Towards a Well-diversified Risk Measure: A DARE Approach[*]

Benjamin HAMIDI[†] Christophe HURLIN[‡]
Patrick KOUONTCHOU[§] Bertrand MAILLET[¶]

October 2013

Abstract

This paper provides a complete framework to aggregate different Quantile and Expectile models for obtaining more diversified Value-at-Risk (VaR) and Expected Shortfall (ES) measures, by applying the diversification principle to model extreme market risks. Following Taylor (2008a) and (2008b), Gouriéroux and Jasiak (2008) and Kuan *et al.* (2009), we introduce a new class of models for the VaR and ES modeling called Dynamic AutoRegressive Expectiles (DARE). We first briefly present the main literature about VaR and ES estimations. We secondly explain the DARE approach and how Expectiles can be used to estimate Quantile risk measures. We thirdly use the main and recent validation tests developed in the literature to compare the DARE approach to other traditional methods for computing extreme risk measures on the French stock market. We evaluate finally, the impact of several conditional weighting functions (in the spirit of Hansen, 2008) of the various risk models aggregated within the DARE Approach, in order to dynamically select the more relevant global Quantile model for extremal risk measure estimation.

Keywords: Expected Shortfall, Value-at-Risk, Expectile, Risk Measures, Backtests.
JEL Classification: C14, C15, C50, C61, G11.

[*]We thank Georges Bresson, Christophe Boucher, Thierry Chauveau, Gilbert Colletaz, Gregory Jannin, Jean-Philippe Médecin and Paul Merlin for their help, suggestions and encouragements when preparing this work. We also thank the co-editor and the editor of the review (Franck Moraux), as well as the the anonymous referee, for their advices. The fourth author thanks the support of the Risk Foundation Chair Dauphine-ENSAE-Groupama "*Behavioral and Household Finance, Individual and Collective Risk Attitudes*" (Louis Bachelier Institute) and all authors acknowledge the support by the Global Risk Institute (www.globalriskinstitute.com).

[†]Neuflize OBC Investments. Email: benjamin.hamidi@fr.abnamro.com.

[‡]Univ. Orleans (LEO/CNRS). Email : christophe.hurlin@univ-orleans.fr.

[§]Variances and Univ. Lorraine (CEREFIGE). Email: patrick.kouontchou@univ-lorraine.fr.

[¶]A.A.Advisors-QCG (ABN AMRO), Variances, Univ. La Reunion and Orleans (CEMOI, LEO/CNRS and LBI). Correspondance to: Pr. Bertrand B. Maillet, Univ. La Reunion, CEMOI, 15 avenue Réné Cassin - CS 92003 - 97744 Saint-Denis Cedex 9 France. Email: bertrand.maillet@univ-reunion.fr.

Towards a Well-diversified Risk Measure: A DARE Approach

October 2013

Abstract

This paper provides a complete framework to aggregate different Quantile and Expectile models for obtaining more diversified Value-at-Risk (VaR) and Expected Shortfall (ES) measures, by applying the diversification principle to model extreme market risks. Following Taylor (2008a) and (2008b), Gouriéroux and Jasiak (2008) and Kuan *et al.* (2009), we introduce a new class of models for the VaR and ES modeling called Dynamic AutoRegressive Expectiles (DARE). We first briefly present the main literature about VaR and ES estimations. We secondly explain the DARE approach and how Expectiles can be used to estimate Quantile risk measures. We thirdly use the main and recent validation tests developed in the literature to compare the DARE approach to other traditional methods for computing extreme risk measures on the French stock market. We evaluate finally, the impact of several conditional weighting functions (in the spirit of Hansen, 2008) of the various risk models aggregated within the DARE Approach, in order to dynamically select the more relevant global Quantile model for extremal risk measure estimation.

Keywords: Expected Shortfall, Value-at-Risk, Expectile, Risk Measures, Backtests.
JEL Classification: C14, C15, C50, C61, G11.

Towards a Well-diversified Risk Measure: A DARE Approach

1 Introduction

Value-at-Risk (VaR) measures the potential loss of a given portfolio over a specified holding period at a specified confidence level, which is commonly fixed at .5%, 1% or 5%. The Basel Committee on Banking Supervision (1996) at the Bank for International Settlements imposes on financial institutions, such as banks and investment firms, to get required adequate capital based on VaR estimates. Losses in most banks' trading book during the financial crisis have been significantly higher than the minimum capital requirements under the former Pillar 1 market risk rules because VaR was underestimated. It leads to the revision of Basel II market risk framework (2009). A stressed VaR requirement taking into account an observation period relating to significant losses was thus introduced. Although VaR became the standard measure of market risk, it has been criticized for reporting only a Quantile, and thus disregarding all the precise nature of outcomes beyond the threshold. In addition, VaR is not a subadditive risk measure. This property concerns the main idea of risk diversification: the aggregated risk of a portfolio should not be greater than the individual risk of its constituent parts (Artzner *et al.*, 1999; Acerbi and Tasche, 2002). The Expected Shortfall (ES) is a risk measure that precisely overcomes this weakness, also giving more accurate information of what the tail of the return distribution is made of. Indeed, the ES is defined as the conditional expectation of the return given that it falls below the VaR (Yamai and Yoshiba, 2002; Artzner *et al.*, 1999; Pflug, 2000).

Based on these definitions of VaR and ES, many estimation methods coexist nowadays. However, none of these methods seems adapted to every market configurations. Besides, no one passes all validation tests proposed in the literature (Hurlin and Tokpavi, 2008). The risk associated to the choice of the extreme risk model is therefore important. In

theory and practice, assessing model risk is a true challenge. While validation procedures for pricing models have traditionally focused on implementation and accuracy, the increasingly important role of models in trading and risk management is such that model misspecification (*i.e.* the "model risk") has become a major risk which cannot be ignored in a validation procedure. It is therefore important to quantify "model uncertainty" and incorporate in order to achieve a robust risk management framework. Moreover, the Basel Committee published in July 2009 a recommendation to Banks to assess and quantify "model Risk" in order to produce adequate reserves. Considering the economic impact of the quality of these risk measures for the banks capital constraints, it seems crucial to limit the risk model. The objective of this paper is to provide a complete framework to aggregate different Quantile and Expectile models (Hansen, 2008), to provide more diversified VaR and ES measures, thus limiting the model risk in a similar way that has been followed for the volatility in a multi-model approach (Visser, 2008).

A recent development in the VaR literature concerns the Conditional AutoRegressive Value-at-Risk (CAViaR) class of models (Engle and Manganelli, 2004), Multi-Quantile CAViaR (Kim *et al.*, 2008) and Quantile GARCH (QGARCH) model (Koenker and Xiao, 2009). These general approaches to the VaR estimation has a strong appeal insofar as they provide an extensive modeling framework and do not rely on strong distributional assumptions. However, only VaR estimation has been investigated until now, and a similar work on the corresponding ES should be done.

Following Taylor (2008a and 2008b), Gouriéroux and Jasiak (2008) and Kuan *et al.* (2009), we present in this paper a new modeling approach that defines estimations for both VaR and ES. The approach involves the use of Asymmetric Least Squares (ALS) regression, whose solution is called an "Expectile". This name was given by Newey and Powell (1987) who note that the ALS solution is determined by the properties of the expectation of exceedances beyond the solution. We use in the following this result to estimate the ES. It has also been shown that there exists a one-to-one mapping from Expectiles to Quantiles. Indeed, Efron (1991) proposes the α-Quantile to be estimated by the Expectile for which the proportion of in-sample observations lying below the Expectile is equal to α. This idea can also be used to estimate the VaR and the ES from Expectiles as shown later on. Aggregating different Quantile and Expectile models, we finally propose in this article a new class of

models for the conditional VaR and ES modeling: the Dynamic AutoRegressive Expectiles (DARE, in short), aiming to provide more diversified extreme risk measures.

The paper is organized as follows. In section 2, we briefly review the main literature about VaR and ES measures and estimation models. In the third section, we describe how Expectiles can be used to estimate VaR and ES, then we introduce the DARE approach. We shortly present and summarize in the fourth section the main and recent backtest procedures for VaR and ES models. According to these validation tests, we provide empirical results obtained on the French Stock Market, related to the accuracy of the proposed method compared to traditional ones. We finally illustrate the impact of several conditional weighting functions to the risk models aggregated into the DARE approach, in order to dynamically select the more appropriated Quantile models for extremal risk measure estimation. The last section concludes.

2 Main Extreme Market Risk Models

Risk management has become in these past few years a central object of interest for researchers, market practitioners and regulators. It involves the computations and follow-ups of some risk measures that can be viewed as single statistics of asset/portfolio returns. We shall first herein focus on extreme risk definitions and estimation methods. We hereafter introduce and define the VaR and the ES risk measures, focusing on the way they allow us to disentangle the behavior of tails. Then, we review the main estimation approaches of these extreme risk measures, which are traditionally classified into three categories: parametric, non-parametric and semi-parametric methods (Engle and Manganelli, 1999).

2.1 Extreme Risk Measure Definitions

The VaR is well known as being one of the benchmark measure for market risk. It is precisely defined as the potential loss a portfolio may suffer from, at a given confidence level over a fixed holding period, such as at time t:

$$Prob[r_t < -VaR_{\alpha,t}] = 1 - \alpha, \tag{1}$$

where r_t is the random variable of asset returns, $Prob\,[.]$ stands for the unconditional probability, and α is the confidence level for the VaR computation.
The Value-at-Risk of a return distribution can thus be viewed as an α-Quantile, such as:

$$q_{1-\alpha}(.) = -VaR_{\alpha}(.) \tag{2}$$

where $q_{1-\alpha}(.)$ is the Quantile Function.
This Quantile Function has to be an increasing function of the risk level α. Quantile estimators have also to fulfill the monotonicity property with α. Thus, a well specified quantile model is expected to provide estimators that behave like true quantiles and to increase for any value of parameters and conditioning variables according to the risk level α. Quantile Functions have other properties as given in the following examples. If $q_{\alpha}(.)$ is a Quantile Function associated to α defined on $[0,\ 1]$, and a is a positive constant, then:

$$q^{*}_{\alpha}(.) = -q_{1-\alpha}(.) \quad \text{and} \quad q^{**}_{\alpha}(.) = q_{\alpha^{a}}(.), \tag{3}$$

are also Quantile Functions.
If $q^{(k)}_{(.)}(.)$ are several Quantile Functions with identical real domains, with $k = [1, ..., n]$ and a_k are positive constants, then:

$$q^{***}_{\alpha}(.) = \sum_{k=1}^{n} a_k\, q^{(k)}_{\alpha}(.), \tag{4}$$

is also a Quantile Function.
These properties can be used to derive new Quantile Functions from other ones.

Although the VaR can be mentioned as a reference and is imposed by regulators, criticisms have been formulated against its generalized use (Beder, 1995; Cheridito and Stadje, 2009). For instance, VaR does not enable us to answer the question: "what is the magnitude of the loss when the VaR limit is exceeded?" Another issue, pointed out by Artzner *et al.* (1999), concerns the "non-coherence" property of this risk measure; it fails indeed to respect the subadditivity property, *i.e.* the VaR of a combined portfolio can be larger than the sum of the VaRs of its components (diversification principle). Prause (1999) also argues that, when bankruptcy in a stake, we should be more interested the distribution of the Maximum Expected Loss. From this point of view, some authors suggest to use other risk measures than the VaR, in order to get a better characterization of extreme events, especially for non-linear portfolio returns.

Another risk measure has therefore recently appeared in the literature, named the Expected Shortfall (ES), that is the value of the expected loss at a given confidence level. Contrary to VaR, this measure satisfies the subadditivity property mentioned above, and provides information about the magnitude of the loss when the VaR is exceeded. More formally, following the definition by Acerbi and Tasche (2002), the ES can be written as such (with the previous notations):

$$ES_{\alpha,t} = -E\left[r_t | r_t \leq -VaR_{\alpha,t}\right]. \tag{5}$$

If we consider that at time t the distribution of r_t is known, we have:

$$ES_{\alpha,t} = -(1-\alpha)^{-1} \int_{-\infty}^{-VaR_{\alpha,t}} r_t f_t(r_t) dr_t, \tag{6}$$

where $f_t(.)$ is the Probability Density Function of r_t at time t.
Thus, VaR and ES can be generically written such as:

$$\begin{cases} VaR_{\alpha,t} = \mu_t + F^{-1}_{1-\alpha,t}(r^*_t)\sigma_t \\ ES_{\alpha,t} = \mu_t + \left\{(1-\alpha)^{-1} f_t[F^{-1}_{1-\alpha,t}(r^*_t)]\right\}\sigma_t, \end{cases} \tag{7}$$

where μ_t and σ_t are respectively the expected return and standard deviation at time t, $f_t(.)$ and $F_t(.)$ are, respectively, some specific Probability Densities and related Cumulative Distribution Functions evaluated at time t, and r^*_t is the standardized return.

In the relation (7), the key for the calculation of empirical estimates of VaR and ES is in the estimation of the distribution function of the returns. The next subsection gives the mainly common alternatives presented in the literature.

2.2 Estimation Methods

Several approaches are used to estimate extreme market risk measures as Value-at-Risk and Expected Shortfall. The choice will depend upon the kind of portfolio, the availability of computational resources and time constraints. The main approaches are traditionally classified into three categories: non-parametric, parametric and semi-parametric methods (Engle and Manganelli, 1999; Jorion, 2006; Nieto and Ruiz, 2008).

2.2.1 Non-parametric Approach of Extreme Risk Measure

The most widely used non-parametric method to estimate a predetermined Quantile is based on the so-called historical approach. It only requires mild distributional assumptions and it implies the estimation of the VaR as the Quantile of the empirical distribution of past historical returns from a dynamic moving window of the most recent period. The empirical VaR can be written such as (with the previous notations):

$$\widehat{VaR}_{\alpha,t} = \widehat{F}^{-1}_{1-\alpha,t}(r_t), \tag{8}$$

where $\widehat{F}^{-1}_{1-\alpha,t}(.)$ is the Quantile Function associated to the probability $1-\alpha$, estimated at time t on past returns.

The main difficulty only concerns the width of considered window for the measure practical computation: a few observations only will lead to an important sampling error, whereas too many observations will slow down the reaction of estimates to changes in the true distribution of financial returns.

Other methods allocate dynamically decreasing weights to the past sample of returns exponentially (the sum being equal to one). Past returns are then ordered in ascending order from the lowest return and the weights are aggregated to reach the given confidence rate; the conditional Quantile estimate is set as the return that corresponds to the final weight used in the previous summation. The forecast is then built from an Exponentially Weighted Average of past observations. If the distribution of returns is moving quickly over time, a relatively fast exponential decay is needed to ensure a reactive adaptation. These exponential smoothing methods are simple and common approaches in practice.

These non-parametric approaches only take into consideration realized past returns on a predefined data sample. These approaches can not model tail returns that have never been realized.

2.2.2 Parametric Approach of Extreme Risk Measure

The parametric approach assumes that returns follow a specific probability distribution, as for example a Normal or a t-Student law. The parameters of the relevant distribution are specified and the risk measure is deduced thanks to the Quantile of the estimated distribution. Consequently, the risk measure mainly depends on the parameter estimation and on the shape

of the chosen distribution. Assuming that $F_t(.)$ is the Cumulative Distribution Function (CDF) of returns at time t, the estimated VaR can be written such as:

$$\widehat{VaR}_{\alpha,t} \;=\; \widehat{\mu}_t + F^{-1}_{1-\alpha,t}(r^*_t|\widehat{\theta}_t)\,\widehat{\sigma}_t, \tag{9}$$

where $\widehat{\theta}_t = (\widehat{\sigma}_t, \widehat{\mu}_t, ...)$ represents the vector of the estimation of the ditribution parameters at time t, r_t is the return and r^*_t is the standardized return.

When modeling asset returns, the Normal distribution remains often used. Assuming at time t a Gaussian Cumulative Distribution Function of returns $\Phi_t(.)$, the specific CDF in equation (9) is replaced by $\Phi_t(.)$. However, Normal distribution suffers from numerous drawbacks. Actually, the Normal distribution has thinner tails than the empirical distribution of returns. Moreover it may be misleading to estimate VaR or ES assuming normally distributed innovations in a model where financial return series are heteroskedastic. Using the t-distribution, the Normal Inverse Gaussian (Barndorff-Nielsen, 1998; Venter and de Jongh, 2002), or applying Extreme Value Theory (EVT, see Embrechts *et al.*, 1997) are other possibilities to avoid these problems and are presented below (see also Bali, 2003; Longin, 2000; Danielsson and de Vries, 2000; Bekiros and Georgoutsos, 2005; Martins-Filho and Yao, 2006; Tolikas, 2008).

In an EVT framework, let us suppose that r_t represents the return at time t and $\underline{r}_s = \max(r_1, r_2, \ldots, r_s)$ is the *maxima* for the variable r_t on the subsample s. Thus, $\underline{r}_s$ represents the highest returns on blocks of data, or (alternatively) peaks over a fixed threshold on a subsample s. According to the Generalized Extreme Value (GEV) or to the Generalized Pareto Distribution (GPD), the *maxima* variable Cumulative Distribution Functions have formal expressions recalled hereafter (Embrechts *et al.*, 1997).

For the GEV family, the Cumulative Distribution Function, $F_t(.)$, can be written at time t, such as (with the previous notations):

$$F_t(\underline{r}_s|\nu_t, \theta_t, \xi_t) = \begin{cases} \exp\left\{-\left[1+\xi_t(\underline{r}_s-\nu_t)\theta_t^{-1}\right]^{-1/\xi_t}\right\} & \text{if } \xi_t \neq 0 \\ \exp\left\{-\exp\left[-(\underline{r}_s-\nu_t)\theta_t^{-1}\right]\right\} & \text{otherwise,} \end{cases} \tag{10}$$

with $[1+\xi_t\,(\underline{r}_s-\nu_t)\,\theta_t{}^{-1}] \geq 0$, where $\nu_t \in \mathbb{R}$, $\theta_t \in \mathbb{R}^*_+$ and $\xi_t \in \mathbb{R}$ are, respectively, location, scale and shape parameters at time t.

For the GPD family, the Cumulative Distribution Function $F_t(.)$ can be written at time t, such as (with the previous notations):

$$F_t(\underline{r}_s|\nu_t,\theta_t,\xi_t) = \begin{cases} 1-\left[1+\xi_t(\underline{r}_s-\nu_t)\theta_t^{-1}\right]^{-1/\xi_t} & \text{if } \xi_t \neq 0 \\ 1-\exp\left[-(\underline{r}_s-\nu_t)\theta_t^{-1}\right] & \text{otherwise.} \end{cases} \tag{11}$$

Finally, following Bali (2003), the estimated VaR can be written such as:

$$\widehat{VaR}_{\alpha,t} = \widehat{\nu}_t + \widehat{F}^{-1}_{1-\alpha,t}(\underline{r}^*_s)\ \widehat{\theta}_t, \tag{12}$$

where $\widehat{F}^{-1}_{1-\alpha,t}(\underline{r}^*_s)$ is the Inverse Standardized Empirical Extreme Value Cumulative Distribution Function given by $\widehat{F}^{-1}_{1-\alpha,t}(\underline{r}^*_s) = \widehat{F}^{-1}_{1-\alpha,t}(\underline{r}_s|0,1,\xi)$ for the $(1-\alpha)$ Quantile, with $\underline{r}^*_s$ being the centered standardized extremal return denoted $\underline{r}_s$.

For VaR estimations, it has been shown that Normal Inverse Gaussian (NIG) based approach can be more robust than the EVT method (Barndorff-Nielsen, 1998). The EVT method (Embrechts *et al.*, 1997) seems however more adapted in large samples and if the NIG distribution does not adequately fit the returns. In the case of symmetric distributions, the t-based approach suits well and is comparable to the NIG-based approach. When focusing on Expected Shortfall, the NIG-based approach can be preferred to the EVT method when the available sample is small (Venter and de Jongh, 2002).

However, the mentioned parametric approaches are affected by the assumed return probability distribution specification, which does not provide a perfect representation of the main stylized items of financial series. The alternative is to more focus on the return characteristics without applying the filter of a pre-determined hypothesis on the return densities as presented in the following sub-section.

2.2.3 Semi-parametric Approach of Extreme Risk Measure

Semi-parametric estimation methods combine the two previous approaches. Expansion-based and Quantile Regression-type approaches belong to this family. Three types of semi-parametric methods are presented below.

We first present the Cornish-Fisher approach (1937) based on statistical series expansions of the returns Probability Density Function around a reference density, which relies on higher-order moments of the estimated distribution. Assuming for example an underlying non-Gaussian distribution of the returns (but not too different from the Gaussian one), the terms of the statistical series expansions aim to adjust the Gaussian reference distribution by including the skewness and excess *kurtosis* of the returns, in order to be as close as possible of the true underlying distribution.

Formally, the Probability Density Function at time t denoted $f_t(.)$ is given such as (with the previous notations):

$$f_t(r_t) = \psi_t(r_t)\ \phi(r_t) + \epsilon_t, \tag{13}$$

with:

$$\psi_t(r_t) = \left\{1 + \frac{\gamma_1[f_t(r_t)]}{3!}P_2(r_t) + \frac{\gamma_2[f_t(r_t)]}{4!}P_3(r_t) + 10\frac{\gamma_1[f_t(r_t)]^2}{36}P_6(r_t)\right\},$$

where $\phi(.)$ is the standard Normal Probability Density Function, γ_1 is the skewness Pearson coefficient, γ_2 is the excess *kurtosis* Pearson coefficient, ϵ_t is the residual of the expansion and $P_i(r_t) = (-1)^i \phi(r_t)^{-1}\left[\frac{\partial^i \phi(r_t)}{\partial r_t^i}\right]$ the Hermite polynomial of order i.

Thus, at time t, the α-Quantile denoted $q_{\alpha,t}(.)$, is then:

$$\begin{aligned} q_{\alpha,t}(r_t) = \Phi^{-1}(\alpha) + \tfrac{\gamma_1[f_t(r_t)]}{3!}P_2\Phi^{-1}(\alpha) + \tfrac{\gamma_2[f_t(r_t)]}{4!}P_3\Phi^{-1}(\alpha) \\ -\tfrac{\gamma_1[f_t(r_t)]^2}{36}\left\{2\left[\Phi^{-1}(\alpha)\right]^3 - 5\Phi^{-1}(\alpha)\right\}, \end{aligned} \tag{14}$$

where $\Phi(.)$ is the standard normal Probability Density Function.

The Cornish-Fisher VaR (Favre and Galeano, 2002), using the principle of a statistical series expansion for the estimation of an α-Quantile, is written such as (with the previous notations):

$$\widehat{VaR}_{\alpha,t} = \widehat{\mu}_t + \widehat{q}_{1-\alpha,t}(r_t)\ \widehat{\sigma}_t, \tag{15}$$

where $\widehat{\sigma}_t$ and $\widehat{\mu}_t$ are respectively the estimations of the returns standard deviation and of the expected return at time t.

The *RiskMetrics* (1996) model is here considered as a second semi-parametric approach. It assumes that asset returns follow a standard Normal distribution, with a time-varying volatility that is estimated by the Exponential Weighted Moving Average method (which

corresponds to a so-called Integrated GARCH model), obtained by using historical data such as:

$$\widehat{\sigma}_t^2 = \lambda\, \widehat{\sigma}_{t-1}^2 \; + (1-\lambda)\; r_{t-1}^2, \tag{16}$$

where $\widehat{\sigma}_t^2$ is the estimation of the returns variance at time t and λ is an *ad hoc* smoothing parameter associated to the *RiskMetrics* model (arbitrary fixed to .94).

Adapting this framework, many other volatility estimation methods, belonging for example to the GARCH family, have been already tested in the literature (Engle and Rangel, 2008; Brownlees and Gallo, 2010). These approaches are also called Conditional Normal methods as we use a Normal distribution to describe the returns associated to a variance estimated by conditional methods. This approach can be adapted to other distribution assumptions.

A third semi-parametric approach presented in the paper is based on Quantile Regression methods. Quantile regression estimation methods need mild distributional assumptions. Conditional AutoRegressive VaR (CAViaR) introduced by Engle and Manganelli (2004) and Quantile GARCH (QGARCH) model (Koenker and Xiao, 2009) is one of them. Koenker and Xiao (2009) propose a Quantile regression estimation of GARCH Models. Quantile regression provides a convenient approach of estimating conditional Quantiles. It has the important virtue of robustness to distributional assumptions and makes no prior presumption about the symmetry of the innovation process.

Engle and Manganelli (2004) directly define the dynamics of risk by the mean of an AutoRegressive model involving the lagged-VaR and the lagged value of the endogenous variable, in a model named CAViaR. They present four CAViaR specifications: a model with a symmetric absolute value, an asymmetric slope, an Indirect GARCH(1,1) and an adaptive form, denoted respectively: $VaR_{\alpha,t}^{SAV}(r_{t-1},\beta)$, $VaR_{\alpha,t}^{AS}(r_{t-1},\beta)$, $VaR_{\alpha,t}^{IG}(r_{t-1},\beta)$, $VaR_{\alpha,t}^{A}(r_{t-1},\beta)$ and properly defined below.

Symmetric Absolute Value CAViaR

The Symmetric Absolute Value CAViaR specification is defined such as (with the previous notations):

$$VaR_{\alpha,t}^{SAV}(r_{t-1},\beta) = \beta_1 + \beta_2 \times VaR_{\alpha,t-1}^{SAV}(r_{t-1},\beta) + \beta_3 \times |r_{t-1}|\,, \tag{17}$$

where β_i, with $i = [1, ..., 3]$, are CAViaR model parameters and r_t is the risky asset return at time t.

In this CAViaR specification, the VaR reacts symmetrically to positive or negative returns.

Asymmetric Slope CAViaR

The Asymmetric Slope CAViaR can be written such as (with the previous notations):

$$\begin{aligned} VaR_{\alpha,t}^{AS}(r_{t-1}, \beta) = \beta_1 + \beta_2 \times VaR_{\alpha,t-1}^{AS}(r_{t-1}, \beta) + \beta_3 \times \max(0, r_{t-1}) \\ + \beta_4 \times [-\min(0, r_{t-1})]. \end{aligned} \tag{18}$$

The Asymmetric Slope specification can thus model the asymmetric leverage effect, which is the tendency for volatility to be greater following a negative return than a positive return of equal size.

Indirect GARCH(1,1) CAViaR

The Indirect GARCH(1,1) CAViaR is defined such as (with the previous notations):

$$VaR_{\alpha,t}^{IG}(r_{t-1}, \beta) = \beta_1 + \beta_2 \times \left[VaR_{\alpha,t-1}^{IG}(r_{t-1}, \beta)\right]^2 + \beta_3 \times r_{t-1}^2. \tag{19}$$

We notice that the Indirect GARCH(1,1) CAViaR model is correctly specified if the underlying data are generated by a GARCH(1,1) model with an Independently and Identically Distributed residual.

Adaptive CAViaR

The Adaptive CAViaR can be expressed such as (with the previous notations):

$$\begin{aligned} VaR_{\alpha,t}^{A}(r_{t-1}, \beta) = VaR_{\alpha,t-1}^{A}(r_{t-1}, \beta) - \alpha \\ + \beta_1 \left[1 + \exp\left\{.5 \times \left[r_{t-1} - VaR_{\alpha,t-1}^{A}(r_{t-1}, \beta)\right]\right\}\right]^{-1}. \end{aligned} \tag{20}$$

The intuition associated to the Adaptive specification is the following: as long as the daily return is not inferior to the VaR estimation, it can increase by a small amount; on the contrary, when the past value is exceeded, it has to decrease. Thus, this model is adapted to larger and larger falls of the risky asset, since it adapts itself to past errors and reduces the probability that the VaR is consecutively under estimated, but it does not guarantee that the VaR is not over-estimated.

Discussing the general CAViaR model without the autoregressive component, the conditional Quantile Function is well defined if parameters can also be considered as Quantile Functions. In fact, CAViaR models weight different baseline Quantile Functions at each date and it can be therefore considered as Quantile Functions. Adding a non-negative autoregressive component of VaR, the CAViaR conditional Quantile Function becomes a linear combination of Quantile Functions weighted by non-negative coefficients. Thus, CAViaR model satisfies the properties of a Quantile Function, even if the indirect GARCH(1,1) or adaptive specifications do not satisfy the monotonicity property. CAViaR model parameters are estimated by using the Quantile regression minimization (denotes QR Sum) presented by Koenker and Bassett (1978):

$$\widehat{\beta}^* = \underset{\widehat{\beta}\in\mathbb{R}^n}{ArgMin} \left\{ \sum_{t=1}^{T} \left\{ \left[\alpha - 1\mathrm{I}_{\left\{r_t < -\widehat{VaR}_{\alpha,t}\left(r_{t-1},\widehat{\beta}\right)\right\}}\right] \left[r_t + \widehat{VaR}_{\alpha,t}\left(r_{t-1},\widehat{\beta}\right)\right] \right\} \right\}, \tag{21}$$

with $\widehat{VaR}_{\alpha,t}\left(r_{t-1},\widehat{\beta}\right)$ is the VaR estimation according to CAViaR models (see equations 17 to 20), $\widehat{\beta}$ is a vector of estimated parameters, $1\mathrm{I}_{\{\cdot\}}$ is the indicator function and T is the size of the data set.

When the Quantile model is linear, this minimization can be formulated as a linear program for which the dual problem is conveniently solved. Koenker and Bassett (1978) show that the resulting Quantile estimator, $\widehat{VaR}_{\alpha,t}$, essentially partitions the r_t observations so that the proportion less than the corresponding Quantile estimate is α.

The procedure proposed by Engle and Manganelli (2004) to estimate their CAViaR models is to generate vectors of parameters from a uniform random number generator between zero and one, or between minus one and zero (depending on the appropriate sign of the parameters). For each of the vectors, the QR Sum is then evaluated. The ten vectors that produced the lowest values for the function are used as initial values in a quasi-Newton algorithm. The QR Sum is calculated for each of the ten resulting vectors, and the one which produces the lowest value of the QR Sum is chosen as the final parameter vector.

CAViaR models only provide a model for the Quantile, but computing by the corresponding ES is not straightforward. The ES can be conveniently computed using conditional Expectile models. Following Taylor (2008a), we use CAViaR models to build Conditional AutoRegressive Expectile (CARE) models (Kuan *et al.*, 2009), replacing the conditional VaR by the conditional Expectile in equations (17), (18), (19) and (20). We estimate CARE

models parameters using the same procedure as for CAViaR models, except that the QR sum is replaced by the Asymmetric Least Square Sum (see equation 23). The τ-Expectile, for which the proportion α of in-sample observations below the Expectile, is the estimator of the α-Quantile (Efron, 1991; Granger and Sin, 2000).

3 The DARE Approach

We propose in this section a way to aggregate well specified Expectiles and Quantile models for obtaining a good estimation of Quantiles based risk measures such as VaR and ES.

We now focus on these two main points to answer this issue. On the one hand, we have to provide a unified framework which enables us to aggregate Quantile estimation methods, involving the introduction hereafter of the DARE approach. On the other hand, we need to evaluate whether Quantiles based on risk measures are well specified. Our evaluation *criteria* are the VaR and ES tests developed in the literature. This section briefly reminds the definition of Expectiles, and their uses to estimate ES and VaR with Conditional AutoRegressive Expectile class of models. Then, we introduce our DARE methodology to aggregate CARE model with a unified Quantile estimation method (the tests used to evaluate these methods are presented in the next section).

VaR and Expected Shortfall can be estimated with Expectiles calculated through the minimization of the following optimization program:

$$\widehat{\mu}^{*}_{\tau,t} = \underset{\widehat{\mu}_{\tau,t}\in\mathbb{R}}{ArgMin} \left\{ E\left[\left|\tau - 1\mathrm{I}_{\{r_t<\widehat{\mu}_{\tau,t}\}}\right| \left(r_t - \widehat{\mu}_{\tau,t}\right)^2 \right] \right\}, \tag{22}$$

where the τ-Expectile of r_t is estimated by the parameter $\widehat{\mu}_{\tau,t}$ and $|.|$ is the absolute value operator.

The τ-Expectile is thus the solution to the minimization of asymmetrically weighted mean squared errors. Due to the quadratic loss function, Expectiles are sensitive to extreme values of the distribution. As such, τ will be referred to as an index of prudentiality. Similar to the definition of VaR, the τ-Expectile is understood as the maximal potential loss within a given holding period under a specific prudentiality level. Expectile can be distinguished from ES

because the latter is determined by a conditional downside mean, which depends only on the tail event and hence is much larger (more conservative) than corresponding Expectile and Quantile. Taking into account the magnitude of potential losses, the Expectile may be a better measure for tail risk.

Parameters of a conditional model for Expectiles can be estimated by using the Asymmetric Least Square (ALS) regression, which is the least square analogue for Quantile regression, such as:

$$\widehat{\beta}^{*} = \underset{\widehat{\beta}\in\mathbb{R}^{n}}{ArgMin}\left\{\sum_{t=1}^{T}\left\{\left|\tau - 1\mathrm{I}_{\left\{r_t<\widehat{\mu}_{\tau,t}\left(r_{t-1};\widehat{\beta}\right)\right\}}\right| \times \left[r_t - \widehat{\mu}_{\tau,t}\left(r_{t-1};\widehat{\beta}\right)\right]^2\right\}\right\}, \tag{23}$$

where $\widehat{\mu}_{\tau,t}$ is the estimation of $\mu_{\tau,t}$ and $\widehat{\beta}$ is the estimated parameter vector of a specific Expectile conditional model.

Expectile can therefore be interpreted as a flexible VaR for the underlying return distribution. One should choose a smaller (larger) α for VaR if the left tail of the return distribution were known to be thicker (thinner). Yet, the shape of a return distribution is rarely known in practice, and α is typically set by regulators and/or the management level.

We can use Expectiles as Quantile estimators, knowing that there is a corresponding α-Quantile for each τ-Expectile (see Efron, 1991; Jones, 1994; Abdous and Remillard, 1995; Yao and Tong, 1996), such as (with the previous notations):

$$\tau = \frac{\alpha \times (-VaR_{\alpha,t}) - \int_{-\infty}^{-VaR_{\alpha,t}} r_t \; df_t(r_t)}{\mu_t - 2\int_{-\infty}^{-VaR_{\alpha,t}} r_t \; df(r_t) + (1-2\alpha)\times(VaR_{\alpha,t})}, \tag{24}$$

where μ_t is the returns expected return at time t and $f_t(.)$ is the return Probability Density Function at time t.

An Expectile at a given τ corresponds to Quantiles with different α under distinct distributions, and hence represents different risk exposures in terms of the probability (frequency) of tail losses. Using Expectile to estimate risk measures, the confidence level α is not *a priori* determined, but the parameter τ can be estimated so as the α-Quantile and the τ-Expectile perfectly match.

Taylor (2008a and 2008b) and Kuan *et al.* (2009) clearly explain the link between the Expectile and Expected Shortfall, that leads to (with the previous notations):

$$ES_{\alpha,t} = \left[1 + \tau\left(1-2\tau\right)^{-1}\alpha^{-1}\right] \times \mu_{\tau,t}\left(r_t;\beta\right). \tag{25}$$

Thus, the conditional Expected Shortfall for a given value of α is proportional to the conditional γ-Quantile model, which is estimated by the τ-Expectile.

Moreover, the Quantile estimations can be linearly combined through the DAQ model proposed by Gouriéroux and Jasiak (2008). This class of dynamic Quantile models is defined by:

$$VaR_{\alpha,t} = \sum_{k=1}^{K} a_k \left(r_{t-1}, y_{t-1}, \beta_k\right) \times VaR_{\alpha,t}^{(k)} \left(\beta_k\right), \tag{26}$$

where $VaR_{\alpha,t}^{(k)}(\cdot)$ are path-independent Quantile Functions, with $k = [1, 2, \ldots, K]$ and $a_k(\cdot)$ are non-negative functions of past returns and of past exogenous variables denoted respectively (r_{t-1}) and (y_{t-1}).

Thus, a DAQ model can use different Quantile Functions to model a given one. We can also combine these functions into a multi-Quantile method (Kim *et al.*, 2008) to increase the accuracy of the conditional model. Actually, every Quantile Function can be extended to define a simple class of parametric dynamic Quantile models.

The VaR and the ES can thus be expressed as a simple combination of Quantiles associated to the same probability:

$$\begin{cases} \underset{[1\times1]}{VaR_{\alpha,t}} & = & \underset{[1\times K]}{\mathbf{W}'_{\alpha,t}} \underset{[K\times1]}{\mathbf{VaR}_{\alpha,t}} \\ \underset{[1\times1]}{ES_{\alpha,t}} & = & \underset{[1\times K]}{\mathbf{W}'_{\alpha,t}} \underset{[K\times1]}{\mathbf{ES}_{\alpha,t}}, \end{cases} \tag{27}$$

with:

$$\begin{cases} \mathbf{W}_{\alpha,t} = \begin{bmatrix} w_1 & w_2 & \ldots & w_K \end{bmatrix}' \\ \mathbf{VaR}_{\alpha,t} = \begin{bmatrix} VaR_{\alpha,t}^{(1)} & VaR_{\alpha,t}^{(2)} & \ldots & VaR_{\alpha,t}^{(K)} \end{bmatrix}' \\ \mathbf{ES}_{\alpha,t} = \begin{bmatrix} ES_{\alpha,t}^{(1)} & ES_{\alpha,t}^{(2)} & \ldots & ES_{\alpha,t}^{(K)} \end{bmatrix}', \end{cases}$$

and:

$$\begin{cases} VaR_{\alpha,t}^{(k)} = VaR_{\alpha,t}^{(k)} \left(\mu_{\tau_k}, \beta_k\right) \\ ES_{\alpha,t}^{(k)} = b_k VaR_{\gamma_k,t}^{(k)} \left(\mu_{\tau_k}, \beta_k\right) \\ b_k = \left[1 + \tau_k \left(1 - 2\tau_k\right)^{-1} \alpha^{-1}\right] \\ \sum_{k=1}^{K} w_k = 1, \end{cases}$$

where $VaR_{\theta,t}^{(k)}(.)$ are a finite number of Quantile Functions associated with probability θ and related to the model k, with $k = [1, 2, ..., K]$.

The estimation of the α-Expected Shortfall can be found by linearly aggregating K Quantile Functions and estimating the right correspondence between the probability α associated

to the Expected Shortfall of each model (denoted $ES_{\alpha}^{(k)}$) and the probability γ_k associated to each Quantile Functions (thanks to the τ_k-Expectile defined by the equation 22).

To clarify the concept of appropriate risk measures, some authors have described a number of properties that a risk measure might or might not have (Frittelli and Rosazza Gianin, 2002). Artzner *et al.* (1997 and 1999) call a risk measure coherent if it verifies the following properties: first order stochastic monotonicity (more profit associated to more risk), positive homogeneity (if your exposition is doubled, the risk associated to your position is also doubled), sub-additivity (diversification principle) and translation invariance (if you add a proportion of cash in your portfolio, its risky asset exposition decreases and the risk difference is exactly due to the proportion of the added cash). The ES respects these *criteria* (Acerbi, 2002). The notion of coherence has been subsequently relaxed. Indeed, the notions of sub-additivity and positive homogeneity can be replaced by the notion of convexity. The class of coherent risk measures is convex (Pflug, 2000; Föllmer and Schied, 2002). The DARE approach can actually be interpreted as a linear aggregation of several ES coherent risk measures. Given some known coherent risk measures, it is possible to generate a new risk measure and it is proved that a convex combination of these coherent risk measures is coherent as well (Acerbi, 2002).

Acerbi (2002) studies a space of coherent risk measures obtained as certain expansions of coherent elementary basis measures. In this space, the concept of "risk aversion function" (*i.e.* the weighting function) arises as the spectral representation of each risk measure in a space of functions of confidence level probabilities. He gives necessary and sufficient conditions on the weighting function to be a coherent measure (admissible risk spectrum *criteria*) and provides for these measures their discrete versions on finite sets of K independent realizations of a random variable, which are not only shown to be coherent measures for any fixed K, but also consistent estimators of the risk measure class for large K.

Therefore, given K coherent risk measures, any convex linear combination is another coherent risk measure, thus the DARE approach provides also a coherent risk measure contained in the generated convex hull. A spectral risk measure can be actually interpreted as a risk measure given as a weighted average of outcomes where bad outcomes are included with larger weights. These weights can be assumed to be equiprobable, but they have at least to be non negative, normalized (their sum has to be equal to one) and respect the monotonic-

ity property. Thus, as long as the weighting function of the DARE approach respects risk spectrum *criteria*, such as (with the previous notations):

$$\begin{cases} \sum_{k=1}^{K} w_k = 1 \\ w_k \geq 0 \\ w_i \geq w_j \text{ with } 1 \leq i \leq j \leq K. \end{cases} \tag{28}$$

This new space of coherent measures allows us to define a coherent spectral measure of risk. According to equation (28), several conditional weighting functions can be applied to the risk models aggregated into the DARE approach. To illustrate our DARE approach, we first adapt a standard point of view, giving equal weights to each risk model. Then we try other weighting function in order to dynamically select the more appropriate Quantile models for the extremal risk measure estimation: the DARE weighting function can then be defined as an increasing function of Quantile evaluation *criteria* (in the spirit of Hansen, 2008).

4 Backtesting Extreme Risk Measures

The DARE approach for extremal risk measure can be interpreted as a specific combination of Quantiles. Appropriate validation tests can therefore be used to define the weighting function associated to the DARE approach. Moreover, we have to compare this aggregation approach with other traditional methods for computing extreme risk measures. A sharp definition and interpretation of extreme risk measure tests is therefore of primary importance in this study.

We present hereafter, the definitions, main intuitions and properties associated to main Quantile evaluation tests developed in the literature and used in this article.

In the following section, we only give a summary of the tests. For more details, see Hendricks (1996), Christoffersen (1998), Berkowitz and Brien (2002), Christoffersen and Pelletier (2004), Ferreira and Lopez (2005), Haas (2005), Hurlin and Tokpavi (2008), Berkowitz *et al.* (2011).

The tests used in the paper evaluate extreme risk measures according to the main characteristics of VaR and ES. The first property that has to be checked to validate a Quantile estimation is to observe whether the violation frequency is not significantly different to the

probability defining the Quantile estimation (Unconditional Coverage test; Kupiec, 1995). Moreover, if the Quantile model is correct, then violations associated to the Quantile forecasting should be independently distributed (and not clustered), this *criterion* is tested by Independence tests (Christoffersen, 1998). Both of these properties (violation frequency and independence) are jointly tested thanks to Conditional Coverage test (Christoffersen, 1998). These three traditional tests have been recently improved using the GMM framework (Candelon *et al.*, 2011), subsampling bootstrap techniques (Escanciano and Olmo, 2010) and a constrained randomization surrogate data technique (Schreiber, 1998). A surrogate data method, explicitly keeping the time patterns for the data along a process of constrained randomization, which is basically based on a re-shuffling of the original data. This procedure thus allows us to generate new samples reproducing some of the time-series characteristics of the original. In the case of returns, the series are known to exhibit both short-term autocorrelation, regime shifts and possibly long memory features.

The amplitude of returns below the Quantile estimations can also be considered as a useful evaluation *criterion*, it can be evaluated thanks to Exception Magnitude test (Berkowitz, 2001). In addition, if a Quantile model is correct, the past Quantile estimations should not allow us to forecast future VaR violations, this property can be tested *via* the Dynamic Quantile test (Engle and Manganelli, 2004). Under specific assumptions the correct properties of returns exceeding the Quantile estimation can be tested using Saddlepoint Technique (Wong, 2008). We detail hereafter this traditional and recent test of extremal risk measures.

For most of these tests, the exception indicator variable associated to the *ex post* observation of a α-VaR violation at the time t, denoted $I_t(\alpha)$, is defined such as:

$$I_t(\alpha) = \begin{cases} 1 & \text{if } r_t < -VaR_{\alpha,t-1} \\ 0 & \text{otherwise,} \end{cases} \tag{29}$$

where r_t is the return at time t and $VaR_{\alpha,t-1}$ is the one-period VaR computed for the α-Quantile at time $t-1$.

The problem of VaR validity can therefore be tested, knowing the violation sequence of $I_t(\alpha)$, for $t = [1, \ldots, T]$, for a given α-Quantile.

4.1 Unconditional and Conditional Exception Tests

The most frequently backtests of VaR models, based on the exception indicator, are the Unconditional and Conditional tests. These VaR tests cover Unconditional Coverage Test (Kupiec, 1995), the Independence test and the Conditional Coverage test (Christoffersen, 1998). The Conditional Coverage test combines the Unconditional Coverage test and the Independence test.

An Unconditional Coverage Test (Kupiec, 1995)

The Unconditional Coverage test of VaR estimates is a count of violations over the entire period (an exception occurs when the *ex post* return is below the opposite of the *ex ante* VaR): for a percentage of exceptions (*i.e.* coverage rate) of α-percent, the *ex ante* VaR is correct if the observed violation frequency is equal to α-percent.

The Unconditional Coverage test of the exceptions (*i.e.* $E[I_t(\alpha)] = \alpha$) can be determined by the following test statistic (Kupiec, 1995; Escanciano and Olmo, 2010), with the previous notations:

$$K = T^{-1/2} \sum_{t=1}^{T} \left[I_t(\alpha) - \alpha \right], \tag{30}$$

This statistic is asymptotically normal $N(0, \alpha(1-\alpha))$ under the independence hypothesis.

Kupiec (1995) and Christoffersen (1998) show that if the VaR model is well specified then the exceptions occurred, can also be modelled as independent draws from a binomial distribution with a probability of occurrence equal to α-percent. The Likelihood Ratio statistic is therefore:

$$LR_{uc} = 2\{\log[\widehat{\alpha}^{N_I}(1-\widehat{\alpha}^{T-N_I})] - \log[\alpha^{N_I}(1-\alpha^{T-N_I})]\}, \tag{31}$$

where T is the number of observations, $N_I = T \times E\left[I_t(\alpha)\right]$ is the number of exceptions and $\widehat{\alpha} = N_I/T$ is the unconditional coverage.

With the hypothesis of $\widehat{\alpha} = \alpha$, the Likelihood Ratio LR_{uc} has an asymptotic distribution such as:

$$LR_{uc} \sim> \chi^2(1), \tag{32}$$

where $\chi^2(n)$ stands for the Chi-square distribution with n degree of freedom.

Escanciano and Olmo (2010) alternatively approximate the critical values of the Unconditional test by using subsampling methodology. They show that the Unconditional VaR backtest is affected by a model misspecification. Thus, they propose to use robust subsampling techniques to approximate the true sampling distribution of this test. They have shown that although the risk estimation can be diversified by choosing a large in-sample size relative to out-of-sample and the risk associated to the model cannot be diversified using subsampling.

The critical values of the test is computed with the assistance of the subsampling methodology. Resampling methods from *iid* sequences have been used extensively in the literature on Quantile regression models (*e.g.* Hahn, 1995; Horowitz, 1998; He and Hu, 2002). The subsampling methodology to approximate the critical values of tests is based on K. Henceforth, we use $G_K(x)$ denotes the Cumulative Distribution Function of the test statistic given by (with the previous notations):

$$G_K(x) = Prob(K \leq x), \tag{33}$$

for any $x \in \mathrm{IR}$.

Let $K_{b,t} = K(r_t, r_{t+1}, \cdots, r_{t+b-1})$, with $t = [1, 2, \cdots, T - b + 1]$, be the test statistic computed with the subsample $(r_t, r_{t+1}, \cdots, r_{t+b-1})$ of size b. Hence, Escanciano and Olmo (2010) approximate the sampling distribution $G_K(x)$ using the distribution of the values of $K_{b,t}$ computed over the $(T-b+1)$ different consecutive subsamples of size b. This subsampling distribution is defined by (with the previous notations):

$$G_{K_b}(x) = (T - b + 1)^{-1} \sum_{t=1}^{T-b+1} 1\mathrm{I}_{\{K_{b,t} \leq x\}}. \tag{34}$$

The $(1 - \tau) - th$ sample Quantile of $G_{K_b}(x)$, is given by:

$$c_{K_b, 1-\tau} = \underset{x \in \mathrm{IR}}{Inf} \left\{ x | \, G_{K_b}(x) \geq 1 - \tau \right\}. \tag{35}$$

We also complement the simple bootstrap results of Escanciano and Olmo (2010) for the test using alternative resampling techniques (surrogate method, see Schreiber, 1998) aiming to keep some of the return and volatility dependence structure. The surrogate method, on *contrario* to the bootstrap method, explicitly keeps the time-patterns for the returns along

a process of constrained randomization, which is basically based on a re-shuffling of the original return data. The artificial series are generated by reproducing some of the main time-series characteristics of the original one. We first randomly re-order, several times, the original series of returns, in order to delete all potential time-structure dependences, while keeping intact the unconditional distribution of the series and the VaR level. This means that compared to the bootstrap method for which data can be drawn several times, each data is only used once with the surrogate method. Secondly, on the new disorganized artificial series, we implement random pairwise permutations until the characteristics of the new series are similar to the original one (short-term autoregressivity, long memory, structural breaks,...).

More precisely, we apply three tests in order to compare their characteristics to the ones of the original series. A short-term autoregressivity test (measuring by the first autocorrelation coefficient of returns and squared returns, see Ljung and Box, 1978 and McLeod and Li, 1983), a long memory one (measuring by the scaling exponent obtained by the Detrended Fluctuation Analysis, see Kantelhardt *et al.*, 2001) and one for breaks (see Inclan and Tiao, 1994). The final surrogate series dataset consists only of series which share the same first autocorrelation coefficients for returns and squared returns (+/-10%), the same scaling exponent (+/-10%) and the same number of detected breaks.

An Independence Test (Christoffersen, 1998)

The independence hypothesis is associated to the idea that if the VaR model is correct then violations associated to VaR forecasting should be independently distributed, it is also called independence of exceptions hypothesis. Actually, this test is to avoid clusters of VaR violations. Thus, past VaR violations should not allow us to forecast future VaR violations.

This test detects the serial independence in the VaR forecast model. VaR violations observed at two different dates for the same coverage rate must be independently distributed. Christoffersen (1998) suggests that the process of $I_t(\alpha)$ violations is modelled with a Markov chain with the following transition probabilities matrix:

$$\Pi = \begin{pmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{pmatrix}, \tag{36}$$

where $\pi_{ij} = Prob\left[I_t(\alpha) = j | I_{t-1}(\alpha) = i\right]$, with $j = \{0, 1\}$ and $i = \{0, 1\}$.

The Likelihood Ratio for the test is:

$$LR_{ind} = 2[\log L(\pi_{01}, \pi_{11}) - \log L(\pi, \pi)], \tag{37}$$

where:

$$L(\pi_{01}, \pi_{11}) = (1 - \pi_{01})^{T_{00}} \, \pi_{01}^{T_{01}} (1 - \ \pi_{11})^{T_{10}} \, \pi_{11}^{T_{11}},$$

and:

$$L(\pi, \pi) = (1 - \pi)^{T_{00}+T_{10}} \, \pi^{T_{01}+T_{11}},$$

with T_{ij} the number of observations in the state j for the current period and at state i for the previous period, $\pi_{01} = T_{01}/(T_{00} + T_{01})$, $\pi_{11} = T_{11}/(T_{10} + T_{11})$ and $\pi = (T_{01} + T_{11})/T$.

The function $L(\pi_{01}, \pi_{11})$ is thus the likelihood under the hypothesis of the first-order Markov dependence, and $L(\pi, \pi)$ is the likelihood under the hypothesis of independence ($\pi_{01} = \pi_{11} = \pi$).
As shown in Christoffersen (1998), LR_{ind} has an asymptotic distribution such as (with the previous notations):

$$LR_{ind} \sim> \chi^2(1). \tag{38}$$

Since for Bernoulli random variables serial independence is equivalent to serial uncorrelation, it is also possible to build a test based on the autocovariances (Candelon *et al.*, 2011; Escanciano and Olmo, 2010; Hurlin and Tokpavi, 2007 and 2008). The proper marginal test of independence is based on (with the previous notations):

$$\zeta_j = (T - j)^{-1/2} \sum_{t=j+1}^{T} \{I_t(\alpha) - E_T[I_t(\alpha)]\} \{I_{t-j}(\alpha) - E_T[I_{t-j}(\alpha)]\}, \tag{39}$$

with:

$$E_T[I_t(\alpha)] = (T - j)^{-1} \left\{ \sum_{t=j+1}^{T} I_t(\alpha) \right\}.$$

If $G_\zeta(x)$ denotes the Cumulative Distribution Function of the test statistic, then (with the previous notations):

$$G_\zeta(x) = Prob(\zeta \leq x), \tag{40}$$

for any $x \in$ IR.

If $\zeta_{b,t} = \zeta(r_t, r_{t+1}, \cdots, r_{t+b-1})$, $t = [1, 2, \cdots, T-b+1]$, be the test statistic computed with the subsample $(r_t, r_{t+1}, \cdots, r_{t+b-1})$ of size b, the subsampling distribution (based on bootstrap or surrogate method) is defined by (with the previous notations):

$$G_{\zeta_b}(x) = (T-b+1)^{-1} \sum_{t=1}^{T-b+1} 1\mathrm{I}_{\{\zeta_{b,t} \leq x\}}, \tag{41}$$

and the $(1-\tau)$-*th* sample Quantile of $G_{\zeta_b}(x)$, $c_{\zeta_b,1-\tau}$, is given by:

$$c_{\zeta_b,1-\tau} = \underset{x \in \mathrm{IR}}{Inf} \left\{ x | \, G_{\zeta_b}(x) \geq 1-\tau \right\}. \tag{42}$$

A Conditional Coverage Test (Christoffersen, 1998)

The Conditional Coverage test, used hereafter, detects at the same time the Independence and Unconditional Coverage tests. This test detects at the same time the serial independence in the VaR forecast model and if the percentage of VaR forecasting violations can be judged different of α-percent. This test of conditional coverage is proposed by Christoffersen (1998).

In this case, if the VaR estimates have a correct conditional coverage, the exception variable must exhibit both correct unconditional coverage and serial independence. The Conditional Coverage test (LR_{cc}) is therefore a joint test of these properties and the relevant statistic is (with the previous notations):

$$LR_{cc} = LR_{uc} + LR_{ind}. \tag{43}$$

As shown by Christoffersen (1998), LR_{cc} has an asymptotic *i.i.d.* distribution such as (with the previous notations):

$$LR_{cc} \sim> \chi^2(2). \tag{44}$$

4.2 An Exception Magnitude Test (Berkowitz, 2001)

The main idea of this test is the magnitude of VaR forecasting violations (exceptions) should be of primary interest to the various users of VaR models (Berkowitz, 2001; Ferreira and Lopez, 2005). Actually, it is useful for investor to investigate the amplitude of returns below the VaR.

In the same way as the LR_{dist} test, Berkowitz (2001) transforms the empirical series into $\breve{r}_t$ series. The $\breve{r}_t$ values are then compared to the normal random variables with the desired coverage level of the VaR estimates.

If the VaR model associated to the probability α generating the empirical Quantiles is correct, the $R_t^G(\alpha)$ series, defined by equation (45), should be identically distributed with the unconditional mean and the standard deviation (0,1) such as:

$$R_t^G(\alpha) = \begin{cases} \breve{r}_t \text{ if } \breve{r}_t < \Phi^{-1}(\alpha) \\ 0 \quad \text{otherwise,} \end{cases} \tag{45}$$

where $\Phi(.)$ is the standard normal Cumulative Distribution Function.
Finally, the corresponding test statistic is (with the previous notations):

$$LR_{mag} = 2[L_{mag}(\mu, \sigma) - L_{mag}(0, 1)], \tag{46}$$

where:

$$\begin{aligned} L_{mag}(\mu, \sigma) \quad = \quad & \textstyle\sum_{t=1}^{T} \log\{1 - \Phi[(\Phi^{-1}(\alpha) - \mu)\sigma^{-1}]\}\, 1\mathrm{I}_{\{R_t^G(\alpha)=0\}} \\ & + \textstyle\sum_{t=1}^{T} \{-(1/2)\log(2\pi\sigma^2) - (R_t^G(\alpha) - \mu)^2(2\sigma^2)^{-1} \\ & - \log[\Phi((\Phi^{-1}(\alpha) - \mu)\sigma^{-1}]\} 1_{\{R_t^G(\alpha)\neq 0\}}, \end{aligned}$$

with $1\mathrm{I}_{\{.\}}$ the indicator function.

As shown in Berkowitz (2001), LR_{mag} is thus asymptotically distributed such as (with the previous notations):

$$LR_{mag} \sim> \chi^2(2). \tag{47}$$

4.3 A Multivariate Portmanteau Test (Hurlin and Tokpavi, 2007)

Hurlin and Tokpavi (2007) propose a multivariate portmanteau test of absence of autocorrelation of violations. The main idea consists in testing the validity of the VaR determination model for a finite sample considering several percentage of violations (coverage rates). Thus, this test exploits a larger information set than those generally used in tests of the Event Probability Forecast Evaluation category, without the drawbacks of tests based on the assessment of return density.

For this test, suppose $h_t(\alpha)$ is associated to the value $(1-\alpha)$ in case of violations and α otherwise:

$$h_t(\alpha) = \begin{cases} 1-\alpha \text{ if } r_t < -VaR_{\alpha,t-1} \\ -\alpha \quad \text{otherwise,} \end{cases} \tag{48}$$

which can be written such as (with the previous notations):

$$h_t(\alpha) = I_t(\alpha) - \alpha. \tag{49}$$

Let $\Pi = \{\pi_1, \ldots, \pi_m\} \subset [0,1]$ be a discrete set of m different coverage rates. Let $H_t = [h_t(\pi_1), h_t(\pi_2), \ldots, h_t(\pi_m)]'$ the vector of dimension $(m \times 1)$ grouping the violation sequences associated to these m coverage rates at time t. The conditional efficiency hypothesis for the vectorial process is then:

$$Cov(H_t, H_{t-l}) = E[H_t \; H'_{t-l}]. \tag{50}$$

The test consists in the joint nullity of the autocorrelations order 1 to T for the vector H_t. Formally, we have:

$$H_0 : E[h_t(\pi_i) h_{t-l}(\pi_j)] = 0, \tag{51}$$

for the lag l, with $l = [1, \ldots, T]$ and the coverage rates $(\pi_i, \pi_j) \in \Pi$.

For the test, Hurlin and Tokpavi (2007) use a multivariate Portmanteau statistics (Hosking, 1980). If $\mathbf{\Omega}_l$ is the matrix of empirical covariances associated to vector H_t such as (with the previous notations):

$$\mathbf{\Omega}_l = \sum_{t=l+1}^{T} H_t \; H'_{t-l}, \tag{52}$$

for the set of lag $l = [1, \ldots, T]$.

The statistic for the test, $Q_m(T)$, is given by:

$$Q_m(T) = T^2 \sum_{l=1}^{T} (T-l)^{-1} Diag(\mathbf{\Omega}'_l \mathbf{\Omega}_0^{-1} \mathbf{\Omega}_l \mathbf{\Omega}_0^{-1})' \mathbf{1}. \tag{53}$$

As shown in Hurlin and Tokpavi (2007), under the null hypothesis of absence of autocorrelation in the vector H_t, $Q_m(T)$ is thus asymptotically distributed such as (with the previous notations):

$$Q_m(T) \sim> \chi^2(Tm^2). \tag{54}$$

4.4 Generalized Method of Moment Duration-based Tests (Candelon *et al.*, 2011)

The Generalized Method of Moment (GMM) Duration-based tests, proposed by Candelon *et al.* (2011), analyze previous VaR forecast tests (Unconditional, Independence and Conditional Coverage Test) as an expression of simple moment conditions, within the well-known GMM framework, these tests extend the framework proposed by Bontemps and Meddahi (2005), and Bontemps (2008).

The GMM Duration Unconditional, Independence and Conditional Coverage tests objective are to choose appropriated moments to respectively verify whether: the VaR observed violation frequency is equal to α-percent, violations associated to the VaR forecasting are independently distributed (testing the hypothesis that duration follows a geometric distribution) and whether the "Exception" and Independence *criteria* are both respected.

The Generalized Method of Moment (GMM) Duration-based Tests proposed by Candelon *et al.* (2011) consist in using a GMM framework in order to test if durations of VaR violations are geometrically distributed. This test analyses previous VaR forecast tests (Unconditional, Independence and Conditional Coverage Test) as an expression of simple moment conditions, within the well-known GMM framework, it extends the framework proposed by Bontemps and Meddahi (2005) and Bontemps (2008).
The objective is to choose the appropriated moments to know whether their empirical expectations are close to 0 or not.

The Generalized Method of Moment (GMM) Duration-based Tests proposed by Candelon *et al.* (2011) analyses VaR forecast tests as an expression of simple moment conditions, which can be tested within the well-known GMM framework. This test extends the framework proposed by Bontemps and Meddahi (2005) and Bontemps (2008), and consists in using a GMM framework in order to test if durations of VaR violations are geometrically distributed. The objective is to choose the appropriated moments to know whether their empirical expectations are close to 0 or not.

A GMM Duration-based Conditional Coverage Test (Candelon *et al.*, 2011)

This test analyzes Conditional Coverage VaR forecast test as an expression of simple moment conditions, within the well-known GMM framework. The objective is to choose the

appropriate moments to know whether their empirical expectations are close to 0 or not.

For a given model and a fixed coverage rate α, we consider a sequence of N_I durations, denoted $d_{\alpha,i}$ for $i = [1, \cdots, N_I]$, observed between two successive violations associated to the $\alpha\%$ VaR forecasts:

$$d^*_{\alpha,i} = t^\alpha_i - t^\alpha_{i-1}, \tag{55}$$

with t^α_i is the date of the ith violation for the level α.

Under the conditional coverage assumption, the duration $d^*_{\alpha,i}$, for $i = [1, \cdots, N_I]$, are *i.i.d.* and has a geometrical distribution with a success probability equals to the coverage rate α.

Candelon *et al.* (2011) define a sequence of moment conditions. $M(d^*_{\alpha,i})$ is a $(p\times 1)$ vector, whose p components are the orthonormal polynomials corresponding to the duration at the level α. p is the number of moment conditions ($p > 1$). These polynomials are a particular case of the Meixner orthonormal polynomials associated to a Pascal (negative Binomial) distribution. The components of $M(d^*_{\alpha,i})$ are denoted $M_j(d^*_{\alpha,i})$ such as, for $j = [1, \ldots, p]$:

$$\begin{aligned} M_{j+1}(d^*_{\alpha,i}) = \; & \left[(1-\alpha)(2j+1) + \alpha(j - d^*_{\alpha,i} + 1)\right] \left[(j+1)(1-\alpha)^{1/2}\right]^{-1} \\ & \times M_j(d^*_{\alpha,i}) - j(j+1)^{-1} M_{j-1}(d^*_{\alpha,i}), \end{aligned} \tag{56}$$

with $M_1(d^*_{\alpha,i}) = 0$ and $M_0(d^*_{\alpha,i}) = 1$.

The null hypothesis of Conditional Coverage (CC) test can be expressed as:

$$\begin{cases} H_0 : E[M_j(d^*_{\alpha,i})] = 0 \\ H_1 : E[M_j(d^*_{\alpha,i})] \neq 0. \end{cases} \tag{57}$$

As shown in Candelon *et al.* (2011), $M_j(d^\alpha_i)$ is thus asymptotically distributed such as (with the previous notations):

$$\left(N^{-1/2} \sum_{i=1}^{N} M_j(d^*_{\alpha,i}) \right)^2 \xrightarrow[N\to\infty]{d} \chi^2(1), \tag{58}$$

for $j = [1, \ldots, p]$.

In an *i.i.d.* context, these moments are asymptotically independent with a unit variance and the Conditional Coverage (CC) test statistic is (with the previous notations):

$$J_{CC}(p) = \left(N^{-1/2} \sum_{i=1}^{N} M(d^*_{\alpha,i}) \right)' \left(N^{-1/2} \sum_{i=1}^{N} M(d^*_{\alpha,i}) \right) \xrightarrow[N\to\infty]{d} \chi^2(p), \tag{59}$$

with $p \in \mathrm{IN}$, the number of orthonormal polynomials used as moment conditions.

A GMM Duration-based Unconditional Coverage Test (Candelon *et al.*, 2011)

This test analyzes Unconditional Coverage VaR forecast test as an expression of simple moment conditions, within the well-known GMM framework. The objective is to choose appropriate moments to know whether their empirical expectations are close to 0 or not.

A GMM Duration-based Unconditional Coverage test is obtained as a special case of the conditional test by considering only the first orthonormal polynomial. This means, when $M(d^*_{\alpha,i}) = M_1(d^*_{\alpha,i})$. The statistic for Unconditional Coverage (UC) test, denoted J_{UC}, is equivalent to $J_{CC}(p)$ with $p = 1$ (with the previous notations):

$$J_{UC} = \left(N^{-1/2} \sum_{i=1}^{N} M_1(d^*_{\alpha,i}) \right) \xrightarrow[N\to\infty]{d} \chi^2(1). \tag{60}$$

A GMM Duration-based Independence Test (Candelon *et al.*, 2011)

Candelon *et al.* (2011) propose a separate test for the independence hypothesis. It consists in testing the hypothesis of a geometric distribution (implying the absence of dependence) with a success probability equal to π; where parameter π can be either fixed *a priori*, or estimated. It is not necessarily equal to the coverage rate α.

The statistic for Independence (Ind) test, denoted J_{Ind}, is defined such as (with the previous notations):

$$J_{Ind}(p) = \left(N^{-1/2} \sum_{i=1}^{N} M(d_{\pi,i}) \right)' \left(N^{-1/2} \sum_{i=1}^{N} M(d_{\pi,i}) \right) \xrightarrow[N\to\infty]{d} \chi^2(p), \tag{61}$$

where $M(d_{\pi,i})$ denotes a $(p \times 1)$ vector whose components are the orthonormal polynomials $M_j(d_{\pi,i})$ for $j = [1, \ldots, p]$, evaluated for a success probability equals to π.

4.5 A Dynamic Quantile Test (Engle and Manganelli, 2004)

The Dynamic Quantile (DQ) test is proposed by Engle and Manganelli (2004). This test is associated to the idea that if the VaR model is correct then past VaR violations should not allow us to forecast future VaR violations. Indeed, this model should produce a sequence of unbiased and uncorrelated indicator variables $h_t(\alpha)$. On the contrary to most of previous

tests, this test is valid using small VaR sample.

Consider the following linear regression model:

$$h_t(\alpha) = \delta_0 + \sum_{l=1}^{L} \delta_l h_{t-l}(\alpha) + \delta_{L+1} VaR_{\alpha,t-1} + \epsilon_t, \tag{62}$$

where $h_t(\alpha)$ is associated to the value $(1-\alpha)$ in case of violations and $(-\alpha)$ otherwise and the δ_l, with $l = [0, \ldots, L+1]$, are real coefficients.

Under the null hypothesis, all coefficients in the regression (62) of the exception indicator variables on its past values and on current VaR estimate, as well as the intercepts are zero. The null hypothesis test of conditional efficiency corresponds to testing the joint nullity of coefficients δ_l.

4.6 A Saddlepoint Technique for an Expected Shortfall Test (Wong, 2008)

The main intuition of the Saddlepoint Technique test for ES (Wong, 2008) is to approximate the tail of the density of the return. So that, under fairly general conditions and under a standard normal null hypothesis, a full expansion can locally approximate the density of the Expected Shortfall (Daniels, 1954), which can be compared to the observed value.

Actually, Edgeworth expansions are frequently used to approximate distributions that lack a convenient closed-form solution but for which moments are available. It is widely recognized that this classic technique works well in the center of a distribution, but can perform very badly in the tails. Indeed, it often produces negative values for densities in the tails. Saddlepoint expansion can be understood as a refinement of Edgeworth expansion on the tails.

Wong (2008) proposes the Saddlepoint Technique (Daniels, 1954 and 1987; Lugannani and Rice, 1980) to derive the small sample asymptotic distribution for the sample Expected Shortfall (ES) under a standard normal null hypothesis.

Let r_t be the portfolio return, suppose that r_t has a standard normal distribution with a Cumulative Distribution Function (CDF) and a Probability Density Function (PDF) denoted respectively by $\Phi(.)$ and $\phi(.)$. Define a random variable R_t such that for a given portfolio

return r:

$$\begin{aligned} Prob(R_t \leq r) &= Prob(r_t \leq r | r_t \leq -VaR_{\alpha,t}^{\Phi(r_t)}), \\ &= Prob(r_t \leq r)/Prob(r_t \leq -VaR_{\alpha,t}^{\Phi(r_t)}), \end{aligned} \tag{63}$$

where $r \leq -VaR_{\alpha,t}^{\Phi(r_t)}$, $VaR_{\alpha,t}^{\Phi(r_t)} = -\Phi^{-1}(\alpha)$, and $VaR^{\Phi(r_t)}$ being VaR level specifically assuming a Gaussian distribution of the returns.

Thus R_t refers to the portfolio return conditional on $r_t \leq -VaR_{\alpha,t}^{\Phi(r_t)}$ and the theoretical Expected Shortfall of r_t at confidence level $1-\alpha$ is given by $-E(R_t)$. Given a sample $\{R_1, R_2, \ldots, R_T\}$ of exceedances below $-VaR_{\alpha,t}^{\Phi(r_t)}$, the estimated Expected Shortfall can be written such as:

$$ES_{\alpha,T} = -T^{-1} \sum_{t=1}^{T} R_t. \tag{64}$$

The PDF of R_t is given by $f(R_t) = \alpha^{-1}\phi(r_t)$ and the corresponding CDF by $F(R_t) = \alpha^{-1}\Phi(r_t)$, where $R_t \in]-\infty, -VaR_{\alpha,t}^{\Phi(r_t)}]$. Therefore, the Moment Generating Function of R_t is then given by:

$$M(R_t) = \alpha^{-1} exp(R_t^2/2) \times \Phi(-VaR_{\alpha,t} - R_t), \tag{65}$$

and its derivatives with respect to R_t are given by:

$$\begin{cases} M^{(1)}(R_t) = R_t \times M(R_t) - exp(-VaR_{\alpha,t} \times R_t) \times \alpha^{-1}\phi(-VaR_{\alpha,t}) \\ M^{(2)}(R_t) = R_t \times M^{(1)}(R_t) + M(R_t) - exp(-VaR_{\alpha,t} \times R_t) \times (-VaR_{\alpha,t}) \\ \qquad\qquad \times \alpha^{-1}\phi(-VaR_{\alpha,t}) \\ \qquad\qquad \vdots \\ M^{(n)}(R_t) = R_t \times M^{(n-1)}(R_t) + (n-1)M^{(n-2)}(R_t) - exp(-VaR_{\alpha,t} \times R_t) \\ \qquad\qquad \times (-VaR_{\alpha,t})^{n-1} \times \alpha^{-1}\phi(-VaR_{\alpha,t}), \end{cases} \tag{66}$$

where $M^{(n)}$ is the order n Moment Generating Function derivative.

In order to define the null and the alternative hypothesis of the test, we need to define the mean and variance of R_t. Posing $R_t = 0$ in $M^{(1)}(R_t)$ and $M^{(2)}(R_t)$, it will respectively give $E(R_t)$ and $E(R_t^2)$. The variance is then given by: $M^{(2)}(0) - [M^{(1)}(0)]^2$. Assuming a Gaussian distribution, the mean of R_t, which is the opposite of $ES_{\alpha,T}$, is then given by (Wong, 2008):

$$\mu_t^{\Phi(R_t)} = -\alpha^{-1}\phi(-VaR_{\alpha,T}^{\Phi(r_t)}), \tag{67}$$

and can thus be expressed such as:

$$\mu_t^{\Phi(R_t)} = -\left(\alpha\sqrt{2\pi}\right)^{-1} exp[-(VaR_{\alpha,T}^{\Phi(r_t)})^2/2]. \tag{68}$$

Its variance is:

$$\begin{aligned}\sigma^2_{R_t} &= 1 + \alpha^{-1} VaR^{\Phi(r_t)}_{\alpha,T} \phi(-VaR^{\Phi(r_t)}_{\alpha,T}) - (\mu^{\Phi(R_t)}_t)^2, \\ &= 1 + \left(\alpha\sqrt{2\pi}\right)^{-1} VaR^{\Phi(r_t)}_{\alpha,T} exp[-(VaR^{\Phi(r_t)}_{\alpha,T})^2/2] - (\mu^{\Phi(R_t)}_t)^2.\end{aligned} \tag{69}$$

Following the central limit theorem, $\sqrt{T}\sigma^{-1}_{R_t}(\widehat{\mu}^{G(R_t)}_t - \mu^{\Phi(R_t)}_t)$ is asymptotically standard normal, with $\widehat{\mu}^{G(R_t)}_t$ the mean of R_t. This asymptotic result is hardly valid for sample sizes one would encounter in practice. For this reason, Wong (2008) uses the Lugannani and Rice (1980) *formula* derived by the saddlepoint technique in order to provide an accurate method to compute under the null hypothesis the required p-value for the backtest of ES. The null and alternative hypotheses can be thus written as:

$$\begin{cases} H_0 : ES_{\alpha,T} = -\alpha^{-1}\phi(-VaR^{\Phi(r_t)}_{\alpha,T}) \\ H_1 : ES_{\alpha,T} > -\alpha^{-1}\phi(-VaR^{\Phi(r_t)}_{\alpha,T}). \end{cases} \tag{70}$$

The p-value of the previous hypothesis test is given by the Lugannani and Rice (1980) *formula* as:

$$p\text{-value} = Prob(\widehat{\mu}^{G(R_t)}_t \leq \widehat{\mu}_t), \tag{71}$$

where $\widehat{\mu}_t$ is the mean of observed return exceedances below the $-VaR_{\alpha,t}$.

The p-value is determined by the tail probability of being less than or equal to the observed sample mean of r_t, denoted $\widehat{\mu}_t$ given by:

$$Prob(\widehat{\mu}^{G(R_t)}_t \leq \widehat{\mu}_t) = \begin{cases} \Phi(\varsigma) - \phi(\varsigma)\left[\frac{1}{\eta} - \frac{1}{\varsigma} + \mathbf{O}(T^{-3/2})\right] & \text{for } \widehat{\mu}_t < -VaR_{\alpha,t} \\ 1 & \text{for } \widehat{\mu}_t \geq -VaR_{\alpha,t}, \end{cases} \tag{72}$$

with $\eta = \varpi\sqrt{TK^{(2)}(\varpi)}$, $\varsigma = \varpi|\varpi|^{-1}\sqrt{2T[\varpi\widehat{\mu}_t - K(\varpi)]}$, $\mathbf{O}(.)$ describes the limiting order function and $K(\varpi)$ the Cumulant Generating Function of ϖ defined by (with the previous notations):

$$K(\varpi) = ln[M(\varpi)], \tag{73}$$

where ϖ is the estimated saddlepoint with the equality $K^{(1)}(\varpi) = \widehat{\mu}_t$ being satisfied and:

$$K^{(1)}(\varpi) = M^{(1)}(\varpi)/M(\varpi). \tag{74}$$

5 Gauging the DARE Risk Measure Accuracy

For investigating the efficiency of the DARE approach for VaR and Expected Shortfall computations, we apply the method to the French Stock Market data and compare it to established methods. We use daily prices from 9th July 1987 until 18th March 2009 of the CAC40 Index. This period delivers 5,659 daily returns. We use a moving window of four years (1,044 daily returns) to re-estimate dynamically parameters for the various methods. Forecasted VaR and ES were computed for each method for the final 4,615 days (about 18 years). This comparison considers daily estimation of the 95% and 99% conditional VaR and ES. These Quantiles were chosen because they are commonly considered and controlled by financial firms and regulators. In the first part of this analysis we use an naive equally weighted DARE Approach, giving the same weights to CARE models (denoted EW DARE).

Then we illustrate the use of optimized weighting function according to the DARE approach. The optimization is based on the forecast combination of the Mallows Model Averaging method (Mallows, 1973; Hansen, 2008). The goal is to obtain the set of weights that minimizes the in-sample Mean-Squared Error (MSE) over the set of feasible forecast combinations. We thus select forecast weights by minimizing a Mallows *criterion*, which is an asymptotically unbiased estimate of the MSE.
Formally, let w_k be a weight assigned to the kth forecast, with $k = [1, 2, \ldots, K]$; K approximating models and associated forecasts. Define the vectors $\mathbf{W} = (w_1, w_2, \ldots, w_K)$ of weights and $\mathbf{VaR}_{\alpha,t} = (VaR^{(1)}_{\alpha,t}, VaR^{(2)}_{\alpha,t}, \ldots, VaR^{(K)}_{\alpha,t})$ of α-VaR at time t. The combination forecast of the coverage rate ($\alpha_t^{\mathbf{W}}$) is (with the previous notations):

$$\hat{\alpha}^{\mathbf{W}}_{t+1} = E[\hat{I}^{\mathbf{W}}_{t+1}(\alpha)], \tag{75}$$

where the exception indicator variable associated to the *ex post* observation of a forecast α-VaR violation at the time $t+1$ is defined such as:

$$I^{\mathbf{W}}_{t+1}(\alpha) = \begin{cases} 1 & \text{if } r_{t+1} < -\mathbf{W}' \, \mathbf{VaR}_{\alpha,t} \\ 0 & \text{otherwise.} \end{cases} \tag{76}$$

Hansen (2008) proposes to select the weights by minimizing the Penalized Sum of Squared Residuals (PSSR, known as the Mallows *criterion*), such as:

$$C_t(\mathbf{W}) = (\alpha - \hat{\alpha}^{\mathbf{W}}_t)^2 + 2K \sum_{k=1}^{K} w_k s_k^2, \tag{77}$$

where $s_k^2 = (t-K)^{-1}(\alpha - \hat{\alpha}_t^{\mathbf{W}})^2$ is an estimate of error variance for a model k.

As described in Timmermann (2006), it may be prudent to minimize relation (77) subject to the convexity constraints (such as: $0 \leq w_k \leq 1$) and additivity constraint (*i.e.*: $\sum_{k=1}^{K} w_k = 1$). The definition of the Mallows weight vector is then:

$$\hat{\mathbf{W}}^* = \underset{w_k \in [0;1],\, \sum_{k=1}^{K} w_k = 1}{ArgMin} \{C_t(\mathbf{W})\}. \tag{78}$$

This optimization procedure will be used hereafter for the computation of the so-called "Hansen-Mallows Optimized-weight DARE-VaR for Value-at-Risk" and "Hansen-Mallows Optimized-weight DARE-ES for Expected Shortfall".

Although there are many VaR estimation methods in use, we restrict our comparison in the following to commonly used benchmark methods. We first consider the most widely used non-parametric method: the historical approach. It requires no distributional assumptions and estimates the VaR as the Quantile of the empirical distribution of historical returns from a moving window.

Parametric approaches, which are then estimated, assume a particular shape of return distribution as Gaussian or t-Student distribution. We respectively consider the Normal, the *RiskMetrics*, the GARCH(1,1) VaR and ES. All these parametric methods also assume a Gaussian distribution by using various conditional volatility forecasts: respectively the empirical volatility, the *RiskMetrics* correlation method based on an Exponential Moving Average estimation, GARCH(1,1) model (the use of the (1,1) specification was based on the general popularity of this order for GARCH models). We also estimate VaR and ES with the NIG and the Student's-t distribution optimizing parameters by using maximum likelihood.

- Please insert Figures 1 and 2 somewhere here -

CAViaR models are presented as benchmark dynamic methods. They are estimated by using the Engle and Manganelli's procedure (2004) (see section 2). We finally compute the DARE approach by aggregating CARE models based on CAViaR models.
As mentioned in the previous section, the DARE approach can be used to estimate the VaR and ES. The relation between VaR and ES in the DARE approach is defined by the equation (25). To be sure to have a good estimation of ES using the DARE approach we have first

to test VaR based on it. We check first whether the VaR using the DARE approach is well specified (see table 1), then we check that the aggregation to compute the ES gives satisfying results (see table 2).

- Please insert Table 1 somewhere here -

Table 1 presents the main test results of VaR validation described in the literature (see the previous section for a detailed description of these tests). For each method and for two different levels of probability (95% and 99%) defining extreme risk measures, this table presents the probability of failing to reject the null hypothesis of a "good" VaR measure, according to the properties studied for each specific test.

For example, according to the "Exception Frequency" test (*i.e.* the Unconditional Coverage test by Kupiec, 1995): the Historical, Gaussian, NIG, *RiskMetrics* and CAViaR Asymmetric Slope models are not suitable for estimating the CAC 40 95% VaR at a 1% significance level. Following the same argument, the Historical, Gaussian, NIG, *RiskMetrics*, Student, Gaussian GARCH and CAViaR Asymmetric Slope models are not appropriate for estimating the CAC40 99% VaR. The Unconditional test performed, using the GMM duration-based method (Candelon *et al.*, 2011), leads to the same conclusion.

We notice that when naively combining all methods (Naive Mean of VaR), this *criterion* is then respected. Moreover, aggregating CAViaR methods based on CARE models in the Naive Equal-weight DARE approach, allows us to respect this test (the hypothesis of a "good" VaR cannot anymore be rejected).

We also observe that the Independence Violation *criterion* (*i.e.* the Independence test by Christoffersen, 1998) is not rejected whatever the VaR method we choose to estimate the 95% VaR. However, this *criterion* is rejected for most of the methods used to estimate the 99% VaR at a significance level of 1% (*i.e.* Historical, Gaussian, Student, NIG, *RiskMetrics* or CAViaR Adaptive). Only extreme non-expected returns seem to be dependent. Using the subsampling approximation technique of Escanciano and Olmo (2010), the independent *criterion* cannot be rejected on this data sample. The framework of the GMM duration based test (Candelon *et al.*, 2011), applied to the independence *criterion*, seems much more powerful and leads to a more contrasted analysis: Historical, Gaussian, Student, NIG and

Adaptive CAViaR are then considered to be inappropriate for VaR estimation methods at a significance level of 1%.

The "Conditional/Unconditional" *criterion* (*i.e.* the Conditional Coverage test by Christoffersen, 1998) combines the "Independence Violation" and "Exception Frequency" *criteria*. According to this test, the results are similar to those of the "Exception Frequency" Test.

The "Dynamic Quantile" test (Engle and Manganelli, 2004) is associated to the idea that if the VaR model is correct then past VaR violations should not allow us to forecast future VaR violations. This test is not so restrictive and it rejects only the Historical, Gaussian and NIG VaR at 95% for a significance level of 1%. At the opposite, the "Exception Magnitude" test (Berkowitz, 2001), that investigates the amplitude of returns below the VaR, is very difficult to pass since it rejects most of the VaR estimation methods. VaR violations seem therefore to be particularly important in size on our data sample.

No benchmark method can be considered as a "good" measure for every test even if some measures are definitively not appropriate. Thus, according to these tests on this data sample, the "worst" methods are the Historical and Normal VaR and the "best" measures seem to be the Naive Equal-weight DARE approach for VaR, CAViaR Indirect GARCH(1,1) and the naive aggregation (Naive Mean of VaR).

The aggregation of VaR approaches seem to be a more robust approach than the classical VaR computations. Thus, naively mixing these different benchmark methods does not allow us to respect all *criteria*, but test results are more stable to the variation of the probability level associated to the VaR. Moreover, using the Naive Equal-weight DARE approach respects most of these tests.

In the proposed Naive Equal-weight DARE approach, the VaR seems to be well specified. We also have checked with the same data set that their aggregation to compute the conditional VaR gives satisfying results. The results are presented in table 2.

- Please insert Table 2 somewhere here -

These complementary tests corroborate the previous results. Table 2 allows us to illustrate that the naive aggregation of 99% ES delivers good and stable results according to

the *criteria* tested. Moreover, it indicates that the risk model is very important for the estimation of the ES associated to small probabilities (*i.e.* 99%). Actually, even a simple naive aggregation leads to diversify the risk model and improve the extreme risk measure estimation. The Naive Equal-weight DARE 99% ES gives the best results (according to the *criteria* tested) and it is the most stable estimation (the stability was checked of performing the same evaluation tests on several subsampling periods). The Naive Equal-weight DARE approach for ES seems to be particularly attractive for extreme probabilities associated to the ES.

- Please insert Table 3 and Table 4 somewhere here -

After having investigate the interest of the DARE approach using an equiprobable weights for the CARE ES and VaR models, we illustrate the use of this approach with a dynamically optimized weight function respecting the *criteria* exposed in section 3. In this illustration, the weights were *ex ante* determined by the minimization of spread between the observed and the theoretical level of risk (see Hansen, 2008). Then we have evaluated this DARE Approach (denoted Hansen DARE) thanks to the tests exposed above. This dynamic DARE approach is compared to the Naive Equal-weight DARE approach and the naive mean of benchmark models. This approach gives comparable results than the Naive Equal-weight DARE approach but seem better than the naive mean.

In conclusion, the DARE approach provides a unified framework to estimate extreme risk measures. The DARE VaR and ES respect most of the Quantile evaluation tests for the different levels of probability studied in this article. This approach can be used to compute extreme measures associated to usual levels of probability. But as an aggregation, it allows us to benefit from risk model diversification, which seems to be particularly efficient for the estimation of extreme Quantiles (*i.e.* 99%).

6 Conclusion

In this paper, we present a new modeling approach for Value-at-Risk and Expected Shortfall, called Dynamic AutoRegressive Expectiles (DARE). This approach is based on

the aggregation of existing extremal risk models and allows us to diversify the risk model associated to each model. Introducing conditional weighting functions to the risk models aggregated into the DARE Approach allows us to select dynamically the more appropriate Quantile models.

We show that the DARE methodology advantageously reveals an underlying structure in data that can be used for risk measure computations. We also show, with different VaR and ES specifications (Normal, Historical, *RiskMetrics*, t-Student, NIG, GARCH and CAViaR) and various backtests (Conditional and Unconditional, Distributional Forecast, Duration-based tests), that the DARE model is well adapted to the financial data. Indeed, the aggregation of VaR approaches is more robust than the classical VaR and ES computations.

To illustrate the DARE approach, we have, first, supposed that the weights associated to Quantile models are equally-weighted; we secondly show, with a grid of weights (defined dynamically), that it was possible to improve the DARE approach estimator. However, to exploit efficiently the aggregation of the Quantile models, it is necessary to define an optimization *criterion*, which is independent from evaluation procedures. Determining and comparing several weighting functions associated to the DARE approach can also be a natural extension of this work.

Finally, the DARE approach, aggregating of several VaR measures, could have some interest in further financial applications, such as stress test assessment or asset pricing.

References

Abdous, B. and B. Remillard (1995): "Relating Quantiles and Expectiles under Weighted-Symmetry", *Annals of the Institute of Statistical Mathematics*, 47, 371-384.

Acerbi, C. (2002): "Spectral Measures of Risk: A Coherent Representation of Subjective Risk Aversion", *Journal of Banking and Finance*, 26, 1505-1518.

Acerbi, C. and D. Tasche (2002): "On the Coherence of Expected Shortfall", *Journal of Banking and Finance*, 26, 1487-1503.

Artzner, P., F. Delbaen, J. Eber and D. Heath (1997): "Thinking Coherently", *Risk*, 10, 68-71.

Artzner, P., F. Delbaen, J. Eber and D. Heath (1999): "Coherent Measures of Risk", *Mathematical Finance*, 9, 203-228.

Bali, T. (2003): "An Extreme Value Approach to estimating Volatility and Value-at-Risk", *Journal of Business*, 76, 83-108.

Barndorff-Nielsen, O. (1998): "Process of Normal Inverse Gaussian Type", *Finance and Stochastics*, 2, 41-68.

Basle Committee on Banking Supervision (1996): *Amendment to the Capital Accord to Incorporate Market Risks*, Bank for International Settlements, 63 pages.

Beder, T. (1995): "VaR: Seductive but Dangerous", *Financial Analysts Journal*, 51, 12-24.

Bekiros, S. and D. Georgoutsos (2005): "Estimation of Value-at-Risk by Extreme Value and Conventional Methods: A Comparative Evaluation of their Predictive Performance", *Journal of International Financial Markets, Institutions and Money*, 13, 209-228.

Berkowitz, J. (2001): "Testing Density Forecasts with Applications to Risk Management", *Journal of Business and Economics Statistics*, 19, 465-474.

Berkowitz, J. and J. Brien (2002): "How Accurate are the Value-at-Risk Models at Commercial Banks?", *Journal of Finance*, 57, 1093-1111.

Berkowitz, J., P. Christoffersen and D. Pelletier (2011): "Evaluating Value-at-Risk Models with Desk-level Data", *Management Science*, 57(12), 2213-2227.

Bontemps, C. (2008): "Moment-based Tests for Discrete Distributions", *IDEI Working Paper - University of Toulouse*, 55 pages.

Bontemps, C. and N. Meddahi (2005): "Testing Normality: A GMM Approach", *Journal of Econometrics*, 124, 149-186.

Brownlees, C. and G. Gallo (2010): "Comparison of Volatility Measures: A Risk Management Perspective", *Journal of Financial Econometrics*, 8, 29-56.

Candelon, B., G. Colletaz, C. Hurlin and S. Tokpavi (2011): "Backtesting Value-at-Risk: A GMM Duration-based Test", *Journal of Financial Econometrics*, 9(2), 314-343.

Cheridito, P. and M. Stadje (2009): "Time-inconsistency of VaR and Time-consistent Alternatives", *Finance Research Letters*, 6, 40-46.

Christoffersen, P. (1998): "Evaluating Interval Forecasts", *International Economic Review*, 39, 841-862.

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

Cornish, E. and R. Fisher (1937): "Moments and Cumulants in the Specification of Distributions", *Review of the International Statistical Institute*, 5, 307-320.

Daniels, H. (1954): "Saddlepoint Approximations in Statistics", *The Annals of Mathematical Statistics*, 25, 631-650.

Daniels, H. (1987): "Tail Probability Approximations", *International Statistical Review*, 55, 37-48.

Danielsson, J. and C. de Vries (2000): "Value-at-Risk and Extreme Returns", *Annales d'Economie et de Statistique*, 60, 239-270.

Efron, B. (1991): "Regression Percentiles using Asymmetric Squared Error Loss", *Statistica Sinica*, 1, 93-125.

Embrechts, P., C. Klüppelberg and T. Mikosch (1997): *Modelling Extremal Events For Insurance and Finance*, Berlin, Springer-Verlag, 655 pages.

Engle, R. and S. Manganelli (1999): "Value-at-Risk Models in Finance", *ECB Working Paper 75*, 40 pages.

Engle, R. and S. Manganelli (2004): "CAViaR: Conditional AutoRegressive Value-at-Risk by Regression Quantile", *Journal of Business and Economic Statistics*, 22, 367-381.

Engle, R. and J. Rangel (2008): "The Spline-GARCH Model for Low-frequency Volatility and its Global Macroeconomic Causes", *Review of Financial Studies*, 21, 1187-1222.

Escanciano, J. and J. Olmo (2010): "Backtesting Parametric Value-at-Risk with Estimation Risk", *Journal of Business and Economic Statistics*, 28, 36-51.

Favre, L. and J. Galeano (2002): "Mean Modified Value-at-Risk Optimization with Hedge Funds", *The Journal of Alternative Investments*, 5, 21-25.

Ferreira, M. and J. Lopez (2005): "Evaluating Interest Rate Covariance Models within a Value-at-Risk Framework", *Journal of Financial Econometrics*, 3, 126-168.

Föllmer, H. and A. Schied (2002): "Convex Measures of Risk and Trading Constraints", *Finance and Stochastics*, 6, 429-447.

Frittelli, M. and E. Rosazza Gianin (2002): "Putting Order in Risk Measures", *Journal of Banking and Finance*, 26, 1473-1486.

Gouriéroux, C. and J. Jasiak (2008): "Dynamic Quantile Models", *Journal of Econometrics*, 147, 198-205.

Granger, C. and C. Sin (2000): "Modelling the Absolute Returns of Different Stock Indices: Exploring the Forecastability of an Alternative Measure of Risk", *Journal of Forecasting*, 19, 277-298.

Haas, M. (2005): "Improved Duration-based Backtesting of Value-at-Risk", *Journal of Risk*, 8, 17-38.

Hahn, J. (1995): "Bootstrapping Quantile Regression Estimators", *Econometric Theory*, 11, 105-121.

Hansen, B. (2008): "Least-squares Forecast Averaging", *Journal of Econometrics*, 146, 342-350.

He, X. and F. Hu (2002): "Markov Chain Marginal Bootstrap", *Journal of the American Statistical Association*, 97, 783-795.

Hendricks, D. (1996): "Evaluation of Value-at-Risk Models using Historical Data", *Economic Policy Review*, 2, 39-69.

Horowitz, J. (1998): "Bootstrap Methods for Median Regression Models", *Econometrica*, 66, 1327-1351.

Hosking, J. (1980): "The Multivariate Portmanteau Statistic", *Journal of the American Statistical Association*, 75, 602-608.

Hurlin, C. and S. Tokpavi (2007): "Un test de validité de la *Value-at-Risk*", *Revue Economique*, 58, 599-608.

Hurlin, C. and S. Tokpavi (2008): "Une évaluation des procédures de *Backtesting*: Tout va pour le mieux dans le meilleur des mondes", *Finance*, 29, 53-80.

Inclan, C. and G. Tiao (1994): "Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance", *Journal of the American Statistical Association*, 89, 913-923.

Jones, M. (1994): "Expectiles and M-Quantiles are Quantiles", *Statistics and Probability Letters*, 20, 149-153.

Jorion, P. (2006): *Value-at-Risk: The New Benchmark for Managing Financial Risk*, McGraw-Hill, 600 pages.

Kantelhardt, J., E. Koscielny-Bunde, H. Rego, S. Havlin and A. Bunde (2001): "Detecting Long-range Correlations with Detrended Fluctuation Analysis", *Physica A*, 295, 441-454.

Koenker, R. and G. Bassett (1978): "Regression Quantiles", *Econometrica*, 46, 33-50.

Koenker, R. and Z. Xiao (2006): "Quantile AutoRegression", *Journal of the American Statistical Association*, 101, 980-990.

Koenker, R. and Z. Xiao (2009): "Conditional Quantile Estimation for GARCH Models", *Boston College Working Papers in Economics*, 725, 37 pages.

Kim, T., S. Manganelli and H. White (2008): "Modeling AutoRegressive Conditional Skewness and *Kurtosis* with Multi-Quantile CAViaR", *ECB Working Paper*, 957, 40 pages.

Kuan, C., J. Yeh and Y. Hsu (2009): "Assessing Value-at-Risk with CARE, the Conditional AutoRegressive Expectile Models", *Journal of Econometrics*, 150, 261-270.

Kupiec, P. (1995): "Techniques for verifying the Accuracy of Risk Measurement Models", *Journal of Derivatives*, 3, 73-84.

Ljung, G. and G. Box (1978): "On a Measure of a Lack of Fit in Time Series Models", *Biometrika*, 65, 297-303.

Longin, F. (2000): "From VaR to Stress Testing: The Extreme Value Approach", *Journal of Banking and Finance*, 24, 1097-1130.

Lugannani, R. and S. Rice (1980): "Saddlepoint Approximation for the Distribution of the Sum of Independent Random Variables", *Advanced in Applied Probability*, 12, 475-490.

Mallows, C. (1973): "Some Comments on C_p", *Technometrics*, 15, 661-675.

Martins-Filho, C. and F. Yao (2006): "Estimation of Value-at-Risk and Expected Shortfall based on Non-linear Models of Return Dynamics and Extreme Value Theory", *Studies in Nonlinear Dynamics & Econometrics*, 10, 41 pages.

McLeod, A. and W. Li (1983): "Diagnostic checking ARMA Time Series Models using Squared-residual Autocorrelations", *Journal of Time Series Analysis*, 4, 269-273.

Newey, W. and J. Powell (1987): "Asymmetric Least Squares Estimation and Testing", *Econometrica*, 55, 819-847.

Nieto, M. and E. Ruiz (2008): "Measuring Financial Risk: Comparison of Alternative Procedures to estimate VaR and ES", *Statistics and Econometrics Working Paper*, University Carlos III, 45 pages.

Pflug, G. (2000): "Some Remarks on the Value-at-Risk and the Conditional Value-at-Risk",

in *Probabilistic Constrained Optimization - Methodology and Applications*, S. Uryasev (Ed), Kluwer Academic Publishers, 272-281.

Prause, K. (1999): "The Generalized Hyperbolic Model: Estimation, Financial Derivatives, and Risk Measures", *Ph.D. Thesis*, University of Freiburg, 168 pages.

RiskMetrics (1996): *Technical Document*, *Morgan Guaranty Trust Company of New York*, 296 pages.

Schreiber, T. (1998): "Constrained Randomization of Time-series Data", *Physical Review Letters*, 80, 2105-2108.

Taylor, J. (2008a): "Estimating Value-at-Risk and Expected Shortfall using Expectiles", *Journal of Financial Econometrics*, 6, 231-252.

Taylor, J. (2008b): "Using Exponentially Weighted Quantile Regression to Estimate Value-at-Risk and Expected Shortfall", *Journal of Financial Econometrics*, 6, 382-406.

Timmermann, A. (2006): "Forecast Combinations", *Handbook of Economic Forecasting 1*, Elliott-Granger-Timmermann (Eds), Elsevier, 135-196.

Tolikas, K. (2008): "Value-at-Risk and Extreme Value Distributions for Financial Returns", *Journal of Risk*, 10, 31-78.

Venter, J. and P. de Jongh (2002): "Risk Estimation using the Normal Inverse Gaussian Distribution", *Journal of Risk*, 4, 1-23.

Visser, M. (2008): "Ranking and Combining Volatility *Proxies* for GARCH and Stochastic Volatility Models", *MPRA Paper*, 4917, 31 pages.

Wong, W. (2008): "Backtesting Trading Risk of Commercial Banks using Expected Shortfall", *Journal of Banking & Finance*, 32, 1404-1415.

Yamai, Y. and T. Yoshiba (2002): "On the Validity of Value-at-Risk: Comparative Analyses with Expected Shortfall", *Monetary and Economic Studies*, 20, 57-85.

Yao, Q. and H. Tong (1996): "Asymmetric Least Squares Regression Estimation: A Nonparametric Approach", *Journal of Nonparametric Statistics*, 6, 273-292.

Table 1: P-statistics related to Null Hypotheses
corresponding to an Out-of-Sample "Good" Property of a Value-at-Risk Model

Tests	**Exception Frequency (1)**	**Indep. (2)**	**Resampling Indep. (3)**	**Cond. / Uncond. (4)**	**Exception Magnitude (5)**	**GMM Duration Uncond. (6)**	**GMM Duration Cond. (7)**	**GMM Duration Indep. (8)**	**Dynamic Quantile (9)**	**Saddlepoint Technique (10)**
Daily VaR 95% Methods:										
Historical	**.00*** **	100.00	100.00	**.00*** **	**.00*** **	**.25*** **	**.00*** **	**.00*** **	**.29*** **	**.11*** **
Normal	**.00*** **	100.00	100.00	**.00*** **	**.00*** **	21.44	**.00*** **	**.00*** **	**.19*** **	**.00*** **
Student	64.69	100.00	99.98	90.04	**.00*** **	70.02	**.00*** **	**.00*** **	3.31**	**.00*** **
NIG	**.00*** **	100.00	100.00	**.00*** **	**.00*** **	**.10*** **	**.00*** **	**.00*** **	**.09*** **	**.52*** **
GEV	**.00*** **	100.00	100.00	**.00*** **	**.00*** **	6.83*	**.00*** **	**.00*** **	1.45**	**.00*** **
GPD	**.00*** **	**.03*** **	**.01*** **	**.00*** **	**.00*** **	**.00*** **	**.00*** **	**.00*** **	42.11	**.00*** **
RiskMetrics	**.00*** **	100.00	100.00	**.00*** **	**.00*** **	17.40	44.13	40.41	98.11	100.00
Normal GARCH(1,1)	53.47	100.00	99.99	82.47	**.00*** **	52.43	56.39	37.27	99.89	99.70
CAViaR Symmetric Absolute Value	30.80	100.00	100.00	59.47	**.00*** **	31.61	38.31	28.10	98.72	100.00
CAViaR Asymmetric Slope	**.00*** **	100.00	100.00	**.00*** **	**.00*** **	5.89*	39.46	63.51	96.12	100.00
CAViaR Indirect GARCH	37.51	100.00	100.00	67.48	**.00*** **	37.80	41.46	23.46	98.29	100.00
CAViaR Adaptive	82.66	100.00	100.00	97.63	**.00*** **	77.14	96.48	90.98	93.50	100.00
Naive Mean of VaR	34.84	100.00	99.91	64.43	**.00*** **	39.23	**.00*** **	**.00*** **	100.00	**.00*** **
Naive Equal-weight DARE	34.04	100.00	100.00	63.48	**.00*** **	34.61	65.64	59.65	98.22	100.00
Daily VaR 99% Methods:										
Historical	**.03*** **	**.66*** **	16.54	**.00*** **	**.00*** **	**.16*** **	**.00*** **	**.00*** **	99.40	**.00*** **
Normal	**.00*** **	**.00*** **	9.68*	**.00*** **	**.00*** **	**.00*** **	**.00*** **	**.00*** **	94.33	**.00*** **
Student	**.01*** **	**.94*** **	14.20	**.00*** **	**.00*** **	**.07*** **	**.00*** **	**.00*** **	99.49	**.00*** **
NIG	**.02*** **	**.12*** **	16.17	**.00*** **	**.00*** **	**.15*** **	**.00*** **	**.00*** **	99.44	**.00*** **
GEV	**.00*** **	**.01*** **	8.62*	**.00*** **	**.00*** **	**.00*** **	**.00*** **	**.00*** **	88.07	3.93**
GPD	15.85	3.46**	6.91*	3.97**	**.00*** **	18.59	**.00*** **	**.00*** **	99.81	**.00*** **
RiskMetrics	**.01*** **	**.84*** **	16.98	**.00*** **	**.00*** **	**.07*** **	2.97*	61.88	99.95	100.00
Normal GARCH(1,1)	**.07*** **	2.82**	20.86	**.03*** **	**.00*** **	**.39*** **	12.06	47.28	99.99	100.00
CAViaR Symmetric Absolute Value	5.02*	5.09*	14.08	2.19**	**.00*** **	8.61*	5.59*	7.37*	100.00	27.16
CAViaR Asymmetric Slope	**.16*** **	2.30**	20.96	**.05*** **	**.00*** **	**.64*** **	**.02*** **	**.11*** **	99.88	99.99
CAViaR Indirect GARCH	32.22	15.22	5.39*	21.98	**.00*** **	38.85	67.00	54.53	99.99	99.18
CAViaR Adaptive	1.82**	**.17*** **	18.69	**.04*** **	**.00*** **	2.92**	**.01*** **	**.01*** **	99.97	**.15*** **
Naive Mean of VaR	1.82**	1.18**	18.17	**.26*** **	**.00*** **	3.62**	**.00*** **	**.00*** **	100.00	**.00*** **
Naive Equal-weight DARE	25.81	2.81**	6.42*	4.73**	**.00*** **	32.17	12.98	9.32*	99.99	64.63

Note: DataStream; daily data from July 9th 1987 until March 18th 2009 of the CAC40 Index; computations by the authors. This period contains 5,659 daily returns. We use a moving window of four years (1,044 daily returns) to dynamically re-estimate parameters for the various methods. Forecasted VaR were computed for each method for the final 4,615 days (about 18 years). This table shows probability of failing to reject (p-values expressed in percentage) the null hypothesis of different tests statistics for "good" VaR model. Significant probabilities of rejection at a significance level of 10%, 5% and 1% are respectively marked with *, ** and ***. Bold probabilities represent risk measures, which are rejected according to the usual 1% significance level. (1) Exception Frequency (Kupiec, 1995) is based on the indicator variable of the good VaR forecast; (2) Independence test (Christoffersen, 1998) detects the serial independence in the VaR model forecast; (3) Resampling Independence test is an Independence test performed using Escanciano and Olmo (2010) subsampling approximation technique; (4) Unconditional Coverage test (Kupiec, 1995) is estimated by counting the exceptions over the entire period and Conditional Coverage test (Christoffersen, 1998) is based on the idea that if the VaR estimates have correct conditional coverage, the exception variable must exhibit both correct unconditional coverage and serial independence; (5) Exception Magnitudes (Berkowitz, 2001) is based on the idea that the magnitudes of exceptions should be of primary interest to the various users of VaR models; (6) to (8) Generalized Method of Moment Duration-based tests (Candelon *et al.*, 2011) are based on the point that VaR forecast tests can be expressed as simple moment conditions; (9) Dynamic Quantile test (Engle and Manganelli, 2004) detects a good model for VaR estimation with the idea that a good model should produce a sequence of unbiased and uncorrelated exception indicator variable; (10) Saddlepoint Technique test (Wong, 2008) is based on a full expansion, which can locally approximate the density of the Expected Shortfall, and can be compared to the observed value.

Table 2: P-statistics related to Null Hypotheses
corresponding to and Out-of-Sample "Good" Property of an Expected Shortfall Model

Tests	**Exception Frequency (1)**	**Indep. (2)**	**Resampling Indep. (3)**	**Cond. / Uncond. (4)**	**Exception Magnitude (5)**	**GMM Duration Uncond. (6)**	**GMM Duration Cond. (7)**	**GMM Duration Indep. (8)**	**Dynamic Quantile (9)**	**Saddlepoint Technique (10)**
Daily ES 95% Methods:										
Historical	**.35*** **	**.00*** **	10.50	**.00*** **	**.00*** **	**.93*** **	**.00*** **	**.00*** **	90.11	**.00*** **
Normal	**.00*** **	**.00*** **	**.05*** **	**.00*** **	**.00*** **	**.00*** **	**.00*** **	**.00*** **	66.35	**.00*** **
Student	2.48**	**.00*** **	11.05	**.00*** **	**.00*** **	4.18**	**.00*** **	**.00*** **	90.32	**.00*** **
NIG	**.00*** **	100.00	**.14*** **	**.00*** **	**.00*** **	**.00*** **	**.00*** **	**.00*** **	47.58	3.02**
GEV	78.55	**.01*** **	10.38	**.04*** **	**.00*** **	77.24	**.00*** **	**.00*** **	96.37	**.00*** **
GPD	**.01*** **	**.00*** **	9.95*	**.00*** **	**.00*** **	**.06*** **	**.00*** **	**.00*** **	88.32	**.00*** **
RiskMetrics	**.19*** **	19.72	9.75*	**.35*** **	**.00*** **	**.57*** **	14.86	48.87	99.76	100.00
Normal GARCH(1,1)	**.48*** **	16.42	10.41	**.71*** **	**.00*** **	1.15**	25.94	90.79	99.73	100.00
CARE Symmetric Absolute Value	3.46**	4.14**	19.40	1.34**	**.00*** **	2.28**	**.00*** **	4.47**	99.98	3.80**
CARE Asymmetric Slope	18.03	7.56*	10.18	8.41*	**.00*** **	18.29	**.00*** **	**.23*** **	99.82	16.93
CARE Indirect GARCH	1.43**	3.12**	19.91	**.49*** **	**.00*** **	**.74*** **	**.42*** **	72.27	99.98	95.26
CARE Adaptive	**.74*** **	**.45*** **	18.30	**.05*** **	**.00*** **	**.68*** **	**.00*** **	**.08*** **	99.94	**.00*** **
Naive Mean of ES	94.96	1.51**	12.95	5.21*	100.00	93.56	**.00*** **	**.00*** **	100.00	**.00*** **
Naive Equal-weight DARE	4.54**	4.54**	14.19	1.83**	**.00*** **	3.18**	**.07*** **	13.26	99.98	16.06
Daily ES 99% Methods:										
Historical	**.05*** **	**.40*** **	**.07*** **	**.00*** **	**.00*** **	**.50*** **	**.00*** **	**.00*** **	99.98	**.00*** **
Normal	**.00*** **	1.06**	12.89	**.00*** **	**.00*** **	**.00*** **	**.00*** **	**.00*** **	99.47	**.00*** **
Student	**.00*** **	1.59**	1.41**	**.00*** **	**.00*** **	**.01*** **	**.00*** **	**.00*** **	99.95	**.00*** **
NIG	**.01*** **	**.63*** **	**.09*** **	**.00*** **	**.00*** **	**.16*** **	**.00*** **	**.00*** **	99.97	**.00*** **
GEV	**.00*** **	**.16*** **	9.37*	**.00*** **	**.00*** **	**.00*** **	**.00*** **	**.00*** **	97.03	99.30
GPD	**.00*** **	1.24**	**.77*** **	**.00*** **	**.00*** **	**.02*** **	**.00*** **	**.00*** **	99.95	**.00*** **
RiskMetrics	**.00*** **	1.24**	**.98*** **	**.00*** **	**.00*** **	**.02*** **	**.00*** **	86.51	99.99	100.00
Normal GARCH(1,1)	**.00*** **	1.41**	1.52**	**.00*** **	**.00*** **	**.01*** **	**.08*** **	48.40	100.00	100.00
CARE Symmetric Absolute Value	62.15	69.18	43.98	81.84	**.00*** **	49.63	56.34	54.64	100.00	**.11*** **
CARE Asymmetric Slope	16.04	1.11**	2.83**	1.48**	**.00*** **	21.64	9.80*	5.75*	100.00	1.24**
CARE Indirect GARCH	78.72	67.65	44.59	88.38	**.00*** **	87.07	51.49	40.60	100.00	99.99
CARE Adaptive	86.76	9.46*	16.08	24.40	**.00*** **	92.40	29.12	18.05	100.00	2.09**
Naive Mean of ES	78.72	67.65	28.90	88.38	**.00*** **	72.07	**.24*** **	**.37*** **	100.00	**.00*** **
Naive Equal-weight DARE	96.02	66.12	15.10	90.73	**.00*** **	84.21	87.83	78.35	100.00	50.61

Note: DataStream; daily data from July 9th 1987 until March 18th 2009 of the CAC40 Index; computations by the authors. This period contains 5,659 daily returns. We use a moving window of four years (1,044 daily returns) to dynamically re-estimate parameters for the various methods. Forecasted ES were computed for each method for the final 4,615 days (about 18 years). This table shows probability of failing to reject (p-values expressed in percentage) the null hypothesis of different tests statistics for "good" ES model. The probability associated to the tested Quantile was determined using the Gaussian Distribution. Significant probabilities of rejection at a significance level of 10%, 5% and 1% are respectively marked with *, ** and ***. Bold probabilities represent risk measures, which are rejected according to the usual 1% significance level. (1) Exception Frequency (Kupiec, 1995) is based on the indicator variable of the good Quantile forecast; (2) Independence test (Christoffersen, 1998) detects the serial independence in the ES model forecast; (3) Resampling Independence test is an Independence test performed using Escanciano and Olmo (2010) subsampling approximation technique; (4) Unconditional Coverage test (Kupiec, 1995) is estimated by counting the exceptions over the entire period and Conditional Coverage test (Christoffersen, 1998) is based on the idea that if the ES estimates have correct conditional coverage, the exception variable must exhibit both correct unconditional coverage and serial independence; (5) Exception Magnitudes (Berkowitz, 2001) is based on the idea that the magnitudes of exceptions should be of primary interest to the various users of ES models; (6) to (8) Generalized Method of Moment Duration-based tests (Candelon *et al.*, 2011) are based on the point that ES forecast tests can be expressed as simple moment conditions; (9) Dynamic Quantile test (Engle and Manganelli, 2004) detects a good model for ES estimation with the idea that a good model should produce a sequence of unbiased and uncorrelated exception indicator variable; (10) Saddlepoint Technique test (Wong, 2008) is based on a full expansion, which can locally approximate the density of the Expected Shortfall, and can be compared to the observed value.

Table 3: P-statistics related to Null Hypotheses
corresponding to an Out-of-Sample "Good" Property of DARE Value-at-Risk Models

Tests	**Exception Frequency (1)**	**Indep. (2)**	**Resampling Indep. (3)**	**Cond. / Uncond. (4)**	**Exception Magnitude (5)**	**GMM Duration Uncond. (6)**	**GMM Duration Cond. (7)**	**GMM Duration Indep. (8)**	**Dynamic Quantile (9)**	**Saddlepoint Technique (10)**
VaR 95%										
Hansen-Mallows Optimized-weight DARE-VaR	27.83	100.00	1.63	55.56	**.00*** **	25.89	59.66	46.37	92.49	100.00
Naive Mean of VaR	92.06	2.70**	1.33**	8.62*	**.00*** **	84.98	**.00*** **	**.00*** **	100.00	**.00*** **
Naive Equal-weight DARE-VaR	27.83	100.00	1.42**	55.56	**.00*** **	27.25	52.79	49.87	98.47	100.00
VaR 99%										
Hansen-Mallows Optimized-weight DARE-VaR	**.43*** **	5.69*	14.85	**.28*** **	**.00*** **	1.53**	2.85**	9.48*	99.99	100.00
Naive Mean of VaR	1.05*	**.71*** **	13.79	**.10*** **	100.00	2.34**	**.00*** **	**.00*** **	100.00	**.00*** **
Naive Equal-weight DARE-VaR	30.58	10.07	1.70*	15.39	**.00*** **	35.60	25.32	16.87	99.99	99.87

Note: DataStream; daily data from July 9th 1987 until March 18th 2009 of the CAC40 Index; computations by the authors. This period contains 5,659 daily returns. We use a moving window of four years (1,044 daily returns) to dynamically re-estimate parameters for the various methods and another moving window of four years (1,044 daily returns) to reestimate the optimized weighting function of the DARE approaches. Forecasted VaR were computed for each method for the final 3,571 days (about 18 years). This table shows probability of failing to reject (p-values expressed in percentage) the null hypothesis of different tests statistics for "good" VaR model. Significant probabilities of rejection at a significance level of 10%, 5% and 1% are respectively marked with *, ** and ***. Bold probabilities represent risk measures, which are rejected according to the usual 1% significance level. (1) Exception Frequency (Kupiec, 1995) is based on the indicator variable of the good VaR forecast; (2) Independence test (Christoffersen, 1998) detects the serial independence in the VaR model forecast; (3) Resampling Independence test is an Independence test performed using Escanciano and Olmo (2010) subsampling approximation technique; (4) Unconditional Coverage test (Kupiec, 1995) is estimated by counting the exceptions over the entire period and Conditional Coverage test (Christoffersen, 1998) is based on the idea that if the VaR estimates have correct conditional coverage, the exception variable must exhibit both correct unconditional coverage and serial independence; (5) Exception Magnitudes (Berkowitz, 2001) is based on the idea that the magnitudes of exceptions should be of primary interest to the various users of VaR models; (6) to (8) Generalized Method of Moment Duration-based tests (Candelon *et al.*, 2011) are based on the point that VaR forecast tests can be expressed as simple moment conditions; (9) Dynamic Quantile test (Engle and Manganelli, 2004) detects a good model for VaR estimation with the idea that a good model should produce a sequence of unbiased and uncorrelated exception indicator variable; (10) Saddlepoint Technique test (Wong, 2008) is based on a full expansion, which can locally approximate the density of the Expected Shortfall, and can be compared to the observed value.

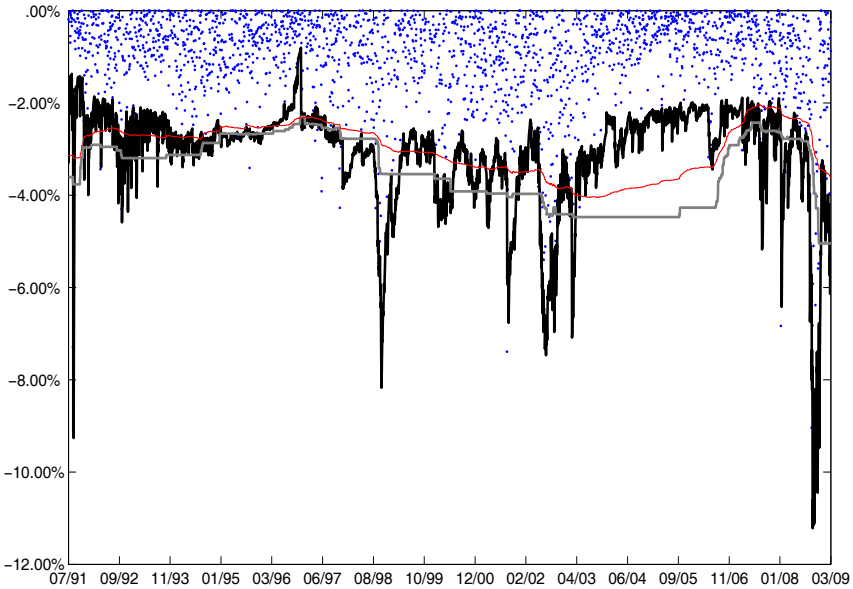
Table 4: P-statistics related to Null Hypotheses
corresponding to an Out-of-Sample "Good" Property of DARE Expected Shortfall Models

Tests	**Exception Frequency (1)**	**Indep. (2)**	**Resampling Indep. (12)**	**Cond. / Uncond. (3)**	**Exception Magnitude (5)**	**GMM Duration Uncond. (7)**	**GMM Duration Cond. (8)**	**GMM Duration Indep. (9)**	**Dynamic Quantile (10)**	**Saddlepoint Technique (13)**
ES 95%										
Hansen-Mallows Optimized-weight DARE-ES	38.30	**.81*** **	14.23	2.06**	**.00*** **	41.88	1.26**	1.38**	99.73	99.92
Naive Mean of ES	43.34	4.24**	10.15	9.39*	**.95*** **	45.31	**.00*** **	**.00*** **	100.00	**.00*** **
Naive Equal-weight DARE-ES	21.44	12.94	19.52	14.65	**.00*** **	19.71	6.63*	15.25	99.97	96.12
ES 99%										
Hansen-Mallows Optimized-weight DARE-ES	27.99	61.83	15.87	49.27	**.00*** **	33.12	69.40	45.79	100.00	94.87
Naive Mean of ES	53.70	65.22	9.57*	74.67	**.00*** **	57.97	2.50**	**.00*** **	99.99	18.04
Naive Equal-weight DARE-ES	53.70	65.22	14.92	74.67	**.00*** **	63.41	81.05	51.69	100.00	34.48

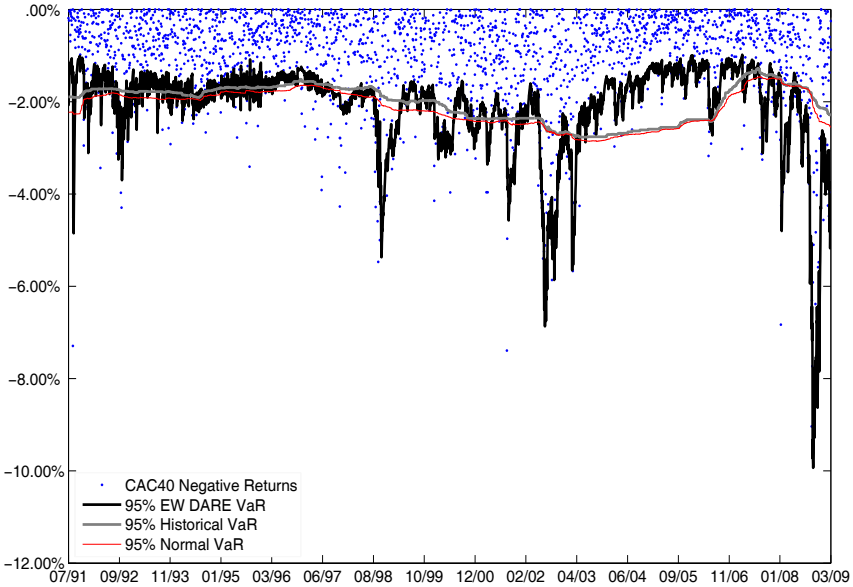
Note: DataStream; daily data from July 9th 1987 until March 18th 2009 of the CAC40 Index; computations by the authors. This period contains 5,659 daily returns. We use a moving window of four years (1,044 daily returns) to dynamically re-estimate parameters for the various methods and another moving window of four years (1,044 daily returns) to reestimate the optimized weighting function of the DARE approaches. Forecasted ES were computed for each method for the final 3,571 days (about 18 years). This table shows probability of failing to reject (p-values expressed in percentage) the null hypothesis of different tests statistics for "good" ES model. The probability associated to the tested Quantile was determined using the Gaussian Distribution. Significant probabilities of rejection at a significance level of 10%, 5% and 1% are respectively marked with *, ** and ***. Bold probabilities represent risk measures, which are rejected according to the usual 1% significance level. (1) Exception Frequency (Kupiec, 1995) is based on the indicator variable of the good Quantile forecast; (2) Independence test (Christoffersen, 1998) detects the serial independence in the ES model forecast; (3) Resampling Independence test is an Independence test performed using Escanciano and Olmo (2010) subsampling approximation technique; (4) Unconditional Coverage test (Kupiec, 1995) is estimated by counting the exceptions over the entire period and Conditional Coverage test (Christoffersen, 1998) is based on the idea that if the ES estimates have correct conditional coverage, the exception variable must exhibit both correct unconditional coverage and serial independence; (5) Exception Magnitudes (Berkowitz, 2001) is based on the idea that the magnitudes of exceptions should be of primary interest to the various users of ES models; (6) to (8) Generalized Method of Moment Duration-based tests (Candelon *et al.*, 2011) are based on the point that ES forecast tests can be expressed as simple moment conditions; (9) Dynamic Quantile test (Engle and Manganelli, 2004) detects a good model for ES estimation with the idea that a good model should produce a sequence of unbiased and uncorrelated exception indicator variable; (10) Saddlepoint Technique test (Wong, 2008) is based on a full expansion, which can locally approximate the density of the Expected Shortfall, and can be compared to the observed value.

Figure 1: 99% and 95% Value-at-Risk Forecasts

- 99% Value-at-Risk Level -



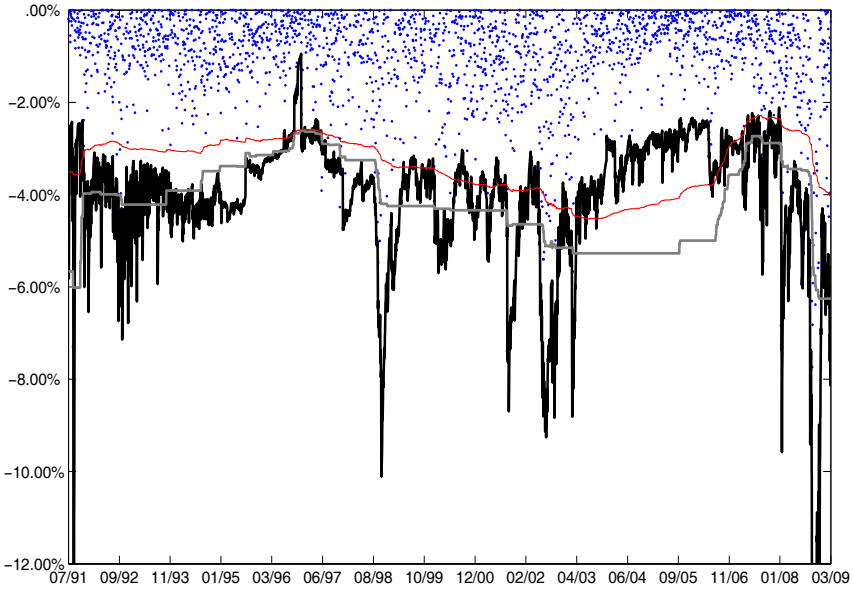
- 95% Value-at-Risk Level -



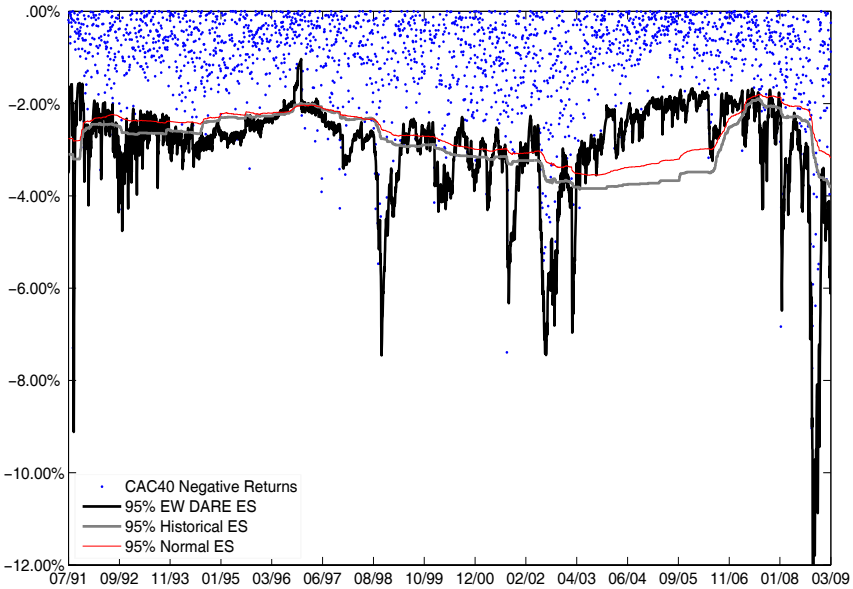
Note: *DataStream*; daily data from July 9th 1987 until March 18th 2009 of the CAC40 Index; computations by the authors. This period contains 5,659 daily returns. We use a moving window of four years (1,044 daily returns) to dynamically re-estimate parameters for the various methods. X-axis gives the Out-of-Sample period and Y-axis gives the Negative returns, DARE-VaR, Historical-VaR and Normal-VaR.

Figure 2: 99% and 95% Expected Shortfall Forecasts

- 99% Expected Shortfall Level -



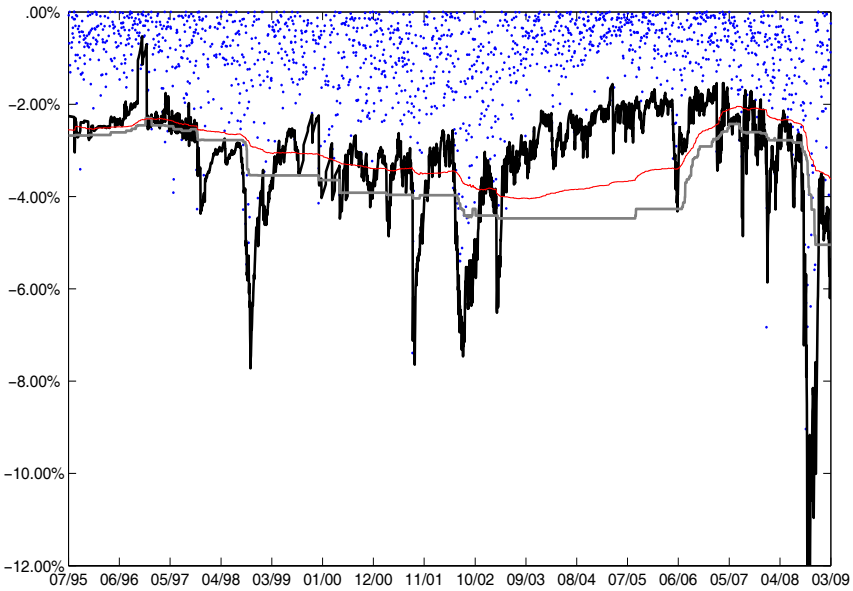
- 95% Expected Shortfall Level -



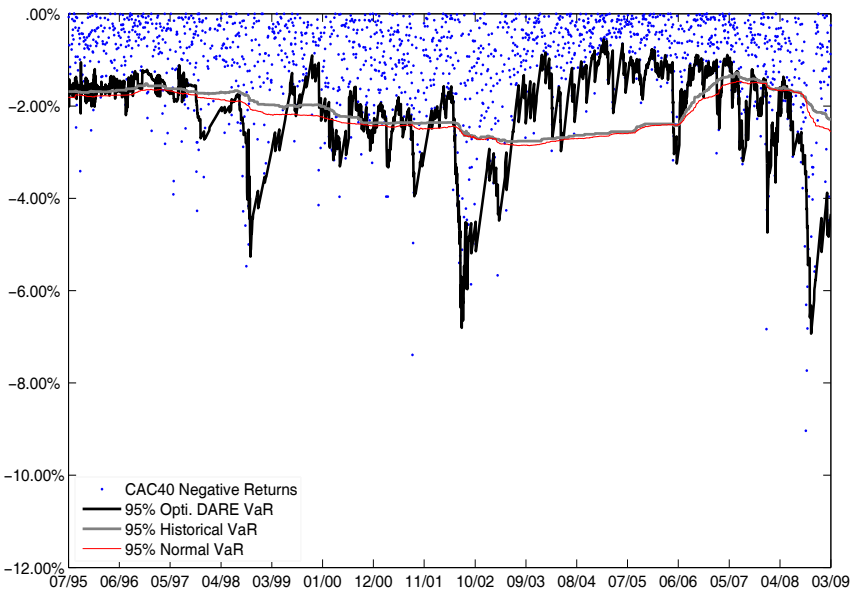
Note: *DataStream*; daily data from July 9th 1987 until March 18th 2009 of the CAC40 Index; computations by the authors. This period contains 5,659 daily returns. We use a moving window of four years (1,044 daily returns) to dynamically re-estimate parameters for the various methods. X-axis gives the Out-of-Sample period and Y-axis gives the Negative returns, DARE-VaR, Historical-VaR and Normal-VaR.

Figure 3: 99% and 95% Value-at-Risk Forecasts

- 99% Value-at-Risk Level -



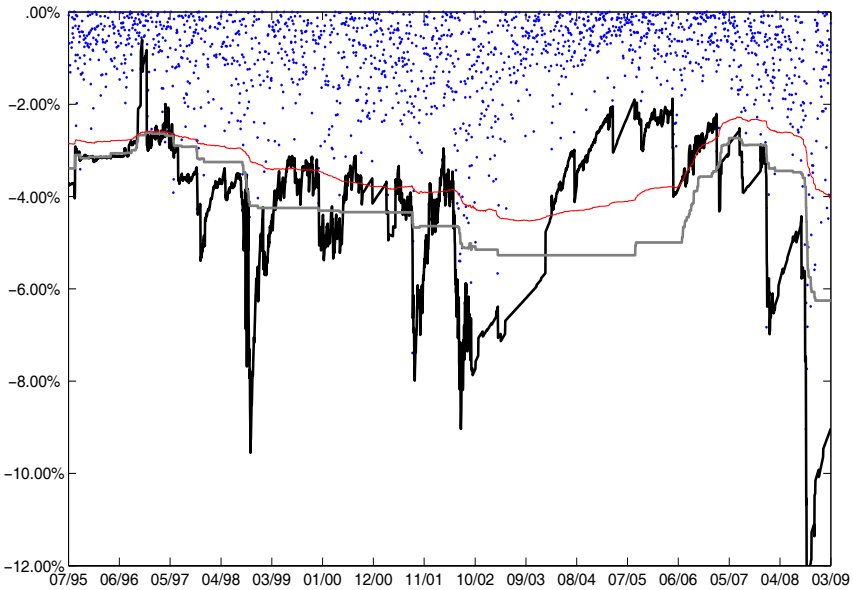
- 95% Value-at-Risk Level -



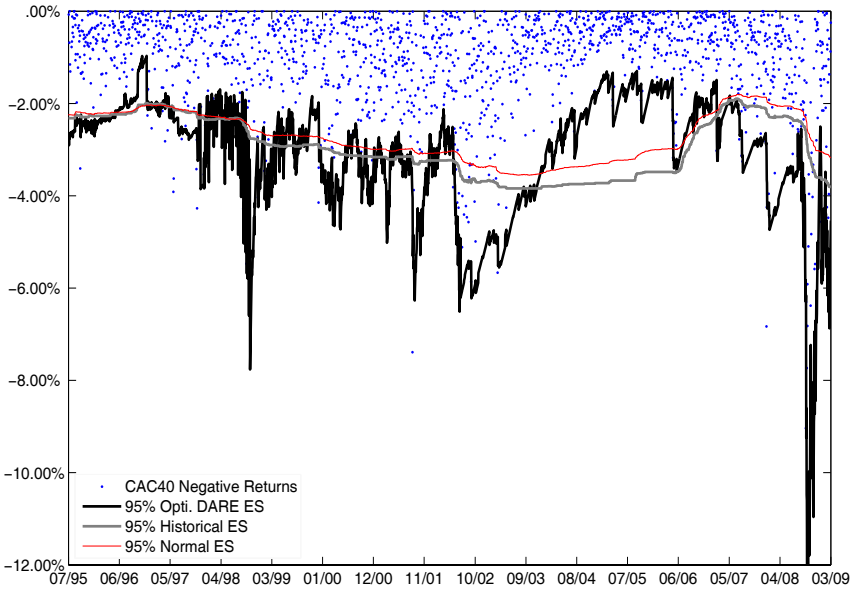
Note: *DataStream*; daily data from July 9th 1987 until March 18th 2009 of the CAC40 Index; computations by the authors. This period contains 5,659 daily returns. We use a moving window of four years (1,044 daily returns) to dynamically re-estimate parameters for the various methods and another moving window of four years (1,044 daily returns) to reestimate the optimized weighting function of the DARE approaches. Forecasted VaR were computed for each method for the final 3,571 days (about 18 years). X-axis gives the Out-of-Sample period and Y-axis gives the Negative returns, Hansen Optimized-weight DARE-VaR, Historical-VaR and Normal-VaR.

Figure 4: 99% and 95% Expected Shortfall Forecasts

- 99% Expected Shortfall Level -



- 95% Expected Shortfall Level -



Note: *DataStream*; daily data from July 9th 1987 until March 18th 2009 of the CAC40 Index; computations by the authors. This period contains 5,659 daily returns. We use a moving window of four years (1,044 daily returns) to dynamically re-estimate parameters for the various methods and another moving window of four years (1,044 daily returns) to reestimate the optimized weighting function of the DARE approaches. Forecasted ES were computed for each method for the final 3,571 days (about 18 years). X-axis gives the Out-of-Sample period and Y-axis gives the Negative returns, Hansen Optimized-weight DARE-ES, Historical-VaR and Normal-VaR.