



HAL
open science

Three Essays in Financial Econometrics

Ophélie Couperier

► **To cite this version:**

Ophélie Couperier. Three Essays in Financial Econometrics. Economics and Finance. Institut Polytechnique de Paris, 2022. English. NNT : 2022IPPAG011 . tel-03978858

HAL Id: tel-03978858

<https://theses.hal.science/tel-03978858>

Submitted on 8 Feb 2023

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

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



INSTITUT
POLYTECHNIQUE
DE PARIS

NNT : 2022IPPAG011

Thèse de doctorat



Three Essays in Financial Econometrics

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à l'École Nationale de la Statistique et de l'Administration Économique

École doctorale n°626 École doctorale de l'Institut Polytechnique de Paris (EDIPP)
Spécialité de doctorat : Sciences Économiques

Thèse présentée et soutenue à Palaiseau, le 2 décembre 2022, par

OPHÉLIE COUPERIER

Composition du Jury :

Valérie Mignon
Professeure, Université Paris Nanterre

Présidente
Rapporteuse

Olivier Darne
Professeur, Université de Nantes

Rapporteur

Sophie Dabo-Niang
Professeure, Université de Lille

Examinatrice

Jean-Paul Renne
Professeur, HEC Lausanne

Examineur

Christian Francq
Professeur, CREST

Directeur de thèse

Christophe Hurlin
Professeur, Université d'Orléans

Co-directeur de thèse

Jean-Michel Zakoïan
Professeur, CREST

Invité
Co-encadrant de thèse

Je dédie cette thèse à mon mari,

Contents

Remerciements	iii
List of Figures	vii
List of Tables	ix
1 Introduction	1
1.1 Modeling Financial Risks	1
1.1.1 Financial Risk Measures	2
1.1.2 Financial Risk Models	3
1.2 Backtesting Financial Risks	7
1.2.1 Traditional Backtests	7
1.2.2 Comparative Backtests	10
1.3 Contribution	12
2 Daily Volatility Forecasting using Intraday Returns and Functional Co- variates	21
2.1 Introduction	21
2.2 Stationarity and existence of moments	23
2.2.1 Stationarity of the first-order model	24
2.2.2 Strict stationarity of the general model	24
2.2.3 Existence of moments and log-moments for the general model	25
2.3 Statistical inference	27
2.3.1 Parametric representation of the functional part	27
2.3.2 Quasi-maximum likelihood estimation	28
2.3.2.1 Strong consistency	29
2.3.2.2 Asymptotic normality	30
2.3.3 Goodness-of-fit portmanteau tests	31
2.4 Simulations	32
2.5 Empirical Study	35
2.5.1 Data description	35

2.5.2	In-sample estimation	35
2.5.3	Model comparison	41
2.5.3.1	Goodness-of-fit	41
2.5.3.2	Out-of-sample forecasting	44
2.6	Conclusion	45
2.7	Appendix	48
2.7.1	Appendix A: Proofs	48
2.7.2	Appendix B: Algorithm for rational representation of the functional parameter	60
2.7.3	Appendix C: Empirical application	61
3	Backtesting Expected Shortfall via Multi-Quantile Regression	63
3.1	Introduction	63
3.2	Multi-quantile regression framework	67
3.2.1	ES as an approximation of VaRs	67
3.2.2	Multi-quantile regression model	68
3.2.3	Parameter estimation and asymptotic properties	69
3.3	Backtesting ES	70
3.3.1	The backtests	70
3.3.2	Finite sample inference	72
3.4	Simulation study	73
3.5	Empirical application	79
3.5.1	Data	79
3.5.2	Empirical results	80
3.5.3	Adjusted ES forecasts	85
3.6	Conclusion	87
3.7	Appendix	88
3.7.1	Appendix A: Application of a finite Riemann sum to ES	88
3.7.2	Appendix B: Assumptions	88
3.7.3	Appendix C: Consistent variance-covariance matrix estimation	89
3.7.4	Appendix D: On the rate of convergence and interplay of \mathbf{p} and \mathbf{T}	90
3.7.5	Appendix E: Proof of consistency under fixed untrue hypothesis	91
3.7.6	Appendix F: Robustness checks of the simulation study	92
3.7.7	Appendix G: Exact calculation method of ES	95
3.7.8	Appendix H: Robustness checks of the empirical application	96
4	Elicitability of Marginal Expected Shortfall and Related Systemic-Risk Measures	103
4.1	Introduction	103

4.2	Marginal expected shortfall and related systemic-risk measures	107
4.3	The scoring function	109
4.3.1	Consistency	110
4.3.2	Quasi-convexity	113
4.4	M-estimation	115
4.5	Numerical illustrations	118
4.5.1	Simulation study	118
4.5.1.1	The experiment	118
4.5.1.2	Simulation results	119
4.5.2	Empirical application	123
4.5.2.1	Empirical setup	123
4.5.2.2	Main results	125
4.6	Conclusion	127
4.7	Appendix	129
4.7.1	Appendix A: Spectral risk measure	129
4.7.2	Appendix B: Assumptions	130
4.7.3	Appendix C: Proofs	132
4.7.4	Appendix D: Some results on \mathbf{e} in conditional volatility model . .	144
4.7.5	Appendix E: Characterization of the asymptotic distribution . . .	146
5	Conclusion	149
	References	153
	Résumé en Français	166

Remerciements

Je souhaite dans un premier temps remercier chaleureusement toutes les personnes qui ont contribué à la réussite de cette thèse, que ce soit d'un point de vue technique et scientifique mais également en temps que soutien moral.

Mes premiers mots sont pour Prof. Christian Francq, Prof. Christophe Hurlin et Prof. Jean-Michel Zakoïan, mes trois directeurs de thèse. Merci pour votre bienveillance, votre patience et les connaissances que vous m'avez transmises. J'ai eu la chance de vous avoir en tant que professeurs dans un premier temps et vous m'avez donné le plus important, l'envie de poursuivre mes études. Vous m'avez ensuite, en tant que mentors, inspirée, guidée et aidée à devenir la jeune chercheuse et enseignante que je suis aujourd'hui. Je tiens également à exprimer ma sincère gratitude au Prof. Olivier Scaillet, un collègue précieux de très bons conseils, qui trouve toujours du temps pour partager ses connaissances, sa motivation et pour m'encourager dans de nouveaux projets. Merci également au Prof. Sylvain Benoit pour les discussions rassurantes et les conseils bienveillants. Je remercie aussi tout particulièrement le Prof. Jean-David Fermanian, le Prof. Gaëlle Le Fol, le Prof. Yannick Lucotte et le Prof. Christian-Yann Robert, pour leurs conseils, échanges sur la recherche et l'encouragement.

Ensuite, je profite de cette occasion pour exprimer ma gratitude à mes amis et collègues de recherche. Chaleureuses pensées à : Alexis, Anmar, Béatrice, Charles, Dylan, Éliisa, Florian, Hajare, Hugo, Inessa, Jalel, Jean-Charles, José, Julien, Olessia, Pauline, Sébastien, Wassim, Yunzie, pour les moments passés ensemble, les pauses cafés et repas à discuter ensemble de recherche et d'enseignement mais également de notre quotidien. Un merci tout particulier à Laura, une collègue qui est devenue mon amie, merci pour tes paroles toujours positives et ton soutien. Un mot tout spécial pour Sullivan. Un immense merci à mon ami que j'ai découvert lors de mes études de master, de thèse mais qui le restera bien après, j'en suis sûre. Merci pour ta présence constante et ta bonne humeur. Aussi, je souhaite remercier Édith et Fanda pour votre aide dans mes tâches administratives mais aussi pour nos discussions toujours agréables. Enfin, un merci tout particulier à Karine, du fond du cœur, je te remercie pour ton aide, ton soutien, ton temps, ta motivation et tes mots d'encouragement, et plus important encore, pour les

bons moments passés ensemble. Tu as su me guider lors de mes premiers pas en thèse et jusqu'au dernier jour de rédaction de mon manuscrit.

Je voudrais maintenant exprimer mes sincères remerciements à ma sœur Lucile et mes parents Béatrice et Patrick. Merci pour votre amour, votre soutien, votre aide, merci d'être toujours à mes côtés dans mes projets de vie. Merci de toujours me permettre de me ressourcer auprès de vous, vous savez combien c'était important pendant toutes ces années. Merci également à toutes les personnes de ma famille qui m'ont encouragée dans cette voie. Merci à ma meilleure amie, ma Léa. Merci d'être toujours à mon écoute et de me pousser à donner sans cesse le meilleur de moi. Je souhaite terminer ces remerciements par la personne qui a le plus contribué à ma réussite, Jérémy. Mon co-auteur mais surtout depuis peu, mon mari. Merci d'être l'épaule sur laquelle je peux me reposer, mon pilier. Je te remercie du fond du cœur pour ton soutien sans faille et le temps que tu as passé à relire mon manuscrit. Merci d'être ma moitié autant sur la dimension professionnelle que dans notre vie quotidienne. Ces longues années n'auraient pas été les mêmes sans ton amour.

Ophélie Couperier, épouse Leymarie. Lille, Septembre 2022

List of Figures

2.1	Gaussian density, with its truncated and rational approximations based on M parameters.	28
2.2	5-minute squared log-returns for the first 25 sample days for the 10 stocks.	36
2.3	Transformation of the intraday variables (black curves) for the first 25 sample days of all stocks and their corresponding approximations (blue curves).	37
2.4	Joint significance test for the exogenous variable.	40
2.5	Estimated intraday coefficient curve $\hat{\delta}(\cdot)$ of each asset.	42
2.6	Rejection dates of goodness-of-fit for $m = 6$ (significant levels of 5% (light shaded area) and 1% (dark shaded area)).	44
2.7	Evolution of the AS-Log-GARCH-fX model in SSM of the MCS procedure (AS-Log-GARCH-fX model alone (dark shaded area) and with other models (light shaded area)).	46
3.1	Empirical size of the tests at 5% significance level (AR(1)-GARCH(1,1) model)	76
3.2	Empirical power of the tests at 5% significance level (AR(1)-GARCH(1,1) model, bootstrap critical values)	77
3.3	S&P500 daily losses (%), and descriptive statistics	80
3.4	In-sample ES estimates issued from the approximation and the exact calculation method (AR(1)-GARCH(1,1) model)	81
3.5	ES forecasts and adjusted ES forecasts over the period 2007-2009 (AR(1)-GARCH(1,1) model)	86
3.6	Empirical power of the tests at 5% significance level (AR(1)-GARCH(1,1) model, asymptotic critical values)	92
3.7	Empirical size of the tests at 5% significance level (GARCH(1,1) model)	93
3.8	Empirical power of the tests at 5% significance level (GARCH(1,1) model, bootstrap critical values)	94
3.9	Empirical power of the tests at 5% significance level (GARCH(1,1) model, asymptotic critical values)	95

3.10	In-sample ES estimates issued from the approximation and the exact calculation method (GARCH(1,1) model)	96
3.11	ES forecasts and adjusted ES forecasts over the period 2007-2012 (AR(1)-GARCH(1,1) model)	98
3.12	ES forecasts and adjusted ES forecasts over the period 2007-2009 (GARCH(1,1) model)	99
3.13	ES forecasts and adjusted ES forecasts over the period 2007-2012 (GARCH(1,1) model)	100
3.14	ES forecasts and adjusted ES forecasts over the periods 2007-2009 (on the left) and 2007-2012 (on the right) with the two BCBS regulatory risk levels (AR(1)-GARCH(1,1) model, GARCH(1,1) model)	101
4.1	Contours of the expected loss $\mathbb{E}[S_0(\theta, Y, e(v))]$ ($\alpha = 0.20, 0.10, 0.05, 0.01$)	115
4.2	One-day ahead out-of-sample MES forecasts	125
4.3	Average out-of-sample losses ($n = 500$)	126

List of Tables

2.1	Simulation results.	33
2.2	Portmanteau statistics (in %).	34
2.3	In-sample estimation results.	39
2.4	Portmanteau test statistics rejection rates (in %) for four models.	43
2.5	SSM of the MCS procedure.	47
2.6	Ticker symbols and company names	62
3.1	p-values of the backtesting tests (AR(1)-GARCH(1,1) model)	82
3.2	QML coefficient estimates ($p = 6$, AR(1)-GARCH(1,1) model)	83
3.3	p-values of the backtesting tests (GARCH(1,1) model)	97
3.4	QML coefficient estimates ($p = 6$, GARCH(1,1) model)	97
4.1	Simulation results for the normal time-invariant MES and VaR ($T = 500$)	120
4.2	Simulation results for the normal time-invariant MES and VaR ($T = 2500$)	121

Chapter 1

Introduction

Financial risk management is a key process for banks, investment funds, and insurance companies to identify risks, analyze them, and make investment decisions based on either accepting or mitigating them. In the aftermath of the 2007-2009 global economic and financial crisis, the credibility of the financial risk management practices was seriously damaged and the Financial Crisis Inquiry Commission pointed out dramatic failures in the risk management process of banks and regulatory agencies. This context has built a momentum to reshape the internal procedures and risk management practices of financial institutions. In particular, the academic research in financial econometrics has contributed to the introduction of new approaches for measuring and modeling risks and the development of comprehensive methods for backtesting the financial risks.

1.1 Modeling Financial Risks

Since the Basel II accords, formal risk modeling is required for all the major international banking institutions by the various national depository institution regulators. In the past, risk analysis was done qualitatively but now with the advent of powerful computing software, and the advances in communication and data collection (e.g., transactions and quotes for stocks, bonds, currencies, options, and other financial instruments), quantitative risk analysis can be performed quickly and effortlessly. Because risk is fundamentally linked with uncertainty, modeling the financial risks implies a representation of risks in terms of random variables. Typically, a risk measure aims to summarize into a single figure a particular dimension of risk (kurtosis, skewness, tails, variance, etc.).

In a first part, we look at two main classes of risk measures, namely market risk and systemic risk measures. Market risk is the risk of losses in positions arising from fluctuations in market prices. Systemic risk is the risk associated with the collapse or failure of a company, industry, financial institution, or an entire economy. In a second part, we give a presentation of the risk models that are typically used for their estimation.

1.1.1 Financial Risk Measures

The notion of market risk is a crucial consideration in modern risk management. Of great importance, the market risk measures have become industry standards for the assessment of risks and the computation of capital requirements in the banking regulatory framework (see BCBS, 2016). A typical measure of market risk is the volatility that measures how rapidly or severely the price of an investment may change. The volatility is a leading indicator of risk in modern portfolio theory for assembling a portfolio of assets (Markowitz, 1952), and for any risk-based investment strategy that involves the computation of the second-order moment (see Roncalli, 2014). For instance, considering only risk parity investment strategies - based on the volatility measure of risk - the asset under management of related investments is estimated at \$400Bn (Financial Times, 2015).

While volatility does not distinguish between downside and upside risks, several measures are specifically devoted to the measurement of extreme downside risks, that is, the risk of experiencing severe losses. Historically, one convention is to use the value-at-risk (VaR), which is defined as 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 cornerstone of internal risk management systems in financial institutions, as a result of the success of the J.P. Morgan (1996) Risk-Metrics system. In credit portfolios, VaR is also used to compute the so-called unexpected credit loss and to determine credit risk capital requirements with the help of the asymptotic single risk factor model (see BCBS, 2005). Furthermore, the VaR is commonly used in central counterparty clearing houses with the standard portfolio analysis of risk to estimate collateral requirements based on a coverage level of potential losses for an individual contract or portfolio of contracts (Chicago Mercantile Exchange, 2012).

While the convention of using VaR is historically well accepted in the risk management industry, the measure displays several weaknesses. Artzner et al. (1999) define a set of four properties a risk measure should satisfy and call the measures verifying these properties "coherent". In particular, VaR (and volatility) are shown to be not coherent. VaR is also criticized for being not sensitive to capture tail risk during periods of significant financial market stress (BCBS, 2016). Recently, there was a shift in favor of 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. Wang and Zitikis (2021) further refine the coherent approach for VaR providing four economic axioms for portfolio risk assessment which are also satisfied by ES.

The 2007-2009 global financial crisis has fostered extensive research on the measurement of systemic risk. Of particular interest is the identification of financial institutions

that contribute strongest to the overall risk of the financial system - the so-called systemically important financial institutions (SIFIs) or global systemically important banks (G-SIBs). The Financial Stability Board (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 they pose a major threat to the system, regulators and policy makers have called for tighter supervision, extra capital requirements, and liquidity buffers for SIFIs, which are set up through the computation of systemic risk measures.

The ultimate goal of a systemic risk measure is to better identify the vulnerabilities of the financial system (see De Bandt and Hartmann, 2002; Benoit et al., 2017, for a survey). We can identify two main families of systemic risk measures. The first set aggregates confidential bank supervisory data. A typical example is the systemic risk score implemented for SIFIs by the Basel Committee on Banking Supervision (BCBS, 2014). Another example is the bank capital shortfall computed from the 2014 regulatory banking stress tests under the supervision of the European Banking Authority (EBA, 2014). The second set of systemic risk measures relies on market data such as stock returns, option prices, or CDS spreads. The most prominent examples are the delta conditional value-at-risk (ΔCoVaR) of Adrian and Brunnermeier (2016), the marginal expected shortfall (MES) and the systemic expected shortfall (SES) of Acharya et al. (2017), and the SRISK of Acharya et al. (2012) and Brownlees and Engle (2017). The ΔCoVaR corresponds to the difference between the conditional VaR (CoVaR) of the financial system conditional on the firm being in distress and the CoVaR conditional on the firm being in its median state. The MES is the firm's expected equity loss when the market falls below a certain threshold over a given horizon. Thus, banks with highest MES and/or ΔCoVaR are likely to be the greatest drivers of systemic risk. The SRISK measures the expected capital shortfall of an institution conditional on a crisis event occurring. The SRISK takes into account the size of the firm via the market capitalization and the book value of total liabilities. The key advantage of these measures is that they can be computed in real time as their implementation only requires public data (see e.g. the V-Lab website of the NYU).

1.1.2 Financial Risk Models

The first volatility risk model was developed 40 years ago, following the 1982 Rob Engle's publication, that introduced the autoregressive conditional heteroskedasticity (ARCH) model, quickly followed by the generalization to GARCH of Bollerslev (1986). The emergence of GARCH-type models is undoubtedly linked to a number of empirical facts found out in almost all return time series. A good conditional heteroskedasticity model should be able to capture most of these empirical facts. The main stylized feature

is volatility clustering. As raised by Mandelbrot (1963), large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes. A second stylized property is the presence of positive excess kurtosis, known as fat tails, saying that the probability of having extreme events (very high or very low returns) is higher than it would be in the Gaussian case. Several additional empirical features were identified and included in new generation risk models. Examples of these generalizations are the Exponential GARCH (Nelson, 1991), the Threshold GARCH (Zakoian, 1994), and the Glosten-Jagannathan-Runkle GARCH (Glosten et al., 1993), that account for the leverage effect, i.e., negative shocks tend to generate an increase in volatility greater than that induced by positive shocks of the same magnitude. Another interesting model is the log-GARCH representation. Because the log-GARCH specifies the log-volatility instead of the volatility, it obviates the usual positivity constraints on the parameters, and allows end-users to test for the nullity of parameters. This property is particularly appealing when exogenous variables are included in the volatility equation (Sucarrat and Escribano, 2012). The class of log-GARCH models, introduced by Geweke (1986), was studied by Milhoj (1987); Francq et al. (2013); Sucarrat et al. (2016), among others. Finally, we can cite the class of generalized autoregressive score (GAS) models where GARCH is a special case when the measurement density is normal (Creal et al., 2013).

The use of estimators of volatility computed with high-frequency data has greatly improved the ability to model financial market volatility. In particular, realized measures of volatility produce a much more accurate measurement of the current level of daily volatility than do the squared returns (see Andersen et al., 2006, for a discussion). Engle (2002b) looks at this possibility first and gives a generalization of GARCH-type models, called multiplicative error model (MEM), where the dependent variable is no longer the daily return but a non-negative time series process typically built using intra-daily data such as a realized measure of volatility, the volume, or the duration between trades. When the dependent variable is the squared return, the MEM model reduces to the GARCH model, and Engle (2002b) proposes to introduce weakly exogenous variables in the volatility equation – typically the lagged realized volatility – leading to the so-called GARCH-X model. Hansen et al. (2012) introduce a slightly different method to GARCH-X model called Realized GARCH where the realized volatility is driven by its own measurement equation to capture the dependence between the realized measure and the underlying conditional variance. Recently, Francq and Thieu (2019) provide quasi-maximum likelihood estimation (QMLE) for volatility models with covariates, including the GARCH-X model. To investigate whether the realized volatility is a good proxy of the volatility, they introduce the actual realized volatility into the volatility equation, and show that the persistent parameter of the conditional variance is always highly significant, suggesting that realized volatility is useful but not a perfect proxy of the actual volatility.

Another direction of this literature looks at risk models that include data sampled at different frequencies. A key advantage of these models is that daily volatility is modeled through intraday predictors without resorting to daily aggregates such as realized volatility. A typical example is mixed data sampling (MIDAS) regression models where future daily volatility is related to past observations of intraday predictors. Since the inclusion of high-frequency data substantially increases the number of lagged forecasting variables, the model embeds a parametric weight function that links intraday predictors with the daily volatility to bypass the problem of parameter proliferation (see Ghysels et al., 2004, 2006). Within the same goal, Corsi (2009) develops the heterogeneous autoregressive (HAR) model where predictors are defined as lagged realized volatilities aggregated at different time horizons to account for long- and short-memory of volatility processes and to afford a natural economic interpretation. Another avenue of research for modeling high frequency time series data is functional data analysis (FDA). In FDA, a time series is a sequence of observations where each variable is defined as a curve. In the context of high-frequency volatility modeling, a curve represents the whole sequence of intraday returns observed on a given day. This area of research has received increasing interest in the recent time series literature. Hörmann et al. (2013) firstly examine this possibility and proposed a functional version of the ARCH process that allows for the modeling of the intraday curve of volatility. This method was then generalized in Aue et al. (2017) to a functional GARCH. Due to their technical complexity, existing estimators of the underlying functional parameters are moment based which is known to be relatively inefficient in this context. To overcome the lack of efficiency, Cerovecki et al. (2019) propose an alternative approach inspired by the classical GARCH QMLE method.

Because of their time-varying dynamics, the tail risk measures such as VaR and ES are generally expressed conditionally on an information set and the forecasts are commonly issued from a dynamic parametric or semiparametric model. Three main approaches are available for their modeling. The first approach recovers tail risk measures indirectly via models of time-varying first and second moments. Typically, a GARCH model or a MIDAS model can be used to build conditional ES and VaR estimates under distributional assumptions. Numerous existing examples can be evoked. McNeil and Frey (2000) combine an autoregressive model with a GARCH model, for modeling conditional mean and variance, with a generalized Pareto distribution, to produce conditional VaR and ES under the extreme value theory. Francq and Zakoïan (2015) provide several ways for the computation of VaR and ES issued from GARCH-type models estimated by QMLE with corresponding confidence intervals. De Roon and Karehnke (2017) use the halves of two normal distributions for modeling skewed returns' distribution and provide closed-form formulas for ES and VaR which only require estimates of first and second moments.

The second approach pools nonparametric methods including the so-called historical simulation method used in the standard approaches of regulation (see e.g. Dowd, 2001, for the case of VaR). More refined nonparametric methods have been also suggested. Gouriéroux et al. (2000) introduce a nonparametric kernel VaR estimator and derive the corresponding asymptotic distribution. Scaillet (2004) and Scaillet (2005) develop nonparametric kernel methods to estimate ES and conditional ES and provide the asymptotic properties for stationary process satisfying strong mixing conditions. Chen and Tang (2005) and Chen (2007) study the statistical properties of the historical simulation and kernel smoothed estimation method for VaR and ES as introduced by Gouriéroux et al. (2000) and Scaillet (2004), respectively. Cai and Wang (2008) consider a weighted double kernel local linear nonparametric estimation of conditional VaR and ES. More recently, Martins-Filho et al. (2018) propose nonparametric estimation of VaR and ES based on the extreme value theory assuming a generalized Pareto distribution for return tails.

The last modeling approach for tail risk measures is based on the existence of a strictly consistent scoring function which enables M -estimation. In case such a scoring function is existing, it is said that the considered risk measure is elicitable (see Osband, 1985; Lambert et al., 2008). For instance, VaR is elicitable because there exists a natural estimation paradigm in regression frameworks for VaR using quantile regression (see Koenker et al., 2018). Building on that property, Engle and Manganelli (2004) introduce the so-called conditional autoregressive value-at-risk (CAViaR) model defined as a semiparametric autoregressive model for quantiles where the parameters are estimated by quantile regression. White et al. (2015) consider a multivariate extension of CAViaR models for modeling the dynamics of the tail interdependence among different assets. Although it is known that ES is not elicitable on its own (Gneiting, 2011a), recent results indicate that VaR and ES are jointly elicitable implying the existence of a strictly consistent scoring function indexed by the pair ES and VaR (Fissler and Ziegel, 2016). Exploiting these results, Patton et al. (2019) introduce new dynamic models for ES and VaR, drawing on the GAS framework of Creal et al. (2013), and show that the estimated parameters of the VaR and ES models are consistent and asymptotically normal. Taylor (2019) introduces a joint model for conditional VaR and ES, inspired by the CAViaR specification, where parameters are estimated by maximizing the likelihood of an asymmetric Laplace distribution. Interestingly, Taylor (2019) shows that the opposite of his objective criterion is a special case of the scoring function of Fissler and Ziegel (2016).

Risk models are also great interest for systemic risk measurement. As for tail risk measures and volatility, a large variety of systemic risk measures are computed with market data. Consequently, the underlying systemic risk models are designed to capture the same kind of stylized features observed for financial time series, such as volatility clustering, fat tails, or leverage effect. Within this context, Adrian and Brunnermeier (2016) consider

a simple quantile regression model to forecast ΔCoVaR , while Girardi and Ergün (2013) estimate it using a multivariate GARCH model. Brownlees and Engle (2017) implement a bivariate GARCH model with a dynamic conditional correlation structure introduced by Engle (2002a) to estimate the long-run MES and SRISK. Bernardi and Catania (2019) develop a new class of flexible copula models where the dependence parameters evolve according to a Markov switching GAS specification. Eckernkemper (2018) considers a model for MES that captures time-varying nonlinear dependence using time-varying copula parameters endowed within a GAS dynamics. Nolde and Zhang (2020) introduce a bivariate GARCH model based on results in extreme value theory that accounts for asymmetry and heavy tails in financial returns. Hoga (2020) propose a dynamic model for bivariate extremes that allows for smooth transitions between regimes of independence and dependence where a GAS framework is used for modeling multivariate extremes.

1.2 Backtesting Financial Risks

Within the current regulatory framework, banks have the possibility to develop and apply their own internal risk models to estimate the financial risk measures (see BCBS, 2005, and BCBS, 2019, for the credit and market risk policy frameworks, respectively). Because the estimated risk measures are explicitly used in the regulatory capital formulas, any underestimation of these measures may threaten the bank's solvability. Moreover, banks are incited to hold the lowest regulatory capital level because reducing capital releases economic resources that can be used for profitable investments. Thus, the assessment of risk models and corresponding risk measures is crucial for financial stability.

Although the financial risk measures may be conceptually valid and have good properties (see Artzner et al., 1999; Chen et al., 2013; Wang and Zitikis, 2021), the risk models used for estimation might be not well-suited, requiring statistical validation procedures. Jorion (2007) defines backtesting as a formal statistical framework that consists in verifying if actual losses are in line with projected losses. A classification has been proposed to distinguish between *traditional* backtests and *comparative* backtests (Fissler et al., 2016). While traditional backtests are suited for risk models' validation, comparative backtests are related with risk models' selection. In the sequel, we review both types of backtests.

1.2.1 Traditional Backtests

Over the past two decades, the need for sound risk management has triggered a sustained effort of academics to contribute to the assessment of risk models. It has led to a large number of contributions for traditional backtesting. The traditional backtests typically rely on a violation process that satisfies a moment condition when the risk model is valid. When backtesting VaR models, the violation is defined as a dummy variable that takes the value one if the portfolio loss exceeds the VaR, and zero otherwise - hence the

term "violation". Following Christoffersen (1998), VaR forecasts are valid if the violation process satisfies two assumptions, namely the unconditional coverage (UC) hypothesis and the independence (IND) hypothesis. The UC hypothesis indicates that the number of exceedances is observed with the right frequency. The IND hypothesis means that VaR violations observed at two different dates for the same coverage rate must be distributed independently. When the UC and IND hypotheses are simultaneously valid, it is said that the VaR forecasts have a correct conditional coverage (CC) and the VaR violation process is a martingale difference sequence (see Campbell, 2007; Christoffersen, 2010, for more details on UC, IND, and CC hypotheses). Differently, Engle and Manganelli (2004) introduce the so-called dynamic quantile test to assess whether the violations are independent of all past information such as the lagged VaRs and the lagged violations. Because of the binary nature of the violations, Dumitrescu et al. (2012) refine the dynamic quantile test using non-linear dynamic binary choice models which improve the finite sample properties of the backtesting tests. Another streamline of the literature uses duration-based tests accounting for the time interval between two consecutive exceedances that follows a geometric distribution with a success probability equal to the coverage rate of VaR (Berkowitz et al., 2011; Candelon et al., 2011; Christoffersen and Pelletier, 2004; Pelletier and Wei, 2016). Some extensions of the traditional backtests have been proposed to assess the validity of VaR models at several coverage rates. For instance, Colletaz et al. (2013) introduce a backtest for 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 substantially below the VaR. To test for the IND hypothesis at several coverage rates, Hurlin and Tokpavi (2006) consider a multivariate portmanteau statistic.

The financial crisis of 2007-2008 and its aftermath led to numerous changes in the financial market regulation. Of great interest, a new emphasis is placed on ES as a measure of risk, complementing the well-known VaR measure. This update has fostered an extensive research on the development of backtests for ES risk 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 that are based on Monte Carlo simulations. 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. Nolde and Ziegel (2017) introduce the so-called conditional calibration tests for backtesting ES. Du and Escanciano (2017) introduce a cumulative violation process for ES defined as the integral of the VaR violation up to the coverage rate of ES which opens up the possibility to devise UC and IND traditional backtests for 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, as an implicit backtest of ES. Costanzino and Curran (2018) provide a traffic

light backtest for ES that extends the so-called traffic light backtest for VaR. Bayer and Dimitriadis (2020) provide a regression-based backtest exploiting the joint elicibility of the pair VaR-ES. 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 traditional backtesting methods for VaR and ES, see the survey of Argyropoulos and Panopoulou (2019).

Until recently, there were no formal backtesting tests for the systemic risk measures. Banulescu-Radu et al. (2021) develop the first formal statistical procedures for assessing the validity of systemic risk measures such as the MES, SES, SRISK, or ΔCoVaR . The tests are built-up in analogy with the ES backtests of Du and Escanciano (2017) including separate tests for the UC and IND hypotheses. In parallel, a large body of academic research surveys the empirical predictive content of the systemic risk measures. Idier et al. (2014) focus on firms with high systemic risk scores and their corresponding likelihood to experience the highest financial losses in a financial crisis. Engle et al. (2015) compare the list of SIFIs produced by the FSB with the SRISK ranking of European financial institutions. Löffler and Raupach (2018) identify nonexotic cases in which a change in a bank's systematic risk, idiosyncratic risk, size, or contagiousness, increases the risk of the system but lowers the measured systemic risk measures of the bank. Considering a panel of Taiwan financial institutions, Lin et al. (2018) show that the industrial production index is Granger-caused by SRISK. They also show that SRISK, MES, and CoVaR, are quite similar in identifying SIFIs. Brownlees et al. (2020) propose a historical assessment of the SRISK and ΔCoVaR based on two dimensions, namely the SIFI ranking challenge and the financial crisis prediction challenge. Using a dataset spanning the 60 years before the introduction of deposit insurance, they conclude that CoVaR and SRISK are helpful in alerting regulators of systemically risky financial institutions. More recently, Duarte and Eisenbach (2021) devise an index of aggregate vulnerability that tracks over time the factors that make the financial system vulnerable to fire sales. They show that the index strongly correlates with measures of systemic risk, including SRISK and ΔCoVaR .

Another kind of systemic risk measures deserving backtesting are the proprietary-based methods embedded in the toolbox of regulatory agencies. A typical example is the systemic risk score implemented by the BCBS and the FSB for SIFIs (BCBS, 2013, 2014). Another example is the bank capital shortfall computed from the 2014 regulatory banking stress tests under the supervision of the EBA (EBA, 2014) where the main goal is assessing the resilience of financial institutions to adverse market scenarios in the European Union. As these measures rely on proprietary data, e.g., balance sheet, cross-positions, size, leverage, liquidity, and interconnectedness, the data are difficult to access, and few research works have been made for their empirical assessment. Philippon et al. (2017) provide the first evaluation of the quality of the banking stress tests. They find that

the estimated exposures issued from the stress tests predict relatively well the realized losses of banks in subsequent years. Using proprietary-based data, Benoit et al. (2019) reproduce the results associated with the systemic risk scoring methodology currently used to track and regulate the SIFIs. They identify a number of shortcomings associated with the current Basel approach and propose several methodological adjustments.

1.2.2 Comparative Backtests

In point forecasting, predictive methods are compared and assessed by means of an error measure or loss function, with the absolute error and the squared error being key examples. Following Gneiting (2011a), the forecasting models are evaluated through the empirical mean of the individual losses computed on a test sample, which results in a summary measure for the comparison of the predictive performance among models, such as the mean absolute error or the mean squared error. Whatever the criterion used, the models are compared and ranked according to the value of the empirical loss function. Because the loss function is generally negatively oriented, a model m is preferred to a model m' as soon as the empirical loss of model m is lower than that of model m' . In risk management, the evaluation of the predictive performance among risk models, or comparative backtests, is no exception to the rule, and is typically based on loss functions, i.e. risk models' selection is issued by the ranking implied by a statistical loss function.

In many applied fields, the preference of forecasters goes to the prediction of the mathematical expectation. See for instance the widely cited "Survey of Professional Forecasters" of the Federal Reserve Bank of Philadelphia where experts are asked for their future expectations of the value of a variety of economic variables. In risk management, the mathematical expectation is not a key functional because the notion of risk generally excludes the central tendency of variables which would then quantify the expected performance of a given financial position and not the risk attached with the position. For that reason, it is common to use different functionals when measuring risks such as the volatility, the value-at-risk, the expected shortfall, the systemic expected shortfall, etc. Because the functional is not the mean, the employed loss function is not necessarily the absolute error or the squared error. If the loss function is not suitable with the functional, the risk models' ranking is non robust, and it will lead to select a sub-optimal risk model.

Volatility is unobservable, even ex post. As a result, the evaluation of volatility models, and the computation of forecasts errors, is performed by means of some ex-post estimators of the volatility (see e.g. Andersen et al., 2006). Formally, the predicted volatility issued by a given risk model is compared with a proxy of the true latent volatility. This problem can be overcome, at least partly, if an unbiased estimator of the latent variable is available. Several conditionally unbiased estimators have been proposed as proxies of the volatility, such as the squared innovations, the realized volatility, and kernels

(see Andersen and Bollerslev 1998; Barndorff-Nielsen and Shephard 2002; Zhang et al. 2005; Zhou 1996; Barndorff-Nielsen et al. 2008, among others, and further developments as in Andersen et al. 2003; Barndorff-Nielsen et al. 2008; Hansen and Lunde 2006b), or considering range-based variance estimators (Parkinson, 1980; Garman and Klass, 1980; Brandt and Diebold, 2006). However, a conditionally unbiased proxy will not necessarily lead to the same ordering between volatility models that would be obtained if the true volatility was observed. Hansen and Lunde (2006a) show that when the evaluation is based on a target affected by errors, the choice of the evaluation criterion is critical to avoid a misleading ranking of volatility models. The problem of consistency of the ordering between two or more volatility forecasts has been further refined in Patton (2011) and Laurent et al. (2013) for univariate and multivariate volatility models, respectively.

Elicitability is a statistical property, that is satisfied by some risk measures, and is useful for ranking and comparing the performance of risk models. The terminology is due to Osband (1985). If a risk measure is elicitable, then there exists a loss function for that risk measure that can be used for comparative backtests on risk models. The risk measure VaR is elicitable implying that a consistent class of loss function is available for comparative VaR backtests. The class of loss functions that is consistent with VaR is known as the generalized piecewise linear class (see Gneiting, 2011b), which nests the well-known piecewise linear function from quantile regression (see e.g. Koenker et al., 2018). For instance using the piecewise linear loss function, Bernardi and Catania (2016) compare the VaR forecasts delivered by alternative model specifications using the model confidence set procedure of Hansen et al. (2011). In contrast to VaR, ES is known to be not elicitable on its own (see Gneiting, 2011a). However, recent results from statistical decision theory show that ES is jointly elicitable with VaR (Fissler and Ziegel, 2016). Thus, there exists a consistent class of loss function with the pair ES-VaR where the elements ES and VaR are not separable into the loss function. This finding opens up the possibility to compare and rank ES and VaR forecasts issued from competing risk models. In this context, the academic research provides comparative backtesting procedures for ES. Ziegel et al. (2020) develop graphical checks of whether one forecast method dominates another under a relevant class of loss function, and propose an associated hypothesis test. Dimitriadis and Schnaitmann (2021) and Dimitriadis et al. (2021) introduce forecast encompassing tests for ES to test whether the forecast of ES from a model m performs not worse than any linear combination of forecasts from several competing risk models.

The systemic risk literature devoted to risk models' selection is surprisingly underdeveloped. As raised by Fissler and Hoga (2021), the systemic risk measures such as the CoVaR, the conditional expected shortfall, or the MES, fail to be elicitable, which is a strong limitation for providing comparative backtests. To bypass the lack of elicibility of the systemic risk measures, Fissler and Hoga (2021) develop a notion of multi-

objective elicibility that enable forecast comparison for systemic risk measures. They devise Diebold-Mariano type tests (Diebold and Mariano, 1995) utilizing two-dimensional scores for the CoVaR, the CoES, and the MES. Within a different context, the traditional backtesting procedure of MES proposed by Banulescu-Radu et al. (2021) overrides the lack of elicibility of MES exploiting the fact that MES can be expressed as an integral of conditional VaRs which are themselves elicitable. Because the backtest relies on a moment condition of a violation process, the procedure can be theoretically adapted to compare the forecasts issued by different systemic risk models through the computation of the empirical mean of the squared centered violations. Brownlees et al. (2018) evaluate the accuracy of tail risk forecasts for systemic risk measurement. They propose two loss functions, namely the tail tick loss and the tail mean square error, to evaluate, respectively, CoVaR and MES forecasts. Because MES and CoVaR can be respectively defined as a truncated expectation and a truncated quantile of a bivariate distribution, they apply the well-known piecewise linear and squared loss function to the firm return given that the market return exceeds the market VaR.

1.3 Contribution

In this renewed context, the scope of our research stands at the juncture between the modeling and the statistical validation of the financial risk measures which are importantly used by academics, regulators, and practitioners. The general aim of this dissertation is to develop and propose advanced methodologies for modeling the financial risk measures and assessing their validity. Our research looks after several risk measures which can be used to quantify a large scope of financial risks, namely, (i) the market risk, (ii) the tail risk, and (iii) the systemic risk. For each of these risk classes, the main goal of our research is to improve the soundness of the banking industry through advanced econometric and statistical tools that quantify the risks held in a market portfolio, or more generally the risks held in banks' balance sheet. For the systematic risk, our dissertation contributes to the development of advanced predictive models that may generate more accurate volatility measures compared to the current models used by practitioners and academics. For the market risk, the goal of our work is to enhance the measurement of tail risk in market portfolios and to promote stronger backtesting practices in the asset management industry against the adverse shocks and potential losses experienced by the investment companies. Finally, our work 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.

In Chapter 2 and Chapter 3, we focus our research on the volatility and the expected shortfall which are both used to quantify the risk in investment portfolios. Chapter 2

develops a pioneering conditional volatility model equipped with a functional variable in the volatility equation. Contrary to the well-known GARCH-X model that includes the realized volatility as a summary of the intra-day price variations, our functional approach enables the inclusion of the whole information contained in the intra-day price movements. We show numerically that the proposed volatility model may capture new sources of information in the intra-day price variations compared to the mainstream volatility models that simply include the realized volatility. Chapter 3 develops backtesting tests for the expected shortfall as the new standard used by banks and their regulators to quantify the tail risk in market portfolios. Compared to the method currently used by practitioners and academics, our backtesting framework greatly simplifies the assessment of ES, by focusing on the validity of several VaRs, which is more intelligible in the context of banking regulation. Our tests 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. Chapter 4 looks after the systemic risk measures that are used to summarize the systemic risk contribution of a given financial institution into a single figure or to rank the financial institutions according to their systemic importance as, e.g., displayed on the Volatility Lab website of the NYU. We focus our study on the systemic risk measures that are expressed as functions of the expected equity loss conditional on a financial crisis, as it is the case for the MES, the SES, and the SRISK, for instance. We establish the elicibility of those indicators leading to the identification of a general class of consistent loss function that can be either used for estimation or to compare and rank the risk models typically used to track in real-time the systemic risk of financial institutions. To the best of our knowledge, it is the first loss function that can be consistently used to compare and rank the forecasts of the systemic risk measures. In the following, we synthesize the contents of each chapter.

Chapter 2: Daily Volatility Forecasting using Intraday Returns and Functional Covariates

Chapter 2, "Daily Volatility Forecasting using Intraday Returns and Functional Covariates" proposes a new volatility model, namely the AS-Log-GARCH(p,q)-pfX model, that explicitly accounts for the influence of intra-day price variations.¹ As raised by Engle and Patton (2001): "*no-one believes that financial asset prices evolve independently of the market around them, and so we expect that other variables may contain relevant information for the volatility of a series.*" In this chapter, we move into this direction and we develop a novel GARCH-type model that incorporates a covariate into the volatility equation itself. Contrary to the widely used GARCH-X model that includes a simple scalar variable into the volatility equation, the AS-Log-GARCH(p,q)-pfX model is

¹This chapter is based on Couperier, Francq, and Zakoian (2021).

equipped with a functional variable into the volatility equation allowing the inclusion of the whole sequence of intra-day returns, and not a simple summary as considered in the GARCH-X model. Furthermore, the AS-Log-GARCH(p,q)-pfX model is explored under a log-volatility representation to avoid positivity constraints on the parameters and to simplify the statistical inference compared to that of the standard GARCH formulations.

Several important theoretical contributions are provided. First, we show the existence of stationarity solutions, moments, and log-moments, for the considered model. Second, we propose a quasi-maximum likelihood estimation (QMLE) procedure for parameter estimation. As a result of QMLE, there is no need to specify the distribution followed by the innovation process of the financial returns, hence the proposed model can be seen as semi-parametric. Third, we derive the asymptotic properties of the QMLE estimates and show that the parameters are asymptotically normally distributed. Finally, we devise a Portmanteau goodness of fit test based on the autocovariances of the squared residuals to assess the validity of the time-varying dynamics of the AS-Log-GARCH(p,q)-pfX model.

Our asymptotic results are illustrated in a simulation study. The parameters of the model are calibrated using the daily returns of Bank of America from December 2000 to December 2008 and we consider the corresponding 5-minutes squared log-returns of Bank of America for the functional variable. Note that we consider a set of seven normalized Legendre polynomials for approximating the functional variable and that the statistical inference is performed using 2000 observations, i.e. about eight years of historical data. First, we apply tests of significance to both the usual parameters in the GARCH equation and the parameter of the functional variable. At 1%, 5%, and 10% nominal levels, our tests of significance provide satisfactory size performance to both the common GARCH parameters and the parameter of the functional variable, hence illustrating the validity of our asymptotic theory. Second, we apply our Portmanteau test on additional volatility models to judge the advantages of the AS-Log-GARCH(p,q)-pfX model against a set of alternative models. We consider three competing models that are nested in the AS-Log-GARCH(p,q)-pfX model, namely, the asymmetric log-GARCH model including the realized volatility as the exogeneous variable, the asymmetric log-GARCH model without exogeneous variable, and the log-GARCH model without exogeneous variable and without asymmetry. Our simulation results are clear-cut. The validity of the three competing models are rejected by the Portmanteau test while it is not for the AS-Log-GARCH(p,q)-pfX model providing evidences of the superiority of the AS-Log-GARCH(p,q)-pfX model to properly account for the intra-day time variation contained in financial market returns.

We apply our model to the daily financial returns of 10 stocks, namely Abbott, American Express, Boeing, Bank of America, Bristol Myers, BP, Citigroup, Caterpillar, Colgate-Palmolive, Cisco Systems. Our data spans from January 1999 to December 2008. We consider the 5-minutes log-returns from 9:35 AM to 4:30 PM for modeling the

functional variable. Similarly to the simulation study, we consider a set of seven normalized Legendre polynomials to properly approximate the functional variable. We fit our model on the database and test the global significance of our exogenous variable. Then, for comparison purposes, we consider three competing models (including either, the realized volatility and a leverage effect, a leverage effect alone, or neither of them) that can be seen as simpler representations of our model. We applied our Portmanteau tests to test the adequacy of the competing models. Then, we apply the model confidence set procedure to select the best forecasting volatility models. Our findings are summarized as follows. First, for seven out of ten assets, we find that our model gives satisfactory results since we conclude to non-autocorrelated residuals in more than 80% of cases. Second, we find that our exogenous variable is useful to model the squared daily log-returns for six out of the ten assets. Finally, our model belongs to the superior set of models for seven out of ten assets more often than the three competitors.

Chapter 3: Backtesting Expected Shortfall via Multi-Quantile Regression

Chapter 3, "Backtesting Expected Shortfall via Multi-Quantile Regression", proposes four easy-to-use tests for assessing the validity of the expected shortfall (ES) based on quantile regression models.² Among the number of fundamental reforms that must be implemented until January 1st, 2022 (BCBS, 2019), the BCBS has substituted value-at-risk (VaR) by ES for the calculation of market risk capital requirements. 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 as the measure satisfies the properties of monotonicity, sub-additivity, homogeneity, and translational invariance (Artzner et al., 1999; Acerbi and Tasche, 2002). Furthermore, ES quantifies the expected size of the loss given that a loss bigger than VaR is experienced, while VaR only captures the likelihood of an incurred loss. In its revised standards for market risk, the BCBS emphasizes the role of ES to avoid the "*perverse incentives to hold positions that featured significant tail risks but were subject to limited risk in "normal" conditions*" (BCBS, 2019, page 3).

Our research aims at providing an original backtesting framework for the ES risk estimates which are computed for the market portfolios in the industry. As defined by Jorion (2007), backtesting is a formal statistical framework that consists of verifying whether actual losses are in line with projected losses. In other words, backtesting implies a systematic comparison of the historical model-generated risk measure forecasts with actual losses. Because the ES parameter is a key constituent for the computation of the regulatory capital charge as part of the market risk, any underestimation of ES that

²This chapter is based on Couperier and Leymarie (2021) and has received the 2019 best paper award of the German Finance Association (DGF).

has not been identified in time may threaten the bank's solvability. For that reason, backtesting is a key requirement for ES to become an industry standard in the long run.

This chapter suggests a natural extension of the standard VaR backtesting procedures enabling the validation of VaR risk estimates at several probability levels jointly. Our approach is as follows. We suggest an approximation of ES defined as a sum of several VaRs at pre-specified probability levels along the tail distribution of the risk model. Our approximation stems from the representation of ES as the limit of a Riemann sum involving a set of several VaRs. Then, we jointly assess the VaRs involved in the Riemann sum as an implicit backtest of ES. Formally, we consider a multi-quantile regression model and show that the model parameters have specific properties under the hypothesis that the VaRs are jointly valid. We propose four backtests which correspond to various linear restrictions on these parameters. These restrictions are implications of a Mincer-Zarnowitz representation (Mincer and Zarnowitz, 1969). Then, we test the resulting parameter restrictions using Wald-type inference. The procedure extends the test of Gaglianone et al. (2011) that allows the backtesting of VaR at a single probability level.

Our approach has many advantages compared to the method currently used by practitioners, regulators, and academics. First, our procedure is flexible since the user may choose the number and values of quantiles for the assessment of ES and can easily focus on various aspects of the tail distribution of the risk model. Second, the method encompasses the recommendation of the Basel III accord of verifying quantiles at risk levels 97.5%, and 99%. 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). Furthermore, the procedure is easy to implement as it is based on the well-established VaR which may complete the toolbox commonly used by risk managers. Finally, our procedure complements the existing literature on regression-based backtests as proposed by Engle and Manganelli (2004, JBES) and Gaglianone et al. (2011, JBES).

A simulation study is considered to assess the finite sample properties of our backtests. To assess the empirical size, we simulate portfolio returns from a AR(1)-GARCH(1,1) model with Student innovations which is defined as the correct data generating process throughout the simulation study. This class of model is widely used for capturing variance dynamics in daily asset returns (see e.g., Berkowitz and O'Brien, 2002; Berkowitz et al., 2011; Du and Escanciano, 2017; Loser et al., 2019). Our simulation results indicate that the use of asymptotic critical values produce substantial size distortions for the four backtests. In order to correct the finite sample size distortions of our backtests, we propose a pairs bootstrap algorithm (Freedman, 1981). The resulting bootstrap critical values provides satisfactory size performances regardless of the sample size and should accordingly be used when asymptotic theory does not apply conveniently. Finally to

investigate the power, we consider several misspecified alternative models considering mean, variance, tail, and dynamic measurement errors. Our tests correctly flag all the misspecified alternatives and we verify that there is a general improvement of powers as the sample size increases, suggesting that these tests are consistent for these alternatives.

An empirical application is also conducted to illustrate the merits of our method. We assess the ES estimates issued from a 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. Our procedure concludes that the forecasts of ES are generally misleading. More interestingly, our empirical results suggest an update of the current regulatory guidelines. We show that the BCBS recommendation of assessing quantiles at risk levels 97.5% and 99% is not necessarily a reasonable guideline to identify misspecified ES models. The use of additional quantiles improve the soundness of the decision. Our results also suggest to limit the number of quantiles in very small samples and to consider higher values if the historical sample covers longer periods. 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.

Chapter 4: Elicitability of Marginal Expected Shortfall and Related Systemic-Risk Measures

Chapter 4, "Elicitability of Marginal Expected Shortfall and Related Systemic-Risk Measures", identifies the first class of strictly consistent scoring function for the market-based systemic risk measures.³ While the debate on the economic usefulness for regulators of market-based systemic risk measures is still at play (Idier, Lame, and Mesonnier, 2014; Zhang et al., 2015; Benoit et al., 2017, Loffler and Raupach, 2018; and Brownlees et al., 2020), answering the remaining questions on estimation, comparison, and evaluation of those indicators should help to finally put an end to all equivocation. If these timely measures are estimated accurately, they react quickly to market changes, offering added value compared to the current yearly systemic-risk scores used by the Financial Stability Board to identify Global Systemically Important Banks (G-SIBs). For instance, the 2020 list of G-SIBs identify the same 30 G-SIBs than the year before even if we are experiencing the COVID-19 pandemic crisis.

Elicitability is a mathematical property, that is satisfied by some risk measures. The terminology is due to Osband (1985). If a risk measure is elicitable, then there exists a scoring function (or loss function) for that risk measure that can be used for estimation or for comparative tests on risk models. Without such a loss function, consistent ranking among methods is not feasible. Importantly, value-at-risk (VaR) is elicitable, while ex-

³This chapter is based on Benoit, Couperier, Leymarie and Scaillet (2021) and has been awarded a research grant sponsored by the Fondation Banque de France (2021-2022).

pected shortfall (ES) is not. According to the internal models approach adopted by most major international banks, the market risk forecasts are issued from internal risk models (BCBS, 2019). Researchers have long debated over the connection between elicibility and backtestability, fuelled by the adoption of ES for the calculation of market risk capital requirements. Fortunately, recent results from statistical decision theory have overcome the lack of elicibility of ES (Fissler and Ziegel, 2016), which finally opens the possibility to compare and rank ES forecasts issued from competing and internal risk models (Patton, Ziegel, and Chen, 2019; Ziegel et al., 2019). So far, however, elicibility has not been established for the market-based systemic risk measures, limiting their inclusion as banking regulation standards, as it is the case for ES and VaR, for instance.

In this chapter, we identify a general class of loss function that is strictly consistent with the market-based systemic risk measures. Our methodological development is as follows. First, we build a class of loss function for the marginal expected shortfall (MES) of a financial institution and for the VaR of the financial market, jointly, establishing the 2-elicibility of the bivariate functional (MES, VaR). Then, we generalize the class of loss function for any systemic-risk measures that can be expressed as a function of MES, as it is the case for SRISK and SES. Our loss function is built up in analogy with the recent advances in statistical theory where a class of loss function for ES is identified (Fissler and Ziegel, 2016). Indeed, ES and MES display many similarities, as they can be both defined as truncated expectations, of a univariate, and of a multivariate distribution, respectively. Furthermore, the MES is the partial derivative of the system ES with respect to the weight of firm in the economy (Scaillet, 2004, Acharya et al., 2017).

Several contributions are provided to the systemic risk literature. First, we apply the loss function as a procedure to estimate the systemic risk measures. Formally, we derive a semi-parametric M -estimator of the MES and VaR defined as the solution to the minimization of the empirical loss function. While the existing estimation methods used for the systemic risk measures, like MES or SRISK, are issued from fully parametric models or models of second-order moment (see Acharya, Engle and Richardson, 2012; Acharya et al., 2017; and Brownlees and Engle, 2017), our framework imposes minimal distributional assumptions on the joint distribution of asset returns. To our knowledge, it is the first time that modeling the MES or SRISK onto a set of explanatory variables is achieved, as in Adrian and Brunnermeier (2016) where ΔCoVaR is explained by macro-state variables. Second, we consider the loss function for models' performance comparison. The loss function can be applied to compare and rank different competing risk models for the systemic risk measures which was not yet available for those indicators. For instance, we can gauge the correctness of the systemic-risk rankings as displayed on the Volatility Lab website of the NYU, or to show that an ill-chosen risk model may skew the evaluation of risk and regulatory capital levels held by banks, as defined by a systemic risk measure.

Our estimation procedure imposes the estimation of a nuisance function into the empirical loss function which is defined as the expected firm return conditional to the market return being equal to its VaR. To ensure the strict consistency of the loss function, the nuisance function is imposed to be strictly increasing with the market-VaR. We use the nonparametric isotone estimator of Dette, Neumeyer, and Pilz (2006) to ensure the strict monotonicity of the nuisance function. Then, we present conditions such that the estimated parameters of the VaR and MES models are consistent in the presence of a nonparametric estimator of a nuisance function in the criterion function (Chen, Linton, and Van Keilegom, 2003; Delsol and Van Keilegom, 2020). Finally, we characterize the asymptotic normality and we show that the nuisance function does affect the asymptotic variance of the estimator (Andrews, 1994; Newey, 1994; Ichimura and Lee, 2010).

We consider a set of Monte Carlo experiments to illustrate the asymptotic theory of our semi-parametric M -estimator for MES and VaR. We consider several sample sizes, namely 250, 500, and 1000 daily observations, representing 1, 2, and 4 years of historical data, respectively. The MES and VaR are computed at several probability levels, namely 0.01, 0.025, 0.05, 0.10, 0.20, that correspond to the largest bank's expected equity loss out of 100, 40, 20, 10, 5, trading days, respectively. First, we observe that the parameters are estimated without bias hence illustrating the consistency of the proposed semi-parametric M -estimator. Second, we confirm numerically that the estimated measures are normally distributed for large sample sizes and we report accurate statistical inference. Finally, we find that the asymptotic variance of MES and VaR is strongly decreasing with the considered probability levels. There is much more estimation uncertainty in the MES and VaR during the most severe market decline (probability level close to 0). Thus, in case of adverse market conditions, systemic risk measures are strongly affected by estimation errors, which in turn may skew the evaluation of risk and the required regulatory capital levels. As a result, the estimation risk should be accordingly taken into account by academics and regulators when evaluating the systemic importance of financial institutions.

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

Chapter 2

Daily Volatility Forecasting using Intraday Returns and Functional Covariates¹

In addition to past daily returns, intraday returns are now well-known to carry valuable information on the daily volatility dynamics. GARCH models with exogenous scalar covariates (like realized volatilities) partially take into account this additional information. In this chapter we propose a volatility model including functional covariates, for handling the whole information conveyed by the low and high-frequency returns. We start by giving general stochastic results (strict stationarity, existence of moments and log-moments). Then we study estimation of the model by Quasi-Maximum Likelihood (QML) and we propose a portmanteau test of goodness-of-fit. Monte Carlo simulations and an empirical application on financial series illustrate the interest of including functional covariates for volatility prediction.

2.1 Introduction

As noted by Engle and Patton (2001): *"no-one believes that financial asset prices evolve independently of the market around them, and so we expect that other variables may contain relevant information for the volatility of a series."* Indeed, in addition to past daily returns, intraday returns are now well-known to carry valuable information on the daily volatility dynamics (see *e.g.* Koopman et al. (2005), Chortareas et al. (2011), Lyócsa et al. (2021) and the references therein). The introduction of scalar exogenous variables, or covariates, in the volatility has given rise to the class of GARCH-X-type processes. The question of which covariates to introduce in the volatility in order to increase the predictive power of the model is of course crucial. As in *e.g.* Fuertes et al.

¹This chapter is based on Couperier, Francq and Zakoïan (2022).

(2009), it seems natural to use *realized volatility measures*, such as realized variance, range, power variation and bipower variation.

In these approaches, the information set does not reduce to the past returns, as in classical GARCH models, but includes covariates and, in particular, a summary of the intraday movements. However, using summaries may entail losses of information for volatility modeling and predicting purposes.

Our objective in this article is to propose a volatility structure including the whole sequence of intraday returns as *functional covariates*. In functional data analysis, observed data are transformed into curves. This area of research has received increasing interest in the recent time series literature. In particular, functional version of GARCH models were studied by Hörmann et al. (2013), Aue et al. (2017), Cerovecki et al. (2019), Kühnert (2020), Sun and Yu (2020). Rice et al. (2020) showed the interest of such models for Value-at-Risk forecasting. For the models studied in the aforementioned references, the response is functional (the curve of the intraday squared returns) and the explanatory variables are lagged values of this random curve. In the model we propose, the response is scalar (daily volatility) and the explanatory variables are mixed scalar and functional.

To be more specific, recall that a functional linear regression model with scalar response y and centered functional covariate $\{X(u), u \in [0, 1]\}$ is a direct extension of the multiple linear regression model, and can be written

$$y = \beta_0 + \int_0^1 \beta(u)X(u)du + \varepsilon$$

with obvious notations. If the response y is positive, as is the case for a volatility or a squared return, it is not obvious to impose tractable conditions on the functional parameter $\beta(\cdot)$ so that the right-hand side of the functional linear regression model be positive with probability one. The log-GARCH specification is particularly relevant in this framework, since no positivity constraint is required. This class, originally introduced by Geweke (1986), was studied by Pantula (1986), Milhoj (1987), Sucarrat et al. (2016), Francq et al. (2013), Francq et al. (2018), among others.

This leads us to study the model

$$\begin{cases} \epsilon_t &= \sigma_t \eta_t \\ \ln \sigma_t^2 &= \omega_t + \sum_{i=1}^q \alpha_{i,t} \ln \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \ln \sigma_{t-j}^2 + \boldsymbol{\pi}' \mathbf{x}_{t-1} \\ &+ \int_0^1 \delta(u)X_{t-1}(u)du \end{cases} \quad (2.1)$$

where (η_t) is a sequence of independent and identically distributed (iid) variables such that $E\eta_1^2 = 1$, $\omega_t = \omega + \sum_{i=1}^q \omega_i \mathbf{1}_{\{\epsilon_{t-i} < 0\}}$, $\alpha_{i,t} = \alpha_{i+} \mathbf{1}_{\{\epsilon_{t-i} > 0\}} + \alpha_{i-} \mathbf{1}_{\{\epsilon_{t-i} < 0\}}$, $\mathbf{x}_t = (x_{1,t}, x_{2,t}, \dots, x_{r,t})'$ is a vector of r exogenous scalar covariates, X_t is an exoge-

nous real-valued functional covariate, $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_r)'$ is a vector of real numbers, and δ is a real-valued function.

In this model, ω and the components of the vectors $\boldsymbol{\omega}_- = (\omega_{1-}, \dots, \omega_{q-})'$, $\boldsymbol{\alpha}_+ = (\alpha_{1+}, \dots, \alpha_{q+})'$, $\boldsymbol{\alpha}_- = (\alpha_{1-}, \dots, \alpha_{q-})'$, and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ are real coefficients, which are not *a priori* subject to positivity constraints.² Moreover, noting that $\alpha_{i,t} = 0$ whence $\epsilon_{t-i} = 0$, it will not be restrictive to assume $\sigma_t > 0$. In this model, the covariates involved in the volatility dynamics are both scalar and functional. A recent reference dealing with the introduction of both scalar and functional covariates in time series regression models is Chen et al. (2020). In the applications we have in mind, $X_t(\cdot)$ could be an intraday functional process. We may define $X_t(u)$ as the price of the asset on day t at intraday time u , where $u \in [0, 1]$. The intraday time is represented by the interval $[0, 1]$. Asymmetries typically encountered in financial series are introduced through coefficients depending on the signs of the past returns. Asymmetries are present in both $\alpha_{i,t}$ and ω_t to ensure stability by scaling,³ a most desirable property (see Francq et al., 2018).

The remainder of the paper is organized as follows. In Section 2.2, we discuss the existence of stationary solutions and the existence of moments and log-moments. In Section 2.3, we introduce our estimation procedure and detail its asymptotic properties. In order to test the goodness-of-fit of our model, we also introduce a portmanteau statistic. In Section 2.4, we conduct Monte Carlo experiments to illustrate our asymptotics results. Section 2.5 is dedicated to the empirical applications. Finally, Section 2.6 concludes. Proofs and technical results are reported in an appendix.

2.2 Stationarity and existence of moments

Before investigating stationarity properties, notice that a convention is required to handle returns equal to zero. Under the convention that $0 \times \ln(0) = 0$, the log-volatility in (2.1) is well-defined whatever the values of the past returns (contrary to the standard log-GARCH model which is not appropriate for series that may contain zeroes). We start by considering particular cases of the general model (2.1) for which an explicit stationarity condition can be established.

²The analysis and interpretation of the impacts of shocks on the volatility may however lead to some plausible restrictions on some of these coefficients. For instance if $\omega_{1-} > 0$, a decrease of the price at time $t - 1$ will increase the volatility by a fixed scaling factor whatever the amplitude of the price drop.

³If (ϵ_t) is an element of this class of models the process (ϵ_t^*) defined by $\epsilon_t^* = c\epsilon_t$ with $c > 0$ also belongs to this class.

2.2.1 Stationarity of the first-order model

When $p = q = 1$, the volatility in Model (2.1) writes, omitting subscripts,

$$\ln \sigma_t^2 = \omega_t + \alpha_t \ln \epsilon_{t-1}^2 + \beta \ln \sigma_{t-1}^2 + \boldsymbol{\pi}' \mathbf{x}_{t-1} + \int_0^1 \delta(u) X_{t-1}(u) du, \quad (2.2)$$

where $\omega_t = \omega + \omega_- \mathbf{1}_{\{\epsilon_{t-1} < 0\}}$, $\alpha_t = \alpha_+ \mathbf{1}_{\{\epsilon_{t-1} > 0\}} + \alpha_- \mathbf{1}_{\{\epsilon_{t-1} < 0\}}$.

We assume that the following assumption holds.

A0: $\{\mathbf{w}_t\} = \{(\eta_t, \mathbf{x}'_t, X_t)\}$ is a strictly stationary and ergodic process.

Note that we do not postulate independence between the innovation η_t and the covariates.

Proposition 1 (Existence of a stationary and ergodic solution). *Assume A0 and the following condition holds*

$$E \left(\ln^+ |\mathbf{1}_{\{\eta_0 \neq 0\}} \ln \eta_0^2| + \ln^+ \|\mathbf{x}_0\| + \ln^+ \left| \int_0^1 \delta(u) X_0(u) du \right| \right) < \infty. \quad (2.3)$$

Then, there exists a unique strictly stationary and ergodic solution (ϵ_t) to Model (2.2) if

$$|\alpha_+ + \beta|^p |\alpha_- + \beta|^q |\beta|^r < 1. \quad (2.4)$$

where $p = P(\eta_0 > 0)$, $q = P(\eta_0 < 0)$, and $r = P(\eta_0 = 0)$. Moreover, this solution is nonanticipative, in the sense that $\epsilon_t = h(\mathbf{w}_t, \mathbf{w}_{t-1}, \dots)$, where h is a measurable function.

Notice that, except when $\alpha_+, \alpha_-, \beta > 0$ or when $\alpha_+ = \alpha_- = 0$, the strict stationarity condition (2.4) does not entail $|\beta| < 1$ as is usually the case in GARCH-type models. For the next result we require an additional independence assumption between the error process and the covariates.

Proposition 2 (Existence of log-moments of order m). *Assume A0 and let $m \geq 1$. Suppose that for any $k > 0$ the random vector (\mathbf{x}'_t, X_t) is independent from the variables η_{t+k} , $k > 0$. Let $E |\mathbf{1}_{\{\eta_0 \neq 0\}} \ln \eta_0^2|^m < \infty$, $E \|\mathbf{x}_0\|^m < \infty$, $E \left| \int_0^1 \delta(u) X_0(u) du \right|^m < \infty$ and*

$$p |\alpha_+ + \beta|^m + q |\alpha_- + \beta|^m + r |\beta|^m < 1. \quad (2.5)$$

Then $E |\ln \sigma_t^2|^m < \infty$ and $E |\ln \epsilon_t^2|^m < \infty$.

2.2.2 Strict stationarity of the general model

Because coefficients equal to zero can always be added, it is not restrictive to assume $p > 1$ and $q > 1$ in Model (2.1). Letting

$$\nu_t = \omega_t + \sum_{i=1}^q \alpha_{it} \ln \eta_{t-i}^2 + \boldsymbol{\pi}' \mathbf{x}_{t-1} + \int_0^1 \delta(u) X_{t-1}(u) du,$$

where $\omega_t = \omega + \sum_{i=1}^q \omega_i \mathbf{1}_{\{\eta_{t-i} < 0\}}$ and $\alpha_{it} = \alpha_{i+} \mathbf{1}_{\{\eta_{t-i} > 0\}} + \alpha_{i-} \mathbf{1}_{\{\eta_{t-i} < 0\}}$ we have $\ln \sigma_t^2 = \nu_t + \sum_{j=1}^{p \vee q} (\alpha_{j,t} + \beta_j) \ln \sigma_{t-j}^2$, with $\alpha_{j,t} = 0$ when $j > q$ and $\beta_j = 0$ when $j > p$. We deduce the matrix representation $\mathbf{z}_t = \mathbf{C}_t \mathbf{z}_{t-1} + \mathbf{b}_t$, where $\mathbf{z}_t = (\ln \sigma_t^2, \dots, \ln \sigma_{t-r+1}^2)' \in \mathbb{R}^{p \vee q}$, $\mathbf{b}_t = (\nu_t, 0, \dots, 0)' \in \mathbb{R}^{p \vee q}$ and

$$\mathbf{C}_t = \begin{pmatrix} \alpha_{1,t} + \beta_1 & \alpha_{2,t} + \beta_2 & \dots & \alpha_{p \vee q,t} + \beta_{p \vee q} \\ 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 1 & 0 \end{pmatrix}.$$

Let $\gamma(\mathbf{C})$ be the top Lyapunov exponent of the sequence $\{\mathbf{C}_t, t \in \mathbb{Z}\}$,

$$\gamma(\mathbf{C}) = \lim_{t \rightarrow \infty} a.s. \frac{1}{t} \ln \|\mathbf{C}_t \mathbf{C}_{t-1} \dots \mathbf{C}_1\| = \inf_{t \geq 1} \frac{1}{t} E (\ln \|\mathbf{C}_t \mathbf{C}_{t-1} \dots \mathbf{C}_1\|). \quad (2.6)$$

The choice of the norm is unimportant for the value of the top Lyapunov exponent. However, in the sequel, the matrix norm will be assumed to be multiplicative. Note that this exponent is well defined since $E \ln^+ \|\mathbf{C}_t\| < \infty$. We also have $E \ln^+ \|\mathbf{b}_t\| < \infty$ using the moment condition (2.3). The proof of the following result follows by the arguments used to establish Theorem 2.1 in Francq et al. (2013).⁴ It is therefore omitted.

Theorem 1 (Strict stationarity). *Assume **A0** and the moment condition (2.3). A sufficient condition for the existence of a strictly stationary and nonanticipative solution to model (2.1) is $\gamma(\mathbf{C}) < 0$. The solution is unique, and it is non anticipative and ergodic.*

It is worth noting that the existence of a stationary solution only depends on the coefficients $\alpha_{i+}, \alpha_{i-}, \beta_j$ and the distribution of η_t . Note also that when $p = q = 1$ we retrieve the strict stationarity condition of Proposition 2.

2.2.3 Existence of moments and log-moments for the general model

The next result provides a sufficient condition for the existence of m -th order log-moments.

Proposition 3 (Existence of integer log-moments). *Let $m \geq 1$. Assume **A0** and the moment condition (2.3). Suppose that $\sum_{i=1}^{p \vee q} |\alpha_{i+} + \beta_i| \vee |\alpha_{i-} + \beta_i| \vee |\beta_i| < 1$, $\|\ln \eta_t^2\|_m < \infty$, $\|x_t\|_m < \infty$, and $\left\| \int_0^1 \delta(u) X_t(u) du \right\|_m < \infty$. Then the strictly stationary solution (ϵ_t) satisfies $E |\ln \epsilon_t^2|^m < \infty$.*

In the proof of the proposition we use the notation $\mathbf{C}^{(\infty)}$ for the (non random) matrix obtained by replacing $\alpha_{it} + \beta_i$ by $|\alpha_{i+} + \beta_i| \vee |\alpha_{i-} + \beta_i| \vee |\beta_i|$ in \mathbf{C}_t , for $i = 1, \dots, p \vee q$.

⁴A difference is that, in the present article, the sequence \mathbf{C}_t is not independent. However, the proof does not require this property.

The existence of small-order moments for the volatility or the log-volatility processes is often useful to derive the asymptotic properties of estimators in GARCH-type models. We start by considering the log-volatility. In the sequel, we use an alternative vector representation based on an iid sequence. Let $\mu_t = \omega_t + \boldsymbol{\pi}'\mathbf{x}_{t-1} + \int_0^1 \delta(u)X_{t-1}(u)du + \ln \eta_t^2$, let $\boldsymbol{\epsilon}_{t,q}^+ = (\mathbb{1}_{\{\epsilon_t > 0\}} \ln \epsilon_t^2, \dots, \mathbb{1}_{\{\epsilon_{t-q+1} > 0\}} \ln \epsilon_{t-q+1}^2)'$, $\boldsymbol{\epsilon}_{t,q}^- = (\mathbb{1}_{\{\epsilon_t < 0\}} \ln \epsilon_t^2, \dots, \mathbb{1}_{\{\epsilon_{t-q+1} < 0\}} \ln \epsilon_{t-q+1}^2)'$, $\mathbf{y}_t = (\{\boldsymbol{\epsilon}_{t,q}^+\}', \{\boldsymbol{\epsilon}_{t,q}^-\}', \ln \sigma_t^2, \dots, \ln \sigma_{t-p+1}^2)'$, $\mathbf{c}_t = (\mu_t \mathbb{1}_{\{\eta_t > 0\}}, \mathbf{0}'_{q-1}, \mu_t \mathbb{1}_{\{\eta_t < 0\}}, \mathbf{0}'_{q-1}, \mu_t, \mathbf{0}'_{p-1})'$ and let the $(2q + p) \times (2q + p)$ matrix

$$\mathbf{A}_t = \begin{pmatrix} \mathbb{1}_{\{\eta_t > 0\}} \boldsymbol{\alpha}_+ & \mathbb{1}_{\{\eta_t > 0\}} \boldsymbol{\alpha}_- & \mathbb{1}_{\{\eta_t > 0\}} \boldsymbol{\beta} \\ \mathbf{I}_{q-1} & \mathbf{0}_{q-1} & \mathbf{0}_{(q-1) \times q} & \mathbf{0}_{(q-1) \times p} \\ \mathbb{1}_{\{\eta_t < 0\}} \boldsymbol{\alpha}_+ & \mathbb{1}_{\{\eta_t < 0\}} \boldsymbol{\alpha}_- & \mathbb{1}_{\{\eta_t < 0\}} \boldsymbol{\beta} \\ \mathbf{0}_{(q-1) \times q} & \mathbf{I}_{q-1} & \mathbf{0}_{q-1} & \mathbf{0}_{(q-1) \times p} \\ \boldsymbol{\alpha}_+ & \boldsymbol{\alpha}_- & \boldsymbol{\beta} \\ \mathbf{0}_{(p-1) \times q} & \mathbf{0}_{(p-1) \times q} & \mathbf{I}_{p-1} & \mathbf{0}_{p-1} \end{pmatrix},$$

where \mathbf{I}_k denotes the $k \times k$ identity matrix. Model (2.1) is rewritten in matrix form as

$$\mathbf{y}_t = \mathbf{A}_t \mathbf{y}_{t-1} + \mathbf{c}_t. \quad (2.7)$$

Let $\gamma(\mathbf{A})$ be the top Lyapunov exponent of the sequence $\{\mathbf{A}_t, t \in \mathbb{Z}\}$.

Proposition 4 (Existence of a fractional log-moment). *Assume that $\gamma(\mathbf{A}) < 0$ and, for some $0 < s < 1$,*

$$E |\mathbb{1}_{\{\eta_0 \neq 0\}} \ln \eta_0^2|^s < \infty, \quad E \|\mathbf{x}_0\|^s < \infty, \quad E \left| \int_0^1 \delta(u)X_0(u)du \right|^s < \infty.$$

Then there exists a strictly stationary solution (ϵ_t) to (2.1) such that $E |\mathbb{1}_{\{\epsilon_t \neq 0\}} \ln \epsilon_t^2|^s < \infty$ and $E |\ln \sigma_t^2|^s < \infty$.

We now turn to the existence of moments.

Proposition 5 (Existence of real-order moments). *Assume $\rho(\mathbf{C}^{(\infty)}) < 1$. Let $c_\infty = \sum_{l \geq 0} \left\| \left(\mathbf{C}^{(\infty)} \right)^l \right\|$, $\lambda_x = \|\boldsymbol{\pi}\| c_\infty$, $\lambda_\eta = \max_{1 \leq i \leq q} (|\alpha_{i+}| \vee |\alpha_{i-}|) c_\infty$. Assume that for some $s > 0$, and some $\tau_1, \tau_2, \tau_3 > 0$ with $\tau_1^{-1} + \tau_2^{-1} + \tau_3^{-1} = 1$,*

$$\begin{aligned} E \exp(s\tau_1(\lambda_\eta \vee 1) |\ln \eta_0^2|) < \infty, \quad E \exp(s\tau_2(\lambda_x \vee 1) \|\mathbf{x}_0\|) < \infty, \\ E \exp\left(s\tau_3(c_\infty \vee 1) \left| \int_0^1 \delta(u)X_0(u)du \right|\right) < \infty. \end{aligned} \quad (2.8)$$

Then the stationary solution of Model (2.1) satisfies $\max \{E|\sigma_t^2|^s, E|\sigma_t^2|^{-s}, E|\epsilon_t^2|^s, E|\epsilon_t^2|^{-s}\} < \infty$.

2.3 Statistical inference

We first introduce several finite-dimensional parametric specifications $\delta(\cdot) = \delta(\cdot; \mathbf{b})$ for the functional parameter of model (2.1).

2.3.1 Parametric representation of the functional part

Consider the Hilbert space of square integrable functions with domain $[0, 1] : H = L^2[0, 1]$. The Hilbert space is equipped with the inner product $\langle f, g \rangle = \int_0^1 f(u)g(u)du$ and the resulting norm $\|\cdot\|$. Let $(\varphi_k)_{k \geq 1}$ be an orthonormal Hilbert basis of H . For instance, one can take the Fourier basis defined recursively by $\varphi_1(x) = 1$, $\varphi_{2k}(x) = \sqrt{2} \cos(2k\pi x)$, $\varphi_{2k+1}(x) = \sqrt{2} \sin(2k\pi x)$ for $k \geq 1$. Another example of Hilbert basis is the sequence of the normalized Legendre polynomials defined by $\varphi_1(x) = 1$ and, for $k \geq 2$, $\varphi_k(x) = c_k(x^k - \sum_{i=1}^{k-1} \langle x^k, \varphi_i \rangle \varphi_i(x))$ with c_k a constant such that $\|\varphi_k\| = 1$. Suppose that the coordinates of $\delta_0 \in H$ in the basis $(\varphi_k)_{k \geq 1}$ are defined by a parameter of finite dimension.

A1: There exists $\mathbf{b}_0 = (b_{1,0}, b_{2,0}, \dots, b_{M,0})'$ in \mathbb{R}^M such that, for all $k \geq 1$, $\langle \delta_0, \varphi_k \rangle = B_k(\mathbf{b}_0)$.

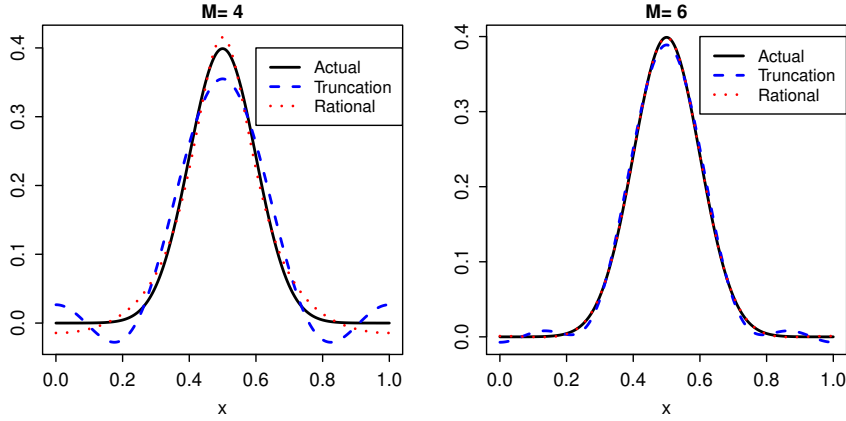
By Bessel's inequality, it is known that $(\langle \delta_0, \varphi_k \rangle)_{k \geq 0}$ is a square-summable sequence. This implies that $\langle \delta_0, \varphi_k \rangle$ is arbitrarily small when k is large. If one assumes there exists k_0 such that $\langle \delta_0, \varphi_k \rangle = 0$ for $k > k_0$, then **A1** holds with $b_{k,0} = \langle \delta_0, \varphi_k \rangle$ and any $M \geq k_0$. Another way to parametrize $\langle \delta_0, \varphi_k \rangle$ is to assume that the power series $\sum_{k=0}^{\infty} \langle \delta_0, \varphi_{k+1} \rangle z^k$ is a *rational function*, i.e. that we have the power series expansions

$$\sum_{k=0}^{\infty} \langle \delta_0, \varphi_{k+2} \rangle z^k = \frac{\Psi_0(z)}{\Phi_0(z)} \quad (2.9)$$

where $\Phi_0(z) = 1 - \sum_{i=1}^{p_1} \phi_{i0} z^i$ and $\Psi_0(z) = \sum_{i=0}^{q_1} \psi_{i0} z^i$, the roots of $\Phi_0(z)$ being outside the unit circle. Since Padé (1892), the approximation of an analytical function by a rational function has proven to be a key technique in numerical analysis. Note, however, that we are not considering here the Padé approximants of the function δ_0 , nor of its dual power series $\sum_{k=0}^{\infty} \langle \delta_0, \varphi_{k+1} \rangle z^k$. Instead we will consider optimal $L^2[0, 1]$ approximations, for which an algorithm is proposed in Appendix. Parameterizations of the form (2.9) are also standard in time series, in particular to parsimoniously represent a Wold decomposition by an ARMA(p_1, q_1) model (see Brockwell and Davis, 1991). Under (2.9), **A1** holds with $\mathbf{b}_0 = (\langle \delta_0, \varphi_1 \rangle, \phi_{10}, \dots, \phi_{p_1 0}, \psi_{00}, \dots, \psi_{q_1 0})'$. In the case $p_1 = q_1 = 1$, we have the explicit relations $B_1(\mathbf{b}) = b_1$ and $B_k(\mathbf{b}) = b_3 b_2^{k-2}$ for $k \geq 2$, with $\mathbf{b} = (b_1, b_2, b_3)'$. Note that we assume $\varphi_1 = 1$ in order to ensure that the parameterization is stable by linear transformations. Indeed, with obvious notation, the linear transformation $a\delta_0 + c$ satisfies **A1** with $b_1 = ab_{1,0} + c$, $b_2 = \phi_{10}, \dots, b_{p_1+1} = \phi_{p_1 0}, b_{p_1+2} = a\psi_{00}, \dots, b_{p_1+q_1+2} = a\psi_{q_1 0}$.

Example 1 (Rational representation of a function). *As an illustration, Figure 2.1 represents the Gaussian density f of mean 0.5 and standard deviation 0.1 (full black line), as well as its approximation by the truncated series $\sum_{k=1}^M \langle f, \varphi_k \rangle \varphi_k$ (dashed blue line) and its approximation by the "dual" rational function $\sum_{k=0}^{\infty} B_k \varphi_k$ (dotted red line), where $B_1 = \langle f, \varphi_1 \rangle = \int_0^1 f(x) dx$ and $\sum_{k=0}^{\infty} B_{k+2} z^k = \frac{\Psi_0(z)}{\Phi_0(z)}$, with a total of $M - 1$ parameters for the polynomials $\Psi_0(z)$ and $\Phi_0(z)$. The algorithm used to fit these parameters is described in Appendix. For this illustration we took the Fourier basis $(\varphi_k)_{k \geq 1}$. As can be seen in the figure, the rational approximation is more accurate than the approximation by truncation. The rational approximation is already quite good with $M = 4$ parameters and it is almost perfect with $M = 6$ parameters.*

Figure 2.1: Gaussian density, with its truncated and rational approximations based on M parameters.



Assuming $X_t \in L^2[0, 1]$ a.s., let the random variables $A_{t,k} = \langle X_t, \varphi_k \rangle$. For \mathbf{b} such that $\sum_{k=1}^{\infty} B_k^2(\mathbf{b}) < \infty$, we have $X_t \stackrel{L^2[0,1]}{=} \sum_{k=1}^{\infty} A_{t,k} \varphi_k$, $\delta \stackrel{L^2[0,1]}{=} \sum_{k=1}^{\infty} B_k(\mathbf{b}) \varphi_k$ and $\int_0^1 \delta(u) X_t(u) du = \sum_{k=1}^{\infty} B_k(\mathbf{b}) A_{t,k}$. The notation $f \stackrel{L^2[0,1]}{=} g$, or simply $f = g$, means that $\|f - g\| = 0$.

2.3.2 Quasi-maximum likelihood estimation

In this section, we estimate the parameters of the volatility model (2.1) without assuming a particular distribution for the iid noise η_t . Let $d = 1 + 3q + p + r + M$ be the total number of unknown parameters. We define the generic parameter

$$\boldsymbol{\theta} = (\omega, \boldsymbol{\omega}'_-, \boldsymbol{\alpha}'_-, \boldsymbol{\alpha}'_+, \boldsymbol{\beta}', \boldsymbol{\pi}', \mathbf{b}')' \in \mathbb{R}^d,$$

with $\boldsymbol{\omega}'_- = (\omega_{1-}, \dots, \omega_{q-})$, $\boldsymbol{\alpha}'_- = (\alpha_{1-}, \dots, \alpha_{q-})$, $\boldsymbol{\alpha}'_+ = (\alpha_{1+}, \dots, \alpha_{q+})$ and $\boldsymbol{\beta}' = (\beta_1, \dots, \beta_p)$. The true unknown parameter is denoted by $\boldsymbol{\theta}_0 = (\omega_0, \boldsymbol{\omega}'_{-,0}, \boldsymbol{\alpha}'_{-,0}, \boldsymbol{\alpha}'_{+,0}, \boldsymbol{\beta}'_0, \boldsymbol{\pi}'_0, \mathbf{b}'_0)'$. Let $(\epsilon_1, \dots, \epsilon_n)$ be a realization of length n of the sta-

tionary solution (ϵ_t) to model (2.1), and let $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ and (X_1, \dots, X_n) be the corresponding observations of the exogenous scalar variables and the exogenous real-valued functional covariate. The quasi-maximum likelihood estimator (QMLE) of $\boldsymbol{\theta}_0$ is defined as follows

$$\hat{\boldsymbol{\theta}}_n = \arg \min_{\boldsymbol{\theta} \in \Theta} \tilde{Q}_n(\boldsymbol{\theta})$$

where $\Theta \subset \mathbb{R}^d$ is the parameter set and

$$\tilde{Q}_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t=1}^n \tilde{\ell}_t, \quad \tilde{\ell}_t = \tilde{\ell}_t(\boldsymbol{\theta}) = \frac{\epsilon_t^2}{\tilde{\sigma}_t^2(\boldsymbol{\theta})} + \ln \tilde{\sigma}_t^2(\boldsymbol{\theta})$$

with $\tilde{\sigma}_t^2(\boldsymbol{\theta})$ recursively defined for $t \geq 1$ by

$$\ln \tilde{\sigma}_t^2(\boldsymbol{\theta}) = \omega_t + \sum_{i=1}^q \alpha_{i,t} \ln \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \ln \tilde{\sigma}_{t-j}^2(\boldsymbol{\theta}) + \boldsymbol{\pi}' \mathbf{x}_{t-1} + \sum_{k=1}^{\infty} B_k(\mathbf{b}) A_{t-1,k}$$

using the initial values $\epsilon_0, \dots, \epsilon_{1-q}, \tilde{\sigma}_0^2 \geq 0, \dots, \tilde{\sigma}_{1-p}^2 \geq 0$.

2.3.2.1 Strong consistency

Let the polynomials $\mathcal{A}_{\boldsymbol{\theta}_+}(z) = \sum_{i=1}^q \alpha_{i+} z^i$, $\mathcal{A}_{\boldsymbol{\theta}_-}(z) = \sum_{i=1}^q \alpha_{i-} z^i$, $\mathcal{W}_{\boldsymbol{\theta}}(z) = \sum_{i=1}^q \omega_{i-} z^i$ and $\mathcal{B}_{\boldsymbol{\theta}}(z) = 1 - \sum_{j=1}^p \beta_j z^j$. Let \mathcal{F}_{t-1} denote the σ -field generated by $\{\epsilon_u, \mathbf{x}_u, X_u, u < t\}$ and $\mathcal{F}_{t-1, t-s-1}^{(\mathbf{x}, X), \eta}$ denote the σ -field generated by $\{\eta_{t-s-k}, \mathbf{x}_{t-k}, X_{t-k}, k > 0\}$. To show the consistency of the QMLE, we need the following assumptions.

A2: $\boldsymbol{\theta}_0 \in \Theta$, Θ is compact.

A3: $E[\mathbf{1}_{\{\epsilon_t \neq 0\}} \ln \epsilon_t^2] < \infty$, $E\|\mathbf{x}_t\| < \infty$ and $\sum_{k=1}^{\infty} E|A_{1,k}| < \infty$.

A4: For all $s \geq 1$, the support of the distribution of η_{t-s} given $\mathcal{F}_{t-1, t-s-1}^{(\mathbf{x}, X), \eta}$ contains at least two positive values and two negative values.

A5: We have $q > 0$ and, if $p > 0$, at least one of the following conditions hold: a) $\omega_{-,0} \neq 0$, $\mathcal{B}_{\boldsymbol{\theta}_0}(z)$ has no common root with $\mathcal{W}_{\boldsymbol{\theta}_0}(z)$ and $(\omega_{q-,0}, \beta_{p,0}) \neq 0$;

b) $\alpha_{+,0} \neq 0$, $\mathcal{B}_{\boldsymbol{\theta}_0}(z)$ has no common root with $\mathcal{A}_{\boldsymbol{\theta}_0+}(z)$ and $(\alpha_{q+,0}, \beta_{p,0}) \neq 0$; c) $\alpha_{-,0} \neq 0$, $\mathcal{B}_{\boldsymbol{\theta}_0}(z)$ has no common root with $\mathcal{A}_{\boldsymbol{\theta}_0-}(z)$ and $(\alpha_{q-,0}, \beta_{p,0}) \neq 0$.

A6: $\gamma < 0$ and, for all $\boldsymbol{\theta} \in \Theta$, $\sum_{j=1}^p \beta_j < 1$ and $\sum_{k=1}^{\infty} B_k^2(\mathbf{b}) < \infty$.

A7: If \mathbf{a} is a non zero vector of \mathbb{R}^r or if $\delta = \sum_{k=1}^{\infty} B_k(\mathbf{b}) \varphi_k \in L^2[0, 1]$ is such that $\|\delta\| \neq 0$ then, conditional on \mathcal{F}_{t-1} , the random variable $\mathbf{a}' \mathbf{x}_t + \int_0^1 \delta(u) X_t(u) du$ is not degenerated.

A8: If $B_k(\mathbf{b}) = B_k(\mathbf{b}_0) \forall k \geq 1$, then $\mathbf{b} = \mathbf{b}_0$.

Assumption **A2** is standard to show the consistency of the QMLE of GARCH models. The first two conditions of Assumption **A6** are also standard. The last condition entails that $\delta := \sum_{k=1}^{\infty} B_k(\mathbf{b}) \varphi_k$ belongs to $L^2[0, 1]$. The first moment condition of Assumption **A3** is studied in Proposition 2. Assumptions **A4**, **A5**, **A7** and **A8** are identifiability conditions. Assumption **A4** prevents taking covariates, for instance of the form $\mathbf{x}_t = \epsilon_{t-s}$,

for which the information given by \mathbf{x}_{t-1} is redundant with that of $\{\epsilon_u, u < t - s\}$. It is similar to Assumption **A4** made by Francq and Thieu (2019) in the case of GARCH-X models. For Log-GARCH models without exogenous variables, Francq et al. (2013) made an assumption similar to, but slightly stronger than, **A5**. Assumptions **A7** is necessary to avoid multicollinearity of the covariates. Under (2.9), **A8** is satisfied if $\Phi_0(z)$ and $\Psi_0(z)$ have no common root and $\psi_{q1,0} \neq 0$ or $\phi_{p1,0} \neq 0$

Theorem 2 (Consistency of the QMLE). *Under Assumptions **A0-A8** the QMLE of $\boldsymbol{\theta}_0$ in Model (2.1) is strongly consistent, i.e. we have $\hat{\boldsymbol{\theta}}_n \rightarrow \boldsymbol{\theta}_0$ almost surely.*

2.3.2.2 Asymptotic normality

To show the asymptotic normality, the following additional assumptions are made.

A9: $\boldsymbol{\theta}_0 \in \mathring{\Theta}$, where $\mathring{\Theta}$ denotes the interior of Θ .

A10: $\kappa_\eta := E\eta_t^4 < \infty$.

A11: For some $\nu > 0$ and $r > 0$, $\sum_{k=1}^{\infty} (EA_{1,k}^{2+\nu})^{\frac{2}{2+\nu}} < \infty$ and

$$E \exp \left\{ r \left(\mathbb{1}_{\{\eta_t \neq 0\}} |\ln \eta_t^2| + |\ln \sigma_t^2| + \|\mathbf{x}_t\| + \sqrt{\int_0^1 X_t^2(u) du} \right) \right\} < \infty. \quad (2.10)$$

A12: The applications $B_k(\cdot)$ are twice continuously differentiable in a neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$, and

$$\sum_{k=1}^{\infty} \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{\partial}{\partial \mathbf{b}} B_k(\mathbf{b}) \right\|^2 < \infty, \quad \sum_{k=1}^{\infty} \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{\partial^2}{\partial \mathbf{b} \partial \mathbf{b}'} B_k(\mathbf{b}) \right\|^2 < \infty.$$

A13: If \mathbf{c} is a non zero vector of \mathbb{R}^M then there exists $k \geq 1$ such that $\mathbf{c}' \frac{\partial}{\partial \mathbf{b}} B_k(\mathbf{b}_0) \neq 0$.

Assumptions **A9** and **A10** are standard. Assumptions **A11** and **A12** are needed to show the existence of the hessian matrix of the objective function in a neighborhood of $\boldsymbol{\theta}_0$. In the symmetric case with $q = 0$ and $p = 1$ it is clear that the Cramer condition (2.10) is necessary for the ratio

$$\frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} = \exp \left\{ \omega^* + \alpha^* \mathbb{1}_{\{\epsilon_t \neq 0\}} \ln \epsilon_{t-1}^2 + \boldsymbol{\pi}^{*'} \mathbf{x}_{t-1} + \int_0^1 \delta^*(u) X_{t-1}(u) du \right\}$$

to admit some moment in a neighborhood of $\boldsymbol{\theta}_0$. Assumption **A12** is very mild. In particular, it is satisfied for the two examples of parameterization that we considered, that is to say when $\mathbf{b} = (\langle \delta, \varphi_1 \rangle, \dots, \langle \delta, \varphi_M \rangle)'$ and when $\mathbf{b} = (\langle \delta, \varphi_1 \rangle, \phi_1, \dots, \phi_p, \psi_0, \dots, \psi_q)'$. **A13** is a necessary identifiability assumption, which is also satisfied in the two parameterizations.

Theorem 3 (Asymptotic normality of the QMLE). *Under the assumptions of Theorem 2 and **A9-A13**, we have $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, (\kappa_\eta - 1)\mathbf{J}^{-1})$ as $n \rightarrow \infty$, where $\mathbf{J} = E\left[\frac{\partial}{\partial \boldsymbol{\theta}} \ln \sigma_t^2(\boldsymbol{\theta}_0) \frac{\partial}{\partial \boldsymbol{\theta}'} \ln \sigma_t^2(\boldsymbol{\theta}_0)\right]$ is a positive definite matrix.*

An important advantage of log-volatility models over standard GARCH is that the coefficients are not constrained to be non negative. Consequently, **A9** may be satisfied when components of $\boldsymbol{\theta}_0$ are equal to zero, and Theorem 3 can be directly used to test the nullity of GARCH coefficients. In other words, a parameter equal to zero—for instance $\pi_{0i} = 0$ —does not cause the difficulties encountered in standard GARCH-X models (see Francq and Thieu, 2019). For tests, consistent estimators of \mathbf{J} and κ_η are however needed. The next proposition shows that one can take empirical estimators. Let $\hat{\eta}_t = \epsilon_t / \hat{\sigma}_t$, where $\hat{\sigma}_t = \tilde{\sigma}_t(\hat{\boldsymbol{\theta}}_n)$.

Proposition 6 (Estimation of the QMLE asymptotic variance). *Under the assumptions of Theorem 3*

$$\hat{\kappa} = \frac{1}{n} \sum_{t=1}^n \hat{\eta}_t^4 \quad \text{and} \quad \hat{\mathbf{J}} = \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \tilde{Q}_n(\hat{\boldsymbol{\theta}}_n)$$

are strongly consistent estimators of κ_η and \mathbf{J} .

2.3.3 Goodness-of-fit portmanteau tests

For testing the adequacy of volatility models, Li and Mak (1994) and Ling and Li (1997) proposed portmanteau tests based on the autocovariances of the squared residuals. The asymptotic distribution of these tests has been further investigated by many authors: see in particular Berkes et al. (2003a) for the standard GARCH models, Carbon and Francq (2011) for the APARCH models, Francq et al. (2018) for Log-GARCH and EGARCH models. Other goodness-of-fit tests exist. In particular, Escanciano and Olmo (2010) proposed diagnostic tests of a general class of conditionally heteroskedastic time series models. Leucht et al. (2015) also proposed specification tests of Cramér-von Mises type for GARCH(1,1) models. Portmanteau tests remain however the most widely used diagnostic tools for time series (see the monograph of Li, 2004).

We thus consider the null hypothesis

$$H_0 : \text{the process } (\epsilon_t) \text{ satisfies Model (2.1),}$$

and define the autocovariances of the squared residuals at lag h , for $1 \leq h < n$, by $\hat{r}_h = n^{-1} \sum_{t=h+1}^n (\hat{\eta}_t^2 - 1)(\hat{\eta}_{t-h}^2 - 1)$. For any fixed integer m , $1 \leq m < n$, consider the statistic $\hat{\mathbf{r}}_m = (\hat{r}_1, \dots, \hat{r}_m)'$. Define the $m \times d$ matrix $\hat{\mathbf{K}}_m$ whose row h , for $1 \leq h \leq m$, is

$$\hat{\mathbf{K}}_m(h, \cdot) = \frac{1}{n} \sum_{t=h+1}^n (\hat{\eta}_{t-h}^2 - 1) \frac{\partial \log \tilde{\sigma}_t^2(\hat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}'}. \quad (2.11)$$

We need to slightly reinforce Assumption **A4**.

A4*: For all $s \geq 1$, the support of the distribution of η_{t-s} given $\mathcal{F}_{t-1, t-s-1}^{(x, X), \eta}$ contains at least three positive values and three negative values.

Theorem 4 (Portmanteau adequacy test). *Under H_0 , the assumptions of Theorem 3 and **A4*** we have*

$$n\hat{\mathbf{r}}_m' \hat{\mathbf{D}}^{-1} \hat{\mathbf{r}}_m \xrightarrow{d} \chi_m^2$$

where $\hat{\mathbf{D}} = (\hat{\kappa} - 1)^2 \mathbf{I}_m - (\hat{\kappa} - 1) \hat{\mathbf{K}}_m \hat{\mathbf{J}}^{-1} \hat{\mathbf{K}}_m'$. Therefore, H_0 is rejected at the nominal level $\underline{\alpha} \in (0, 1)$ when $\{n\hat{\mathbf{r}}_m' \hat{\mathbf{D}}^{-1} \hat{\mathbf{r}}_m \geq \chi_m^2(1 - \underline{\alpha})\}$.

2.4 Simulations

The aim of this section is to evaluate the finite sample properties of both the QMLE of model (2.1) and the portmanteau test statistics introduced in the previous section. To do so, we start by simulating a sequence of returns with log-volatility given by

$$\begin{aligned} \ln \sigma_t^2 = & \omega + \omega_- 1_{\epsilon_{t-1} < 0} + (\alpha_- 1_{\epsilon_{t-1} < 0} + \alpha_+ 1_{\epsilon_{t-1} > 0}) \ln \epsilon_{t-1}^2 \\ & + \beta \ln \sigma_{t-1}^2 + \int_0^1 \delta(u) X_{t-1}(u) du. \end{aligned} \quad (2.12)$$

We consider two sample sizes in the simulation study, namely $n = 1000$ and $n = 2000$. To build the functional exogenous variable, we collect the 5-minute log-returns of Bank of America from December 29, 2004 to December 30, 2008 and from December 13, 2000 to December 30, 2008, corresponding to $n = 1000$ and $n = 2000$, respectively, and we use the data from 9:35 am to 4:00 pm. To obtain a smooth functional variable X_t , we take the logarithm of a rolling window average on the 5-minute squared log-returns using 24 observations. Then, we project X_t on the first $M = 7$ normalized Legendre polynomials φ_k and consider the coefficients $A_{t,k} = \langle X_t, \varphi_k \rangle$ for $k = 1, \dots, M$ and $t = 1, \dots, n$.

Our simulation procedure is as follows. First, we simulate 1000 samples of size $n = 1000, 2000$ with standard Gaussian innovations ($\eta_t \sim \mathcal{N}(0, 1)$). Note that the model parameters displayed in Table 2.1 are calibrated via the daily log-returns of Bank of America using the same periods than that used for the exogenous variable. Second, we estimate the parameters of the 1000 time-series generated by QMLE as detailed in Section 2.3.2. Our estimation results are displayed in Table 2.1. The second and third rows disclose the mean and median estimated parameters across simulations. The fourth and fifth rows report the corresponding Mean Absolute Errors (MAE) and Mean Squared Errors (MSE). The sixth row displays the standard deviation across simulations. Three remarks should be raised. First, we find that the parameter estimates are close to their expected values. Second, we observe that the average bias of estimates decreases with the sample size n . Finally, the standard deviation of estimates is decreasing with n confirming

our previous findings that the developed QMLE estimators are consistent. In the last three rows of both panels we display the coverage rates associated with 90%, 95% and 99% asymptotic confidence intervals, that is the relative frequency of simulations for which the true parameter lies in the confidence interval. Our simulation results are satisfactory. We find that the statistical inference is accurate for realistic sample sizes. In average, considering a 90% confidence level, the coverage rates are equal to 87.5% and 89.1% for $n = 1000$ and $n = 2000$, respectively. Similarly, at a 95% and 99% confidence level, the coverage rates are equal to 93.2% and 98.1% for $n = 1000$, and 93.8% and 98.4% for $n = 2000$, respectively.

Table 2.1: Simulation results.

	ω	ω_-	α_-	α_+	β	b_1	b_2	b_3	b_4	b_5	b_6	b_7
$n = 1000$												
θ_0	-2.181	0.229	0.018	-0.020	0.608	0.193	0.106	-0.431	0.132	0.307	1.869	0.331
Mean	-2.256	0.220	0.015	-0.023	0.598	0.199	0.105	-0.436	0.133	0.308	1.879	0.332
Median	-2.218	0.209	0.014	-0.023	0.605	0.197	0.103	-0.429	0.138	0.305	1.867	0.372
MAE	0.403	0.273	0.019	0.020	0.062	0.032	0.089	0.143	0.200	0.320	0.420	0.419
MSE	0.275	0.116	0.001	0.001	0.007	0.002	0.013	0.033	0.062	0.160	0.287	0.276
Std	0.519	0.340	0.024	0.024	0.081	0.041	0.112	0.183	0.248	0.401	0.536	0.525
$IC_{90\%}$	86.3%	87.4%	87.1%	88.0%	85.3%	87.2%	87.5%	86.2%	89.5%	87.4%	89.1%	89.1%
$IC_{95\%}$	92.5%	93.4%	92.8%	93.3%	91.3%	92.0%	93.8%	92.2%	95.6%	94.0%	93.4%	94.6%
$IC_{99\%}$	97.1%	98.9%	98.7%	98.1%	96.2%	96.8%	98.4%	97.9%	99.0%	98.2%	98.8%	98.9%
$n = 2000$												
θ_0	-2.283	0.218	0.032	0.001	0.556	0.239	0.052	0.078	-0.368	0.545	0.506	-0.392
Mean	-2.314	0.217	0.031	-0.001	0.552	0.242	0.046	0.084	-0.376	0.551	0.496	-0.390
Median	-2.290	0.225	0.031	-0.001	0.555	0.240	0.048	0.084	-0.368	0.549	0.496	-0.395
MAE	0.305	0.187	0.014	0.013	0.055	0.030	0.061	0.097	0.152	0.228	0.282	0.298
MSE	0.153	0.056	0.000	0.000	0.005	0.001	0.006	0.015	0.037	0.077	0.131	0.144
Std	0.390	0.236	0.017	0.017	0.070	0.038	0.077	0.123	0.193	0.278	0.362	0.380
$IC_{90\%}$	87.6%	89.3%	89.8%	90.4%	86.4%	86.2%	89.5%	88.6%	89.0%	90.3%	90.3%	91.7%
$IC_{95\%}$	92.6%	93.9%	94.8%	94.4%	91.4%	91.9%	94.2%	94.1%	93.4%	94.9%	94.5%	95.3%
$IC_{99\%}$	97.2%	98.6%	99.5%	98.4%	96.8%	97.7%	98.7%	98.3%	98.6%	98.9%	98.9%	99.2%

Let us now look at the finite sample size properties of the portmanteau test statistics introduced in Section 2.3.3. We consider three competing models that can be regarded as sub-models of the representation (2.12). The first alternative model, hereafter denoted by A_1 , is of the form (2.12), but without the functional covariate X_t and with the realized volatility as scalar exogenous variable⁵. The alternative model A_2 is Model (2.12) without

⁵The realized volatility can be seen as a summary of the intraday data since it is the sum of the 5-minutes squared log-returns on a day.

covariates (*i.e.* with $\mathbf{b} \equiv 0$). The last model, henceforward denoted A_3 , is a Log-GARCH(1,1) model without exogenous variable and without asymmetry. Therefore, it is equivalent to the model (2.12) with $\omega_- = 0$, $\alpha_- = \alpha_+$ and $\mathbf{b} = 0$.

For the 1000 replications, $n = 1000$ and $n = 2000$, we estimate by QML the parameters associated with the true model (2.12) and the three misspecified models A_1 , A_2 , and A_3 . Then, we apply the corresponding portmanteau tests on the squared residuals. Table 2.2 reports the rejection rates at various nominal levels, namely $\alpha = 10\%, 5\%, 1\%$. We consider a number of lags $m = 6, 12, 18, 24, 30$ to build the empirical auto-covariances used in the portmanteau tests. The left panel of Table 2.2 reports the rejection rates for $n = 1000$ and the right panel is for $n = 2000$. The empirical sizes are close to their nominal values, meaning that the portmanteau test is well-sized. Several remarks should be raised regarding the power experiments. First, we find that the rejection rates associated with the alternative models A_1 , A_2 , and A_3 , are substantially higher than the considered nominal levels, as displayed in the last nine rows. Second, we confirm that the frequencies of rejection are increasing with the sample size n , indicating that the test is consistent for the considered alternatives. Finally, we verify that there is a general improvement of the empirical powers from A_1 to A_3 given that A_3 is the most restricted model and A_1 is the closest model, compared to model (2.12). In other words, the tighter the constraints imposed to the true model, the higher the rejection rates.

Table 2.2: Portmanteau statistics (in %).

	m	6	12	18	24	30	6	12	18	24	30
		$n = 1000$					$n = 2000$				
True model	10%	9.1	9.0	9.0	9.6	8.8	7.6	8.8	9.1	10.3	9.8
	5%	4.3	3.9	4.4	5.1	4.9	3.8	3.8	4.3	5.2	5.3
	1%	0.4	0.7	0.7	0.8	0.9	0.6	1.0	1.0	1.4	1.2
A_1	10%	19.3	18.4	23.1	24.5	24.4	24.5	27.0	31.4	31.6	32.8
	5%	12.6	11.4	15.3	16.3	16.0	17.2	18.8	20.9	22.4	24.2
	1%	4.6	4.4	5.1	6.5	7.2	6.0	8.0	9.7	10.8	11.6
A_2	10%	65.5	72.4	80.4	81.8	81.3	90.6	93.1	95.5	96.6	96.9
	5%	57.5	65.8	73.5	76.8	77.0	86.5	90.9	93.2	95.3	95.2
	1%	44.8	53.5	62.0	67.9	68.3	74.1	82.5	87.3	90.8	92.4
A_3	10%	72.6	79.2	85.0	86.8	82.8	94.7	96.4	97.2	97.8	97.7
	5%	66.7	74.8	80.9	83.8	80.2	92.6	94.8	95.9	96.9	96.9
	1%	52.9	63.1	71.6	75.7	74.6	85.4	90.6	92.1	93.8	94.3

2.5 Empirical Study

We now estimate log-GARCH models, including or not functional covariates, on a panel of ten stocks.

2.5.1 Data description

We consider ten large capitalizations from the NYSE and the NASDAQ. The list of tickers and company names is reported in 2.7.3. Our sample period spans from January 4, 1999 to December 31, 2008. For the dependent variable, we collect the daily log-returns of the adjusted closing prices, yielding 2489 observations per series. For the exogenous variable, we consider the 5-minute log-returns. The 5-minute log-returns are collected every trading days from 9:35 am to 4:00 pm, yielding 77 observations per day.⁶ The daily log-returns are collected from Yahoo Finance and the 5-minute log-returns are issued from TickData.

The daily log-returns of the ten assets under consideration display several cluster of volatilities and a general increase of volatility for the ten stocks at the end of the sample corresponding to the early stages of the 2007-2009 global financial crisis. The daily log-returns hide a great heterogeneity at the intraday level. To emphasize this, Figure 2.2 displays the 5-minutes squared log-returns of the stocks for the first 25 trading days of the sample period, hence corresponding to January 1999. Vertical dotted lines materialize the separation between consecutive trading days. We observe that within a given day the profile of the 5-minutes squared log returns is quite irregular and noisy.

To reduce the noise and ease the implementation of the functional approach, we smooth the 5-minutes squared log returns. Figure 2.3 displays the smooth 5-minutes squared log returns (in black). As in the simulation study, we consider the logarithm of a rolling window average with 24 observations to smooth the 5-minutes squared log-returns. Then, we apply the normalized Legendre polynomials with $M = 7$ to calculate the coefficients that will be used in our model. The corresponding approximation is reported in blue in Figure 2.3. The approximated curves seem to capture the main information contained in the intraday variables.

2.5.2 In-sample estimation

Our main goal is to investigate whether the intraday functional variable is useful to explain the future squared daily log-returns. In the sequel, we consider a rolling window estimation scheme for the in-sample estimation. More formally, we use the n last observations (i.e. $\epsilon_{t-n+1}, \dots, \epsilon_t$) of the stock return to estimate the parameter. We set $n = 1000$,

⁶We do not include the first observation of the opening trading session, i.e. 9:30 am, because of the unusual business activity at that time.

Figure 2.2: 5-minute squared log-returns for the first 25 sample days for the 10 stocks.

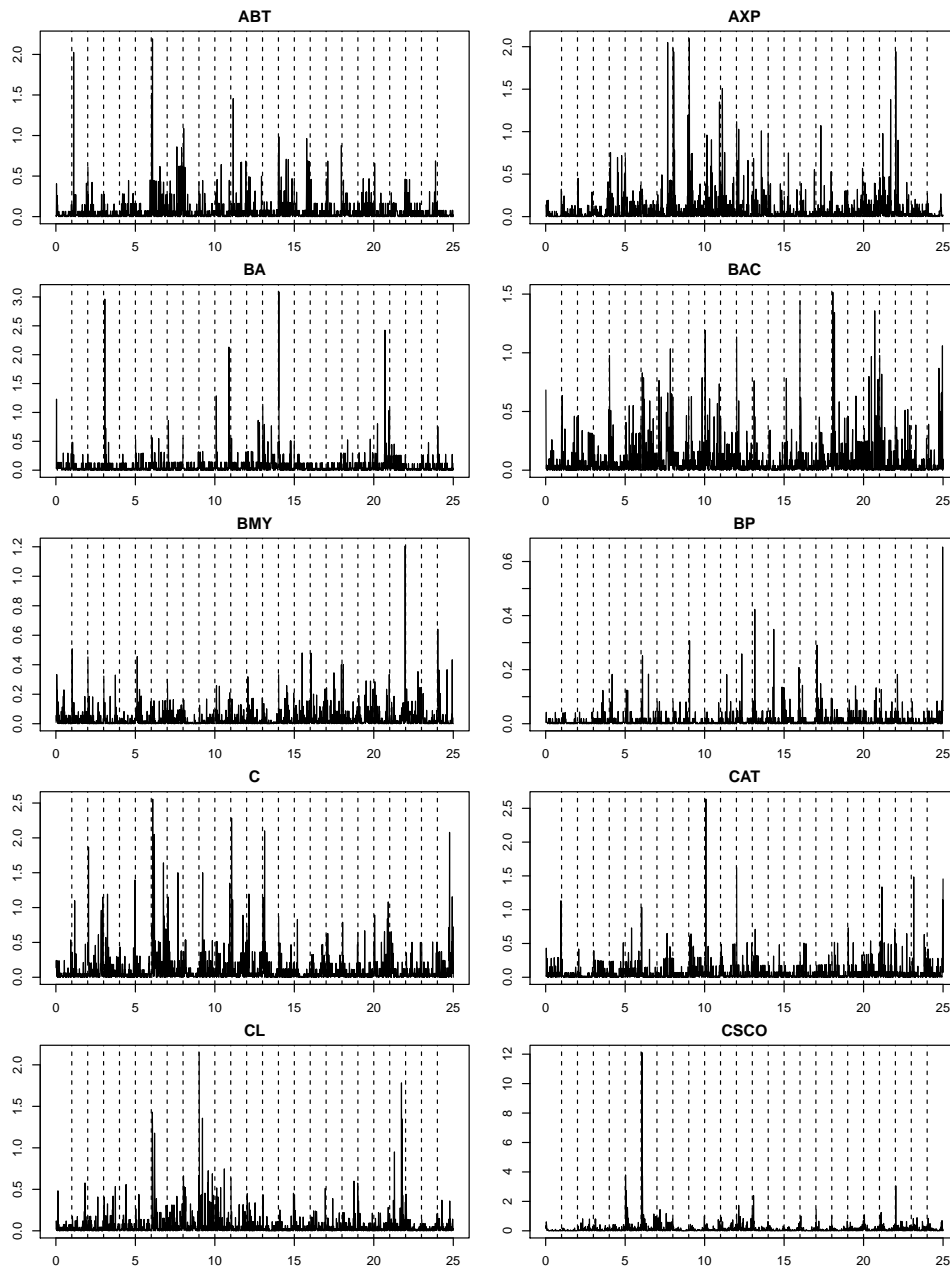
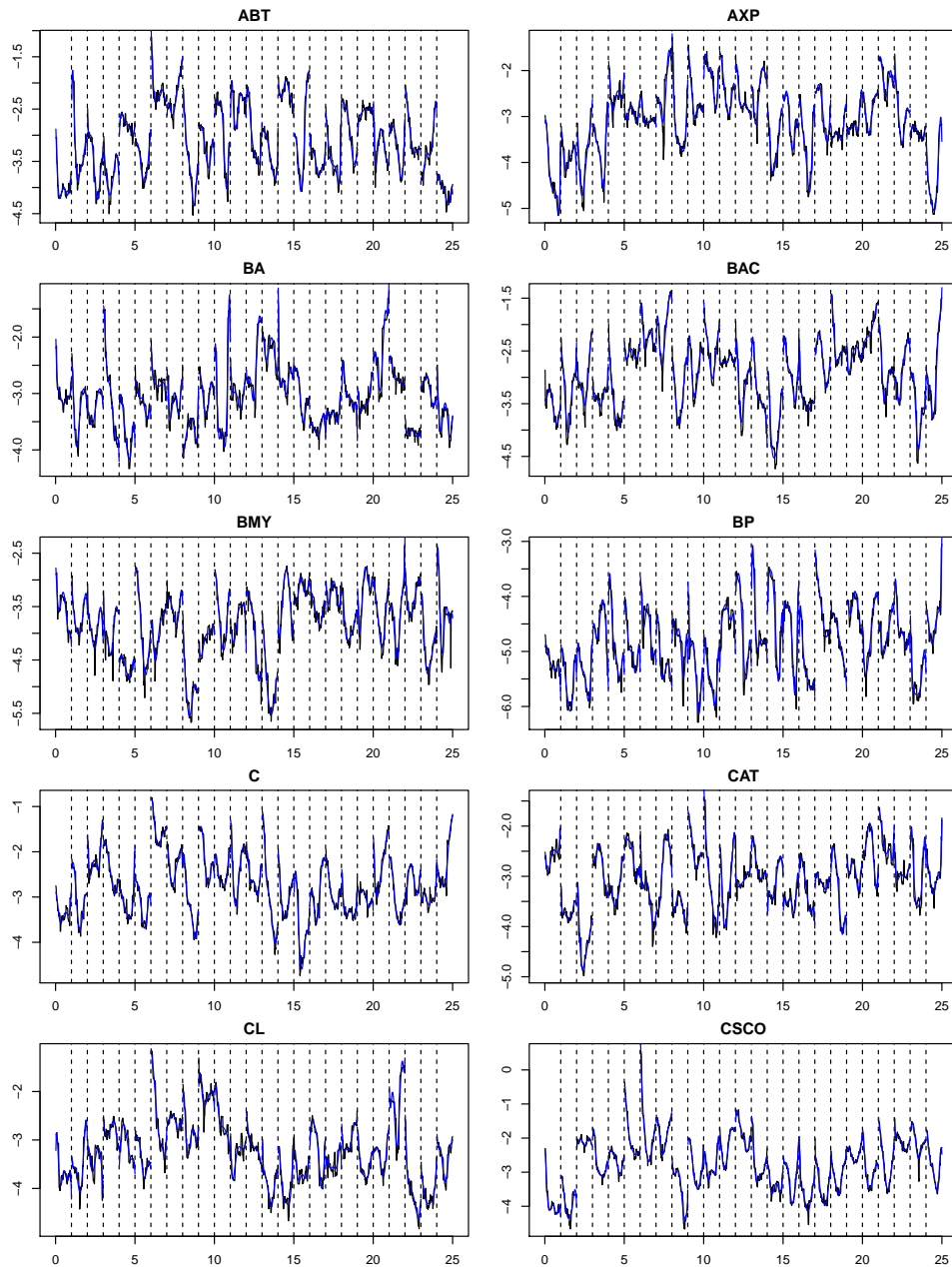


Figure 2.3: Transformation of the intraday variables (black curves) for the first 25 sample days of all stocks and their corresponding approximations (blue curves).



resulting in a 4-year rolling window. The parameter estimates are updated every day and we then obtain a series of 1489 estimated coefficients per stock.

Table 2.3 gives a summary of the parameter estimates. The top three rows of each panel report the mean, the median and the standard deviation of the estimated parameters. In the fourth row referred as IRR (for Individual Rejection Rate), we test for the individual significance of the parameters. In the last row referred as JRR (Joint Rejection Rate), we look at the relevance of the functional variable, *i.e.* we test $H_0 : \delta_0 \equiv 0$ (or equivalently $\mathbf{b}_0 = 0$). In both cases, we build Wald test statistics using the estimated standard errors issued from the asymptotic theory and we report the rejection frequency at the 5% significance level.

Two important results stand out. First, we find that the individual components of the functional variable (taken in isolation) are informative for modeling the dependent variable. For instance, we observe that the b_1 coefficient is statistically significant in more than 70% of cases, for eight out of ten assets (that is for all series but ABT and CL). Looking at b_2 , we find that the coefficient is significant in more than 20% of cases for four out of ten assets. Second, we find that the introduction of the intraday variable is overall very useful to explain future daily volatility. Looking at the test for nullity of δ_0 , we find that the corresponding rejection frequencies exceed 60% for six out of the ten assets considered.

To go a step further, we report in Figure 2.4 the dates of rejection of the joint test at 1% (dark shaded area) and 5% (light shaded area) significance levels, as well as corresponding daily log-returns. Overall, we find that the introduction of the functional variable is of main interest in period of large increase in volatility. We observe large clusters of rejection of the null hypothesis at the end of the sample where volatility is high due to the impact of the financial crisis on financial markets. Consequently, the introduction of intraday log-returns in the volatility equation is likely to improve the model depending whether the uncertainty in future prices is high or low.

Figure 2.5 reports the intraday estimated curve associated with the ten tickers. The curve is deduced from the $\hat{\mathbf{b}}$ estimated coefficients and the functional basis that has been used to rebuild the $\hat{\delta}(u)$ estimated intraday curve of each asset.⁷ We can distinguish three different types of profiles. First, BAC, BP, CAT and CL assets admit almost U-shape curves. Consequently, the most informative intraday log-returns are those happening at the beginning and the end of the day whereas the middle of the day does not really matter. Second, BMY and CSCO display curves with almost constant values along the day and higher values at the end of the day, while C depicts the reverse profile. In other words, the most indicative period in a day to explain daily volatility for BMY and CSCO

⁷To save space, we only report the average of estimated coefficients computed over the 1489 estimations.

Table 2.3: In-sample estimation results.

		ω	ω_-	α_-	α_+	β	b_1	b_2	b_3	b_4	b_5	b_6	b_7	
ABT	Mean	-1.800	0.652	0.053	-0.007	0.727	0.100	0.111	-0.067	-0.093	0.358	0.335	-0.118	
	Median	-1.161	0.683	0.057	-0.010	0.823	0.090	0.079	-0.015	-0.076	0.225	0.056	-0.235	
	Std	1.625	0.252	0.024	0.021	0.248	0.065	0.115	0.184	0.165	0.779	0.697	0.558	
	IRR	79.1%	30.3%	32.3%	0.0%	92.5%	23.5%	0.0%	0.0%	0.0%	0.0%	15.1%	0.6%	0.0%
	JRR										17.3%			
AXP	Mean	-2.506	0.818	0.026	-0.057	0.585	0.267	0.031	0.301	-0.085	-0.096	-0.417	-0.908	
	Median	-2.314	0.829	0.030	-0.051	0.654	0.239	0.023	0.329	0.047	-0.041	-0.353	-0.916	
	Std	1.107	0.316	0.017	0.022	0.196	0.108	0.091	0.263	0.351	0.365	0.389	0.327	
	IRR	100.0%	64.4%	2.1%	58.7%	96.4%	100.0%	0.0%	41.0%	0.2%	5.8%	0.0%	13.6%	
	JRR										77.2%			
BA	Mean	-1.971	0.110	0.036	0.036	0.603	0.177	0.083	0.273	0.028	0.561	0.132	-0.552	
	Median	-1.507	0.230	0.025	0.044	0.699	0.138	0.087	0.256	0.063	0.474	0.160	-0.585	
	Std	1.264	0.348	0.033	0.022	0.244	0.115	0.093	0.239	0.184	0.441	0.485	0.513	
	IRR	99.9%	0.0%	28.8%	30.0%	84.3%	95.1%	1.7%	17.9%	0.0%	15.5%	0.7%	4.3%	
	JRR										23.4%			
BAC	Mean	-2.245	0.924	0.026	-0.062	0.652	0.219	0.018	0.108	0.220	0.843	-0.244	0.518	
	Median	-2.364	1.018	0.028	-0.082	0.628	0.241	0.012	0.113	0.210	0.913	-0.129	0.608	
	Std	0.758	0.426	0.015	0.046	0.086	0.068	0.162	0.118	0.136	0.299	0.513	0.769	
	IRR	99.9%	59.0%	5.4%	66.8%	99.7%	97.5%	6.6%	0.0%	0.0%	6.4%	0.0%	1.1%	
	JRR										64.6%			
BMY	Mean	-2.380	0.785	0.048	-0.021	0.631	0.182	0.141	0.761	0.685	1.124	1.287	0.896	
	Median	-2.071	0.908	0.039	-0.035	0.714	0.140	0.110	0.838	0.740	1.193	1.395	0.488	
	Std	1.663	0.551	0.033	0.036	0.262	0.125	0.162	0.338	0.379	0.445	0.655	0.814	
	IRR	76.4%	26.5%	1.2%	22.3%	72.2%	71.6%	24.3%	75.8%	38.9%	46.0%	31.3%	31.5%	
	JRR										61.7%			
BP	Mean	-2.038	0.402	0.032	-0.002	0.665	0.131	0.054	0.095	-0.007	0.277	-0.043	0.305	
	Median	-1.294	0.417	0.033	0.002	0.761	0.116	-0.021	0.054	-0.099	0.212	0.006	0.355	
	Std	1.603	0.192	0.012	0.016	0.211	0.048	0.192	0.186	0.302	0.299	0.574	0.552	
	IRR	100.0%	13.2%	19.6%	0.0%	85.8%	86.7%	21.2%	0.6%	0.0%	0.4%	0.0%	0.0%	
	JRR										9.9%			
C	Mean	-2.170	0.554	0.007	-0.023	0.611	0.269	-0.065	0.347	-0.410	0.578	-0.594	-0.255	
	Median	-2.054	0.567	0.020	-0.031	0.627	0.257	-0.047	0.326	-0.488	0.555	-0.940	-0.453	
	Std	0.634	0.242	0.030	0.023	0.120	0.080	0.139	0.177	0.375	0.240	0.798	0.753	
	IRR	100.0%	23.1%	0.0%	27.5%	99.7%	100.0%	0.9%	33.3%	26.7%	12.0%	12.8%	1.5%	
	JRR										82.9%			

Figure 2.4: Joint significance test for the exogenous variable.

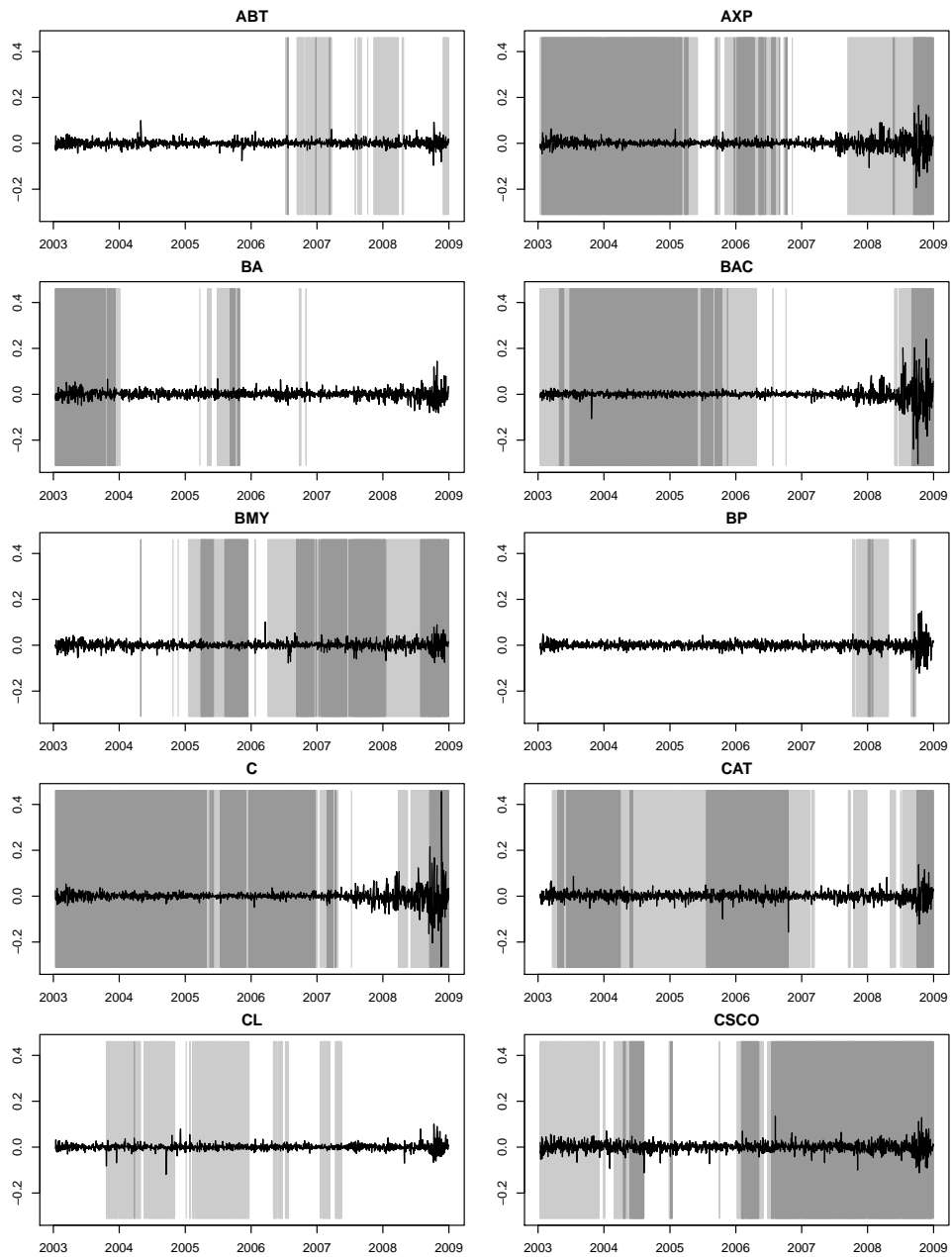


Table 2.3: (continued)

		ω	ω_-	α_-	α_+	β	b_1	b_2	b_3	b_4	b_5	b_6	b_7
CAT	Mean	-4.585	0.040	0.003	-0.020	0.284	0.210	0.381	0.124	0.367	0.647	-0.009	0.836
	Median	-3.11	-0.019	0.012	-0.019	0.424	0.187	0.365	0.107	0.465	0.726	0.546	0.441
	Std	2.805	0.306	0.025	0.015	0.366	0.073	0.129	0.299	0.287	0.898	1.430	1.065
	IRR	100.0%	0.4%	0.0%	0.0%	63.4%	73.3%	63.0%	10.9%	2.2%	23.3%	22.7%	0.0%
	JRR										78.4%		
CL	Mean	-3.126	1.163	0.060	-0.047	0.587	0.134	-0.009	-0.055	0.600	1.594	0.424	2.059
	Median	-2.578	1.203	0.065	-0.038	0.616	0.110	-0.005	-0.122	0.790	1.746	0.466	2.529
	Std	1.294	0.386	0.016	0.039	0.097	0.076	0.092	0.221	0.478	0.509	0.428	1.572
	IRR	99.9%	25.5%	4.7%	10.3%	95.8%	37.2%	0.0%	0.0%	5.4%	57.3%	0.0%	61.0%
	JRR										36.2%		
CSCO	Mean	-3.315	0.113	0.015	0.008	0.398	0.276	0.266	0.341	0.178	-0.017	0.589	0.082
	Median	-1.416	0.130	0.018	0.011	0.709	0.212	0.228	0.341	0.104	0.222	0.398	-0.126
	Std	2.776	0.316	0.016	0.022	0.454	0.145	0.253	0.112	0.209	0.618	0.654	0.730
	IRR	99.9%	15.7%	0.0%	4.7%	63.0%	99.9%	39.9%	30.5%	0.0%	3.7%	5.8%	2.6%
	JRR										72.8%		

is market close while it is market opening for C. Finally, we find that the profile of the ABT, AXP and BA curves is almost sinusoidal. For this type of curve, the interpretation is more complex as there are no clear stylized periods where intraday price is indicative of future daily volatility.

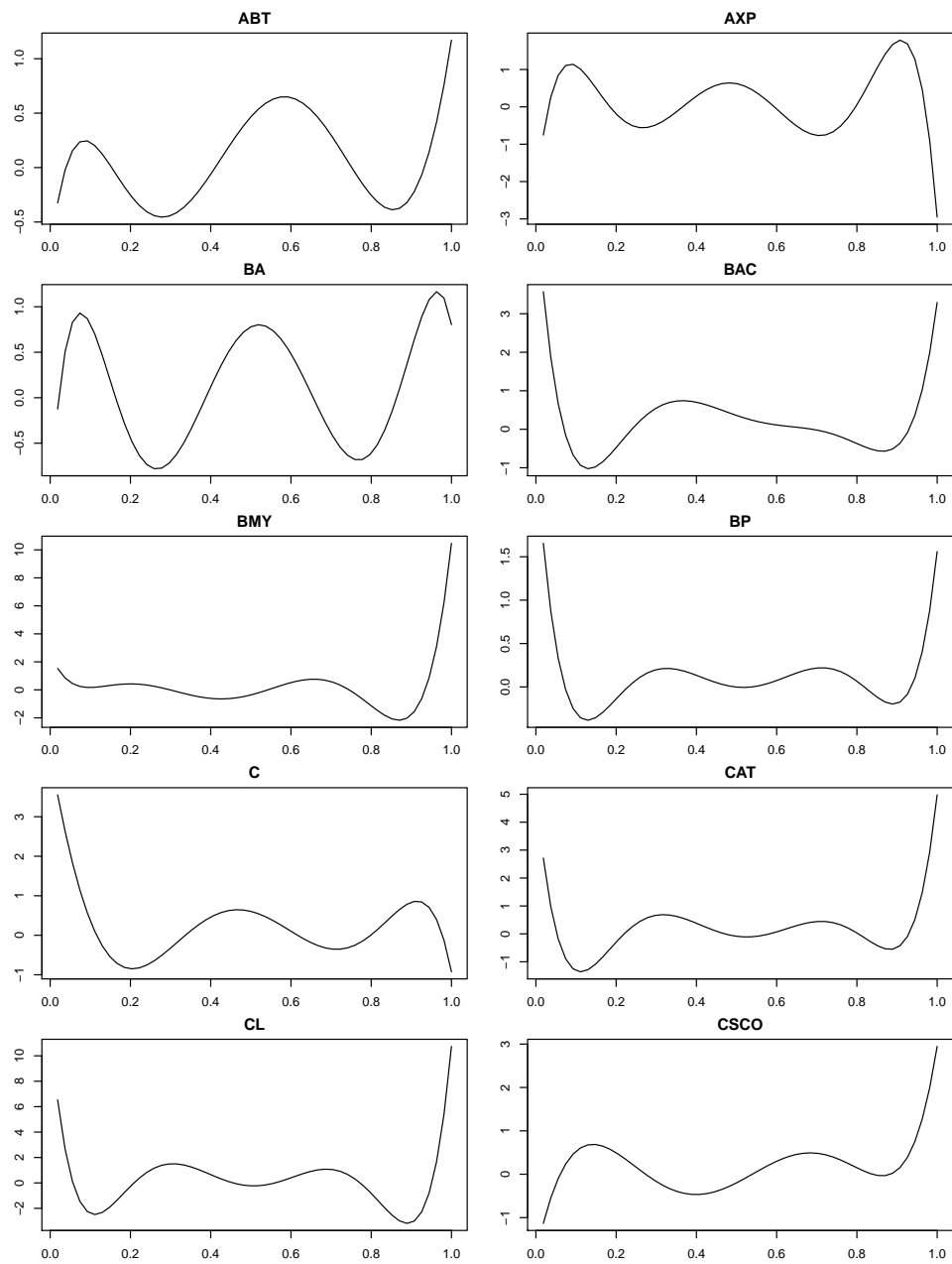
2.5.3 Model comparison

In this section, we consider a horse race between Model (2.12) and some nested competitors from the existing literature. The logic is to examine whether the more complex version of the model, that we call AS-Log-GARCH-fX for Asymmetric log-GARCH model with partially functional covariates model, may offer added value compared to simpler volatility time-series models. We consider three competing models: the Log-GARCH model (without asymmetries and exogenous variables), the AS-Log-GARCH model that accounts for the leverage effect (but does not include exogenous variables), and the AS-Log-GARCH-X model including both a leverage effect and an exogenous scalar variable. We use the daily realized volatility for the exogenous variable defined as the sum of the 5-minutes squared log-returns. In the first part, we apply in-sample portmanteau goodness of fit tests. In a second part, we assess the out-of-sample forecast performance of the four volatility models.

2.5.3.1 Goodness-of-fit

Table 2.4 reports the rejection rates of the null hypothesis of the portmanteau test statistics at the nominal level $\alpha = 5\%$. Except for BMV, CL, and ABT (when m is large), the AS-Log-GARCH-fX model overall passes the portmanteau tests, and generally

Figure 2.5: Estimated intraday coefficient curve $\hat{\delta}(\cdot)$ of each asset.



displays smaller rejection rates compared to the three competitors. Note that the rolling window scheme produces dependent statistics, which may explain an empirical frequency of rejection much smaller than the nominal level.

Table 2.4: Portmanteau test statistics rejection rates (in %) for four models.

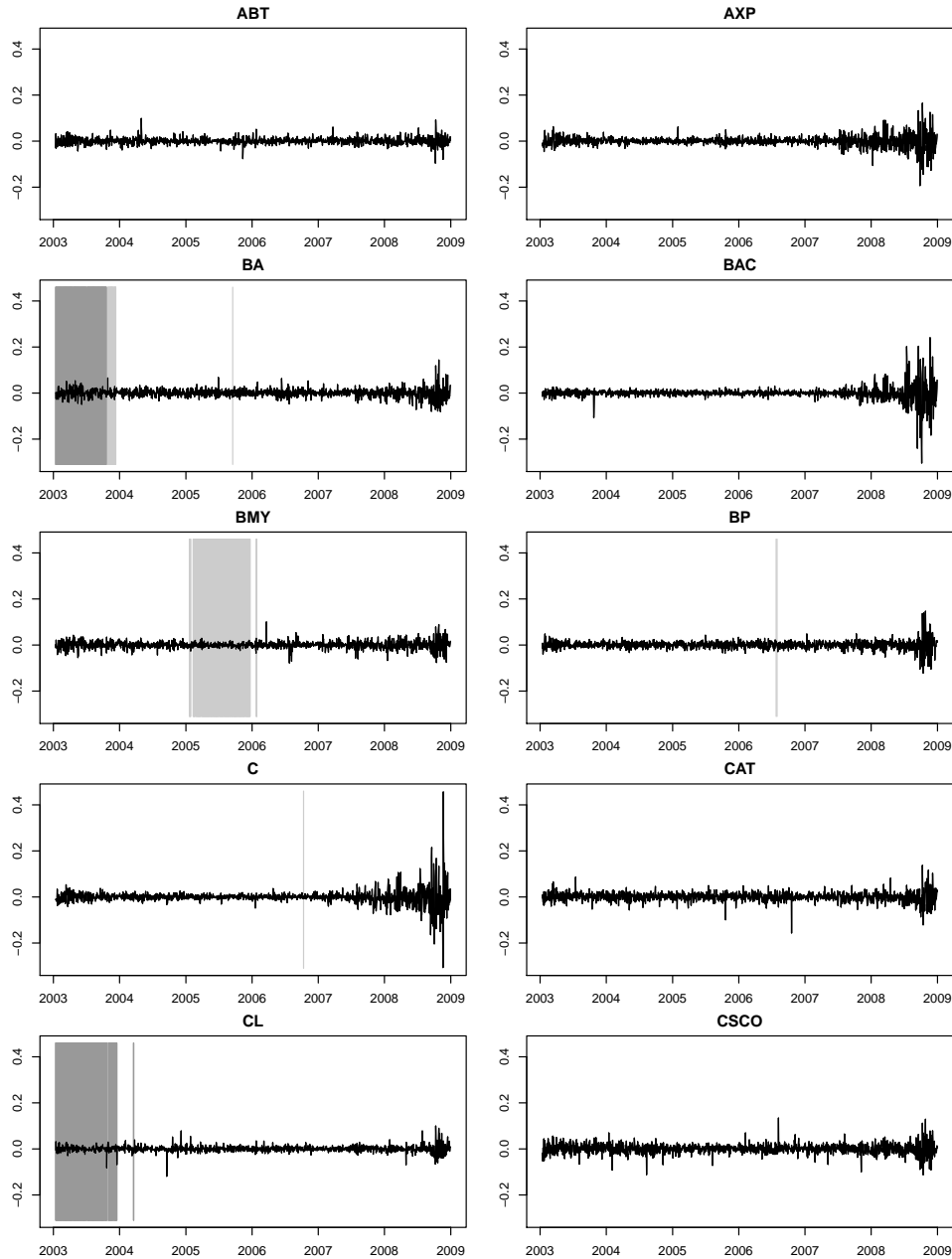
m	6	12	18	24	30	m	6	12	18	24	30
ABT	0.0	0.4	44.3	28.0	26.7	BP	0.1	0.0	0.0	0.4	0.1
	0.1	0.1	56.2	53.8	39.8		1.8	1.3	0.1	7.5	4.9
	0.7	0.3	56.1	40.6	38.8		7.8	16.9	15.8	16.6	15.1
	4.8	1.4	56.6	56.2	56.2		65.4	67.3	69.8	72.6	71.3
AXP	0.0	15.5	10.1	0.0	0.0	C	0.1	0.1	0.1	0.0	0.0
	41.9	24.6	23.4	23.1	23.1		1.7	2.2	2.3	1.8	1.6
	91.9	79.8	65.3	51.6	61.2		8.5	9.7	12.7	9.5	9.1
	88.5	86.5	85.5	84.6	85.3		73.7	82.1	80.1	73.3	70.3
BA	15.1	16.1	11.2	0.0	0.0	CAT	0.0	0.0	0.0	0.0	0.0
	24.5	33.7	12.3	6.5	3.8		0.0	0.0	0.0	0.0	0.0
	33.8	36.6	30.5	12.6	52.8		0.0	0.0	0.0	0.0	0.0
	45.4	49.1	50.7	46.1	53.2		0.0	0.0	0.0	0.7	0.0
BAC	0.0	0.5	0.3	3.6	13.3	CL	15.8	15.5	13.6	34.4	36.7
	0.0	8.8	0.0	9.1	12.0		0.0	1.1	0.0	12.0	10.0
	0.1	2.3	1.1	1.5	5.8		47.7	25.6	14.3	30.6	19.2
	20.2	15.3	8.8	9.1	12.2		59.5	21.4	13.6	37.4	26.7
BMY	15.1	59.5	57.1	52.89	52.0	CSCO	0.0	0.0	0.0	0.0	0.0
	0.0	53.3	53.0	52.82	52.6		0.0	0.0	0.0	0.0	0.0
	0.3	78.0	53.0	52.82	51.0		0.1	0.3	0.0	0.0	0.0
	37.7	93.5	93.3	92.34	91.8		0.6	0.5	0.5	0.5	0.3

For each series, the 1st row corresponds to AS-Log-GARCH-fX, the 2nd to AS-Log-GARCH-X, the 3rd to AS-Log-GARCH, and the 4th to Log-GARCH

Figure 2.6 reports the rejection dates of the null hypothesis of non-autocorrelated residuals with $m = 6$ lags. To highlight the dates of rejection, we consider light and dark shaded bars representing rejection dates at 5% and 1% significance levels, respectively. We find that the rejection dates are sparse. It is worth noting however that we observe a cluster of rejection at the beginning of the sample (2003-2004) for BA and CL and in

the middle of the sample (2005-2006) for *BMY*. Importantly, our model is never rejected at the end of the sample, that is during the financial crisis. Similar conclusions hold for larger values of *m*.

Figure 2.6: Rejection dates of goodness-of-fit for $m = 6$ (significant levels of 5% (light shaded area) and 1% (dark shaded area)).



2.5.3.2 Out-of-sample forecasting

We now turn to the out-of-sample (OOS) forecast performance evaluations. We estimate the model parameters using the rolling-window estimation scheme used in Section 2.5.2. Then, we compute one-step ahead forecasts of the conditional variance. As a re-

sult, for each stock we obtain a series of 1488 forecasts spanning from January 2003 to December 2008.

We apply the model confidence set (MCS) procedure of Hansen et al. (2011) to compare the volatility forecasts issued from the AS-Log-GARCH-fX model and the three competitors. At a given confidence level, the MCS procedure makes it possible to identify buckets of models where OOS forecast performance is equivalent within the same bucket but is statistically different between distinct buckets. The bucket including the best outperforming models is called the superior set of models (SSM). In the sequel, we consider a 75%-confidence level for the MCS procedure. We apply the MCS procedure on forecasting series of size $s = 500$ using a rolling-window scheme. We then compute the procedure 989 times and report the corresponding result at the end of the forecasting series (for graphics). Formally, the result of the MCS procedure displayed at time t is produced using the forecasts from time $t - s + 1$ to time t .

Table 2.5 reports the frequency at which a given model enters into the SSM. The AS-Log-GARCH-fX model belongs to the SSM for six assets out of ten assets, i.e., ABT, AXP, BA, BMY, C, CSCO, regardless of the considered period. Furthermore, the AS-Log-GARCH-fX model provides the highest rate into the SSM for BAC compared to the competitors although the model is not always into the best bucket. Finally, for the remaining stocks, i.e., BP, CAT, CL, the AS-Log-GARCH-fX model is dominated by the AS-Log-GARCH-X model because the latter displays the highest frequency rates into the SSM, suggesting for these stocks that a summary of the day (instead of the entire curve) is self-sufficient to forecast volatility.

Figure 2.7 highlights the dates where the AS-Log-GARCH-fX model enters the SSM with at least one competitor (light shaded area) and where it enters the SSM alone (dark shaded area). Two remarks stand out. First, at the end of the sample period, the AS-Log-GARCH-fX is ranked among the SSM for the ten tickers suggesting that the model is generally competitive to predict volatility in times of crisis. Second, the AS-Log-GARCH-fX enters the SSM alone regularly. For instance, the AS-Log-GARCH-fX model enters the SSM alone in more than 25% of times for five assets and accordingly outperforms significantly the three competitors in terms of OOS forecasting performance.

2.6 Conclusion

In this article, we proposed a new volatility model taking advantage of the functional information extracted from intraday data, together with the usual low frequency (daily) information. The choice of a log-GARCH formulation allowed to introduce such extra information without facing problems due to positivity constraints, which are particularly important in the case of functional data. On the other hand, the absence of a lower bound for the volatility entailed challenging difficulties in the treatment of initial values

Figure 2.7: Evolution of the AS-Log-GARCH-fX model in SSM of the MCS procedure (AS-Log-GARCH-fX model alone (dark shaded area) and with other models (light shaded area)).

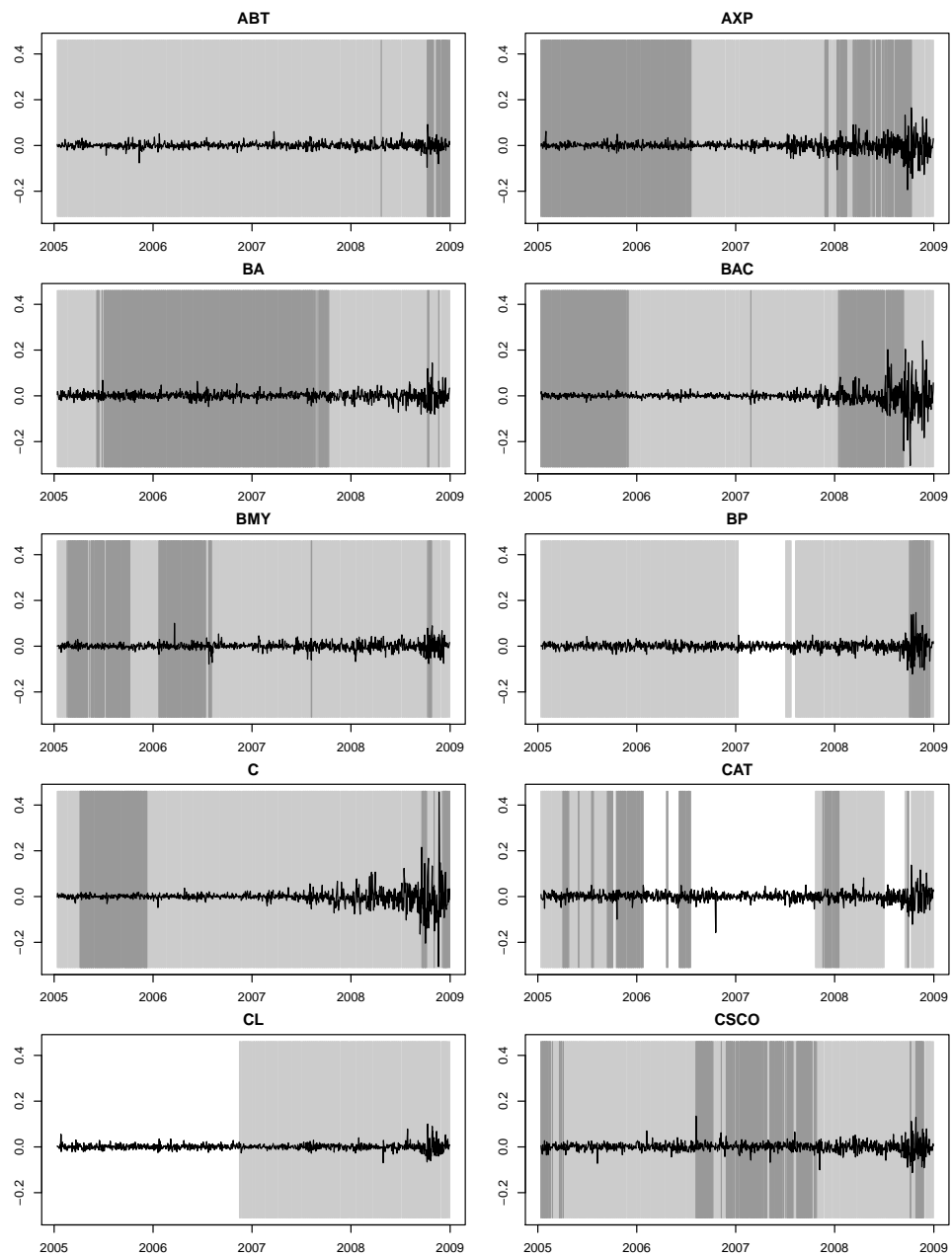


Table 2.5: SSM of the MCS procedure.

	AS-Log-GARCH-fX	AS-Log-GARCH-X	AS-Log-GARCH	Log-GARCH
ABT	100.0%	87.6%	90.3%	94.8%
AXP	100.0%	32.8%	39.7%	42.7%
BA	100.0%	37.9%	25.3%	36.1%
BAC	99.7%	49.2%	25.9%	61.4%
BMY	100.0%	69.0%	69.9%	67.6%
BP	86.6%	93.6%	72.7%	63.2%
C	100.0%	37.9%	72.9%	74.1%
CAT	34.8%	95.6%	59.8%	50.9%
CL	53.2%	100.0%	44.3%	43.4%
CSCO	100.0%	32.5%	59.5%	68.3%

for the derivation of asymptotic properties of the QML estimator. Despite these technical difficulties, the probability structure and the asymptotic theory of estimation for the proposed formulation have been established, and the model is simple to use and implement. Confronting the model to real financial data, we found that the functional data are more informative than the commonly used summaries (such as realized volatility) for volatility prediction. Further issues concern the use of this structure for financial risk management, or the multivariate extensions.

2.7 Appendix

2.7.1 Appendix A: Proofs

Proof of Proposition 1. Noticing that the signs of ϵ_t and η_t coincide, and letting $a_t = \beta + \alpha_+ \mathbb{1}_{\{\eta_{t-1} > 0\}} + \alpha_- \mathbb{1}_{\{\eta_{t-1} < 0\}}$, $b_t = \omega + \omega_- \mathbb{1}_{\{\eta_{t-1} < 0\}} + (\alpha_+ \mathbb{1}_{\{\eta_{t-1} > 0\}} + \alpha_- \mathbb{1}_{\{\eta_{t-1} < 0\}}) \ln \eta_{t-1}^2 + \boldsymbol{\pi}' \mathbf{x}_{t-1} + \int_0^1 \delta(u) X_{t-1}(u) du$, the log-volatility equation can be rewritten as a stochastic recurrence equation of the form

$$\ln \sigma_t^2 = a_t \ln \sigma_{t-1}^2 + b_t. \quad (2.13)$$

Using Lemma 4.1 in Francq and Zakoïan (2019), the existence of a unique strictly stationary, ergodic and nonanticipative solution holds provided that $E \ln^+ |a_t| < \infty$, $E \ln^+ |b_t| < \infty$ and $E \ln |a_t| < 0$. These conditions are met under the assumptions of the proposition. From a strictly stationary solution $(\ln \sigma_t^2)$ of (2.13), we easily deduce a strictly stationary solution (ϵ_t) of (2.2). \blacksquare

Proof of Proposition 2. First note that the moment conditions of the proposition entail (2.3), while (2.5) entails (2.4) by Jensen's inequality (i.e. the condition $E |a_t|^m < 1$ entails the condition $E \ln |a_t| = m^{-1} E \ln |a_t|^m < 0$). Thus the strict stationary solution exists and Equation (2.13) can be expanded as

$$\ln \sigma_t^2 = b_t + \sum_{i=0}^{\infty} \left(\prod_{j=0}^i a_{t-j} \right) b_{t-i-1}.$$

The moment conditions in the proposition ensure that $\|b_0\|_m < \infty$, while (2.5) is equivalent to $\|a_0\|_m < 1$. It follows that, using the independence,

$$\|\ln \sigma_t^2\|_m \leq \|b_0\|_m + \sum_{j=1}^{\infty} \{\|a_0\|_m\}^j \|b_0\|_m < \infty.$$

The conclusion follows. \blacksquare

Proof of Proposition 3. For any matrix $\mathbf{M} = (M_{ij})$, let $\text{Abs}(\mathbf{M}) = (|M_{ij}|)$, and, if \mathbf{M} is a square matrix, let $\rho(\mathbf{M})$ denote the spectral radius of \mathbf{M} .

We first note that the conditions of the proposition ensure the existence of a strictly stationary solution. Indeed, the latter requires that the infinite sum

$$\mathbf{z}_t = \sum_{\ell=0}^{\infty} \mathbf{C}_t \dots \mathbf{C}_{t-\ell} \mathbf{b}_{t-\ell-1} + \mathbf{b}_t \quad (2.14)$$

exists a.s., which holds because

$$\left\| \sum_{\ell=0}^{\infty} \mathbf{C}_t \dots \mathbf{C}_{t-\ell} \mathbf{b}_{t-\ell-1} + \mathbf{b}_t \right\| \leq \sum_{\ell=0}^{\infty} \|\mathbf{C}^{(\infty)}\|^\ell \|\mathbf{b}_{t-\ell}\| < \infty, \quad a.s. \quad (2.15)$$

where $\mathbf{C}^{(\infty)}$ denotes the (non random) matrix obtained by replacing $\alpha_{it} + \beta_i$ by $|\alpha_{i+} + \beta_i| \vee |\alpha_{i-} + \beta_i| \vee |\beta_i|$ in \mathbf{C}_t , for $i = 1, \dots, p \vee q$. The latter inequality follows from the Cauchy rule, the moment condition (2.3) and the fact that $\sum_{i=1}^{p \vee q} |\alpha_{i+} + \beta_i| \vee |\alpha_{i-} + \beta_i| \vee |\beta_i| < 1$ entails $\rho(\mathbf{C}^{(\infty)}) < 1$ (see for instance Corollary 2.2 in Francq and Zakoïan, 2019).

By (2.14), we have, componentwise

$$\text{Abs}(\mathbf{z}_t) \leq \sum_{\ell \geq 0} \left(\mathbf{C}^{(\infty)} \right)^\ell \text{Abs}(\mathbf{b}_{t-\ell}). \quad (2.16)$$

For any matrix \mathbf{M} , let $\|\mathbf{M}\|_m = (E\|\mathbf{M}\|^m)^{1/m}$ where $\|\mathbf{M}\|$ is the sum of the absolute values of the elements of \mathbf{M} . For any conformable non random matrix \mathbf{A} , it follows that $\|\mathbf{A}\mathbf{M}\|_m \leq \|\mathbf{A}\| \|\mathbf{M}\|_m$. Thus,

$$\|\text{Abs}(\mathbf{z}_t)\|_m \leq \sum_{\ell \geq 0} \left\| \mathbf{C}^{(\infty)} \right\|^\ell \|\text{Abs}(\mathbf{b}_0)\|_m < \infty$$

using again $\rho(\mathbf{C}^{(\infty)}) < 1$ and the moment conditions. ■

Proof of Proposition 4. The strict stationarity under the condition $\gamma(\mathbf{A}) < 0$ follows as in Theorem 1. Under this condition, there exists $k \geq 1$ such that $E(\ln \|\mathbf{A}_k \dots \mathbf{A}_1\|) < 0$. Therefore, we have $\delta := E\|\mathbf{A}_k \dots \mathbf{A}_1\|^s < 1$ for some $s \in (0, 1)$ by Lemma 2.3 in Berkes et al. (2003b). In view of

$$\mathbf{y}_t = \mathbf{c}_t + \sum_{n=0}^{\infty} \mathbf{A}_t \mathbf{A}_{t-1} \dots \mathbf{A}_{t-n} \mathbf{c}_{t-n-1} \quad (2.17)$$

the c_r -inequality and standard arguments (see *e.g.* Corollary 2.3 in Francq and Zakoïan, 2019) entail that $E\|\mathbf{y}_t\|^s < \infty$, provided $E\|\mathbf{c}_t\|^s < \infty$, which holds true under the assumptions of the proposition. The conclusion follows. ■

Proof of Proposition 5. First notice that the existence of a strictly stationary solution is guaranteed. Indeed, (2.15) is satisfied. Note that $|\ln \sigma_t^2| \leq \|\mathbf{z}_t\| = \|\text{Abs}(\mathbf{z}_t)\|$. In view of (2.16) it follows that, as in Francq et al. (2013, proof of Proposition 3.1),

$$\max \{E|\sigma_t^2|^s, E|\sigma_t^2|^{-s}\} \leq E\{\exp(s\|\text{Abs}(\mathbf{z}_1)\|)\} = \sum_{k=0}^{\infty} \frac{s^k \|\text{Abs}(\mathbf{z}_1)\|_k^k}{k!}$$

$$\begin{aligned} &\leq \sum_{k=0}^{\infty} \frac{s^k \|\text{Abs}(\mathbf{b}_0)\|_k^k \left\{ \sum_{\ell \geq 0} \|(\mathbf{C}^{(\infty)})^\ell\| \right\}^k}{k!} \\ &= E \exp \left\{ s |\nu_0| \sum_{\ell \geq 0} \|(\mathbf{C}^{(\infty)})^\ell\| \right\}. \end{aligned}$$

Moreover $|\nu_0| \leq |\omega_0| + \max_{1 \leq i \leq q} (|\alpha_{i+}| \vee |\alpha_{i-}|) \sum_{j=1}^q |\ln \eta_{-j}^2| + \|\boldsymbol{\pi}\| \|\mathbf{x}_{-1}\| + \left| \int_0^1 \delta(u) X_{-1}(u) du \right|$. Letting $\bar{\omega} = |\omega| + \sum_{i=1}^q |\omega_{i-}|$, we have

$$\begin{aligned} \max \{E|\sigma_t^2|^s, E|\sigma_t^2|^{-s}\} &\leq \exp \{s \bar{\omega} c_\infty\} E \left[\left(\prod_{j=1}^q \exp \{s \lambda_\eta |\ln \eta_{-j}^2|\} \right) \times \right. \\ &\quad \left. \exp \{s \lambda_x \|\mathbf{x}_{-1}\|\} \exp \left\{ s \left| \int_0^1 \delta(u) X_{-1}(u) du \right| c_\infty \right\} \right] < \infty, \end{aligned}$$

where the last inequality follows from the Hölder inequality and (2.8). ■

Proof of Theorem 2. First note that **A3** and the Cauchy-Schwarz and Jensen inequalities entail

$$E \left| \sum_{k=1}^{\infty} B_k(\mathbf{b}) A_{t-1,k} \right| \leq \sum_{k=1}^{\infty} E |A_{t-1,k}| \sqrt{\sum_{k=1}^{\infty} B_k^2(\mathbf{b})} < \infty.$$

Therefore **A3** and **A6** ensure the a.s. existence of

$$\ln \sigma_t^2(\boldsymbol{\theta}) = \mathcal{B}_\theta^{-1}(B) \left\{ \omega_t + \sum_{i=1}^q \alpha_{i,t} \ln \epsilon_{t-i}^2 + \boldsymbol{\pi}' \mathbf{x}_{t-1} + \sum_{k=1}^{\infty} B_k(\mathbf{b}) A_{t-1,k} \right\}. \quad (2.18)$$

We can thus set

$$Q_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t=1}^n \ell_t(\boldsymbol{\theta}), \quad \ell_t(\boldsymbol{\theta}) = \frac{\epsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})} + \ln \sigma_t^2(\boldsymbol{\theta}). \quad (2.19)$$

The consistency of $\hat{\boldsymbol{\theta}}_n$ will be obtained by showing the following intermediate results:

- (i) $\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\theta} \in \Theta} |Q_n(\boldsymbol{\theta}) - \tilde{Q}_n(\boldsymbol{\theta})| = 0$ a.s.;
- (ii) if $\sigma_1^2(\boldsymbol{\theta}) = \sigma_1^2(\boldsymbol{\theta}_0)$ a.s. then $\boldsymbol{\theta} = \boldsymbol{\theta}_0$;
- (iii) $E |\ell_t(\boldsymbol{\theta}_0)| < \infty$ and if $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, $E \ell_t(\boldsymbol{\theta}) > E \ell_t(\boldsymbol{\theta}_0)$;
- (iv) any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ has a neighborhood $V(\boldsymbol{\theta})$ such that $\liminf_{n \rightarrow \infty} \inf_{\boldsymbol{\theta}^* \in V(\boldsymbol{\theta})} \tilde{Q}_n(\boldsymbol{\theta}^*) > E \ell_1(\boldsymbol{\theta}_0)$ a.s.

In the sequel, K denotes a generic positive constant, or a positive random variable which is measurable with respect to \mathcal{F}_0 . We thus have

$$\sup_{\boldsymbol{\theta} \in \Theta} |\ln \sigma_t^2(\boldsymbol{\theta}) - \ln \tilde{\sigma}_t^2(\boldsymbol{\theta})| < K, \quad \text{a.s. for } t = q - p + 1, \dots, q. \quad (2.20)$$

In the case $p = q = 1$, we have:

$$\ln \sigma_t^2(\boldsymbol{\theta}) - \ln \tilde{\sigma}_t^2(\boldsymbol{\theta}) = \beta^{t-1} (\ln \sigma_1^2(\boldsymbol{\theta}) - \ln \tilde{\sigma}_1^2(\boldsymbol{\theta})), \quad \forall t \geq 1.$$

In the sequel, $\rho \in (0, 1)$ denotes a generic constant whose exact value is unimportant and may vary from line to line. In the general case, using (2.20) and **A6** one can show that for almost all trajectories,

$$\sup_{\boldsymbol{\theta} \in \Theta} |\ln \sigma_t^2(\boldsymbol{\theta}) - \ln \tilde{\sigma}_t^2(\boldsymbol{\theta})| \leq K \rho^t. \quad (2.21)$$

First, we complete the proof of (i) in the case $p = q = 1$, $\omega_{1-} = 0$ and $\alpha_{1+} = \alpha_{1-} = \alpha$, for which the notation is more explicit. In view of the multiplicative form of the volatility

$$\sigma_t^2(\boldsymbol{\theta}) = e^{\beta^t \ln \sigma_0^2(\boldsymbol{\theta})} \prod_{l=0}^{t-1} e^{\beta^l (\omega + \alpha \mathbf{1}_{\{\epsilon_{t-l-1} \neq 0\}} \ln \epsilon_{t-l-1}^2 + \boldsymbol{\pi}' \mathbf{x}_{t-l-1} + \sum_{k=1}^{\infty} B_k(\mathbf{b}) A_{t-l-1,k})}, \quad (2.22)$$

we have

$$\begin{aligned} & \frac{1}{t} \ln \left| \frac{1}{\sigma_t^2(\boldsymbol{\theta})} - \frac{1}{\tilde{\sigma}_t^2(\boldsymbol{\theta})} \right| \\ &= \frac{-1}{t} \sum_{l=0}^{t-1} \beta^l \left(\omega + \alpha \mathbf{1}_{\{\epsilon_{t-l-1} \neq 0\}} \ln \epsilon_{t-l-1}^2 + \boldsymbol{\pi}' \mathbf{x}_{t-l-1} + \sum_{k=1}^{\infty} B_k(\mathbf{b}) A_{t-l-1,k} \right) \\ & \quad + \frac{1}{t} \ln \left| e^{-\beta^t \ln \sigma_0^2(\boldsymbol{\theta})} - e^{-\beta^t \ln \tilde{\sigma}_0^2(\boldsymbol{\theta})} \right|. \end{aligned}$$

It is known that for a sequence of random variables (Y_t) such that $\sup_t E|Y_t| < \infty$, we have almost surely $t^{-1}Y_t \rightarrow 0$ as $t \rightarrow \infty$ (see exercise 2.13 in Francq and Zakoïan, 2019). According to this property, the first term of the right-hand side of the last equality tends almost surely to zero because it is bounded by a variable of the form $t^{-1}|Y_t|$ with $E|Y_t| = E|Y_1| < \infty$, under **A3**⁸. The second term is equal to

$$t^{-1} \ln \left| \beta^t (\ln \sigma_0^2(\boldsymbol{\theta}) - \ln \tilde{\sigma}_0^2(\boldsymbol{\theta})) e^{-\beta^t x^*} \right|,$$

⁸Note that the variable Y_t is successively equal to $\sum_{l=0}^{\infty} \beta^l \alpha \mathbf{1}_{\{\epsilon_{t-l-1} \neq 0\}} \ln \epsilon_{t-l-1}^2$, $\sum_{l=0}^{\infty} \beta^l \boldsymbol{\pi}' \mathbf{x}_{t-l-1}$ and $\sum_{l=0}^{\infty} \beta^l \sum_{k=1}^{\infty} B_k(\mathbf{b}) A_{t-l-1,k}$.

where x^* is between $\ln \sigma_1^2(\boldsymbol{\theta})$ and $\ln \tilde{\sigma}_1^2(\boldsymbol{\theta})$. This second term tends to $\ln |\beta| < 0$ when $t \rightarrow \infty$. It follows that

$$\sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{1}{\sigma_t^2(\boldsymbol{\theta})} - \frac{1}{\tilde{\sigma}_t^2(\boldsymbol{\theta})} \right| \leq K \rho^t. \quad (2.23)$$

Now consider general values for p and q . By **A2** and the second condition of **A6**, we have $\mathcal{B}_\theta^{-1}(z) = \sum_{i=0}^{\infty} c_i(\boldsymbol{\theta}) z^i$ where $c_i(\boldsymbol{\theta})$ tends to zero at an exponential rate as $i \rightarrow \infty$, uniformly in Θ . We thus have

$$\begin{aligned} \ln \sigma_t^2(\boldsymbol{\theta}) &= \sum_{i=0}^{t-1} \left[c_i(\boldsymbol{\theta}) \left\{ \omega + \boldsymbol{\pi}' \mathbf{x}_{t-i-1} + \sum_{k=1}^{\infty} B_k(\mathbf{b}) A_{t-i-1,k} \right\} + c_{i,\omega}(\boldsymbol{\theta}) \mathbb{1}_{\{\epsilon_{t-i-1} < 0\}} \right. \\ &\quad \left. + c_{i,+}(\boldsymbol{\theta}) \mathbb{1}_{\{\epsilon_{t-i-1} > 0\}} \ln \epsilon_{t-i-1}^2 + c_{i,-}(\boldsymbol{\theta}) \mathbb{1}_{\{\epsilon_{t-i-1} < 0\}} \ln \epsilon_{t-i-1}^2 \right] \\ &\quad + \sum_{j=1}^p c_{t,j}(\boldsymbol{\theta}) \ln \sigma_{1-j}^2(\boldsymbol{\theta}) \end{aligned}$$

with

$$\sup_{\boldsymbol{\theta} \in \Theta} \max \{ |c_i(\boldsymbol{\theta})|, |c_{i,+}(\boldsymbol{\theta})|, |c_{i,-}(\boldsymbol{\theta})|, |c_{i,\omega}(\boldsymbol{\theta})| \} \leq K \rho^i, \quad (2.24)$$

$$\sup_{\boldsymbol{\theta} \in \Theta} \max \{ |c_{t,1}(\boldsymbol{\theta})|, \dots, |c_{t,p}(\boldsymbol{\theta})| \} \leq K \rho^t. \quad (2.25)$$

We then obtain a multiplicative form for $\sigma_t^2(\boldsymbol{\theta})$ which generalizes (2.22), and we deduce that

$$t^{-1} \ln \left| \frac{1}{\sigma_t^2(\boldsymbol{\theta})} - \frac{1}{\tilde{\sigma}_t^2(\boldsymbol{\theta})} \right| = a_1 + a_2,$$

where, in view of (2.24) and already given arguments,

$$\begin{aligned} a_1 &= -t^{-1} \sum_{i=0}^{t-1} \left[c_i(\boldsymbol{\theta}) \left\{ \omega + \boldsymbol{\pi}' \mathbf{x}_{t-i-1} + \sum_{k=1}^{\infty} B_k(\mathbf{b}) A_{t-i-1,k} \right\} + c_{i,\omega}(\boldsymbol{\theta}) \mathbb{1}_{\{\epsilon_{t-i-1} < 0\}} \right. \\ &\quad \left. + c_{i,+}(\boldsymbol{\theta}) \mathbb{1}_{\{\epsilon_{t-i-1} > 0\}} \ln \epsilon_{t-i-1}^2 + c_{i,-}(\boldsymbol{\theta}) \mathbb{1}_{\{\epsilon_{t-i-1} < 0\}} \ln \epsilon_{t-i-1}^2 \right] = o(1) \text{ a.s.} \end{aligned}$$

For x_j^* 's between $\ln \sigma_{1-j}^2(\boldsymbol{\theta})$ and $\ln \tilde{\sigma}_{1-j}^2(\boldsymbol{\theta})$, using (2.25) we have

$$\begin{aligned} a_2 &= t^{-1} \ln \left| e^{-\sum_{j=1}^p c_{t,j}(\boldsymbol{\theta}) \ln \sigma_{1-j}^2(\boldsymbol{\theta})} - e^{-\sum_{j=1}^p c_{t,j}(\boldsymbol{\theta}) \ln \tilde{\sigma}_{1-j}^2(\boldsymbol{\theta})} \right| \\ &= t^{-1} \ln \left| - \sum_{j=1}^p c_{t,j}(\boldsymbol{\theta}) \{ \ln \sigma_{1-j}^2(\boldsymbol{\theta}) - \ln \tilde{\sigma}_{1-j}^2(\boldsymbol{\theta}) \} e^{-\sum_{j=1}^p c_{t,j}(\boldsymbol{\theta}) x_j^*} \right| \\ &= t^{-1} \ln \left| - \sum_{j=1}^p c_{t,j}(\boldsymbol{\theta}) \right| + o(1) \text{ a.s.} \end{aligned}$$

Using again (2.25), it follows that $\limsup_{n \rightarrow \infty} a_2 \leq \ln \rho < 0$. We conclude that (2.23) holds true in the general case. The proof of (i) then easily follows from (2.21) and (2.23).

We now show (ii). Assume that $\sigma_t(\boldsymbol{\theta}) = \sigma_t(\boldsymbol{\theta}_0)$ a.s. and, by stationarity, $\ln \sigma_t(\boldsymbol{\theta}) = \ln \sigma_t(\boldsymbol{\theta}_0)$ for all t . The polynomials $\mathcal{B}_\theta(B)$ and $\mathcal{B}_{\theta_0}(B)$ are invertible using the second part of **A6**. Since $\ln \sigma_t^2(\boldsymbol{\theta}) = \ln \sigma_t^2(\boldsymbol{\theta}_0)$ we have

$$\begin{aligned} & \left(\frac{\mathcal{A}_\theta^+(B)}{\mathcal{B}_\theta(B)} - \frac{\mathcal{A}_{\theta_0}^+(B)}{\mathcal{B}_{\theta_0}(B)} \right) \mathbb{1}_{\{\epsilon_t > 0\}} \ln \epsilon_t^2 + \left(\frac{\mathcal{A}_\theta^-(B)}{\mathcal{B}_\theta(B)} - \frac{\mathcal{A}_{\theta_0}^-(B)}{\mathcal{B}_{\theta_0}(B)} \right) \mathbb{1}_{\{\epsilon_t < 0\}} \ln \epsilon_t^2 \\ & + \left(\frac{\boldsymbol{\pi}'}{\mathcal{B}_\theta(B)} - \frac{\boldsymbol{\pi}'_0}{\mathcal{B}_{\theta_0}(B)} \right) \mathbf{x}_{t-1} + \sum_{k=1}^{\infty} \left(\frac{B_k(\mathbf{b})}{\mathcal{B}_\theta(B)} - \frac{B_k(\mathbf{b}_0)}{\mathcal{B}_{\theta_0}(B)} \right) A_{t-1,k} \\ & + \left(\frac{\mathcal{W}_\theta(B)}{\mathcal{B}_\theta(B)} - \frac{\mathcal{W}_{\theta_0}(B)}{\mathcal{B}_{\theta_0}(B)} \right) \mathbb{1}_{\{\epsilon_t < 0\}} = \frac{\omega_0}{\mathcal{B}_{\theta_0}(1)} - \frac{\omega}{\mathcal{B}_\theta(1)} \quad \text{a.s.} \end{aligned} \quad (2.26)$$

If

$$\frac{\mathcal{A}_\theta^+(B)}{\mathcal{B}_\theta(B)} \neq \frac{\mathcal{A}_{\theta_0}^+(B)}{\mathcal{B}_{\theta_0}(B)} \quad \text{or} \quad \frac{\mathcal{A}_\theta^-(B)}{\mathcal{B}_\theta(B)} \neq \frac{\mathcal{A}_{\theta_0}^-(B)}{\mathcal{B}_{\theta_0}(B)} \quad \text{or} \quad \frac{\mathcal{W}_\theta(B)}{\mathcal{B}_\theta(B)} \neq \frac{\mathcal{W}_{\theta_0}(B)}{\mathcal{B}_{\theta_0}(B)}, \quad (2.27)$$

then there exist $c_{i,+} \neq 0$ or $c_{i,-} \neq 0$ or $c_i \neq 0$ such that

$$\sum_{i=i_0}^{\infty} [c_{i,+} \mathbb{1}_{\{\epsilon_{t-i} > 0\}} \ln \epsilon_{t-i}^2 + c_{i,-} \mathbb{1}_{\{\epsilon_{t-i} < 0\}} \ln \epsilon_{t-i}^2 + c_i \mathbb{1}_{\{\epsilon_{t-i} < 0\}}] \in \mathcal{F}_{t-1}^{\mathbf{x}, X} \quad (2.28)$$

where $\mathcal{F}_{t-1}^{\mathbf{x}, X}$ denotes the sigma-field generated by $\{\mathbf{x}_{t-k}, X_{t-k}, k > 0\}$. Since $\mathbb{1}_{\{\epsilon_t > 0\}} \ln \epsilon_t^2 = \mathbb{1}_{\{\eta_t > 0\}} (\ln \eta_t^2 + \ln \sigma_t^2)$ and $\mathbb{1}_{\{\epsilon_t < 0\}} \ln \epsilon_t^2 = \mathbb{1}_{\{\eta_t < 0\}} (\ln \eta_t^2 + \ln \sigma_t^2)$, there exist $(c_+, c_-, c)' \in \mathbb{R}^3 \setminus (0, 0, 0)$ such that: $c_+ \mathbb{1}_{\{\eta_{t-i_0} > 0\}} (\ln \eta_{t-i_0}^2 + \ln \sigma_{t-i_0}^2) + c_- \mathbb{1}_{\{\eta_{t-i_0} < 0\}} (\ln \eta_{t-i_0}^2 + \ln \sigma_{t-i_0}^2) + c \mathbb{1}_{\{\eta_{t-i_0} < 0\}} = c_{t,i_0}$ where c_{t,i_0} is a measurable function of $\{\eta_{t-k-i_0}, \mathbf{x}_{t-k}, X_{t-k}, k > 0\}$. This equation is equivalent to the two equations $(c_+ (\ln \eta_{t-i_0}^2 + \ln \sigma_{t-i_0}^2) - c_{t,i_0}) \mathbb{1}_{\{\eta_{t-i_0} > 0\}} = 0$ and $(c_- (\ln \eta_{t-i_0}^2 + \ln \sigma_{t-i_0}^2) - c_{t,i_0} + c) \mathbb{1}_{\{\eta_{t-i_0} < 0\}} = 0$. Note that if an equation of the form $a \ln x^2 \mathbb{1}_{\{x > 0\}} + b \mathbb{1}_{\{x > 0\}} = 0$ admits two positive solutions then $a = 0$ and $b = 0$. This result and **A4** imply that $c_+ = 0$. Similarly, we obtain that $c_- = 0$ and $c = 0$, which leads to a contradiction. We conclude that (2.27) cannot hold true, and by **A5**, we obtain $\mathcal{A}_\theta^+(B) = \mathcal{A}_{\theta_0}^+(B)$, $\mathcal{A}_\theta^-(B) = \mathcal{A}_{\theta_0}^-(B)$, $\mathcal{W}_\theta(B) = \mathcal{W}_{\theta_0}(B)$ and $\mathcal{B}_\theta(B) = \mathcal{B}_{\theta_0}(B)$. Then (2.26) becomes

$$(\boldsymbol{\pi}' - \boldsymbol{\pi}'_0) \mathbf{x}_{t-1} + \sum_{k=1}^{\infty} (B_k(\mathbf{b}) - B_k(\mathbf{b}_0)) A_{t-1,k} \in \mathcal{F}_{t-1}^{\mathbf{x}, X}$$

which entails $\boldsymbol{\pi}' = \boldsymbol{\pi}'_0$ and $B_k(\mathbf{b}) = B_k(\mathbf{b}_0)$ for all k , under **A7**, and thus $\mathbf{b} = \mathbf{b}_0$ under **A8**. We then obtain $\omega = \omega_0$. Hence (ii) is proved.

Finally, the proofs of (iii) and (iv) are identical to those given by Francq and Zakoian (2004) for the standard GARCH. By a standard compactness argument, we complete the proof of the Theorem 2. \blacksquare

Proof of Theorem 3. Denote by θ_i the i -th element of $\boldsymbol{\theta}$. Taylor expansions give

$$\frac{\partial}{\partial \theta_i} Q_n(\hat{\boldsymbol{\theta}}_n) - \frac{\partial}{\partial \theta_i} Q_n(\boldsymbol{\theta}_0) = \frac{\partial^2}{\partial \theta_i \partial \boldsymbol{\theta}'} Q_n(\tilde{\boldsymbol{\theta}}_{n,i}) (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), \quad 1 \leq i \leq d,$$

where the $\tilde{\boldsymbol{\theta}}_{n,i}$'s are such that $\|\tilde{\boldsymbol{\theta}}_{n,i} - \boldsymbol{\theta}_0\| \leq \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|$. As in Section 5 of Bardet and Wintenberger (2009), the asymptotic normality is obtained by showing, as $n \rightarrow \infty$:

- (i) $\sqrt{n} \frac{\partial}{\partial \boldsymbol{\theta}} Q_n(\boldsymbol{\theta}_0) \rightarrow \mathcal{N}(\mathbf{0}, (\kappa_\eta - 1) \mathbf{J})$ with \mathbf{J} is invertible,
- (ii) $\left\| \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} Q_n(\tilde{\boldsymbol{\theta}}_n) - \mathbf{J} \right\| \rightarrow 0$ a.s. for any sequence $(\tilde{\boldsymbol{\theta}}_n)$ converging a.s. to $\boldsymbol{\theta}_0$,
- (iii) $\sqrt{n} \left\| \frac{\partial}{\partial \boldsymbol{\theta}} \tilde{Q}_n(\hat{\boldsymbol{\theta}}_n) - \frac{\partial}{\partial \boldsymbol{\theta}} Q_n(\hat{\boldsymbol{\theta}}_n) \right\| \rightarrow 0$ a.s.

We begin to show the existence of \mathbf{J} . First note that, in view of (2.18), **A2**, **A3** and **A6**, the process $(\ln \sigma_t^2(\boldsymbol{\theta}))_t$ is stationary and ergodic, and satisfies $E \sup_{\boldsymbol{\theta} \in \Theta} |\ln \sigma_t^2(\boldsymbol{\theta})| < \infty$. Note also that the Cramer condition (2.10) entails

$$E \mathbf{1}_{\{\epsilon_t \neq 0\}} |\ln \epsilon_t^2|^k + E \|\mathbf{x}_t\|^k + E \left(\sum_{i=1}^{\infty} A_{t,i}^2 \right)^k < \infty \text{ for all } k.$$

Moreover $\left\| \sum_{k=1}^{\infty} B_k(\mathbf{b}) A_{t-1,k} \right\|_{2+\nu} \leq \sqrt{\sum_{k=1}^{\infty} B_k^2(\mathbf{b}) \sum_{k=1}^{\infty} \|A_{t-1,k}\|_{2+\nu}^2}$. Under **A11**, we thus have $E \sup_{\boldsymbol{\theta} \in \Theta} |\ln \sigma_t^2(\boldsymbol{\theta})|^{2+\nu} < \infty$. By the same arguments, the random vector $\frac{\partial}{\partial \boldsymbol{\theta}} \ln \sigma_t^2(\boldsymbol{\theta})$ is the stationary solution of the equation

$$\frac{\partial}{\partial \boldsymbol{\theta}} \ln \sigma_t^2(\boldsymbol{\theta}) = \sum_{j=1}^p \beta_j \frac{\partial}{\partial \boldsymbol{\theta}} \ln \sigma_{t-j}^2(\boldsymbol{\theta}) + \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}), \quad (2.29)$$

with

$$\begin{aligned} \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}) &= \left(1, \mathbf{1}_{t-1,q}^-, \boldsymbol{\epsilon}_{t-1,q}^-, \boldsymbol{\epsilon}_{t-1,q}^+, \boldsymbol{\sigma}'_{t-1,p}(\boldsymbol{\theta}), \mathbf{x}'_{t-1}, \sum_{k=1}^{\infty} \frac{\partial}{\partial \mathbf{b}'} B_k(\mathbf{b}) A_{t-1,k} \right)', \\ \mathbf{1}_{t-1,q}^- &= (\mathbf{1}_{\epsilon_{t-1} < 0}, \dots, \mathbf{1}_{\epsilon_{t-q} < 0})', \quad \boldsymbol{\epsilon}_{t-1,q}^- = (\mathbf{1}_{\epsilon_{t-1} < 0} \ln \epsilon_{t-1}^2, \dots, \mathbf{1}_{\epsilon_{t-q} < 0} \ln \epsilon_{t-q}^2)', \\ \boldsymbol{\epsilon}_{t-1,q}^+ &= (\mathbf{1}_{\epsilon_{t-1} > 0} \ln \epsilon_{t-1}^2, \dots, \mathbf{1}_{\epsilon_{t-q} > 0} \ln \epsilon_{t-q}^2)', \quad \boldsymbol{\sigma}'_{t-1,p}(\boldsymbol{\theta}) = (\ln \sigma_{t-1}^2(\boldsymbol{\theta}), \dots, \ln \sigma_{t-p}^2(\boldsymbol{\theta}))'. \end{aligned}$$

Under **A11** and **A12**, the Hölder inequality implies

$$\left\| \sum_{k=1}^{\infty} \frac{\partial}{\partial \mathbf{b}} B_k(\mathbf{b}) A_{t-1,k} \right\|_{2+\nu} \leq \sqrt{\sum_{k=1}^{\infty} \left\| \frac{\partial}{\partial \mathbf{b}} B_k(\mathbf{b}) \right\|^2 \sum_{k=1}^{\infty} \|A_{t-1,k}\|_{2+\nu}^2} < \infty.$$

Under **A11**, the other terms of the vector $\Upsilon_t(\boldsymbol{\theta})$ also admit moments of order $2 + \nu$ uniformly in Θ . We thus have

$$E \sup_{\boldsymbol{\theta} \in \Theta} \left\| \frac{\partial}{\partial \boldsymbol{\theta}} \ln \sigma_t^2(\boldsymbol{\theta}) \right\|^{2+\nu} < \infty, \quad (2.30)$$

which entails the existence of **J**.

We need to establish the existence of other moments. Differentiating (2.29) and using the previous arguments, it can be seen that the processes $(\partial^2 \ln \sigma_t^2(\boldsymbol{\theta}) / \partial \theta_i \partial \theta_j)_t$ are stationary and ergodic, with

$$E \sup_{\boldsymbol{\theta} \in \Theta} \left\| \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \ln \sigma_t^2(\boldsymbol{\theta}) \right\|^{2+\nu} < \infty. \quad (2.31)$$

From (2.18), we can write

$$\ln \sigma_t^2(\boldsymbol{\theta}_0) - \ln \sigma_t^2(\boldsymbol{\theta}) = \mathcal{B}_{\boldsymbol{\theta}}^{-1}(B) u_{t-1}(\boldsymbol{\theta}). \quad (2.32)$$

For instance, in the case $p = q = 1$ we have

$$\begin{aligned} u_t(\boldsymbol{\theta}) = & (\beta_0 - \beta) \ln \sigma_t^2(\boldsymbol{\theta}_0) + \omega_0 - \omega + (\omega_{-,0} - \omega_-) \mathbf{1}_{\epsilon_t < 0} + (\alpha_{-,0} - \alpha_-) \mathbf{1}_{\epsilon_t < 0} \ln \epsilon_t^2 \\ & + (\alpha_{+,0} - \alpha_+) \mathbf{1}_{\epsilon_t > 0} \ln \epsilon_t^2 + (\boldsymbol{\pi}_0 - \boldsymbol{\pi})' \mathbf{x}_t + \int_0^1 (\delta_0 - \delta)(u) X_t(u) du. \end{aligned}$$

In the general case, we have (2.32) with

$$\begin{aligned} |u_t(\boldsymbol{\theta})| \leq & K \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \left\{ 1 + \sum_{i=1}^q (\mathbf{1}_{\eta_{t+1-i} < 0} + \mathbf{1}_{\eta_{t+1-i} \neq 0} |\ln \eta_{t+1-i}^2|) \right. \\ & \left. + \sum_{i=1}^{p \vee q} |\ln \sigma_{t+1-i}^2| + \|\mathbf{x}_t\| + \sqrt{\int_0^1 X_t^2(u) du} \right\}. \end{aligned} \quad (2.33)$$

For $\mathcal{B}_{|\boldsymbol{\theta}|}^{-1}(z) = \sum_{j=0}^{\infty} \pi_j(\boldsymbol{\theta}) z^j$ and $V(\boldsymbol{\theta}_0) \subset \Theta$, let $K_{V(\boldsymbol{\theta}_0)} = \sum_{j=0}^{\infty} \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} |\pi_j(\boldsymbol{\theta})|$. For any stationary sequence (u_t) of random variables of L^k , we have $\|\sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \mathcal{B}_{\boldsymbol{\theta}}^{-1}(B) u_t\|_k \leq K_{V(\boldsymbol{\theta}_0)} \|u_1\|_k$. Therefore, for all $r > 0$, there exists a neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that

$$\begin{aligned} E \exp \left\{ r \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \mathcal{B}_{\boldsymbol{\theta}}^{-1}(B) u_{t-1}(\boldsymbol{\theta}) \right\} &= \left\| \sum_{k=0}^{\infty} \frac{\{r \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \mathcal{B}_{\boldsymbol{\theta}}^{-1}(B) u_{t-1}(\boldsymbol{\theta})\}^k}{k!} \right\|_1 \\ &\leq \sum_{k=0}^{\infty} \frac{\left\{ r \|\sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \mathcal{B}_{\boldsymbol{\theta}}^{-1}(B) \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} |u_{t-1}(\boldsymbol{\theta})|\|_k \right\}^k}{k!} \end{aligned}$$

$$\begin{aligned} &\leq \sum_{k=0}^{\infty} \frac{\{rK_{V(\theta_0)} \sup_{\theta \in V(\theta_0)} \|u_1(\theta)\|_k\}^k}{k!} \\ &\leq E \sum_{k=0}^{\infty} \frac{\{rK_{V(\theta_0)} \sup_{\theta \in V(\theta_0)} \|u_1(\theta)\|\}^k}{k!} = E \exp \left\{ rK_{V(\theta_0)} \sup_{\theta \in V(\theta_0)} \|u_1(\theta)\| \right\} \end{aligned}$$

which, in view of (2.33), is finite under **A11** when $V(\theta_0)$ is small enough. Noting that $E \sup_{\theta \in V(\theta_0)} \left(\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)} \right)^r = E \exp \left\{ r \sup_{\theta \in V(\theta_0)} \log \frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)} \right\}$, we deduce that

$$\text{for all } r > 0, \text{ there exists } V(\theta_0) \text{ such that } E \sup_{\theta \in V(\theta_0)} \left(\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)} \right)^r < \infty. \quad (2.34)$$

Let us show that \mathbf{J} is non singular. Suppose that there exists a real vector $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d)'$ such that $\boldsymbol{\lambda}' \frac{\partial}{\partial \boldsymbol{\theta}} \ln \sigma_t^2(\theta_0) = 0$ a.s. By stationarity, in view of (2.29), this implies $\boldsymbol{\lambda}' \boldsymbol{\Upsilon}_t = 0$ a.s., with $\boldsymbol{\Upsilon}_t = \boldsymbol{\Upsilon}_t(\theta_0)$. First note that $\lambda_i = 0$ for $i = 2, \dots, 3q + 1$, otherwise an equation of the form (2.28) would hold true (with non zero coefficients) and we have already shown this is impossible. Let $\boldsymbol{\lambda}_\pi = (\lambda_{3q+p+2}, \dots, \lambda_{3q+p+r+1})'$ and let $\boldsymbol{\lambda}_b$ be the vector of the M last components of $\boldsymbol{\lambda}$. Note that

$$\boldsymbol{\lambda}'_b \sum_{k=1}^{\infty} \frac{\partial}{\partial \mathbf{b}} B_k(\mathbf{b}_0) A_{t-1,k} = \int_0^1 \delta_c(u) X_{t-1}(u) du, \quad \delta_c(u) = \sum_{k=1}^{\infty} \boldsymbol{\lambda}'_b \frac{\partial}{\partial \mathbf{b}} B_k(\mathbf{b}_0) \varphi_k(u).$$

By **A7** we must have $\boldsymbol{\lambda}_\pi = \mathbf{0}$ and $\delta_c = 0$, that is $\boldsymbol{\lambda}'_b \frac{\partial}{\partial \mathbf{b}} B_k(\mathbf{b}_0) = 0$ for all k . By **A13**, this entails $\boldsymbol{\lambda}_b = \mathbf{0}$. Finally, the other components of $\boldsymbol{\lambda}$ are also zero because $\ln \sigma_t$ cannot be equal to a function of its past values under **A4**. This completes the proof of the invertibility of \mathbf{J} .

Now, using (2.19) we have

$$\frac{\partial}{\partial \boldsymbol{\theta}} Q_n(\theta_0) = \frac{1}{n} \sum_{t=1}^n (1 - \eta_t^2) \frac{\partial}{\partial \boldsymbol{\theta}} \ln \sigma_t^2(\theta_0).$$

We know that η_t and $\ln \sigma_t^2(\theta_0)$ are independent and $E \eta_t^2 = 1$. Therefore, using **A10** and the existence of \mathbf{J} , the Central Limit Theorem for martingale differences (see Billingsley, 1961) completes the proof of (i).

To show (ii), first note that

$$\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} Q_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t=1}^n \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \ell_t(\boldsymbol{\theta}),$$

with

$$\frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \left(1 - \frac{\eta_t^2 \sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \right) \frac{\partial^2 \ln \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} + \frac{\eta_t^2 \sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2(\boldsymbol{\theta})} \frac{\partial \ln \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}.$$

Using in particular (2.31) and (2.34), it follows that

$$E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\| < \infty. \quad (2.35)$$

Moreover, by the ergodic theorem

$$\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} Q_n(\boldsymbol{\theta}_0) \rightarrow \mathbf{J} \quad \text{a.s.}$$

Let $V_k(\boldsymbol{\theta})$ be the ball of centre $\boldsymbol{\theta}$ and radius $1/k$ such that $V_k(\boldsymbol{\theta}_0) \subset V(\boldsymbol{\theta}_0)$. Using again the ergodic theorem, we obtain

$$\begin{aligned} & \sup_{\boldsymbol{\theta} \in V_k(\boldsymbol{\theta}_0)} \left\| \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} Q_n(\boldsymbol{\theta}) - \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} Q_n(\boldsymbol{\theta}_0) \right\| \leq \frac{1}{n} \sum_{t=1}^n \sup_{\boldsymbol{\theta} \in V_k(\boldsymbol{\theta}_0)} \left\| \frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \frac{\partial^2 \ell_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\| \\ & \rightarrow E \sup_{\boldsymbol{\theta} \in V_k(\boldsymbol{\theta}_0)} \left\| \frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \frac{\partial^2 \ell_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\|. \quad \text{a.s.} \end{aligned} \quad (2.36)$$

By the dominated convergence theorem, (2.35) and the continuity of the second derivatives, the expectation in (2.36) is smaller than any arbitrary $\varepsilon > 0$ when k is large enough. This completes the proof of (ii).

From (2.29) and an equivalent representation for $\frac{\partial}{\partial \boldsymbol{\theta}} \ln \tilde{\sigma}_t^2(\boldsymbol{\theta})$, we have

$$\begin{aligned} \frac{\partial \ln \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial \ln \tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \sum_{j=1}^p \beta_j \left(\frac{\partial \ln \sigma_{t-j}^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial \ln \tilde{\sigma}_{t-j}^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \\ &+ \begin{pmatrix} \mathbf{0}_{3q+1} \\ \boldsymbol{\sigma}_{t-1,p}(\boldsymbol{\theta}) - \tilde{\boldsymbol{\sigma}}_{t-1,p}(\boldsymbol{\theta}) \\ \mathbf{0}_{r+M} \end{pmatrix} \end{aligned} \quad (2.37)$$

with $\tilde{\boldsymbol{\sigma}}_{t,p}(\boldsymbol{\theta})$ defined as $\boldsymbol{\sigma}_{t,p}(\boldsymbol{\theta})$. By the second condition of **A6**, we thus have $|\partial \ln \sigma_t^2(\boldsymbol{\theta})/\partial \theta_k - \partial \ln \tilde{\sigma}_t^2(\boldsymbol{\theta})/\partial \theta_k| \leq K \rho^t$ for $k = 1, \dots, 3q+1$ and $k = 3q+p+2, \dots, d$. Now consider the case $k = 3q+1+k'$ with $k' \in \{1, \dots, p\}$. Recalling the notation $\left(1 - \sum_{j=1}^p \beta_j z^j\right)^{-1} = \sum_{i=0}^{\infty} c_i(\boldsymbol{\theta}) z^i$ and noting that the k -th element of the last vector of the right-hand side of (2.37) is $\ln \sigma_{t-k'}^2(\boldsymbol{\theta}) - \ln \tilde{\sigma}_{t-k'}^2(\boldsymbol{\theta})$, we obtain

$$\begin{aligned} \frac{\partial \ln \sigma_t^2(\boldsymbol{\theta})}{\partial \theta_k} - \frac{\partial \ln \tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial \theta_k} &= \sum_{i=0}^{t-1} c_i(\boldsymbol{\theta}) (\ln \sigma_{t-k'-i}^2(\boldsymbol{\theta}) - \ln \tilde{\sigma}_{t-k'-i}^2(\boldsymbol{\theta})) \\ &+ \sum_{j=1}^p c_{t,k,j}(\boldsymbol{\theta}) \left\{ \frac{\partial \ln \sigma_{1-j}^2(\boldsymbol{\theta})}{\partial \theta_k} - \frac{\partial \ln \tilde{\sigma}_{1-j}^2(\boldsymbol{\theta})}{\partial \theta_k} \right\}. \end{aligned}$$

with $c_i(\boldsymbol{\theta})$ satisfying (2.24) and

$$\sup_{\boldsymbol{\theta} \in \Theta} \max \{ |c_{t,k,1}(\boldsymbol{\theta})|, \dots, |c_{t,k,p}(\boldsymbol{\theta})| \} \leq K \rho^t.$$

These relations and (2.21) show that $\sup_{\boldsymbol{\theta} \in \Theta} \left\| \frac{\partial}{\partial \boldsymbol{\theta}} \ln \sigma_t^2(\boldsymbol{\theta}) - \frac{\partial}{\partial \boldsymbol{\theta}} \ln \tilde{\sigma}_t^2(\boldsymbol{\theta}) \right\| \leq K \rho^t$, for almost all trajectories. Point (iii) easily follows and the asymptotic normality is proved. ■

Proof of Proposition 6. Let $\eta_t(\boldsymbol{\theta}) = \epsilon_t/\sigma_t(\boldsymbol{\theta})$ and $\tilde{\eta}_t(\boldsymbol{\theta}) = \epsilon_t/\tilde{\sigma}_t(\boldsymbol{\theta})$. Clearly (2.23) holds when the power 2 is replaced by any other power. We thus have

$$\begin{aligned} \sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{1}{n} \sum_{t=1}^n \eta_t^4(\boldsymbol{\theta}) - \frac{1}{n} \sum_{t=1}^n \tilde{\eta}_t^4(\boldsymbol{\theta}) \right| &\leq \frac{1}{n} \sum_{t=1}^n \epsilon_t^4 \sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{1}{\sigma_t^4(\boldsymbol{\theta})} - \frac{1}{\tilde{\sigma}_t^4(\boldsymbol{\theta})} \right| \\ &\leq \frac{K}{n} \sum_{t=1}^n \epsilon_t^4 \rho^t = o(1) \quad \text{a.s.} \end{aligned} \quad (2.38)$$

Now a Taylor expansion yields

$$\frac{1}{n} \sum_{t=1}^n \eta_t^4(\hat{\boldsymbol{\theta}}_n) = \frac{1}{n} \sum_{t=1}^n \eta_t^4 - \frac{2}{n} \sum_{t=1}^n \eta_t^4 \frac{\sigma_t^4(\boldsymbol{\theta}_0)}{\sigma_t^4(\boldsymbol{\theta}_n)} \frac{\partial \ln \sigma_t^2(\boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta}'} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$$

for some $\boldsymbol{\theta}_n$ between $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$. The strong consistency of $\hat{\boldsymbol{\theta}}_n$, (2.30), (2.34) and **A10** then entail that

$$\frac{1}{n} \sum_{t=1}^n \hat{\eta}_t^4 = \frac{1}{n} \sum_{t=1}^n \eta_t^4 + o(1) \quad \text{a.s.}$$

The strong consistency of $\hat{\kappa}$ now follows from **A10** and the ergodic theorem. The consistency of $\hat{\mathbf{J}}$ is shown by the arguments used to establish (ii) and (iii) in the proof of Theorem 3. ■

Proof of Theorem 4. Introduce the vector $\mathbf{r}_m = (r_1, \dots, r_m)'$ where

$$r_h = n^{-1} \sum_{t=h+1}^n s_t s_{t-h}, \quad \text{with } s_t = \eta_t^2 - 1 \text{ and } 0 < h < n.$$

Let $s_t(\boldsymbol{\theta})$ (respectively $\tilde{s}_t(\boldsymbol{\theta})$) be the random variable obtained by replacing η_t by $\eta_t(\boldsymbol{\theta})$ (respectively $\tilde{\eta}_t(\boldsymbol{\theta})$) in s_t . Let $r_h(\boldsymbol{\theta})$ (respectively $\tilde{r}_h(\boldsymbol{\theta})$) be obtained by replacing η_t by $\eta_t(\boldsymbol{\theta})$ (respectively $\tilde{\eta}_t(\boldsymbol{\theta})$) in r_h . Let the vectors $\mathbf{r}_m(\boldsymbol{\theta}) = (r_1(\boldsymbol{\theta}), \dots, r_m(\boldsymbol{\theta}))'$, $\mathbf{r}_m = \mathbf{r}_m(\boldsymbol{\theta}_0)$ and $\tilde{\mathbf{r}}_m(\boldsymbol{\theta}) = (\tilde{r}_1(\boldsymbol{\theta}), \dots, \tilde{r}_m(\boldsymbol{\theta}))'$. Note that $\hat{\mathbf{r}}_m = \tilde{\mathbf{r}}_m(\hat{\boldsymbol{\theta}}_n)$.

By the arguments used to prove (2.38), it can be shown that the unknown initial values have no asymptotic impact on the statistic $\hat{\mathbf{r}}_m$, in the sense that

$$\sup_{\boldsymbol{\theta} \in \Theta} \sqrt{n} \|\mathbf{r}_m(\boldsymbol{\theta}) - \tilde{\mathbf{r}}_m(\boldsymbol{\theta})\| = o(1), \quad \text{a.s.} \quad (2.39)$$

Now Taylor expansions show that, for $h = 1, \dots, m$

$$\sqrt{n}r_h(\hat{\boldsymbol{\theta}}_n) = \sqrt{n}r_h(\boldsymbol{\theta}_0) + \frac{\partial r_h(\boldsymbol{\theta}_n^*)}{\partial \boldsymbol{\theta}'} \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$$

for some $\boldsymbol{\theta}_n^*$ between $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$. Using (2.30), (2.31) and (2.34), it can be shown that for some neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$

$$E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left| \frac{\partial^2 s_t(\boldsymbol{\theta}) s_{t-h}(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right| < \infty \quad \text{for all } i, j \in \{1, \dots, d\}.$$

Using the almost sure convergence of $\boldsymbol{\theta}_n^*$ to $\boldsymbol{\theta}_0$, the ergodic theorem and standard arguments entail

$$\frac{\partial r_h(\boldsymbol{\theta}_n^*)}{\partial \boldsymbol{\theta}} \rightarrow E \frac{\partial s_t s_{t-h}}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0) = -E \left\{ s_{t-h} \frac{\partial \ln \sigma_t^2(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\}.$$

We have shown that

$$\sqrt{n}\hat{\mathbf{r}}_m = \sqrt{n}\mathbf{r}_m - \mathbf{K}_m \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o(1) \quad \text{a.s.}, \quad (2.40)$$

where

$$\mathbf{K}_m = E \left\{ \mathbf{s}_{t-1:t-m} \frac{\partial \ln \sigma_t^2(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \right\}, \quad \mathbf{s}_{t-1:t-m} = (s_{t-1}, \dots, s_{t-m})'.$$

We now derive the asymptotic distribution of $\sqrt{n}(\mathbf{r}_m, \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$. Note that

$$\sqrt{n}\mathbf{r}_m = \frac{1}{\sqrt{n}} \sum_{t=1}^n s_t \mathbf{s}_{t-1:t-m} + o(1) \xrightarrow{d} \mathcal{N} \{ \mathbf{0}, (\kappa_\eta - 1)^2 \mathbf{I}_m \}.$$

We also have seen in the proof of Theorem 3 that

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -\mathbf{J}^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^n (1 - \eta_t^2) \frac{\partial \ln \sigma_t^2}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0) + o_P(1).$$

The central limit theorem applied to the martingale difference

$$\left\{ \left(s_t \frac{\partial \ln \sigma_t^2}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}_0), s_t \mathbf{s}'_{t-1:t-m} \right)'; \sigma(\eta_u, u \leq t) \right\}$$

then shows that

$$\sqrt{n} \begin{pmatrix} \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \\ \mathbf{r}_m \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{t=1}^n s_t \begin{pmatrix} \mathbf{J}^{-1} \frac{\partial \ln \sigma_t^2}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0) \\ \mathbf{s}_{t-1:t-m} \end{pmatrix} + o_P(1)$$

$$\xrightarrow{d} \mathcal{N} \left\{ \mathbf{0}, \begin{pmatrix} (\kappa_\eta - 1)\mathbf{J}^{-1} & (\kappa_\eta - 1)\mathbf{J}^{-1}\mathbf{K}'_m \\ (\kappa_\eta - 1)\mathbf{K}_m\mathbf{J}^{-1} & (\kappa_\eta - 1)^2\mathbf{I}_m \end{pmatrix} \right\}. \quad (2.41)$$

By (2.40) and (2.41), we obtain

$$\sqrt{n}\hat{\mathbf{r}}_m \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{D}), \quad \mathbf{D} = (\kappa_\eta - 1)^2\mathbf{I}_m - (\kappa_\eta - 1)\mathbf{K}_m\mathbf{J}^{-1}\mathbf{K}'_m.$$

It can be shown that $\hat{\mathbf{D}} \rightarrow \mathbf{D}$ almost surely. It remains to show that \mathbf{D} is invertible. Note that

$$\mathbf{D} = (\kappa_\eta - 1)\mathbf{E}\mathbf{V}_t\mathbf{V}'_t, \quad \mathbf{V}_t = \mathbf{s}_{t-1:t-m} - \mathbf{K}_m\mathbf{J}^{-1}\frac{\partial}{\partial\boldsymbol{\theta}}\ln\sigma_t^2(\boldsymbol{\theta}_0).$$

If this matrix were singular then there would exist $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)'$ such that $\boldsymbol{\lambda} \neq \mathbf{0}$ and

$$\boldsymbol{\lambda}'\mathbf{V}_t = \boldsymbol{\lambda}'\mathbf{s}_{t-1:t-m} - \sum_{i=0}^{\infty} \boldsymbol{\mu}'_i\Upsilon_{t-i} = \mathbf{0} \quad \text{a.s.}, \quad (2.42)$$

with $\boldsymbol{\mu}'_i = c_i(\boldsymbol{\theta}_0)\boldsymbol{\lambda}'\mathbf{K}_m\mathbf{J}^{-1}$, using (2.29). Let $s \leq m$ such that $\lambda_s \neq 0$ and $\lambda_h = 0$ if $h < s$. Denote by $R_{t,s}$ a generic random variable which is $\mathcal{F}_{t-1, t-s-1}^{(x, X), \eta}$ -mesurable. Equations (2.42) and (2.29) entail

$$\eta_{t-s}^2 + \sum_{h=1}^s a_h \mathbb{1}_{\eta_{t-h} < 0} + b_h^- \mathbb{1}_{\eta_{t-h} < 0} \ln \epsilon_{t-h}^2 + b_h^+ \mathbb{1}_{\eta_{t-h} > 0} \ln \epsilon_{t-h}^2 + c_h \ln \sigma_{t-h}^2 = R_{t,s}. \quad (2.43)$$

If $s \geq 2$, (2.43) entails

$$\mathbb{1}_{\eta_{t-1} > 0} \{b_1^+ \ln \eta_{t-1}^2 + R_{t,1}\} = 0 \quad \text{and} \quad \mathbb{1}_{\eta_{t-1} < 0} \{b_1^- \ln \eta_{t-1}^2 + R_{t,1}\} = 0 \quad \text{a.s.}$$

which implies $b_1^+ = b_1^- = 0$ by **A4**. By the same argument, it can be seen that $a_h = b_h^- = b_h^+ = c_{h-1} = 0$ for $h = 1, \dots, s-1$. From (2.43) we then obtain $\mathbb{1}_{\eta_{t-s} > 0} \{\eta_{t-s}^2 + b_s^+ \ln \eta_{t-s}^2 + R_{t,s}\} = 0$ and $\mathbb{1}_{\eta_{t-s} < 0} \{\eta_{t-s}^2 + b_s^- \ln \eta_{t-s}^2 + R_{t,s}\} = 0$ a.s. which is impossible under **A4***. Therefore \mathbf{D} is invertible and the conclusion follows. \blacksquare

2.7.2 Appendix B: Algorithm for rational representation of the functional parameter

Let $f(x) = \sum_{k=1}^{\infty} \langle f, \varphi_k \rangle \varphi_k(x) \in H$. We want to approximate optimally this function, in the $L^2[0, 1]$ sense, by

$$\hat{f}_M(x; \mathbf{b}) = \sum_{k=1}^{\infty} B_k(\mathbf{b})\varphi_k(x)$$

where $\mathbf{b} = (b_1, \phi_1, \dots, \phi_{p_1}, \psi_0, \dots, \psi_{q_1})'$ is a vector of M parameters such that

$$B_1(\mathbf{b}) = b_1 = \int_0^1 f(x)dx, \quad \sum_{k=2}^{\infty} B_k(\mathbf{b})z^{k-2} = \frac{\sum_{i=0}^{q_1} \psi_i z^i}{1 - \sum_{i=1}^{p_1} \phi_i z^i}, \quad (2.44)$$

with $p_1 \in \{0, \dots, M-2\}$, $q_1 \in \{0, \dots, M-2\}$ and $p_1 + q_1 + 1 = M-1$. If, instead of an optimal L^2 approximation, one wanted to approximate the analytic power series $R(z) := \sum_{k=2}^{\infty} \langle f, \varphi_k \rangle z^{k-2}$ around the origin, in such a way that

$$R(z) = \frac{Q_{q_1}(z)}{P_{p_1}(z)} + O(z^{p_1+q_1+1}), \quad Q_{q_1}(z) = \sum_{i=0}^{q_1} a_i z^i, \quad P_{p_1}(z) = 1 - \sum_{i=1}^{p_1} c_i z^i,$$

then the solution would be given by the Padé approximant $Q_{q_1}(\cdot)/P_{p_1}(\cdot)$ of $R(\cdot)$. Note also that the Padé approximant coincides with the optimal L^2 solution when $p_1 = 0$. We thus propose the following algorithm:

1. Set $p_1^* = 0$, $q_1^* = M-2$, $\mathbf{b}_i^* = (\langle f, \varphi_1 \rangle, \dots, \langle f, \varphi_M \rangle)'$ and the initial error of approximation $E^* = \int_0^1 f^2(x)dx - \sum_{i=1}^M \langle f, \varphi_i \rangle^2$.
2. For $p_1 = 1, \dots, M-1$, set $q_1 = M - p_1 - 1$, and take the initial value $\mathbf{b} = (b_1, c_1, \dots, c_{p_1}, a_0, \dots, a_{q_1})'$ where $\frac{\sum_{i=0}^{q_1} a_i z^i}{1 - \sum_{i=1}^{p_1} c_i z^i}$ is the Padé approximant of R . Starting with this initial value, solve the optimisation problem

$$\hat{\mathbf{b}}_{p_1} = \arg \min_{\mathbf{b}} \|f - \hat{f}_M(x; \mathbf{b})\|^2 = \arg \min_{\mathbf{b}} \sum_{k=2}^{\infty} \{\langle f, \varphi_k \rangle - B_k(\mathbf{b})\}^2$$

where the minimum is taken for \mathbf{b} defined by (2.44), p_1 and q_1 being fixed. Let $E(p_1) = \|f - \hat{f}_M(x; \hat{\mathbf{b}}_{p_1})\|^2$. If $E(p_1) < E^*$, set $E^* = E(p_1)$, $p_1^* = p_1$, $q_1^* = q_1$ and $\mathbf{b}^* = \hat{\mathbf{b}}_{p_1}$.

At the output of the algorithm, $\sum_{k=1}^{\infty} B_k(\mathbf{b}^*)\varphi_k(x)$ is the optimal L^2 rational approximation of f , where the coefficients $(B_k(\mathbf{b}^*))_{k \geq 1}$ are defined by M numbers $\mathbf{b}^* = (b_1, \phi_1^*, \dots, \phi_{p_1^*}^*, \psi_0^*, \dots, \psi_{q_1^*}^*)'$ such that

$$B_1(\mathbf{b}^*) = b_1 = \int_0^1 f(x)dx, \quad \sum_{k=2}^{\infty} B_k(\mathbf{b}^*)z^{k-2} = \frac{\sum_{i=0}^{q_1^*} \psi_i^* z^i}{1 - \sum_{i=1}^{p_1^*} \phi_i^* z^i}.$$

Note that Step 1 gives the approximation of f by truncation. Therefore the rational approximation given by the algorithm cannot be less accurate than the approximation by truncation (based on the same number M of parameters).

2.7.3 Appendix C: Empirical application

Table 2.6: Ticker symbols and company names

Symbol	Issue name
ABT	ABBOTT LABORATORIES
AXP	AMERICAN EXPRESS CO
BA	BOEING CO
BAC	BANK OF AMERICA
BMJ	BRISTOL MYERS SQ
BP	BP PLC
C	CITIGROUP
CAT	CATERPILLAR
CL	COLGATE-PALMOLIVE CO
CSCO	CISCO SYSTEMS

Chapter 3

Backtesting Expected Shortfall via Multi-Quantile Regression¹

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, 2022 (BCBS, 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 L_t is the (integrable) ex-post loss on a portfolio at time t , Ω_{t-1} is the

¹This chapter is based on Couperier and Leymarie (2021).

information at time $t - 1$, and $Q_{L_t}(\cdot)$ is the quantile function of L_t , the τ -level ES and VaR are given by

$$\begin{aligned} ES_t(\tau) &= \mathbb{E}[L_t \mid L_t \geq VaR_t(\tau); \Omega_{t-1}], \\ VaR_t(\tau) &= Q_{L_t}(\tau; \Omega_{t-1}). \end{aligned}$$

As an alternative tail risk measure, ES offers a number of appealing properties that overcomes the deficiencies of the more-familiar VaR. In particular, ES is *coherent* meaning that it satisfies the properties of monotonicity, sub-additivity, homogeneity, and translational invariance (see Artzner et al., 1999; Acerbi and Tasche, 2002). Furthermore, ES gives information about the expected size of the 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 to overcome the "*perverse incentives to hold positions that featured significant tail risks but were subject to limited risk in "normal" conditions.*" (BCBS, 2019, page 3).

Although ES is considered as the new standard for risk management and regulatory requirements, there are still outstanding questions about the modeling of ES (see e.g., De Roon and Karehnke, 2017; Patton et al., 2019; Taylor, 2019; Hautsch and Herrera, 2020), and the validation of the ES forecasts, or backtesting. Jorion (2006) defines backtesting as a formal statistical framework that consists of verifying if actual losses are in line with projected losses. Because ES is unobservable, its evaluation cannot be done conventionally as a direct comparison of the observed value with its forecast, and thus generally relies on the elicibility property. A risk measure is *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 elicibility (Gneiting, 2011a), but satisfies narrower properties such as conditional elicibility (Emmer et al., 2015), or joint elicibility 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, 2020).

To circumvent the lack of elicibility 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. Assume the law of L_t is continuous. The definition of a conditional probability and a change of variable yield a useful representation of ES in terms of VaR

$$ES_t(\tau) = \frac{1}{1 - \tau} \int_{\tau}^1 VaR_t(u) du. \quad (3.1)$$

Using this connection, 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 that generalizes the violation process for VaR and propose two backtests of ES. Starting with the same process, Löser et al. (2019) develop a backtest of ES that is theoretically valid in finite out-of-sample size and that can be easily extended to a multivariate setting. Costanzino and Curran (2018) provide a Traffic 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). See the survey of Argyropoulos and Panopoulou (2016) for more details.

This article reviews the relation between ES and VaR to provide an original way for backtesting bank's internal ES models that has the advantage of examining a limited number of VaRs in the tail distribution of risk models. Indeed, the definition of a Riemann sum gives a handy approximation of ES,

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

where the risk level u_j is defined by $u_j = \tau + (j - 1)\frac{1-\tau}{p}$ for $j = 1, 2, \dots, p$. This representation suggests that p quantiles with appropriate risk levels would be convenient to assess the performance of an ES risk 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 in line with the guidelines of financial supervisors. Indeed, backtesting bank's internal ES models at the bank-wide level "*must be based on a VaR measure calibrated at a 99th percentile confidence level*" (BCBS, 2019, page 81) while it should be based at the trading desk level on "*each desk's one-day VaR measure [...] at both the 97.5th percentile and the 99th percentile*" (BCBS, 2019, page 83). This means that backtesting requirements within the Basel framework involve the ex-post validation of up to two quantiles issued by bank's internal ES models.

The main contribution of this article is to propose an original backtesting methodology to ES based on the theory of multi-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. The method extends the seminal idea of Gaglianone et al. (2011) to evaluate the validity of a single VaR relying on a single quantile regression.

Our procedure has many advantages. First, it encompasses the regulatory standards that consist of verifying the validity of one or two quantiles in the tail distribution of risk models which makes various comparisons possible. Second, it offers flexibility since the risk manager or the supervisor may select both the number of risk levels and their magni-

tude depending on the objective in mind (regulatory guidelines, ES statistical approximation, etc.). Third, it enters the category of regression-based backtests and complements the existing literature on regression-based risk forecast evaluation (see e.g., Engle and Manganelli, 2004; Christoffersen, 2011; Bayer and Dimitriadis, 2020). It also represents an alternative to the multiple VaR exceptions backtests (see e.g., Colletaz et al., 2013; Kratz et al., 2018). Finally, the procedure may be useful for systemic risk measures that involve VaR and ES estimation (see e.g., Benoit et al., 2015; Tarashev et al., 2016; Van Oordt and Zhou, 2019).

Formally, we show that the parameters of the multi-quantile regression model have specific properties under the hypothesis of valid ES forecasts. We propose four backtests which correspond to various linear restrictions on these parameters. These restrictions are implications of a Mincer-Zarnowitz representation (Mincer and Zarnowitz, 1969). Then, we test the resulting parameter restrictions using Wald-type inference. We apply QML estimation for the quantile regression parameters and we implement a pairs bootstrap algorithm (Freedman, 1981) to correct the finite sample size distortions of our backtests. Finally, we introduce a procedure deduced from our regression framework to adjust the invalid risk forecasts.

Several Monte Carlo experiments are reported and an empirical application with the S&P500 series is conducted. Our backtests deliver good performances to flag misleading ES risk models. We also find that the use of asymptotic critical values is prone to substantial size distortions, while the implementation of bootstrap critical values provides satisfactory size performances regardless of the sample size. The latter should hence be preferred when asymptotic theory does not apply conveniently.

Our empirical results suggest an update of the current regulatory guidelines. First, we show that the BCBS recommendation of assessing quantiles at risk levels 97.5% and 99% is not always sufficient to identify misspecified ES models. The use of additional quantiles is recommended to improve the soundness of the decision. Second, our results suggest to limit the number p of quantiles in small samples (with typically $p \leq 6$) and to consider higher values if the historical sample covers longer periods. 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 the paper is structured as follows. In Section 2, we introduce the multi-quantile regression framework. Section 3 describes the null hypotheses, the test statistics, their asymptotic properties, and the implementation of the bootstrap critical values. Section 4 studies the finite sample properties of our backtests from a set of Monte Carlo experiments. Section 5 applies our methodology to the S&P500 index and introduce a procedure to adjust the imperfect ES risk forecasts. We conclude the paper in Section 6.

3.2 Multi-quantile regression framework

This section describes 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. 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 approximated as an average 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 approximation is defined as a finite Riemann sum involving p VaRs such as*

$$ES_t(\tau) \approx \frac{1}{p} \sum_{j=1}^p VaR_t(u_j), \quad (3.2)$$

where risk levels u_j , $j = 1, 2, \dots, p$, 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 accuracy. In particular, $p = 1$ involves a single VaR at coverage level τ , while increasing p to infinity leads Equation (3.2) to converge to the theoretical ES. As we rely on a Riemann sum, the approximation assigns equal weights $1/p$ to each element in the sum, and the risk levels u_j , $j = 1, 2, \dots, p$, are determined so that the interval is equally partitioned between the two boundaries τ and 1. Several alternatives for the approximation of a definite integral are available. Here, we rely on a Riemann sum for its simplicity and ease of implementation. We show how to derive the above formula in Appendix A.

In practice, p may be chosen small as the interval of the definite integral is restricted to the extreme upper tail distribution. For instance, Gouriéroux and Liu (2012) identify for a large class of distributions a common linear conversion pattern between VaR and ES, so that a few VaRs are generally enough to get a good approximation of ES. Daniélsion and Zhou (2016) show that VaR and ES are in most cases related by a small constant and are hence almost equally informative.

Our approximation is useful for at least two reasons. First, this simple formula is appealing in a regulatory and risk management viewpoint since the estimation of VaR is well-established and its computation is easier compared to ES. Secondly, and it is the

purpose of this paper, the above relationship greatly simplifies the assessment of ES, by focusing on the validity of several VaRs, and is more intelligible in the context of banking regulation. This approach is fully consistent with the BCBS guidelines on ES assessment stating that backtesting bank's internal ES models at the bank-wide level "*must be based on a VaR measure calibrated at a 99th percentile confidence level*" (BCBS, 2019, page 81) and must be based at the trading desk level on "*each desk's one-day VaR measure [...] at both the 97.5th percentile and the 99th percentile*" (BCBS, 2019, page 83)

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 $t = 1, 2, \dots, T$. In addition, we denote by Ω_{t-1} the information set available at time $t - 1$, with $(L_{t-1}, L_{t-2}, \dots) \subseteq \Omega_{t-1}$. 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

$$\Pr(L_t \geq VaR_t(u_j) | \Omega_{t-1}) = u_j. \quad (3.3)$$

A VaR model is said to be correctly specified (at coverage level u_j) as soon as Equation (3.3) 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 (3.3) 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 $VaR_t(u_j)$, $t = 1, 2, \dots, T$, issued by the ES model satisfies Equation (3.3) jointly for $j = 1, 2, \dots, p$.

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 $\{L_t, t = 1, 2, \dots, T\}$ on the p VaR forecasts $\{VaR_t(u_j), t = 1, 2, \dots, T\}_{j=1,2,\dots,p}$ in a multi-quantile regression model as follows:

$$L_t = \beta_0(u_j) + \beta_1(u_j) VaR_t(u_j) + \epsilon_{j,t} \quad \forall j = 1, 2, \dots, p, \quad (3.4)$$

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) and is related to the multi-quantile CaViAR model (MQ-CaViAR) of White et al. (2015) which allows a joint modeling of multiple conditional VaRs. Given the multi-quantile regression model of Equation (3.4), the u_j -th conditional quantile of L_t is defined as

$$Q_{L_t}(u_j; \Omega_{t-1}) = \beta_0(u_j) + \beta_1(u_j) VaR_t(u_j) \quad \forall j = 1, 2, \dots, p. \quad (3.5)$$

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 $j = 1, 2, \dots, p$. For these 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 $j = 1, 2, \dots, p$. Under the hypothesis that a sequence of VaR is valid, coefficients satisfy $\beta_0(u_j) = 0$, and $\beta_1(u_j) = 1$, for $j = 1, 2, \dots, p$. 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 $\beta = (\beta(u_1)', \beta(u_2)', \dots, \beta(u_p)')$ the stacked vector of $2p$ coefficients. We assume that the sequence $\{u_j, j = 1, 2, \dots, p\}$ is ordered in the sense that $u_1 < u_2 < \dots < u_p < 1$.

In order to estimate β , we consider a QML estimator dedicated to multi-quantile regression, given by

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

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, it is shown that this estimator is consistent and asymptotically normally distributed. The conditions are described in Appendix B and a discussion is provided on how these assumptions are fulfilled.

Under Assumptions A0-A2 in Appendix B, the asymptotic distribution of the QML estimator is given by

$$\sqrt{T} (\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

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 $V = \mathbb{E}[\eta_t\eta_t']$, $\eta_t = \sum_{j=1}^p \nabla Q_{L_t}(u_j; \Omega_{t-1}) \psi_{u_j}(\epsilon_{j,t})$, $A = \sum_{j=1}^p \mathbb{E}[f_{j,t}(0) \nabla Q_{L_t}(u_j; \Omega_{t-1}) \nabla' Q_{L_t}(u_j; \Omega_{t-1})]$, where $\nabla Q_{L_t}(u_j; \Omega_{t-1})$ denotes the $2p$ gradient vector differentiated with respect to β , $\epsilon_{j,t} = L_t - Q_{L_t}(u_j; \Omega_{t-1})$, and $f_{j,t}(0)$ denotes the pdf of $\epsilon_{j,t}$ evaluated at zero. In Appendix C, we provide a consistent estimator $\hat{\Sigma}$ of Σ that will be used to compute our test statistics.

Finally, Appendix D provides a discussion on the rate of convergence and interplay of p and T when T tends to infinity. Under this asymptotic framework, it is shown that p is increasing with T . We consider a simple illustration assuming p takes the form of a power function. Under this assumption, T needs to diverge faster than p in order to maintain the asymptotic theory. This condition is not importantly restrictive but suggests the existence of an (asymptotic) upper limit for p which depends on the sample size. Section 4 provides several Monte-Carlo experiments with various values for p and T to give guidelines on how to choose these parameters jointly in finite samples. In Section 5, we suggest an avenue of research using extremal conditional quantile models (Chernozhukov, 2005; Chernozhukov and Fernández-Val, 2011) that allows examining deeper the tail distribution of ES risk models.

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 risk levels u_j , $j = 1, 2, \dots, p$. 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 $j = 1, 2, \dots, p$. As highlighted by Gaglianone et al. (2011) for a unique quantile regression, the aforementioned set of restrictions retains a Mincer and Zarnowitz (1969) interpretation for each quantile regression in (3.4). Here, 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:*

$$H_{0,J_1} : \sum_{j=1}^p (\beta_0(u_j) + \beta_1(u_j)) = p, \quad (3.6)$$

$$H_{0,J_2} : \sum_{j=1}^p \beta_0(u_j) = 0, \text{ and, } \sum_{j=1}^p \beta_1(u_j) = p, \quad (3.7)$$

$$H_{0,I} : \sum_{j=1}^p \beta_0(u_j) = 0, \quad (3.8)$$

$$H_{0,S} : \sum_{j=1}^p \beta_1(u_j) = p, \quad (3.9)$$

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.

Equations (3.6)-(3.9) of 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 risk forecasts are valid. The coefficients are summed across risk levels u_j , $j = 1, 2, \dots, p$. This aggregation substantially reduces the number of constraints. H_{0,J_2} is hence characterized by two constraints, and 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 $j = 1, 2, \dots, p$. 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 $\beta_0(u_j)$, $j = 1, 2, \dots, p$, and $H_{0,S}$ is devoted to the analysis of the slope parameters $\beta_1(u_j)$, $j = 1, 2, \dots, p$. These additional null hypotheses complement the joint backtests to identify the nature of the misspecification. If the joint hypotheses are rejected, separate tests for these two types of measurement error should be considered. They are inspired by the prediction-realization framework of Mincer and Zarnowitz (1969). When $H_{0,I}$ is rejected, the intercept parameters, $\beta_0(u_j)$, $j = 1, 2, \dots, p$, do not sum to 0, and hence, the average of VaR risk forecasts either underestimate or overestimate the true quantiles, if the sign of the sum is positive or negative, respectively. The rejection of $H_{0,S}$ indicates that the sum of the slope parameters $\beta_1(u_j)$, $j = 1, 2, \dots, p$, does not equal p , which highlights correlation between the forecasting errors and the quantile series.

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*

$$W = T \left(R_W \hat{\beta} - q_W \right)' \left(R_W \hat{\Sigma} R_W' \right)^{-1} \left(R_W \hat{\beta} - q_W \right), \quad (3.10)$$

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

To assess our null hypotheses we consider Wald-type inference. Equation (3.10) of Definition 3 gives the general expression of the test statistics. According to 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 7 (Chi-squared distribution). *Consider the multi-quantile regression model in Equation (3.4), Assumptions A0-A3 in Appendix B, 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 7 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. Note that we provide in Appendix E 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 we can employ them if the asymptotic conditions are fulfilled for realistic sample sizes. However, in the case of ES assessment, the focus is on the extreme tail distribution, that is for risk levels above the regulatory coverage level, i.e. $\tau = 0.975$. This may induce scarce information and affect the inference when the sample size is not large enough. Furthermore, our asymptotic framework implicitly assumes that $(1 - u_p)T$ diverges to infinity, where u_p denotes the highest level of the multi-quantile regression. Chernozhukov (2005) and Chernozhukov and Fernández-Val (2011) provide a refinement of this assumption based on the extreme value theory allowing $(1 - u_p)T \rightarrow k < \infty$. However, to date this literature has only considered single conditional extremal quantile models and it is not obvious how the results for the single quantile models extend to multi-quantile models. To overcome these typical deficiencies, we implement a bootstrap procedure to adjust the critical values of our test statistics 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 , $j = 1, 2, \dots, p$, 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 $\{L_t, VaR_t(u_j)\}_{j=1,2,\dots,p}$, $t = 1, 2, \dots, T$, to obtain $\hat{\beta}$ and $\hat{\Sigma}$, and compute the unconstrained test statistic W given by

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

2. Build a bootstrap sample by drawing with replacement T pairs of observations from the original data $\{L_t, VaR_t(u_j)\}_{j=1,2,\dots,p}$, $t = 1, 2, \dots, T$.
3. Estimate the model on the bootstrap sample, to obtain $\hat{\beta}^b$ and $\hat{\Sigma}^b$, and compute the bootstrapped test statistic W^b under the null hypothesis as follows:

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

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

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 $R_W \beta - R_W \hat{\beta}$ rather than $R_W \beta - q$. Since the estimate of β from the bootstrap samples should, on average, be equal to $\hat{\beta}$, 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 , $b = 1, 2, \dots, B$. 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 $P = B^{-1} \sum_{b=1}^B \mathbb{1}(W^b > W)$.

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 sample sizes $T = 250, 500, 1000, 2500$. The results associated with the bootstrap critical values are based on $B = 1000$ bootstrap samples. Finally, the backtests are computed with $\tau = 0.975$ that is the current banking regulation coverage level.

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 $p = 1, 2, 4, 6, 8, 10, 12$. The p risk levels u_1, u_2, \dots, 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) and the number of quantiles of the bank-wide level regulatory guidances (in case $\tau = 0.99$). With $p = 2$ risk levels, our backtests are in accordance with the regulatory guidances at the trading-desk risk management level. Finally, the case $p = 4$ corresponds to the framework considered by Emmer et al. (2015).

The correct data generating process is given by the AR(1)-GARCH(1,1) specification with Student innovations. This class of model is some of the most widely used for capturing variance dynamics in daily asset returns (see e.g., Berkowitz and O'Brien, 2002; Berkowitz et al., 2011; Du and Escanciano, 2017; Löser et al., 2019). The ex-post portfolio loss L_t , $t = 1, 2, \dots, T$, is given by

$$\begin{aligned} L_t &= \delta_0 + \delta_1 L_{t-1} + \epsilon_t, \\ \epsilon_t &= \sigma_t \eta_t, \quad \eta_t \sim t_v, \\ \sigma_t^2 &= \gamma_0 + \gamma_1 \epsilon_{t-1}^2 + \gamma_2 \sigma_{t-1}^2, \end{aligned} \tag{3.11}$$

where t_v denotes the Student's t distribution with v degrees of freedom. Given the model in Equation (3.11), the true ES and VaR at coverage level τ are given by

$$ES_t(\tau) = \delta_0 + \delta_1 L_{t-1} + \sigma_t m(\tau), \tag{3.12}$$

$$VaR_t(\tau) = \delta_0 + \delta_1 L_{t-1} + \sigma_t F_v^{-1}(\tau), \tag{3.13}$$

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. As a robustness check of the above model, Appendix F provides simulation results for the simple case of a GARCH(1,1) model that excludes the conditional mean component with $L_t = \epsilon_t$ where ϵ_t is as in Equation (3.11). Both models are calibrated using the opposite of the daily log-returns of the S&P500 index over the period from January 2, 2013 to December 29, 2017, with $(\hat{\delta}_0, \hat{\delta}_1, \hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2, \hat{v}) = (-0.085, -0.093, 0.034, 0.214, 0.748, 5)$ and $(\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2, \hat{v}) = (0.034, 0.197, 0.763, 5)$, respectively for the AR(1)-GARCH(1,1) model and the GARCH(1,1) model. 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 (3.11), with $\sigma_t^2 = (\gamma_0 + \gamma_1 \epsilon_{t-1}^2 + \gamma_2 \sigma_{t-1}^2) \times (1 - \kappa)$, where $\kappa = 0.25, 0.50, 0.75$, respectively.

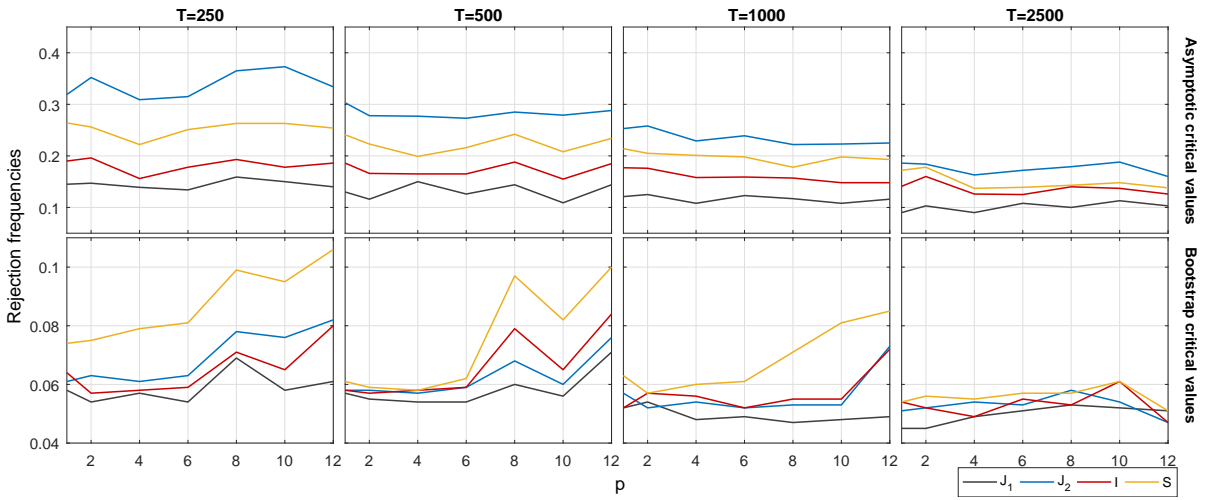
A_2 : GARCH in mean model: $L_t = \kappa \times \sigma_t^2 + \epsilon_t$, $\epsilon_t = \sigma_t \eta_t$, $\sigma_t^2 = \gamma_0 + \gamma_1 \epsilon_{t-1}^2 + \gamma_2 \sigma_{t-1}^2$, $\eta_t \sim t_v$, where $\kappa = +2.5, -2.5$, respectively.

A_3 : AR(1)-GARCH(1,1) model with mixed normal innovations: L_t satisfies Equation (3.11), with $\eta_t \sim (0.5X^+ + 0.5X^-) / \sqrt{10}$, 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 $VaR_t(\tau) = \text{percentile}(\{L_{t-i}\}_{i=1}^{250}, 100\tau)$, and $ES_t(\tau) = \frac{1}{\sum_{i=1}^{250} \mathbb{1}_{(L_{t-i} \geq VaR_{t-i}(\tau))}} \sum_{i=1}^{250} L_{t-i} \times \mathbb{1}_{(L_{t-i} \geq VaR_{t-i}(\tau))}$.

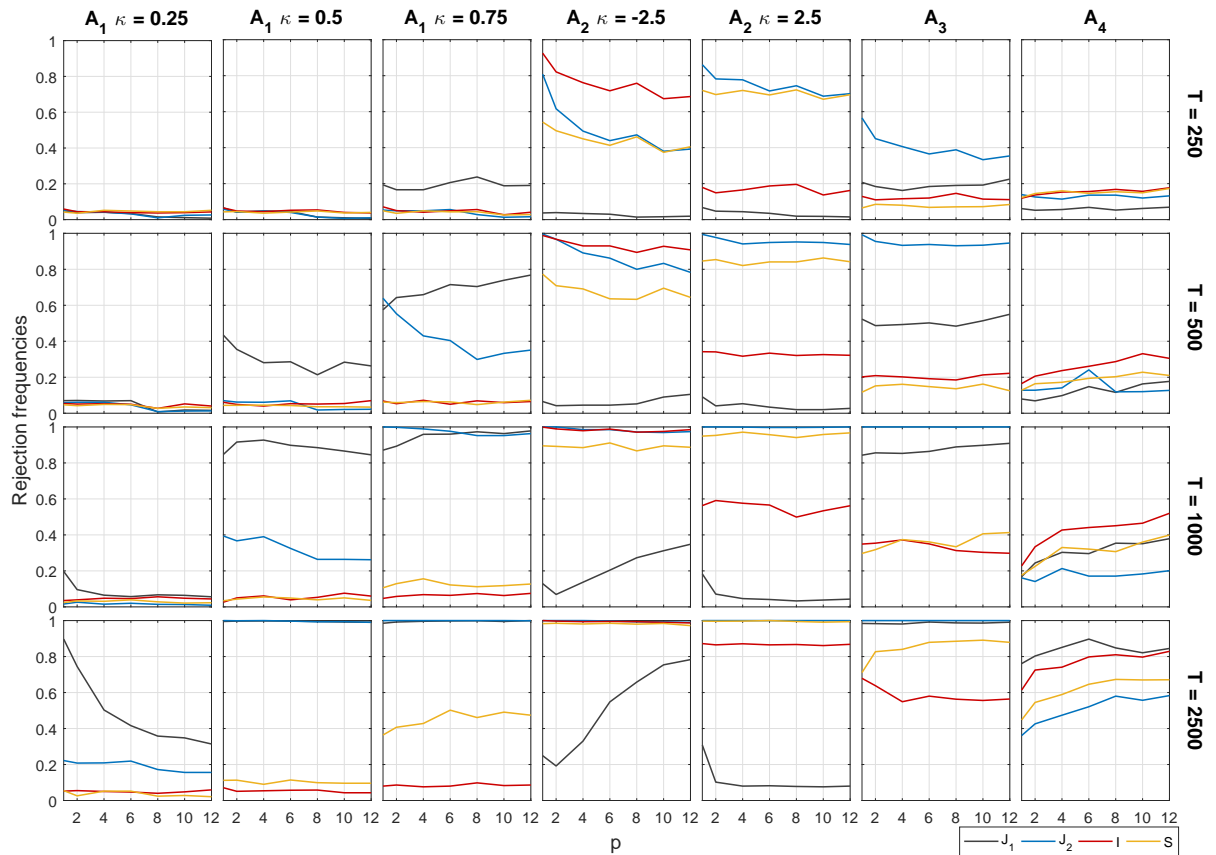
In A_1 , the conditional variance of the series σ_t is alternately underestimated of 25%, 50%, and 75% to examine whether our tests are able to detect an underestimation of ES stemming from a misleading appreciation of volatility. In A_2 , the misspecification occurs in the conditional mean by assuming 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 tail risk assessment. Bayer and Dimitriadis (2020) 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 (2020), Gaglianone et al. (2011), among others.

Figure 3.1: Empirical size of the tests at 5% significance level (AR(1)-GARCH(1,1) model)



Note: Size of the four backtests are displayed as a function of p . The first row reports the results computed with the asymptotic critical values, and the second row those computed with the bootstrap critical values. The columns correspond to different sample sizes T .

Figure 3.2: Empirical power of the tests at 5% significance level (AR(1)-GARCH(1,1) model, bootstrap critical values)



Note: Power of the four backtests are displayed as a function of p . The rows correspond to different sample sizes T , and the columns to the different misspecified alternatives A_1 - A_4 . Reported powers are size corrected.

Figure 3.1 displays graphically empirical sizes of the tests at 5% significance level. The first row reports the results of the asymptotic tests and the second row embeds those of the bootstrap based tests. Each column is for a given sample size T , and the results are shown as a function of p for comparison. As previously discussed, the use of conventional 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. On the contrary, the backtests based on bootstrap critical values display empirical sizes that are close to the nominal size of 5% for all reported sample sizes and risk levels. For large coverage levels and moderate samples, it is hence recommended to use bootstrap critical values rather than the standard asymptotic ones.

To go further about the interrelationship between p and T , we note that the size of the four bootstrap-based backtests slightly deteriorates for $p > 6$ with $T = 250$ and $T = 500$ revealing that the tests are sensitive to the choice of p in small samples. In details, the slope backtest is the most affected by these distortions, while the J_1 backtest is well-sized most of the time. On the contrary, for larger sample sizes, typically $T = 1000$ and $T = 2500$, these distortions are negligible. Consequently, it would make sense to restrict the number p of quantiles when applying the tests in small samples, with typically $p \leq 6$, and to consider higher values if the historical sample covers longer periods.

To provide robustness check of these results, Figure 3.7 in Appendix F reports empirical sizes when the data generating process is given by a GARCH(1,1) model. We observe the same findings as those provided with the AR(1)-GARCH(1,1) model. The asymptotic tests are largely oversized, while the bootstrap tests are close to the nominal size of 5% for all reported sample sizes and risk levels. Finally, there is also an asymptotic refinement of the empirical sizes as T increases for both asymptotic and bootstrap tests.

Figure 3.2 reports the empirical powers (size-corrected) associated with our seven alternatives. Here, we only present the simulation results associated with the bootstrap critical values. The simulation results obtained with the asymptotic critical values are overall the same (see Figure 3.6 in Appendix F). Overall, the tests correctly flag the misspecified alternatives A_1 , A_2 , A_3 , A_4 , and we verify that there is a general improvement of powers as the sample size T increases (from row 1 to row 4), 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. Although the intercept and slope backtests exhibit lower power performances, they provide useful informations on the type of misspecification. In details, the slope backtest performs better in alternatives A_1 and A_3 , while the intercept backtest is superior for alternative A_4 . Thus, A_1 and A_3 mainly affect the expected value of the slope parameters meaning that the errors are correlated and proportional to the true quantiles. In contrast, alternative A_4 induces distortions in the expected value of the intercept coefficients suggesting that the origin of errors is more global as they are not related to the true quantiles.

Third, we observe that the selection of number p of risk levels is difficult to link with the rejection frequencies in alternatives A_1 , A_2 , A_3 , since reported powers are slightly affected by p in general. This finding is explained by the nature of these alternatives for which the misspecification is relatively uniform along the tail, and does not require many levels. On the contrary, in alternative A_4 , an increase of p is beneficial for detecting the misleading one-year historical simulation method as power is unequivocally increasing with p , especially when T is large. 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. In the same spirit as in the Portmanteau tests when selecting an optimal number of lags to flag serial correlation (Ljung and Box, 1978; Box and Pierce, 1970), it follows that considering successive values $p = 1, 2, \dots, p_{\max}$ with $p_{\max} = 12$, as illustrated above, may offer useful information to improve the statistical decision.

Finally, we provide a robustness check of the powers with the GARCH(1,1) model (see Figures 3.8 and 3.9 in Appendix F). The rejection frequencies are very close to those associated with the AR(1)-GARCH(1,1) model. Consequently, the decision whether to introduce or not a conditional mean in the risk model does not affect the power performances.

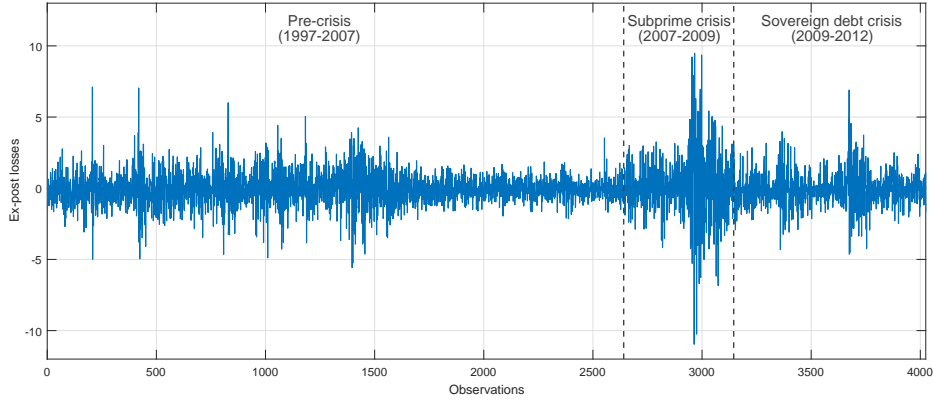
3.5 Empirical application

In this section, we apply our backtests to the daily returns of the S&P500 index. In addition, we provide a method for the adjustment of imperfect risk 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 , $j = 1, 2, \dots, p$, are calculated accordingly with Definition 1. In addition, we consider the risk levels suggested by the BCBS at the trading-desk level, 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 $p = 1, 2, 4, 6, 8, 10, 12$.

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 to June 30, 2007, and we use 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 the subprime mortgage crisis and the European sovereign debt crisis, two major episodes of 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 3.3: 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
Kurt. exc.	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 to December 31, 2012. Source: *finance.yahoo.com* website.

The S&P500 series is depicted in Figure 3.3 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) display more severe levels of volatility, with several extreme events. Figure 3.3 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, the series is right-skewed and has a kurtosis excess.

To predict the ES risk measure, we fit an AR(1)-GARCH(1,1) model with Student innovations, as defined in (3.11), using the S&P500 daily losses of the in-sample period. The ES and VaR forecasts are defined as in Equations (3.12) and (3.13), respectively. The set of unknown parameters is estimated by maximum likelihood. We obtain the following coefficient estimates $\{\hat{\delta}_0, \hat{\delta}_1, \hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2, \hat{v}\} = \{-0.057, -0.032, 0.007, 0.060, 0.936, 9\}$. As a robustness check, we also fit a GARCH(1,1) model on the same period as defined in the simulation study and for which we obtain the following estimates $\{\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2, \hat{v}\} = \{0.007, 0.059, 0.937, 9\}$.

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 $p = 1, 2, 4, 6, 8, 10, 12$ 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 G.

Figure 3.4: In-sample ES estimates issued from the approximation and the exact calculation method (AR(1)-GARCH(1,1) model)

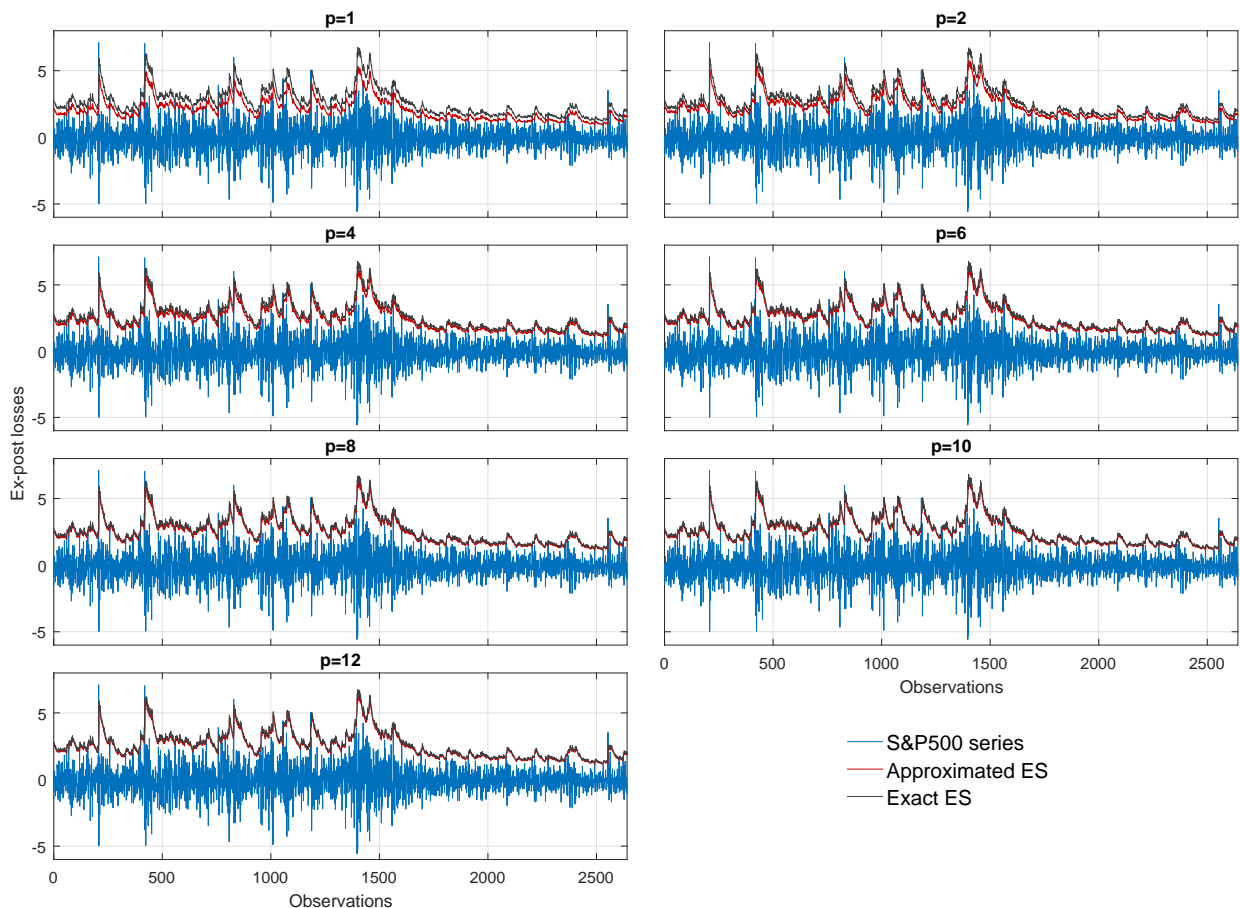


Figure 3.4 reports the in-sample ES estimates obtained with the approximation and the exact calculation method. Two remarks should be made here. First, the ES risk forecasts issued from the approximation and the exact method strongly correlate regardless of the value p . The approximation performs very well to capture the ex-post losses information. Second, we observe that the approximation is substantially improved when p slightly increases and coincides almost completely with the exact ES using six (or more) quantiles.

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. Daniélsson and Zhou (2016) argue that the two measures of risk are related by a small constant and are conceptually equally informative. This similarity also comes from the structure of the model used to compute the risk measure. For instance, VaR and ES

issued by an AR(1)-GARCH(1,1) model have common conditional mean and variance across risk levels implying that these risk measures are closely related (see Equations (3.12) and (3.13)). Finally, Figure 3.10 in Appendix H displays the same results using a GARCH(1,1) model. Removing the conditional mean component does not affect the approximation accuracy as the two computation methods match almost perfectly for $p \geq 6$. For its ease of implementation and accuracy, the approximation is appealing to compute and evaluate the performance of ES risk forecasts.

Table 3.1: p-values of the backtesting tests (AR(1)-GARCH(1,1) model)

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
8	0.099	0.049	0.154	0.564
10	0.029	0.061	0.053	0.432
12	0.023	0.052	0.038	0.223
2 (<i>regulatory levels</i>)	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
8	0.008	0.008	0.041	0.538
10	0.007	0.010	0.021	0.410
12	0.004	0.006	0.008	0.245
2 (<i>regulatory levels</i>)	0.006	0.012	0.032	0.448

Note: p-values of the four backtests computed with $p = 1, 2, 4, 6, 8, 10, 12$ 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.

Table 3.1 reports the p-values of the backtests. For a sake of clarity, we only report the p-values obtained with the bootstrap critical values and the results are discussed at 5% significance level. Panel A provides the results over the sample 2007-2009. The test statistic J_1 leads to reject the validity of the ES predictions regardless of the number p of quantiles (except for $p = 8$ where the rejection occurs at a 10% significance level). Interestingly, we observe that the larger p , the smaller the p-value until $p = 6$, indicating that the rejections are more severe when the number of risk levels increases until an optimal number p . It supports the existence of an upper limit for p relative to the sample size as T is relatively small ($T = 504$), and p should not be chosen too large. The test statistic J_2 displays higher p-values in general. 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 $p = 2, 4, 6, 12$, we tend to reject the expected value on the intercept coefficients, and there is a global bias in the quantile estimates issued by the ES model. On the contrary, the statistic S leads to the conclusion that the slope parameters are as expected under the null hypothesis, and thus, the magnitude of errors is not related to the true quantiles. Panel B contains the p-values for the period 2007-2012. Overall, we obtain similar results, but the rejections are more severe in this enlarged sample. Rejections of J_1 are now experienced at a 1% significance level and even for $p > 6$, as opposed to panel A. This highlights the underlying link between p and T as panel B uses $T = 1384$ observations enabling the use of additional quantiles. Table 3.3 of Appendix H displays the p-values of the backtests when applying a GARCH(1,1) model. The results are similar. Note however, for $p = 1$, that the p-values are generally higher with the GARCH(1,1) model than for the AR(1)-GARCH(1,1) model. For instance, the p-value of J_1 in panel B is 0.056 with the AR(1)-GARCH(1,1) model, while it reaches 0.199 with the GARCH(1,1) model. For that model, additional quantiles are needed to increase the rejection capabilities of the tests.

In sum, we should be 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. Furthermore, the results issued from the regulatory guidelines are contrasted. One or two risk levels are not always enough to provide a sound conclusion about the validity of bank's internal ES models. At least in our sample, looking at additional risk levels beyond the regulatory coverage level $\tau = 0.975$ improves the decision reliability on whether or not using that model.

Table 3.2: QML coefficient estimates ($p = 6$, AR(1)-GARCH(1,1) model)

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

Note: Standard errors are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5% and 1% level, respectively, and are obtained with the pairs bootstrap algorithm. °, °°, and °°, indicate statistical significance at the same levels and are obtained with the procedure of Chernozhukov and Fernández-Val (2011). Panel A gives estimation results for the period 2007-2009 and Panel B provides estimation results for the period 2007-2012.

Table 3.2 displays the coefficient estimates of the multi-quantile regression of Equation (3.4) for $p = 6$ risk levels, to help understand the reasons that explain the rejections of the ES forecasts. Panels A and B provide the results for periods 2007-2009 and 2007-2012, respectively. 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. We observe in both panels that the coefficients β_0 are overestimated for all the risk levels u_1, u_2, \dots, 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, \dots, 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.

Furthermore, we provide in Table 3.2 one by one inference on the regression parameters with the pairs bootstrap algorithm. The results are depicted with the symbol "*" and are discussed at a 5% significance level. We observe that the intercept parameters are statistically not equal to zero for the intermediary 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. In addition, we report 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.

We conclude the study by suggesting an avenue of research to backtesting tail risk measures under the extreme value (EV) theory. As pointed out in Section 3.2, our baseline regression implicitly assumes that $(1 - u_p)T \rightarrow \infty$ when $T \rightarrow \infty$. Chernozhukov (2005) shows that EV laws apply under the condition that $(1 - u_p)T \rightarrow k < \infty$ when $T \rightarrow \infty$. We consider the extremal conditional quantile model of Chernozhukov and Fernández-Val (2011) which is valid under that extremal condition. The procedure allows testing individual restrictions while assuming the above EV theory. In the same spirit, Bee et al. (2018) have proposed a new conditional quantile model with EV theory refinements that overcomes the issue of unstable inference in the tails where data are sparse. The results are depicted in Table 3.2 with the symbol "o". Overall, we find similar results between EV theory and pairs bootstrap. Rejection of the null is mostly experienced at the same

levels in panel A and panel B. However, we observe that the procedure of Chernozhukov and Fernández-Val (2011) is generally more powerful than pairs bootstrap at the highest risk levels. For instance, the expected value of the intercept parameters $\beta_0(u_j)$ in panel A is rejected at level 1% for $j = 3, 4, 5, 6$ with the EV procedure, while the pairs bootstrap rejects the null at larger levels (5% or 10%). This robustness check greatly illustrates the superiority of EV theory when applied to multi-quantile regression models and tail risks. We get overall the same results with the GARCH(1,1) model (see Table 3.4 in Appendix H).

3.5.3 Adjusted ES forecasts

In this section, we devise an adjustment for the ES risk forecasts deduced from our backtests. Our routine accounts for both misspecification and estimation uncertainty, without having to change the misspecified risk model. The procedure serves at identifying whether the model overestimates or underestimates the true unknown ES, by comparing the initial forecast with its adjusted counterpart, which may be useful for risk managers and regulatory agencies.

The correction of imperfect risk forecasts is not new in the financial literature. Gouriéroux and Zakoïan (2013) adjust VaR forecasts affected by estimation uncertainty. Boucher et al. (2014) adjust imperfect VaR forecasts based on backtesting frameworks and Lazar and Zhang (2019) recently apply the same strategy to adjust imperfect ES forecasts. The typical method consists of modifying the coverage level τ of the risk measure to meet the null hypothesis of valid risk forecasts. Here, the originality stems from the fact that we use a regression-based method to correct the ex-ante forecasts, while available techniques are based on the concept of violation (Christoffersen, 1998; Kupiec, 1995). This allows adjusting 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 Equation (3.5), 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 ES approximation as follows:

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

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 pa-

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, \dots, u_p .

Figure 3.5: ES forecasts and adjusted ES forecasts over the period 2007-2009 (AR(1)-GARCH(1,1) model)

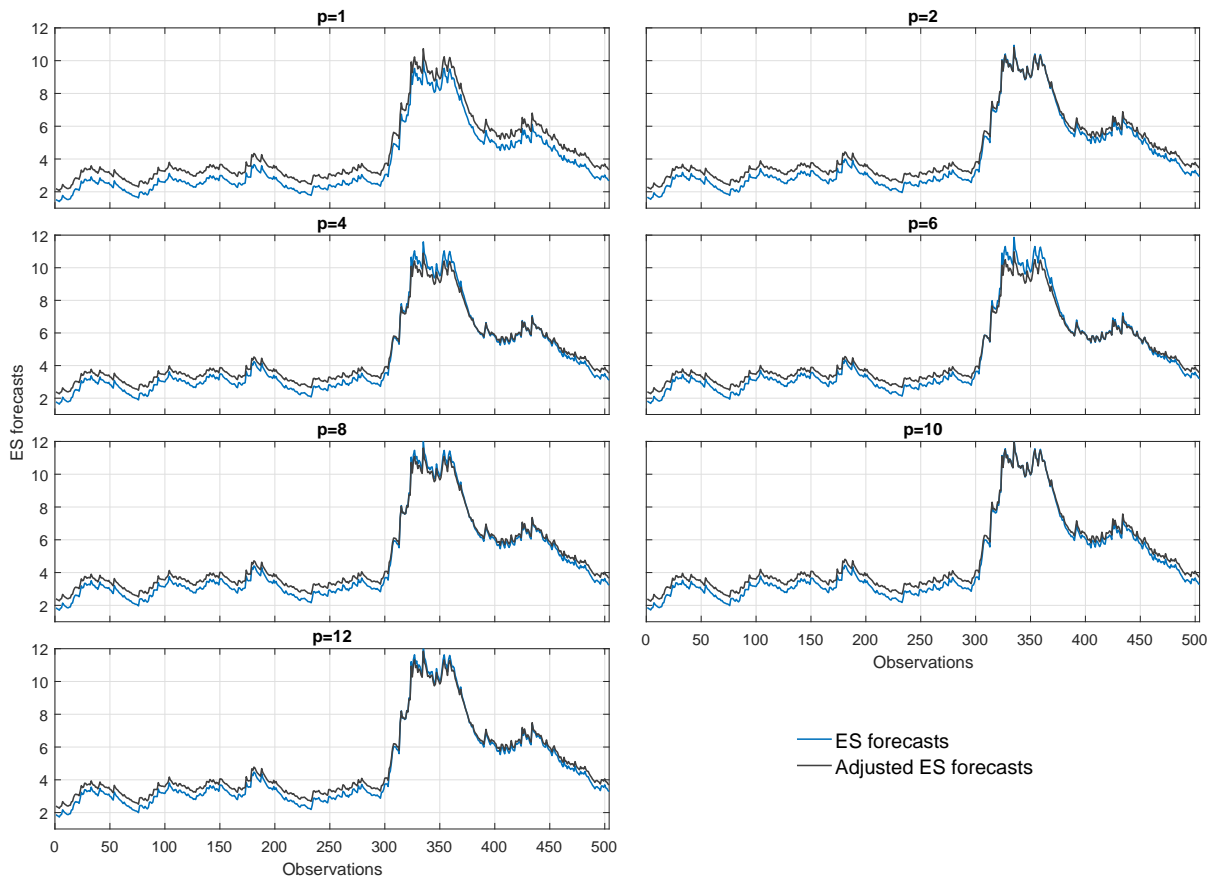


Figure 3.5 reports the ES predictions and adjusted ES predictions for the period 2007-2009 and $p = 1, 2, 4, 6, 8, 10, 12$ quantiles. The adjusted ESs are issued from the parameters $\beta_0(u_j)$ and $\beta_1(u_j)$, $\forall j$, that are estimated using the observations over 2007-2009, enabling us to disentangle whether the ES forecasts are overestimated or underestimated. We observe that the AR(1)-GARCH(1,1) model generally provides underestimated forecasts compared to the adjusted predictions. This finding is consistent with Begley et al. (2017) for the VaR who calculate the average number of exceptions of bank's self-reported level of VaR and show that the number of exceptions increases considerably during the period 2007-2009. Differently, De Nicolò and Lucchetta (2017) show that AR-type models often underestimate tail risks for horizons up to one year ahead, while there is probably a room for improvement using factor-augmented quantile models. Note also that the skewness of the daily losses is positive in our sample (see Figure 3.3), and it would be useful to account for asymmetry considering an asymmetric power distribution as proposed by

Komunjer (2007). Using the AR(1)-GARCH(1,1), we observe that the underestimation is more pronounced for the smallest predictions, the error being more severe when the risk forecasts are originally small. Finally, the ES forecasts are slightly overestimated when the variance of the series is larger, suggesting that the risk model may overestimate the true volatility in turbulent financial times. This is due to the volatility persistence in the GARCH component. Our findings are robust to (1) the use of a simple GARCH(1,1) model, (2) the use of the two BCBS regulatory levels, and (3) the extended period 2007-2012 (see Figures 3.11, 3.12, 3.13, and 3.14 in Appendix H).

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. One of the major obstacle to its implementation was the deficit of simple tools for the evaluation of its forecasts. This article introduces four easy-to-use regression-based backtests of ES. Our 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, 2020; Kratz et al., 2018, etc.). Our methodology detects 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 to the S&P500 index over the period 2007-2012. During this period of financial turmoil, our backtests clearly reject the validity of the ES forecasts issued by a AR(1)-GARCH(1,1) and a GARCH(1,1) model. We also highlight the importance of choosing a sufficient number of quantiles to assess ES risk models. The use of one or two quantiles is not always enough to identify whether the tail risk in bank's internal ES models is properly accounted for. On the contrary, four or more quantiles (until an optimal number) deliver much more sound decisions, suggesting an update of the regulatory guidelines.

3.7 Appendix

3.7.1 Appendix A: Application of a finite Riemann sum to ES

In the sequel, we show how to derive the approximation of ES suggested in Definition 1. Consider the following improper Riemann integral,

$$\int_a^b f(t)dt, \quad (3.14)$$

where $f(\cdot)$ is given by the increasing function $\frac{1}{1-\tau}VaR_t(\cdot)$ and where a and b are respectively τ and 1 so that the above expression is identical to the ES defined in Equation (3.1). Definition of a Riemann sum yields a useful approximation of Equation (3.14),

$$S_p(f) = \frac{b-a}{p} \sum_{j=1}^p f\left(a + (j-1) \frac{b-a}{p}\right),$$

where p is the number of subdivisions or quantiles taken in the definite integral to approximate ES. Replacing a , b , and $f(\cdot)$, by their corresponding quantities leads,

$$\frac{1}{1-\tau} \int_{\tau}^1 VaR_t(u)du \approx \frac{1}{1-\tau} S_p(VaR_t) = \frac{1}{p} \sum_{j=1}^p VaR_t\left(\tau + (j-1) \frac{1-\tau}{p}\right).$$

This verifies the ES formula of Definition 1 where risk levels u_j are given by $\tau + (j-1) \frac{1-\tau}{p}$.

3.7.2 Appendix B: Assumptions

This section introduces the assumptions needed to establish the asymptotic normality and the consistency of the QML estimator and to ensure the validity of Proposition 7.

Assumption A0: $\{L_t, VaR_t(u_j)\}_{j=1}^p$ is a stationary and ergodic process and measurable with respect to Ω_{t-1} .

Assumption A1: L_t has conditional (on Ω_{t-1}) distribution function F_t , with continuous and positive density f_t at conditional quantile $Q_{L_t}(u; \Omega_{t-1}) = F_t^{-1}(u|\Omega_{t-1})$ for all $u \in (0, 1)$.

Assumption A2: We have $\mathbb{E}[|L_t|] < \infty$. Furthermore, consider the quantity $D_{0,t} = \max_{t=1, \dots, T} \max_{j=1, \dots, p} \sup |Q_{L_t}(u_j; \Omega_{t-1})|$, then we have $\mathbb{E}[D_{0,t}] < \infty$.

Assumption A3: The matrices $A = \sum_{j=1}^p \mathbb{E}[f_{j,t}(0) \nabla Q_{L_t}(u_j; \Omega_{t-1}) \nabla' Q_{L_t}(u_j; \Omega_{t-1})]$ and $V = \mathbb{E}[\eta_t \eta_t']$ are positive definite.

Assumption A0 is standard in modeling financial times series. It is broadly accepted that asset prices are integrated at order one, so that financial returns are stationary. This data assumption is hence satisfied. Assumption A1 allows for nonidentical distributions as we enable L_t to be conditional on an unknown information set Ω_{t-1} . Assumption A2

imposes moment conditions, and in particular ensures finite expectation for L_t . This is satisfied by the vast majority of financial time series models, including stationary and invertible ARMA processes, GARCH processes, etc. Assumption A3 is standard in Wald-type inference to ensure that the variance-covariance matrix Σ is positive definite. Furthermore, Assumptions A0 through A2 are standard in QML estimation (e.g., White, 1994), and are also widely used in the literature on quantile regression models (e.g., Koenker and Machado, 1999; Koenker and Xiao, 2002). They are of great importance to establish consistency and to apply the central limit theorem of White (2001, theorem 5.24) based on the method proposed by Huber (1967).

3.7.3 Appendix C: 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. (2015). A consistent estimate of Σ can be obtained from the decomposition of the Huber (1967) sandwich form and is thus given by $\hat{\Sigma} = \hat{A}^{-1}\hat{V}\hat{A}^{-1}$. In the sequel, we provide consistent estimators \hat{A} and \hat{V} . To obtain \hat{V} , we apply a simple plug-in estimator as follows:

$$\hat{V} = T^{-1} \sum_{t=1}^T \hat{\eta}_t \hat{\eta}_t'$$

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

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 implemented by Engle and Manganelli (2004) for instance to estimate the variance-covariance matrix of a set of coefficients issued from the so-called CaViaR model. Then, \hat{A} is given by

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

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

3.7.4 Appendix D: On the rate of convergence and interplay of p and T

Let us consider the highest risk level u_p issued from the sequence $u_j, j = 1, 2, \dots, p$. Given Definition 1, we have

$$u_p = 1 - (1 - \tau)/p,$$

where p is the number of VaRs used to approximate ES and τ is a constant number representing the coverage level of ES. For a number p of subdivisions large enough, the approximation of Definition 1 is close to the theoretical ES and u_p is close to one. In what follows, we study this limiting case. Let us define u_p as a function of the sample size T such as,

$$u_p = 1 - \epsilon_T, \tag{3.15}$$

where the nonstochastic positive sequence $\epsilon_T = (1 - \tau)/p$ satisfies $\epsilon_T \rightarrow 0$ when $T \rightarrow \infty$. Equation (3.15) is a common representation to extremal quantile regression (Chernozhukov, 2005; Chernozhukov and Fernández-Val, 2011). Given the definition of ϵ_T and since τ is a constant parameter, it follows that p is increasing with T . To illustrate this point, assume p takes the form of a power function,

$$p = T^\gamma, \tag{3.16}$$

with $\gamma > 0$. Next, we consider a QML estimator which implicitly assumes that $T(1 - u_p) \rightarrow \infty$ as T goes to infinity. Chernozhukov (2005) and Chernozhukov and Fernández-Val (2011) relate this condition to an intermediate order quantile regression. Our goal is to identify a suitable rate of convergence of p and T which ensures the above condition. We have

$$T\epsilon_T \rightarrow \infty. \tag{3.17}$$

Then, combining Equations (3.16) and (3.17), we get

$$(1 - \tau)T^{1-\gamma} \rightarrow \infty. \tag{3.18}$$

Equation (3.18) is only satisfied when $\gamma < 1$. Looking at Equation (3.16), this condition implies that T needs to diverge faster than p to guarantee asymptotic theory.

3.7.5 Appendix E: 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

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

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

$$R_W \hat{\beta} - q_W \xrightarrow{p} R_W \beta - q_W \neq 0.$$

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

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

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

$$\lim_{T \rightarrow +\infty} W = +\infty,$$

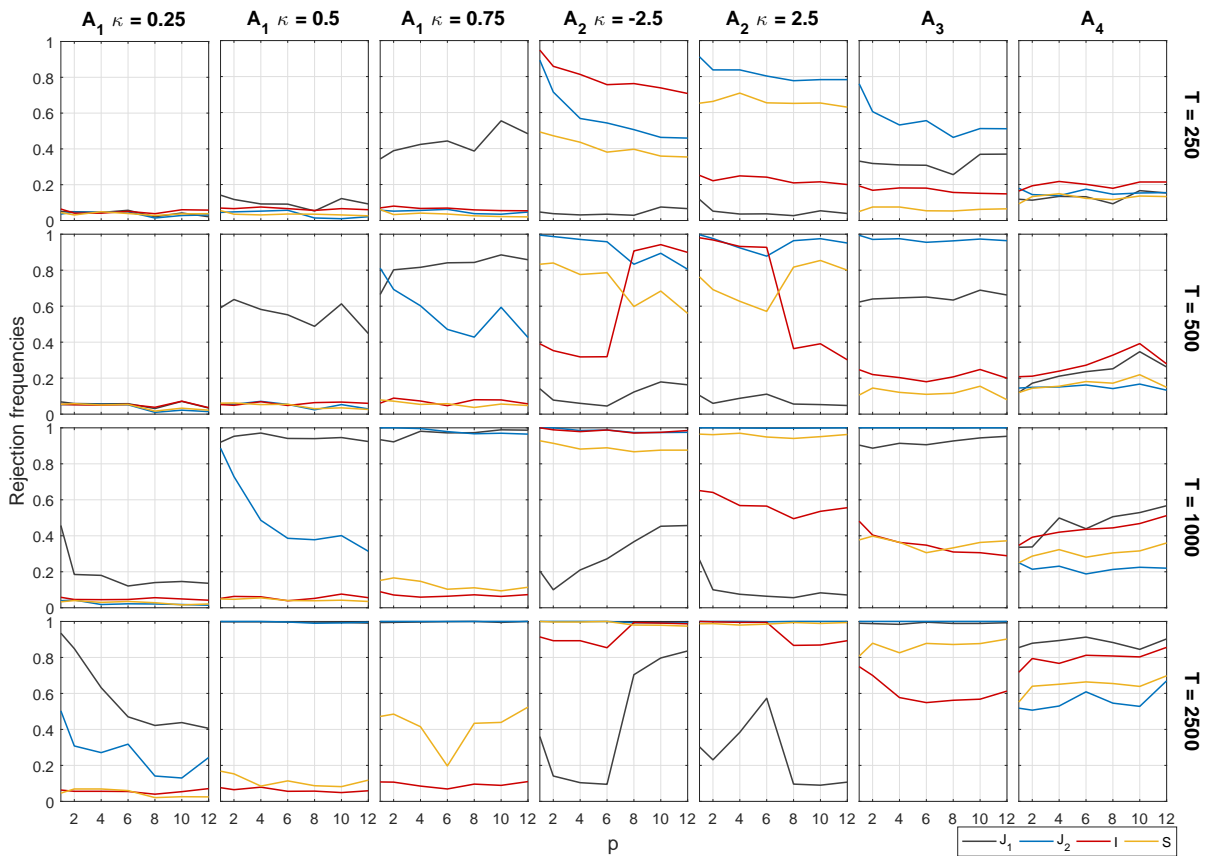
and therefore we get

$$\lim_{T \rightarrow +\infty} \mathbb{P}(W > \chi_{1-\alpha}^2(d_W) | H_{1,W}) = 1,$$

where $\chi_{1-\alpha}^2(d_W)$ 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. ■

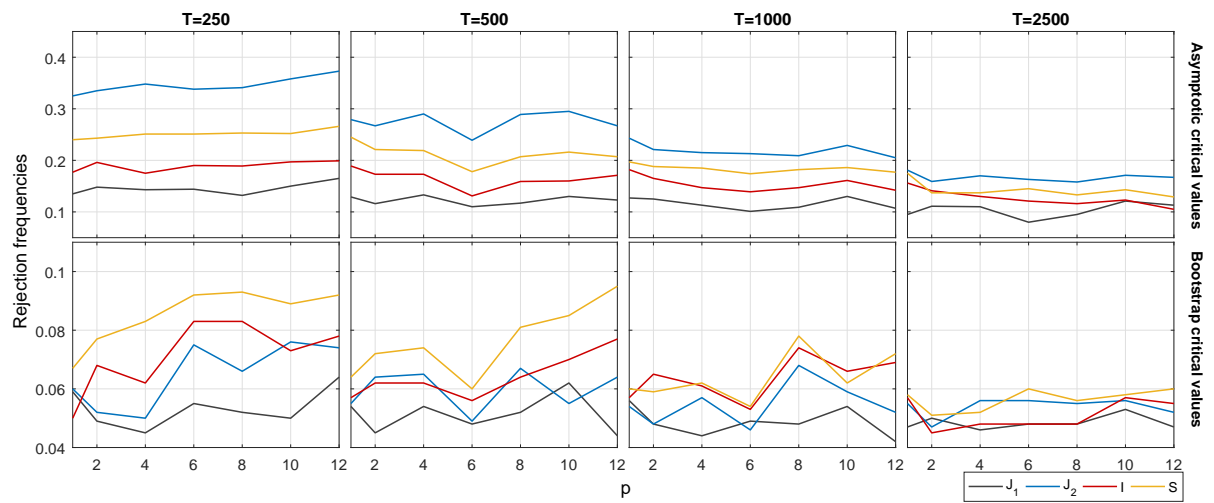
3.7.6 Appendix F: Robustness checks of the simulation study

Figure 3.6: Empirical power of the tests at 5% significance level (AR(1)-GARCH(1,1) model, asymptotic critical values)



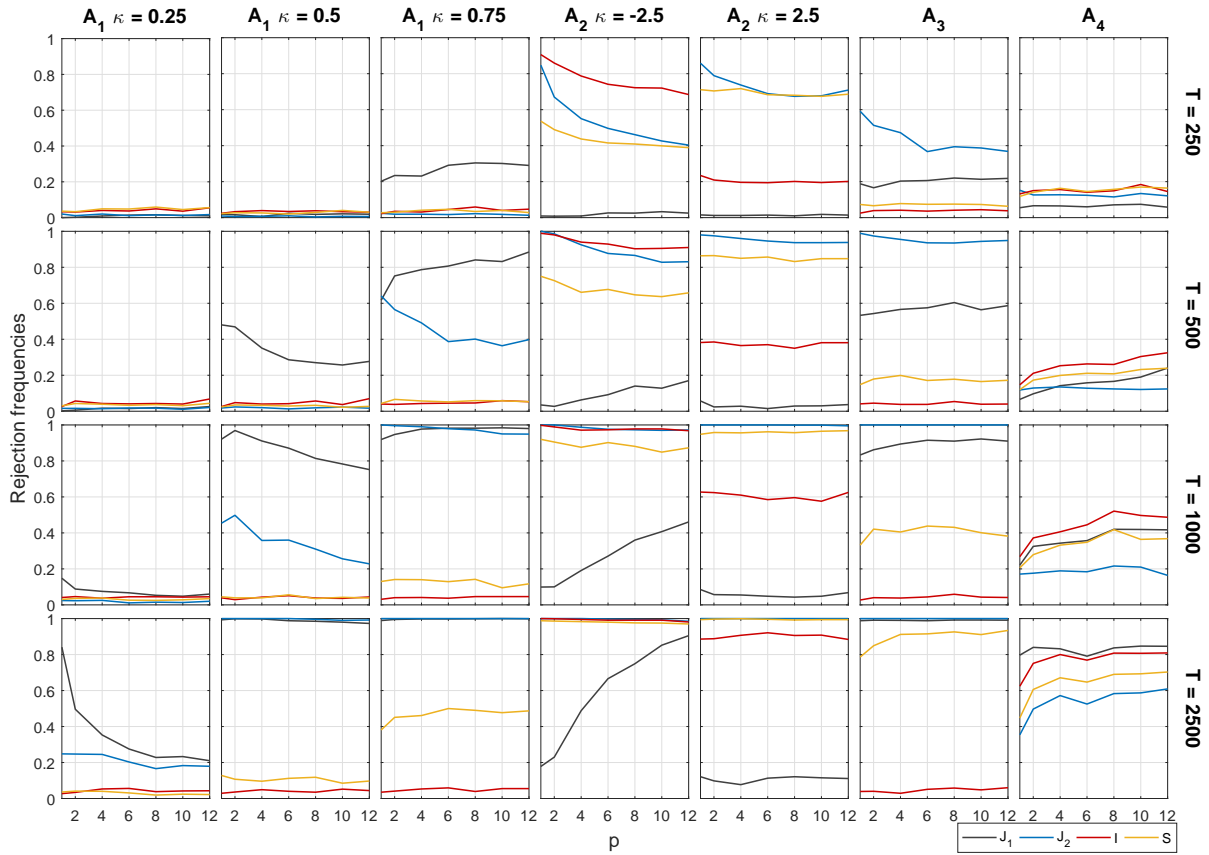
Note: Power of the four backtests are displayed as a function of p . The rows correspond to different sample sizes T , and the columns to the different misspecified alternatives A_1 - A_4 . Reported powers are size corrected.

Figure 3.7: Empirical size of the tests at 5% significance level (GARCH(1,1) model)



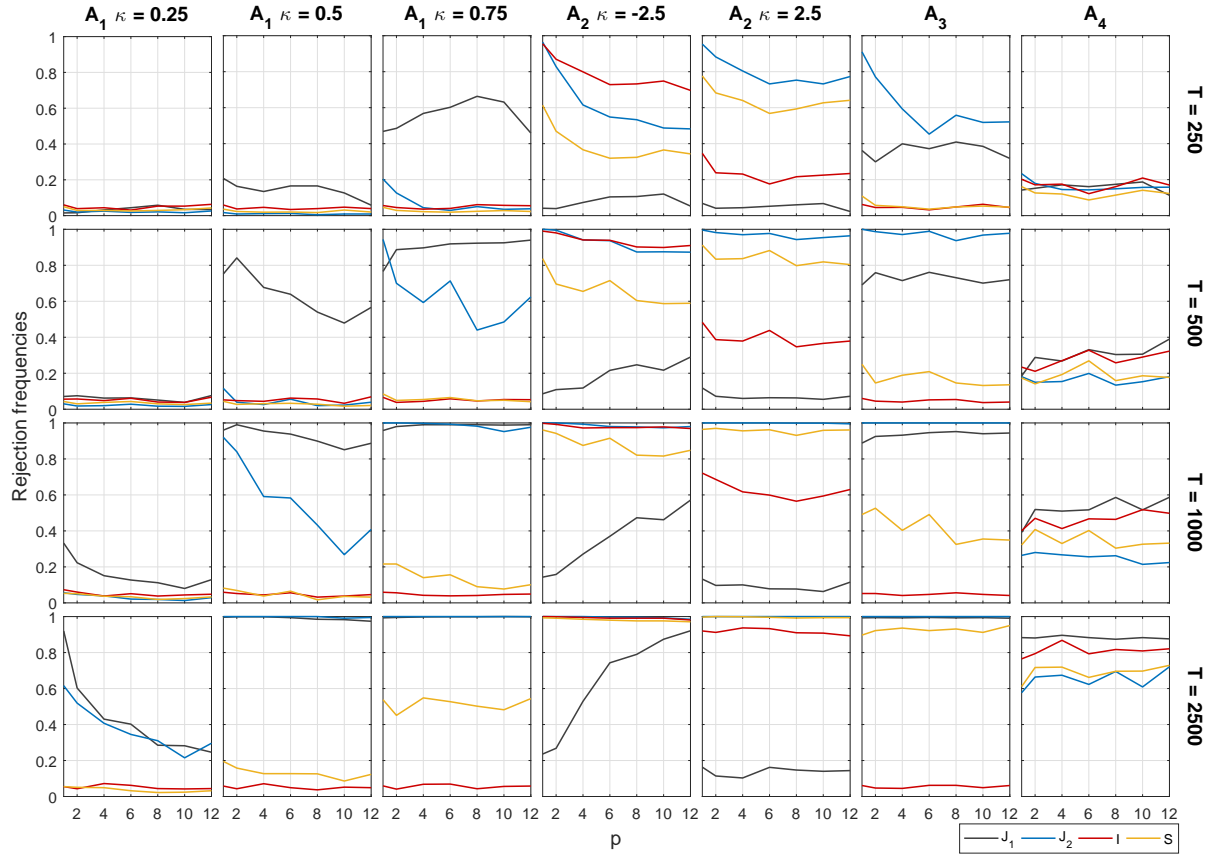
Note: Size of the four backtests are displayed as a function of p . The first row reports the results computed with the asymptotic critical values, and the second row those computed with the bootstrap critical values. The columns correspond to different sample sizes T .

Figure 3.8: Empirical power of the tests at 5% significance level (GARCH(1,1) model, bootstrap critical values)



Note: Power of the four backtests are displayed as a function of p . The rows correspond to different sample sizes T , and the columns to the different misspecified alternatives A_1 - A_4 . Reported powers are size corrected.

Figure 3.9: Empirical power of the tests at 5% significance level (GARCH(1,1) model, asymptotic critical values)



Note: Power of the four backtests are displayed as a function of p . The rows correspond to different sample sizes T , and the columns to the different misspecified alternatives A_1 - A_4 . Reported powers are size corrected.

3.7.7 Appendix G: Exact calculation method of ES

This section describes the methodology for the exact computation of ES forecasts at coverage level τ . Several techniques are available in practice. As 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 such that $m(\tau) = \frac{1}{\sum_{s=1}^S \mathbf{1}(\eta_t^s \geq F_v^{-1}(\tau))} \sum_{s=1}^S \eta_t^s \times \mathbf{1}(\eta_t^s \geq F_v^{-1}(\tau))$, where $F_v^{-1}(\tau)$ is the τ -quantile of the innovation distribution and is obtained as $F_v^{-1}(\tau) = \text{percentile}(\{\eta_t^s\}_{s=1}^S, 100\tau)$.

3. Compute the ES at time t as $ES_t(\tau) = \delta_0 + \delta_1 L_{t-1} + \sigma_t m(\tau)$.

3.7.8 Appendix H: Robustness checks of the empirical application

Figure 3.10: In-sample ES estimates issued from the approximation and the exact calculation method (GARCH(1,1) model)

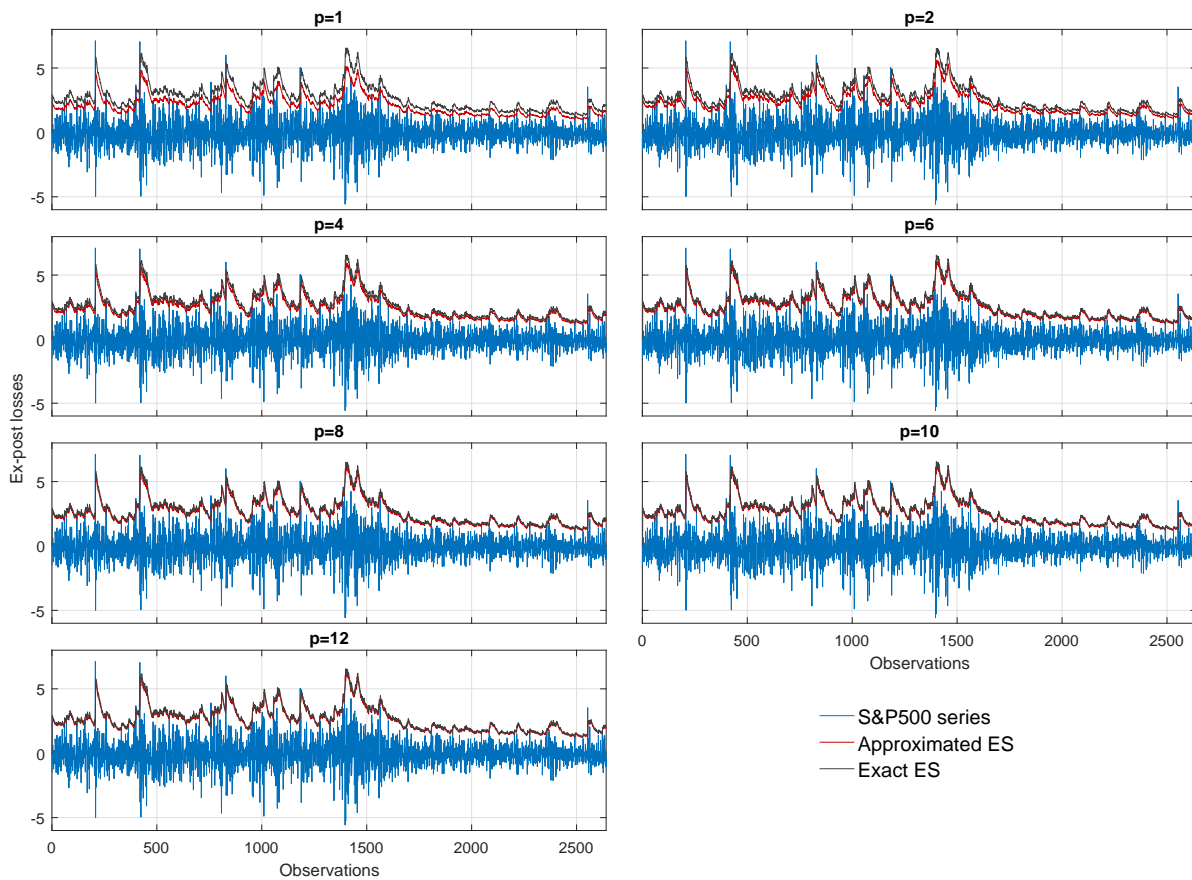


Table 3.3: p-values of the backtesting tests (GARCH(1,1) model)

p	$J_1^{(b)}$	$J_2^{(b)}$	$I^{(b)}$	$S^{(b)}$
Panel A. 2007-2009				
1	0.060	0.082	0.141	0.853
2	0.027	0.054	0.045	0.233
4	0.014	0.048	0.027	0.106
6	0.016	0.046	0.026	0.135
8	0.153	0.053	0.229	0.673
10	0.036	0.067	0.052	0.448
12	0.026	0.058	0.042	0.223
2 (regulatory levels)	0.036	0.064	0.085	0.406
Panel B. 2007-2012				
1	0.199	0.076	0.401	0.839
2	0.011	0.029	0.050	0.458
4	0.005	0.012	0.015	0.196
6	0.008	0.011	0.036	0.317
8	0.029	0.015	0.102	0.563
10	0.010	0.020	0.021	0.360
12	0.012	0.020	0.043	0.415
2 (regulatory levels)	0.005	0.015	0.014	0.230

Note: p-values of the four backtests computed with $p = 1, 2, 4, 6, 8, 10, 12$ 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.

Table 3.4: QML coefficient estimates ($p = 6$, GARCH(1,1) model)

	u_1	u_2	u_3	u_4	u_5	u_6
Panel A. 2007-2009						
β_0	0.600 (0.307)	0.683 * (0.298)	0.769 ** (0.264)	0.811 ** _{ooo} (0.257)	0.972 * _{ooo} (0.446)	1.065 _{ooo} (0.266)
β_1	1.011 (0.093)	0.955 (0.089)	0.917 (0.059)	0.853 ** (0.055)	0.804 (0.143)	0.692 ** _{ooo} (0.043)
<i>joint</i>	*	*	*	**		**
Panel B. 2007-2012						
β_0	0.338 (0.303)	0.388 _{ooo} (0.198)	0.601 ** _{ooo} (0.197)	0.743 *** _{ooo} (0.189)	0.753 ** _{ooo} (0.293)	0.603 (0.799)
β_1	1.025 (0.122)	0.987 (0.069)	0.911 (0.066)	0.860 ** (0.056)	0.829 (0.109)	0.832 (0.308)
<i>joint</i>	*	*	*	**	**	

Note: Standard errors are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5% and 1% level, respectively, and are obtained with the pairs bootstrap algorithm. _o, _{oo}, and _{ooo}, indicate statistical significance at the same levels and are obtained with the procedure of Chernozhukov and Fernández-Val (2011). Panel A gives estimation results for the period 2007-2009 and Panel B provides estimation results for the period 2007-2012.

Figure 3.11: ES forecasts and adjusted ES forecasts over the period 2007-2012 (AR(1)-GARCH(1,1) model)

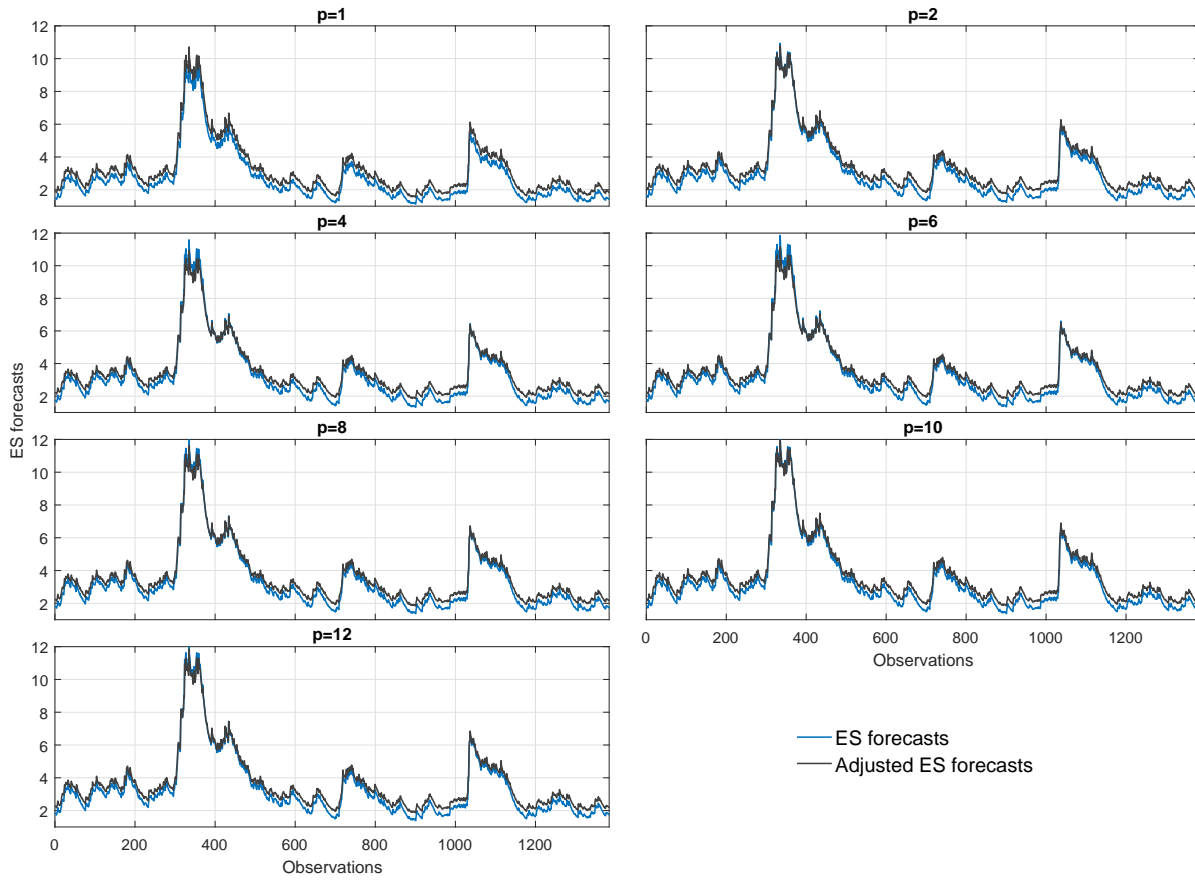


Figure 3.12: ES forecasts and adjusted ES forecasts over the period 2007-2009 (GARCH(1,1) model)

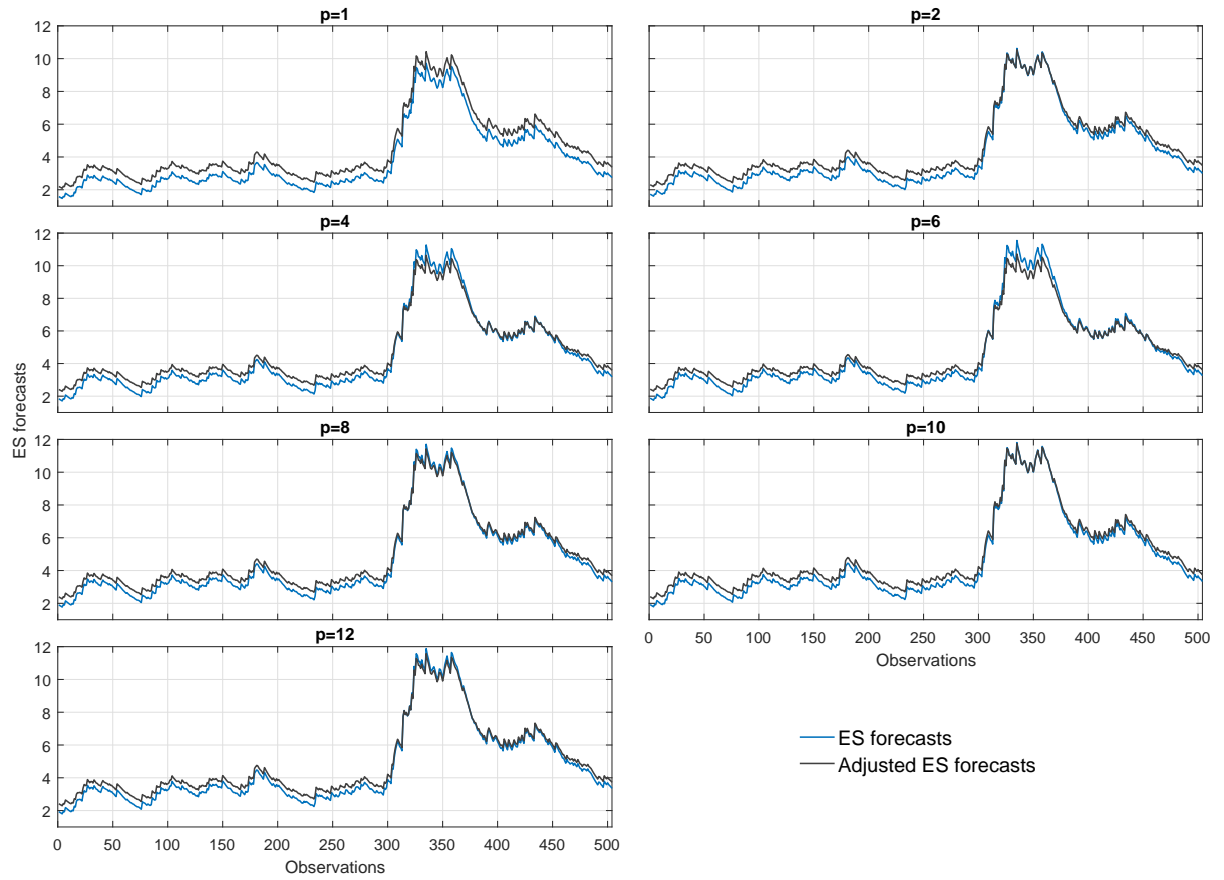


Figure 3.13: ES forecasts and adjusted ES forecasts over the period 2007-2012 (GARCH(1,1) model)

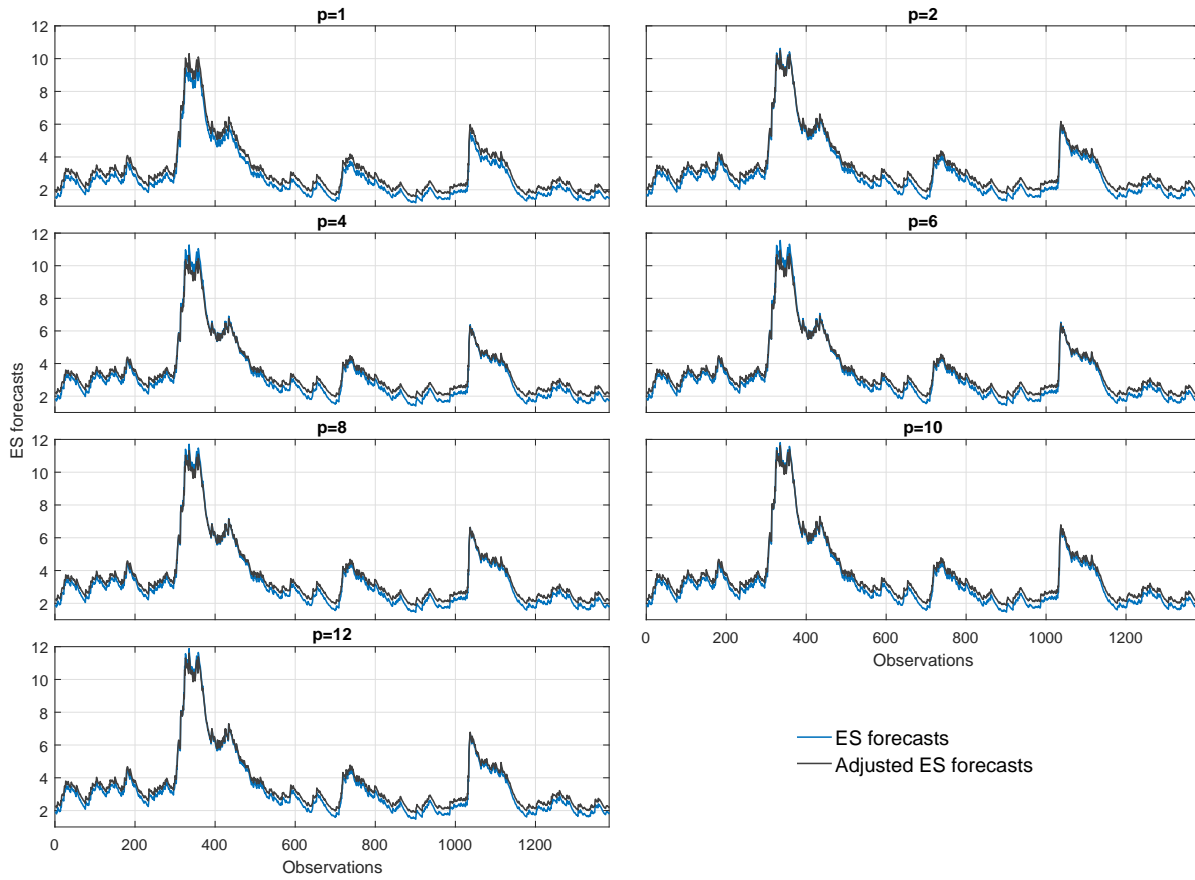
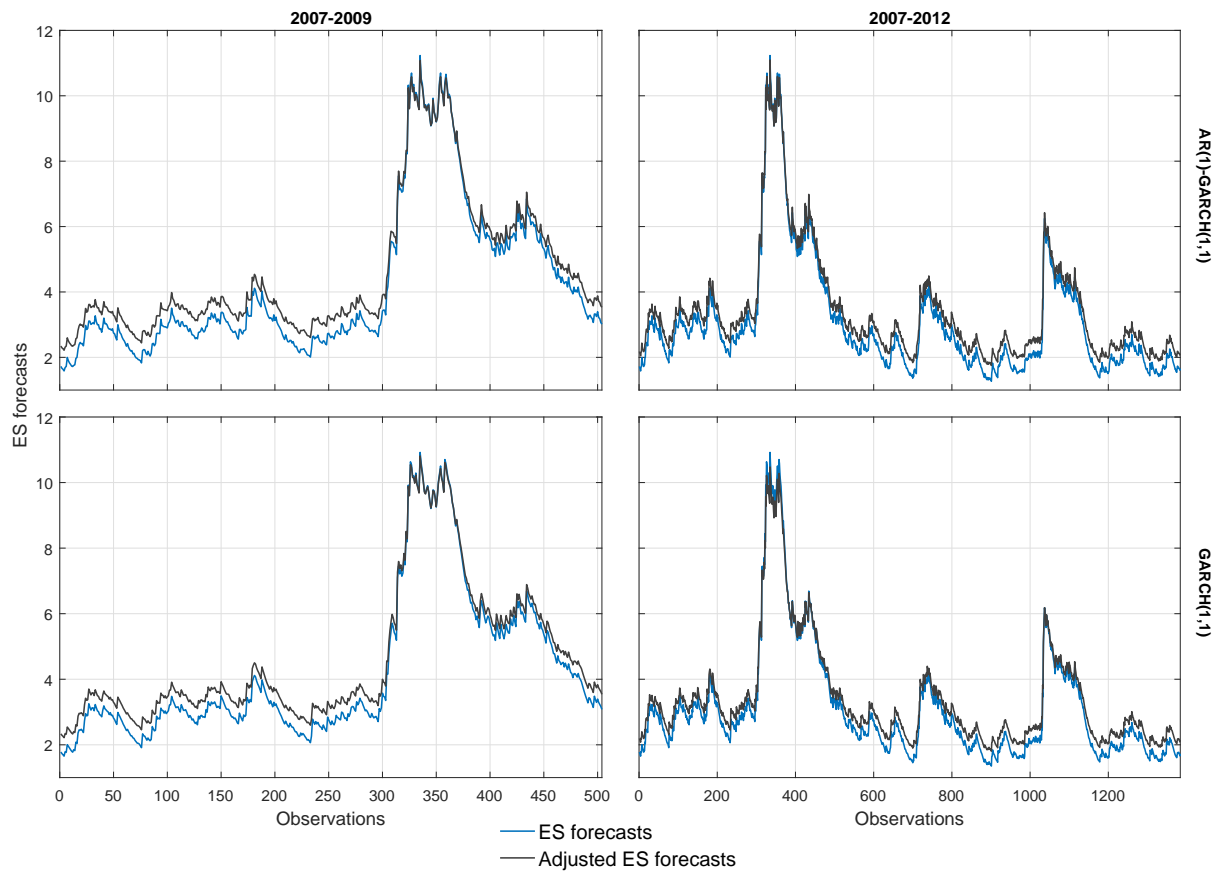


Figure 3.14: ES forecasts and adjusted ES forecasts over the periods 2007-2009 (on the left) and 2007-2012 (on the right) with the two BCBS regulatory risk levels (AR(1)-GARCH(1,1) model, GARCH(1,1) model)



Chapter 4

Elicitability of Marginal Expected Shortfall and Related Systemic-Risk Measures¹

A risk measure, or more generally a statistical functional, is called elicitable if it can be defined as the minimizer of a suitable expected scoring function. This article explores the notion of elicibility (and identifiability) for systemic-risk measures that are used to identify the financial institutions contributing the most to the overall risk in the financial system. Our elicitation framework applies to systemic-risk measures that are expressed as a function of the expected equity loss conditional on a financial crisis, such as the marginal expected shortfall (MES), the systemic expected shortfall (SES), or the systemic-risk measure SRISK. This property paves the way to the implementation of semiparametric M-estimation for the systemic-risk measures or to the comparison and backtesting of the systemic-risk models used by academics and policy makers to rank the systemically important financial institutions (SIFIs) whose failure might trigger a crisis in the entire financial system.

4.1 Introduction

The financial crisis of 2007–2009 and its aftermath has made systemic risk a focal point of research and policy. Very few crisis-related papers made a higher impact both in academia and among regulators than the series of papers proposing global measures of systemic risk based on market data (see Bisias et al., 2012; Benoit et al., 2017, for a survey on systemic-risk measures). The most prominent examples of return-based systemic-risk measures are the marginal expected shortfall (MES) and the systemic ex-

¹This chapter is based on Benoit, Couperier, Leymarie and Scaillet (2022) and has been awarded a research grant sponsored by the *Fondation Banque de France*.

pected 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).² If these timely measures are estimated accurately, they react quickly to market changes, offering added value compared to the current yearly systemic-risk scores used by the Financial Stability Board to identify Global Systemically Important Banks (G-SIBs). For instance, the 2020 list of G-SIBs identifies the same 30 G-SIBs than the year before even if we are experiencing the COVID-19 pandemic crisis. Indeed, the list published last November is based on historical data (as of December 2019).

While the debate on the economic usefulness for regulators of return-based systemic-risk measures is still at play (Idier et al., 2014; Zhang et al., 2015; Löffler and Raupach, 2018; Brownlees et al., 2020), solving the issues on estimation, comparison, and evaluation of these indicators should help to finally put an end to all equivocation. These issues can be solved if the systemic-risk indicators are shown to be elicitable.

The term elicibility is due to Savage (1971), Osband (1985), and Lambert et al. (2008). Elicitability is achieved through the identification of a consistent scoring function for the functional of interest. A scoring function is said to be (strictly) consistent for a given statistical functional (e.g., the mean, median, etc.), if the expected scoring function takes its minimum when the statistical functional is used as the forecast (Gneiting, 2011a). Conversely, we say that a functional is elicitable if there exists a strictly consistent scoring function for it. For example, a scoring function is consistent for the mean if no other quantity than the mean induces a lower expected function. The class of scoring functions that is consistent for the mean is known as the Bregman class (Bregman, 1967) and includes the well-known squared-error function. The class of scoring functions that is consistent for the quantile is known as the generalized piecewise linear class, including the piece-wise linear function for quantile regression (Koenker and Bassett, 1978).

The main contribution of this work is to identify a general class of scoring function which is strictly consistent for return-based systemic-risk measures. This class is defined for the MES of a financial institution and for the Value-at-Risk (VaR) of the financial market, providing a formal framework for the joint elicibility of the bivariate functional (VaR, MES). The class is then extended to any measures expressed as a function of the expected equity loss of a financial institution conditional on a financial crisis, as for the SRISK for instance. As the Expected Shortfall (ES) and the MES can be both defined as a truncated mean, our class of scoring function shares common points with that of Fissler and Ziegel (2016) to elicit the ES and VaR associated with a single asset.

²The three articles by Acharya et al. (2012), Acharya et al. (2017), Brownlees and Engle (2017) that defined the MES, the SES, and the SRISK, have been cited more than 4,000 times since their publication (source: Google Scholar). For online computation of some of these measures, see the Stern-NYU V-Lab initiative website.

Elicitability has been proven to be a necessary axiom for estimation and for comparing and ranking the models' performance (Gneiting, 2011a; Ziegel, 2016; Nolde and Ziegel, 2017; Taylor, 2019, 2020). The literature on the elicibility of systemic-risk measures is rather scant. Fissler et al. (2019) show the elicibility of a particular class of systemic-risk measures where systemic risk is measured by the set of allocations of additional capital that lead to acceptable outcomes (Feinstein et al., 2017). Even if these measures exhibit interesting properties, they are not applied in practice since determining ex-ante acceptable output is quite challenging. More recently, Fissler and Hoga (2021) introduce the notion of multivariate scoring functions that can be applied to the MES, the VaR, and the CoVaR. However, the notion implies a cascade system of scoring functions which has no direct use for M-estimation and (VaR,MES)-model comparison. Fissler and Hoga (2021) focus on getting general results for conditional identifiability and conditional elicibility of functionals of the distribution of the data while we focus on semiparametric M-estimation of the MES and the VaR based on a univariate scoring function. The univariate nature of our scoring function allows the use of the canonical total order on \mathbb{R} . On the contrary, a multivariate scoring function requires the introduction of a lexicographic order, for example on \mathbb{R}^2 in the bivariate case. Using a bivariate scoring function, Fissler and Hoga (2021) show that we can get joint identification for the MES and the VaR. Using a counterexample based on independence between the return of the financial institution and the market return, they show that the MES and the VaR fail to be jointly elicitable. In Section 3, we show that we can restore joint elicibility if we assume instead that the return of the financial institution and the market return are strictly positively related, an empirical feature of our data for banks (see Section 4.5).

Our framework only imposes the introduction of an external parameter into the scoring function, that is the mean of firm return conditional on the market return being equal to its α -VaR. This quantity stems from the first-order derivative (with respect to the market α -VaR) of the MES of the i -th financial institution times the cumulative distribution function (cdf) of the market return. To ensure the strict consistency of our scoring function, this parameter is imposed to be strictly increasing with the market α -VaR.³ Following Andrews (1994); Newey (1994); Ichimura and Lee (2010), we regard this quantity as a nuisance function issued from a criterion function. We apply an infinite dimensional parameter estimation for the nuisance function using the non-parametric isotone estimation of Dette et al. (2006) to ensure strict monotonicity of the conditional mean.

Our work is related to the calculation of moments for multivariate truncated distribution. MES is defined as the first-order moment of a one-sided truncated bivariate

³This assumption suggests that the firm and market are positively correlated which is broadly accepted in the systemic risk literature. For the capital asset pricing model, this assumption is equivalent to assuming a strictly positive beta between firm and market returns.

distribution, and to that extent, this paper suggests avenues in statistical decision theory. Useful results for truncated moments are available for some popular parametric distribution. Rosenbaum (1961) gives explicit formulas to first and second moments of a truncated bivariate standard normal, when the truncation is one-sided and on a single variable. Ang and Chen (2002) extend to two-sided truncated bivariate standard normal and provide test for asymmetries of correlation in different financial market regimes. Nadarajah (2007) calculates moments of one-sided truncated bivariate standard Student t -distribution, while Ho et al. (2012) give general formulas for computing first and second moments of a truncated multivariate t under the double truncation.

We develop semiparametric M-estimation for the MES and the VaR. While the existing estimation methods used for the systemic-risk measures, like MES or SRISK, are issued from fully parametric models or models of second-order moment (e.g., M-GARCH model, dynamic bivariate copula, etc.), our framework imposes minimal distributional assumptions on the joint distribution of asset returns. We present conditions such that the estimated parameters of the VaR and MES models are consistent in the presence of a non-parametric first-step estimator in the criterion function (Chen et al., 2003; Delsol and Van Keilegom, 2020). Then, we characterize the asymptotic normality and we show that the estimated function does affect the asymptotic variance of the estimator (Andrews, 1994; Newey, 1994; Ichimura and Lee, 2010).⁴

Our work also contributes to (i) forecast validation and (ii) forecast comparison for the systemic-risk measures. (i) In the banking industry, forecast validation is known as backtesting. According to Jorion (2007), backtesting is a formal statistical framework that consists in verifying if actual losses are in line with projected losses. Theorem 5 establishes the existence of a strict identification function for the functional of interest (VaR, MES).⁵ Using the identification function and without added difficulty, it is possible to provide backtesting tests similar to those used for the standard market risk measures such as VaR that relies on the so-called violation process (Kupiec, 1995; Christoffersen, 1998; Berkowitz et al., 2011, among others).⁶ (ii) As regards forecast comparison, we exploit the scoring function to rank and compare ex-post the econometric models used to forecast the systemic-risk measures. A necessary condition for a scoring function to consistently rank forecasting models is convexity. Proposition 8 shows that the scoring function is quasi-convex under mild assumptions except when the nuisance function is

⁴The current version of this paper only deals with the case of i.i.d. sample. We leave for a latter version the case of non i.i.d. sample that can be treated under strong mixing conditions for the nuisance function estimate.

⁵Identification functions are known as moment restrictions when considering the generalized method of moments (Newey and McFadden, 1994).

⁶To date, only one formal backtesting procedure for MES-based systemic-risk measures is available (Banulescu-Radu et al., 2021), which typically overcomes the lack of identifiability exploiting the fact that MES can be expressed as an integral of conditional VaR which are themselves identifiable.

not strictly linear and/or not strictly increasing. In practice, the linearity is ensured for most available econometrics models including multivariate conditional volatility models (e.g., M-GARCH-type, realized-volatility-type).

We provide two numerical illustrations. In a simulation study, we illustrate the consistency and asymptotic normality of the proposed semiparametric M-estimator for MES and VaR. We show that our estimator has good finite sample properties and that the statistical inference gives satisfactory results. Interestingly, we show that the finite sample performance of the estimator is affected by the choice of the nuisance parameter estimator. Then in an empirical study, we perform out-of-sample forecast evaluation using the scoring function and we rank competing MES and VaR forecasts. Our study focuses on four major US banks over the period 2000-2016. As in Acharya et al. (2017), we consider time-invariant MES and VaR which we estimate using three competing estimation schemes: fixed, recursive, and rolling. Using the scoring function for MES and VaR, we show that the rolling estimation scheme mainly outperforms the fixed and recursive estimation schemes. Furthermore, we find that the scoring function takes its highest values during the 2007-2009 financial crisis suggesting that the systemic-risk measures are subject to significant model risk in periods of financial distress.

The remainder of the paper is organized as follows. In Section 4.2, we define the MES and introduce the concept of MES-based indicators which include SES and SRISK as special cases. In Section 4.3, we present the class of scoring function dedicated to MES and MES-based indicators. In Section 4.4, we introduce a semiparametric M-estimator for MES and VaR and its corresponding asymptotic properties. In Section 4.5, we run a simulation study to illustrate the finite sample properties of our M-estimator. Then, we perform a short empirical application to illustrate our elicitation framework for forecast comparison. Finally, we conclude the paper in Section 4.6.

4.2 Marginal expected shortfall and related systemic-risk measures

Section 4.2 defines the marginal expected shortfall (hereinafter MES) and discuss its relationship with the other existing systemic-risk measures. In the sequel, we consider the following notations. Let $Y = (Y_M, Y_i)'$ be the vector of two equity returns. Within the systemic risk framework, Y_i denotes the stock return of the i -th financial institution and Y_M denotes the market return. In addition, let us denote by $F_{Y_i}(\cdot)$ and $F_{Y_M}(\cdot)$ the cumulative distribution function (cdf) of Y_i and Y_M , and by $F(\cdot)$ the joint cdf of Y . Finally, we assume that the cdfs are continuous function, and the bivariate random variable Y takes its values in \mathbb{R}^2 .

Following Acharya et al. (2017), we define the MES of a financial firm as its short-run expected equity loss conditional on the market facing a loss greater than its value-at-risk (hereinafter VaR). Formally, the α -level MES of the i -th financial institution is given by,

$$MES_i^\alpha = \mathbb{E} [Y_i | Y_M \leq VaR_M^\alpha], \quad (4.1)$$

where VaR_M^α denotes the α -level VaR of Y_M with $F_{Y_M}(VaR_M^\alpha) = \alpha$ and $\alpha \in (0, 1)$. In (4.1), we use the convention that the VaR corresponds to a lower quantile of the return distribution and has a negative sign for a loss while it receives a positive sign in the banking regulation to compute capital requirements. Accordingly, the MES has also a negative sign for a loss. The MES captures the expected return of the i -th financial institution given that the market is in times of distress, and as such constitutes an indicator of the systemic risk contribution of the i -th institution. In others words, MES measures how the financial institution adds to the financial system overall risk. Interestingly, in the case Y_i and Y_M are independent, the truncated and marginal probability distribution function (pdf) of Y_i are equal, and consequently, the MES simplifies to a non-truncated expectation for Y_i . Reversely, in the case $Y_i = Y_M$, the MES simplifies to the ES. Furthermore, MES can be expressed as the partial derivative of the system ES with respect to the weight of firm i in the economy (Scaillet, 2004).

Several systemic-risk indicators are built-up in analogy with MES in a way that both their formulas and economic interpretations are closely related to it. They represent the expected capital shortfall of the i -th institution when the market falls below its α -level VaR. These similarities with MES comes from the MES being a key constituent of them. It is used to compute the expected market value of the i -th institution given a market decline below a given threshold. This quantity is calculated through the long-run marginal expected shortfall (LRMES). The LRMES is simply a MES defined in terms of cumulative returns after h periods that we define as,

$$MES_{i,t}^\alpha(h) = \mathbb{E}_t \left[\tilde{Y}_{i,t+h} | \tilde{Y}_{M,t+h} \leq \widetilde{VaR}_{M,t+h}^\alpha \right], \quad (4.2)$$

where \mathbb{E}_t denotes the conditional expectation operator with respect to an information set Ω_t , $\tilde{Y}_{t+h} = (\tilde{Y}_{i,t+h}, \tilde{Y}_{M,t+h})'$ is the vector of multi-period arithmetic firm and market returns between $t+1$ and $t+h$, and $\widetilde{VaR}_{M,t+h}^\alpha$ is the corresponding α -VaR of $\tilde{Y}_{M,t+h}$ used as the market threshold.

Definition 4 (MES-type risk measure). *A MES-type risk measure RM is defined as a deterministic function of the MES with $RM_{i,t} = g_t(MES_{i,t}^\alpha(h), X_t)$ where $MES_{i,t}^\alpha(h)$ denotes the MES defined in terms of cumulative returns after h periods as in Equation (4.2), g_t denotes a monotonic function in MES, and X_t represents a set of variables belonging to Ω_t .*

This classification has been proposed by Banulescu-Radu et al. (2021) and many systemic-risk indicators comply with Definition 4. The most prominent are the SRISK (Acharya et al., 2012; Brownlees and Engle, 2017) and the systemic expected shortfall (SES) (Acharya et al., 2017).⁷ In the following, we give a definition of these two indicators. Acharya et al. (2017) show that the SES can be expressed as a linear function of the MES as follows,

$$SES_{i,t} = (kL_{i,t} - 1 + \theta MES_{i,t}^\alpha(h) + \Delta_i) W_{i,t}, \quad (4.3)$$

where $L_{i,t}$ is the leverage ($A_{i,t}/W_{i,t}$), $A_{i,t}$ denotes the total assets, and $W_{i,t}$ is the market capitalization or market value of equity, θ and Δ_i are constant terms. SES represents the amount that a bank equity drops below its target level (defined as a fraction k of assets) conditional on a systemic crisis and when the aggregate capital is less than k times the aggregate assets.

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 of the firm i on day t , denoted $CS_{i,t}$, is defined as the capital reserves the firm needs to hold for regulation and/or prudential management minus the firm's equity, $CS_{i,t} = k(D_{i,t} + W_{i,t}) - W_{i,t}$, with $D_{i,t}$ the book value of debt and k the prudential capital ratio. We follow Acharya et al. (2012) and Brownlees and Engle (2017) to end up with,

$$SRISK_{i,t} = W_{i,t} [k(L_{i,t} - 1) - (1 - k)(1 + MES_{i,t}^\alpha(h))]. \quad (4.4)$$

Equation (4.4) is similar to the definition reported by Brownlees and Engle (2017) in Equation (1) of their paper on 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 according to our Equation (4.1).

4.3 The scoring function

Section 4.3 introduces the scoring function defined for the VaR of the market and the MES of the institution jointly. In Section 4.3.1, we give the class of scoring function and we present the conditions under which the expected score is strictly consistent. In Section 4.3.2, we look at additional desirable properties related to the shape of the expected score and in particular to the quasi-convexity.

⁷Definition 4 also encompasses the Component Expected Shortfall (CES) of Banulescu and Dumitrescu (2015), and the Δ -Conditional Expected Shortfall (Δ CoES) of Ferreira (2018), among others. This definition does not include the class of network systemic-risk measures such as proposed by Billio et al. (2012) or Hué et al. (2019).

4.3.1 Consistency

The MES offers a number of desirable properties. It is easy to show that MES satisfies the properties of monotonicity, sub-additivity, homogeneity, and translational invariance. Hence MES can be recognized as a coherent risk measure (see Artzner et al., 1999; ?). Furthermore, MES is comonotonically additive and law-invariant implying that we can view the MES as a spectral risk measure as proved in 4.7.1 (Kusuoka, 2001; Acerbi, 2002; Tasche, 2002; Jouini et al., 2006).

However, it has been shown that any spectral risk measures fail to be elicitable stand-alone (Ziegel, 2016), the sole exception is the mean. The mean is elicitable stand-alone using the Bregman class of scoring function (Bregman, 1967; Savage, 1971; Banerjee et al., 2005). Therefore, the MES is not elicitable stand-alone which means that there is no scoring function for which the MES uniquely minimizes in expectation. As a spectral risk measure, the ES is also not elicitable stand-alone (see the proof of Theorem 11 in Gneiting, 2011a).

This result has two main consequences. First, modeling the conditional MES given a set of explanatory variables using a regression model is infeasible since the estimation of the regression parameters through an M-estimator requires a suitable scoring function. Second, consistent ranking of competing forecasts for the MES is infeasible. Consequently, to date, there exists no such a regression framework modeling the MES based on a set of covariates nor model comparison methods. Even though the MES is not elicitable stand-alone, we show that the quantile, i.e., the VaR of the market and the MES are jointly elicitable by introducing a class of scoring functions, whose expectation is minimized by these two functionals. In that view, the MES is higher order elicitable as there exists a strictly consistent scoring function for the bi-dimensional functional (VaR, MES) (see Fissler and Ziegel, 2016, for the notion of higher order elicibility).

Theorem 5 (\mathcal{F} -consistent scoring function). *Let \mathcal{F} be a class of distribution functions on \mathbb{R}^2 with finite first moments, unique α -quantiles and continuous densities. Define the set of unknown parameters $\theta = (v, m)'$ associated to the functional $\mathcal{T} = (\text{VaR}_M^\alpha, \text{MES}_i^\alpha)'$, $\mathcal{T} : \mathcal{F} \rightarrow \Theta \subseteq \mathbb{R}^2$ with $\alpha \in (0, 1)$, $\Theta = \Theta_v \times \Theta_m = \{\theta \in \mathbb{R}^2 : m < e(v)\}$, where $e : \Theta_v \rightarrow \mathbb{R}$, $v \mapsto e(v) = \mathbb{E}[Y_i | Y_M = v]$. Then the following assertions are true:*

(i) *The functional \mathcal{T} is 2-elicitable with respect to \mathcal{F} .*

(ii) *Let $S : \Theta \times \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}$ be the scoring function defined by:*

$$S(\theta, y, e(v)) = (G_1(y_M) - G_1(v))(\alpha - \mathbb{1}(y_M \leq v)) - G_2(m) + G(y) + G_2'(m) \left(m + \frac{e(v)}{\alpha} (\mathbb{1}(y_M \leq v) - \alpha) - \frac{1}{\alpha} y_i \mathbb{1}(y_M \leq v) \right), \quad (4.5)$$

where $G : \mathbb{R}^2 \rightarrow \mathbb{R}$ is \mathcal{F} -integrable, $G_1 : \Theta_v \rightarrow \mathbb{R}$ is such that $\mathbb{1}(-\infty, v] G_1$ is \mathcal{F} -integrable for all $v \in \Theta_v$, $G_2 : \Theta_m \rightarrow \mathbb{R}$.

Let $G_3 : \Theta_v \rightarrow \mathbb{R}$ be the functions defined by:

$$G_3(v) = \frac{e(v)}{\alpha} G_2'(m) + G_1(v), \quad (4.6)$$

for all $m \in \Theta_m$. If G_2 is convex and the functions given in Equation (4.6) are increasing, then S is \mathcal{F} -consistent for \mathcal{T} . S is strictly \mathcal{F} -consistent for \mathcal{T} as soon as G_2 is strictly convex and the functions given in Equation (4.6) are strictly increasing.

(iii) The function $V : \Theta \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$ with components

$$\begin{aligned} V_1(\theta, y) &= \alpha - \mathbb{1}(y_M \leq v), \\ V_2(\theta, y) &= m - \frac{1}{\alpha} y_i \mathbb{1}(y_M \leq v), \end{aligned} \quad (4.7)$$

is a strict \mathcal{F} -identification function for \mathcal{T} which satisfies Assumption (A3) (4.7.2). Every strictly \mathcal{F} -consistent scoring function $S : \Theta \times \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}$ for \mathcal{T} which satisfies Assumptions (A5) and (A7) (4.7.2) is necessarily of the form given in Equation (4.5) almost everywhere as soon as Assumptions (A1) and (A6) (4.7.2) are satisfied, G_2 is strictly convex and the functions in Equation (4.6) are strictly increasing.

Theorem 5 presents the main result of the paper. The proof for strict consistency of the loss $S(\theta, y, e(v))$ is reported in 4.7.3. We provide the general class of scoring functions indexed by functions G , G_1 and G_2 that is strictly \mathcal{F} -consistent for the bivariate functional $\mathcal{T} = (VaR_M^\alpha, MES_i^\alpha)'$. The first line in Equation (4.5) refers to the generalized piecewise linear (GPL) function which is strictly consistent with the quantile, namely VaR_M^α . The second and third lines involve both MES_i^α and VaR_M^α through the unknown parameters m and v . It is interesting to note that the scoring function is not separable in m and v , illustrating why MES is only elicitable with VaR.

Assertions (ii) and (iii) of Theorem 5 require that G_2 is strictly convex and the functions of Equation (4.6) are strictly increasing to ensure the strict \mathcal{F} -consistency of S . These conditions entail that G_2 is strictly increasing and strictly convex and G_1 is increasing. These conditions are similar to those considered for the bivariate functional (VaR, ES) in Fissler and Ziegel (2016).

The function e defined in Assertion (ii) represents the conditional expectation of the firm return, Y_i , given that the market return, Y_M , is equal to its α -VaR. This function has a direct interpretation in Fissler and Ziegel (2016) where it is exactly equal to v . Following Equation (4.6), e is a strictly increasing function. We regard this condition as

having a strictly positive beta for the firm return in a CAPM model, which is satisfied with banks equity returns.⁸ Interestingly, if we define the market return y_M as the value weighted average of firm returns (for all the firms that belong to the financial system), then $e(v)$ corresponds to the first-order derivative of v with respect to the firm market share, hence measuring how the financial institution adds to the overall risk of the financial system, through the market VaR.⁹ Here, we must ensure that $m < e(v)$ since we have $MES_i^\alpha = \mathbb{E}[Y_i | Y_M \leq VaR_M^\alpha] < \mathbb{E}[Y_i | Y_M = VaR_M^\alpha] = e(VaR_M^\alpha)$, by definition.

Theorem 5 entails that minimizing the expected loss using any scoring functions S of the form of Equation (4.5) yields MES_i^α and VaR_M^α , such that:

$$(VaR_M^\alpha, MES_i^\alpha) \equiv \arg \min_{\theta \in \Theta} \mathbb{E}[S(\theta, Y, e(v))]. \quad (4.8)$$

The loss function of Theorem 5 is strictly \mathcal{F} -consistent for the functionals VaR and MES. This property remains valid under any MES-type risk measures as given in Definition 4. The proof follows from the continuous mapping theorem that is applied on the continuous and conditionally deterministic function g_t . Consequently, the class of scoring functions S of Equation (4.5) can be used to compute any MES-based systemic-risk measures including SRISK and SES.

Assertion (iii) gives the strict \mathcal{F} -identification function V for $\mathcal{T} = (VaR_M^\alpha, MES_i^\alpha)'$. The function can be directly employed for forecasting validation, often known as backtesting in the financial risk management industry (Jorion, 2007). We recognize, in place of V_1 , the so-called violation process used to backtesting the VaR (see e.g. Kupiec, 1995; Christoffersen, 1998; Berkowitz et al., 2011). The function V_1 is only a function of the unknown parameter v that stands for the VaR. That is why the VaR is conventionally backtested alone using that function. Contrary to V_1 , V_2 is a function of m and v and can be used to backtesting the MES and the VaR jointly. As V_2 is under-identified alone because of the inclusion of two unknown parameters m and v , it is best practice to evaluate V_1 and V_2 jointly to ensure the system is just-identified. Applying same argument as for standard backtesting procedures, the \mathcal{F} -identification function V is a martingale difference sequence and it is possible to invoke unconditional coverage test and independence test for these financial risk measures (see Christoffersen, 2010, for a survey).

⁸In a CAPM model, e has a closed form expression given by $e(v) = \frac{cov(Y_M, Y_i)}{V[Y_M]}v$ and is thus strictly increasing if the covariance between firm and market returns is strictly positive.

⁹The proof is as follows. Let us define $Y_M = \sum_{j=1}^N w_j Y_j$, where w_j and Y_j are weight and return of the j -institution in the financial system, respectively. Applying expectation, we have $v \equiv \mathbb{E}[Y_M | Y_M = v] = \sum_{j=1}^N w_j \mathbb{E}[Y_j | Y_M = v]$. Taking derivative with respect to w_i , i.e., share of the i -institution in the system, it follows $\frac{\partial v}{\partial w_i} = \mathbb{E}[Y_i | Y_M = v] \equiv e(v)$. With a similar reasoning, Scaillet (2004) shows that $\frac{\partial ES_M^\alpha}{\partial w_i} = MES_i^\alpha$, implying that e and MES can be used in risk-based investment strategy to select asset positions in terms of their contribution to the total portfolio risk measured by the VaR and the ES, respectively.

It is worth noting that if we consider a single stock return Y , i.e., $Y_i = Y_M \equiv Y$, then the quantities $MES_i^\alpha = \mathbb{E}[Y_i | Y_M \leq VaR_M^\alpha]$ and $e(v) = \mathbb{E}[Y_i | Y_M = v]$ can be simplified to $ES^\alpha = \mathbb{E}[Y | Y \leq VaR^\alpha]$ and v , respectively. A direct consequence is that the corresponding class of loss functions of Theorem 5 encompasses the loss function of Fissler and Ziegel (2016) dedicated to the bivariate functional $\mathcal{T} = (VaR^\alpha, ES^\alpha)'$, with $\theta = (v, e)'$ the set of unknown parameters.

4.3.2 Quasi-convexity

We now give a number of results related to the quasi-convexity of the scoring function of Equation (4.5). Of interest, we can select G , G_1 , and G_2 such that we obtain a \mathcal{F} -consistent and \mathcal{F} -quasi-convex scoring function for \mathcal{T} . Note that \mathcal{F} -quasi-convexity is a desirable property as it allows ranking competing forecasting models. If the set of competing models under comparison includes the correctly specified model, the ranking induced by a consistent scoring function is self sufficient. However, when comparing misspecified models, the preference order is trickier to establish and the convexity becomes fundamental. Under convexity, the predicted values that are far away from actual values are penalized heavily in comparison to less deviated prediction values. Furthermore, if the scoring function S is \mathcal{F} -quasi-convex, then Equation (4.8) is a quasi-convex optimization problem making the numerical optimization much easier (see e.g., Diewert et al., 1981).

Proposition 8 (\mathcal{F} -consistent and \mathcal{F} -quasi-convex scoring function). *Let \mathcal{F} be a class of distribution functions on \mathbb{R}^2 with finite first moments, unique α -quantiles, continuous densities, and $MES_i^\alpha(F) < 0$ for all $F \in \mathcal{F}$. Then, any function $S : \Theta^- \times \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}$, with $\Theta^- = \Theta_v^- \times \Theta_m^- = \{\theta \in \mathbb{R}^2 : m < e(v) \text{ and } m < 0\}$, where $e : \Theta_v^- \rightarrow \mathbb{R}$, $v \mapsto e(v) = \mathbb{E}[Y_i | Y_M = v]$, which is of equivalent form as ¹⁰*

$$S_0(\theta, y, e(v)) = \frac{1}{m\alpha} \left[(\alpha - \mathbb{1}(y_M \leq v)) e(v) + y_i \mathbb{1}(y_M \leq v) \right] + \ln(-m) \quad (4.9)$$

is a strictly \mathcal{F} -consistent and \mathcal{F} -quasi-convex scoring function for $\mathcal{T} = (VaR_M^\alpha, MES_i^\alpha)'$, where e is a strictly increasing and a strictly linear function of v .

Proposition 8 gives the form of the scoring function that is \mathcal{F} -consistent and \mathcal{F} -quasi-convex for \mathcal{T} . The proof for quasi-convexity of the loss $S_0(\theta, y, e(v))$ is reported in 4.7.3. To ensure the \mathcal{F} -consistency and the \mathcal{F} -quasi-convexity of Equation (4.9) for \mathcal{T} , we assume that e is a strictly increasing and strictly linear function of v . The condition for e is stronger than the one used in Theorem 5 ensuring \mathcal{F} -consistency only. This additional condition might be felt as strong but at least satisfied in asset pricing models where the market is one of the factor. It is also assumed in conditional volatility models such as

¹⁰If S is strictly \mathcal{F} -consistent, then the function defined by $\tilde{S}(\theta, y, e(v)) := \delta S(\theta, y, e(v)) + \varphi(y)$ is also a strictly \mathcal{F} -consistent scoring function for \mathcal{T} , for any δ strictly positive and any \mathcal{F} -integrable function $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$. We say that S and \tilde{S} are of equivalent form (or just equivalent).

multivariate GARCH models that are broadly used in systemic-risk modeling (see e.g., Engle et al., 2015; Brownlees and Engle, 2017; Acharya et al., 2012, etc.). Note that the linearity of e is deduced from the multiplicative form used in GARCH-type process and the Cholesky decomposition that can be applied to the variance-covariance matrix of Y . We report the proof in 4.7.4. Furthermore, we assume that MES_i^α is strictly negative to ensure that the function S_0 is well defined. This is a weak assumption in risk management as the level α is conventionally small (usually, 1% or 5%). Consequently, we restrict the sign of m to be strictly negative so that the function $\ln(-m)$ used in Equation (4.9) is well defined.

The function S_0 of Proposition 8 is a special case of the function S defined in Equation (4.5). More formally, G , G_1 , and G_2 have explicit forms with $G = -1$, $G_1 = 0$, and $G_2(x) = -\ln(-x)$, $\forall x < 0$. The forms are similar to those considered by Patton et al. (2019) for the pair (VaR, ES) who stressed that the resulting scoring function is such that the difference between the losses of two forecasts $S_0(\theta_1, y, e(v_1))$ and $S_0(\theta_2, y, e(v_2))$ is homogeneous of degree zero. It would lead to higher power in the Diebold-Mariano tests (see Diebold and Mariano, 1995; Patton and Sheppard, 2009).¹¹ Similarly, Nolde and Ziegel (2017) show that for equivalent choices, there exists unique functions G_1 and G_2 leading to a difference between two losses that is homogeneous of degree zero.

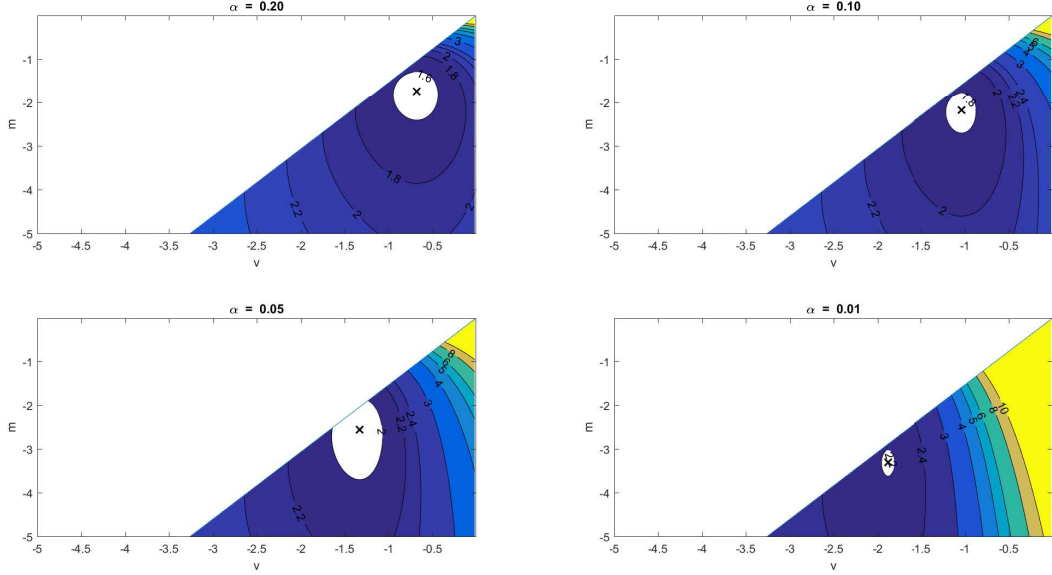
In the sequel, we provide a brief illustration of Proposition 8 through a simple example. We evaluate the expected score $\mathbb{E}[S_0(\theta, Y, e(v))]$ under the normality of Y . We assume $Y \sim \mathcal{N}(\mathbf{0}, \Sigma)$, with $\Sigma = \begin{pmatrix} \sigma_M^2 & \rho_i \sigma_i \sigma_M \\ \rho_i \sigma_i \sigma_M & \sigma_i^2 \end{pmatrix}$, where σ_i , σ_M , and ρ_i , denote the volatility of firm returns, the volatility of market returns, and the correlation between firm and market returns. For the parameter calibration, we consider $\sigma_i = 1.830$, $\sigma_M = 0.809$, and $\rho_i = 0.678$, which are the empirical counterparts of the daily log-return of Bank of America and the CRSP market value-weighted index from January 1, 2012 to December 30, 2016. Under the conditions for G_1 , G_2 , and the normality of Y (ensuring linearity of e), the expected score reads as,

$$\mathbb{E}[S_0(\theta, Y, e(v))] = -\frac{1}{m} \left(\rho_i \sigma_i \frac{\phi(v/\sigma_M)}{\alpha} - v \frac{\rho_i \sigma_i}{\sigma_M} \left(1 - \frac{\Phi(v/\sigma_M)}{\alpha} \right) \right) + \ln(-m), \quad (4.10)$$

where ϕ and Φ denote the pdf and cdf of the standard normal distribution.

Figure 4.1 gives the contours of $\mathbb{E}[S_0(\theta, Y, e(v))]$ as defined in (4.10) considering a grid for m and v . Only values where $m < e(v)$ are considered. The true MES_i^α and VaR_M^α are marked with a black cross and we verify that the minimum of $\mathbb{E}[S_0(\theta, Y, e(v))]$ is attained

¹¹Homogeneity of degree 0 of a function implies that the value of this function is unchanged if we multiply by a strictly positive constant both the competing forecasts and the observations. Consequently, the ranking of the competing forecasts is invariant by a change in the measurement unit.

Figure 4.1: Contours of the expected loss $\mathbb{E}[S_0(\theta, Y, e(v))]$ ($\alpha = 0.20, 0.10, 0.05, 0.01$)

at MES_i^α and VaR_M^α . Two remarks should be raised. First, the minimum of the expected loss is moving from the upper-right side to the lower-left side of the contour plot as α goes to zero. With our set of calibrated parameters, we have $(VaR_M^{0.20}, MES_i^{0.20}) = (-0.6809, -1.7368)$, $(VaR_M^{0.10}, MES_i^{0.10}) = (-1.0368, -2.1775)$, $(VaR_M^{0.05}, MES_i^{0.05}) = (-1.3307, -2.5593)$, $(VaR_M^{0.01}, MES_i^{0.01}) = (-1.8820, -3.3068)$. Formally, as we consider a positive correlation between the firm and market returns, lowering the probability level α leads VaR_M^α to decrease and MES_i^α as well. Second, we observe that the contour plots suggest that the level sets of the expected loss function are boundaries of convex sets, which supports Proposition 8. This property is helpful to enable consistent ranking of forecasting models.

4.4 M-estimation

This Section uses the scoring function for MES_i^α and VaR_M^α to do M-estimation. We introduce a semiparametric M-estimator $\hat{\theta} = (\hat{v}, \hat{m})'$ for VaR and MES denoted by $\theta_0 = (v_0, m_0)'$ hereinafter. We show that $\hat{\theta}$ is consistent and asymptotically normally distributed under some assumptions. To establish the asymptotic theory, we make use of a set of results and proofs for semiparametric M-estimators involving the presence

of an estimated nuisance parameter. We consider Chen et al. (2003) and Delsol and Van Keilegom (2020) for the consistency and Andrews (1994), Newey and McFadden (1994), and Ichimura and Lee (2010) for the asymptotic normality.

For a sake of completeness, we introduce additional notations at the estimation step. We denote by $\Theta (\equiv (\Theta_v \times \Theta_m)) \subseteq \mathbb{R}^2$ a finite dimensional parameter space. We denote by \mathcal{E} an infinite dimensional parameter set defined as the space of smooth (twice differentiable) strictly increasing functions. Thus, $\theta_0 \in \Theta$ and $e : \Theta_v \rightarrow \mathcal{E}; v \mapsto e(v) = \mathbb{E} [Y_i | Y_M = v]$ denote the true unknown finite and infinite dimensional parameters of the estimation problem, respectively. We assume that the function $e \in \mathcal{E}$ depends on the unknown parameter v . For a sake of simplicity, we will consider the notation $(\theta, e) \equiv (\theta, Y, e(v))$ hereafter.

Denote by S a non-random measurable vector-valued function $S : \Theta \times \mathbb{R}^2 \times \mathcal{E} \rightarrow \mathbb{R}$ verifying,

$$\theta_0 = \arg \min_{\theta \in \Theta} \bar{S}(\theta, e), \quad (4.11)$$

with $\bar{S}(\theta, e) = \mathbb{E} [S(\theta, Y, e(v))]$. Assume the sets Θ and \mathcal{E} are metric spaces and denote by d and $d_{\mathcal{E}}$ their respective metrics. We can consider the usual L_1 (Manhattan) or L_2 (Euclidean) norms for d and $d_{\mathcal{E}}$.

Assume that the observed data $\{Y_t\}_{t=1}^T$ are a random sample of Y . For each v , $\hat{e}(v)$ is the non-parametric estimator of $e(v)$ on that sample. The nuisance parameter being allowed to depend on v , we implicitly define $d_{\mathcal{E}}(e, \hat{e})$ uniformly over v , i.e. $d_{\mathcal{E}}(e, \hat{e}) := \sup_{v \in \Theta_v} d_{\mathcal{E}}^1(e(v), \hat{e}(v))$ for some metric $d_{\mathcal{E}}^1$. We estimate θ_0 by $\hat{\theta} \in \Theta$. A natural sample analog estimator of θ_0 , that exploits the scoring function defined in Theorem 5, is an M-estimator that minimizes $S_T(\theta, \hat{e}) = T^{-1} \sum_{t=1}^T S(\theta, y_t, \hat{e}(v))$. The resulting estimator $\hat{\theta}$ is defined such that,

$$\hat{\theta} = \arg \min_{\theta \in \Theta} S_T(\theta, \hat{e}). \quad (4.12)$$

As for many criterion functions (e.g., check function), the criterion function $S_T(\theta, \hat{e})$ is not differentiable (at $Y_M = v$). Simulated annealing, particle swarm, interior-point, and simplex methods are suitable solvers for the estimator in Equation (4.12) as they do not rely on derivative calculation.

Theorem 6 (Consistency of $\hat{\theta}$). *Under Theorem 5 and Assumptions (B1)-(B4) given in 4.7.2, the M-estimator of θ_0 defined in (4.12) is weakly consistent, i.e., we have $\hat{\theta} \rightarrow \theta_0$, in probability.*

Theorem 6 states that $\hat{\theta}$ is a consistent estimator of θ_0 . We display the proof in 4.7.3. To establish the proof, we use Assumptions (B1)-(B4) given in 4.7.2, which are simple and primitive conditions in order to exploit the high-level arguments of Chen et al. (2003) and Delsol and Van Keilegom (2020). The stated result highlights that, even though the true function e is unknown and replaced by a non-parametric consistent estimator \hat{e} to make

the criterion minimization feasible, it does not affect the consistency. Then, the estimator in (4.12) provides consistent estimates of the two risk measures MES_i^α and Var_M^α . Here, MES_i^α and Var_M^α are time-invariant. It differs from the time-varying setting of Patton et al. (2019) where the semiparametric estimation targets the parameters driving the specification of the dynamics of the risk measures.

A particular attention must be paid on the choice of the non-parametric estimator \hat{e} of e . In Assumption (B2) of 4.7.2, we indicate that \hat{e} must share the same conditions of the parameter space than those defined for e in Theorem 5 in order to guarantee the strict \mathcal{F} -consistency of the scoring function and then the validity of the estimator in (4.12). Consequently, \hat{e} must be a strictly increasing function of v and we cannot use standard kernel regression because this estimator does not ensure that \hat{e} is strictly increasing in v . As a result, we resort on isotone kernel estimators (Brunk, 1955; Hall and Huang, 2001; Dette et al., 2006, etc.).

To go a step further, we investigate the limiting distribution of the estimator $\hat{\theta}$. Andrews (1994), Newey and McFadden (1994), and Ichimura and Lee (2010) focus on two-step semiparametric estimators, where the first-stage involves some unknown infinite dimensional parameters. We can apply Theorem 3.2 in Ichimura and Lee (2010) developed for the i.i.d. setting under additional regularity assumptions than the ones listed in 4.7.2 to obtain the asymptotic normality of the estimator $\hat{\theta}$ in a time-series setting. Here, we do not provide a formal theorem of the limiting distribution under a suitable list of assumptions and skip its proof. Still, in 4.7.5, we provide the computational steps to characterize the asymptotic variance of the limiting Gaussian distribution. Assume that Ω_0 exists. Then,

$$\sqrt{T} \left(\hat{\theta} - \theta_0 \right) \xrightarrow{d} \mathcal{N} \left(0, V_0^{-1} \Omega_0 V_0^{-1} \right), \quad (4.13)$$

with V_0 the Hessian matrix of $\bar{S}(\theta, e)$ with respect to θ , evaluated at $\theta = \theta_0$,

$$V_0 = \begin{pmatrix} f_{Y_M}(v_0) \left(G_1'(v_0) + \frac{G_2'(m_0)e'(v_0)}{\alpha} \right) & 0 \\ 0 & G_2''(m_0) \end{pmatrix}, \quad (4.14)$$

and with Ω_0 defined by:

$$\Omega_0 = \lim_{T \rightarrow \infty} V \left[\frac{1}{T} \sum_{t=1}^T \Gamma_0(Y_t) \right],$$

$$\Gamma_0(y) = \begin{pmatrix} \left(G_1'(v_0) + \frac{G_2'(m_0)e'(v_0)}{\alpha} \right) \left(\mathbb{1}(y_M \leq v_0) - \alpha \right) + \frac{G_2'(m_0)}{\alpha} (e(v_0) - y_i) (\delta_{Y_M}(v_0) - f_{Y_M}(v_0)) \\ G_2''(m_0) \left(m_0 - e(v_0) + \frac{\mathbb{1}(y_M \leq v_0)}{\alpha} (e(v_0) - y_i) \right) \end{pmatrix},$$

where $\delta_{Y_M}(v)$ corresponds to the dirac function of Y_M evaluated at v . We can estimate Ω_0 by a standard HAC estimator (Newey and West, 1987).

As highlighted in Equation (4.13), the estimator $\hat{\theta}$ is asymptotically normally distributed. The matrix V_0 in Equation (4.14) corresponds to the Hessian matrix of the expected scoring function with respect to θ evaluated at the true parameter θ_0 . It is worth noting that its first diagonal element equals $f_{Y_M}(v)$ times the first-order derivative of G_3 with respect to v (see Theorem 5). The latter is strictly positive since G_3 is supposed to be a strictly increasing function of v . Furthermore, the second element is also strictly positive as G_2 is strictly increasing and strictly convex. Thus, V_0 is positive definite.

Furthermore, the expression for V_0 is importantly simplified compared to the one obtained in Ichimura and Lee (2010). Surprisingly, our expression is equivalent to the one we would have obtained if the nuisance parameter was not dependent on θ , i.e., $e(v) \equiv e$. When the nuisance parameter is a function of θ , the general formula is given in Equation (3.6) of Theorem 3.2 in Ichimura and Lee (2010). If we compare that formula with ours, only the first summand is different from zero and V_0 becomes the Hessian matrix when e is not a function of v . The other terms are here zeros because they can all be expressed as a product of $F_{Y_M}(v) - \alpha$. Because $F_{Y_M}(v) - \alpha$ is also the expectation of the identification function for the parameter v , i.e., $\mathbb{E}[V_1(\theta, Y)]$ in Equation (4.7), its value is zero when evaluated at θ_0 and vanishes asymptotically when evaluated at $\hat{\theta}$.

4.5 Numerical illustrations

Section 4.5 provides real-life applications either dedicated to MES and VaR estimation or to the comparison of competing MES and VaR forecasts. In Section 4.5.1, we study the finite sample performance of the M-estimator of Section 4.4. In Section 4.5.2, we build MES and VaR forecasts on real data and give out-of-sample performance comparisons for three competing forecasting approaches.

4.5.1 Simulation study

In what follows, we run a simulation experiment to illustrate the consistency and asymptotic normality of the M-estimator introduced in (4.12) with the corresponding Theorem 6 and Equation (4.13).

4.5.1.1 The experiment

As in Acharya et al. (2017), we focus on a time-invariant MES. We consider a bivariate normal distribution for the daily demeaned return, $Y_t = (Y_{M,t}, Y_{i,t})'$, such that $Y_t = \Sigma^{1/2} z_t$, with $Y_{M,t}$ the market return and $Y_{i,t}$ the firm return, and where z_t denotes an i.i.d. Gaussian vector error process with $\mathbb{E}[z_t] = \mathbf{0}$ and $\mathbb{E}[z_t z_t'] = I_2$. We define the time-

invariant variance-covariance matrix Σ as,

$$\Sigma = \begin{pmatrix} \sigma_M^2 & \rho_i \sigma_M \sigma_i \\ \rho_i \sigma_M \sigma_i & \sigma_i^2 \end{pmatrix},$$

where σ_M^2 and σ_i^2 represent the volatility of market and firm returns, and ρ_i is the correlation between both returns. The parameters of the covariance matrix are calibrated using the daily log-returns of Bank of America and the CRSP market value-weighted index over the period January 1, 2012 to December 30, 2016. The estimated values for unconditional variances of market and firm returns and for the unconditional correlation are equal to 0.722, 3.506, and 0.663, respectively.

The simulation procedure is as follows. For each Monte Carlo replication $b = 1, \dots, B$, we simulate the series $\{Y_{M,t}^{(b)}, Y_{i,t}^{(b)}\}_{t=1}^T$ using the time-invariant specification. Then, we use the resulting simulated time-series to estimate $\theta_0 = (v_0, m_0)'$ using the M-estimator introduced in (4.12) and we get $\hat{\theta}^{(b)} = (\hat{v}^{(b)}, \hat{m}^{(b)})'$. We also estimate the asymptotic variance-covariance matrix of $\hat{\theta}^{(b)}$ defined in Equation (4.13) using its sample analog. We consider $B = 1000$ replications and each time the estimation procedure is run for sample sizes $T = 500, 2500$ and probability levels $\alpha = 0.01, 0.05, 0.10, 0.20$.

As discussed in Section 4.4, we need a consistent estimator of the nuisance parameter e_0 to guaranty the consistency of $\hat{\theta}$. We consider the isotone non-parametric estimator proposed by Dette et al. (2006). This estimator is a strictly increasing function of v ensuring the validity of Theorem 6 and Equation (4.13) for the asymptotic properties of $\hat{\theta}$. We consider two different kernels for the non-parametric estimation, namely the Gaussian and Epanechnikov density. We set the bandwidths of the kernels using the Silverman rule of thumb. For a sake of comparison, we estimate the nuisance parameter using a simple plug-in OLS estimation which also delivers a consistent estimate of θ_0 . Indeed, the considered time-invariant representation implies that e is a linear function of v .

4.5.1.2 Simulation results

Tables 4.1 and 4.2 report the simulation results of \hat{m} and \hat{v} for $\alpha = 0.01, 0.05, 0.10, 0.20$ and the sample sizes $T = 500$ and $T = 2500$, respectively. The first and second columns display the estimation results using the non-parametric isotone estimation of e while the last column is for the OLS estimation of e . In the row [1] of each panel, we report the true values of the parameters. Then, the rows [2] and [3] report the median parameter estimates and the average bias of the estimated parameters computed across simulations.

Table 4.1: Simulation results for the normal time-invariant MES and VaR ($T = 500$)

		Non-parametric isotone				OLS	
		Gaussian		Epanechnikov		m	v
		m	v	m	v	m	v
$\alpha = 0.01$							
[1]	True	-3.307	-1.882	-3.307	-1.882	-3.307	-1.882
[2]	Median	-3.110	-1.868	-3.299	-1.870	-2.894	-1.825
[3]	Avg bias	0.242	0.029	0.085	0.006	0.452	0.013
[4]	Emp variance	0.235	0.038	0.163	0.038	0.404	0.089
[5]	Asy variance	0.533	0.041	0.521	0.044	0.759	0.026
[6]	Emp covariance	0.038	-	0.029	-	0.119	-
[7]	Asy covariance	0.047	-	0.052	-	0.033	-
[8]	Emp coverage	0.865	0.960	0.945	0.965	0.863	0.960
[9]	Asy coverage	0.897	0.923	0.789	0.930	0.778	0.898
$\alpha = 0.05$							
[1]	True	-2.559	-1.331	-2.559	-1.331	-2.559	-1.331
[2]	Median	-2.442	-1.355	-2.500	-1.352	-2.398	-1.316
[3]	Avg bias	0.116	-0.011	0.055	-0.017	0.167	0.008
[4]	Emp variance	0.076	0.022	0.070	0.018	0.083	0.033
[5]	Asy variance	0.096	0.034	0.099	0.031	0.103	0.022
[6]	Emp covariance	0.014	-	0.011	-	0.020	-
[7]	Asy covariance	0.022	-	0.021	-	0.017	-
[8]	Emp coverage	0.937	0.978	0.968	0.981	0.913	0.970
[9]	Asy coverage	0.924	0.974	0.961	0.974	0.904	0.896
$\alpha = 0.10$							
[1]	True	-2.178	-1.037	-2.178	-1.037	-2.177	-1.037
[2]	Median	-2.100	-1.059	-2.100	-1.067	-2.102	-1.030
[3]	Avg bias	0.082	-0.019	0.070	-0.021	0.084	0.005
[4]	Emp variance	0.047	0.016	0.043	0.017	0.056	0.232
[5]	Asy variance	0.049	0.027	0.049	0.026	0.054	0.016
[6]	Emp covariance	0.006	-	0.008	-	0.020	-
[7]	Asy covariance	0.011	-	0.011	-	0.010	-
[8]	Emp coverage	0.918	0.971	0.943	0.981	0.922	0.947
[9]	Asy coverage	0.910	0.980	0.934	0.967	0.930	0.914
$\alpha = 0.20$							
[1]	True	-1.737	-0.681	-1.737	-0.681	-1.737	-0.681
[2]	Median	-1.681	-0.717	-1.680	-0.719	-1.687	-0.679
[3]	Avg bias	0.050	-0.044	0.054	-0.038	0.056	0.003
[4]	Emp variance	0.028	0.013	0.025	0.011	0.023	0.012
[5]	Asy variance	0.025	0.021	0.025	0.020	0.025	0.014
[6]	Emp covariance	0.004	-	0.004	-	0.005	-
[7]	Asy covariance	0.004	-	0.004	-	0.004	-
[8]	Emp coverage	0.918	0.947	0.937	0.974	0.946	0.952
[9]	Asy coverage	0.913	0.986	0.927	0.982	0.937	0.902

Note: This table reports the simulation results of \hat{m} and \hat{v} for a sample size $T = 500$. The first and second columns give the estimation results using the non-parametric isotone estimation for e with Gaussian and Epanechnikov kernels, respectively. The third column uses OLS estimation for e . Row [1] of each panel displays the values of the true parameters. Rows [2] and [3] present the median estimated parameters and the average bias of the estimated parameters. Rows [4] and [5] report the empirical variances of \hat{m} and \hat{v} (across simulations) and the resulting estimated asymptotic variances of the estimators. Rows [6] and [7] report the same results for the covariance between \hat{m} and \hat{v} . Rows [8] and [9] give the coverage rates for 95% confidence intervals built using the empirical variances and the asymptotic variances, respectively.

Table 4.2: Simulation results for the normal time-invariant MES and VaR ($T = 2500$)

		Non-parametric isotone				OLS	
		Gaussian		Epanechnikov		m	v
		m	v	m	v	m	v
$\alpha = 0.01$							
[1]	True	-3.307	-1.882	-3.307	-1.882	-3.307	-1.882
[2]	Median	-3.200	-1.896	-3.260	-1.901	-3.132	-1.861
[3]	Avg bias	0.113	-0.007	0.059	-0.010	0.191	0.017
[4]	Emp variance	0.061	0.016	0.044	0.015	0.072	0.025
[5]	Asy variance	0.093	0.021	0.095	0.020	0.120	0.013
[6]	Emp covariance	0.009	-	0.006	-	0.018	-
[7]	Asy covariance	0.019	-	0.019	-	0.016	-
[8]	Emp coverage	0.935	0.972	0.978	0.974	0.890	0.978
[9]	Asy coverage	0.951	0.956	0.977	0.947	0.901	0.859
$\alpha = 0.05$							
[1]	True	-2.559	-1.331	-2.559	-1.331	-2.559	-1.331
[2]	Median	-2.510	-1.355	-2.511	-1.352	-2.509	-1.330
[3]	Avg bias	0.049	-0.024	0.039	-0.018	0.052	0.007
[4]	Emp variance	0.015	0.008	0.017	0.007	0.010	0.015
[5]	Asy variance	0.018	0.013	0.018	0.012	0.020	0.008
[6]	Emp covariance	0.002	-	0.002	-	0.005	-
[7]	Asy covariance	0.005	-	0.005	-	0.005	-
[8]	Emp coverage	0.947	0.971	0.945	0.976	0.928	0.945
[9]	Asy coverage	0.931	0.978	0.940	0.985	0.937	0.913
$\alpha = 0.10$							
[1]	True	-2.178	-1.037	-2.178	-1.037	-2.177	-1.037
[2]	Median	-2.150	-1.058	-2.140	-1.059	-2.152	-1.030
[3]	Avg bias	0.032	-0.018	0.032	-0.021	0.029	0.005
[4]	Emp variance	0.009	0.006	0.009	0.006	0.009	0.008
[5]	Asy variance	0.010	0.009	0.009	0.009	0.010	0.007
[6]	Emp covariance	0.001	-	0.001	-	0.002	-
[7]	Asy covariance	0.002	-	0.002	-	0.002	-
[8]	Emp coverage	0.939	0.974	0.932	0.973	0.944	0.941
[9]	Asy coverage	0.932	0.986	0.933	0.975	0.949	0.926
$\alpha = 0.20$							
[1]	True	-1.737	-0.681	-1.737	-0.681	-1.737	-0.681
[2]	Median	-1.720	-0.698	-1.720	-0.698	-1.714	-0.678
[3]	Avg bias	0.018	-0.018	0.019	-0.021	0.019	0.004
[4]	Emp variance	0.005	0.005	0.005	0.005	0.005	0.005
[5]	Asy variance	0.005	0.007	0.005	0.007	0.005	0.006
[6]	Emp covariance	0.001	-	0.001	-	0.001	-
[7]	Asy covariance	0.001	-	0.001	-	0.001	-
[8]	Emp coverage	0.951	0.973	0.953	0.979	0.941	0.943
[9]	Asy coverage	0.946	0.994	0.948	0.980	0.939	0.949

Note: This table reports the simulation results of \hat{m} and \hat{v} for a sample size $T = 2500$. The first and second columns give the estimation results using the non-parametric isotone estimation for e with Gaussian and Epanechnikov kernels, respectively. The third column uses OLS estimation for e . Row [1] of each panel displays the values of the true parameters. Rows [2] and [3] present the median estimated parameters and the average bias of the estimated parameters. Rows [4] and [5] report the empirical variances of \hat{m} and \hat{v} (across simulations) and the resulting estimated asymptotic variances of the estimators. Rows [6] and [7] report the same results for the covariance between \hat{m} and \hat{v} . Rows [8] and [9] give the coverage rates for 95% confidence intervals built using the empirical variances and the asymptotic variances, respectively.

Two remarks should be raised. First, we observe that the parameters m_0 and v_0 are estimated without bias, except for the smallest probability level $\alpha = 0.01$ and sample size $T = 500$ where the estimation is more unstable due to a lack of observations. As expected, the observed bias vanishes for more central probability levels and when the sample size increases (from $T = 500$ to $T = 2500$) confirming numerically that the parameters are estimated consistently. Second, we note that the non-parametric isotone estimation with an Epanechnikov kernel gives the best finite sample performances. We find that even for small probability levels, the bias is very limited for that isotone estimator of the nuisance parameter. Hence, the choice of a nuisance parameter estimator over competitors affects the finite sample performance of the M-estimator.

In the subsequent rows of Tables 4.1 and 4.2, we report the simulation results of variances and covariances for \hat{m} and \hat{v} . Rows [4] and [5] display the empirical variances of \hat{m} and \hat{v} (computed across simulations) and the estimated asymptotic variances of \hat{m} and \hat{v} issued by Equation (4.13), respectively. Rows [6] and [7] report the same information but for the covariance between \hat{m} and \hat{v} . First, we find that the theoretical variances and covariance that are estimated with the sample analogs are close to the empirical variances and covariance which illustrates the validity of our asymptotic theory. Second, we find that the estimated variances are decreasing with the sample size and increasing as we move further into market tail events (i.e., as α decreases), as expected.

The last rows [8] and [9] report statistical inference where we print the coverage rates. A coverage rate corresponds to the relative frequency of simulations for which the true parameter lies in the confidence interval. Row [8] reports the coverage rates for 95% confidence intervals constructed using the empirical variances of \hat{m} and \hat{v} . Row [9] gives the feasible coverage rates for 95% confidence intervals using the theoretical estimated variances according to Equation (4.13). Overall, the results for statistical inference are satisfactory. We find that the coverage rates calculated with the empirical

and theoretical variances are very close to 95% which suggests that our asymptotic theory is valid. However, our simulation results are more shaded for the smallest probability level $\alpha = 0.01$ and sample size $T = 500$. We find that the coverage rates are below 95% implying that statistical tests on parameters would be over-sized and a correction would be possibly needed.

4.5.2 Empirical application

In the sequel, we illustrate the merit of the scoring function in (4.9) for out-of-sample forecasting evaluation. We build the daily MES and VaR forecasts of four major US banks and we compute average scoring functions to provide a comparison across three competing forecasting approaches. Section 4.5.2.1 describes the empirical setup. Then, Section 4.5.2.2 reports our main results.

4.5.2.1 Empirical setup

As in Acharya et al. (2017), we consider a simple time-invariant MES at probability level $\alpha = 0.05$. We assume that the daily firm and market returns $Y_t = (Y_{M,t}, Y_{i,t})'$, for $t = 1, \dots, T$, are normally distributed. Under the normality assumption, the MES for the i -bank return and the VaR of the market take closed-form expressions (Brownlees and Engle, 2017) given by $MES_{i,t}^\alpha = -\rho_{i,t}\sqrt{\sigma_{i,t}^2}\lambda(\Phi^{-1}(\alpha))$ and $VaR_{M,t}^\alpha = \sqrt{\sigma_{M,t}^2}\Phi^{-1}(\alpha)$. The term $\lambda(x)$ corresponds to the Mills ratio satisfying $\lambda(x) = \phi(x)/\Phi(x)$ where ϕ and Φ denote the pdf and cdf of the standard normal distribution, $\sigma_{i,t}^2$ and $\sigma_{M,t}^2$ denote the variance at time t of $Y_{i,t}$ and $Y_{M,t}$, and $\rho_{i,t}$ is the correlation at time t between $Y_{i,t}$ and $Y_{M,t}$. Given our statistical assumptions, the MES solely depends on the volatility of the firm and its correlation with the market but not on the volatility of the market $\sigma_{M,t}^2$. To ease the notation, we write m_t for $MES_{i,t}^\alpha$ and v_t for $VaR_{M,t}^\alpha$ in the following.

As we consider time-invariant MES and VaR, the parameters $\sigma_{i,t}^2$, $\sigma_{M,t}^2$, and $\rho_{i,t}$ are constant for all t . We estimate them using their empirical counterparts $\hat{\rho}_{i,t} =$

$\frac{\sum_{t=1}^T (Y_{i,t} - \bar{Y}_i)(Y_{M,t} - \bar{Y}_M)}{\sqrt{\sum_{t=1}^T (Y_{i,t} - \bar{Y}_i)^2} \sqrt{\sum_{t=1}^T (Y_{M,t} - \bar{Y}_M)^2}}$, $\hat{\sigma}_{i,t}^2 = T^{-1} \sum_{t=1}^T (Y_{i,t} - \bar{Y}_i)^2$, and $\hat{\sigma}_{M,t}^2 = T^{-1} \sum_{t=1}^T (Y_{M,t} - \bar{Y}_M)^2$, where \bar{Y}_i and \bar{Y}_M denotes the empirical means of $Y_{i,t}$ and $Y_{M,t}$, respectively, and T corresponds to the considered sample size.

To provide meaningful out-of-sample forecast evaluation, we need to perform a comparison of MES and VaR predictions across different competing approaches. To that end, we select three types of estimation schemes: (i) a fixed estimation scheme, (ii) a recursive estimation scheme, (iii) a rolling-window estimation scheme. Thus, the parameters $\sigma_{i,t}^2$, $\sigma_{M,t}^2$, and $\rho_{i,t}$ – used to compute the MES and VaR forecasts – are estimated using different sample lengths and periods. In the fixed scheme, we estimate the parameters using all available information from January 3, 2000, to December 31, 2004 once and for all. Hence, the MES and VaR forecasts are time-invariant along the sample period. In the recursive scheme, we estimate the parameters using all available information from January 3, 2000, up to the end of each month. In the rolling-window scheme, we also update the forecasts up to the end of each month but we only use the most recent T daily observations for estimation. In the following, we will consider a two-year sample length, i.e. $T = 500$.

The out-of-sample MES and VaR forecasts $\{\hat{v}_t, \hat{m}_t\}_{t=T+1}^{T+n}$ issued from these three competing approaches are subsequently assessed through the computation of the empirical scoring function,

$$S_n(\theta, \hat{\beta}v) = \frac{1}{n} \sum_{t=T+1}^{T+n} \frac{1}{\hat{m}_t \alpha} \left[(\alpha - \mathbb{1}(y_{M,t} \leq \hat{v}_t)) \hat{\beta}_t \hat{v}_t + y_{i,t} \mathbb{1}(y_{M,t} \leq \hat{v}_t) \right] + \ln(-\hat{m}_t), \quad (4.15)$$

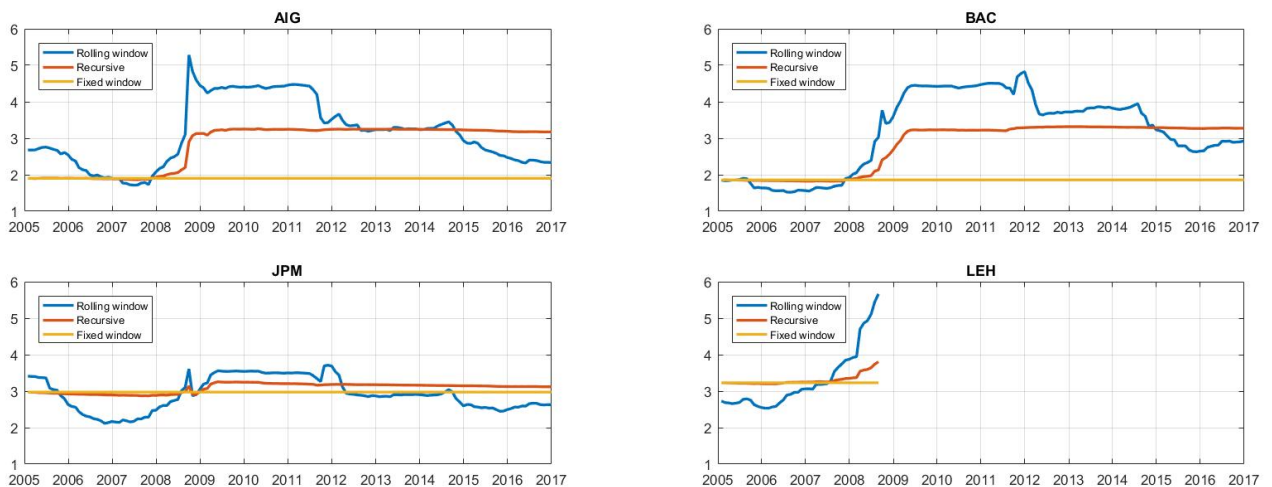
where $\hat{\beta}_t$ is the slope OLS estimate of the simple linear regression of $Y_{i,t}$ onto $Y_{M,t}$ using the same estimation scheme and sample period as the ones used for the computation of \hat{m}_t and \hat{v}_t . Our choice of a simple OLS estimator for the nuisance parameter is consistent with our model assumptions. Furthermore, the nuisance parameter estimator is strictly

linear with respect to v through β and we confirm empirically that the values of $\hat{\beta}_t$ are always strictly positive implying that the nuisance parameter estimator is strictly increasing, thereby ensuring the quasi-convexity of the scoring function. We set the out-of-sample size to $n = 500$, i.e., about two years of daily returns. Hereinafter, we refer to the values of the empirical scoring function in (4.15) as average losses.

4.5.2.2 Main results

Our empirical results are reported for four major US banks, namely Bank of America (BAC), JP Morgan (JPM), American International Group (AIG), and Lehman Brothers (LEH). We display our results for the period 2005-2017. Note that for Lehman Brothers, our results are only available till September 15, 2008 because of the collapse of Lehman Brothers.

Figure 4.2: One-day ahead out-of-sample MES forecasts

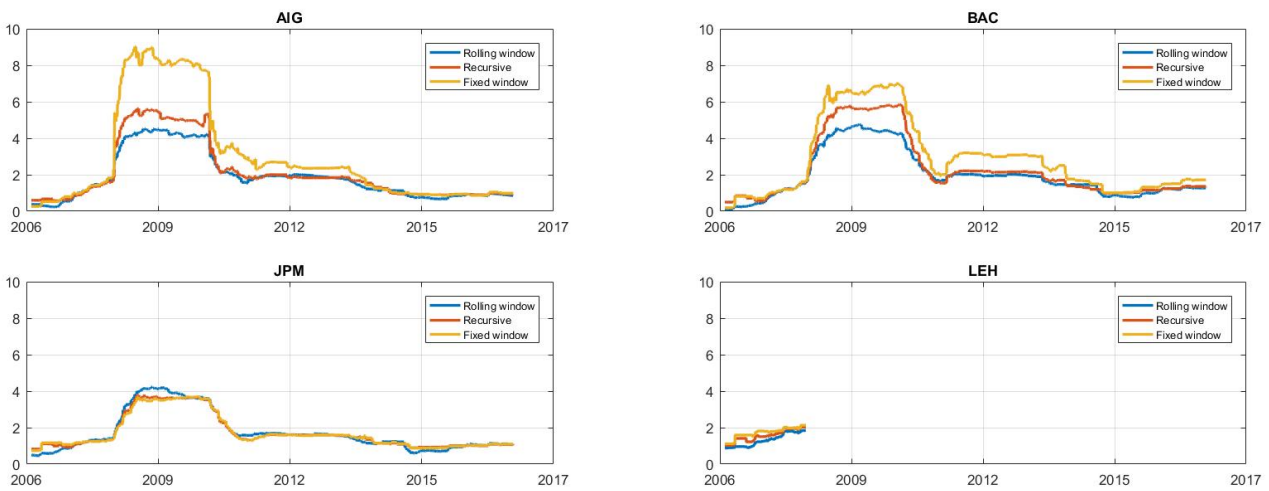


Note: This figure plots the 5% MES and VaR for American International Group (AIG), Bank of America (BAC), JP Morgan (JPM), and Lehman Brothers (LEH), over the period January 2005 to December 2016. The estimates are based on a rolling window, a recursive window, and a fixed estimation scheme.

Figure 4.2 displays the one-day ahead out-of-sample forecasts of the MES issued by the three competing approaches over the period January 2005 to December 2017. The MES forecasts associated with the fixed estimation scheme is time-invariant which is naturally expected given that a 2000-2004 fixed sample period is used to estimate $\sigma_{i,t}^2$,

$\sigma_{M,t}^2$, and $\rho_{i,t}$. In contrast, the MES produced by the rolling and recursive estimation schemes are time-varying and reveal a bump as of 2008 which coincides with the 2007-2009 financial crisis. The MES forecasts produced by the rolling estimation scheme with $T = 500$ observations reacts more quickly compared to the one produced by the recursive estimation scheme that does not overweight the earliest financial information. Interestingly within the rolling scheme, we observe a large increase of MES for Lehman Brothers just before its bankruptcy. Once we account for dynamics in these financial time series, the MES captures (at least partially) the impact of the financial crisis on the capital shortfalls of these four institutions.

Figure 4.3: Average out-of-sample losses ($n = 500$)



Note: This figure plots the out-of-sample losses of the 5% MES and VaR forecasts of American International Group (AIG), Bank of America (BAC), JP Morgan (JPM), and Lehman Brothers (LEH), over the period January 2005 to December 2016. We use $n = 500$ to compute the average losses. The losses are based on three different forecasting approaches of MES and VaR, namely a rolling window, a recursive window, and a fixed estimation scheme.

We now turn to the out-of-sample forecast evaluation of the competing approaches discussed in Section 4.5.2.1. Figure 4.3 displays the average out-of-sample losses that are computed using Equation (4.15). We raise two main remarks. First, we observe that the out-of-sample forecast performance fluctuates considerably over time. For all

the three competing approaches, we find that the out-of-sample losses are much higher during the 2007-2009 financial crisis than for any other periods. We conclude that the out-of-sample forecast performance worsens a lot in times of financial instability, which is unfortunately when systemic-risk measures are most needed. Our findings support the statistical evidences highlighted in Banulescu-Radu et al. (2021) that the systemic-risk measures are misleading before and during financial crisis episodes. Our loss may serve for the early identification of systemic events. To identify structural breaks, we might compare the out-of-sample performances of a forecasting model, judged by the average losses, with the in-sample performances (Giacomini and Rossi, 2009). Second, although there is a decline in forecast accuracy in times of crisis, the MES and VaR of the rolling window estimation scheme generally outperform the ones calculated with a fixed and recursive estimation schemes, suggesting that incorporating the most recent data before crisis lead to higher performance in predicting future systemic crises.

4.6 Conclusion

Elicitability is useful for model selection, estimation, generalized regression, forecast ranking (Gneiting, 2011a) and comparative backtesting as well (Nolde and Ziegel, 2017). This mathematical property is growing in popularity in the banking industry while being extensively used by quantitative risk managers for backtesting standard risk measures like the value-at-risk (VaR) or the expected shortfall (ES) (see Acerbi and Szekely, 2014; Fissler and Ziegel, 2016).

This article establishes the elicibility and the identifiability of the systemic-risk measures that are widely used by academics and regulators to quantify how much an entity, be it a bank or hedge fund or sovereign, contributes to the vulnerability of the financial system. We derive a strict consistent scoring function for the marginal expected shortfall (MES) and the value-at-risk (VaR). Our framework is then extended to any

systemic-risk measures that can be expressed as an expected equity conditional to a financial crisis, as it is the case for the SRISK for instance.

This finding paves the road to M-estimation, forecast comparison, and backtesting, of those systemic-risk parameters, and thereby has immediate relevance for European regulation aimed at containing systemic risk, i.e., Basel IV with the yearly identification of global systemically important banks. We exploit the scoring function to develop a joint semiparametric M-estimator for the MES and VaR. We show the consistency and we characterize the asymptotic distribution of the estimator enabling modeling and statistical inference for the systemic-risk indicators. We also give a discussion on how to perform model comparison and backtesting for systemic-risk measures under our framework. A small empirical study applied to four large US banks is provided to highlight the merit of our elicitation framework.

4.7 Appendix

4.7.1 Appendix A: Spectral risk measure

In this section, we show that MES is a spectral risk measure and discuss a number of desirable properties that the systemic indicator satisfies. First, MES is a *coherent risk measure* (Artzner et al., 1999) as it satisfies the following properties:

- Monotonicity:

$$Y_i \leq Y_j \implies MES_i^\alpha = \mathbb{E}[Y_i | Y_M \leq VaR_M^\alpha] \leq \mathbb{E}[Y_j | Y_M \leq VaR_M^\alpha] = MES_j^\alpha,$$

- Super-additivity:

$$MES_{i+j}^\alpha = \mathbb{E}[Y_i + Y_j | Y_M \leq VaR_M^\alpha] \geq \mathbb{E}[Y_i | Y_M \leq VaR_M^\alpha] + \mathbb{E}[Y_j | Y_M \leq VaR_M^\alpha] = MES_i^\alpha + MES_j^\alpha,$$

- Positive homogeneity:

$$MES_{i,\lambda}^\alpha = \mathbb{E}[\lambda Y_i | Y_M \leq VaR_M^\alpha] = \lambda \mathbb{E}[Y_i | Y_M \leq VaR_M^\alpha] = \lambda MES_i^\alpha, \quad \forall \lambda \in \mathbb{R}, \lambda > 0,$$

- Translation invariance:

$$MES_{i,+a}^\alpha = \mathbb{E}[Y_i + a | Y_M \leq VaR_M^\alpha] = \mathbb{E}[Y_i | Y_M \leq VaR_M^\alpha] + a = MES_i^\alpha + a, \quad \forall a \in \mathbb{R}.$$

Furthermore, MES satisfies additional properties that we give below:

- Comonotonic additivity:

$$MES_{i,j}^\alpha = \mathbb{E}[Y_i + Y_j | Y_M \leq VaR_M^\alpha] = \mathbb{E}[Y_i | Y_M \leq VaR_M^\alpha] + \mathbb{E}[Y_j | Y_M \leq VaR_M^\alpha] = MES_i^\alpha + MES_j^\alpha,$$

- Law-invariance: if Y_i and Y_j have the same conditional (on $Y_M \leq VaR_M^\alpha$) distribution (i.e., $F_{Y_i | Y_M \leq VaR_M^\alpha}$ and $F_{Y_j | Y_M \leq VaR_M^\alpha}$, respectively), then $MES_i^\alpha = \mathbb{E}[Y_i | Y_M \leq VaR_M^\alpha] = \mathbb{E}[Y_j | Y_M \leq VaR_M^\alpha] = MES_j^\alpha$.

A coherent, comonotonically additive, and law-invariant risk measure is said to be a *spectral risk measure* (Kusuoka, 2001; Acerbi, 2002; Tasche, 2002; Jouini et al., 2006). More formally, MES can be expressed as a function of (conditional) quantiles as follows:

$$MES_i^\alpha = \int_0^1 CoVaR_i^\alpha(\beta) d\beta,$$

where $CoVaR_i^\alpha(\beta) = F_{Y_i|Y_M \leq VaR_{Y_M}^\alpha}^{-1}(\beta)$ following the notation in Banulescu-Radu et al. (2021).

4.7.2 Appendix B: Assumptions

To derive the \mathcal{F} -consistency and \mathcal{F} -quasi-convexity of the scoring function, we introduce the following assumptions.

Assumption (A1). Let \mathcal{F} be a convex class of distributions functions on \mathbb{R}^2 . Assume that for all $\theta \in \mathring{\Theta}$, there are $F_1, F_2, F_3 \in \mathcal{F}$ such that

$$0 \in \text{int} \left(\text{conv} \left(\{ \bar{V}(\theta, F_1), \bar{V}(\theta, F_2), \bar{V}(\theta, F_3) \} \right) \right),$$

where \bar{V} designates the expected \mathcal{F} -identification function, $\text{int}(M)$ and $\text{conv}(M)$ denote the interior and the convex hull of M , respectively, with $\text{conv}(M) := \left\{ \sum_{i=1}^n \lambda_i x_i \mid n \in \mathbb{N}; x_1, x_2, \dots, x_n \in M; \lambda_1, \lambda_2, \dots, \lambda_n > 0, \sum_{i=1}^n \lambda_i = 1 \right\}$, for a set $M \subseteq \mathbb{R}^2$.

Assumption (A2). The expected \mathcal{F} -identification function $\bar{V}(\cdot, F) : \Theta \rightarrow \mathbb{R}^2, \theta \mapsto \bar{V}(\theta, F)$ is continuous, for all $F \in \mathcal{F}$.

Assumption (A3). The expected \mathcal{F} -identification function $\bar{V}(\cdot, F)$ is continuously differentiable, for all $F \in \mathcal{F}$.

Assumption (A4). The expected scoring function $\bar{S}(\cdot, F, e(\cdot)) : \Theta \rightarrow \mathbb{R}, \theta \mapsto \bar{S}(\theta, F, e(v))$, is continuously differentiable, for all $F \in \mathcal{F}$, with $\theta = (v, m)'$.

Assumption (A5). The expected scoring function $\bar{S}(\cdot, F, e(\cdot))$ is continuously differentiable and the gradient is locally Lipschitz continuous, for all $F \in \mathcal{F}$. Additionally, the expected scoring function $\bar{S}(\cdot, F, e(\cdot))$ is twice continuously differentiable at $\tau = \mathcal{T}(F) \in \mathring{\Theta}$.

Assumption (A6). For all $Y \in \mathbb{R}^2$, there exists a sequence $(F_n)_{n \in \mathbb{N}}$ of distributions $F_n \in \mathcal{F}$ that converges weakly to the Dirac-measure δ_Y such that the support of F_n is contained in a compact set K for all n .

Assumption (A7). The complement of the sets

$$C_V = \left\{ (\theta, y) \in \Theta \times \mathbb{R}^2 \mid V(\theta, \cdot) \text{ is continuous at the point } y \right\}$$

$$C_S = \left\{ (\theta, y, e(v)) \in \Theta \times \mathbb{R}^2 \times \mathbb{R} \mid S(\theta, \cdot, e(v)) \text{ is continuous at the point } y \right\}$$

has 4-dimensional Lebesgue measure zero.

For all $\theta \in \mathring{\Theta}$, Assumption (A1) ensures that there exist $F_1, F_2, F_3 \in \mathcal{F}$ such that $\bar{V}(\theta, F_1), \bar{V}(\theta, F_2), \bar{V}(\theta, F_3)$ are linearly independent vectors. It is a way to guarantee that the functional \mathcal{T} varies sufficiently to apply Proposition 9 so as to obtain a necessary form of the scoring function S (i.e., that the class \mathcal{F} is sufficiently “rich”). Such assumptions like (A1) are common in the literature (see Osband, 1985; Steinwart et al., 2014; Lambert, 2013).

In the case of the \mathcal{F} -identification function V is continuous in its first argument θ , Assumption (A2) is satisfied. The same argument is valid if V is continuously differentiable in its first argument with Assumption (A3). However, Assumptions (A2) and (A3) relate to the expected \mathcal{F} -identification function \bar{V} and therefore are much weaker requirements.

Assumptions (A4) and (A5) are equivalent conditions to Assumptions (A2) and (A3) but in terms of scoring functions (instead of identification functions). Assumption (A5) and the use of Rademacher’s theorem entail that the gradient of \bar{S} is (totally) differen-

tiable (in its first argument) for almost all $\theta \in \Theta$. Therefore using Schwarz's theorem, the Hessian of $\bar{S}(\cdot, F, e(\cdot))$ exists for almost all $\theta \in \Theta$ and is symmetric.

To derive the consistency of the estimator $\hat{\theta}$, we introduce the following assumptions.

Assumption (B1). $\theta_0 \in \Theta$, Θ is a compact subset of \mathbb{R}^2 .

Assumption (B2). $\hat{e} \in \mathcal{E}$ and \hat{e} converges to e in probability.

Assumption (B3). Y_t is strictly stationary and has a continuous distribution.

Assumption (B4). $\mathbb{E} | Y_{i,t} | < \infty$, $\mathbb{E} | G_1(Y_{M,t}) | < \infty$ and $\mathbb{E} | G(Y_t) | < \infty$.

4.7.3 Appendix C: Proofs

In order to make our proofs, we introduce some intermediate Lemma, Propositions and Corollary.

Lemma 1. *A scoring function $S : \Theta \times \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}$ is strictly \mathcal{F} -consistent for $\mathcal{T} : \mathcal{F} \rightarrow \Theta$ if and only if the function*

$$\psi : D \rightarrow \mathbb{R}, s \mapsto \mathbb{E} \left[S(\tau + su, Y, e(\tau_1 + su_1)) \right]$$

has a global unique minimum at $s = 0$ for all $F \in \mathcal{F}$, $\tau = \mathcal{T}(F)$ and $u \in \mathbb{S}^1$, with \mathbb{S}^1 the unit 1-sphere such that $\mathbb{S}^1 = \{x \in \mathbb{R}^2 : \|x\| = 1\}$, and where $D = \{s \in \mathbb{R} : \tau + su \in \Theta\}$.

Proposition 9 (Osband's principle). *Let $\mathcal{T} : \mathcal{F} \rightarrow \Theta \subseteq \mathbb{R}^2$ be a surjective, elicitable and identifiable functional with a strict \mathcal{F} -identification function $V : \Theta \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$ and a strictly \mathcal{F} -consistent scoring function $S : \Theta \times \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}$. There exists a matrix-valued function $h : \mathring{\Theta} \rightarrow \mathbb{R}^{2 \times 2}$ such that for all $\theta \in \mathring{\Theta}$ and $F \in \mathcal{F}$*

$$\begin{aligned} \partial_1 \mathbb{E} [S(\theta, Y, e(v))] &= h_{11}(\theta) \mathbb{E} [V_1(\theta, Y)] + h_{12}(\theta) \mathbb{E} [V_2(\theta, Y)], \\ &\text{and} \end{aligned} \tag{4.16}$$

$$\partial_2 \mathbb{E} [S(\theta, Y, e(v))] = h_{21}(\theta) \mathbb{E} [V_1(\theta, Y)] + h_{22}(\theta) \mathbb{E} [V_2(\theta, Y)].$$

as soon as the assumptions (A1) and (A4) in 4.7.2 hold, where $\overset{\circ}{\Theta}$ denotes the interior of Θ and $\partial_1 \mathbb{E} [S(\theta, Y, e(v))]$ and $\partial_2 \mathbb{E} [S(\theta, Y, e(v))]$ correspond to the partial derivatives of $\mathbb{E} [S(\theta, Y, e(v))]$ with respect to v and m , respectively. Additionally, h is continuous if Assumption (A2) in 4.7.2 is valid. The function h is locally Lipschitz continuous in the case Assumptions (A3) and (A5) in 4.7.2 are also satisfied.

Corollary 1. *Let $\mathcal{T} : \mathcal{F} \rightarrow \Theta \subseteq \mathbb{R}^2$ be a surjective, elicitable and identifiable functional with a strict \mathcal{F} -identification function $V : \Theta \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$ and a strictly \mathcal{F} -consistent scoring function $S : \Theta \times \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}$. Suppose Assumptions (A1), (A3) and (A5) in 4.7.2 hold. Then, for all $F \in \mathcal{F}$ and almost all $\theta \in \overset{\circ}{\Theta}$, we have the following identities for the second-order derivatives:*

$$\begin{aligned}
\partial_1 \partial_1 \mathbb{E} [S(\theta, Y, e(v))] &= \partial_1 h_{11}(\theta) \mathbb{E} [V_1(\theta, Y)] + h_{11}(\theta) \partial_1 \mathbb{E} [V_1(\theta, Y)] \\
&\quad + \partial_1 h_{12}(\theta) \mathbb{E} [V_2(\theta, Y)] + h_{12}(\theta) \partial_1 \mathbb{E} [V_2(\theta, Y)], \\
\partial_2 \partial_2 \mathbb{E} [S(\theta, Y, e(v))] &= \partial_2 h_{21}(\theta) \mathbb{E} [V_1(\theta, Y)] + h_{21}(\theta) \partial_2 \mathbb{E} [V_1(\theta, Y)] \\
&\quad + \partial_2 h_{22}(\theta) \mathbb{E} [V_2(\theta, Y)] + h_{22}(\theta) \partial_2 \mathbb{E} [V_2(\theta, Y)], \\
\partial_1 \partial_2 \mathbb{E} [S(\theta, Y, e(v))] &= \partial_1 h_{21}(\theta) \mathbb{E} [V_1(\theta, Y)] + h_{21}(\theta) \partial_1 \mathbb{E} [V_1(\theta, Y)] \\
&\quad + \partial_1 h_{22}(\theta) \mathbb{E} [V_2(\theta, Y)] + h_{22}(\theta) \partial_1 \mathbb{E} [V_2(\theta, Y)] \\
&= \partial_2 h_{11}(\theta) \mathbb{E} [V_1(\theta, Y)] + h_{11}(\theta) \partial_2 \mathbb{E} [V_1(\theta, Y)] \\
&\quad + \partial_2 h_{12}(\theta) \mathbb{E} [V_2(\theta, Y)] + h_{12}(\theta) \partial_2 \mathbb{E} [V_2(\theta, Y)] \\
&= \partial_2 \partial_1 \mathbb{E} [S(\theta, Y, e(v))],
\end{aligned} \tag{4.17}$$

where h is the matrix-valued function appearing in Equation (4.16). In particular, Equation (4.17) holds for $\tau = \mathcal{T}(F) \in \overset{\circ}{\Theta}$.

Proposition 10. *Assume that $\overset{\circ}{\Theta} \subseteq \mathbb{R}^2$ is a star domain. Let $\mathcal{T} : \mathcal{F} \rightarrow \Theta$ be a surjective, elicitable and identifiable functional with a strict \mathcal{F} -identification function $V : \Theta \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$ and a strictly \mathcal{F} -consistent scoring function $S : \Theta \times \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}$. Suppose Assumptions (A1), (A2), (A4), (A6) and (A7) in 4.7.2 hold. Let h be the matrix-valued function*

appearing in Equation (4.16). Then the scoring function S is necessarily of the form

$$S(\theta, y, e(v)) = \int_{z_1}^v \left(h_{11}(u, z_2)V_1(u, z_2, y) + h_{12}(u, z_2)V_2(u, z_2, y) \right) du + \int_{z_2}^m \left(h_{21}(v, u)V_1(v, u, y) + h_{22}(v, u)V_2(v, u, y) \right) du + b(y), \quad (4.18)$$

for almost all $\theta \in \Theta$, for almost all $y \in \mathbb{R}^2$, for some star point $z = (z_1, z_2)' \in \mathring{\Theta}$ and some \mathcal{F} -integrable function $b : \mathbb{R}^2 \rightarrow \mathbb{R}$. In particular, regarding the expected score $\mathbb{E}[S(\theta, Y, e(v))]$, Equation (4.18) holds for all $\theta \in \mathring{\Theta}$, $F \in \mathcal{F}$.

The proof of Proposition 9 is the same than the proof of Fissler and Ziegel (2016) for the pair (VaR, ES) and is thus omitted (see Fissler and Ziegel, 2016, Theorem 3.2). For the same reason, the proof of Proposition 10 is not reported (see Fissler and Ziegel, 2016, Proposition 3.4).

Propositions 9 and 10 and Corollary 1 determine necessary conditions for strictly \mathcal{F} -consistent scoring functions for a functional \mathcal{T} . When a strict \mathcal{F} -identification function V for \mathcal{T} is available, they offer a way on how to build a strictly \mathcal{F} -consistent scoring function S from V .

Proposition 11. *Let $\Theta \subseteq \mathbb{R}^2$ be an open convex set and let the expected scoring function $\bar{S}(\cdot, F, e(\cdot)) : \Theta \rightarrow \mathbb{R}$, $\theta \mapsto \bar{S}(\theta, F, e(v))$ be a twice continuously differentiable function with gradient $\partial \bar{S}$ and Hessian $\partial^2 \bar{S}$. Then \bar{S} is quasi-convex if and only if for all $\theta = (v, m)' \in \Theta$, for all $F \in \mathcal{F}$ and for all $u \in \mathbb{S}^1$,*

$$u' \partial \bar{S}(\theta, F, e(v)) = 0$$

implies

(i) $u' \partial^2 \bar{S}(\theta, F, e(v))u > 0$; or

(ii) $u' \partial^2 \bar{S}(\theta, F, e(v))u = 0$ and $\psi : D \rightarrow \mathbb{R}$, $\psi(s) = \bar{S}(\theta + su, F, e(v + su_1))$ is quasi-convex,

with \mathbb{S}^1 the unit 1-sphere such that $\mathbb{S}^1 = \{x \in \mathbb{R}^2 : \|x\| = 1\}$, and where $D = \{s \in \mathbb{R} : \theta + su \in \Theta\}$. Moreover, \bar{S} is strictly quasi-convex if and only if this implication holds where in (ii) ψ does not attain a local maximum at $s = 0$.

Proposition 11 is drawn from Schaible and Ziemba (1981).

Theorem 5. This proof is inspired from Fissler and Ziegel (2016).

(i) By definition of elicibility, we know that if the second part of Theorem 5 (ii) is satisfied, then the functional \mathcal{T} is 2-elicitable if the distributions in \mathcal{F} have unique α -quantiles.

(ii) Let $S : \Theta \times \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}$ be of the form given in Equation (4.5) with \mathcal{G}_2 a convex function. Let the functions G_3 given in Equation (4.6) be increasing functions. Let $F \in \mathcal{F}$, $\theta \in \Theta$, with $\theta = (v, m)'$ and set $\tau = \mathcal{T}(F)$, with $\tau = (\tau_1, \tau_2)'$ and define $w = \min(m, \tau_2)$. Then we can rewrite (4.5) as

$$\begin{aligned} S(\theta, y, e(v)) &= (\mathbb{1}(y_M \leq v) - \alpha) \left(G_1(v) + G'_2(w) \frac{e(v) - y_i}{\alpha} \right) + G_1(y_M) (\alpha - \mathbb{1}(y_M \leq v)) \\ &\quad + (G'_2(m) - G'_2(w)) \left(m + \frac{e(v)}{\alpha} (\mathbb{1}(y_M \leq v) - \alpha) - \frac{1}{\alpha} y_i \mathbb{1}(y_M \leq v) \right) \\ &\quad + G'_2(w)(m - y_i) - G_2(m) + G(y). \end{aligned} \tag{4.19}$$

Taking the expectation of Equation (4.19) yields that $\mathbb{E}[S(\theta, Y, e(v))] - \mathbb{E}[S(\tau, Y, e(\tau_1))] = U_1 + U_2$ with

$$\begin{aligned} U_1 &= (F_{Y_M}(v) - \alpha) \left(G_1(v) + G'_2(w) \frac{e(v)}{\alpha} \right) \\ &\quad - \mathbb{E} \left[(\mathbb{1}(Y_M \leq v) - \mathbb{1}(Y_M \leq \tau_1)) \left(G_1(Y_M) + G'_2(w) \frac{Y_i}{\alpha} \right) \right], \end{aligned} \tag{4.20}$$

and

$$\begin{aligned} U_2 &= (G'_2(m) - G'_2(w)) \left(m + \frac{e(v)}{\alpha} (F_{Y_M}(v) - \alpha) - \frac{1}{\alpha} \mathbb{E}[Y_i \mathbb{1}(Y_M \leq v)] \right) \\ &\quad + G'_2(w)(m - \tau_2) + G_2(\tau_2) - G_2(m). \end{aligned} \tag{4.21}$$

Suppose that $\tau_1 < v$. Rewriting the expectation term in Equation (4.20), we obtain

$$\begin{aligned}
 & \mathbb{E} \left[\left(\mathbb{1}(Y_M \leq v) - \mathbb{1}(Y_M \leq \tau_1) \right) \left(G_1(Y_M) + G'_2(w) \frac{Y_i}{\alpha} \right) \right] \\
 &= \int_{\mathbb{R}} \int_{\tau_1}^v \left(G_1(x_1) + G'_2(w) \frac{x_2}{\alpha} \right) f(x_1, x_2) dx_1 dx_2 \\
 &= \int_{\mathbb{R}} \int_{\tau_1}^v \left(G_1(x_1) + G'_2(w) \frac{x_2}{\alpha} \right) f_{(Y_i|Y_M=x_1)}(x_2) f_{Y_M}(x_1) dx_1 dx_2 \\
 &= \int_{\tau_1}^v \left(G_1(x_1) + G'_2(w) \frac{e(x_1)}{\alpha} \right) f_{Y_M}(x_1) dx_1 \\
 &= \left[\left(G_1(x_1) + G'_2(w) \frac{e(x_1)}{\alpha} \right) F_{Y_M}(x_1) \right]_{\tau_1}^v - \int_{\tau_1}^v \left(G'_1(x_1) + G'_2(w) \frac{e'(x_1)}{\alpha} \right) F_{Y_M}(x_1) dx_1,
 \end{aligned} \tag{4.22}$$

where the last line comes from integrating by parts. Using that $F_{Y_M}(\tau_1) = \alpha$ if \mathcal{F} has a unique α -quantile, we obtain by plugging Equation (4.22) in Equation (4.20)

$$\begin{aligned}
 U_1 &= \alpha \left(G_1(\tau_1) + G'_2(w) \frac{e(\tau_1)}{\alpha} - G_1(v) - G'_2(w) \frac{e(v)}{\alpha} \right) + \int_{\tau_1}^v \left(G'_1(x_1) + G'_2(w) \frac{e'(x_1)}{\alpha} \right) F_{Y_M}(x_1) dx_1 \\
 &= -\alpha \int_{\tau_1}^v \left(G'_1(x_1) + G'_2(w) \frac{e'(x_1)}{\alpha} \right) dx_1 + \int_{\tau_1}^v \left(G'_1(x_1) + G'_2(w) \frac{e'(x_1)}{\alpha} \right) F_{Y_M}(x_1) dx_1 \\
 &= \int_{\tau_1}^v \left(G'_1(x_1) + G'_2(w) \frac{e'(x_1)}{\alpha} \right) (F_{Y_M}(x_1) - \alpha) dx_1.
 \end{aligned}$$

Since the term $G_1(x) + G'_2(w) \frac{e(x)}{\alpha}$ is increasing in $x \in [\tau_1, v]$ by assumption and $F_{Y_M}(x) \geq \alpha$ for $x \in [\tau_1, v]$, it implies that $U_1 \geq 0$. Similarly, $v < \tau_1$ entails that $U_1 \geq 0$. If the function $G_1(x) + G'_2(w) \frac{e(x)}{\alpha}$ is strictly increasing in x , then we get the strict positivity $U_1 > 0$, if $v \neq \tau_1$.

We will now focus on the term U_2 . We use the fact that $F_{Y_M}(\tau_1) = \alpha$ if \mathcal{F} has a unique α -quantile and the definition of τ_2 to rewrite the expectation term in Equation

(4.21) such that

$$\begin{aligned}
\mathbb{E} [Y_i \mathbb{1}(Y_M \leq v)] &= \int_{\mathbb{R}} \int_{-\infty}^v x_2 f(x_1, x_2) dx_1 dx_2 \\
&= \int_{\mathbb{R}} \int_{-\infty}^{\tau_1} x_2 f(x_1, x_2) dx_1 dx_2 + \int_{\mathbb{R}} \int_{\tau_1}^v x_2 f(x_1, x_2) dx_1 dx_2 \\
&= \alpha \tau_2 + \int_{\tau_1}^v e(x_1) f_{Y_M}(x_1) dx_1 \\
&= \alpha \tau_2 + [e(x_1) F_{Y_M}(x_1)]_{\tau_1}^v - \int_{\tau_1}^v e'(x_1) F_{Y_M}(x_1) dx_1,
\end{aligned} \tag{4.23}$$

where the last line comes from integrating by parts. We plug Equation (4.23) in Equation (4.21) to obtain:

$$\begin{aligned}
U_2 &= (G'_2(m) - G'_2(w)) \left(m + \frac{e(v)}{\alpha} (F_{Y_M}(v) - \alpha) - \tau_2 - \frac{e(v)}{\alpha} F_{Y_M}(v) + e(\tau_1) \right. \\
&\quad \left. + \frac{1}{\alpha} \int_{\tau_1}^v e'(x_1) F_{Y_M}(x_1) dx_1 \right) + G'_2(w)(m - \tau_2) + G_2(\tau_2) - G_2(m) \\
&= (G'_2(m) - G'_2(w)) \left(m - \tau_2 + \frac{1}{\alpha} \int_{\tau_1}^v e'(x_1) (F_{Y_M}(x_1) - \alpha) dx_1 \right) \\
&\quad + G'_2(w)(m - \tau_2) + G_2(\tau_2) - G_2(m) \\
&\geq G'_2(m)(m - \tau_2) + G_2(\tau_2) - G_2(m) \\
&\geq 0.
\end{aligned}$$

We have that G'_2 and e are increasing function and $F_{Y_M}(x) \geq \alpha$ for $x \in [\tau_1, v]$ which justify the first inequality. If G'_2 and e are strictly increasing, the inequality is strict. We exploit the convexity of G_2 to have the last inequality. Additionally, when $m \neq \tau_2$ and G_2 is strictly convex, we obtain a strict inequality.

(iii) Recall that the MES satisfies the following expression :

$$MES_i^\alpha = \frac{1}{\alpha} \int_{\mathbb{R}} \int_{-\infty}^{VaR_M^\alpha} x_2 f(x_1, x_2) dx_1 dx_2. \tag{4.24}$$

We focus on the assertions on V given in Equation (4.7). Let $F \in \mathcal{F}$, $\tau = \mathcal{T}(F)$. Then we have that $\mathbb{E} [V_1(\theta, Y)] = F_{Y_M}(v) - \alpha$, for $\theta \in \Theta$, with $\theta = (v, m)'$. This quantity is null if and only if $v = \tau_1$. Also, if we consider the expression given in Equation (4.24),

we have that

$$\mathbb{E} [V_2(\tau_1, m, Y)] = m - \frac{1}{\alpha} \int_{\mathbb{R}} \int_{-\infty}^{\tau_1} x_2 f(x_1, x_2) dx_1 dx_2 = m - \tau_2.$$

Thus, V is only null if $\theta = \tau$ which corresponds to a strict \mathcal{F} -identification function for \mathcal{T} . Moreover, V satisfies Assumption (A3) in 4.7.2, and we have for $\theta \in \Theta$ that

$$\begin{aligned} \partial_1 \mathbb{E} (V_1(\theta, Y)) &= f_{Y_M}(v), & \partial_2 \mathbb{E} (V_1(\theta, Y)) &= 0, \\ \partial_1 \mathbb{E} (V_2(\theta, Y)) &= -f_{Y_M}(v) \frac{e(v)}{\alpha}, & \partial_2 \mathbb{E} (V_2(\theta, Y)) &= 1. \end{aligned}$$

Let us assume that $\tau = \mathcal{T}(F) \in \mathring{\Theta}$. Let S be a strictly \mathcal{F} -consistent scoring function for \mathcal{T} satisfying Assumption (A5) in 4.7.2. Then we can apply Proposition 9 and Corollary 1 to get that there is locally Lipschitz continuous functions $h : \mathring{\Theta} \rightarrow \mathbb{R}^{2 \times 2}$ such that Equations (4.16) and (4.17) hold. If we evaluate the last equality of Equation (4.17) (i.e., the second cross derivatives) at the point $\theta = \tau$, we get

$$h_{21}(\tau) \partial_1 \mathbb{E} [V_1(\tau, Y)] + h_{22}(\tau) \partial_1 \mathbb{E} [V_2(\tau, Y)] = h_{12}(\tau) \partial_2 \mathbb{E} [V_2(\tau, Y)], \quad (4.25)$$

which takes the form $h_{21}(\tau) f_{Y_M}(\tau_1) - h_{22}(\tau) f_{Y_M}(\tau_1) \frac{e(\tau_1)}{\alpha} = h_{12}(\tau)$. The second cross derivative of the expected scoring function at the point $\theta = \tau$ has the expression

$$\partial_2 \partial_1 \mathbb{E} [S(\tau, Y, e(\tau_1))] = \partial_2 \partial_1 \mathbb{E} [S(\tau, Y, e(\tau_1))] = \frac{G_2''(\tau_2)}{\alpha} e'(\tau_1) (F_{Y_M}(\tau_1) - \alpha) = 0. \quad (4.26)$$

From the third equality in (4.17), combining Equation (4.25) and Equation (4.26), we deduce that $h_{12}(\tau) = 0$, by identification. This entails $h_{21}(\tau) = h_{22}(\tau) \frac{e(\tau_1)}{\alpha}$. Hence, the surjectivity of \mathcal{T} entails that, for $\theta \in \mathring{\Theta}$,

$$h_{12}(\theta) = 0, \quad h_{21}(\theta) = h_{22}(\theta) \frac{e(v)}{\alpha}. \quad (4.27)$$

From Equation (4.27), we deduce that $h_{21}(\theta)\partial_1\mathbb{E}[V_1(\theta, Y)] + h_{22}(\theta)\partial_1\mathbb{E}[V_2(\theta, Y)] = 0$ since $\partial_1\mathbb{E}[V_1(\theta, Y)] = f_{Y_M}(v)$ and $\partial_1\mathbb{E}[V_2(\theta, Y)] = -f_{Y_M}(v)\frac{e'(v)}{\alpha}$. From this zero equality, by taking the difference between $\partial_1\partial_2\mathbb{E}[S(\theta, Y, e(v))]$ and $\partial_2\partial_1\mathbb{E}[S(\theta, Y, e(v))]$ in (4.17), and using $\partial_2\mathbb{E}[V_1(\theta, Y)] = 0$ and $h_{12}(\theta) = 0$ in Equation (4.27), we deduce

$$(\partial_1h_{21}(\theta) - \partial_2h_{11}(\theta))\mathbb{E}[V_1(\theta, Y)] + (\partial_1h_{22}(\theta) - \partial_2h_{12}(\theta))\mathbb{E}[V_2(\theta, Y)] = 0. \quad (4.28)$$

From Equation (4.28), Assumption (A1) in 4.7.2 implies that the two coefficients of the convex combination are zero, this gives $\partial_1h_{21}(\theta) = \partial_2h_{11}(\theta)$ and $\partial_1h_{22}(\theta) = \partial_2h_{12}(\theta)$. We know that $\partial_2h_{12}(\theta) = 0$ from Equation (4.27) and we get

$$\partial_1h_{22}(\theta) = 0. \quad (4.29)$$

From Equation (4.27), we have $\partial_1h_{21}(\theta) = \partial_1h_{22}(\theta)\frac{e'(v)}{\alpha} + h_{22}(\theta)\frac{e'(v)}{\alpha}$, then using Equation (4.29) we get

$$\partial_2h_{11}(\theta) = \partial_1h_{21}(\theta) = h_{22}(\theta)\frac{e'(v)}{\alpha}. \quad (4.30)$$

From Equation (4.29), we get that there is a locally Lipschitz continuous function $\varphi_2 : \Theta_m \rightarrow \mathbb{R}$ such that for all $\theta \in \mathring{\Theta}$ we have $h_{22}(\theta) = \varphi_2(m)$. Equations (4.29) and (4.30) entail that $h_{11}(\theta) = \frac{e'(v)}{\alpha}G'_2(m) + \varphi_1(v)$ for $\theta \in \mathring{\Theta}$, where $\varphi_1 : \Theta_v \rightarrow \mathbb{R}$ is locally Lipschitz continuous and $G'_2 : \Theta_m \rightarrow \mathbb{R}$ is such that $G''_2 = \varphi_2$.

We can apply Proposition 10 using the form of the matrix-valued function h that we found. Let $z \in \mathring{\Theta}$ be some star point. Then there is some \mathcal{F} -integrable function $b : \mathbb{R} \rightarrow \mathbb{R}$

such that

$$\begin{aligned}
S(\theta, y, e(v)) &= \int_{z_1}^v \left(G_2'(z_2) \frac{e'(u)}{\alpha} + \varphi_1(u) \right) (\mathbb{1}(y_M \leq u) - \alpha) du \\
&\quad + \int_{z_2}^m \varphi_2(u) \left(\frac{e(v)}{\alpha} (\mathbb{1}(y_M \leq v) - \alpha) + u - \frac{1}{\alpha} y_i \mathbb{1}(y_M \leq v) \right) du + b(y) \\
&= \int_{z_1}^v \left(G_2'(z_2) \frac{e'(u)}{\alpha} + \varphi_1(u) \right) (\mathbb{1}(y_M \leq u) - \alpha) du \\
&\quad + (G_2'(m) - G_2'(z_2)) \left(\frac{e(v)}{\alpha} (\mathbb{1}(y_M \leq v) - \alpha) - \frac{1}{\alpha} y_i \mathbb{1}(y_M \leq v) \right) \\
&\quad + mG_2'(m) - G_2(m) - (z_2G_2'(z_2) - G_2(z_2)) + b(y),
\end{aligned} \tag{4.31}$$

for almost all $\theta \in \Theta$ and $y \in \mathbb{R}^2$, where $G_2 : \Theta_m \rightarrow \mathbb{R}$. By a straightforward computation, we can check that the representation of S given in Equation (4.31) is equivalent to the expression of S given in Equation (4.5) upon choosing an appropriate \mathcal{F} -integrable function $G : \mathbb{R} \rightarrow \mathbb{R}$.

Lastly, we will show that G_2 is strictly convex and that the functions G_3 given in Equation (4.6) are strictly increasing. Let us define $D = \{s \in \mathbb{R} : \tau + su \in \mathring{\Theta}\}$, with $u = (u_1, u_2) \in \mathbb{S}^1$ and $\psi : D \rightarrow \mathbb{R}$ by $\psi(s) = \mathbb{E}[S(\bar{s}, Y, e(\bar{s}_1))]$, with $\bar{s} = \tau + su$. Without loss of generality, we suppose that $u_2 \geq 0$. We obtain that

$$\begin{aligned}
\psi(s) &= \int_{z_1}^{\bar{s}_1} \left(G_2'(z_2) \frac{e'(r)}{\alpha} + \varphi_1(r) \right) (F_{Y_M}(r) - \alpha) dr \\
&\quad + (G_2'(\bar{s}_2) - G_2'(z_2)) \left(\frac{e(\bar{s}_1)}{\alpha} (F_{Y_M}(\bar{s}_1) - \alpha) - \frac{1}{\alpha} \int_{\mathbb{R}} \int_{-\infty}^{\bar{s}_1} x_2 f(x_1, x_2) dx_1 dx_2 \right) \\
&\quad + \bar{s}_2 G_2'(\bar{s}_2) - G_2(\bar{s}_2) - (z_2 G_2'(z_2) - G_2(z_2)) + \mathbb{E}[b(Y)].
\end{aligned}$$

Using Lemma 1, we know that the function ψ has a minimum at $s = 0$, implying that there is $\epsilon > 0$ such that $\psi'(s) < 0$ for $s \in (-\epsilon, 0)$ and $\psi'(s) > 0$ for $s \in (0, \epsilon)$. If $u_2 = 0$, then

$$\psi'(s) = u_1 (F_{Y_M}(\bar{s}_1) - \alpha) \left(G_2'(\tau_2) \frac{e'(\bar{s}_1)}{\alpha} + \varphi_1(\bar{s}_1) \right).$$

Since the quantities $\psi'(s)$ and $u_1(F_{Y_M}(\bar{s}_1) - \alpha)$ are strictly positive for $s > 0$ whatever the sign of u_1 , we obtain that $G'_2(\tau_2)\frac{e'(\bar{s}_1)}{\alpha} + \varphi_1(\bar{s}_1) > 0$. Exploiting the surjectivity of \mathcal{T} we can deduce that the functions in Equation (4.6) are strictly increasing. On the other hand, if $u_1 = 0$, we obtain that $\psi'(s) = s\varphi_2(\bar{s}_2)$. Again using the surjectivity of \mathcal{T} , we get that $\varphi_2 > 0$ which shows the strict convexity of G_2 . \blacksquare

Proposition 8. By definition, the equivalence of scoring functions preserves the strict \mathcal{F} -consistency and the \mathcal{F} -quasi-convexity of the scoring function S . Consequently, if S_0 is an \mathcal{F} -consistent and \mathcal{F} -quasi-convex scoring function for \mathcal{T} , then these two properties will be shared by equivalent scoring functions.

First, the loss function S_0 coincides with the function S given in Theorem 5 with specific functions G, G_1 and G_2 such that $G \equiv -1$, $G_1 \equiv 0$ is increasing and \mathcal{F} -integrable, and $\forall x < 0$, $G_2(x) = -\ln(-x)$ is strictly increasing and strictly convex. Then we can apply Theorem 5 and S_0 is strictly \mathcal{F} -consistent for \mathcal{T} .

To show that S_0 is also \mathcal{F} -quasi-convex, we verify the criterion given in Proposition 11. In the sequel, we consider the general form of the scoring function S given in Theorem 5. Let $F \in \mathcal{F}$ with continuous density $f = F'$, $\theta \in \Theta^-$, $u \in \mathbb{S}^1$, and $u' \partial \mathbb{E}[S(\theta, Y, e(v))] = 0$. The first order and second order partial derivatives of the expected score $\mathbb{E}(S(\theta, Y, e(v)))$ have the form

$$\begin{aligned} \partial_1 \mathbb{E}[S(\theta, Y, e(v))] &= \left(G'_1(v) + \frac{G'_2(m)}{\alpha} e'(v) \right) (F_{Y_M}(v) - \alpha), \\ \partial_2 \mathbb{E}[S(\theta, Y, e(v))] &= G''_2(m) \left[m + (F_{Y_M}(v) - \alpha) \frac{e(v)}{\alpha} - \frac{1}{\alpha} \int_{-\infty}^v \int_{\mathbb{R}} z f(y, z) dz dy \right], \\ \partial_1 \partial_1 \mathbb{E}[S(\theta, Y, e(v))] &= \left(G''_1(v) + \frac{G''_2(m)}{\alpha} e''(v) \right) (F_{Y_M}(v) - \alpha) + f_{Y_M}(v) \left(G'_1(v) + \frac{G'_2(m)}{\alpha} e'(v) \right), \\ \partial_2 \partial_2 \mathbb{E}[S(\theta, Y, e(v))] &= G''_2(m) + G'''_2(m) \left[m + (F_{Y_M}(v) - \alpha) \frac{e(v)}{\alpha} - \frac{1}{\alpha} \int_{-\infty}^v \int_{\mathbb{R}} z f(y, z) dz dy \right], \\ \partial_1 \partial_2 \mathbb{E}[S(\theta, Y, e(v))] &= \partial_2 \partial_1 \mathbb{E}[S(\theta, Y, e(v))] = \frac{G''_2(m)}{\alpha} e'(v) (F_{Y_M}(v) - \alpha). \end{aligned} \tag{4.32}$$

Using Equation (4.32), this is equivalent to

$$\begin{aligned} & u_1 (F_{Y_M}(v) - \alpha) \left(G_1'(v) + \frac{G_2'(m)}{\alpha} e'(v) \right) \\ &= -u_2 G_2''(m) \left[m + (F_{Y_M}(v) - \alpha) \frac{e(v)}{\alpha} - \frac{1}{\alpha} \int_{-\infty}^v \int_{\mathbb{R}} z f(y, z) dz dy \right], \end{aligned}$$

and then we obtain

$$\begin{aligned} u' \partial^2 \mathbb{E} [S(\theta, Y, e(v))] u &= u_1^2 f_{Y_M}(v) \left(G_1'(v) + \frac{G_2'(m)}{\alpha} e'(v) \right) + u_2^2 G_2''(m) \\ &\quad + u_1 (F_{Y_M}(v) - \alpha) \left[u_1 \left(G_1''(v) + \frac{G_2'(m)}{\alpha} e''(v) \right) + 2u_2 \frac{G_2''(m)}{\alpha} e'(v) \right. \\ &\quad \left. - u_2 G_2'''(m) \frac{G_1'(v) + \frac{G_2'(m)}{\alpha} e'(v)}{G_2''(m)} \right]. \end{aligned} \tag{4.33}$$

According to the Proposition 11, S is \mathcal{F} -quasi convex if Equation (4.33) is strictly positive. The first term is strictly positive since it is a product of a density function and the derivative of the functions introduced in Equation 4.6. The second summand is also strictly positive, the function G_2 being strictly convex. However, concerning the third summand, the sign of the first term of the product (i.e., $u_1 (F_{Y_M}(v) - \alpha)$) is unknown. Consequently the term in brackets must be null entailing that $G_1' \equiv 0$, $G_1'' \equiv 0$, $2(G_2''(x))^2 = G_2'(x)G_2'''(x), \forall x < 0$ and $e'' \equiv 0$. Therefore, G_1 is a constant function, and G_2 is equivalent of the function $c_1 \ln(-x) + c_0$ with $c_1, c_0 \in \mathbb{R}$ with $c_1 < 0$ and e is linear. ■

Proof. Theorem 6. Following Chen et al. (2003) and Delsol and Van Keilegom (2020), the consistency of $\hat{\theta}$ are obtained by showing the following intermediate results:

- (i) $\hat{\theta} \in \Theta$ and $S_T(\hat{\theta}, \hat{e}) \leq S_T(\theta_0, \hat{e}) + o_p(1)$.
- (ii) $\forall \delta > 0, \exists \epsilon(\delta) > 0$ such that $d(\theta, \theta_0) > \delta \implies S(\theta, e) - S(\theta_0, e) > \epsilon(\delta)$.
- (iii) $P(\hat{e} \in \mathcal{E}) \rightarrow 1$ as $T \rightarrow \infty$ and $d_{\mathcal{E}}(\hat{e}, e) = o_p(1)$.

(iv) $\lim_{d_{\mathcal{E}}(\tilde{e}, e) \rightarrow 0} \sup_{\theta \in \Theta} | \bar{S}(\theta, \tilde{e}) - \bar{S}(\theta, e) | = 0$, with $\tilde{e} \in \mathcal{E}$.

(v) $\sup_{\theta \in \Theta, \tilde{e} \in \mathcal{E}} \frac{|S_T(\theta, \tilde{e}) - S_T(\theta_0, \tilde{e}) - \bar{S}(\theta, \tilde{e}) + \bar{S}(\theta_0, \tilde{e})|}{1 + |S_T(\theta, \tilde{e}) - S_T(\theta_0, \tilde{e})| + |S(\theta, \tilde{e}) - S(\theta_0, \tilde{e})|} = o_p(1)$.

As pointed out by Delsol and Van Keilegom (2020), if statements (i)-(v) hold, we get consistency irrespective of the data being i.i.d. or exhibiting serial correlation.

To show that (i) is correct, we use a Taylor approximation of $S_T(\theta_0, \hat{e})$ at $\hat{\theta}$,

$$S_T(\theta_0, \hat{e}) = S_T(\hat{\theta}, \hat{e}) + (\theta_0 - \hat{\theta})' \left. \frac{\partial S_T(\theta, \hat{e})}{\partial \theta} \right|_{\theta = \hat{\theta}} + R(\hat{\theta}).$$

By definition, $\hat{\theta}$ is the global minimum of $S_T(\hat{\theta}, \hat{e})$ implying that $R(\hat{\theta})$ is strictly positive and $\left. \frac{\partial S_T(\theta, \hat{e})}{\partial \theta} \right|_{\theta = \hat{\theta}}$ is an $o_p(1)$. We thus obtain

$$S_T(\hat{\theta}, \hat{e}) \leq S_T(\theta_0, \hat{e}) + o_p(1).$$

To assess (ii), we consider a Taylor development of $S(\theta, e)$ at θ_0 and then we have:

$$S(\theta, e) - S(\theta_0, e) = (\theta - \theta_0)' \left. \frac{\partial S(\theta, e)}{\partial \theta} \right|_{\theta = \theta_0} + \frac{1}{2} (\theta - \theta_0)' \left. \frac{\partial^2 S(\theta, e)}{\partial \theta \partial \theta'} \right|_{\theta = \theta_0} (\theta - \theta_0) + o((\theta - \theta_0)'(\theta - \theta_0)),$$

θ_0 being the global minimum of $S(\theta, e)$, $\left. \frac{\partial S(\theta, e)}{\partial \theta} \right|_{\theta = \theta_0}$ is null and $\left. \frac{\partial^2 S(\theta, e)}{\partial \theta \partial \theta'} \right|_{\theta = \theta_0}$ is positive definite. Therefore we can conclude that $\forall \delta > 0, \exists \epsilon(\delta) > 0$ such that $d(\theta, \theta_0) > \delta$ implies

$$S(\theta, e) - S(\theta_0, e) > \epsilon(\delta).$$

Part (iii) is automatically satisfied by Assumption (B2). Then we need to use a non parametric estimator belonging in \mathcal{E} , i.e., that is strictly increasing. Consequently, a standard non parametric kernel regression estimator is not a good candidate. Also, we have to consider a nonparametric estimator \hat{e} of e that is weakly consistent.

To show (iv), we rewrite the expression of the expected scoring function and find that $\bar{S}(\theta, \tilde{e}) - \bar{S}(\theta, e) = (\tilde{e} - e) \frac{G'_2(m)}{\alpha} (F_{Y_M}(v) - \alpha)$ which directly entails:

$$\lim_{d_{\mathcal{E}}(\tilde{e}, e) \rightarrow 0} \bar{S}(\theta, \tilde{e}) = \bar{S}(\theta, e).$$

Lastly, we need to assess the point (v). First, note that this condition is implied by $\sup_{\theta \in \Theta, \tilde{e} \in \mathcal{E}} |S_T(\theta, \tilde{e}) - \bar{S}(\theta, \tilde{e})| = o_p(1)$. This condition is satisfied if $S_T(\theta, \tilde{e}) = T^{-1} \sum_{t=1}^T S(\theta, y_t, \tilde{e}(v))$ converges uniformly in probability to $\bar{S}(\theta, \tilde{e})$, i.e., if a Uniform Law of Large Numbers (ULLN) can be applied.

A ULLN can be applied if these additional assumptions are valid:

- (vi) $S(\theta, y, \tilde{e}(v))$ is continuous at each $\theta \in \Theta$ with probability one,
- (vii) $S(\theta, y, \tilde{e}(v))$ is dominated by a function $U(y)$, i.e., $|S(\theta, y, \tilde{e}(v))| \leq U(y)$,
- (viii) $\mathbb{E}[U(Y_t)] < \infty$.

Points (vi) is satisfied by Assumption (B3).

For (vii), suppose the parameter space Θ is $[-C, C]$ for some positive constant $C = (c_v, c_m)$. The dominated function $U(y)$ is the following:

$$U(y) = 2(|G_1(y_M)| + |G_1(c_v)|) + |G'_2(c_m)| \frac{2|\mathbb{E}[Y_i]| + |y_i|}{\alpha} + |G_2(c_m)| + |G(y)|.$$

Then, Assumption (viii) is correct using Assumption (B4). ■

4.7.4 Appendix D: Some results on e in conditional volatility model

Proof. Let us consider the notation $Y_t = (Y_{M,t}, Y_{i,t})'$ for the daily demeaned returns. Denote by Ω_{t-1} the information set available at time $t - 1$, with $(Y_{t-1}, Y_{t-2}, \dots) \subseteq \Omega_{t-1}$. Assume the following conditional volatility model for Y_t ,

$$Y_t = \Sigma_t^{1/2} \eta_t,$$

where $\eta_t = (\eta_{M,t}, \eta_{i,t})'$ denotes a bivariate i.i.d. vector error process with $E[\eta_t] = \mathbf{0}$ and $E[\eta_t \eta_t'] = I_2$, and where $\Sigma_t = \begin{pmatrix} \sigma_{M,t}^2 & \rho_{i,t} \sigma_{i,t} \sigma_{M,t} \\ \rho_{i,t} \sigma_{i,t} \sigma_{M,t} & \sigma_{i,t}^2 \end{pmatrix}$ is the conditional variance-covariance matrix of Y_t with $\sigma_{M,t}^2$, $\sigma_{i,t}^2$, and $\rho_{i,t}$, the conditional variances of $Y_{M,t}$ and $Y_{i,t}$, and conditional correlation between $Y_{M,t}$ and $Y_{i,t}$, respectively. Note that there exists a broad range of specification for modeling the conditional variances $\sigma_{M,t}^2$, $\sigma_{i,t}^2$ and conditional correlation $\rho_{i,t}$ (see e.g. Bauwens et al., 2006, for a survey). Let us consider the Cholesky decomposition of Σ_t such as,

$$\Sigma_t^{1/2} = \begin{pmatrix} \sigma_{M,t} & 0 \\ \rho_{i,t} \sigma_{i,t} & \sigma_{i,t} \sqrt{1 - \rho_{i,t}^2} \end{pmatrix}.$$

The multiplicative form of the model for Y_t allows us to write,

$$\begin{aligned} Y_{M,t} &= \sigma_{M,t} \eta_{M,t}, \\ Y_{i,t} &= \rho_{i,t} \sigma_{i,t} \eta_{M,t} + \sigma_{i,t} \sqrt{1 - \rho_{i,t}^2} \eta_{i,t}. \end{aligned}$$

Taking conditional expectation, it follows,

$$\mathbb{E}[Y_{i,t} | Y_{M,t} = v_t, \Omega_{t-1}] = \rho_{i,t} \sigma_{i,t} \mathbb{E}[\eta_{M,t} | Y_{M,t} = v_t, \Omega_{t-1}] + \sigma_{i,t} \sqrt{1 - \rho_{i,t}^2} \underbrace{\mathbb{E}[\eta_{i,t} | Y_{M,t} = v_t, \Omega_{t-1}]}_{=0 \text{ as } \eta_{i,t} \perp \eta_{M,t}}.$$

Then, substituting $\eta_{M,t}$ by $Y_{M,t}/\sigma_{M,t}$, we get,

$$\mathbb{E}[Y_{i,t} | Y_{M,t} = v_t, \Omega_{t-1}] = \frac{\rho_{i,t} \sigma_{i,t}}{\sigma_{M,t}} \mathbb{E}[Y_{M,t} | Y_{M,t} = v_t, \Omega_{t-1}] = \frac{\rho_{i,t} \sigma_{i,t}}{\sigma_{M,t}} v_t.$$

Applying the law of iterated expectations, we end with,

$$\mathbb{E}[Y_{i,t} | Y_{M,t} = v_t] = \mathbb{E}[\mathbb{E}[Y_{i,t} | Y_{M,t} = v_t, \Omega_{t-1}] | Y_{M,t} = v_t] = v_t \mathbb{E}\left[\frac{\rho_{i,t} \sigma_{i,t}}{\sigma_{M,t}} | Y_{M,t} = v_t\right] = v_t \mathbb{E}\left[\frac{\rho_{i,t} \sigma_{i,t}}{\sigma_{M,t}}\right].$$

To ensure $\mathbb{E}[Y_{i,t} | Y_{M,t} = v_t]$ is strictly linear and strictly increasing, we need $\mathbb{E}\left[\frac{\rho_{i,t} \sigma_{i,t}}{\sigma_{M,t}}\right]$ to be strictly positive, which is satisfied as soon as the correlation is strictly positive. It is a realistic assumption in the systemic risk literature and asset pricing theory as the betas capture the systematic risk of the banking sector. The result applies to a large range of

volatility models including M-GARCH models. Furthermore, it encompasses the case of a time-invariant variance-covariance matrix Σ . ■

4.7.5 Appendix E: Characterization of the asymptotic distribution

Let $\bar{S}(\theta, e)$ denote the expectation of the scoring function $S(\theta, Y, e(v))$. Let $\partial_1 \bar{S}(\theta, e)$ denote a vector of the usual partial derivatives of $\bar{S}(\theta, e)$ with respect to the first argument θ ; and let $\partial_1^2 \bar{S}(\theta, e)$ correspond to the usual Hessian matrix of $\bar{S}(\theta, e)$ with respect to θ , for a constant e . Furthermore, let $D_\theta \bar{S}(\theta, e)$ and $D_e \bar{S}(\theta, e)$ denote the partial Fréchet derivatives of $\bar{S}(\theta, e)$ with respect to θ and e , respectively. Also, let $D_{\theta\theta} \bar{S}(\theta, e)$, $D_{\theta e} \bar{S}(\theta, e)$, and $D_{ee} \bar{S}(\theta, e)$ correspond to second-order partial Fréchet derivatives of $\bar{S}(\theta, e)$. Finally, let $\delta_{Y_M}(v)$ correspond to the dirac function of Y_M evaluated at v .

We apply Theorem 3.2 in Ichimura and Lee (2010): Assume that the observed data $\{Y_t\}_{t=1}^T$ are a random sample of Y . Let Assumptions 3.1 – 3.6 in Ichimura and Lee (2010) hold. Assume that Ω_0 exists and V_0 is a positive definite matrix. Then $\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, V_0^{-1} \Omega_0 V_0^{-1})$, with V_0 being the Hessian matrix of $\bar{S}(\theta, e)$ with respect to θ , evaluated at $\theta = \theta_0$, i.e., $V_0 = \frac{d^2}{d\theta d\theta'} \bar{S}(\theta, e) \Big|_{\theta=\theta_0}$, and Ω_0 such that $\Omega_0 = \mathbb{E}[\Gamma_0(Y)' \Gamma_0(Y)]$ with

$$\Gamma_0(y) = \Delta_{10}(y) - \mathbb{E}[\Delta_{10}(Y)] + \Delta_{20}(y, e'(v_0)) - \mathbb{E}[\Delta_{20}(Y, e'(v_0))] + \Gamma_1(y). \quad (4.34)$$

The terms of Equation (4.34) are calculated as follows:

$$\Delta_{10}(y) = \frac{\partial}{\partial \theta} S(x, y, e(v_0)) \Big|_{x=\theta_0} = \begin{pmatrix} G_1'(v_0) (\mathbb{1}(y_M \leq v_0) - \alpha) + \delta_{Y_M}(v_0) \frac{G_2'(m_0)}{\alpha} (e(v_0) - y_i) \\ G_2''(m_0) \left(m_0 - e(v_0) + \frac{\mathbb{1}(y_M \leq v_0)}{\alpha} (e(v_0) - y_i) \right) \end{pmatrix},$$

$$\mathbb{E}[\Delta_{10}(Y)] = 0,$$

$$\Delta_{20}(y, e'(v_0)) = D_e S(\theta_0, y, e(v_0)) e'(v_0) = \frac{G_2'(m_0)}{\alpha} (\mathbb{1}(y_M \leq v_0) - \alpha) e'(v_0),$$

$$\mathbb{E}[\Delta_{20}(Y, e'(v_0))] = 0.$$

Under Assumption 3.2 in Ichimura and Lee (2010), we get

$$D_e \bar{S}(\theta, e)(\tilde{e}(v) - e(v)) = \mathbb{E}[\Delta_2(Y, \tilde{e}(v) - e(v))] = \mathbb{E}[g(Y, \theta)(\tilde{e}(v) - e(v))]$$

with $\tilde{e} \in \mathcal{E}$, and $g(y, \theta) = \frac{G_2'(m)}{\alpha} (\mathbb{1}(y_M \leq v) - \alpha)$.

Finally, we provide an explicit form for $\Gamma_1(y)$ in Equation (4.34) when the first-stage estimator is a kernel regression estimator of $e(v)$,

$$\Gamma_1(y) = (y_i - e(v_0)) \mathbb{E} \left[\frac{\partial g(Y, \theta_0)}{\partial \theta} \right] = (y_i - e(v_0)) \begin{pmatrix} f_{Y_M}(v_0) \frac{G_2'(m_0)}{\alpha} \\ 0 \end{pmatrix}.$$

Then, using all these elements, we can rewrite Equation (4.34) and we obtain:

$$\Gamma_0(y) = \begin{pmatrix} \left(G_1'(v_0) + \frac{G_2'(m_0)e'(v_0)}{\alpha} \right) (\mathbb{1}(y_M \leq v_0) - \alpha) + \frac{G_2'(m_0)(e(v_0) - y_i)}{\alpha} (\delta_{Y_M}(v_0) - f_{Y_M}(v_0)) \\ G_2''(m_0) \left(m_0 - e(v_0) + \frac{\mathbb{1}(y_M \leq v_0)}{\alpha} (e(v_0) - y_i) \right) \end{pmatrix}.$$

We now focus on the expression of V_0 . By the chain rule, we get

$$\begin{aligned} V_0 &= \frac{d^2}{d\theta d\theta'} \bar{S}(\theta, e) \Big|_{\theta=\theta_0} \\ &= \partial_1^2 \bar{S}(\theta_0, e) + D_{ee} \bar{S}(\theta_0, e) \frac{\partial e(v_0)}{\partial \theta} \frac{\partial e(v_0)}{\partial \theta'} + 2D_e [\partial_1 \bar{S}(\theta_0, e)] \frac{\partial e(v_0)}{\partial \theta'} + D_e \bar{S}(\theta_0, e) \frac{\partial^2 e(v_0)}{\partial \theta \partial \theta'}. \end{aligned} \quad (4.35)$$

Focusing on each element, we have

$$\partial_1^2 \bar{S}(\theta_0, e) = \begin{pmatrix} f_{Y_M}(v_0) \left(G_1'(v_0) + \frac{G_2'(m_0)e'(v_0)}{\alpha} \right) & 0 \\ 0 & G_2''(m_0) \end{pmatrix},$$

$$D_{ee} \bar{S}(\theta_0, e) \frac{\partial e(v_0)}{\partial \theta} \frac{\partial e(v_0)}{\partial \theta'} = 0 \text{ since } \bar{S}(\theta, e) \text{ is a linear function of } e,$$

$$D_e [\partial_1 \bar{S}(\theta_0, e)] \frac{\partial e(v_0)}{\partial \theta} = 0 \text{ since } F_{Y_M}(v) - \alpha \text{ is null when it is evaluated at } \theta_0,$$

$$D_e \bar{S}(\theta_0, e) \frac{\partial^2 e(v_0)}{\partial \theta \partial \theta'} = 0 \text{ since } F_{Y_M}(v) - \alpha \text{ is null when it is evaluated at } \theta_0.$$

Consequently, all the terms in Equation (4.35) are null except the first one, leading V_0 to satisfy the following expression,

$$V_0 = \begin{pmatrix} f_{Y_M}(v_0) \left(G_1'(v_0) + \frac{G_2'(m_0)e'(v_0)}{\alpha} \right) & 0 \\ 0 & G_2''(m_0) \end{pmatrix}.$$

Chapter 5

Conclusion

This dissertation contributes to the academic research in econometrics and risk management. Our research's goal is twofold: *(i)* to quantify the financial risks incurred by financial institutions and *(ii)* to assess the validity of the risk measures commonly used in the financial industry or by regulators. We focus on three kind of financial risks, *(i)* credit risk, *(ii)* market risk, and *(iii)* systemic risk. In Chapter 2 and Chapter 3, we develop new methods for modeling and backtesting the volatility and the expected shortfall, two measures typically used to quantify the risk in investment portfolios. In Chapter 4, we provide new estimation methods for the systemic risk measures that are used to identify the financial institutions contributing the most to the overall risk in the financial system.

It is now broadly accepted that intraday returns are meaningful predictors of the daily volatility dynamics. Usual approach to account for the intraday movements is to introduce a summary of the intraday returns through realized volatility measures. The introduction of scalar exogenous variables in the volatility equation has given rise to the class of GARCH-X-type processes. In Chapter 2, we develop a volatility structure that groups the whole sequence of intraday returns as functional covariates. Contrary to the GARCH-X model that includes a summary of the intraday returns, our approach makes possible to account for the whole information contained in the intraday price movements via functional data analysis (FDA). The latter makes possible to transform

observed data into curves. Our contributions are the following. First, we develop a functional log-GARCH model where the response variable is scalar and the explanatory variables consist of a mix of scalar and functional variables. Second, we prove the existence of stationarity solutions and the existence of moments and log-moments. Third, we introduce a quasi-maximum likelihood estimator and detail the corresponding asymptotic properties. Finally, we develop a portmanteau statistic to test the goodness-of-fit of the model. In the empirical part, we apply our model to the intraday returns for a panel of ten NYSE/NASDAQ assets. Our functional variable is sampled at 5-minutes frequency. Our findings are as follows. First, we conclude to non-autocorrelated residuals in the model in more than 80% of cases and for seven out of ten assets. Second, we find that our functional variable is generally statistically significant which confirms that the whole intraday sequence must be used to explain the squared daily log-returns. Finally, we fit three competing models (including either, the realized volatility and a leverage effect, a leverage effect, or neither of them) that are nested representations of our model. We apply the model confidence set procedure to select the best volatility models. Our model belongs to the superior set of models for seven out of ten assets more often than the three competitors. A natural step forward of this work would consist in considering the multivariate time-series extension.

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. Among the number of fundamental reforms, the Value-at-Risk (VaR) has been substituted by the Expected Shortfall (ES) for the calculation of market risk capital requirements. ES offers a number of appealing properties that overcomes the deficiencies of the VaR. In particular, ES is a coherent risk measure. However, ES is not elicitable contrary to VaR, making its evaluation trickier. Chapter 3 introduces an econometric methodology to test for the validity of ES forecasts for market portfolios.

We exploit the fact that ES can be approximated using a finite sequence of VaRs. Our methodology examines jointly the validity of the VaR forecasts along the tail distribution of the risk model. Our procedure encompasses the BCBS recommendation of verifying quantiles at two specific risk levels. Our 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. Our regression framework makes possible to derive an analytical correction to adjust the misleading forecasts. In an empirical application, we highlight the ability of our backtests to reject a misspecified ES model. We also show the importance of choosing a sufficient number of quantiles to assess ES risk models. In particular, we find that the use of one or two quantiles is not always enough to identify whether the tail risk in bank's internal ES models is properly accounted for. On the contrary, four or more quantiles (until an optimal number) deliver much more sound decisions, suggesting an update of the regulatory guidelines. Further issues concern the same approach but for assessing the whole return distribution (interval and density forecasts).

The elicibility property paves the road to many useful applications: for model selection, estimation, generalized regression, forecast ranking and backtesting. A risk measure is said to be elicitable if there exists a loss function or a scoring function such that, the risk measure itself is the solution to minimize the expected loss. Similarly, if such a loss function exists we say that it is a consistent scoring function for the considered risk measure. This mathematical property is growing in popularity in the banking industry while being extensively used by quantitative risk managers for backtesting industry standard risk measures like the VaR or more recently the ES. As it is the case for ES, the marginal expected shortfall (MES) is not elicitable if we consider the risk measure stand-alone. However, in Chapter 4, we establish the elicibility of the MES jointly with the VaR of the market return. More precisely, we identify a strictly consistent scoring

function for the pair of risk measures (VaR, MES). Our framework is then extended to any systemic-risk measures that can be expressed as an expected equity conditional to a financial crisis, as it is the case for the SRISK for instance. These systemic-risk measures are widely used by academics and regulators to quantify how much an entity, be it a bank or hedge fund or sovereign, contributes to the vulnerability of the financial system. We exploit the scoring function to develop a semi-parametric M-estimator for the pair (VaR, MES). We show the consistency of the estimator. We also characterize the asymptotic distribution of the estimator which makes possible to model or to perform statistical inference for the systemic-risk indicators. We also give a discussion on how to perform model comparison and backtesting for systemic-risk measures under our framework. A small empirical study applied to four large US banks is provided to highlight the merit of our elicitation framework. A natural step forward of this work would be to propose a backtesting procedure based on the identification functions (derivatives of the scoring function) of the pair (VaR, MES) in order to assess the validity of the MES forecasts.

Bibliography

- Acerbi, C., 2002. Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking and Finance* 26, 1505–1518.
- Acerbi, C., Szekely, B., 2014. Backtesting expected shortfall. *Risk* 27, 76–81.
- Acerbi, C., Tasche, D., 2002. On the coherence of expected shortfall. *Journal of Banking & Finance* 26, 1487–1503.
- Acharya, V., Engle, R., Richardson, M., 2012. Capital shortfall: A new approach to ranking and regulating systemic risks. *American Economic Review* 102, 59–64.
- Acharya, V., Pedersen, L., Philippon, T., Richardson, M., 2017. Measuring systemic risk. *Review of Financial Studies* 30, 2–47.
- Adrian, T., Brunnermeier, M., 2016. Covar. *American Economic Review* 106, 1705–1741.
- Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International economic review* , 885–905.
- Andersen, T.G., Bollerslev, T., Christoffersen, P.F., Diebold, F.X., 2006. Volatility and correlation forecasting. *Handbook of economic forecasting* 1, 777–878.
- Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. *Econometrica* 71, 579–625.
- Andrews, D.W.K., 1994. Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica* 62, 43–72. URL: <http://www.jstor.org/stable/2951475>.
- Ang, A., Chen, J., 2002. Asymmetric correlations of equity portfolios. *Journal of financial Economics* 63, 443–494.
- Argyropoulos, C., Panopoulou, E., 2016. A survey on risk forecast evaluation. Working paper .
- Argyropoulos, C., Panopoulou, E., 2019. Backtesting var and es under the magnifying glass. *International Review of Financial Analysis* 64, 22–37.
- Artzner, P., Delbaen, F., Eber, J.M., Heath, D., 1999. Coherent measures of risk. *Mathematical Finance* 9, 203–228.
- Aue, A., Horváth, L., F. Pellatt, D., 2017. Functional generalized autoregressive conditional heteroskedasticity. *Journal of Time Series Analysis* 38, 3–21.

- Banerjee, A., Merugu, S., Dhillon, I.S., Ghosh, J., 2005. Clustering with bregman divergences. *Journal of Machine Learning Research* 6, 1705–1749. URL: <http://www.jmlr.org/papers/volume6/banerjee05b/banerjee05b.pdf>.
- 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.
- Banulescu-Radu, D., Hurlin, C., Leymarie, J., Scaillet, O., 2021. Backtesting marginal expected shortfall and related systemic risk measures. *Management Science* 67, 5301–5967.
- Bardet, J.M., Wintenberger, O., 2009. Asymptotic normality of the quasi maximum likelihood estimator for multidimensional causal processes. *Annals of Statistics* 37, 2730–2759.
- Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2008. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica* 76, 1481–1536.
- Barndorff-Nielsen, O.E., Shephard, N., 2002. Estimating quadratic variation using realized variance. *Journal of Applied econometrics* 17, 457–477.
- 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, 2013. Global systemically important banks: Updated assessment methodology and the higher loss absorbency requirements. Report, July .
- Basel Committee on Banking Supervision, 2014. The g-sib assessment methodology - score calculation. Report, November .
- Basel Committee on Banking Supervision, 2016. Minimum capital requirements for market risk. Consultation paper, January .
- Basel Committee on Banking Supervision, 2019. Minimum capital requirements for market risk. Consultation paper, February .
- Bauwens, L., Laurent, S., Rombouts, J.V., 2006. Multivariate garch models: a survey. *Journal of applied econometrics* 21, 79–109.
- Bayer, S., Dimitriadis, T., 2020. Regression-Based Expected Shortfall Backtesting. *Journal of Financial Econometrics* .
- Bee, M., Dupuis, D.J., Trapin, L., 2018. Realized extreme quantile: A joint model for conditional quantiles and measures of volatility with evt refinements. *Journal of Applied Econometrics* 33, 398–415.
- Begley, T., Purnanandam, A., Zheng, K., 2017. The strategic underreporting of bank risk. *Review of Financial Studies* 30, 3376–3415.
- Benoit, S., Colliard, J.E., Hurlin, C., Pérignon, C., 2017. Where the risks lie: A survey on systemic risk. *Review of Finance* 21, 109–152.

- Benoit, S., Couperier, O., Leymarie, J., Scaillet, O., 2022. Elicitability of marginal expected shortfall and related systemic-risk measures. Working paper .
- Benoit, S., Hurlin, C., Pérignon, C., 2015. Implied risk exposures. *Review of Finance* 19, 2183–2222.
- Benoit, S., Hurlin, C., Pérignon, C., 2019. Pitfalls in systemic-risk scoring. Forthcoming in *Journal of Financial Intermediation* .
- Berkes, I., Horváth, L., Kokoszka, P., 2003a. Asymptotics for garch squared residual correlations. *Econometric Theory* 19, 515–540.
- Berkes, I., Horváth, L., Kokoszka, P., 2003b. Garch processes: structure and estimation. *Bernoulli* 9, 201–227.
- Berkowitz, J., 2001. Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics* 19, 465–474.
- Berkowitz, J., Christoffersen, P., Pelletier, D., 2011. Evaluating value at risk models with desk level data. *Management Science* 57, 2213–2272.
- Berkowitz, J., O'Brien, J., 2002. How accurate are value-at-risk models at commercial banks? *Journal of Finance* 57, 1093–1111.
- Bernardi, M., Catania, L., 2016. Comparison of value-at-risk models using the mcs approach. *Computational Statistics* 31, 579–608.
- Bernardi, M., Catania, L., 2019. Switching generalized autoregressive score copula models with application to systemic risk. *Journal of Applied Econometrics* 34, 43–65.
- Billingsley, P., 1961. The lindeberg-levy theorem for martingales. *Proceedings of the American Mathematical Society* 12, 788–792.
- 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, 535–559.
- Bisias, D., Flood, M., Lo, A., Valavanis, S., 2012. A survey of systemic risk analytics. *Annual Review of Financial Economics* 4, 255–296.
- Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics* 31, 307–327.
- Boucher, C., Daniélfsson, 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, 1509–1526.
- Brandt, M., Diebold, F., 2006. A no-arbitrage approach to range-based estimation of return covariances and correlations. *Journal of Business* 79, 61–73.

- Bregman, L., 1967. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics* 7, 200 – 217. URL: <http://www.sciencedirect.com/science/article/pii/0041555367900407>, doi:10.1016/0041-5553(67)90040-7.
- Brockwell, P., Davis, R., 1991. in: *Time series: theory and methods*. New-York: Springer, Second Edition.
- Brownlees, C., Cavaliere, G., Monti, A., 2018. Evaluating the accuracy of tail risk forecasts for systemic risk measurement. *Annals of Financial Economics* 13, 1850009.
- Brownlees, C., Chabot, B., Ghysels, E., Kurz, C., 2020. Back to the future: Backtesting systemic risk measures during historical bank runs and the great depression. *Journal of Banking & Finance* 113, 105736.
- Brownlees, C., Engle, R., 2017. Srisk: A conditional capital shortfall measure of systemic risk. *Review of Financial Studies* 30, 48–79.
- Brunk, H.D., 1955. Maximum likelihood estimates of monotone parameters. *The Annals of Mathematical Statistics* , 607–616.
- Cai, Z., Wang, X., 2008. Nonparametric estimation of conditional var and expected shortfall. *Journal of Econometrics* 147, 120–130.
- Campbell, S.D., 2007. A review of backtesting and backtesting procedures. *Journal of Risk* 9, 1–18.
- Candelon, B., Colletaz, G., Hurlin, C., Tokpavi, S., 2011. Backtesting value-at-risk: A gmm duration-based test. *Journal of Financial Econometrics* 9, 314–343.
- Carbon, M., Francq, C., 2011. Portmanteau goodness-of-fit test for asymmetric power garch models. *Austrian Journal of Statistics* 40, 55–64.
- Cerovecki, C., Francq, C., Hörmann, S., Zakoïan, J.M., 2019. Functional garch models: the quasi-likelihood approach and its applications. *Journal of econometrics* 209, 353–375.
- Chen, C., Iyengar, G., Moallemi, C.C., 2013. An axiomatic approach to systemic risk. *Management Science* 59, 1373–1388.
- Chen, S.X., 2007. Nonparametric estimation of expected shortfall. *Journal of Financial Econometrics* 6, 87–107.
- Chen, S.X., Tang, C.Y., 2005. Nonparametric inference of value-at-risk for dependent financial returns. *Journal of Financial Econometrics* 3, 227–255.
- Chen, X., Linton, O., Van Keilegom, I., 2003. Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* 71, 1591–1608.
- Chen, Y., Xu, X., Koch, T., Zhang, G., 2020. Regularized partially functional autoregressive model. Available at SSRN 3482262 .
- Chernozhukov, V., 2005. Extremal quantile regression. *Annals of Statistics* 33, 806–839.

- Chernozhukov, V., Fernández-Val, I., 2011. Inference for extremal conditional quantile models, with an application to market and birthweight risks. *Review of Economic Studies* 78, 559–589.
- Chicago Mercantile Exchange, 2012. Cme clearing risk management and financial safeguards. Report .
- Chortareas, G., Jiang, Y., Nankervis, J., 2011. Forecasting exchange rate volatility using high-frequency data: Is the euro different? *International Journal of Forecasting* 27, 1089–1107.
- 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, 84–108.
- Christoffersen, P.F., 1998. Evaluating interval forecasts. *International Economic Review* , 841–862.
- Colletaz, G., Hurlin, C., Pérignon, C., 2013. The risk map: A new tool for validating risk models. *Journal of Banking & Finance* 37, 3843–3854.
- Corsi, F., 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7, 174–196.
- Costanzino, N., Curran, M., 2015. Backtesting general spectral risk measures with application to expected shortfall. *Journal of Risk Model Validation* 9, 21–31.
- Costanzino, N., Curran, M., 2018. A simple traffic light approach to backtesting expected shortfall. *Risks* 6, 1–7.
- Couperier, O., Francq, C., Zakoïan, J.M., 2022. Daily volatility forecasting using intraday returns and functional covariates. Working paper .
- Couperier, O., Leymarie, J., 2021. Backtesting expected shortfall via multi-quantile regression. Working paper .
- Creal, D., Koopman, S.J., Lucas, A., 2013. Generalized autoregressive score models with applications. *Journal of Applied Econometrics* 28, 777–795.
- Daniëlsson, J., Zhou, C., 2016. Why risk is so hard to measure. DNB working paper .
- 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 .
- De Nicolò, G., Lucchetta, M., 2017. Forecasting tail risks. *Journal of Applied Econometrics* 32, 159–170.
- De Roon, F., Karehnke, P., 2017. A simple skewed distribution with asset pricing applications. *Review of Finance* 21, 2169–2197.
- Delsol, L., Van Keilegom, I., 2020. Semiparametric m-estimation with non-smooth criterion functions. *Annals of the Institute of Statistical Mathematics* 72, 577–605.

- Dette, H., Neumeyer, N., Pilz, K.F., 2006. A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli* 12, 469–490. URL: <https://www.jstor.org/stable/25464816>.
- Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13, 253–263.
- Diewert, W., Avriel, M., Zang, I., 1981. Nine kinds of quasiconcavity and concavity. *Journal of Economic Theory* 25, 397 – 420. URL: <http://www.sciencedirect.com/science/article/pii/0022053181900399>, doi:10.1016/0022-0531(81)90039-9.
- Dimitriadis, T., Liu, X., Schnaitmann, J., 2021. Encompassing tests for value at risk and expected shortfall multi-step forecasts based on inference on the boundary. Forthcoming in *Journal of Financial Econometrics* .
- Dimitriadis, T., Schnaitmann, J., 2021. Forecast encompassing tests for the expected shortfall. *International Journal of Forecasting* 37, 604–621.
- Dowd, K., 2001. Estimating var with order statistics. *The Journal of Derivatives* 8, 23–30.
- Du, Z., Escanciano, J.C., 2017. Backtesting expected shortfall: Accounting for tail risk. *Management Science* 63, 901–1269.
- Duarte, F., Eisenbach, T.M., 2021. Fire-sale spillovers and systemic risk. *The Journal of Finance* 76, 1251–1294.
- Dumitrescu, E.I., Hurlin, C., Pham, V., 2012. Backtesting value-at-risk: From dynamic quantile to dynamic binary tests. *Finance* 33, 79–111.
- Eckernkemper, T., 2018. Modeling systemic risk: time-varying tail dependence when forecasting marginal expected shortfall. *Journal of Financial Econometrics* 16, 63–117.
- Emmer, S., Kratz, M., Tasche, D., 2015. What is the best risk measure in practice? a comparison of standard measures. *Journal of Risk* 18, 31–60.
- Engle, R., 2002a. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics* 20, 339–350.
- Engle, R., 2002b. New frontiers for arch models. *Journal of Applied Econometrics* 17, 425–446.
- Engle, R., Jondeau, E., Rockinger, M., 2015. Systemic risk in europe. *Review of Finance* 19, 145–190.
- Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica* , 987–1007.
- Engle, R.F., Manganelli, S., 2004. Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics* 22, 367–381.
- Engle, R.F., Patton, A.J., 2001. What good is a volatility model. *Quantitative Finance* 1, 237–245.
- Escanciano, J.C., Olmo, J., 2010. Backtesting parametric value-at-risk with estimation risk. *Journal of Business & Economic Statistics* 28, 36–51.

- European Banking Authority, 2014. Results of 2014 eu-wide stress test. Report, October 26 .
- Feinstein, Z., Rudloff, B., Weber, S., 2017. Measures of systemic risk. *SIAM Journal on Financial Mathematics* 8, 672–708. doi:10.1137/16M1066087.
- 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 .
- Fissler, T., Hlavinová, J., Rudloff, B., 2019. Elicitability and identifiability of systemic risk measures. arXiv , arXiv-1907.
- Fissler, T., Hoga, Y., 2021. Backtesting systemic risk forecasts using multi-objective elicibility. Working paper .
- Fissler, T., Ziegel, J., 2016. Higher order elicibility and osband’s principle. *Annals of Statistics* 44, 1680–1707.
- Fissler, T., Ziegel, J.F., Gneiting, T., 2016. Expected shortfall is jointly elicitable with value-at-risk: Implications for backtesting. *Journal of Risk* , 58–61.
- Francq, C., Thieu, L.Q., 2019. Qml inference for volatility models with covariates. *Econometric Theory* 35, 37–72.
- Francq, C., Wintenberger, O., Zakoïan, J.M., 2013. Garch models without positivity constraints: Exponential or log garch? *Journal of Econometrics* 177, 34–46.
- Francq, C., Wintenberger, O., Zakoïan, J.M., 2018. Goodness-of-fit tests for log-garch and egarch models. *Test* 27, 27–51.
- Francq, C., Zakoïan, J.M., 2004. Maximum likelihood estimation of pure garch and arma-garch. *Bernoulli* 10, 605–637.
- Francq, C., Zakoïan, J.M., 2015. Risk-parameter estimation in volatility models. *Journal of Econometrics* 184, 158–173.
- Francq, C., Zakoïan, J.M., 2019. in: *GARCH Models: structure, statistical inference and financial applications*. John Wiley, Second Edition.
- Freedman, D., 1981. Bootstrapping regression models. *Annals of Statistics* 9, 1218–1228.
- Fuertes, A., Izzeldin, M., Kalotychou, E., 2009. On forecasting daily stock volatility: The role of intraday information and market conditions. *International Journal of Forecasting* 25, 259–281.
- 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, 150–160.
- Garman, M.B., Klass, M.J., 1980. On the estimation of security price volatilities from historical data. *Journal of business* , 67–78.

- Geweke, J., 1986. Modelling persistence in conditional variances: A comment. *Econometric Review* 5, 57–61.
- Ghysels, E., Santa-Clara, P., Valkanov, R., 2004. The midas touch: Mixed data sampling regression models. Working paper .
- Ghysels, E., Santa-Clara, P., Valkanov, R., 2006. Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics* 131, 59–95.
- Giacomini, R., Rossi, B., 2009. Detecting and predicting forecast breakdowns. *The Review of Economic Studies* 76, 669–705. URL: <http://www.jstor.org/stable/40247617>, doi:10.1111/j.1467-937X.2009.00545x.
- Girardi, G., Ergün, T., 2013. Systemic risk measurement: Multivariate garch estimation of covar. *Journal of Banking & Finance* 37, 3169–3180.
- Glosten, L.R., Jagannathan, R., Runkle, D.E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance* 48, 1779–1801.
- Gneiting, T., 2011a. Making and evaluating point forecasts. *Journal of the American Statistical Association* 106, 746–762.
- Gneiting, T., 2011b. Quantiles as optimal point forecasts. *International Journal of forecasting* 27, 197–207.
- Gourieroux, C., Laurent, J.P., Scaillet, O., 2000. Sensitivity analysis of values at risk. *Journal of empirical finance* 7, 225–245.
- Gouriéroux, C., Liu, W., 2012. Converting tail-var to var: An econometric study. *Journal of Financial Econometrics* 10, 233–264.
- Gouriéroux, C., Zakoïan, J.M., 2013. Estimation-adjusted var. *Econometric Theory* 29, 735–770.
- Hall, P., Huang, L.S., 2001. Nonparametric kernel regression subject to monotonicity constraints. *The Annals of Statistics* 29, 624–647.
- Hansen, P.R., Huang, Z., Shek, H.H., 2012. Realized garch: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics* 27, 877–906.
- Hansen, P.R., Lunde, A., 2006a. Consistent ranking of volatility models. *Journal of Econometrics* 131, 97–121.
- Hansen, P.R., Lunde, A., 2006b. Realized variance and market microstructure noise. *Journal of Business & Economic Statistics* 24, 127–161.
- Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. *Econometrica* 79, 453–497.
- Hautsch, N., Herrera, R., 2020. Multivariate dynamic intensity peaks-over-threshold models. *Journal of Applied Econometrics* 35, 248–272.
- Ho, H.J., 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, 25–40.

- Hoga, Y., 2020. Modeling time-varying tail dependence, with application to systemic risk forecasting. *Journal of Financial Econometrics* .
- Hörmann, S., Horváth, L., Reeder, R., 2013. A functional version of the arch model. *Econometric Theory* 29, 267–288.
- 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., Tokpavi, S., 2006. Backtesting var accuracy: a new simple test. *Journal of Risk* 9, 19–37.
- Ichimura, H., Lee, S., 2010. Characterization of the asymptotic distribution of semi-parametric m-estimators. *Journal of Econometrics* 159, 252 – 266. URL: <https://www-sciencedirect-com.proxy.bu.dauphine.fr/science/article/pii/S0304407610001302>, doi:10.1016/j.jeconom.2010.05.005.
- 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 .
- Jouini, E., Schachermayer, W., Touzi, N., 2006. Law invariant risk measures have the fatou property, in: Kusuoka, S., Yamazaki, A. (Eds.), *Advances in Mathematical Economics*. Springer, Tokyo. volume 9. URL: https://doi.org/10.1007/4-431-34342-3_4.
- Kerkhof, J., Melenberg, B., 2004. Backtesting for risk-based regulatory capital. *Journal of Banking & Finance* 28, 1845–1865.
- Koenker, R., Bassett, G., 1978. Regression quantiles. *Econometrica: journal of the Econometric Society* , 33–50.
- Koenker, R., Chernozhukov, V., He, X., Peng, L., 2018. *Handbook of quantile regression*. Chapman and Hall/CRC Handbooks of Modern Statistical Methods .
- Koenker, R., Machado, J., 1999. Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistiscal Association* 94, 1296–1310.
- Koenker, R., Xiao, Z., 2002. Inference on the quantile regression process. *Econometrica* 70, 1583–1612.
- Komunjer, I., 2007. Asymmetric power distribution: Theory and applications to risk measurement. *Journal of Applied Econometrics* 22, 891–921.
- Koopman, S., Jungbacker, B., Hol, E., 2005. Forecasting daily variability of the s&p 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance* 12, 445–475.

- 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.
- Kühnert, S., 2020. Functional arch and garch models: A yule-walker approach. *Electronic Journal of Statistics* 14, 4321–4360.
- Kupiec, P., 1995. Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives* 3, 73–84.
- Kusuoka, S., 2001. On law invariant coherent risk measures, in: Kusuoka, S., Maruyama, T. (Eds.), *Advances in Mathematical Economics*. Springer, Tokyo. volume 3. URL: https://doi.org/10.1007/978-4-431-67891-5_4.
- Lambert, N.S., 2013. Elicitation and Evaluation of Statistical Forecasts. Stanford University. URL: https://web.stanford.edu/~nlambert/papers/elicitacion_july2013.pdf.
- Lambert, N.S., Pennock, D.M., Shoham, Y., 2008. Eliciting properties of probability distributions, in: *Proceedings of the 9th ACM Conference on Electronic Commerce*, pp. 129–138.
- Laurent, S., Rombouts, J.V., Violante, F., 2013. On loss functions and ranking forecasting performances of multivariate volatility models. *Journal of Econometrics* 173, 1–10.
- Lazar, E., Zhang, N., 2019. Model risk of expected shortfall. *Journal of Banking & Finance* 105, 74–93.
- Leucht, A., Kreiss, J.P., Neumann, M., 2015. A model specification test for garch(1,1) processes. *Scandinavian Journal of Statistics* 42, 1167–1193.
- Li, W., 2004. in: *Diagnostic checks in time series*. Boca Raton, Florida: Chapman and Hall.
- Li, W., Mak, T., 1994. On the squared residual autocorrelations in non-linear time series with conditional heteroscedasticity. *Journal of Time Series Analysis* 15, 627–636.
- Lin, E.M., Sun, E.W., Yu, M.T., 2018. Systemic risk, financial markets, and performance of financial institutions. *Annals of Operations Research* 262, 579–603.
- Ling, S., Li, W., 1997. On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *Journal of the American Statistical Association* 92, 1184–1194.
- Ljung, G., Box, G., 1978. On a measure of lack of fit in time series models. *Biometrika* 65, 297–303.
- Löffler, G., Raupach, P., 2018. Pitfalls in the use of systemic risk measures. *Journal of Financial and Quantitative Analysis* 53, 269–298.
- Löser, R., Wied, D., Ziggel, D., 2019. New backtests for unconditional coverage of expected shortfall. *Journal of Risk* 21, 39–59.
- Lyócsa, S., Molnár, P., Výrost, T., 2021. Stock market volatility forecasting: Do we need high-frequency data? *International Journal of Forecasting* 37, 1092–1110.
- Mandelbrot, B., 1963. The variation of certain speculative prices. *Journal of Business* 36, 394–419.
- Markowitz, H., 1952. Portfolio selection. *Journal of Finance* 7, 77–91.

- Martins-Filho, C., Yao, F., Torero, M., 2018. Nonparametric estimation of conditional value-at-risk and expected shortfall based on extreme value theory. *Econometric Theory* 34, 23–67.
- 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, 271–300.
- Milhoj, A., 1987. A multiplicative parameterization of arch models. university of copenhagen, department of statistics. Working paper Department of Statistics, University of openhagen.
- Mincer, J.A., Zarnowitz, V., 1969. The evaluation of economic forecasts, in: *Economic forecasts and expectations: Analysis of forecasting behavior and performance*. NBER, pp. 3–46.
- Morgan, J., 1996. *Riskmetrics*. Technical Document (4th ed.), New York: Morgan Guaranty Trust Company of New York .
- Nadarajah, S., 2007. A truncated bivariate t distribution. *Stochastics and Quality Control* 22, 303–313.
- Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* , 347–370.
- Newey, K., McFadden, D., 1994. Large sample estimation and hypothesis. *Handbook of Econometrics, IV*, Edited by RF Engle and DL McFadden , 2112–2245.
- Newey, W., West, K., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
- Newey, W.K., 1994. The asymptotic variance of semiparametric estimators. *Econometrica* 62, 1349–1382. URL: <http://www.jstor.org/stable/2951752>.
- Nolde, N., Zhang, J., 2020. Conditional extremes in asymmetric financial markets. *Journal of Business & Economic Statistics* 38, 201–213.
- Nolde, N., Ziegel, J.F., 2017. Elicitability and backtesting: Perspectives for banking regulation. *Annals of Applied Statistics* 11, 1833–1874.
- Osband, K.H., 1985. *Providing Incentives for Better Cost Forecasting (Prediction, Uncertainty Elicitation)*. Ph.D. thesis. University of California, Berkeley.
- Padé, H., 1892. Sur la représentation approchée d’une fonction par des fractions rationnelles. *Annales scientifiques de l’École Normale Supérieure* 9, 3–93.
- Pantula, S., 1986. Modelling the persistence of conditional variances: a comment. *Econometric Review* 5, 71–74.
- Parkinson, M., 1980. The extreme value method for estimating the variance of the rate of return. *Journal of business* , 61–65.
- Patton, A.J., 2011. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* 160, 246–256.
- Patton, A.J., Sheppard, K., 2009. Optimal combinations of realised volatility estimators. *International Journal of Forecasting* 25, 218 – 238. doi:10.1016/j.ijforecast.2009.01.011. forecasting Returns and Risk in Financial Markets using Linear and Nonlinear Models.

- 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.
- Philippon, T., Pessarossi, P., Camara, B., 2017. Backtesting european stress tests. NBER working paper .
- Powell, J.L., 1984. Least absolute deviations estimation for the censored regression model. *Journal of Econometrics* 25, 303–325.
- Rice, G., Wirjanto, T., Zhao, Y., 2020. Forecasting value at risk with intra-day return curves. *International Journal of Forecasting* 36, 1023–1038.
- Roncalli, T., 2014. Introduction to risk parity and budgeting. Chapman and Hall/CRC Financial Mathematics Series .
- Rosenbaum, S., 1961. Moments of a truncated bivariate normal distribution. *Journal of the Royal Statistical Society: Series B (Methodological)* 23, 405–408.
- Savage, L.J., 1971. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association* 66, 783–801. URL: <https://www-jstor-org.proxy.bu.dauphine.fr/stable/2284229>, doi:10.2307/2284229.
- Scaillet, O., 2004. Nonparametric estimation and sensitivity analysis of expected shortfall. *Mathematical Finance* 14, 115–129.
- Scaillet, O., 2005. Nonparametric estimation of conditional expected shortfall. *Insurance and Risk Management Journal* 74, 639–660.
- Schaible, S., Ziemba, W., 1981. Generalized concavity in optimization and economics. 1, Academic Press. URL: <https://books.google.fr/books?id=sE0gAAAAMAAJ>.
- Sender, H., Wigglesworth, R., 2015. Investing: Whatever the weather? *Financial Times* .
- Steinwart, I., Pasin, C., Williamson, R., Zhang, S., 2014. Elicitation and identification of properties. *Journal of Machine Learning Research* 35, 482–526. URL: <http://proceedings.mlr.press/v35/steinwart14.pdf>.
- Sucarrat, G., Escibano, A., 2012. Automated model selection in finance: General-to-specific modelling of the mean and volatility specifications. *Oxford Bulletin of Economics and Statistics* 74, 716–735.
- Sucarrat, G., Grønneberg, S., Escibano, A., 2016. Estimation and inference in univariate and multivariate log-garch-x models when the conditional density is unknown. *Computational statistics & data analysis* 100, 582–594.
- Sun, H., Yu, B., 2020. Volatility asymmetry in functional threshold garch model. *Journal of Time Series Analysis* 41, 95–109.
- Tarashev, N., Tsatsaronis, K., Borio, C., 2016. Risk attribution using the shapley value: Methodology and policy applications. *Review of Finance* 20, 1189–1213.

- Tasche, D., 2002. Expected shortfall and beyond. *Journal of Banking & Finance* 26, 1519 – 1533. URL: <https://www-sciencedirect-com.proxy.bu.dauphine.fr/science/article/pii/S0378426602002728>, doi:10.1016/S0378-4266(02)00272-8.
- Taylor, J., 2020. Forecast combinations for value at risk and expected shortfall. *International Journal of Forecasting* 36, 428 – 441. doi:10.1080/07350015.2017.1281815.
- 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, 121–133.
- Van Oordt, M., Zhou, C., 2019. Systemic risk and bank business models. *Journal of Applied Econometrics* 34, 365–384.
- Wang, R., Zitikis, R., 2021. An axiomatic foundation for the expected shortfall. *Management Science* 67, 1413–1429.
- White, H., 1994. *Estimation, inference and specification analysis* Cambridge University Press, New York.
- White, H., 2001. *Asymptotic theory for econometricians* Academic Press, San Diego.
- White, H., Kim, T.H., Manganelli, S., 2015. Var for var: Measuring tail dependence using multivariate regression quantiles. *Journal of Econometrics* 187, 169–188.
- Wong, W.K., 2008. Backtesting trading risk of commercial banks using expected shortfall. *Journal of Banking & Finance* 32, 1404–1415.
- Zakoian, J.M., 1994. Threshold heteroskedastic models. *Journal of Economic Dynamics and control* 18, 931–955.
- Zhang, L., Mykland, P.A., Ait-Sahalia, Y., 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100, 1394–1411.
- Zhang, Q., Vallascas, F., Keasy, K., Cai, C.X., 2015. Are market-based measures of global systemic importance of financial institutions useful to regulators and supervisors? *Journal of Money, Credit and Banking* 47, 1403–1442. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/jmcb.12249>, doi:10.1111/jmcb.12249, arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1111/jmcb.12249>.
- Zhou, B., 1996. High-frequency data and volatility in foreign-exchange rates. *Journal of Business & Economic Statistics* 14, 45–52.
- Ziegel, J.F., 2016. Coherence and elicibility. *Mathematical Finance* 26, 901–918. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/mafi.12080>, doi:10.1111/mafi.12080, arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1111/mafi.12080>.
- Ziegel, J.F., Krüger, F., Jordan, A., Fasciati, F., 2020. Robust forecast evaluation of expected shortfall. *Journal of financial econometrics* 18, 95–120.

Résumé en Français

Cette thèse contribue à la recherche académique en économétrie et en gestion des risques. Nous poursuivons un double objectif : (i) quantifier les risques financiers encourus par les institutions financières et (ii) évaluer la validité des mesures de risque couramment utilisées dans l'industrie financière ou par les régulateurs. Nous nous concentrons sur trois types de risques financiers, (i) le risque de crédit, (ii) le risque de marché et (iii) le risque systémique. Dans le Chapitre 2 et le Chapitre 3, nous développons de nouvelles méthodes de modélisation et d'évaluation adaptées à la volatilité et à l'*Expected Shortfall*. Ces deux mesures de risque sont généralement utilisées pour quantifier le risque dans les portefeuilles d'investissement. Dans le Chapitre 4, nous proposons de nouvelles méthodes d'estimation dédiées aux mesures du risque systémique. Ces dernières ont pour but d'identifier les institutions financières qui contribuent le plus au risque global du système financier.

Il est maintenant largement admis que les rendements intra-journaliers sont des prédicteurs significatifs de la dynamique de la volatilité quotidienne. L'approche habituelle pour tenir compte des mouvements intra-journaliers consiste à introduire un résumé des rendements intra-journaliers par le biais de mesures de la volatilité réalisée. L'ajout de variables exogènes scalaires dans l'équation de la volatilité a donné naissance à la classe des processus de type GARCH-X. Dans le Chapitre 2, nous développons une structure de volatilité qui regroupe l'ensemble de la séquence des rendements intra-journaliers en tant que covariables fonctionnelles. Contrairement au modèle GARCH-X qui inclut un ré-

sumé des rendements intra-journaliers, notre approche permet de rendre compte de toute l'information contenue dans les mouvements de prix intra-journaliers en utilisant l'analyse des données fonctionnelles (ADF). Cette dernière permet de transformer les données observées en courbes. Nos apports sont les suivants. Premièrement, nous développons un modèle log-GARCH fonctionnel dans lequel la variable de réponse est un scalaire et les variables explicatives peuvent comporter des variables de type scalaires et fonctionnelles. Deuxièmement, nous montrons l'existence de solutions de stationnarité, moments et log-moments, pour le modèle considéré. Troisièmement, nous introduisons un estimateur du quasi-maximum de vraisemblance et détaillons les propriétés asymptotiques correspondantes. Enfin, nous développons une statistique portemanteau pour tester la qualité de l'ajustement du modèle. Dans la partie empirique, nous appliquons notre modèle aux rendements intra-journaliers d'un panel de dix actifs NYSE/NASDAQ. Notre variable fonctionnelle est échantillonnée à une fréquence de 5 minutes. Nos constatations sont les suivantes. Premièrement, nous concluons à des résidus non-autocorrélés dans le modèle dans plus de 80% des cas et pour sept actifs sur dix. Deuxièmement, nous constatons que notre variable fonctionnelle est généralement statistiquement significative, ce qui confirme que toute la séquence intra-journalière doit être utilisée pour expliquer les rendements carrés logarithmiques quotidiens. Enfin, nous ajustons trois modèles concurrents (comportant soit la volatilité réalisée et un effet de levier, soit un effet de levier, soit aucun des deux) qui sont des représentations imbriquées de notre modèle. Nous sélectionnons les meilleurs modèles de volatilité en utilisant la procédure *model confidence set (MCS)*. Notre modèle appartient à l'ensemble constitué des meilleurs modèles de prévisions de la volatilité à une fréquence plus élevée que pour les modèles concurrents pour sept actifs sur dix. Une avancée naturelle de ce travail consisterait à développer l'extension multivariée de ce modèle.

Le Comité de Bâle sur le contrôle bancaire a adopté les accords de Bâle III pour améliorer la capacité du secteur bancaire à absorber les chocs résultants de tensions financières et économiques. Parmi les réformes fondamentales, la *Value-at-Risk* (VaR) a été remplacée par l'*Expected Shortfall* (ES) pour le calcul des exigences de fonds propres pour le risque de marché. L'ES offre un certain nombre de propriétés attrayantes qui répondent à des défauts de la VaR. En particulier, l'ES est une mesure de risque cohérente. Cependant, l'ES n'est pas élicitable contrairement à la VaR, ce qui rend son évaluation plus délicate. Le Chapitre 3 présente une méthodologie économétrique pour tester la validité des prévisions d'ES pour les portefeuilles de marché. Nous exploitons le fait que l'ES peut être approximée en utilisant une séquence finie de VaR. Notre méthodologie examine conjointement la validité des prévisions de VaR le long de la queue de distribution du modèle de risque. Notre procédure englobe la recommandation du Comité de Bâle qui vise à valider l'ES en examinant deux VaRs à des niveaux de risque spécifiques. Notre approche économétrique consiste à régresser les pertes ex post sur les prévisions de VaR dans un modèle de régression multi-quantiles, puis à tester les paramètres estimés résultants à l'aide d'inférences de type Wald. Notre cadre de régression permet de dériver une correction analytique pour ajuster les prévisions incorrectes. Dans une application empirique, nous mettons en évidence la capacité de nos backtests à rejeter un modèle d'ES mal spécifié. Nous montrons également qu'il est crucial de choisir un nombre suffisant de quantiles pour évaluer les modèles de risque d'ES. En particulier, nous constatons qu'utiliser un ou deux quantiles n'est pas toujours suffisant pour déterminer si le risque extrême est correctement pris en compte dans les modèles internes d'ES de la banque. Au contraire, quatre quantiles ou plus (jusqu'à un nombre optimal) délivrent des décisions beaucoup plus judicieuses, suggérant une mise à jour des recommandations réglementaires. Ce Chapitre pourrait être étendu pour évaluer l'ensemble de la distribution des rendements (dans le cadre de prévisions d'intervalle et de densité).

La propriété d'élicitabilité ouvre la voie à de nombreuses applications utiles : la sélection de modèles, l'estimation, la régression, le classement de modèles de prévisions et le backtesting. Une mesure de risque est dite élicitable s'il existe une fonction de perte ou une fonction de score telle que la mesure de risque elle-même est la solution pour minimiser la perte attendue. De même, si une telle fonction de perte existe, nous disons qu'il s'agit d'une fonction de perte cohérente pour la mesure de risque considérée. Cette propriété mathématique gagne en popularité dans le secteur bancaire tout en étant largement utilisée par les gestionnaires de risques quantitatifs pour le backtesting des mesures de risque standards de l'industrie comme la VaR ou plus récemment l'ES. Comme c'est le cas pour l'ES, la *Marginal Expected Shortfall* (MES) n'est pas élicitable si l'on considère la mesure de risque seule. Cependant, dans le Chapitre 4, nous montrons que la MES est élicitable conjointement avec la VaR du rendement du marché. Plus précisément, nous identifions une fonction de perte strictement cohérente pour le couple de mesures de risque (VaR, MES). Notre cadre est ensuite étendu à toutes les mesures de risque systémique qui peuvent être exprimées en fonds propres attendus conditionnels à une crise financière, comme c'est le cas pour la SRISK par exemple. Ces mesures du risque systémique sont largement utilisées par les universitaires et les régulateurs pour quantifier dans quelle mesure une entité, qu'il s'agisse d'une banque, d'un fond spéculatif ou d'un État, contribue à la vulnérabilité du système financier. Nous exploitons la fonction de perte pour développer un M-estimateur semi-paramétrique pour la paire (VaR, MES). Nous montrons la consistance de l'estimateur. Nous caractérisons également sa distribution asymptotique qui permet de modéliser ou d'effectuer des inférences statistiques pour les indicateurs de risque systémique. Nous discutons également sur la façon d'effectuer une comparaison de modèles et un backtesting pour les mesures de risque systémique en exploitant la fonction de perte. Une petite étude empirique appliquée à quatre grandes banques américaines est fournie pour souligner le mérite de notre cadre d'étude. Une

avancée naturelle de ce travail serait de proposer une procédure de backtesting basée sur les fonctions d'identification (dérivées de la fonction de perte) du couple (VaR, MES) afin d'évaluer la validité des prévisions de MES.

Titre : Trois Essais en Économétrie Financière

Mots clés : Économétrie financière, Gestion des risques, Réglementation bancaire, Mesures de risque.

Résumé : Cette thèse contribue à la recherche académique en économétrie et en gestion des risques financiers. L'objectif est double : (i) quantifier les risques financiers encourus par les institutions financières et (ii) évaluer la validité des mesures de risque couramment utilisées dans l'industrie financière ou par les régulateurs. Nous nous concentrons sur trois types de risques financiers, (i) le risque de crédit, (ii) le risque de marché et (iii) le risque systémique. Dans les Chapitres 2 et 3, nous développons de nouvelles méthodes de modélisation et d'évaluation adaptées à la volatilité et à l'*Expected Shortfall* (ES), habituellement utilisées pour mesurer le risque de pertes en capital dans les portefeuilles d'investissement. Dans le Chapitre 4, nous nous intéressons aux mesures du risque systémique dont le but est d'identifier les institutions financières qui contribuent le plus au risque global du système financier.

Dans le Chapitre 2, nous développons une structure de volatilité qui regroupe l'ensemble de la séquence des rendements intra-journaliers en tant que covariables fonctionnelles. Contrairement au modèle GARCH à variables exogènes déjà existant (GARCH-X), notre approche permet de rendre compte de

toute l'information contenue dans les mouvements de prix intra-journaliers en utilisant l'analyse des données fonctionnelles. Le Chapitre 3 présente une méthodologie économétrique pour tester la validité des prévisions d'ES pour les portefeuilles de marché, mesure à présent utilisée pour le calcul des exigences de fonds propres pour le risque de marché suite à l'adoption des accords de Bâle III par le Comité de Bâle. Nous exploitons la relation entre l'ES et la *Value-at-Risk* (VaR). L'approche proposée est cohérente avec la recommandation du Comité de Bâle qui vise à valider l'ES en examinant deux VaRs à des niveaux de risque spécifiques. Dans le Chapitre 4, nous nous intéressons à la propriété d'élicitabilité des mesures de risque systémique construites à partir de données de marchés. Une mesure de risque est élicitable s'il existe une fonction de perte pour laquelle la mesure de risque est la quantité qui permet de minimiser la perte attendue. Nous identifions une fonction de perte strictement cohérente pour la *Marginal Expected Shortfall* (MES) et la VaR du rendement du marché conjointement et nous l'exploitons pour développer un M-estimateur semi-paramétrique de ce couple de mesures (VaR, MES).

Title : Three Essays in Financial Econometrics

Keywords : Financial econometrics, Risk management, Banking regulation, Risk measures.

Abstract : This dissertation contributes to the academic research in econometrics and financial risk management. Our research's goal is twofold: (i) to quantify the financial risks incurred by financial institutions and (ii) to assess the validity of the risk measures commonly used in the financial industry or by regulators. We focus on three kind of financial risks, (i) credit risk, (ii) market risk, and (iii) systemic risk. In chapters 2 and 3, we develop new methods for modeling and backtesting the volatility and the Expected Shortfall (ES), two measures typically used to quantify the risk of incurred losses in investment portfolios. In Chapter 4, we provide new estimation methods for the systemic risk measures that are used to identify the financial institutions contributing the most to the overall risk in the financial system.

In Chapter 2, we develop a volatility structure that groups the whole sequence of intraday returns as functional covariates. Contrary to the well-known GARCH model with exogenous variables (GARCH-X), our approach makes possible to account for the

whole information contained in the intraday price movements via functional data analysis. Chapter 3 introduces an econometric methodology to test for the validity of ES forecasts in market portfolios. This measure is now used to calculate the market risk capital requirements following the adoption of the Basel III accords by the Basel Committee on Banking Supervision (BCBS). Our method exploits the existing relationship between ES and Value-at-Risk (VaR) and complies - as a special case - with the BCBS recommendation of verifying the VaR at two specific risk levels. In Chapter 4, we focus on the elicibility property for the market-based systemic risk measures. A risk measure is said to be elicitable if there exists a loss function such that the risk measure itself is the solution to minimize the expected loss. We identify a strictly consistent scoring function for the Marginal Expected Shortfall (MES) and the VaR of the market return jointly and we exploit the scoring function to develop a semi-parametric M-estimator for the pair (VaR, MES).