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

Benjamin Hamidi† Patrick Kouontchou‡ Bertrand Maillet§

December 2009

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) and Gouriéroux and Jasiak (2008), 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, and we secondly explain the DARE approach and how expectiles can be used to estimate quantile risk measures. We 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. Finally, we illustrate several conditional weighting functions (Hansen, 2008) to the risk models aggregated into the DARE Approach in order to select dynamically the more appropriated quantile models 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, Christophe Hurlin, Gilbert Colletaz, Gregory Jannin, Jean-Philippe Médecin, Paul Merlin and Tristan Roger for help, suggestions and encouragements when preparing this work. Preliminary version: do not quote or diffuse without permission. The third author thanks the Europlace Institute of Finance for financial support. The usual disclaimers apply.

†A.A.Advisors-QCG (ABN AMRO), Variances and University of Paris-1 (CES/CNRS). Email: benjamin.hamidi@univ-paris1.fr.

‡Variances and University of Paris-1 (CES/CNRS). Email: patrick.kouontchou@univ-paris1.fr.

§A.A.Advisors-QCG (ABN AMRO), Variances and University of Paris-1 (CES/CNRS and EIF). Corresponding author: Dr B. B. Maillet, CES/CNRS, MSE, 106-112 Bd de l'Hôpital F-75647 Paris Cedex 13. Tel: +33 144078189/70 (fax). Email: bmaillet@univ-paris1.fr.

"Towards a Well-diversified Risk Measure: A DARE Approach"

December 2009

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) and Gouriéroux and Jasiak (2008), 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, and we secondly explain the DARE approach and how expectiles can be used to estimate quantile risk measures. We 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. Finally, we illustrate several conditional weighting functions (Hansen, 2008) to the risk models aggregated into the DARE Approach in order to select dynamically the more appropriated quantile models 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 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 funds based on VaR estimates. Although VaR became the standard measure of market risk, it has been criticized for reporting only a quantile, and thus disregarding all the outcomes beyond the threshold. In addition, VaR is not a subadditive risk measure. This property concerns the main idea which is the aggregated risk of a portfolio should not be greater than the individual risk of its constituent parts (see Artzner *et al.*, 1999; Acerbi and Tasche, 2002). The Expected Shortfall (ES) is a risk measure that overcomes these weaknesses. Indeed, the ES is defined as the conditional expectation of the return given that it falls below the VaR (see Yamai and Yoshiba, 2002).

Many VaR and ES estimation methods coexists nowadays. However, none of these methods seems adapted to every market configurations. Besides, none of these methods always 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. Considering the economic impact of the quality of these risk measures for the banks capital constraints, it seems crucial to limit this risk model. The objective of this paper is to provide a complete framework to aggregate different quantile and expectile models (see Hansen, 2008)), to provide more diversified VaR and ES measures, limiting the model risk in the same way, that has been done for the volatility in a multi-model approach (see Visser, 2008).

A recent development in the VaR literature concerns the Conditional AutoRegressive Value-at-Risk (CAViaR) class of models (see Engle and Manganelli, 2004), Multi-quantile CAViaR (see Kim *et al.*, 2008) and Quantile GARCH (QGARCH) model (see Koenker and Xiao, 2009). This approach to the VaR estimation has a strong appeal insofar as it provides an extensive modeling framework and does not rely on strong distributional assumptions. However, the focus is merely on VaR estimation, and it is not obvious to estimate the corresponding ES.

Following Taylor (2008a and 2008b), Kuan *et al.* (2009) and Gouriéroux and Jasiak (2008), 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. The solution of an ALS regression is known as 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 propose in this article a new class of models for the conditional VaR and ES modeling: the Dynamic AutoRegressive Expectiles (DARE, in short). We thus provide more diversified VaR and ES measures.

In the first section, we review the literature about VaR and ES measures and estimation models. In the second section, we describe how expectiles can be used to estimate VaR and ES, then we introduce the DARE approach. We recall in the third section main and recent backtests procedure for VaR and ES models. According to these VaR and ES tests, we present the empirical results regarding the accuracy of the proposed method compared to traditional VaR and ES methods. We finally illustrate several conditional weighting functions to the risk models aggregated into the DARE approach in order to select dynamically the more appropriated quantile models for extremal risk measure estimation. The last section concludes.

2 Main Extreme Market Risk Models

Extreme market risk measures as Value-at-Risk (VaR) and Expected Shortfall (ES) can be estimated using several approaches.

We first define and introduce hereafter the VaR and the ES extreme market 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. Actually they can be classified into three categories: parametric, non-parametric and semi-parametric methods (see Engle and Manganelli, 1999).

2.1 Extreme Risk Measures Definitions

Risk management has become in these past few years a central object of interest for researchers, market practitioners and regulators. The VaR is well known as being one of the benchmark measure for market risk.

The VaR is defined as the loss a portfolio may suffer at a given confidence level over a fixed holding period such as:

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

where r_t is the random variable of asset returns, and α is the confidence level for the VaR computation.

The distribution of r_t can be set here as an empirical non-continuous distribution, or as a theoretical specified continuous distribution.

The Value-at-Risk of an empirical distribution can also be viewed as the α-quantile. Although the VaR can be mentioned as a reference and is imposed by regulators, criticisms have been formulated against its generalized use (see Beder, 1995; Cheridito and Stadje, 2009). For instance, the VaR does not enable us to answer the question: "what is the magnitude of the loss when the VaR limit is exceeded?". Another issue regarding the VaR, pointed out by the article of Artzner *et al.* (1999), concerns the "non-coherence" property of this risk measure; it fails to respect the subadditivity property, *i.e.* the VaR of two assets in a portfolio can be greater than these two aggregated individual VaRs. Prause (1999) also argues that, to avoid bankruptcy, we should forecast the distribution of the maximum expected loss. From this point of view, regulators should use other risk measures than the VaR in order to get a better characterization of extreme events, especially for nonlinear portfolio returns.

Therefore, another risk measure has appeared in the literature, 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 Acerbi and Tasche (2002) the ES can be written as such (with previous notations):

$$ES_{\alpha,t} = -E\left[r_t | r_t \leq -VaR_{\alpha,t}\right]. \tag{2}$$

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

$$ES_{\alpha,t} = -(1-\alpha)^{-1} \int_{-\infty}^{-VaR_{\alpha,t}} r_t f(r_t) dr_t, \tag{3}$$

where $f(.)$ is the Probability Density Function of r_t.

For all estimation methods, VaR and ES can be written as:

$$\begin{cases} VaR_{\alpha,t} = \mu_t + F_{1-\alpha}^{-1}(r_t^*)\sigma_t \\ ES_{\alpha,t} = \mu_t + \left\{(1-\alpha)^{-1} f[F_{1-\alpha}^{-1}(r_t^*)]\right\}\sigma_t, \end{cases} \tag{4}$$

where μ_t and σ_t are respectively the expected return and standard deviation at time t, $f(.)$ and $F(.)$ are respectively a Probability Density and Cumulative Distribution Function, and r_t^* is the standardized return.

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 on the kind of portfolio, the availability of computational resources and time constraints. The main approaches can be classified into three categories: parametric, non-parametric and semi-parametric methods (see Engle and Manganelli, 1999; Jorion, 2006; Nieto and Ruiz, 2008).

2.2.1 Parametric Approach of Extreme Risk Measure Estimations

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 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 parameters estimation and on the shape of the chosen distribution. Assuming that $F(.)$ is the Cumulative Distribution Function (CDF) of returns, the estimated VaR can be written such as:

$$\widehat{VaR}_{\alpha,t} \;=\; \widehat{\mu}_t + F^{-1}_{1-\alpha}(r_t^*)\,\widehat{\sigma}_t, \tag{5}$$

where $\widehat{\sigma}_t$ and $\widehat{\mu}_t$ are respectively the estimation of the return standard deviation and the expected return 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 a Gaussian Cumulative Distribution Function of returns $\Phi(.)$, the specific CDF in equation (5) is replaced by $\Phi(.)$. However, Normal distribution suffer 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 heteroscedastic. Using the t-distribution, the Normal Inverse Gaussian, (Barndorff-Nielsen, 1998) or applying Extreme Value Theory (EVT) have been proposed as possible solutions to avoid these problems and are presented below.

In an EVT framework, let us suppose now that r_t represents extreme movements in risk factors 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 exceedances 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 Function $H(.)$, has a formal expression recalled hereafter (see Embrechts *et al.*, 1997).
For the GEV family, the Cumulative Distribution Function $H(.)$ can be written, such as

(with previous notations):

$$H(\underline{r}_s|\nu,\theta,\xi) = \begin{cases} \exp\left\{-\left[1+\xi(\underline{r}_s-\nu)\theta^{-1}\right]^{-1/\xi}\right\} & \text{if } \xi \neq 0 \\ \exp\left\{-\exp\left[-(\underline{r}_s-\nu)\theta^{-1}\right]\right\} & \text{otherwise,} \end{cases} \tag{6}$$

with $[1+\xi\,(\underline{r}_s-\nu)\,\theta^{-1}] \geq 0$, where $\nu \in \mathbb{R}$, $\theta \in \mathbb{R}_+^*$ and $\xi \in \mathbb{R}$ are, respectively, location, scale and shape parameters For the GPD family, the Cumulative Distribution Function $H(.)$ can be written, such as (with previous notations):

$$H(\underline{r}_s|\nu,\theta,\xi) = \begin{cases} 1-\left[1+\xi(\underline{r}_s-\nu)\theta^{-1}\right]^{-1/\xi} & \text{if } \xi \neq 0 \\ 1-\exp\left[-(\underline{r}_s-\nu)\theta^{-1}\right] & \text{otherwise.} \end{cases} \tag{7}$$

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

$$\widehat{VaR}_{\alpha,t} \; = \; \widehat{\nu}_t + \widehat{H}^{-1}_{1-\alpha}(\underline{r}^*_s)\,\widehat{\theta}_t, \tag{8}$$

where $H^{-1}_{1-\alpha}(\cdot)$ is the Inverse Standardized Cumulative Distribution Function given by $H(\underline{r}^*_s) = H(\underline{r}^*_s|0,1,\xi)$ for the $(1-\alpha)$ quantile, and $\underline{r}^*_s$ is the standardized $\underline{r}_s$.

For VaR estimation, it has been shown that NIG-based approach is more robust than the EVT method (see Barndorff-Nielsen, 1998). The EVT method should only be used in large samples if the NIG distribution does not adequately fit 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 provides accurate results and is the preferred method when the available sample is small.
At any rate, the 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.

2.2.2 Non-parametric Approach of Extreme Risk Measure Estimations

The most widely used non-parametric method to estimate a predetermined quantile is based on the so-called historical simulations. The latter only requires mild distributional assumptions and it implies the estimation of the VaR as the quantile of the empirical distribution of historical returns from a moving window of the most recent period. The VaR can be written such as:

$$VaR_{\alpha,t} \; = F^{-1}_{1-\alpha}(r_t), \tag{9}$$

where $F^{-1}_{1-\alpha}(.)$ is the r_t unconditional quantile function associated to the probability $1-\alpha$.

The main problem is the way to get the width of this window: few observations will lead to an important sampling error, whereas too many observations will slow down estimates to react to changes in the true distribution of financial returns.

Other methods allocate to the sample of returns exponentially decreasing weights (the sum is equal to one). 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.

2.2.3 Semi-parametric Approach of Extreme Risk Measure Estimations

Semi-parametric estimation methods combine the two previous approaches. Cornish-Fisher and Quantile regression approaches belong to this family.

The Cornish-Fisher approach (1937) is based on statistic expansions of the returns Probability Density Function around a reference density, which integrate the higher moments of distribution. Assuming for example a Gaussian distribution of the returns, the Cornish-Fisher expansions adjust the distribution to the true returns distribution according to skewness and excess *kurtosis*, such as:

$$\begin{cases} f(r_t) = \psi(r_t)\ \phi(r_t) + \epsilon_t \\ \psi(r_t) = \left\{1 + \frac{\gamma_1[f(r_t)]}{3!} H_2(r_t) + \frac{\gamma_2[f(r_t)]}{4!} H_3(r_t) + 10\frac{\gamma_1[f(r_t)]^2}{36} H_6(r_t)\right\}, \end{cases} \tag{10}$$

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

Thus the quantile at level α (denoted $Q_\alpha(.)$) is then:

$$\begin{aligned} Q_\alpha\left[f\left(r_t\right)\right] = Q_\alpha\left[\Phi\left(r_t\right)\right] + \tfrac{\gamma_1[f(r_t)]}{3!} H_2\left\{Q_\alpha\left[\Phi\left(r_t\right)\right]\right\} + \tfrac{\gamma_2[f(r_t)]}{4!} H_3\left\{Q_\alpha\left[\Phi\left(r_t\right)\right]\right\} \\ -\tfrac{\gamma_1[f(r_t)]^2}{36}\left\{2Q_\alpha\left[\Phi\left(r_t\right)\right]^3 - 5Q_\alpha\left[\Phi\left(r_t\right)\right]\right\}, \end{aligned} \tag{11}$$

where $\Phi(.)$ is the standard normal Probability Density Function.
The Cornish-Fisher VaR estimation at level α is then (with previous notations):

$$\widehat{VaR}_{\alpha,t} = \widehat{\mu}_t + \widehat{Q}_{(1-\alpha)}\left[f\left(r_t\right)\right]\widehat{\sigma}_t, \tag{12}$$

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* model can be considered as another semi-parametric approach. It assumes that asset returns follow a centered Normal distribution with a volatility estimated by the Exponential Weighted Moving Average method, which corresponds to an Integrated GARCH model, obtained by using historical data, such as:

$$\widehat{\sigma}_t^2 \;=\; \lambda\,\widehat{\sigma}_{t-1}^2 \;+\; (1-\lambda)\,r_{t-1}^2, \tag{13}$$

where $\widehat{\sigma}_t^2$ is the estimation of the returns variance at time t and λ is the parameter associated to the *RiskMetrics* model (equals to .94).
Many other volatility estimation methods, belonging for example to the GARCH family, have been already tested in the literature. 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.

The main 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) have directly defined the dynamics of risk, by means of an AutoRegression involving the lagged-VaR and the lagged value of endogenous variable, called 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: $VaR_{\alpha,t}^{SAV}(r_{t-1},\beta)$, $VaR_{\alpha,t}^{AS}(r_{t-1},\beta)$, $VaR_{\alpha,t}^{IG}(r_{t-1},\beta)$, $VaR_{\alpha,t}^{A}(r_{t-1},\beta)$.

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

$$VaR_{\alpha,t}^{SAV}(r_{t-1},\beta) = \beta_1 + \beta_2 \times VaR_{\alpha,t-1}^{SAV}(r_{t-1},\beta) + \beta_3 \times |r_{t-1}|, \tag{14}$$

where β_i 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.
The Asymmetric Slope CAViaR can be written such as (with previous notations):

$$\begin{aligned} VaR_{\alpha,t}^{AS}(r_{t-1},\beta) = \beta_1 + \beta_2 \times VaR_{\alpha,t-1}^{AS}(r_{t-1},\beta) \\ +\beta_3 \times \max(0,r_{t-1}) + \beta_4 \times [-\min(0,r_{t-1})]. \end{aligned} \tag{15}$$

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

$$VaR_{\alpha,t}^{IG}(r_{t-1},\beta) = \beta_1 + \beta_2 \times \left[VaR_{\alpha,t-1}^{IG}(r_{t-1},\beta)\right]^2 + \beta_3 \times r_{t-1}^2. \tag{16}$$

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

$$\begin{aligned} VaR_{\alpha,t}^{A}(r_{t-1},\beta) = VaR_{\alpha,t-1}^{A}(r_{t-1},\beta) - \alpha \\ +\beta_1 \left[1 + \exp\left\{.5 \times \left[r_{t-1} - VaR_{\alpha,t-1}^{A}(r_{t-1},\beta)\right]\right\}\right]^{-1}. \end{aligned} \tag{17}$$

The intuition associated to the Adaptive specification is the following: as long as the daily return is not inferior to the VaR estimation, the VaR can increase by a small amount; on the contrary, when the VaR is exceeded, the VaR has to decrease. Thus this model is adapted to larger and larger falls of the risky asset. This model adapts itself to its 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 be considered as quantile functions too. 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):

$$\widehat{\beta}^{*} = \underset{\widehat{\beta}\in\mathbb{R}^{n}}{ArgMin}\left\{\sum_{t=1}^{T}\left\{\left[\alpha - 1\mathrm{I}_{\left\{r_t < -\widehat{VaR}_{\alpha,t}\left(r_{t-1},\widehat{\beta}\right)\right\}}\right]\left[r_t + \widehat{VaR}_{\alpha,t}\left(r_{t-1},\widehat{\beta}\right)\right]\right\}\right\}, \qquad (18)$$

with $\widehat{VaR}_{\alpha,t}\left(r_{t-1},\widehat{\beta}\right)$ is the VaR estimation according to CAViaR models (see equations 14 to 17), $\widehat{\beta}$ is a vector of estimated parameters and $1\mathrm{I}_{\{.\}}$ is the indicator function.

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, $\widehat{VaR}_{\alpha,t}(\alpha)$, 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 (see Kuan *et al.*, 2008), replacing the conditional VaR by the conditional expectile in equations (14), (15), (16) and (17). 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 20). The τ-expectile, for which the proportion α of in-sample observations below the expectile, is the estimator of the α-quantile (see 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 focus on these two main points to answer this issue. On 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 by the minimization of:

$$\widehat{\mu}^*_{\tau,t} = \underset{\widehat{\mu}_{\tau,t}\in\mathbb{R}}{ArgMin}\left\{E\left[\left|\tau - 1\mathrm{I}_{\{r_t<\widehat{\mu}_{\tau,t}\}}\right|(r_t-\widehat{\mu}_{\tau,t})^2\right]\right\}, \tag{19}$$

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

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:

$$\widehat{\beta}^{*} = \underset{\widehat{\beta}\in\mathbb{R}^n}{ArgMin}\left\{\sum_{t=1}^{T}\left\{\left|\tau - 1\mathrm{I}_{\left\{r_t<\widehat{\mu}_{\tau,t}\left(r_{t-1};\widehat{\beta}\right)\right\}}\right| \times \left[r_t - \widehat{\mu}_{\tau,t}\left(r_{t-1};\widehat{\beta}\right)\right]^2\right\}\right\}, \qquad (20)$$

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.

We can use expectiles as quantile estimators (and VaR) given 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 previous notations):

$$\tau = \frac{\alpha\left(-VaR_{\alpha,t}\right) \; - \; \int_{-\infty}^{-VaR_{\alpha,t}} \; r_t \, df(r_t)}{E(r_t) \; - \; 2\int_{-\infty}^{-VaR_{\alpha,t}} \; r_t \, df(r_t) \; + \; \left(VaR_{\alpha,t}\right)(1-2\alpha)}, \qquad (21)$$

where, $f(.)$ is the return Probability Density Function.
Using expectile to estimate risk measures, the confidence level α is not determined *à priori* 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.* (2008) explain the link between the Expectile and Expected Shortfall, that leads to (with previous notations):

$$ES_{\alpha,t} = \left[1 + \tau\left(1-2\tau\right)^{-1}\alpha^{-1}\right]\mu_{\tau,t}\left(r_t;\beta\right). \qquad (22)$$

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:

$$VaR_{\alpha,t} = \sum_{k=1}^{K} a_k\left(r_{t-1}, y_{t-1}, \beta_k\right) \times VaR_{\alpha,t}^{(k)}\left(\beta_k\right), \qquad (23)$$

where $VaR_{\alpha,t}^{(k)}(\cdot)$ are a finite number of 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, DAQ model can use different quantile functions to model a given one. We can also combine these functions into a multi-quantile method (see 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 can thus be expressed as a simple combination of quantiles associated to the same probability:

$$\underset{[1\times 1]}{VaR_{\alpha,t}} = \underset{[1\times K]}{\mathbf{W}'_{\alpha,t}} \underset{[K\times 1]}{\mathbf{VaR}_{\alpha,t}}, \tag{24}$$

with:

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

and:

$$\begin{cases} VaR^{(k)}_{\alpha,t} = VaR^{(k)}_{\alpha,t}\left(\mu_{\tau_k}, \beta_k\right) \\ \sum_{k=1}^{K} w_k = 1, \end{cases}$$

where $VaR^{(k)}_{\alpha,t}(.)$ are a finite number of quantile functions associated with probability α and related to the model k, with $k = [1, 2, ..., K]$.

The Expected Shortfall can therefore be expressed as a combination of quantiles whose associated probabilities are defined through the estimation of equation (20):

$$\underset{[1\times 1]}{ES_{\alpha,t}} = \underset{[1\times K]}{\mathbf{W}'_{\alpha,t}} \underset{[K\times 1]}{\mathbf{VaR}_{\gamma,t}}, \tag{25}$$

with:

$$\begin{cases} \mathbf{W}_{\alpha,t} = \begin{bmatrix} w_1 \times b_1 & w_2 \times b_2 & ... & w_K \times b_K \end{bmatrix}' \\ \mathbf{VaR}_{\gamma,t} = \begin{bmatrix} VaR^{(1)}_{\gamma_1,t} & VaR^{(2)}_{\gamma_2,t} & ... & VaR^{(K)}_{\gamma_K,t} \end{bmatrix}', \end{cases}$$

and:

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

where $VaR^{(k)}_{\gamma_k,t}(.)$ are a finite number of quantile functions associated with probability γ_k and related to the model k, with $k = [1, 2, ..., K]$.

The estimation of the α-Expected Shortfall can be found by aggregating linearly K quantile functions and estimating the right correspondence between the probability α associated to the Expected Shortfall of each model (denoted $ES^{(k)}_{\alpha}$) and the probability γ_k associated to each quantile functions (thanks to the τ_k-expectile defined by the equation 19).

Artzner *et al.* (1999) call a risk measure coherent, if it verifies the following properties: first order stochastic monotonicity, positive homogeneity, sub-additivity and translation invariance. The ES respects these *criteria* (see Acerbi, 2002). The class of coherent risk measures is convex (see Pflug, 2000). 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) 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 acting 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. As long as the weighting function of the DARE approach respects risk spectrum *criteria*, such as (with previous notations):

$$\begin{cases} \sum_{k=1}^{K} w_k = 1 \\ w_k \geq 0 \\ w_i \geq w_j \text{ with } 0 \leq i \leq j \leq K. \end{cases} \tag{26}$$

This new space of coherent measures allows us to define a coherent spectral measure of risk. According to equation (26), several conditional weighting functions can be applied to the risk models aggregated into the DARE approach (see Hansen, 2008). In order to select dynamically the more appropriated quantile models for the extremal risk measure estimation, the DARE weighting function can be defined as an increasing function of the main evaluation tests results developed in the literature.

4 Backtesting Extreme Risk Measures

The DARE approach for extremal risk measure can be interpreted as a specific combination of quantile. 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.* (2009).

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.*, 2008) or subsampling techniques (Escanciano and Olmo, 2009).

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:

$$I_t(\alpha) = \begin{cases} 1 & \text{if } r_t < -VaR_{\alpha,t-1} \\ 0 & \text{otherwise,} \end{cases} \tag{27}$$

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 $t = [1, \dots, T]$, for a given α-quantile.

4.1 Unconditional and Conditional Exceptions 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.

Therefore, Kupiec (1995) shows therefore that if the VaR model is well specified then the exceptions occurred, can be modeled as independent draws from a binomial distribution with a probability of occurrence equal to α-percent.

The Likelihood Ratio statistic is:

$$LR_{uc} = 2\{\log[\widehat{\alpha}^{N_I}(1 - \widehat{\alpha}^{T-N_I})] - \log[\alpha^{N_I}(1 - \alpha^{T-N_I})]\}, \tag{28}$$

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.

As shown by Kupiec (1995), with the hypothesis of $\widehat{\alpha} = \alpha$, the Likelihood Ratio LR_{uc} has an asymptotic distribution such as:

$$LR_{uc} \sim> \chi^2(1), \tag{29}$$

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

The Unconditional Coverage test for the independence of the exceptions $(E[I_t(\alpha)] = \alpha)$ can also be determined by the following test statistic (Kupiec, 1995; Escanciano and Olmo, 2009), using 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.

Escanciano and Olmo (2009) 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 test resampling statistic for the unconditional coverage hypothesis is given by (using previous notations):

$$K_S \equiv S^{-1/2} \sum_{t=U+1}^{T} \left[I_t(\alpha) - \alpha \right], \tag{31}$$

with U being the in-sample size, S the corresponding out-of-sample testing period, and T is the total sampling size.

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, see, *e.g.*, Hahn (1995), Horowitz (1998) or He and Hu (2002).

The subsampling methodology to approximate the critical values of tests is based on K_S. Henceforth, we use $G_{K_S}(w)$ denotes the Cumulative Distribution Function of the test statistic given by (using previous notations):

$$G_{K_S}(w) = Prob(K_S \leq w), \tag{32}$$

for any w.

Let $K^b_{S,t} = K_S(r_t, r_{t+1}, \cdots, r_{t+b-1})$, 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 (2009) approximate the sampling distribution $G_{K_S}(w)$ using the distribution of the values of $K^b_{S,t}$ computed over the $(T-b+1)$ different consecutive subsamples of size b. This subsampling distribution is defined by (using previous notations):

$$G_{K^b_S}(w) = (T-b+1)^{-1} \sum_{t=1}^{T-b+1} 1\mathrm{I}_{\{K^b_{S,t} \leq w\}}. \tag{33}$$

Let $c_{K^b_S,1-\tau}$ be the $(1-\tau)$-th sample quantile of $G_{K^b_S}(w)$, *i.e.*:

$$c_{K^b_S,1-\tau} = \inf_w \left\{ w | \; G_{K^b_S}(w) \geq 1-\tau \right\}. \tag{34}$$

The rejection region in this case is determined by $K_S > c_{K^b_S,1-\tau}$.

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 modeled with a Markov chain with the following transition probabilities matrix:

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

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:

$$LR_{ind} = 2[\log L(\pi_{01}, \pi_{11}) - \log L(\pi, \pi)], \tag{35}$$

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 previous notations):

$$LR_{ind} \sim> \chi^2(1). \tag{36}$$

Since for Bernoulli random variables serial independence is equivalent to serial uncorrelation, it is also possible to build a test based on the autocovariances (see Candelon et *et al.*, 2008; Escanciano and Olmo, 2009, Hurlin and Tokpavi, 2007 and 2008). The proper marginal test of independence is based on (using previous notations):

$$\zeta_{S,j} = (S - j)^{-1/2} \sum_{t=U+j+1}^{T} \{I_t(\alpha) - E_T[I_t(\alpha)]\} \{I_{t-j}(\alpha) - E_T[I_{t-j}(\alpha)]\}, \tag{37}$$

with:

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

If $G_{\zeta_S}(w)$ denotes the cumulative distribution function of the test statistic, then (using previous notations):

$$G_{\zeta_S}(w) = Prob(\zeta_S \leq w), \tag{38}$$

for any w.

If $\zeta^b_{S,t} = \zeta_S(r_t, r_{t+1}, \cdots, r_{t+b-1})$, $t = [1, 2, \cdots, T - b + 1]$, be the test statistic computed with the subsample $(r_t, r_{t+1}, \cdots, r_{t+b-1})$ of size b, the subsampling distribution is defined by (using previous notations):

$$G_{\zeta_S^b}(w) = (T-b+1)^{-1} \sum_{t=1}^{T-b+1} 1\mathrm{I}_{\{\zeta_{S,t}^b \leq w\}}, \tag{39}$$

and the $(1-\tau)$-th sample quantile of $G_{\zeta_S^b}(w)$, $c_{\zeta_S^b,1-\tau}$, is given by:

$$c_{\zeta_S^b,1-\tau} = \inf_w \left\{ w | \, G_{\zeta_S^b}(w) \geq 1-\tau \right\}. \tag{40}$$

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 previous notations):

$$LR_{cc} = LR_{uc} + LR_{ind}. \tag{41}$$

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

$$LR_{cc} \sim> \chi^2(2). \tag{42}$$

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 (see 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 (43), should be identically distributed with the unconditional mean and the standard deviation (0,1) such as:

$$R_t^G(\alpha) = \begin{cases} \breve{r}_t \text{ if } \breve{r}_t < \Phi^{-1}(\alpha) \\ 0 \quad \text{otherwise} \end{cases}, \tag{43}$$

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

$$LR_{mag} = 2[L_{mag}(\mu, \sigma) - L_{mag}(0, 1)], \tag{44}$$

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 previous notations):

$$LR_{mag} \sim> \chi^2(2). \tag{45}$$

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:

$$h_t(\alpha) = \begin{cases} 1 - \alpha \text{ if } r_t < -VaR_{\alpha,t-1} \\ -\alpha \quad \text{otherwise} \end{cases}, \tag{46}$$

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

$$h_t(\alpha) = I_t(\alpha) - \alpha.$$

Let $\Pi = \{\pi_1, \dots, \pi_m\} \subset [0,1]$ be a discrete set of m different coverage rates. Let $H_t = [h_t(\pi_1), h_t(\pi_2), \dots, h_t(\pi_m)]'$ 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:

$$Cov(H_t, H_{t-l}) = E[H_t \; H'_{t-l}]. \tag{47}$$

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

$$H_0 : E[h_t(\pi_i) h_{t-l}(\pi_j)] = 0, \tag{48}$$

for the lag $l = [1, \dots, T]$ 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 covariance associated to vector H_t such as (with previous notations):

$$\mathbf{\Omega}_l = \sum_{t=l+1}^{T} H_t \; H'_{t-l}, \tag{49}$$

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

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

$$Q_m(T) = T^2 \sum_{l=1}^{T} (T-l)^{-1} Diag(\mathbf{\Omega}'_l \mathbf{\Omega}_0^{-1} \mathbf{\Omega}_l \mathbf{\Omega}_0^{-1})' \mathbf{1}. \tag{50}$$

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 previous notations):

$$Q_m(T) \sim> \chi^2(Tm^2). \tag{51}$$

4.4 Generalized Method of Moment Duration-based Tests (Candelon *et al.*, 2008)

The Generalized Method of Moment (GMM) Duration-based tests, proposed by Candelon *et al.* (2008), 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.

A GMM Duration-based Conditional Coverage Test (Candelon *et al.*, 2008)

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_i for $i = [1, \cdots, N_I]$, observed between two successive violations associated to the α% VaR forecasts:

$$d_i^\alpha = t_i^\alpha - t_{i-1}^\alpha, \tag{52}$$

with t_i^α is the date of the i^{th} violation for the level α.

Under the conditional coverage assumption, the duration d_i^α, 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.* (2008) define a sequence of moment conditions. $M(d_i^\alpha)$ 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 $M(d_i^\alpha)$ are denoted $M_j(d_i^\alpha)$ such as, for $j = [1, \ldots, p]$:

$$\begin{aligned} M_{j+1}(d_i^\alpha) = \quad & [(1-\alpha)(2j+1) + \alpha(j - d_i^\alpha + 1)] \left[(j+1)(1-\alpha)^{1/2}\right]^{-1} \\ & \times M_j(d_i^\alpha) - j(j+1)^{-1} M_{j-1}(d_i^\alpha), \end{aligned} \tag{53}$$

with $M_1(d_i^\alpha) = 0$ and $M_0(d_i^\alpha) = 1$.

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

$$\begin{cases} H_0 : E[M(d_i^\alpha)] = 0 \\ H_1 : E[M(d_i^\alpha)] \neq 0. \end{cases} \tag{54}$$

As shown in Candelon *et al.* (2008), $M_j(d_i^\alpha)$ is thus asymptotically distributed such as (with previous notations):

$$\left(N^{-1/2}\sum_{i=1}^{N} M_j(d_i^\alpha)\right)^2 \xrightarrow[N\to\infty]{d} \chi^2(1), \tag{55}$$

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

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

$$J_{CC}(p) = \left(N^{-1/2}\sum_{i=1}^{N} M(d_i^\alpha)\right)' \left(N^{-1/2}\sum_{i=1}^{N} M(d_i^\alpha)\right) \xrightarrow[N\to\infty]{d} \chi^2(p), \tag{56}$$

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

A GMM Duration-based Unconditional Coverage Test (Candelon *et al.*, 2008)

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 $M(d_i^\alpha) = M_1(d_i^\alpha)$. The statistic for Unconditional Coverage (UC) test, denoted J_{UC}, is equivalent to $J_{CC}(p)$ with $p = 1$ (using previous notations):

$$J_{UC} = \left(N^{-1/2}\sum_{i=1}^{N} M_1(d_i^\alpha)\right) \xrightarrow[N\to\infty]{d} \chi^2(1). \tag{57}$$

A GMM Duration-based Independence Test (Candelon *et al.*, 2008)

Candelon *et al.* (2008) 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 previous notations):

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

where $M(d_i^\pi)$ denotes a $(p \times 1)$ vector whose components are the orthonormal polynomials $M_j(d_i^\pi)$ 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:

$$h_t(\alpha) = \delta_0 + \sum_{l=1}^{L} \delta_l h_{t-l}(\alpha) + \delta_{L+1} VaR_{\alpha,t-1} + \epsilon_t, \tag{58}$$

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 (58) 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 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 (see 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. Wong (2008) proposes the Saddlepoint Technique (see 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:

$$\begin{aligned} Prob(R_t \leq r) &= Prob(r_t \leq r | r_t \leq -VaR^G_{\alpha,t}), \\ &= Prob(r_t \leq r)/Prob(r_t \leq -VaR^G_{\alpha,t}), \end{aligned} \tag{59}$$

where $r \leq -VaR^G_{\alpha,t}$ and $VaR^G_{\alpha,t} = -\Phi^{-1}(\alpha)$. VaR^G is thus the VaR assuming a Gaussian distribution of the returns.

Thus R_t refers to the portfolio return conditional on $r_t \leq -VaR^G_{\alpha,t}$ 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 $-VaR^G_{\alpha,t}$, the estimated Expected Shortfall can be written such as:

$$ES_{\alpha,T} = -T^{-1} \sum_{t=1}^{T} R_t. \tag{60}$$

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

$$M(r) = \alpha^{-1} exp(r^2/2) \times \Phi(-VaR_{\alpha,t} - r), \tag{61}$$

and its derivatives with respect to r are given by:

$$\begin{cases} M^{(1)}(r) = r \times M(r) - exp(-VaR_{\alpha,t} \times r) \times \alpha^{-1}\phi(-VaR_{\alpha,t}) \\ M^{(2)}(r) = r \times M'(r) + M(r) - exp(-VaR_{\alpha,t} \times r) \times (-VaR_{\alpha,t}) \\ \qquad\qquad \times \alpha^{-1}\phi(-VaR_{\alpha,t}) \\ \vdots \\ M^{(n)}(r) = r \times M^{(n-1)}(r) + (n-1)M^{n-2}(r) - exp(-VaR_{\alpha,t} \times r) \\ \qquad\qquad \times(-VaR_{\alpha,t})^{(n-1)} \times \alpha^{-1}\phi(-VaR_{\alpha,t}), \end{cases} \tag{62}$$

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

In order to define the null and alternative hypotheses of the test, we need to define the mean and variance of R_t. Posing $r = 0$ in $M^{(1)}(r)$ and $M^{(2)}(r)$, 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 (see Wong, 2008):

$$\mu_{R_t} = -\alpha^{-1}\phi(-VaR^G_{\alpha,T}), \tag{63}$$

and can thus be expressed such as:

$$\mu_{R_t} = -\left(\alpha\sqrt{2\pi}\right)^{-1} exp[-(VaR^G_{\alpha,T})^2/2]. \tag{64}$$

Its variance is:

$$\begin{aligned} \sigma^2_{R_t} &= 1 + \alpha^{-1}VaR^G_{\alpha,T}\phi(-VaR^G_{\alpha,T}) - \mu^2_{R_t}, \\ &= 1 + \left(\alpha\sqrt{2\pi}\right)^{-1} VaR^G_{\alpha,T}exp[-(VaR^G_{\alpha,T})^2/2] - \mu^2_{R_t}. \end{aligned} \tag{65}$$

Following the central limit theorem, $\sqrt{T}\sigma^{-1}_{R_t}(\bar{R}_t - \mu_{R_t})$ is asymptotically standard normal, with $\bar{R}_t$ 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 therefore written as:

$$\begin{cases} H_0 : ES_{\alpha,T} = -\alpha^{-1}\phi(-VaR^G_{\alpha,T}) \\ H_1 : ES_{\alpha,T} > -\alpha^{-1}\phi(-VaR^G_{\alpha,T}). \end{cases} \tag{66}$$

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

$$p\text{-value} = Prob(\bar{R}_t \leq \bar{r}), \tag{67}$$

where $\bar{r}$ 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 $\bar{r}$ given by:

$$Prob(\bar{R}_t \leq \bar{r}) = \begin{cases} \Phi(\varsigma) - \phi(\varsigma)\left[\frac{1}{\eta} - \frac{1}{\varsigma} + \mathbf{O}(T^{-3/2})\right] & \text{for } \bar{r} < -VaR_{\alpha,t} \\ 1 & \text{for } \bar{r} \geq -VaR_{\alpha,t}, \end{cases} \tag{68}$$

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

$$K(\varpi) = ln[M(\varpi)],$$

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

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

5 Gauging the DARE Risk Measures

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 9^{th} July 1987 until 18^{th} 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.

Although there are many VaR estimation methods, we restricted our comparison to commonly used benchmark methods.
We first consider the most widely used non-parametric method: the historical simulation. 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.

- Insert Figures 1 and 2 here -

CAViaR models are presented as benchmark 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 (22). 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).

- Insert Table 1 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.*, 2008), 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 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 (2008), the independent *criterion* cannot be rejected on this data sample. The framework of the GMM duration based test (Candelon *et al.*, 2008), 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 inapropriate 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 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 DARE approach respects most of these tests.

In the proposed 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.

- *Insert Table 2 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 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 DARE approach for ES seems to be particularly attractive for extreme probabilities associated to the ES.

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 Preliminary 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 every 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 by the p-values of backtests procedures), 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(2)*, 371-384.

Acerbi C., (2002), "Spectral Measures of Risk: A Coherent Representation of Subjective Risk Aversion", *Journal of Banking and Finance 26(7)*, 1505-1518.

Acerbi C. and D. Tasche, (2002), "On the Coherence of Expected Shortfall", *Journal of Banking and Finance 26(7)*, 1487-1503.

Artzner P., F. Delbaen, J. Eber and D. Heath, (1999), "Coherent Measures of Risk", *Mathematical Finance 9(3)*, 203-228.

Bali T., (2003), "An Extreme Value Approach to estimating Volatility and Value-at-Risk", *Journal of Business 76(1)*, 83-108.

Barndorff-Nielsen O., (1998), "Process of Normal Inverse Gaussian Type", *Finance and Stochastics 2(1)*, 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(5)*, 12-24.

Berkowitz J., (2001), "Testing Density Forecasts with Applications to Risk Management", *Journal of Business and Economics Statistics19(4)*, 465-474.

Berkowitz J. and J. Brien, (2002), "How Accurate are the Value-at-Risk Models at Commercial Banks?", *Journal of Finance 57(3)*, 1093-1111.

Berkowitz J., P. Christoffersen and D. Pelletier, (2009), "Evaluating Value-at-Risk Models with Desk-level Data", *Management Science*, forthcoming, 33 pages.

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

Bontemps C. and N. Meddahi, (2005), "Testing Normality: A GMM Approach", *Journal of Econometrics 124(1)*, 149-186.

Candelon B., G. Colletaz, C. Hurlin and S. Tokpavi, (2008), "Backtesting Value-at-Risk: A GMM Duration-based Test", *Working Paper*, 40 pages.

Cheridito P. and M. Stadje, (2009), "Time-inconsistency of VaR and Time-consistent Alternatives", *Finance Research Letters 6(1)*, 40-46.

Christoffersen P., (1998), "Evaluating Interval Forecasts", *International Economic Review 39(4)*, 841-862.

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

Cornish E. and R. Fisher, (1937), "Moments and Cumulants in the Specification of Distributions", *Review of the International Statistical Institute 5(4)*, 307-320.

Daniels H., (1954), "Saddlepoint Approximations in Statistics", *The Annals of Mathematical Statistics 25(4)*, 631-650.

Daniels H., (1987), "Tail Probability Approximations", *International Statistical Review 55(3)*, 37-48.

Diebold F., T. Gunther and A. Tay, (1998), "Evaluating Density Forecasts with Applications to Financial Risk Management", *International Economic Review 39(3)*, 863-883.

Efron B., (1991), "Regression Percentiles using Asymmetric Squared Error Loss", *Statistica Sinica 1(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(4)*, 367-381.

Escanciano J. and J. Olmo, (2009), "Backtesting Parametric Value-at-Risk with Estimation Risk", *Journal of Business and Economic Statistics*, forthcoming, 48 pages.

Ferreira M. and J. Lopez, (2005), "Evaluating Interest Rate Covariance Models within a Value-at-Risk Framework", *Journal of Financial Econometrics 3(1)*, 126-168.

Gouriéroux C. and J. Jasiak, (2008), "Dynamic Quantile Models", *Journal of Econometrics 147(1)*, 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(4)*, 277-298.

Haas M., (2005), "Improved Duration-based Backtesting of Value-at-Risk", *Journal of Risk 8(2)*, 17-38.

Hahn J., (1995), "Bootstrapping Quantile Regression Estimators", *Econometric Theory 11(1)*, 105-121.

Hansen B., (2008), "Least-squares Forecast Averaging", *Journal of Econometrics 146(2)*, 342-350.

He X. and F. Hu, (2002), "Markov Chain Marginal Bootstrap", *Journal of the American Statistical Association 97(459)*, 783-795.

Hendricks D., (1996), "Evaluation of Value-at-Risk Models using Historical Data", *Economic Policy Review 2(1)*, 39-69.

Horowitz J., (1998), "Bootstrap Methods for Median Regression Models", *Econometrica 66(6)*, 1327-1351.

Hosking J., (1980), "The Multivariate Portmanteau Statistic", *Journal of the American Statistical Association 75(371)*, 602-608.

Hurlin C. and S. Tokpavi, (2007), "Un test de validité de la Value-at-Risk", *Revue Economique 58(3)*, 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(1)*, 53-80.

Jones M., (1994), "Expectiles and M-quantiles are Quantiles", *Statistics and Probability Letters 20(2)*, 149-153.

Jorion P., (2006), *Value-at-Risk: The New Benchmark For Managing Financial Risk*, 3^{rd} ed, McGraw-Hill, 600 pages.

Koenker R. and G. Bassett, (1978), "Regression Quantiles", *Econometrica 46(1)*, 33-50.

Koenker R. and Z. Xiao, (2006), "Quantile Autoregression", *Journal of the American Statistical Association 101(475)*, 980-990.

Koenker R. and Z. Xiao, (2009), "Conditional Quantile Estimation for GARCH Models", *Working Paper*, 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(2)*, 261-270.

Kupiec P., (1995), “Techniques for Verifying the Accuracy of Risk Measurement Models”, *Journal of Derivatives 3(2)*, 73-84.

Lugannani R. and S. Rice, (1980), “Saddlepoint Approximation for the Distribution of the Sum of Independent Random Variables”, *Advanced in Applied Probability 12(2)*, 475-490.

Mittnik S. and M. Paolella, (2000), “Conditional Density and Value-at-Risk Prediction of Asian Currency Exchange Rates”, *Journal of Forecasting 19(4)*, 313-333.

Newey W. and J. Powell, (1987), “Asymmetric Least Squares Estimation and Testing”, *Econometrica 55(4)*, 819–847.

Nieto M. and E. Ruiz, (2008), “Measuring Financial Risk: Comparison of Alternative Procedures to estimate VaR and ES”, *Working Paper*, 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.

Poon S. and C. Granger, (2003), “Forecasting Volatility in Financial Markets: A Review”, *Journal of Economic Literature 41(2)*, 478-639.

Taylor J., (2008a), “Estimating Value-at-Risk and Expected Shortfall using Expectiles”, *Journal of Financial Econometrics 6(2)*, 231-252.

Taylor J., (2008b), “Using Exponentially Weighted Quantile Regression to Estimate Value-at-Risk and Expected Shortfall”, *Journal of Financial Econometrics 6(3)*, 382-406.

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(7)*, 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(1)*, 57-85.

Yao Q. and H. Tong, (1996), “Asymmetric Least Squares Regression Estimation: A Nonparametric Approach”, *Journal of Nonparametric Statistics 6(2)*, 273-292.

Table 1: P-statistics related to Null Hypotheses corresponding to a "Good" Property of a VaR 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	**.29*****	**.00*****	**.00*****	**.25*****	**.00*****	**.00*****	100.00	**.11*****
Normal	**.00*****	100.00	100.00	**.00*****	**.00*****	21.44	**.00*****	**.00*****	**.19*****	**.00*****
Student	64.68	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*****
Normal GARCH(1,1)	53.47	100.00	100.00	82.47	**.00*****	52.43	56.37	37.27	99.89	99.70
RiskMetrics	**.00*****	100.00	100.00	**.00*****	**.00*****	17.40	44.14	40.41	98.11	100.00
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	93.28	100.00	99.99	99.64	**.00*****	89.29	**.01*****	**.00*****	100.00	**.00*****
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*****	19.35	**.00*****	**.00*****	**.16*****	**.00*****	**.00*****	99.40	**.00*****
Normal	**.00*****	**.00*****	9.69*	**.00*****	**.00*****	**.00*****	**.00*****	**.00*****	94.33	**.00*****
Student	**.01*****	**.94*****	13.82	**.00*****	**.00*****	**.07*****	**.00*****	**.00*****	99.49	**.00*****
NIG	**.02*****	**.12*****	18.02	**.00*****	**.00*****	**.15*****	**.00*****	**.00*****	99.44	**.00*****
Normal GARCH(1,1)	**.07*****	2.82**	22.99	**.03*****	**.00*****	**.39*****	12.06	47.28	99.99	100.00
RiskMetrics	**.01*****	**.84*****	17.13	**.00*****	**.00*****	**.07*****	2.97*	61.88	99.95	100.00
CAViaR Symmetric Absolute Value	5.02*	5.09*	14.33	2.19**	**.00*****	8.61*	5.59*	7.37*	100.00	27.16
CAViaR Asymmetric Slope	**.16*****	2.30**	21.41	**.05*****	**.00*****	**.64*****	**.02*****	**.11*****	99.88	99.99
CAViaR Indirect GARCH	32.22	15.22	5.36*	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	6.84*	4.64**	12.72	2.61**	**.00*****	10.03	**.00*****	**.00*****	100.00	**.00*****
DARE	25.81	2.81**	6.76*	4.73**	**.00*****	32.17	12.98	9.32*	99.99	64.63

Source: DataStream; daily data from July 9^{th} 1987 until March 18^{th} 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 resepectively 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 (2009) 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.*, 2008) 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.

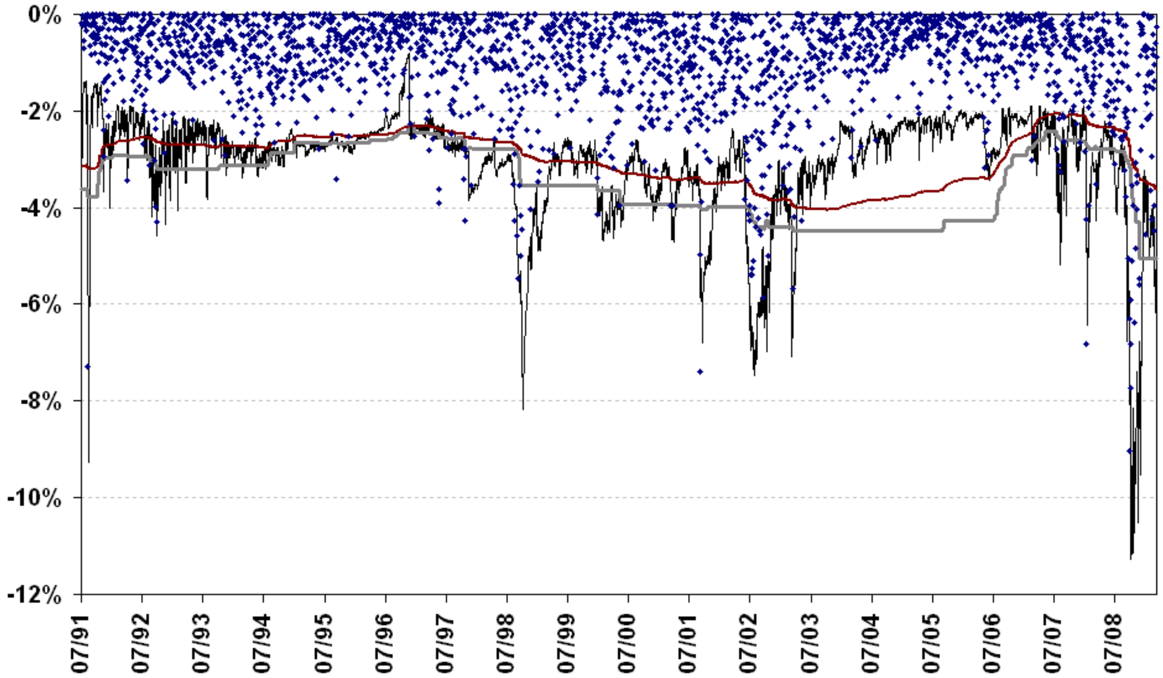
Table 2: P-statistics related to Null Hypotheses
corresponding to a "Good" Property of an Expected Shortfall Model

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***	10.01	.00***	.00***	.93***	.00***	.00***	90.11	.00***
Normal	.00***	.00***	.01***	.00***	.00***	.00***	.00***	.00***	66.35	.00***
Student	2.48**	.00***	10.41	.00***	.00***	4.18**	.00***	.00***	90.32	.00***
NIG	.01***	.00***	9.14*	.00***	.00***	.06***	.00***	.00***	88.32	.00***
Normal GARCH(1,1)	.48***	16.42	8.90*	.71***	.00***	1.15**	25.94	90.79	99.73	100.00
RiskMetrics	.19***	19.72	10.55	.35***	.00***	.57***	14.86	48.87	99.76	100.00
CARE Symmetric Absolute Value	3.46**	4.14**	17.43	1.34**	.00***	2.28**	.00***	4.47**	99.98	3.80**
CARE Asymmetric Slope	18.03	7.56*	9.44*	8.41*	.00***	18.29	.00***	.23***	99.82	16.93
CARE Indirect GARCH	1.43**	3.12**	19.54	.49***	.00***	.74***	.42***	72.27	99.98	95.26
CARE Adaptive	.74***	.45***	18.77	.05***	.00***	.68***	.00***	.08***	99.94	.00***
Naive Mean of ES	42.31	3.02**	9.45*	6.93*	.00***	40.59	.00***	.00***	100.00	.00***
DARE	4.54**	4.54**	13.54	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**	10.93	.00***	.00***	.00***	.00***	.00***	99.47	.00***
Student	.00***	1.59**	2.09	.00***	.00***	.01***	.00***	.00***	99.95	.00***
NIG	.01***	.63***	.20***	.00***	.00***	.16***	.00***	.00***	99.97	.00***
Normal GARCH(1,1)	.00***	1.41**	1.37**	.00***	.00***	.01***	.08***	48.40	100.00	100.00
RiskMetrics	.00***	1.24**	.83***	.00***	.00***	.02***	.21***	86.51	99.99	100.00
CARE Symmetric Absolute Value	62.15	69.18	36.13	81.84	.00***	49.63	56.34	54.64	100.00	.11***
CARE Asymmetric Slope	16.04	1.11**	1.59**	1.48**	.00***	21.64	9.80*	7.75*	100.00	1.24**
CARE Indirect GARCH	78.72	67.65	21.50	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	42.07	70.73	36.15	71.84	.00***	38.87	6.09*	16.18	100.00	.00***
DARE	96.02	66.12	15.13	90.73	.00***	84.21	87.83	78.35	100.00	50.61

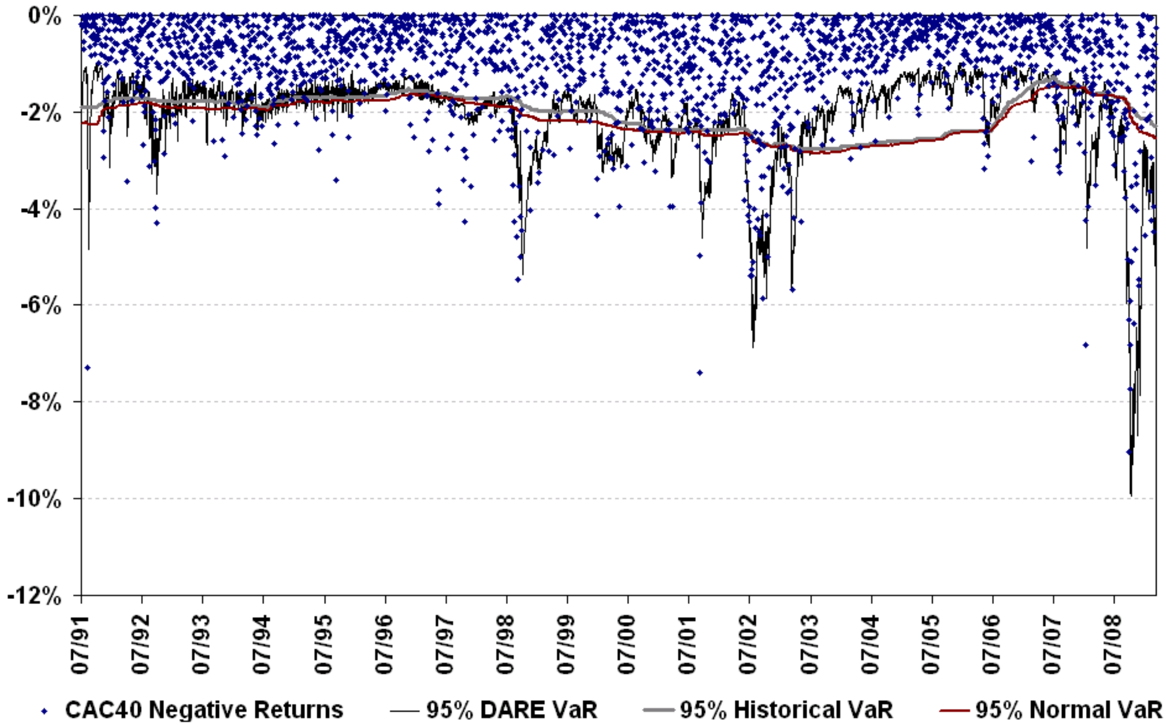
Source: DataStream; daily data from July 9^{th} 1987 until March 18^{th} 2009 of the CAC40 Index; computations by the authors. This period contains 5,659 daily returns. We use a moving window of four years (1,044 daily returns) to dynamically re-estimate parameters for the various methods. Forecasted 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 resepectively 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 (2009) 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.*, 2008) 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



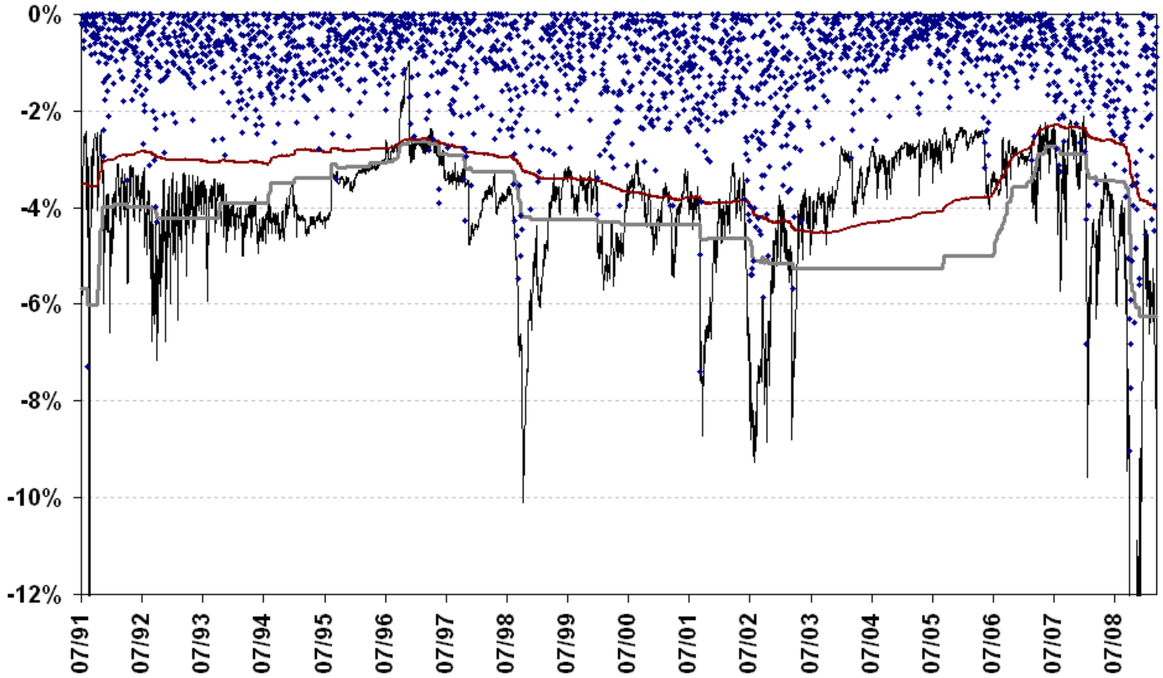
95% Value-at-Risk



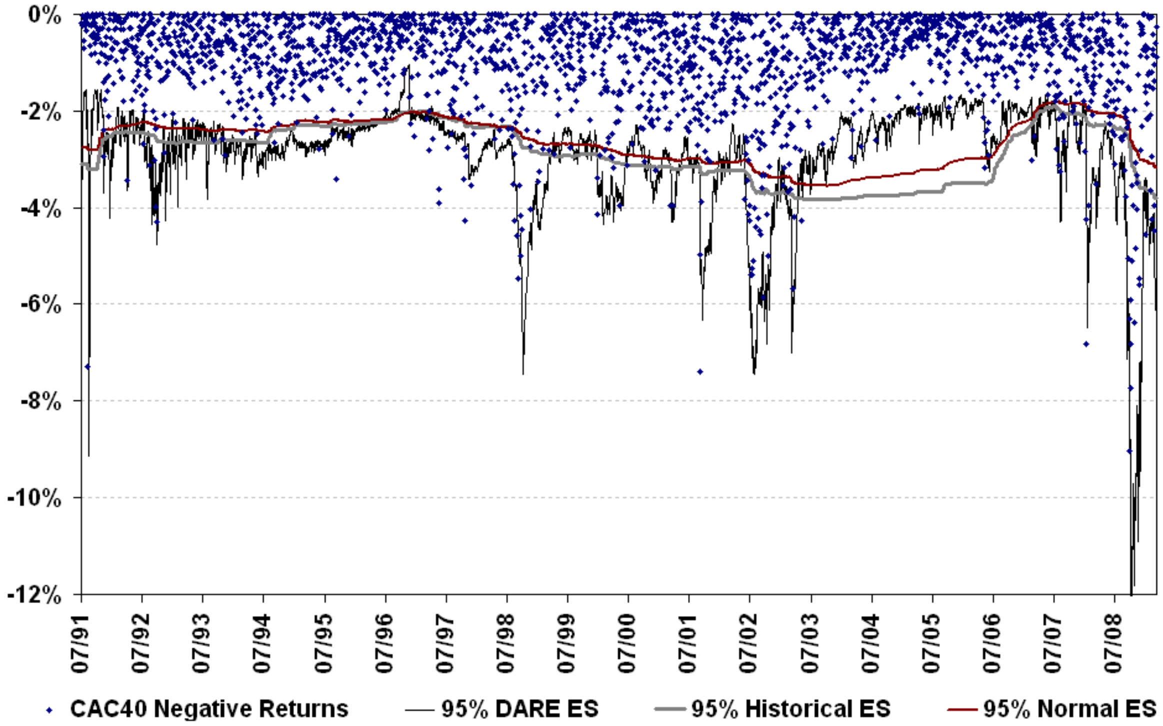
Source: *DataStream*; daily data from July 9^{th} 1987 until March 18^{th} 2009 of the CAC40 Index; computations by the authors. This period contains 5,659 daily returns. We use a moving window of four years (1,044 daily returns) to dynamically re-estimate parameters for the various methods.

Figure 2: 99% and 95% Expected Shortfall Forecasts

99% Expected Shortfall



95% Expected Shortfall



Source: *DataStream*; daily data from July 9^{th} 1987 until March 18^{th} 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.