UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE CIÊNCIAS ECONÔMICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

VICTOR FREITAS FEIO DE LEMOS

THE EDGE OF EXPECTIL: MODELLING EXPECTIL WITH EVT AND COMPARE IT USING COMPARATIVE BACKTEST

Porto Alegre

2018

VICTOR FREITAS FEIO DE LEMOS

THE EDGE OF EXPECTIL: MODELLING EXPECTIL WITH EVT AND COMPARE IT USING COMPARATIVE BACKTEST

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título de Mestre em Economia, com ênfase em Economia Aplicada.

Orientador: Hudson da Silva Torrent

Porto Alegre

2018

CIP - Catalogação na Publicação

Lemos, Victor Freitas Feio de
The Edge of expectil: modelling expectil with evt and compare it using comparative backtest / Victor Freitas Feio de Lemos. -- 2018.
49 f.
Orientador: Hudson da Silva Torrent.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Faculdade de Ciências Econômicas, Programa de Pós-Graduação em Economia, Porto Alegre, BR-RS, 2018.

1. Expectil. 2. Teoria do valor extremo. 3. Backtesting. 4. GJR-GARCH. 5. Mercado de ações. I. Torrent, Hudson da Silva, orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

VICTOR FREITAS FEIO DE LEMOS

THE EDGE OF EXPECTIL: MODELLING EXPECTIL WITH EVT AND COMPARE IT USING COMPARATIVE BACKTEST

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título de Mestre em Economia, com ênfase em Economia Aplicada.

Trabalho aprovado em: Porto Alegre, ____________ de ____________ de ____________

BANCA EXAMINADORA:

Hudson da Silva Torrent - Orientador
PPGE/UFRGS

Guilherme Valle Moura
PPGECO/UFSC

Marcelo Brutti Righi
PPGA/UFRGS

ACKNOWLEDGEMENTS

To the Graduate Program in Economics (PPGE) of the Federal University of Rio Grande do Sul - UFRGS.

To the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES*, for the funding support during my studies.

A very special gratitude goes to my forever encouraging family, who have provided me through moral and emotional support in my life. I am also grateful to all my classmates, what a such opportunity to work with all of you.

I am also grateful to the faculty of the Program for their continuous guidance, and for all university staff for their support.

A sincere gratitude for Professor Guilherme Valle Moura and Professor Marcelo Brutti Righi, for serving as my committee members.

Finally, I would like to express my special appreciation to my advisor Professor Hudson da Silva Torrent, for his valuable guidance and encouragement during my Master's degree.

RESUMO

Expectiles são uma família de parâmetros de medidas de risco coerentes que foram recentemente sugeridas como uma alternativa aos *quantiles* (VaR) e ao *expected shortfall* (ES). Neste trabalho, os *expectiles* são usados como uma medida de risco (EVaR), discutimos seu significado financeiro e os comparamos com VaR e ES. Destinamos a desenvolver uma mudança de *expectile* ao longo do tempo que mede o risco inerente de um ativo durante todo o período de gerenciamento. Por fim, realizamos uma análise empírica para o índice brasileiro (Ibovespa) por meio do modelo GJR-GARCH com inovações estimadas student-t utilizando o algoritmo de *Maximum Likelihood* (MLE) e *Markov Chain Monte Carlo* (MCMC)/*Bayesian* por Metropolis-Hastings e avaliamos a precisão das previsões por meio de uma função de ganho consistente. Resultados numéricos indicam que as expectativas são alternativas perfeitamente razoáveis para as medidas de risco VaR e ES.

Palavras-chaves: Expectil. Teoria do valor extremo. Backtesting. GJR-GARCH. Mercado de ações.

ABSTRACT

Expectiles are a one-parameter family of coherent risk measures that have been recently suggested as an alternative to quantiles (VaR) and to expected shortfall (ES). In this work, the expectiles are used as a measure of risk (EVaR), we discuss their financial meaning and compare them with VaR and ES. It is intended to develop a changing expectile over time that measures the inherent risk of an asset throughout the management period. Lastly, an empirical analysis is carried out for Brazilian index (Ibovespa) by means of GJR-GARCH model with student-t innovations estimated using Maximum Likelihood (MLE) and Markov Chain Monte Carlo (MCMC)/Bayesian by Metropolis-Hastings algorithm and we assess the accuracy of the forecasts by means of a consistent scoring function. Numerical results indicate that expectiles are perfectly reasonable alternatives to VaR and ES risk measures.

Keywords: Expectil. Extreme value theory (EVT). Backtesting. GJR-Garch. Stock market.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 INTRODUCTION

Risk management is a very wide and complex field that is at the core of every financial activity. The importance of this area of expertise is emphasized by the regulatory framework which is currently at the center of many discussions and negotiations. As a matter of fact, in December 2017 the Basel Committee on Banking Supervision (BCBS) published a package of proposed reforms for the global regulatory framework of our industry which is frequently referred to as 'Basel IV'. The Committee's aim is to make the capital framework more robust and to improve confidence in the system.

The complexity of risk management relies on the combination of very advanced quantitative methods that should always be combined to a deep qualitative understanding of the market and the notions that are hidden behind every computation. The last years have been characterized by significant instabilities in financial markets worldwide. This has led to numerous criticisms about the existing risk management systems and motivated the search for more appropriate methodologies able to cope with rare events that have heavy consequences. The typical question one would like to answer is: "If things go wrong, how wrong can they go? "The problem is then how to model the rare phenomena that lie outside the range of available observations and how can I compare and verify competing estimation procedures. In such a situation it seems essential to rely on Extreme value theory (EVT) once provides a firm theoretical foundation on which we can build statistical models describing extreme events and since Expectil Value at Risk (EVaR) has the property of elictability, we will use this new coherent risk measure that has been recently suggested in the literature as an alternative to Value at Risk (VaR) and Expected Shortfall (ES) and compare to them.

EVT is a statistical technique for estimating extreme events with low frequency but high severity. This technique is widely used in financial risk management since empirical evidence from various studies (see Sheikh e Qiao (2010); Berkowitz et al. (2011)) shows that in the majority of cases, financial asset return distributions are heavy-tailed, especially in times of financial instability.

Recent studies show that estimates of GARCH-type models can be strongly affected by structural breaks in the volatility dynamics (see, e.g., Lamoureux e Lastrapes (1990); Bauwens, Dufays e Rombouts (2014)). Estimating a GARCH–type model on data displaying structural breaks yields a non-stationary model and implies poor risk predictions. A way to cope with this problem is provided by Markov-switching GARCH (MSGARCH) models, once the parameters of which vary over time according to a latent discrete Markov process. These models can quickly adapt to variations in the unconditional volatility level, which improves risk predictions (ARDIA et al., 2008).

Research on changing volatility using time series models has been active since the pioneer paper by Engle (1982). From that time, ARCH (AutoRegressive Conditional Heteroscedasticity) and GARCH (Generalized ARCH) type models grew rapidly into a rich family of empirical models for volatility forecasting during the 80's. These models are widespread and essential tools in financial econometrics. In most empirical applications it turns out that the simple specification p = q = 1 is able to reproduce the volatility dynamics of financial data. This has led the GARCH(1,1) model to become the workhorse model by both academics and practitioners. Given a model specification for conditional variance (ht), the log-returns are then modelled as $y_t = \varepsilon_t h_t^{1/2}$, where ε_t are i.i.d disturbances. Common choices for ε_t are Normal and Student-t disturbances. The Student-t specification is particularly useful, since it can provide the excess kurtosis in the conditional distribution that is often found in financial time series processes (unlike models with Normal innovations). Until recently, GARCH models have mainly been estimated using the classical Maximum Likelihood technique. The Bayesian approach offers an attractive alternative which enables small sample results, robust estimation, model discrimination, model combination, and probabilistic statements on functions of the model parameters (ARDIA; HOOGERHEIDE, 2010).

Fissler, Ziegel e Gneiting (2015), Nolde, Ziegel et al. (2017) have recently proposed to replace traditional backtests by comparative backtests based on strictly consistent scoring functions since they allow for conservative tests and are sensitive with respect to increasing information sets. Roughly, this means that a risk measurement procedure that correctly incorporates more risk factors will always be preferred over a simpler procedure that uses less information. However, comparative backtests necessitate an elicitable risk measure. Examples of elicitable risk measures are VaR and expectiles, while expected shortfall (ES) is not elicitable. However, ES turns out to be jointly elicitable with VaR, which allows for comparative backtests also for ES.

In our discussion, we are focusing on the following three risk measures: VaR, a popular risk measure that is elicitable; expectiles, the only coherent and elicitable risk measures; and ES, a coherent and comonotonically additive risk measure, which is jointly elicitable together with VaR, and which is the new standard measure in banking regulation. We consider the following three approaches to handle the forecasting procedure: fully parametric (FP), filtered historical simulation (FHS), and a semi-parametric estimation based on extreme value theory (EVT). VaR at level $\alpha \in (0,1)$, denoted

$$VaR_\alpha(X) = inf[x|F_X(x) \geq \alpha], \tag{1}$$

where F_X is the cumulative distribution function of X. From the statistical perspective, VaR_α is simply the α-quantile of the underlying distribution. Positive values of X are interpreted as losses in this manuscript, hence we are interested in VaR_α for values of α close to one. Committee et al. (2013), p.103–108, specifically requests VaR_α values for $\alpha = 0.99$, which we refer to as the standard Basel VaR level. ES of an integrable random variable X at level $\nu \in (0,1)$ is given by

$$ES_\nu(X) = \frac{1}{1-\nu}\int_\nu^1 VaR_\alpha(X)\partial\alpha. \tag{2}$$

The Bank for International Settlements [2014] proposes $\nu = 0.975$ as the standard Basel ES level, as $ES_{0.975}$ should yield a similar magnitude of risk as $VaR_{0.99}$ under the standard normal distribution. As introduced by Newey and Powell [1987], the τ-expectile $e_\tau(X)$ of X with finite mean is the unique solution $x = e_\tau(X)$ to the equation

$$\tau\int_x^\infty (y-x)\partial F_X(y) = (1-\tau)\int_{-\infty}^x (x-y)\partial F_X(y). \tag{3}$$

As shown in Bellini et al. (2014), Ziegel (2016), τ-expectiles are elicitable coherent risk measures for $\tau \in [1/2, 1)$. Expectiles generalize the expectation just as quantiles generalize the median. Considering the level $\tau = 0.99855$ leads to a comparable magnitude of risk as $VaR_{0.99}$ and $ES_{0.975}$ under the standard normal distribution; see Bellini e Bernardino (2017).

It is necessary to deepen the study on methodological applications for expectile forecasting, since it is configured as a measure of risk recently studied in this type of literature, due to its properties, being the only measure of risk coherent and elicitable. According to what was exposed above, it is necessary to analyze the canonical forecasting methodologies such as historical simulation, fully parametric and compare them with other that have a high performance to predict extreme events like Extreme Value Theory (EVT). The present study attempts to fill this gap by forecasting expectile with EVT and compares it with models and canonical risk measures applied to the Brazilian stock market, something never observed in the Brazilian studies of risk measurement.

The questions to be answered with the realization of this work are: Is the Expectile a good risk measure? Which models and / or distributions are more suitable for prediction? And what is the Expectile contribution to the benchmark, VaR and ES measures? Does the Nolde, Ziegel et al. (2017) suggestion to re-examining suitability of the GARCH-type actually have the ability to improve the results in backtesting? EVaR is a good measure of risk for the Ibovespa?

The paper is organized as follows. In section 2, we introduce the concept of Expectile and how it can be applied. In section 3, we introduce EVT and the POT method. Section 4 elucidates the theoretical specification of the Markov-switching GJR-GARCH model with t-student innovations and the transition regime process. Section 5 contains the theoretical discussion of backtesting risk measures following the methodology of Nolde, Ziegel et al. (2017). In section 6 we will explain the models to forecast the mentioned risk measures. Section 7 contains an application to the returns on the Ibovespa Composite index and conclusion in section 8 with a summary and a discussion of the findings.

2 EXPECTILE

It is well known that the left and right quantiles x_α^- and x_α^+ of a random variable X can be defined through the minimization of an asymmetric, piecewise linear loss function

$$[x_\alpha^-(X), x_\alpha^+(X)] = \underset{x\in R}{Argmin}\ \alpha E[(X-x)_+] + (1-\alpha)E[(X-x)_-]\ for\ \alpha \in (0,1) \tag{4}$$

where $x^+ = max(x,0)$ and $x^- = max(-x,0)$; see, for example Koenker e Hallock (2001). Expectiles $e_q(X)$ have been introduced by Newey e Powell (1987) as the minimizers of an asymmetric quadratic loss

$$e_q(X) = \underset{x\in R}{Argmin}\ qE[(X-x)_+^2] + (1-q)E[(X-x)_-^2]\ for\ q \in (0,1) \tag{5}$$

When $q = \frac{1}{2}$, it is well known that $e_q(X) = E[X]$, thus expectiles can be seen as an asymmetric generalization of the mean. The term ‘expectiles’ has probably been suggested as a combination of ‘expectation’ and ‘quantiles’. Expectiles are uniquely identified by the first-order condition (f.o.c.)

$$qE[(X - e_q(X))_+] = (1-q)E[(X - e_q(X))_-] \tag{6}$$

Since Equation (6) is well defined for each $X \in L^1$, which is the natural domain of definition of the expectiles, we take it as the definition of $e_q(X)$. Letting

$$\ell_q(x) := qx_+ - (1-q)x_- \tag{7}$$

we see that Equation (6) can be rewritten as:

$$E[\ell_q(X - e_q(X))] = 0 \tag{8}$$

Hence, expectiles are an example of shortfall risk measures in the sense of Föllmer e Schied (2002), also known as zero utility premia in the actuarial literature. From this point of view, they had been considered in Weber (2006), although the connection with the minimization problem (4) and with the statistical notion of expectiles emerged only in the more recent literature. In general, a statistical functional that can be defined as the minimizer of a suitable expected loss function as in Equation (5) is said to be elicitable; we refer to Gneiting (2011), Bellini e Bignozzi (2015), Ziegel (2016), Embrechts, Klüppelberg e Mikosch (2013), Davis (2016) and Acerbi e Szekely (2014) for further information about the elicitability property and its financial relevance. See also the discussion in Section 4 on the relationship between elicitability and backtesting.

$EVaR_q$ is the financial risk measure associated with expectiles, in the same way as VaR_α is the financial risk measure associated with the quantiles. For q $\leq \frac{1}{2}$, $EVaR_q$ is a coherent risk measure, since it satisfies the well-known axioms introduced by Artzner et al. (1999). Indeed, it is easy to see that

- $EVaR_q(X+h) = EVaR_q(X) - h$, *for* $h \in R$ *(translation invariance),*
- $X \leq Y$ *as* $\Rightarrow EVaR_q(X) \geq EVaR_q(Y)$ *(monotonicity),*
- $EVaR_q(\lambda X) = \lambda EVaR_q(X)$, *for* $\lambda \geq 0$ *(positive homogeneity) and*
- $EVaR_q(X+Y) \leq EVaR_q(X) + EVaR_q(Y)$ *(subadditivity)*

Moreover, it has been shown in several papers that $EVaR_q$ with q $\leq \frac{1}{2}$ is the only coherent risk measure that is also elicitable (see Weber (2006);) Ben-Tal e Teboulle (2007); Bellini e Bignozzi (2015); Bellini et al. (2014); Delbaen et al. (2016); Ziegel (2016)). In order to better understand the financial meaning of $EVaR_q$, it is interesting to compare its acceptance set with VaR_α and with ES_α. Recall that the acceptance set of a translation invariant risk measure ρ is defined as

$$\mathcal{A}_\rho = \{X \mid \rho(X) \leq 0\} \tag{9}$$

and that ρ can be recovered by $\mathcal{A}_\rho$ by the formula

$$\rho(X) = inf\{m \in R \mid X + m \in \mathcal{A}_\rho\} \tag{10}$$

In the case of VaR_α

$$\mathcal{A}_{VaR_\alpha} = \{X \mid \mathcal{P}(X < 0) \leq \alpha\} \tag{11}$$

Notice that we can equivalently write

$$\mathcal{A}_{VaR_\alpha} = \left\{X \middle| \frac{\mathcal{P}(X > 0)}{\mathcal{P}(X \leq 0)} \geq \frac{1-\alpha}{\alpha}\right\} \tag{12}$$

In the case of ES_α, we have

$$\mathcal{A}_{ES_\alpha} = \left\{X \middle| \frac{1}{\alpha}\int_0^\alpha x_u(X)\partial u \geq 0\right\} \tag{13}$$

In the case of $EVaR_q$, the acceptance set can be written as

$$\mathcal{A}_{EVaR_q} = \left\{ X \middle| \frac{E[X_+]}{E[X_-]} \geq \frac{1-q}{q} \right\} \tag{14}$$

The $EVaR_q$ is then the amount of money that should be added to a position in order to have a prespecified, sufficiently high gain-loss ratio. We recall that the gain-loss ratio or Ω-ratio is a popular performance measure in portfolio management. It is sometimes argued that $EVaR_q$ is "difficult to explain"to the financial community, but this is probably due to the fact (2) is usually taken as starting point instead of Equation (14), which has a transparent financial meaning: in the case of VaR_α, a position is acceptable if the ratio of the probability of a gain with respect to the probability of a loss is sufficiently high (12); in the case of $EVaR_q$, a position is acceptable if the ratio between the expected value of the gain and the expected value of the loss is sufficiently high (14). In Section 6, we provide an application of Ibovespa daily returns for the computation of expectiles by means of a fully parametric model, an historical method (non parametric) and extreme-value-theory (semi-parametric model). Choosing q = 0.00145, the magnitude of $VaR_{0.01}$, $ES_{0.025}$ and $EVaR_{0.00145}$ is closely comparable. In conclusion, we believe that $EVaR_q$ is a perfectly reasonable risk measure, displaying many similarities with VaR_α and ES_α, surely worth of deeper study and practical experimentations by risk managers, regulators and portfolio managers.

3 EXTREME VALUE THEORY (EVT)

There are two methods for modeling extreme events with low frequency but high severity; the block maxima method and the Peaks Over Threshold (POT) method. For financial time series data, the POT method is often used to model extreme events. The block maxima method is not commonly used to do statistical inference on financial time series data because: the method does not make sufficient use of data as it uses only data from sub-period maxima, the choice of sub-period length is not clearly defined, the method is unconditional and does not take into account the effects of other explanatory variables (TSAY, 2014). In this work we use the POT method based on the generalized Pareto distribution (GPD). The POT method focuses on modeling the exceedances of the loss above a certain threshold η and the time of occurrence. The threshold is selected such that there are enough data points to do meaningful statistical analysis.

Let $\{x_i\}_{i=1}^{T}$ represent the loss variables of an asset return, then as $T \rightarrow \infty$, $\{x_i\}_{i=1}^{T}$ is assumed to be independent and identically distributed, and $(x-\mu)/\sigma$ follows a generalized extreme value (GEV) distribution:

$$F_{\xi,\mu,\sigma(x)} = \begin{cases} exp[-(1+\xi x)^{-1/\xi}], & \text{for } \xi \neq 0, \\ \\ exp[-e^{-x}] & \text{for } \xi = 0, \end{cases}$$

where ξ is the shape parameter and $1/\xi$ is the tail index of the GEV distribution. $x < -1/\xi$ if $\xi < 0$ and $x > -1/\xi$ if $\xi > 0$. Also, let the conditional distribution of the excesses over the threshold, i.e. $xi - \eta = y/xi > \eta$, be given by

$$Pr(x-\eta \leq y/x > \eta) = \frac{Pr(\eta \leq x \leq y+\eta)}{Pr(x > \eta)} = \frac{Pr(x \leq y+\eta) - Pr(x \leq \eta)}{1 - Pr(x \leq \eta)} \tag{15a}$$

$$= \frac{F(y+\eta) - F(\eta)}{1 - F(\eta)} = F_{\eta}(y). \tag{15b}$$

Again, as $T \to \infty$, $(y + \eta - \mu)/\sigma$ follows a GEV distribution. Therefore,

$$Pr(x - \eta \leq y/x > \eta) = \frac{F(y+\eta) - F(\eta)}{1 - F(\eta)}$$

$$= \frac{exp\left[-\left(1 + \frac{\xi(y+\eta-\mu)}{\sigma}\right)^{-1/\xi}\right] - exp\left[-\left(1 + \frac{\xi(\eta-\mu)}{\sigma}\right)^{-1/\xi}\right]}{1 - exp\left[-\left(1 + \frac{\xi(\eta-\mu)}{\sigma}\right)^{-1/\xi}\right]} \tag{16}$$

$$\approx 1 - \left(1 + \frac{\xi y}{\sigma + \xi(\eta - \mu)}\right)^{-1/\xi}$$

where $y > 0$ and $\sigma + \xi(\eta - \mu) > 0$. If we let $\Psi(\eta) = \sigma + \xi(\eta - \mu)$, and as $\eta \to \infty$, Eqn.(16) is approximated by the generalized Pareto distribution (GPD)

$$G_{\xi,\Psi(\eta)}(y) = \begin{cases} 1 - \left[1 + \frac{\xi y}{\Psi(\eta)}\right]^{-1/\xi}, & \text{for } \xi \neq 0, \\ \\ 1 - exp[-y/\Psi(\eta)] & \text{for } \xi = 0, \end{cases}$$

with shape parameter ξ and scale parameter $\Psi(\eta)$. $\Psi(\eta) > 0, y \in [0, x - \eta]$ when $\xi \geq 0$, and $y \in [0, -\Psi(\eta)\xi]$ when $\xi < 0$. If $\xi = 0$, then the equations above becomes an exponential distribution with parameter $1/\sigma$ (TSAY, 2014). $y = x - \eta$, then Eqn.(15b) can be written as

$$\frac{F(y+\eta) - F(\eta)}{1 - F(\eta)} = \frac{F(x) - F(\eta)}{1 - F(\eta)} \approx G_{\xi,\Psi(\eta)}(x - \eta) \tag{17}$$

$$\Rightarrow F(x) = F(\eta) + [1 - F(\eta)]G_{\xi,\Psi(\eta)}(x - \eta)$$

We can now state the tail estimator for the underlying distribution $F(x/\xi, \Psi(\eta)$ using the empirical estimate of $F(\eta)$. i.e., $\hat{F}(\eta) = (T - N_\eta)/T$ as

$$F(x/\xi, \Psi(\eta)) \approx \frac{T - N_\eta}{T}\left[1 + \frac{\xi(x - \eta)}{\Psi(\eta)}\right]^{-1/\xi} \tag{18}$$

where N_η is the number of observations above the threshold. After deciding on the choice of η, and assuming that the number of points above η are independent and identically distributed, then the parameters $\Psi(\eta)$ and ξ can be estimated by means of maximum likelihood estimation with likelihood function

$$\ell(x_i, ..., x_{N_\eta}/\xi, \sigma, \mu) = \prod_{i=1}^{N_\eta} \{(x_i) \ for \ x_i > \eta. \tag{19}$$

The choice of the threshold η is an important step in the POT method because Eqn.(18) is dependent on η and the number of points (i.e. exceedances) above η since the parameters are estimated based on the exceedances. Thus, it is very important to find the proper threshold value. Until this day, there is no clear cut satisfactory method in determining a proper threshold. Danielsson e Vries (1998) developed a semi-parametric estimator for the tails of the distribution and estimate the threshold through a bootstrap of the mean square error (MSE) of the tail index, and by minimizing MSE through the choice of the threshold. Danielsson et al. (2001) further used a two-step subsample bootstrap method to determine the threshold that minimizes the asymptotic MSE. Hill (1975) and Davison e Smith (1990) propose graphical tools to help identify the proper threshold known as the Hill plot and the mean excess plot respectively. As such methods require judgement at every time step at which conditional forecasts of risk measures are to be made, they are prohibitive for our purposes. Hence, in this paper we adopt a pragmatic approach as in McNeil e Frey (2000), and take k = 60 in samples of size n = 500.

4 MARKOV-SWITCHING GARCH MODELS

We define $y_t \epsilon IR$ as the log-return at time t = 1,...,T. We consider the case when y_t has zero mean and is not serially correlated, that is, the following moment conditions are assumed: $E[y_t] = 0$ and $E[y_t, y_{t-h}] = 0$ for all $t > 0$ and $h > 0$. The general Markov–switching GARCH specification can then be expressed as:

$$y_t|(s_t = k, I_{t-1}) \sim D(0, h_{k,t}, \xi_k), \tag{20}$$

where $D(0, h_{k,t}, \xi_k)$ is a continuous distribution with zero mean, time-varying variance $h_{k,t}$, and additional shape parameters gathered in the vector ξ_k. The additional integer-valued stochastic variable s_t, defined on the discrete space (1,...,K), characterizes the Markov-switching GARCH model. We assume that s_t evolves according to an unobserved first order ergodic homogeneous Markov chain with KxK transition probability matrix P:

$$P = \begin{pmatrix} P_{1,1} & \dots & P_{1,k} \\ \vdots & \ddots & \vdots \\ P_{k,1} & \dots & P_{k,k} \end{pmatrix}$$

where $p_{i,j} \equiv P[s_t = j|s_{t-1} = i]$ is the probability of a transition from state $s_{t-1} = i$ to state $s_t = j$. Obviously, the following constraints hold: $0 < p_{i,j} < 1 \forall\ i, j \epsilon 1, ..., K$, and $\sum_{j=1}^{K} p_{i,j} = 1, \forall i\ \epsilon\ 1, ..., K$. In (1) we denote by I_{t-1} the information set observed up to time $t-1$, that is, $I_{t-1} \equiv [y_{t-i}, i > 0]$. Given the parametrization of $D()$, we have $E[y_t^2|s_t = k, I_{t-1},] = h_{k,t}$, that is, $h_{k,t}$ is the variance of y_t conditional on the realization of $s_t = k$.

*4.1 **Bayesian GJR-GARCH model with Student's-t innovations***

The GJR model of Glosten, Jagannathan e Runkle (1993) with Student-t innovations for the log-returns yt is also able to capture the asymmetry in the conditional volatility process. This model is given by:

$$r_t = \varepsilon_t \left(\frac{v-2}{v} w_t h_t \right)^{1/2} \tag{21a}$$

$$h_{k,t} = \alpha_{0,k} + (\alpha_{1,k} + \alpha_{2,k}\ 1_{[y_{t-1}<0]}) y_{t-1}^2 + \beta_k h_{k,t-1}, \tag{21b}$$

$$\varepsilon_t \sim N(0,1); \quad w_t \sim IG(\frac{v}{2}, \frac{v}{2}); t = 1, ..., T \tag{21c}$$

for $k = 1, ..., K$, where I is the indicator function taking value one if the condition holds, and zero otherwise. In this case, we have $\psi_k = (\alpha_{0,k}, \alpha_{1,k}, \alpha_{2,k}, \beta_k)^T$. The parameter $\alpha_{2,k}$ controls the degree of asymmetry in the conditional volatility response to the past shock in regime k. To ensure positivity, we require that $\alpha_{0,k} > 0, \alpha_{1,k} > 0, \alpha_{2,k} \geq 0, \beta k \geq 0$.

IG and N(0,1) symbolise the inverted gamma and standard normal distributions, respectively. The degrees of freedom parameter $v > 2$ guarantees finite conditional variance (GEWEKE, 1993). Eqns.(21a) and (21b) are the mean and variance equations respectively; where r_t are the log-return series and $\alpha_0, \alpha_i, \beta_i$ are the GARCH parameters. $\varepsilon_{i,t}$ is a series of independent and identically distributed random variables with zero mean and unit variance, and can assume a standard normal, Student's-t, generalised error (GED), or skewed distributions (TSAY, 2014). The conditional variance is stationary and almost surely positive if and only if $\alpha_{0,k} > 0, \alpha_{1,k} \geq 0, \alpha_{2,k} \geq 0, \beta_k \geq 0, and(\alpha_{1,k} + \alpha_{2,k} E[\eta_{k,t}^2\ 1_{[\eta_{k,t}<0]}] + \beta_k) < 1$

$$\Sigma = \Sigma(\psi, w) = diag(w_t \frac{v-2}{v} h_t(\alpha, \beta)_{t=1}^{T}) \tag{22}$$

where $h_t = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta h_{t-1}(\alpha, \beta)$, we can express the likelihood of (ψ, v) as

$$L(\psi, w|y) \propto (det\Sigma)^{-1/2} exp \left[\frac{-1}{2} y' \Sigma^{-1} y \right] \tag{23}$$

The Bayesian approach considers (ϕ, w) as a random variable which is characterized by a prior density denoted by $p(\phi, w)$. The prior is specified with the help of parameters called hyperparameters which are initially assumed to be known and constant. Moreover, depending on the researcher's prior information, this density can be more or less informative. Then, by coupling the likelihood function of the model parameters with the prior density, we can transform the probability density using Bayes' rule to get the posterior density $p(\phi, w|y)$ as follows

$$p(\psi, v|y) = \frac{L(\psi, w|y) p(\psi, w)}{\int L(\psi, w|y) p(\psi, w) \partial\psi \partial w} \tag{24}$$

This posterior is a quantitative, probabilistic description of the knowledge about the model parameters after observing the data. We use truncated normal priors on the GJR-GARCH parameters α and β

$$p(\alpha) \propto \phi N_2(\alpha|\mu_\alpha, \sigma_\alpha) \quad 1\left[\alpha \: \epsilon \: IR_+^2\right] \tag{25a}$$

$$p(\beta) \propto \phi N_1(\beta|\mu_\beta, \sigma_\beta) \quad 1\left[\alpha \: \epsilon \: IR_+\right] \tag{25b}$$

where μ and σ are the hyperparameters,1$\Delta$is the indicator function and ϕN_d is the d-dimensional normal density. The prior distribution of vector w conditional on v is found by noting that the components v_t are independent and identically distributed from the inverted gamma density, which yields

$$p(w|v) = \left(\frac{v}{2}\right)^{\frac{Tv}{2}} \left[\Gamma\left(\frac{v}{2}\right)\right]^{-T} \left(\prod_{t=1}^{T} w_t\right)^{-\frac{v}{2}-1} exp\left[-\frac{1}{2}\Sigma_{t=1}^{T}\frac{v}{w_t}\right] \tag{26}$$

We follow Deschamps (2006) and Geweke (1993) in the choice of the prior distribution on the degrees of freedom parameter. The distribution is a translated exponential with parameters $\lambda > 0$ and $\delta \geq 2$

$$p(v) = \lambda exp[-\lambda(v-\delta)] \; 1[v > \delta] \tag{27}$$

For large values of λ, the mass of the prior is concentrated in the neighborhood of δ and a constraint on the degrees of freedom can be imposed in this manner. Normality of the errors is assumed when δ is chosen large. As pointed out by Deschamps (2006), this prior density is useful for two reasons. First, it is potentially important, for numerical reasons, to bound the degrees of freedom parameter away from two to avoid explosion of the conditional variance. Second, we can approximate the normality of the errors while maintaining a reasonably tight prior which can improve the convergence of the sampler.

4.2 ***Metropolis-Hastings***

The joint prior distribution is then formed by assuming prior independence between the parameters, i.e. $p(\psi, v) = p(\alpha)p(\beta)p(w|v)p(v)$. The recursive nature of the GARCH(1,1) variance equation implies that the joint posterior and the full conditional densities cannot be expressed in closed form. There exists no (conjugate) prior that can remedy this property. Therefore, we cannot use the simple Gibbs sampler and need to rely on a more elaborated Markov Chain Monte Carlo (MCMC) simulation strategy to approximate the posterior density. The idea of MCMC sampling was first introduced by Metropolis et al. (1953) and was subsequently generalized by Hastings (1970). The sampling strategy relies on the construction of a Markov chain with realizations $(\psi^{[0]}, v^{[j]}), ..., (\psi^{[j]}, v^{[j]})$,... in the parameter space. Under appropriate regularity conditions, asymptotic results guarantee that as j tends to infinity,$\psi^{[j]}, v^{[j]}$ tends in distribution to a random variable whose density is 3. Hence, after discarding a burn-in of the first draws, the realized values of the chain can be used to make inference about the joint posterior.

Following Chib e Greenberg (1995), suppose we have a density that can generate candidates. Since we are dealing with Markov chains, however, we permit that density to depend on the current state of the process. Accordingiy, the candidate-generating density is denoted q(x,y), where $\int q(x, J.)\partial y = 1$. This density is to be interpreted as saying that when a process is at the point x, the density generates a value y from q(x,3,). If it happens that q(x,y) itself satisfies the reversibility condition (22) for all x, y, our search is over. But most likely it will not. We might find, for example, that for some x, y

$$\pi(x)q(x, y) > \pi(y)q(y, x) \tag{28}$$

In this case, speaking somewhat loosely, the process moves from x to y too often and from J, to x too rarely. A convenient way to correct this condition is to reduce the number of moves from x to y by introducing a probability $\alpha(x, y) < 1$ that the move is made. We refer to $\alpha(x, y)$ as the probability of move. If the move is not made, the process again returns x as a value from the target distribution. Thus transitions from x to y $(y \neq x)$ are made according to

$$p_{MH}(x, y) = q(x, y)\alpha(x, y), \quad x \neq y, \tag{29}$$

where $\alpha(x, y)$ is yet to be determined. Consider again inequality (27). It tells us that the movement from y to x is not made often enough. We should therefore define a(y,x) to be as large as possible, and since it is a probability, its upper limit is 1. But now the probability of move a(x, y) is determined by requiring that $p_{MH}(x, y)$ satisfies the reversibility condition, because then

$$\begin{aligned} \pi(x)q(x, y)\alpha(x, y) &= \pi(y)q(y, x)\alpha(y, x) \\ &= \pi(y)q(y, x) \end{aligned} \tag{30}$$

We now see that $\alpha(x,y) = \pi(y)q(y,x)/\pi(x)q(x,y)$. Of course, if the inequality in (27) is reversed, we set $\alpha(x,y) = 1$ and derive $\alpha(y,x)$ as above. The probabilities $\alpha(x,y)$ and $\alpha(y,x)$ are thus introduced to ensure that the two sides of (27) are in balance or, in other words, that $p_{MH}(x,y)$ satisfies reversibility. Thus we have shown that in order to be reversible, the probability of move must be set to

$$\begin{aligned} \alpha(x,y) = min\left[\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right] \quad &, if \quad \pi(x)q(x,y) > 0 \\ &= 1, \quad otherwise. \end{aligned} \tag{31}$$

5 BACKTESTING RISK MEASURES

5.1 ***Score Functions***

A risk measure ρ is usually defined on some space of random variables. If ρ is law-invariant, it can alternatively be viewed as a map from some collection of probability distributions P to the real line R. Law-invariance means that for two random variables X and Y that have the same distribution, we have $\rho(X) = \rho(Y)$. All risk measures considered in this manuscript are law-invariant. Let $\Theta(X) = (\rho_1(X), ..., \rho_k(X))$ be a vector of $k \geq 1$ risk measures.

Definition 1. A scoring function S : $R_k x R \rightarrow R$ is called strictly consistent for Θ with respect to P if

$$E(S(\Theta(X), X)) < E(S(r, X)) \tag{32}$$

for all $r = (r_1, ..., r_k) \neq \Theta(X) = (\rho_1(X), ..., \rho_k(X))$ and all X with distribution in P. The scoring function S is consistent if equality is allowed in (Eq 32). The vector of risk measures Θ is called elicitable with respect to P if there exists a strictly consistent scoring function for it.

Elicitability is useful for model selection, estimation, generalized regression, forecast ranking, and, as we will detail in this paper, allows for comparative backtesting. A comprehensive literature review on elicitability can be found in Gneiting (2011), Frongillo e Kash (2015) and Fissler, Ziegel e Gneiting (2015). The question of elicitability of risk measures has recently received considerable attention. All available results in the case k = 1 are based on the simple but powerful observation that a necessary requirement of elicitability are convex level sets in a distributional sense. Weber (2006) was the first to study risk measures with convex level sets. Bellini e Bignozzi (2015) used his results to study elicitability for the broad class of monetary risk measures. Under weak regularity assumptions, they show that elicitable monetary risk measures are so-called shortfall risk measures. For more specific classes of risk measures, such as coherent, convex or distortion risk measures, the same result can be shown without any additional regularity assumptions. While expected shortfall is itself not elicitable, Fissler, Ziegel e Gneiting (2015) have shown that the pair $\Theta = (VaR_\alpha, ES_\alpha)$ is elicitable. The classes of (strictly) consistent scoring functions for VaR_α, τ-expectiles and (VaR_v, ES_v) have been characterized. The following three propositions state sufficient conditions for (strict) consistency.

Proposition 1 (Thomson (1979), Saerens (2000)). All scoring functions of the form

$$S(r,x) = (1 - \alpha - 1\{x > r\})G(r) + 1\{x > r\}G(x) \tag{33}$$

where G is an increasing function on R, are consistent for VaR_α, $\alpha \in (0,1)$, with respect to P_0. The scoring functions of the above form are stricly consistent for VaR_α with respect to $P' \subseteq P_0$ if G is stricly increasing, $G(X)$ is integrable for all X with distribution in P', and all distributions in P' have a unique α-quantile.

Proposition 2.

$$S(r,x) = 1\{x > r\}(1-2\tau)(\phi(r)-\phi(x)-\phi'(r)(r-x))-(1-\tau)(\phi(r)-\phi'(r)(r-x)), \tag{34}$$

where ϕ is a convex function with subgradient ϕ', are consistent for the τ-expectile, $\tau \in (0,1)$, with respect to P_1. If ϕ is strictly convex, then the scoring functions of the above form are strictly consistent for the τ-expectile relative to the class $P' \subseteq P_1$ such that ϕ(X) is integrable for all X with distribution in P'.

Proposition 3 (Fissler, Ziegel e Gneiting (2015)). All scoring functions of the form

$$\begin{aligned} S(r_1,r_2,x) = 1\{x > r_1\}(-G_1(r_1) + G_1(x) - G_2(r_2)(r_1 - x)) \\ +(1-v)(G_1(r_1) - G_2(r_2)(r_2 - r_1) + \mathcal{G}_2(r_2)) \end{aligned} \tag{35}$$

In risk management applications, it may be useful to allow only for strictly positive risk measure predictions, once this opens up the possibility for attractive choices of homogeneous scoring functions in the above propositions. If r $\in (0,\infty)$ is assumed in (Eq 33) or (Eq 34), then, for strict consistency, we only need that G or ϕ are defined on $(0,\infty)$, and that they are strictly increasing or strictly convex on this domain, respectively.

Closely connected to elicitability is the concept of identifiability.

Definition 2. The vector of risk measures Θ is called identifiable with respect to P, if there is a function $V : R_k x R \to R_k$ such that

$$E(V(r,X)) = 0 \quad \Leftrightarrow \quad r = \Theta(X), \tag{36}$$

for all X with distribution in P.

5.2 ***Calibration and traditional backtests***

We fix the following notation. Suppose that $\Theta(X) = (\rho_1(X), ..., \rho_k(X))$ is an identifiable functional with identification function V with respect to $\mathcal{P}$. Let $\{X_t\}_{t\in N}$ be a series of negated log-returns adapted to the filtration $\mathcal{F} = \{\mathcal{F}_t\}_{t\in N}$ and $\{R_t\}_{t\in N}$ a sequence of predictions of Θ, which are $\mathcal{F}_{t-1}$-measurable. Hence, the predictions are based on the information about $\{X_t\}_{t\in N}$ available at time t-1 represented by the sigma-algebra $\mathcal{F}_{t-1}$.

Definition 3. The sequence of predictions $\{R_t\}_{t\in N}$ is calibrated for Θ on average if

$$E(V(R_t, X_t)) = 0 \ for\ all\ t\ \in\ N; \tag{37}$$

it is super-calibrated for Θ on average if $E(V(R_t, X_t)) \geq 0$ component-wise, for all $t \in N$. The sequence of predictions $\{R_t\}_{t\in N}$ is conditionally calibrated for Θ if

$$E(V(R_t, X_t) \mid \mathcal{F}_{t-1}) = 0,\ almost\ surely,\ for\ all\ t\ \in\ N; \tag{38}$$

it is conditionally super-calibrated for Θ if $E(V(R_t, X_t)|F_{t-1}) \geq 0$ component-wise, almost surely, for all $t \in N$. Sub-calibration is defined analogously.

If one knows the conditional distributions $L(X_t|F_{t-1})$ and strives for the best possible prediction of Θ based on the information in F_{t-1}, it is natural to use

$$\Theta(\mathcal{L}(X_t \mid \mathcal{F}_{t-1})) \tag{39}$$

as a predictor, which we term the optimal F-conditional forecast for Θ. For the same reason, we call $\Theta(X_t) = \Theta(L(X_t))$ the optimal unconditional forecast.

Calibration characterizes optimal forecasts in the following sense. The optimal unconditional forecast is the only deterministic forecast that is calibrated for Θ on average.

Following Fissler, Ziegel e Gneiting (2015), we call any backtest that considers a null hypothesis of the type “The risk measurement procedure is correct” a traditional backtest. Traditional backtests are similar to goodness-of-fit tests, that is, they allow to demonstrate that the risk measurement procedure under consideration is making incorrect predictions, if the respective null hypothesis can be rejected. Despite the somewhat misleading terminology that a traditional backtest is passed if the null hypothesis is not rejected, this does not mean that in this case, one can be sure that the null hypothesis is correct (with a pre-specified small probability of error) as this would necessitate that we control the power of the test explicitly.

Testing the null hypothesis

H0 : The sequence of predictions $\{R_t\}_{t \in N}$ is calibrated for Θ on average.

Conditional calibration is a stronger notion than average calibration, and it appears more natural in a dynamic risk management context. A traditional backtest for conditional calibration considers the null hypothesis

H0 : The sequence of predictions $\{R_t\}_{t \in N}$ is conditionally calibrated for Θ.

The requirement $E(V(R_t, X_t)|F_{t-1}) = 0$, almost surely, is equivalent to stating that $E(h_t' V(R_t, X_t)) = 0$ for all $\Im_{t-1}$-measurable R^k-valued functions h_t. Following Giacomini e White (2006), we consider an $\Im$ -predictable sequence $\{h_t\}_{t \in N}$ of qxk-matrices h_t called test functions to construct a Wald-type test statistic:

$$t_1 = n \left(\frac{1}{n} \sum_{t=1}^{n} h_t V(R_t, X_t) \right)' \hat{\Omega_n}^{-1} \left(\frac{1}{n} \sum_{t=1}^{n} h_t V(R_t, X_t) \right) \tag{40}$$

In applications, the choice of the test functions is motivated by the principle that they should represent the most important information available at time point t-1. We call this type of traditional backtests as conditional calibration tests. In cases where $h_t = 1$, we refer to these tests as simple conditional calibration tests.

Commonly used backtests for VaR_α and ES_v are closely related to conditional calibration tests for specific choices of the test functions h_t. In fact, choosing $h_t = 1$ in the case of VaR_α, the conditional calibration test for VaR_α is closely related to the standard backtest for VaR_α based on the number of VaR exceedances. In the case of ES_v, the conditional calibration test for (VaR_v,ES_v) is related to the backtest for ES_v of McNeil e Frey (2000) based on exceedance residuals.

5.3 ***Comparative backtests***

Suppose now that the functional $\Theta = (\rho_1, ..., \rho_k)$ is elicitable with respect to $\mathcal{P}$. Let $\{X_t\}_{t \in N}$ be a series of negated log-returns adapted to the filtration $\mathcal{F} = \{\mathcal{F}_t\}_{t \in N}$ as well as to the filtration $\mathcal{F}^* = \{\mathcal{F}_t^*\}_{t \in N}$. Let $\{\mathcal{R}_t\}_{t \in N}$ and $\{\mathcal{R}_t^*\}_{t \in N}$ be two sequences of predictions of Θ, which are $\mathcal{F}$ and $\mathcal{F}^*$-predictable, respectively. We assume that all conditional distributions $\mathcal{L}(X_t \mid \mathcal{F}_{t-1})$, $\mathcal{L}(X_t \mid \mathcal{F}_{t-1}^*)$ and all unconditional distributions $\mathcal{L}(X_t)$ belong to $\mathcal{P}$ almost surely. We refer to the predictions $\{\mathcal{R}_t^*\}_{t \in N}$ as the standard procedure, while $\{\mathcal{R}_t\}_{t \in N}$ is the internal model. The two filtrations $\mathcal{F}$ and $\mathcal{F}^*$ acknowledge the fact that the internal model and the standard model may be based on different information sets. For example, one model may include more risk factors than the other, or, certain expert opinion may be used to adjust one model but not the other.

Definition 4. Let S be a consistent scoring function for Θ with respect to $\mathcal{P}$. Then, $\{\mathcal{R}_t\}_{t\in N}$ S-dominates $\{\mathcal{R}^*_t\}_{t\in N}$ (on average) if

$$E(S(R_t, X_t) - S(R^*_t, X_t)) \leq 0, \ for\ all\ t\ \in\ N. \tag{41}$$

Furthermore, $\{\mathcal{R}_t\}_{t\in N}$ conditionally S-dominates $\{\mathcal{R}^*_t\}_{t\in N}$ if

$$E(S(R_t, X_t) - S(R^*_t, X_t) \mid \mathcal{F}^*_{t-1}) \leq 0, \ almost\ surely,\ for\ all\ t \in\ N \tag{42}$$

The definition of conditional dominance is asymmetric in terms of the role of the standard procedure and the internal procedure. The standard procedure and the information $\mathcal{F}^*$ it is based on are considered as a benchmark of predictive ability, which is why we condition on $\mathcal{F}^*_{t-1}$ and not on $\mathcal{F}_{t-1}$. Any method that dominates the benchmark has superior predictive ability relative to this benchmark.

There are several reasons why the predictions $\{\mathcal{R}_t\}_{t\in N}$ should be preferred over $\{\mathcal{R}^*_t\}_{t\in N}$ if the former dominates the latter. Firstly, comparison of forecasts with respect to the described dominance relations is consistent with respect to increasing information sets. That is, if $\mathcal{F}^*_t \subseteq \mathcal{F}_t$ for all t and $\{\mathcal{R}_t\}_{t\in N}$, $\{\mathcal{R}^*_t\}_{t\in N}$ are the optimal conditional forecasts with respect to their filtrations as defined at (Eq 27), then the internal procedure dominates the standard procedure, both, conditionally and on average.

The condition for conditional S-dominance in (2.15) can be formulated equivalently as

$$E((S(R_t, X_t) - S(R^*_t, X_t))h_t) \leq 0, \ for\ all\ h_t \geq 0,\ \mathcal{F}^*_{t-1}\ -\ measurable, \tag{43}$$

It is tempting to work with a vector h_t of $\mathcal{F}^*$-predictable test functions in order to test for conditional S-dominance as suggested in the conditional calibration tests. However, we are interested in comparing the standard procedure to the internal procedure and reach a definite answer as to which one is to be preferred. If $E((S(R_t, X_t) - S(R^*_t, X_t))h_{t,i}) > 0$ but $E((S(R_t, X_t) - S(R^*_t, X_t))h_{t,j}) < 0$ for different components $h_{t,i}, h_{t,j}$ of the vector h_t, no clear preference for either method can be given. Therefore, we do not pursue this approach further.

In comparative backtesting we are interested in the null hypotheses

H^-_0 : The internal model predicts at least as well as the standard model.

H_0^+ : The internal model predicts at most as well as the standard model.

The null hypothesis H_0^- is analogous to the null hypothesis of a correct model and estimation procedure but now adapted to a comparative setting. As mentioned in the introduction, considering a backtest as passed if the null hypothesis cannot be rejected is anti-conservative or aggressive in nature and may therefore be problematic in regulatory practice. On the other hand, the null hypothesis H_0^+ is such that the comparative backtest is passed if we can reject H_0^+ . This means that we can explicitly control the type I error of allowing an inferior internal model over an established standard model.

We assume now that the limit

$$\lambda := \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} E((S(R_t, X_t) - S(R_t^*, X_t)) \;\in\; [-\infty, +\infty] \tag{44}$$

exists (while we allow it to take the values $\pm\infty$). It is clear that S-dominance on average implies $\lambda \leq 0$. If the sequence of score differences $\{S(R_t, X_t) - S(R_t^*, X_t)\}_{t\in N}$ is first-order stationary, then $\lambda \leq 0$ implies S-dominance on average. Under (Eq 44), we can compare any two sequences of risk measure estimates with respect to their predictive performance. If the limit λ in (Eq 44) is non-positive, then the internal procedure is at least as good as the standard procedure, whereas the internal procedure predicts at most as well as the standard procedure if $\lambda \geq 0$. Ordering risk measurement procedures is a compromise in the quest for conditional dominance. On the one hand, it is clearly a weaker notion than conditional dominance, but on the other hand it introduces a meaningful total order on all risk measurement procedures given a sensible choice of the scoring function S;

Therefore, we reformulate our comparative backtesting hypotheses as

$$H_0^- : \lambda \leq 0$$

$$H_0^+ : \lambda \geq 0$$

The test statistic

$$\Delta_n \overline{S} := \frac{1}{n} \sum_{t=1}^{n} (S(R_t, X_t) - S(R_t^*, X_t)), \tag{45}$$

for n large enough, has expected value less or equal to zero under H_0^- , whereas under H_0^+ its expectation is non-negative. Tests of H_0^+ or H_0^- based on a suitably rescaled version of $\Delta_n \overline{S}$ are so-called Diebold-Mariano tests; see Diebold e Mariano (2002).

5.4 Choice of the scoring function

Based on (Eq 33), (Eq 34) and (Eq 35), one has a large number of choices for strictly consistent scoring functions for VaR, expectiles and (VaR, ES). In the case of VaR_α, the standard choice is to take $G(r) = r$ in (Eq 33) leading to the classical asymmetric piecewise linear loss. In the case of expectiles, one could argue that a natural choice is taking $\phi(r) = r^2$ in (Eq 34), which simplifies to the squared error function for the mean (up to equivalence). This is also the scoring function suggested by Newey e Powell (1987) for expectile regression. Consistent scoring functions for (VaR, ES) have only recently been discovered; see Acerbi e Szekely (2014), Fissler, Ziegel e Gneiting (2015). Therefore, there is no natural classical choice for the functions G_1, G_2 in (Eq 35).

A scoring function S is called positive homogeneous of degree b (or b-homogeneous) if for all $r = (r_1, ..., r_k)$ and all x

$$S(cr, cx) = c^b S(r, x),\ \ for\ all\ c > 0.$$

The crucial property of a scoring function is to be positive homogeneous. Patton (2011) underlines the importance of positive homogeneity of the scoring function for forecast ranking. Positive homogeneous scoring functions are also favorable because they are so-called "unit consistent". For example, changing the units, from, say, R$ to million R$, will not change the ordering of forecasts assessed by this scoring function, and will thus also leave the results of comparative backtests unchanged. Concerning the choice of the degree b of homogeneity, Patton (2006) shows that in the case of volatility forecasts, $b = 0$ requires weaker moment conditions than a larger choice of b for the validity of Diebold-Mariano tests which are used in comparative backtesting.

For the comparative backtests for VaR that we investigate in Section 4.3, we consider the classical 1-homogeneous choice obtained by choosing G(r) = r in (Eq 33) leading to the scoring function

$$S(r, x) = (1 - \alpha - 1\{x > r\})r + 1\{x > r\})x \tag{46}$$

Guided by the arguments given above, we alternatively consider the 0-homogeneous score differences by choosing $G(r) = logr$, $r > 0$ which leads to the score

$$S(r, x) = (1 - \alpha - 1\{x > r\})logr + 1\{x > r\})logx \tag{47}$$

The choice $\phi(r) = r^2$ in (Eq 34) leads to the strictly consistent scoring function

$$S(r,x) = -1\{x > r\}(1-2\tau)(x-r)^2 + (1-\tau)r(r-2x) \tag{48}$$

for the τ-expectile e_τ. Besides this 2-homogeneous choice, we also investigate the 0-homogeneous alternative that arises by choosing $\phi(r) = -log(r)$, $r > 0$, hence we obtain the scoring function

$$S(r,x) = 1\{x > r\}(1-2\tau)\left(log\frac{x}{r} + 1 - \frac{x}{r}\right) + (1-\tau)\left(logr - 1 + \frac{x}{r}\right) \tag{49}$$

For (VaR_v, ES_v), we consider the (1/2)- homogeneous scoring function given by choosing $G_1(x) = 0$, $\mathcal{G}_2(x) = x^{1/2}$, $x > 0$ in (Eq 35) for comparative backtesting in Section 4.3. It is given by

$$S(r_1, r_2, x) = 1\{x > r_1\}\frac{x - r_1}{2\sqrt{r_2}} + (1-v)\frac{r_1 + r_2}{2\sqrt{r_2}} \tag{50}$$

As for the other risk measures, we also consider the 0-homogeneous alternative by choosing $G_1(x) = 0$, $G_2(x) = logx$,$x > 0$ which yields the scoring function

$$S(r_1, r_2, x) = 1\{x > r_1\}\frac{x - r_1}{r_2} + (1-v)\left(\frac{r_1}{r_2} - 1 + log(r_2)\right) \tag{51}$$

6 FORECASTING OF RISK MEASURES

In this section we discuss a number of estimation procedures for producing conditional forecasts of the three risk measures discussed in this paper, namely the VaR, expectile and ES. Owing to the widespread use of VaR in the banking sector, a great number of methods exist to produce its point forecasts. In contrast, estimation and forecasting of expectiles in the risk measurement context is a relatively recent topic. However, in many cases, similar methods as those used for VaR forecasting can be adopted for expectiles.

For illustrative purposes, we consider the following framework for forecasting of the risk measures. Suppose the series of negated log-returns $\{X_t\}_{t\in N}$ can be modeled as

$$X_t = \mu_t + \sigma_t Z_t \tag{52}$$

where $\{Z_t\}_{t\in N}$ is a sequence of i.i.d. random variables with zero mean and unit variance, and μ_t and σ_t are measurable with respect to the sigma algebra $\mathcal{F}_{t-1}$, representing the information about the process $\{X_t\}_{t\in N}$ available up to time t-1. In order to capture typical time dynamics of financial time series, one possibility is to assume that the conditional mean μ_t follows an ARMA process, while the condition variance σ_t^2 evolves according to a GARCH model specification.

Let ρ denote any of the three risk measures we consider. In the above setting, conditionally on the information up to time t-1, the one-step ahead forecast of ρ is

$$\rho(X_t|\mathcal{F}_{t-1}) = \mu_t + \sigma_t \rho(Z) \tag{53}$$

where Z is used to denote a generic random variable with the same distribution as the $Z_t's$. Following McNeil e Frey (2000) and Diebold, Schuermann e Stroughair (2000), one can adopt a two-stage estimation procedure for the forecast $\rho(X_t|\mathcal{F}_{t-1})$. First μ_t and σ_t are estimated via the maximum likelihood procedure under a specific assumption on the distribution of the innovations Z_t in (Eq 52). The second stage involves estimation of $\rho(Z)$, the risk measure for i.i.d. sequence $\{Z_t\}_{t\in N}$, based on the sample of standardized residuals

$$\{\hat{z}_t = (x_t - \hat{\mu}_t)/\hat{\sigma}_t\} \tag{54}$$

We consider the following three approaches to handle the second stage in the forecasting procedure: fully parametric (FP), filtered historical simulation (FHS), and a semi-parametric estimation based on extreme value theory (EVT).

6.1 ***Fully parametric estimation***

Under the fully parametric approach, a specific (parametric) model is assumed for the sequence of innovations $\{Z_t\}_{t\in N}$. Examples of typically used probability distributions include the normal, Student's t and a skewed t distribution. Parameters of the assumed distribution for Zt's, denoted F_Z, can be estimated based on the standardized residuals $\{\hat{z}_t\}$ in (Eq 54) using, for example, the maximum likelihood method. If the model for Zt's coincides with the one used to estimate the filter in the first stage, then no additional estimation is required at the second stage with all model parameters coming directly from the first stage estimation. The fitted distribution is used to compute the estimate of a given risk measure.

6.2 ***Filtered historical simulation***

The method employs a non-parametric estimation of the risk measures based on the standardized residuals $\{\hat{z}_t\}$, which can be seen as representing a filtered time series. In particular, we draw a sample $\{\hat{z}_i^*; \leq i \leq N\}$ of a large size N from $\{\hat{z}_i^*; \leq t \leq n\}$ and then take the empirical estimate of a given risk functional as the estimate for $\rho(Z)$. The empirical α-quantile gives the VaR estimate $\widehat{VaR}_\alpha^{FHS}(Z)$. The empirical τ-expectile $\hat{e}_\tau^{FHS}$ is obtained using the least asymmetric weighted squares via iterative minimization of

$$\sum_{i=1}^{N} w_i(\tau)(\hat{z}_i^* - e_\tau)^2, \qquad w_i(\tau) = \tau 1\{\hat{z}_i^* > e_\tau\} + (1-\tau)1\{\hat{z}_i^* < e_\tau\} \quad \text{with respect to } e_\tau$$

The ES is estimated by the empirical version of the conditional expectation given that the residual exceeds the corresponding VaR estimate

$$\widehat{ES}_v^{FHS}(Z) = \frac{1}{\#\{i : i = 1, ..., N, \hat{z}_i^* > \widehat{VaR}_\alpha^{FHS}(Z)\}} \sum_{i=1}^{N} \hat{z}_i^* 1\{\hat{z}_i^* > \widehat{VaR}_\alpha^{FHS}(Z)\}$$

6.3 ***EVT-based semi-parametric estimation***

Risk is naturally associated with extremal events, and hence risk measure estimates rely on accurate estimation of a tail of the underlying distribution. However, inference about the distributional tails is notoriously difficult as there are frequently not enough data points in the tail regions neither to give a proper justification for a parametric model nor to obtain reliable empirical estimates. Hence, unless a sufficiently long time series is available relative to the desired risk level for risk measure estimation, the two methods fully parametric and historical simulation are unlikely to produce accurate forecasts. An alternative is to base estimation on asymptotic results of extreme value theory (EVT). For a detailed account, refer to, e.g., Embrechts, Klüppelberg e Mikosch (1997). The main premise is that, for a sufficiently high threshold u, conditional excesses of random variable Z satisfy

$$Z - u \mid Z > u \sim GP(\beta_u, \xi) \tag{55}$$

where $GP(\beta, \xi)$ denotes the generalized Pareto distribution with scale $\beta > 0$ and shape parameter $\xi \in R$. It is common in applications to set the threshold at an upper order statistic; i.e., $u = z_{(k+1)}$ for some $k < n$, where $z_(1) > z_(2) > \Delta\Delta\Delta > z_(n)$ are the decreasing order statistics of the sample $z_1, ..., z_n$ from F_Z. This leads to the following EVT-based estimates of $VaR_\alpha(Z)$ and $ES_\nu(Z)$ (see McNeil e Frey (2000))

$$\widehat{VaR_\alpha}^{EVT}(Z) = u + \frac{\hat{\beta}_u}{\hat{\xi}} \left(\left(\frac{k}{\alpha n} \right)^{\hat{\xi}} - 1 \right), \ \hat{\xi} \neq 0, \tag{56}$$

and

$$\widehat{ES_v}^{EVT}(Z) = \widehat{VaR_v}^{EVT}(Z) \left(\frac{1}{1-\hat{\xi}} + \frac{\hat{\beta} - \hat{\xi} u}{(1-\hat{\xi})\widehat{VaR_v}^{EVT}(Z)} \right), \tag{57}$$

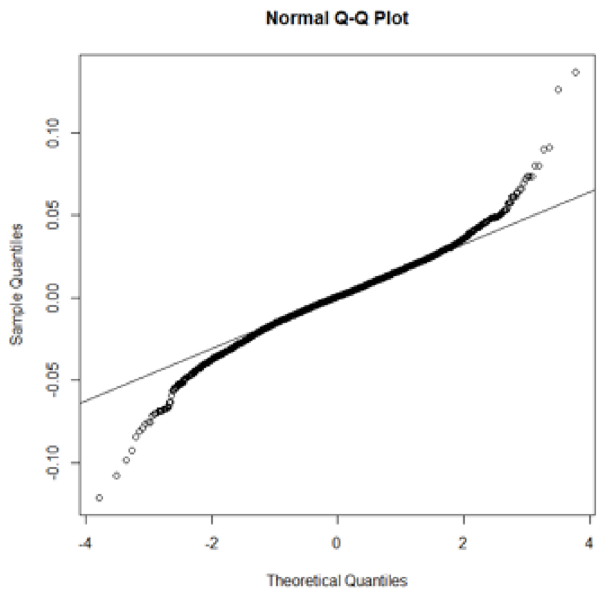
with $(\hat{\beta}_u, \hat{\xi})$ being parameter estimates of the GP distribution fitted to excesses over u. In the spirit of the above EVT-based estimators for VaR and ES, we derive an estimator for the τ-expectile. The details are provided in Appendix A.1,2.

In the discussion above we assume that threshold u or equivalently k, the number of upper order statistics, is given so as to ensure adequacy of the approximation in (3.4). However, in practice, an accurate choice has to be made to balance the bias-variance trade-off as a too large value of u increases variability of the parameter estimates of β_u and ξ, while insufficiently large u introduces the bias due to invalidity of (3.4). Various techniques have been proposed to assist with the choice of threshold such as graphical tools based on linearity of the mean excess function. As such methods require judgement at every time step at which conditional forecasts of risk measures are to be made, they are prohibitive for our purposes. Hence, we adopt a pragmatic approach as in McNeil e Frey (2000), and take k = 60 in samples of size n = 500.

7 APPLICATION IN BRAZILIAN STOCK MARKET

We have fitted an GJR-GARCH model to the negated log-returns of the IBOVESPA Composite index using a moving estimation window of 500 data points. The time series we consider is from Jan. 3, 2000 until March 1, 2018, which gives us an out-of-sample size n = 3.995 to perform backtesting.

The plot of the time series is presented in Figure 1. Well-known stylized facts observed in financial time series, such as volatility clustering and presence of outliers, are evident from Figure 1. Furthermore, we also note that large (absolute) returns are more frequent at the start (2000-2006) and at the end (2012-2018) of the sample, than in the middle (2006-2012). This suggests that the shape of the log-returns distribution may be time-varying and analyzing the qqplot we see that the distribution of returns has fat tails, supporting the use of student-t distribution. So we consider the asymmetric two-state MSGARCH model implemented by Ardia et al. (2008) and Mullen et al. (2011). This is an extension of the MSGARCH model introduced in Haas et al. (2004a), where a GJR variance specification with a Student-t distribution is defined in each regime.



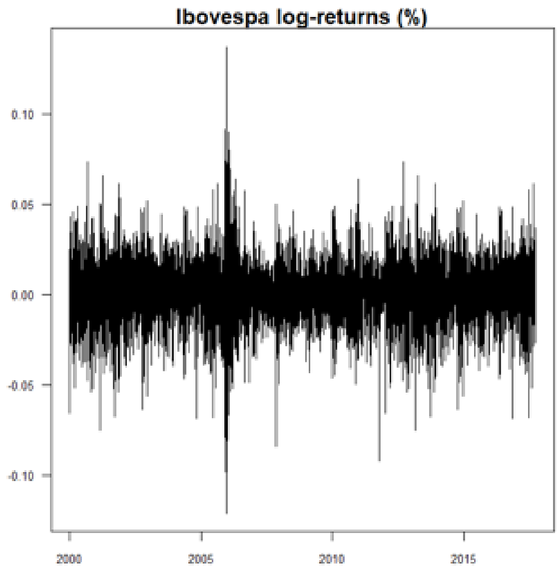


Figure 1 – Norm Q-Q and log-returns Ibovespa

Table 1 – Estimation via Maximum Likelihood (ML)

Fitted Parameters	Estimate	Std. Error	t value	Pr(> $\|t\|$)
α_{01}	0.0000	0.0000	302.9414	$< 1e^{-16}$
α_{11}	0.0000	0.0000	3.6872	$1.134e^{-04}$
α_{21}	0.3253	0.0011	305.2837	$< 1e^{-16}$
β_1	0.6394	0.0010	612.1276	$< 1e^{-16}$
ν_1	16.9629	0.0365	465.2007	$< 1e^{-16}$
α_{02}	0.0000	0.0000	329.1283	$< 1e^{-16}$
α_{12}	0.0001	0.0000	7.0516	$8.841e^{-13}$
α_{22}	0.0824	0.0001	629.8198	$< 1e^{-16}$
β_2	0.9396	0.0001	9016.1452	$< 1e^{-16}$
P_{11}	0.9912	0.0000	23597.0768	$< 1e^{-16}$
P_{21}	0.0035	0.0000	184.3012	$< 1e^{-16}$

Transition Matrix	$t+1\|k=1$	$t+1\|k=2$	**Stable Probabilities**	
$t\|k=1$	0.9912	0.0088	*State 1*	*State 2*
$t\|k=2$	0.0035	0.9965	0.2853	0.7147

Parameter estimates indicate that the evolution of the volatility process is heterogeneous across the two regimes. As well as a different reactions to past negative returns: $\alpha_{21} \approx 0.32$ vs. $\alpha_{22} \approx 0.08$. Also the volatility persistence in the two regimes is different. The first regime reports $\alpha_{1,1} + 1/2\alpha_{2,1} + \beta_1 \approx 0.80$ while the second regime reports $\alpha_{1,2} + 1/2\alpha_{2,2} + \beta_2 \approx 0.98$

In summary, the first regime is characterized by: i) low unconditional volatility, ii) strong volatility reaction to past negative returns, and iii) low persistence of the volatility process. Differently, the second regime is characterized by: i) high unconditional volatility, ii) weak volatility reaction to past negative returns, and iii) high persistence of the volatility process.

Clearly, regime one would be identified by market operators as "calm market conditions" with low volatility levels, low persistence and high reaction to past negative returns, while regime two as "turbulent market conditions" with high volatility level, strong persistence and lower reaction to past negative returns.

Figure 2 displays the smoothed probabilities of being in regime two (high unconditional volatility regime), $P[S_t = 2|I_T]\ fort = 1, ..., T$, superimposed on the Ibovespa returns (top graph) as well as the filtered volatility of the overall process (bottom graph).

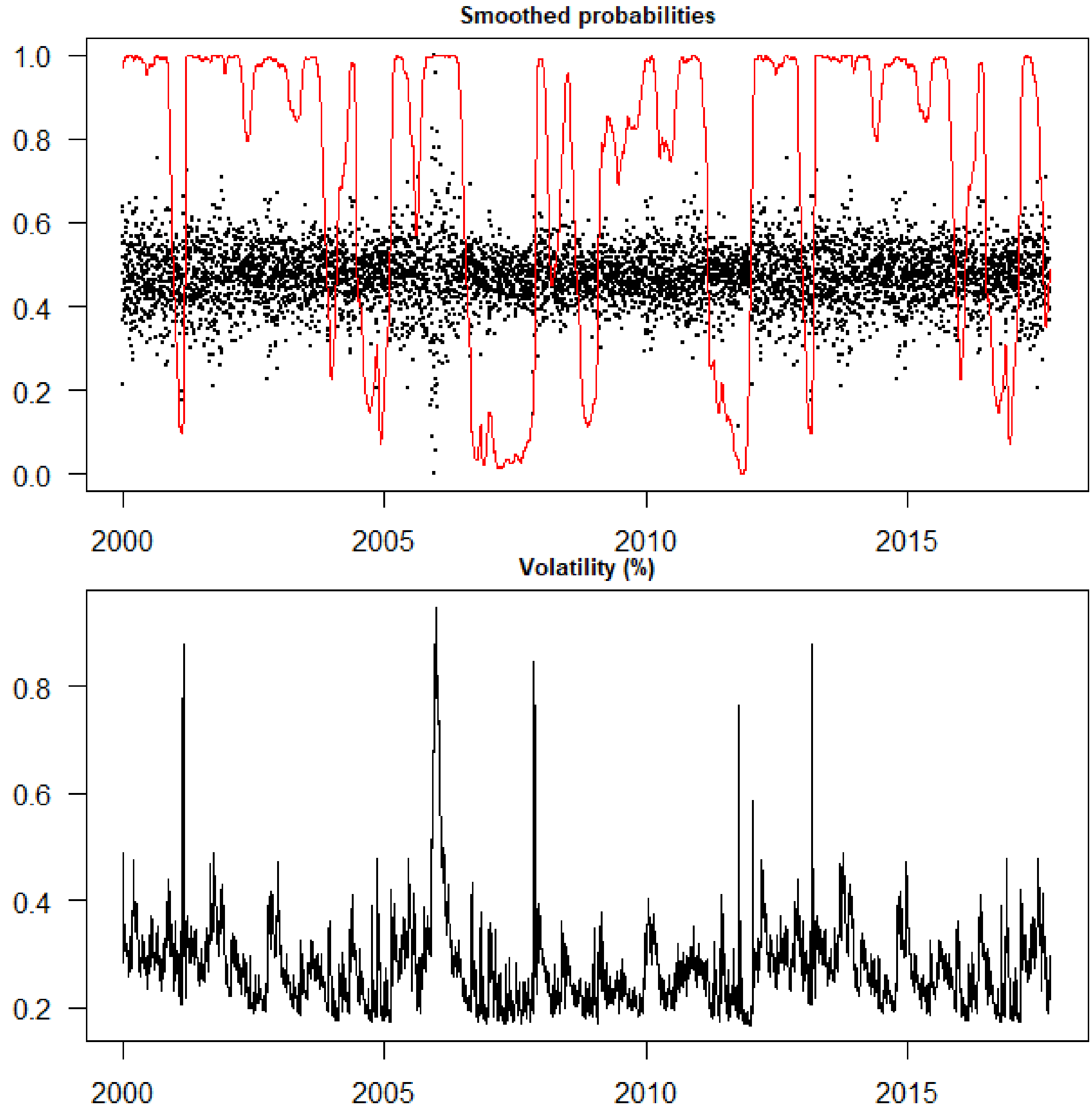


Figure 2 – Estimated smoothed probabilities of the second regime. The small black circles depict the Ibovespa log-returns. Bottom: Filtered conditional volatilities.

Interestingly, we further note that the Markov Chain evolves in a transitional way over time and that, in the limit, as reported by the probabilities of being in the two states are about 28% and 72%.

As documented by Ardia et al. (2008) and Mullen et al. (2011), ML estimation can be difficult for MSGARCH–type models. Fortunately, MCMC procedures can be used to explore the joint posterior distribution of the model parameters, thus avoiding convergence to local maxima commonly encountered via ML estimation. The Bayesian approach offers additional advantages. Specifically, the estimation method is a random–walk

Metropolis–Hasting algorithm with coerced acceptance rate. The controls parameters may be defined as follows:

- par: *A vector of starting parameters* ⇒ We use ML parameter estimates.
- n.burn: *Number of discarded draws* ⇒ We use n.burn = 5000.
- *Number of MCMC draws* ⇒ We use n.mcmc = 12500.
- n.thin: *Thinning factor* ⇒ We use n.thin = 5.

We observed excellent performance in the context of (identified) mixture models. Using the ML parameter estimates as starting values, we can estimate the model by MCMC using the R package MSGARCH, Ardia et al. (2016).

Table 2 – Estimation via MCMC

Posterior Sample	Mean	SD	SE	TSSE	RNE
α_{01}	0.0000	0.0000	0.0000	0.0000	0.0919
α_{11}	0.0005	0.0017	0.0000	0.0001	0.2543
α_{21}	0.3413	0.1098	0.0022	0.0090	0.0591
β_1	0.6145	0.1138	0.0023	0.0082	0.0768
ν_1	18.2083	3.5608	0.0712	0.1882	0.1432
α_{02}	0.0000	0.0000	0.0000	0.0000	0.0976
α_{12}	0.0002	0.0002	0.0000	0.0000	0.0228
α_{22}	0.0852	0.0232	0.0005	0.0010	0.2006
β_2	0.9365	0.0183	0.0004	0.0008	0.1986
P_{11}	0.9899	0.0049	0.0001	0.0003	0.1222
P_{21}	0.0034	0.0018	0.0000	0.0001	0.1164

Transition Matrix	$t+1\|k=1$	$t+1\|k=2$	**Stable Probabilities**	
$t\|k=1$	0.9899	0.0101	*State 1*	*State 2*
$t\|k=2$	0.0034	0.9966	0.2529	0.7471

Finally, Figure 3 displays the estimated filtered probabilities of the first state (high unconditional volatility state), implied by the best model parameters of solnp (in red solid line). In addition, we report in solid blue lines, the 50% area of the paths obtained over the $(n.mcmc - n.burn)/n.thin = (12500 - 5000)/5 = 1500$ runs. The parameters obtained with MCMC estimation lead to a not clear separation of regimes in the filtering probabilities with ups and downs over time. Although the transience, we may associate the beginning of the serie with low unconditional volatility state. Then, from the second half of 2005 to 2010, the returns are more associated with the high unconditional volatility regime. From 2010 to 2018, the model remains in the low unconditional volatility regime.

But, we clearly conclude, as already seen in the Stable Probabilities in table above, that the regimes are very transience.

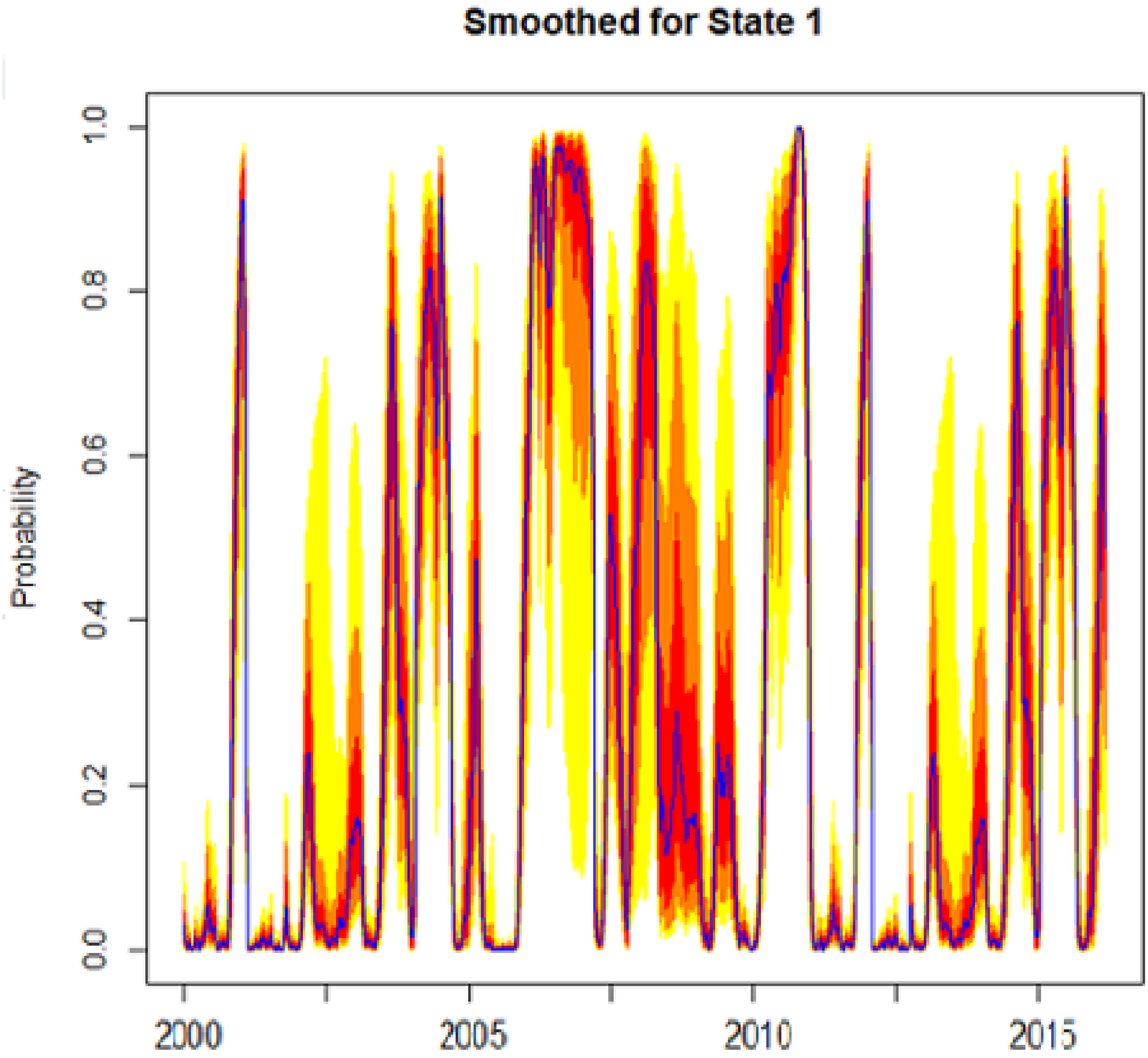


Figure 3 – Filtered probabilities of the first regime obtained by MCMC for the two–state Markov–switching GJR model with skewed Student–t innovations. Blue line indicates the median.

It is interesting to note that the RW adaptive estimation leads to results similar to the more complex MCMC estimation strategy. The posterior distribution of mixture and Markov–switching models often exhibits non-elliptical shapes which lead to non-reliable estimation of the uncertainty of model parameters. This invalidates the use of the Gaussian asymptotic distribution for inferential purposes infinite samples. Our results display this characteristic as shown in Figure 4 where we plot 2,500 draws of the posterior sample for the parameters $\alpha_{1,1}$ and $\alpha_{1,2}$. The blue square reports the posterior mean while the red triangle reports the ML estimate.

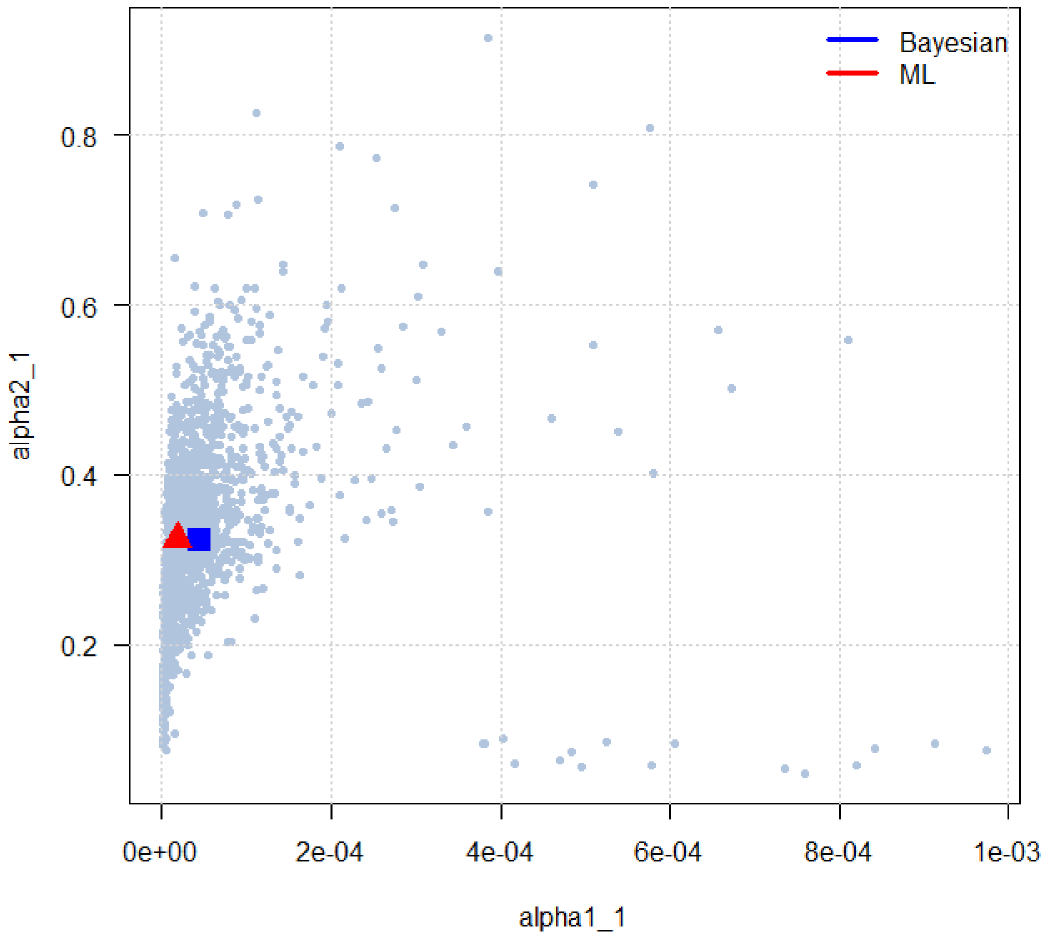


Figure 4 – Scatter plot of posterior draws from the marginal distribution of $(\alpha_{1,1}, \alpha_{1,2})^T$ obtained with the adaptive random walk strategy. The blue square reports the posterior mean, and the red triangle reports the ML estimate. The graph is based on 2,500 draws from the joint posterior sample.

An interesting aspect of the Bayesian estimation is that we can make distributional (probabilistic) statements on any (possibly nonlinear) function of the model parameters. This is achieved by simulation. For instance, for each draw in the posterior sample we can compute the unconditional volatility in each regime, to get its posterior distribution. Figure 5 displays the posterior distributions of the unconditional annualized volatility in each regime. In the low-volatility regime, the distribution is centered around 0.003*sqrt(250)= 4,7% per annum. For the high-volatility regime, the distribution is centered around

0.006*sqrt(250)= 9,4% per annum. Notice that both distributions exhibit positive skewness (0.19and0.18, respectively). Hence, relying on the asymptotic Normal approximation would yield erroneous overestimates of the 95% confidence band of the unconditional volatility in each regime.

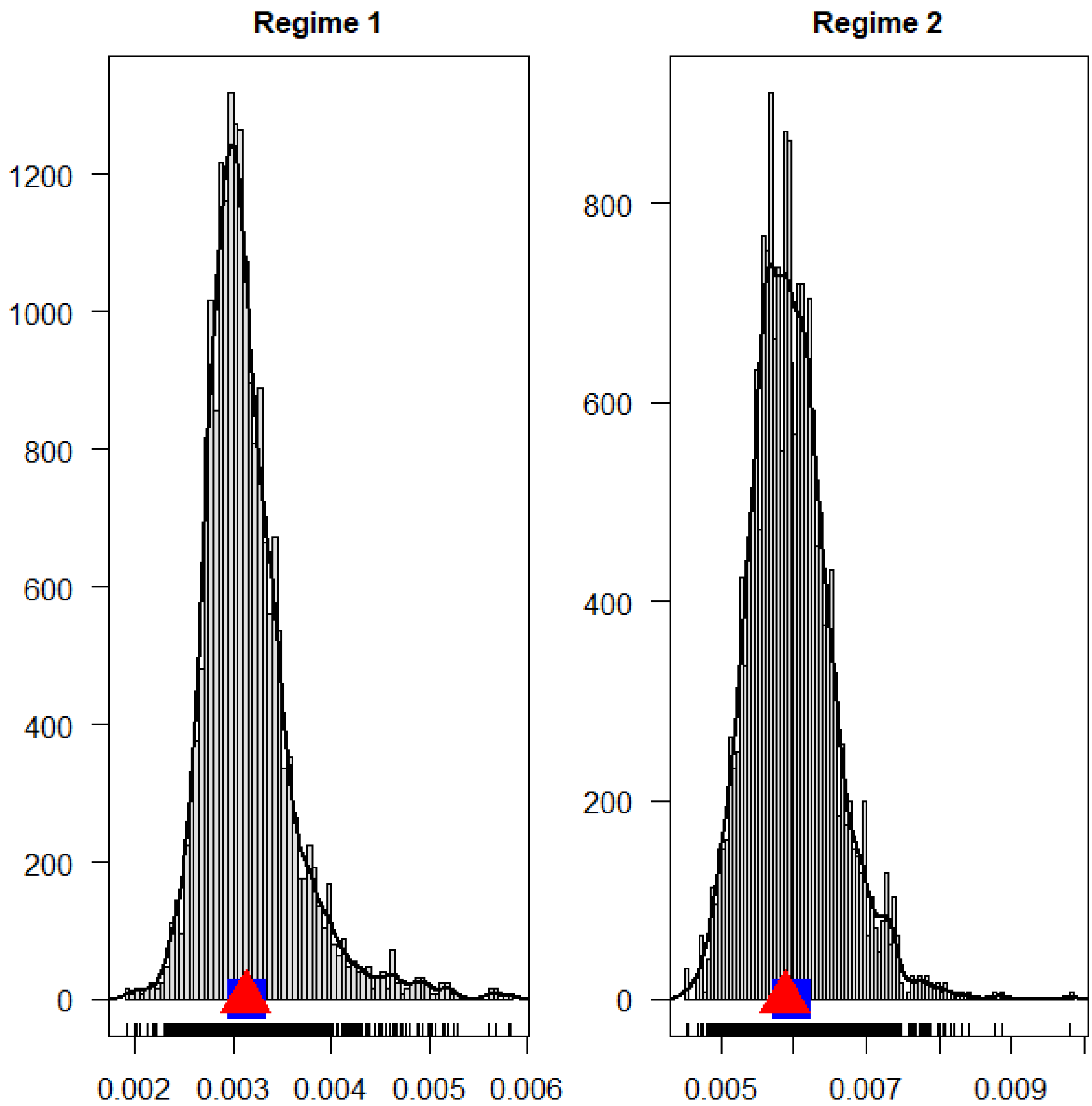


Figure 5 – Histograms of the posterior distribution for the unconditional volatility in each regime. Both graphs are based on 2,500 draws from the joint posterior sample. The blue square reports the posterior mean, and the red triangle reports the ML estimate.

7.1 ***Backtesting***

In order to link to the previously used approaches to assess the quality of VaR forecasts (and to make comparisons between the methods), we computed the percentage of times the observations exceeded the VaR_{α} forecasts, commonly referred to as the percentage of violations. Based on the values reported under the column "% Viol."in Table 1 , we observe that some of the misspecified models were actually able to hit nearly exactly the expected proportion of violations by matching the risk measure level $(1-\alpha)$. This is the case, for instance, for "n-EVT"and "t-EVT"methods at $\alpha = 0.99$. Although large deviations from the risk measure confidence level do suggest substantial method deficiencies (as in the case of "n-FP"and "t-FP"methods), these values also highlight that the deviations from the $(1-\alpha)$ level alone are unlikely to provide a good basis for differentiating the methods' performance in terms of prediction.

Table 3 – Summary of traditional and comparative backtesting based on the negated log-returns on the IBOVESPA Composite index with an GJR-GARCH(1,1) filter fitted over moving estimation window of 500 observations, and the out-of-sample size of n =3.995. The second column reports the average risk measure forecasts. "% Viol." gives the percentage of $VaR_{0.99}$ forecast exceedances. The simple CCT and general CCT columns contain the p-values for two-sided simple and general conditional calibration tests, respectively. The final two columns show the average scores, scaled by one minus the risk measure confidence level for presentation purposes, based on the specified scoring functions along with the corresponding method ranks (in brackets). In bold the methods that cct tests reject at a level of 5%.

Method	$\overline{VaR}_{0.99}$	% Viol.	simple CCT	General CCT	$\bar{S}$[eq.46]	$\bar{S}$[eq 47]
n-FP	4.077	1.302	**0.000**	**0.000**	5.268 (4)	1.622 (4)
n-FHS	4.390	1.101	**0.016**	**0.021**	5.283 (5)	1.626 (5)
n-EVT	4.311	1.101	0.539	0.212	5.214 (2)	1.609 (2)
st-FP	4.294	1.126	**0.044**	**0.012**	5.230 (3)	1.612 (3)
st-FHS	4.377	1.227	**0.039**	0.189	5.297 (6)	1.629 (6)
st-EVT	4.302	1.076	0.640	0.398	5.121 (1)	1.601 (1)

	$\overline{e}_{0.99855}$	simple CCT	General CCT	$\overline{S}$[eq.48]	$\overline{S}$[eq.49]
n-FP	4.078	**0.000**	**0.000**	31.604 (5)	0.676 (6)
n-FHS	4.447	**0.045**	**0.002**	30.732 (2)	0.658 (5)
n-EVT	4.427	**0.034**	**0.008**	31.084 (3)	0.653 (3)
st-FP	4.543	0.083	**0.031**	31.727 (6)	0.646 (1)
st-FHS	4.446	0.446	0.194	30.905 (2)	0.655 (4)
st-EVT	4.422	0.330	0.178	30.255 (1)	0.649 (2)

	$\overline{ES}_{0.975}$	simple CCT	General CCT	$\overline{S}$[eq.50]	$\overline{S}$[eq.51]
n-FP	4.077	**0.000**	**0.000**	5.268 (4)	1.622 (4)
n-FHS	4.390	**0.022**	**0.021**	5.283 (5)	1.626 (5)
n-EVT	4.311	0.539	**0.021**	5.214 (1)	1.609 (1)
st-FP	4.294	**0.004**	**0.012**	5.230 (3)	1.612 (3)
st-FHS	4.377	0.193	0.189	5.297 (6)	1.629 (6)
st-EVT	4.302	0.640	0.398	5.221 (2)	1.610 (2)

Table 3 summarizes results of traditional and comparative backtesting for six forecasting methods and, as before, for the three risk measures (VaR, expectile and the (VaR, ES) pair) at their standard Basel levels.

In the case of $VaR_{0.99}$, the traditional backtests based on the two-sided simple conditional calibration tests are passed only under the n-EVT and st-EVT methods. At this relatively high risk measure level, the EVT-based methods outperform their other

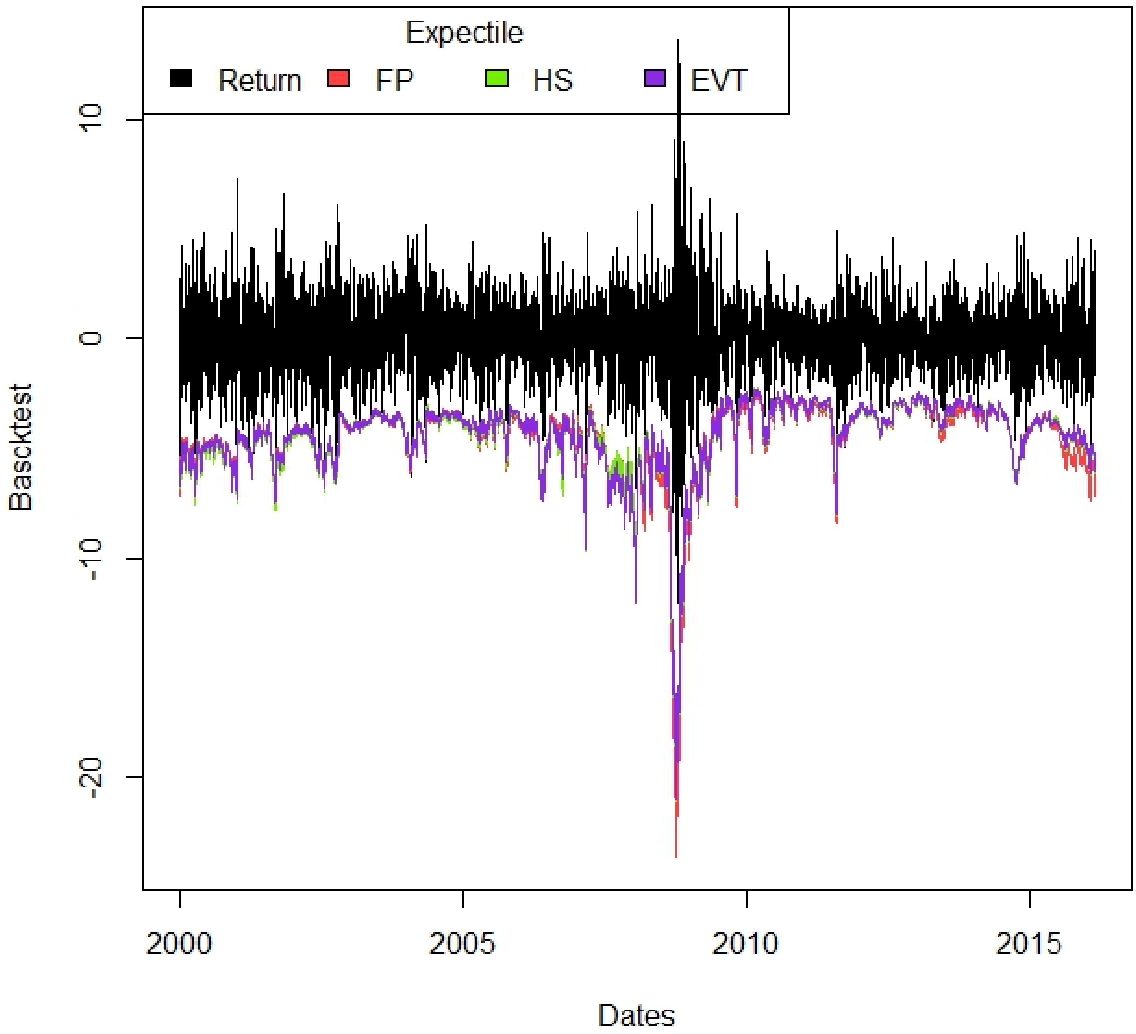


Figure 6 – Expectile$_{0.99855}$ for the Ibovespa log-return series, evaluated using the three considered methods based on rolling windows of length N = 500: Fully Parametric (red line), historical method (green line) and Extreme Value Theory (purple line)

competitors based on both the traditional backtests and the average scores. It should also be noted that the two scoring functions have lead to the same rankings of the forecasting procedures. The historical simulation methods (n-FHS and st-FHS) show the worst performance in terms of their predictive ability. On the other hand, for the 0.99855-expectile, the tests of simple conditional calibration are rejected (at 5% level) for all the methods that use the normal likelihood.

For $VaR_0.99$, we performed the conditional calibration tests also with the test function $h_t = (1, V(r_{t-1}, x_{t-1}))'$. They lead to conclusions similar to those based on the simple conditional calibration tests. This example underlines the importance of further studies on appropriate choices of test functions. The results for (VaR_V, ES_v) with $v = 0.975$ suggest better performance when a more flexible model such as the st is used to fit the GJR-Garch filter, although the use of EVT-based methods has a potential to compensate for likelihood mis-specifications. Again, fully parametric methods (n-FP and st-FP) faill in the comparative backtests against most of the other more flexible alternatives. The outcomes show one interesting aspect which is not in contradiction with the theory but may be puzzling and merit further investigation in future studies: The conditional calibration test rejects all methods using a normal likelihood but the scoring functions rank the n-EVT method as the best or second best performing method. It seems that the test function used in the conditional calibration test is sensitive to the likelihood function used in fitting the GJR-GARCH filter whereas the scoring functions are more sensitive to the method at the second stage giving preference to the EVT methods.

8 CONCLUSION

In this research paper, Expectil is analyzed as a measure of the inherent risk of a financial asset, with the aim of answering the following questions: is the Expectil a good measure of financial risk? Which method is best for your prediction?

Expectil is defined as a matrix for capital requirements, that is, as the amount of capital that must be added to a position or portfolio to obtain a sufficiently high profit and loss ratio. In order to respond as widely as possible to the rest of the questions posed, an analysis was carried out with Ibovespa returns, using a fully parametric, non parametric and semi-parametric methods with the GJR-GARCH. In response to the first question, in this paper we bet on the Expectil since it is a risk measure with the capacity to absorb the possible losses of capital in the left tail of the distribution of yields. to allow the real risk to be adjusted as much as possible by considering the information available for the entire distribution. With respect to the second question, Extreme Value Theory tend to be preferred in backtesting.

It is expected that the expectiles will better capture the risk compared to the VaR and the ES by: i) being sensitive to the magnitudes of the values that exceed the VaR and also taking into account the magnitudes of the most extreme values of the distribution, ii) take into account the information of both tails, so if the profile of the right queue is modified, the expectiles will be affected, unlike ES that does not, iii) because it is a coherent and elicitable measure.

Finally, this work could be extended with its extension to the portfolios using multivariable distributions, variable correlations over time, non-linear dependency, in-depth analysis of the Expectil validation and further development.

BIBLIOGRAPHY

ACERBI, C.; SZEKELY, B. Back-testing expected shortfall. *Risk*, Incisive Media Limited, v. 27, n. 11, p. 76–81, 2014. 13, 30

ARDIA, D. et al. Markov-switching garch models in r: The msgarch package. 2016. 38

ARDIA, D.; HOOGERHEIDE, L. F. Efficient bayesian estimation and combination of garch-type models. 2010. 10

ARDIA, D. et al. Financial risk management with bayesian estimation of garch models. *Lecture notes in economics and mathematical systems*, Springer, v. 612, 2008. 10, 35, 37

ARTZNER, P. et al. Coherent measures of risk. *Mathematical finance*, Wiley Online Library, v. 9, n. 3, p. 203–228, 1999. 14

BAUWENS, L.; DUFAYS, A.; ROMBOUTS, J. V. Marginal likelihood for markov-switching and change-point garch models. *Journal of Econometrics*, Elsevier, v. 178, p. 508–522, 2014. 9

BELLINI, F.; BERNARDINO, E. D. Risk management with expectiles. *The European Journal of Finance*, Taylor & Francis, v. 23, n. 6, p. 487–506, 2017. 11

BELLINI, F.; BIGNOZZI, V. On elicitable risk measures. *Quantitative Finance*, Taylor & Francis, v. 15, n. 5, p. 725–733, 2015. 13, 14, 24

BELLINI, F. et al. Generalized quantiles as risk measures. *Insurance: Mathematics and Economics*, Elsevier, v. 54, p. 41–48, 2014. 11, 14

BEN-TAL, A.; TEBOULLE, M. An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*, Wiley Online Library, v. 17, n. 3, p. 449–476, 2007. 14

CHIB, S.; GREENBERG, E. Understanding the metropolis-hastings algorithm. *The american statistician*, Taylor & Francis Group, v. 49, n. 4, p. 327–335, 1995. 22

COMMITTEE, B. et al. Fundamental review of the trading book: A revised market risk framework. *Consultative Document, October*, 2013. 11

DANIELSSON, J. et al. Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate analysis*, Elsevier, v. 76, n. 2, p. 226–248, 2001. 18

DANIELSSON, J.; VRIES, C. G. D. *Beyond the sample: Extreme quantile and probability estimation.* [S.l.], 1998. 18

DAVIS, M. H. Verification of internal risk measure estimates. *Statistics & Risk Modeling*, De Gruyter Oldenbourg, v. 33, n. 3-4, p. 67–93, 2016. 13

DAVISON, A. C.; SMITH, R. L. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 393–442, 1990. 18

DELBAEN, F. et al. Risk measures with the cxls property. *Finance and stochastics*, Springer, v. 20, n. 2, p. 433–453, 2016. 14

DESCHAMPS, P. J. A flexible prior distribution for markov switching autoregressions with student-t errors. *Journal of Econometrics*, Elsevier, v. 133, n. 1, p. 153–190, 2006. 21

DIEBOLD, F. X.; MARIANO, R. S. Comparing predictive accuracy. *Journal of Business & economic statistics*, Taylor & Francis, v. 20, n. 1, p. 134–144, 2002. 29

DIEBOLD, F. X.; SCHUERMANN, T.; STROUGHAIR, J. D. Pitfalls and opportunities in the use of extreme value theory in risk management. *The Journal of Risk Finance*, MCB UP Ltd, v. 1, n. 2, p. 30–35, 2000. 32

EMBRECHTS, P.; KLÜPPELBERG, C.; MIKOSCH, T. *Modelling extremal events, volume 33 of Applications of Mathematics*. [S.l.]: New York). Springer-Verlag, Berlin, 1997. 34

EMBRECHTS, P.; KLÜPPELBERG, C.; MIKOSCH, T. *Modelling extremal events: for insurance and finance*. [S.l.]: Springer Science & Business Media, 2013. v. 33. 13

ENGLE, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, JSTOR, p. 987–1007, 1982. 10

FISSLER, T.; ZIEGEL, J. F.; GNEITING, T. Expected shortfall is jointly elicitable with value at risk-implications for backtesting. *arXiv preprint arXiv:1507.00244*, 2015. 10, 24, 25, 26, 30

FÖLLMER, H.; SCHIED, A. Convex measures of risk and trading constraints. *Finance and stochastics*, Springer, v. 6, n. 4, p. 429–447, 2002. 13

FRONGILLO, R.; KASH, I. A. Vector-valued property elicitation. In: *Conference on Learning Theory*. [S.l.: s.n.], 2015. p. 710–727. 24

GEWEKE, J. Bayesian treatment of the independent student-t linear model. *Journal of applied econometrics*, Wiley Online Library, v. 8, n. S1, p. S19–S40, 1993. 20, 21

GIACOMINI, R.; WHITE, H. Tests of conditional predictive ability. *Econometrica*, Wiley Online Library, v. 74, n. 6, p. 1545–1578, 2006. 27

GLOSTEN, L. R.; JAGANNATHAN, R.; RUNKLE, D. E. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, Wiley Online Library, v. 48, n. 5, p. 1779–1801, 1993. 19

GNEITING, T. Making and evaluating point forecasts. *Journal of the American Statistical Association*, Taylor & Francis, v. 106, n. 494, p. 746–762, 2011. 13, 24

HASTINGS, W. K. Monte carlo sampling methods using markov chains and their applications. Oxford University Press, 1970. 22

HILL, B. M. A simple general approach to inference about the tail of a distribution. *The annals of statistics*, JSTOR, p. 1163–1174, 1975. 18

KOENKER, R.; HALLOCK, K. F. Quantile regression. *Journal of economic perspectives*, v. 15, n. 4, p. 143–156, 2001. 13

LAMOUREUX, C. G.; LASTRAPES, W. D. Persistence in variance, structural change, and the garch model. *Journal of Business & Economic Statistics*, Taylor & Francis, v. 8, n. 2, p. 225–234, 1990. 9

MCNEIL, A. J.; FREY, R. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of empirical finance*, Elsevier, v. 7, n. 3-4, p. 271–300, 2000. 18, 27, 32, 34

METROPOLIS, N. et al. Equation of state calculations by fast computing machines. *The journal of chemical physics*, AIP, v. 21, n. 6, p. 1087–1092, 1953. 22

NEWEY, W. K.; POWELL, J. L. Asymmetric least squares estimation and testing. *Econometrica: Journal of the Econometric Society*, JSTOR, p. 819–847, 1987. 13, 30

NOLDE, N.; ZIEGEL, J. F. et al. Elicitability and backtesting: Perspectives for banking regulation. *The annals of applied statistics*, Institute of Mathematical Statistics, v. 11, n. 4, p. 1833–1874, 2017. 10, 12

PATTON, A. J. Modelling asymmetric exchange rate dependence. *International economic review*, Wiley Online Library, v. 47, n. 2, p. 527–556, 2006. 30

PATTON, A. J. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, Elsevier, v. 160, n. 1, p. 246–256, 2011. 30

SHEIKH, A. Z.; QIAO, H. Non-normality of market returns: A framework for asset allocation decision making. *The Journal of Alternative Investments*, Euromoney Institutional Investor PLC, v. 12, n. 3, p. 8, 2010. 9

TSAY, R. S. *An introduction to analysis of financial data with R*. [S.l.]: John Wiley & Sons, 2014. 16, 17, 20

WEBER, S. Distribution-invariant risk measures, information, and dynamic consistency. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, Wiley Online Library, v. 16, n. 2, p. 419–441, 2006. 13, 14, 24

ZIEGEL, J. F. Coherence and elicitability. *Mathematical Finance*, Wiley Online Library, v. 26, n. 4, p. 901–918, 2016. 11, 13, 14