het verwerven, beheren en analyseren van data om risico's te monitoren en om de complexere en razendsnelle bedrijfsomgeving te schetsen. Deze context heeft geleid tot de ontwikkeling van geavanceerde financiële instrumenten en nieuwe technieken voor risicobeheer. Het wetenschappelijke onderzoek in financiële econometrie heeft met name een impuls en richting gegeven aan (*i*) de ontwikkeling van nieuwe financiële risicomaatstaven, (*ii*) de introductie van adequate ramings- en inferentiemethoden en (*iii*) de implementatie van validatietechnieken die zouden moeten zijn gewijd aan die indicatoren.

In deze vernieuwde context richt ons onderzoek zich op financiële-risicomaatregelen en de validatietechnieken die zijn toegewijd aan hun voorspellende modellen. Ons methodologische ontwikkelingen omvatten drie belangrijke categorieën van financieel risico, namelijk (*i*) kredietrisico, (*ii*) marktrisico en (*iii*) systeemrisico. Ons proefschrift draagt wat betreft het kredietrisico bij aan het verbeteren van de betrouwbaarheid van verliesramingen van kredietportefeuilles en het effectiever maken van de wettelijke kapitaalallocatie. Wat betreft het marktrisico is ons werk erop gericht om vermogensbeheerpraktijken gezonder te maken zodat beleggingsmaatschappijen een goede dekking hebben tegen ongunstige marktschokken en mogelijke verliezen. Tot slot draagt dit proefschrift ook bij aan de versterking van de financiële stabiliteit als geheel door het verbeteren van de monitoring van banken door de systeemrelevante financiële instellingen (*systemically important financial institutions*, SIFI's) nauwkeurig te identificeren middels systeemrisicomaatregelen. Dit werk is concreet gemaakt in drie hoofdstukken (artikelen) die onafhankelijk van elkaar kunnen worden bestudeerd. Hieronder beschrijven we de inhoud van elk hoofdstuk.

In het eerste hoofdstuk worden onderwerpen besproken die gerelateerd zijn aan kredietrisicobeoordeling. Het wettelijk kader voor kredietrisico bepaalt het vereiste niveau voor het wettelijk kapitaal om onverwachte kredietverliezen van een bank af te dekken (BCBS, 2005). Om de hoogte van dit onverwachte verlies af te leiden, geeft het Basel Committee een theoretisch kader dat is gebaseerd op het *asymptotic single risk factor* (ASRF)-model, dat is geïnspireerd op het baanbrekende Metron-Vasicek "*model of the firm*" (Merton, 1974; Vasicek, 2002). Op basis van het ASRF-model en enkele externe geschatte risicoparameters is het dan mogelijk om het vereiste kapitaal voor kredietrisico te berekenen. In deze formule is de standaardwaarde bij verlies *loss given default* (LGD) een van belangrijkste parameters. De LGD kan globaal worden gedefinieerd als de verhouding tussen de verliezen (uitgedrukt als percentage van de openstaande kredieten bij wanbetaling) die nooit door de kredietgever zullen worden teruggevorderd of gelijkwaardig als één minus de terugvorderingsratio. Omdat de LGB-formule de hoogte van het wettelijk kapitaal op een lineaire manier benaderd, zal elke onderschatting van deze risicoparameter leiden tot een onderschatting van het wettelijk kapitaal en tot de laagste solvabiliteit van een bank. In dit hoofdstuk wordt een nieuwe, originele vergelijksmethodiek voorgesteld waarin het LGD-voorspellende model wordt geselecteerd dat de laagste schattingsfouten op de hoogte van het wettelijk kapitaal veroorzaakt. Aan de hand van een steekproef van krediet- en leasecontracten, die aan ons beschikbaar zijn gesteld door een internationale bank, illustreren we het belang van onze methode door de rangorde van zes concurrerende LGD-modellen te vergelijken. Onze empirische bevindingen tonen duidelijk aan dat rangordes van modellen op basis van kapitaalverliezen aanzienlijk verschillen van rangordes die gebaseerd zijn op de LGD-verliesfuncties, die op dit moment worden gebruikt door toezichthouders, banken en wetenschappers. We laten vooral zien dat de voorgestelde methode de solvabiliteit van de bank verbetert en de draagkracht en stabiliteit van het bancaire systeem versterkt.

In het tweede hoofdstuk wordt een nieuwe backtesting-procedure voor het verwachte tekort (*expected shortfall*, ES) voorgesteld. We raden een logische voortzetting aan van standaard backtesting-procedures voor *value-at-risk* (VaR) waarmee we VaR-schattingen gezamenlijk kunnen testen op verschillende waarschijnlijkheidsniveaus. Omdat ES breed kan worden gedefinieerd als een functie van VaR op verschillende waarschijnlijkheidsniveaus langs de kansverdeling van de staart (*tail distribution*) van het portefeuilleverlies, kan onze aanpak worden beschouwd als een impliciete backtest voor ES. Deze nieuwe klasse van statistische toetsen is consistent met de huidige wettelijke Basel-richtlijnen die het backtesten van ES aanbevelen door de geldigheid van verschillende VaR's in de *tail distribution* van het portefeuilleverlies te verifiëren (BCBS, 2016, pagina 57). Om onze onderzoeksstrategie te implementeren, hebben we een validatieraamwerk ontwikkeld dat is gebaseerd op multivariate kwantiele regressie. De procedure verruimt

de test van Gaglianone et al. (2011) waarmee VaR op één waarschijnlijkheidsniveau kan worden getoetst. Om de voordelen van onze methode te illustreren, beoordelen we ES-schattingen middels een AR(1)-GARCH(1,1)-model, waarbij wordt verondersteld dat de portefeuillerendementen van een belegger wordt gegeven door de S&P 500-index voor de periode 2007-2012. De procedure concludeert dat de schattingen van ES misleidend zijn in deze periode van financiële onrust. Onze resultaten suggereren ook dat men zeer voorzichtig moet zijn in het gebruik van een enkele VaR om de *tail distribution* van het portefeuilleverlies te bepalen. Bovendien is het gebruik van twee VaR's, zoals wordt geadviseerd door financiële toezichthouders, niet altijd voldoende om onjuiste risicovoorspellingen te identificeren, wat derhalve kan leiden tot een foutief niveau voor de marktrisico-kapitaalvereisten, wat de financiële stabiliteit bedreigt. Een algemeen resultaat is dat we aantonen dat vier tot zes VaR's voor het bepalen van de *tail distribution* een betere mogelijkheid geeft om een onjuist ES-model te identificeren, waar financiële toezichthouders dienovereenkomstig rekening mee moeten houden. Ten slotte gebruiken we onze onderzoeksstrategie om een nieuwe, originele techniek te bieden die de imperfecte ES-voorspellingen aanpast en deze voorspelling schoont van schattingsrisico en modelrisico.

In het derde hoofdstuk ontwikkelen wij de eerste statistische procedure voor het bepalen van markt-gebaseerde systeemrisicomaatstaven. Onze onderzoeksstrategie volgt de uitgangspunten van de backtesten die standaard worden gebruikt voor de VaR. Deze backtesten maken gebruik van de *martingale difference sequence* (mds)-eigenschap van een schendingsproces (*violation process*) (onder andere Kupiec, 1995; Christoffersen, 1998, 2010; Berkowitz et al., 2011). De belangrijkste technische noviteit van onze benadering is het gebruik van de mds-eigenschap voor, zoals wij dat noemen, het cumulatieve gezamenlijke schendingsproces (*cumulative joint violation process*). Deze nieuwe schendingsklasse is specifiek aanbevolen om de geldigheid van systeemrisicomaatregelen te bepalen en breidt de bivariate casus uit naar het cumulatieve schendingsproces van Du en Escanciano (2017) voor het backtesten van ES. Wij richten ons onderzoek eerst op het marginaal verwachte tekort en daarna breiden we onze backtesting-procedure uit naar andere systeemrisicomaatstaven (SRISK, SES, CoVaR). Door gebruikmaking van de mds-eigenschap van een *cumulative joint violation process* kunnen verschillende soorten backtesten van systeemrisico-indicatoren worden geïmplementeerd. We stellen twee toetsen voor die zijn gebaseerd op de zogenaamde UC-hypthese en de IND-hypothese (Christoffersen, 1998). Deze toetsen zijn bedoeld om te verifiëren of de voorspellingen van systeemrisicomaatstaven in overeenstemming zijn met de ex-post verliezen door het aantal en de correlatie van de cumulatieve gezamenlijke schendingen te controleren. Onze empirische resultaten, die zijn gebaseerd op Amerikaanse banken, tonen aan dat de voorspellingen voor het systeemrisico voor één dag vooruit misleidend zijn voor een grote

subset van banken, voordat de systeemrisicocrisis zich voordoet. Echter, wanneer we rekening houden met een langere voorspellingshorizon (één maand), dan kan worden geconcludeerd dat de systeemrisicovoorspellingen geldig zijn. Daarnaast suggereren onze toetsen dat deze indicatoren geschikter zijn om de dynamiek van systeemrisico op lange termijn vast te leggen. Tot slot laten wij zien dat onze procedure ook kan worden gebruikt als een vroegtijdig waarschuwingssysteem (*early warning system*, EWS) voor een systeemrisicocrisis. Daartoe introduceren wij een EWS-indicator die wordt gedefinieerd als het verschil tussen de systeemrisicovoorspelling afgegeven door een mogelijk verkeerd gespecificeerd risicomodel en de gecorrigeerde tegenhanger. Onze EWS-indicator toont voor de meeste Amerikaanse banken een scherpe stijging vóór de eerste tekenen van de crisis. Onze EWS-indicator kan nuttige inzichten geven voor het realtime monitoren van het financiële systeem en voor het completeren van de toolbox die wetenschappers en wetgevers gebruiken om de opbouw van systeemrisico's in rustige tijden vast te leggen.

Curriculum Vitae

Jérémy Leymarie was born on July 21, 1989, in Brive-La-Gaillarde, France. In 2013, he obtained a Bachelor with honors in Economics at the University of Auvergne (France). Then, he integrated the Master in Econometrics at the University of Orléans (France). He completed this master with honors in 2015.

Upon graduation, Jérémy started his Ph.D. research at the University of Orléans and Maastricht University, under the supervision of Prof. Christophe Hurlin and Prof. Alain Hecq. His work has been concretized so far in three papers (currently either submitted or published in well-known international journals such as the *European Journal of Operational Research* or the *Journal of Business and Economic Statistics*). Besides, he received a grant from the Banque de France Foundation for his chapter entitled "Backtesting Marginal Expected Shortfall and Related Systemic Risk Measures" coauthored with C. Hurlin, O. Scaillet, and D. Banulescu. He also received the best paper award at the annual meeting of the German Finance Association in 2019 for his chapter entitled "Backtesting Expected Shortfall via Multi-Quantile Regression" coauthored with O. Couperier.

Jérémy has presented his research in more than 40 international conferences, workshops, and seminars, including recently the 26th Annual Meeting of the German Finance Association (Germany, 2019), the 72nd European Meeting of the Econometric Society (UK, 2019), the 36th International Conference of the French Finance Association (Canada, 2019), the 12th Annual Conference of the Society for Financial Econometrics (China, 2019), the 4th International Workshop on "Financial Markets and Nonlinear Dynamics" (France, 2019), the 12th Financial Risks International Forum (France, 2019), the 12th International Conference on Computational and Financial Econometrics (Italy, 2018).

In January 2020, Jérémy will join the Department of Statistics and Operations Research at the University of Vienna (Austria) on a postdoctoral position.