A Stochastic Dominance Approach to the Basel III Dilemma: Expected Shortfall or VaR?*

Chia-Lin Chang
Department of Applied Economics
Department of Finance
National Chung Hsing University
Taichung, Taiwan

Juan-Ángel Jiménez-Martín
Department of Quantitative Economics and
Complutense Institute of Economic Analysis (ICAE)
Complutense University of Madrid, Spain

Esfandiar Maasoumi
Department of Economics
Emory University, USA

Michael McAleer
Department of Quantitative Finance
National Tsing Hua University, Taiwan
and
Econometric Institute, Erasmus School of Economics
Erasmus University Rotterdam
and
Tinbergen Institute, The Netherlands
and
Department of Quantitative Economics
Complutense University of Madrid, Spain

Teodosio Pérez-Amaral
Department of Quantitative Economics and
Complutense Institute of Economic Analysis (ICAE)
Complutense University of Madrid, Spain

EI2015-14
Revised: May 2015

* The first author wishes to thank the National Science Council, Taiwan, the second and fifth authors acknowledge the Ministerio de Ciencia y Tecnología of Spain through the research project ECO2012-31941 and Comunidad de Madrid, and the fourth author is grateful to the National Science Council, Taiwan and the Australian Research Council.

Abstract

The Basel Committee on Banking Supervision (BCBS) (2013) recently proposed shifting the quantitative risk metrics system from Value-at-Risk (VaR) to Expected Shortfall (ES). The BCBS (2013) noted that "a number of weaknesses have been identified with using VaR for determining regulatory capital requirements, including its inability to capture tail risk" (p. 3). For this reason, the Basel Committee is considering the use of ES, which is a coherent risk measure and has already become common in the insurance industry, though not yet in the banking industry. While ES is mathematically superior to VaR in that it does not show "tail risk" and is a coherent risk measure in being subadditive, its practical implementation and large calculation requirements may pose operational challenges to financial firms. Moreover, previous empirical findings based only on means and standard deviations suggested that VaR and ES were very similar in most practical cases, while ES could be less precise because of its larger variance. In this paper we find that ES is computationally feasible using personal computers and, contrary to previous research, it is shown that there is a stochastic difference between the 97.5% ES and 99% VaR. In the Gaussian case, they are similar but not equal, while in other cases they can differ substantially: in fat-tailed conditional distributions, on the one hand, 97.5%-ES would imply higher risk forecasts, while on the other, it provides a smaller down-side risk than using the 99%-VaR. It is found that the empirical results in the paper generally support the proposals of the Basel Committee.

Key words and phrases: Stochastic dominance, Value-at-Risk, Expected Shortfall, Optimizing strategy, Basel III Accord.

JEL Classifications: G32, G11, G17, C53, C22.

1. Introduction

In the financial market industry, it is well known that the Basel III Accord requires that banks and other Authorized Deposit-taking Institutions (ADIs) communicate their daily risk forecasts to the appropriate monetary authorities at the beginning of each trading day, using one of a range of alternative financial risk models to forecast Value-at-Risk (VaR). The risk estimates from these models are used to determine the daily capital charges (DCC) and associated capital costs of ADIs, depending in part on the number of previous violations, whereby realized losses exceed the estimated VaR (for further details see, for example, Chang et al. (2011)).

Recently, the Basel Committee on Banking Supervision (BCBS) (2013) published a consultative document which presents the Basel Committee's initial proposals in regard to trading book capital requirement policies. A key element of the proposal is moving the quantitative risk metrics system from VaR to expected shortfall (ES), and lowering the confidence level from 99% to 97.5%. The Basel Committee (2013) observed that "a number of weaknesses have been identified in using Value-at-Risk (VaR) for determining regulatory capital requirements, including its inability to capture tail risk" (p. 3). For this reason, the Basel Committee has considered the use of Expected Shortfall (ES). ES is a coherent risk measure and has already become common in the insurance industry, although not yet in the banking industry. Artzner et al. (1997) proposed the use of ES to alleviate the problems inherent in VaR: ES considers losses beyond the VaR level and is shown to be sub-additive, while VaR disregards losses beyond the percentile and is not sub-additive. However, although ES is mathematically superior to VaR in that it does not show "tail risk" and is a coherent risk measure in being subadditive, its practical implementation and greater computational requirements may pose operational challenges to financial firms.

Danielsson (2013) examines the quantitative impact of such a proposal through analytical calculations, Monte Carlo simulations, and empirical results from observed data. He tries to analyze one of the main issues raised from this change, namely that estimating ES conditional on VaR might be such that estimation and model risk for ES will be strictly higher than for VaR. Having found that 97.5% ES and 99% VaR are exactly the same for conditionally normal procedures, and a slightly greater ES than VaR for conditional Student-t, the analysis concludes that the 97.5% ES risk forecasts are generally more volatile than their 99% VaR counterparts.

Prior methods for testing the main concerns of moving from VaR to ES have focused on the first and second moments of both risk measures distributions. It is important to observe that mean-variance type assessments are justified by a joint consideration of a quadratic risk function, as well as the full characterization of the Gaussian case by the first and second moments. Absent a Gaussian setting, justification of a quadratic loss function itself becomes questionable. There should be concern with higher moments (if they exist), and often asymmetrical tail-area behaviour, especially when tail functions such as VaR and ES are of interest?

In this paper we recommend and analyze the Stochastic Dominance (SD) approach for comparing the two popular risk measurements, namely ES and VaR, as a complementary robust alternative that seeks weak uniform rankings over entire classes of evaluation functions, based on nonparametric distributions of VaR and ES. SD tests use a large amount of information from the empirical probability distribution of VaR and ES. Consequently, the partial ordering is more precise. Such an approach can incorporate useful information about the likelihood of specific levels of VaR or ES, and provide bank regulators with greater information about the uncertainty associated with each different option.

The concept of SD arose from decision theory and has been widely used in finance providing, for instance, pairwise comparisons of risky prospects such that all individuals whose utility functions belong to some set *U* will prefer one prospect over another. Chang et al. (2015) use SD tests for choosing among several VaR forecasting models to analyze whether the DCC charges produced for such a model would stochastically dominate the DCC charges produced for an alternative model.

Assuming F and G are the distribution functions of ES and VaR, respectively, ES SD VaR iff $\mathrm{F}(x) \leq \mathrm{G}(x)$, with strict inequality over some values of *x*. This means that ES is dominant over all increasing evaluation functions since, at all percentiles, the probability of ES is greater than VaR. In particular, ES will have a higher median value than VaR. Similarly, each and every percentile of the F distribution will be at a higher *x* level than the corresponding percentile of the G distribution.

In this paper we generate probability distributions of both VaR and ES and apply SD to compare both risk measures. We examine several conditional volatility models for forecasting VaR and ES, including GARCH, EGARCH and GJR, paired with Gaussian and Student-t distributions. The results show that the ES FSD VaR for the three models and two distributions. This would imply that the likelihood of a greater mean of ES is higher than the mean of VaR. In addition, we also find that ES SSD VaR. This implies that the uncertainty inherent in the estimated VaR is greater than the uncertainty in ES, at least in the case of extreme events, which is contrary to Danielson's (2013) results.

The remainder of the paper is organized as follows. Section 2 describes VaR and ES risk measures. Section 3 presents alternative conditional volatility models to produce VaR and ES. In Section 4, the definition, notation and properties of SD are presented. Section 5 introduces the data, describes the block bootstrapping method to simulate time series, and illustrates the application of SD to analyze if ES stochastically dominates VaR. Section 6 presents the main results. Section 7 gives some concluding comments.

2. Forecasting Value-at-Risk and Expected Shortfall

In this section we introduce the calculation and forecasting of Value-at-Risk (VaR) and Expected Shortfall (ES).

2.1. Value-at-Risk

VaR refers to the lower bound of a confidence interval for a (conditional) mean, that is, a "worst case scenario on a typical day". If interest lies in modelling the random variable, Y_t, it can be decomposed as follows:

$$Y_t = \mu_t + \varepsilon_t. \qquad (1)$$

This decomposition states that Y_t comprises a predictable component, μ_t, which is the conditional mean, and a random component, ε_t. The variability of Y_t, and hence its distribution, is determined by the variability of ε_t. If it is assumed that ε_t follows a conditional distribution, $\varepsilon_t \sim D(\mu_t, \sigma_t^2)$, where μ_t and σ_t are the time-varying conditional mean and standard

deviation of ε_t, respectively, these can be estimated using a variety of parametric, semi-parametric or non-parametric methods.

The VaR for a given confidence level $q \in (0,1)$ and time *t* is given by the smallest number y_q such that the lost Y_{t+1} at time *t+1*will fall below y_q with probability *q:*

$$VaR_t^q = \inf\left\{y_q \in \Re : P\left(Y_{t+1} \le y_q\right) \ge q\right\} = \inf\left\{y_q \in \Re : P\left(Y_{t+1} > y_q\right) \le 1-q\right\}. \qquad (2)$$

Thus, VaR is a quantile of the distribution of the loss function, and *q* is usually taken to be in the range [0.9, 1). For example, the Basel II accord refers to the "*99%-VaR*". Sometimes the level of significance or coverage rate, $\alpha = 1-q$, is used instead. If the random variable ε_t is normally distributed then, for $q \in (0,1)$, the VaR of ε_t is given by:

$$VaR_{t+1}^q = \mu_{t+1} + \sigma_{t+1}\Phi^{-1}(q), \qquad (3)$$

where Φ is the cumulative distribution function of a standard normal variable.

If instead the normalized random variable $\tilde{Z}_t = (\varepsilon_t - \mu_t)/\sigma_t$ has a standardized t-distribution with $\nu > 2$ degrees of freedom, that is, the Student t distribution with mean 0 and variance 1, the VaR of $\varepsilon_t = \mu_t + \sigma_t \tilde{Z}_t$ would be:

$$VaR_{t+1}^q = \mu_{t+1} + \sigma_{t+1}\sqrt{\nu^{-1}(\nu-2)}\, t_\nu^{-1}(q), \qquad (4)$$

where $t_\nu^{-1}(q)$ is the *q* quantile of the *standardized* Student-t distribution. As quantiles translate under monotonic transformations, the *q* quantile with mean 0 and variance 1 is given as $t_\nu^{-1}(q)\sqrt{\nu^{-1}(\nu-2)}$.

2.2. Expected Shortfall

The expected shortfall at level *q* is the expected value at time *t* of the loss in the next period, Y_{t+1}, conditional on the loss exceeding VaR_t^q:

$$ES_{t+1}^{q} = E_t\left[Y_{t+1} \middle| Y_{t+1} > VaR_{t+1}^{q}\right]. \tag{5}$$

For $q \in (0,1)$, the expected shortfall for a normally distributed random variable, $\varepsilon_t \sim N(\mu_t, \sigma_t^2)$, is given as:

$$ES_{t+1}^{q} = \mu_{t+1} + \sigma_{t+1}\frac{\phi\left(\Phi^{-1}(q)\right)}{1-q}, \tag{6}$$

where ϕ is the density of a standard normal variable.

If instead the normalized random variable $\tilde{Z}_t = (\varepsilon_t - \mu_t)/\sigma_t$ has a *standardized* t-distribution with $\nu > 2$ degrees of freedom, then the expected shortfall of ε_t is given by:

$$ES_{t+1}^{q} = \mu_{t+1} + \sigma_{t+1}\frac{f^*_{\nu}\left(t^{*-1}_{\nu}(q)\right)}{1-q}\frac{\left(\nu - 2 + \left(t^{*-1}_{\nu}(q)\right)^2\right)}{q}, \tag{7}$$

Where $t^{*-1}_{\nu}(q)$ denotes the q quantile of the *standardized* Student t distribution (that is, with zero mean and unit variance) having ν degrees of freedom, and $f^*_{\nu}\left(t^{*-1}_{\nu}(q)\right)$ is the value of its density function at that point. The standardized Student t density function is given as:

$$f^*_{\nu}(x) = \left((\nu-2)\pi\right)^{-1/2}\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{\nu+1}{2}\right)\left(1+(\nu-2)^{-1}x^2\right)^{-(1+\nu)/2}, \tag{8}$$

where the *gamma function,* Γ, is an extension of the factorial function to non-integer values (Alexander, 2009, p. 130).

Analogous to the VaR calculations, for the calculation of ES it is possible for σ_t to be replaced by alternative estimates of the conditional standard deviation in order to obtain an appropriate VaR (for useful reviews of theoretical results for conditional volatility models, see Li et al. (2002) and McAleer (2005), where several univariate and multivariate, conditional, stochastic and realized volatility models are discussed).

The next section describes several volatility models that are widely used to forecast the 1-day ahead conditional variances and VaR thresholds.

3. Models for Forecasting VaR

ADIs can use internal models to determine their VaR thresholds. There are alternative time series models for estimating conditional volatility. In what follows, we present several well-known conditional volatility models that can be used to evaluate strategic market risk disclosure, namely GARCH, GJR and EGARCH, with Gaussian and Student-*t* distributions. These univariate models are chosen because they are widely used in the literature. For an extensive discussion of the theoretical properties of several of these models see, for example, Ling and McAleer (2002a, 2002b, 2003a), McAleer (2005), Caporin and McAleer (2012), McAleer and Hafner (2014), and McAleer (2014).

3.1 GARCH

For a wide range of financial data series, time-varying conditional variances can be explained empirically through the autoregressive conditional heteroskedasticity (ARCH) model, which was proposed by Engle (1982). When the time-varying conditional variance has both autoregressive and moving average components, this leads to the generalized ARCH(*p*,*q*), or GARCH(*p*,*q*), model of Bollerslev (1986). It is very common in practice to impose the widely estimated GARCH(1,1) specification in advance.

Consider the stationary AR(1)-GARCH(1,1) model for daily returns, y_t :

$$y_t = \varphi_1 + \varphi_2 y_{t-1} + \varepsilon_t, \qquad |\varphi_2| < 1 \tag{9}$$

For t =1,…, n, where the shocks to returns are given by:

$$\begin{aligned} \varepsilon_t &= \eta_t \sqrt{h_t}, \quad \eta_t \sim iid(0,1) \\ h_t &= \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}, \end{aligned} \tag{10}$$

and $\omega > 0, \alpha \geq 0, \beta \geq 0$ are sufficient conditions to ensure that the conditional variance $h_t > 0$. The stationary AR(1)-GARCH(1,1) model can be modified to incorporate a non-stationary ARMA(*p*,*q*) conditional mean and a stationary GARCH(*r*,*s*) conditional variance, as in Ling and McAleer (2003b). Tsay (1987) shows that $\alpha > 0$ in the derivation of the GARCH model.

3.2 EGARCH

An alternative model to capture asymmetric behaviour in the conditional variance is the Exponential GARCH, or EGARCH(1,1), model of Nelson (1991), namely:

$$log h_t = \omega + \alpha \left| \frac{\varepsilon_{t-1}}{h_{t-1}} \right| + \gamma \frac{\varepsilon_{t-1}}{h_{t-1}} + \beta log h_{t-1}, \quad |\beta| < 1, \tag{11}$$

where the parameters α, β and γ have different interpretations from those in the GARCH(1,1) and GJR(1,1) models.

EGARCH captures asymmetries differently from GJR. The parameters α and γ in EGARCH(1,1) represent the magnitude (or size) and sign effects of the standardized residuals, respectively, on the conditional variance, whereas α and $\alpha + \gamma$ represent the effects of positive and negative shocks, respectively, on the conditional variance in GJR(1,1). Unlike GJR, EGARCH can accommodate leverage, depending on the restrictions imposed on the size and sign parameters, though leverage is not guaranteed (further details can be found in McAleer et al. (2007)). McAleer and Hafner (2014) showed that $\alpha > 0$ and $\gamma > 0$ in the derivation of the EGARCH model.

3.3 GJR

In the symmetric GARCH model, the effects of positive shocks (or upward movements in daily returns) on the conditional variance, h_t, are assumed to be the same as the effects of negative shocks (or downward movements in daily returns) of equal magnitude. In order to accommodate asymmetric behaviour, Glosten, Jagannathan and Runkle (1992) proposed a model (hereafter GJR), for which GJR(1,1) is defined as follows:

$$h_t = \omega + (\alpha + \gamma I(\eta_{t-1}))\varepsilon_{t-1}^2 + \beta h_{t-1}, \tag{12}$$

where $\omega > 0, \alpha \geq 0, \alpha + \gamma \geq 0, \beta \geq 0$ are sufficient conditions for $h_t > 0,$ and $I(\eta_t)$ is an indicator variable defined by:

$$I(\eta_t) = \begin{cases} 1, & \varepsilon_t < 0 \\ 0, & \varepsilon_t \geq 0 \end{cases} \tag{13}$$

as η_t has the same sign as ε_t. The indicator variable differentiates between positive and negative shocks, so that asymmetric effects in the data are captured by the coefficient γ. For financial data, it is expected that $\gamma \geq 0$ because negative shocks have a greater impact on risk than do positive shocks of similar magnitude. The asymmetric effect, γ, measures the contribution of shocks to both short run persistence, $\alpha + \gamma/2$, and to long run persistence, $\alpha + \beta + \gamma/2$.

Although GJR permits asymmetric effects of positive and negative shocks of equal magnitude on conditional volatility, the special case of leverage, whereby negative shocks increase volatility while positive shocks decrease volatility (see Black (1976) for an argument using the debt/equity ratio), cannot be accommodated, in practice (for further details on asymmetry versus leverage in the GJR model, see Caporin and McAleer (2012)). McAleer (2014) showed that $\alpha > 0$ and $\gamma > 0$ in the derivation of the GJR model.

In the empirical analysis, the three conditional volatility models given above are estimated under the following distributional assumptions on the conditional shocks: (1) Gaussian and (2) Student-*t*, with estimated degrees of freedom. As the models that incorporate the *t* distributed errors are estimated by QMLE, the resulting estimators are consistent and asymptotically normal, so they can be used for estimation, inference and forecasting.

4. Stochastic Dominance[1]

[1]This section is based on Donald and Hsu (2013), and Linton, Maasoumi, and Whang (2007).

The objective is to evaluate the stochastic properties of 99%-VaR against that of 97.5%-ES using alternative conditional volatility models and error distributions for forecasting volatility. The stochastic dominance concept is applied to determine whether 97.5%-ES stochastically dominates 99%-VaR, which means that using ES for forecasting risk, not only would imply a higher likelihood of greater values of risk forecast, but also a lower uncertainty. Below we briefly describe the SD tests that are used in this paper.

4.1 Definitions and Hypothesis Formulation

Let X and Ybe two random variables with cumulative distribution functions (CDF) F_X and F_Y, respectively. For first order stochastic dominance (SD1), Y SD1 X, if $F_Y(z) \leq F_X(z)$ for all $z \in$ R. Let $W_U(F)$ denote an evaluation function of the form $W_U(F) = \int U(z)\, dF(z)$, where F *is* the distribution of an underlying variable, and U is any "utility" function. SD1 is defined over monotonically increasing utility functions, that is, $W_U(F_Y) \geq W_U(F_X)$ for all $U(z)$ such that $U'(z) \geq 0$.

The technical assumptions for the underlying statistical theory include the following (see Linton, Maasoumi and Whang (2005) (hereafter LMW), Linton, Song and Whang (2010), and Donald and Hsu (2013) for further details):

Assumption 4.1:

1. $Z = [0, \overline{z}]$, *where* $\overline{z} < \infty$.

2. F_X *and* F_Y *are continuous functions on Z such that* $F_X(z) = F_Y(z) = 0$ *iff* $z = 0$, *and* $F_X(z) = F_Y(z) = 1$ *iff* $z = \overline{z}$.

In order to test if Y SD1 X, Donald and Hsu(2013) formulate their hypotheses as:

$$H_0 : F_Y(z) \leq F_X(z) \text{ for all } z \in Z, \tag{14}$$

$$H_1 : F_Y(z) > F_X(z) \text{ for all } z \in Z, \tag{15}$$

This test is different from that in Linton, Maasoumi and Whang (2005), who provide a simultaneous test of either Y SD1 X or X SD1 Y.

Assumption 4.2*:*

1. $\{X_i\}_{i=1}^{N}$ *and* $\{Y_i\}_{i=1}^{M}$ *are samples from distributions with CDF's* F_X *and* F_Y*, respectively. It is possible to deal with independent samples and observations. Linton, Maasoumi and Whang (2005)allow dependent time series, and possibly dependent X and Y.*

2. M is a function of N satisfying that $M(N) \to \infty$ *and* $N/(N+M(N)) \to \lambda \sim (0, 1)$ *when* $N \to \infty$.

The CDF's F_X and F_Y are estimated by empirical CDFs:

$$\hat{F}_{X,N}(z) = \frac{1}{N}\sum_{i=1}^{N} 1(X_i \le z),\ \hat{F}_{Y,M}(z) = \frac{1}{M}\sum_{i=1}^{N} 1(Y_i \le z),$$

where $1(\cdot)$ denotes the indicator function. The Kolmogorov-Smirnov test statistic is given by:

$$\hat{S}_N = \sqrt{\frac{NM}{N+M}} \sup_{z \in Z} (\hat{F}_{Y,N}(z) - \hat{F}_{X,M}(z)).$$

Let Ψ_{h_2} denote a zero mean Gaussian process with covariance kernel equal to $h_2 \in H_2$, where H_2 denotes the collection of all covariance kernels on $Z \times Z$. For F_X and F_Y satisfying Assumption 4.1, let $h_2^{X,Y}$ denote the covariance kernel on $Z \times Z$, such that:

$$h_2^{X,Y}(z_1, z_2) = \lambda \cdot F_X(z_1)(1 - F_X(z_2)) + (1 - \lambda) \cdot F_Y(z_1)(1 - F_Y(z_2)) \text{ for } z_1 \le z_2, \quad (16)$$

with λ defined in Assumption 4.2. It follows that:

$$\sqrt{NM/(N+M)}\,(\hat{F}_{Y,M}(z) - \hat{F}_{X,N}(z) - (F_Y(z) - F_X(z))) \Rightarrow \Psi_{h_2 X,Y}$$

(see, for example, Linton, Maasoumi and Whang (2005), or Donald and Hsu (2013)). Typical limiting results are:

1. Under H_0, $\hat{S}_N \xrightarrow{D} \sup_{z \in Z} \Psi_{h_2 X,Y}$.

2. Under H_1, $\hat{S}_N \xrightarrow{D} \infty$.

Several approaches for resampling and subsampling implementation of SD tests have been proposed. Some methods simulate the limiting Gaussian process, in the spirit of Barrett and Donald (2003), using the multiplier method, bootstrap with separate samples, or bootstrap with combined samples. Simulated processes weakly converge to the same process as the limit process, conditional on the sample path with probability approaching 1.

4.2 Re-centering Functions

These simulation methods do not work well enough when the data are weakly dependent, as for time series samples that are used in this paper. In these cases, one has to appeal to either the subsampling technique of Linton, Maasoumi and Whang (2005), or a variant of the block bootstrap. Donald and Hsu (2013) provide a comparative examination of these alternatives.

Donald and Hsu (2013) and Linton, Maasoumi and Whang (2005) propose re-centering methods introduced by Hansen (2005) to construct critical values for Kolmogorov-Smirnov type tests. This approach provides a test with improved size and power properties compared with the unadjusted test mounted at the composite boundary of the null and alternative spaces, the so-called Least Favorable Case (LFC).

The re-centering function proposed by Donald and Hsu (2013) is conditioned to apply on the null space:

$$\hat{\mu}_N(z) = \left(\hat{F}_{Y,N}(z) - \hat{F}_{X,N}(z)\right) \cdot 1\left(\sqrt{N}\ \left(F_{Y,N}(z) - F_{X,N}(z)\right) < aN\right).$$

For $\alpha < 1/2$, let:

$$\hat{c}^{bb}_{\eta,N} = \max\{\tilde{c}^{bb}_N,\ \eta\},$$
$$\tilde{c}^{bb}_N = \sup\left\{\ c \mid P^u\left(\sup_{z\in Z}\sqrt{N}\left(\hat{D}^{bb}_N(z) + \hat{\mu}_N(z)\right) \le c\right) \le 1-\alpha\right\}$$

where $\hat{D}^{bb}_N(z)$ is the "S" statistic computed using the b-th resample/subsample block. If the decision rule is to reject the null hypothesis, H_0: $F_Y(z) \le F_X(z)$ for all $z \sim Z$ when $\hat{S}_N > \hat{c}^{bb}_{\eta,N}$, then the corresponding test will have the same size properties as in the independent random samples case.

4.3 Weakly Dependent Data

Let $\{(X_i, Y_i)\}_{i=1}^N$ be a strictly stationary time series sequence with joint distribution function F_{XY} on Z^2 and marginal CDF's F_X and F_Y, respectively. Suppose that Assumption 1 of Linton, Maasoumi and Whang (2005) holds. Then under the null hypothesis that H_0: $F_Y(z) \leq F_X(z)$ for all $z \sim Z$, $\hat{S}_N \xrightarrow{D} \sup_{z \in Z} \Psi_{h_2}(z)$, where:

$$h_2(z_1, z_2) = \lim_{N \to \infty} Cov\Bigg(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} (1(Y_i \leq z_1) - 1(X_i \leq z_1) - F_Y(z_1) + F_X(z_1)), \\ \frac{1}{\sqrt{N}} \Sigma_{i=1}^{N} (1(Y_i \leq z_2) - 1(X_i \leq z_2) - F_Y(z_2) + F_X(z_2))\Bigg), \tag{17}$$

which is the long-run covariance kernel function. In order to simulate Ψ_{h_2}, Donald and Hsu (2013) propose the blockwise bootstrap as in Linton, Maasoumi and Whang (2005) because the multiplier method and the bootstrap methods do not account for the weak dependence of the data.

Under the same conditions as in Linton, Maasoumi and Whang (2005), it follows that $\sqrt{N} D_N^{bb}(z) \overset{p}{\Rightarrow} \Psi_{h_2}(z)$, where h_2 is defined in (17). We adopt the Donald and Hsu (2013) recentering procedure and critical values as they are less conservative under the null, and at least as powerful under the alternative. In order to appreciate the role played by the choice of critical values, consider the critical value in either the multiplier method (mp), bootstrap with separate samples (bs), or bootstrap with combined samples (bc), given below for *k* = *mp*, *bs*, *bc*, as follows:

$$\hat{q}_N^k = \sup\left\{ q \mid P^u\left(\sqrt{\frac{NM}{N+M}} \sup_{z \in Z} \hat{D}_N^k(z) \leq q \right) \leq 1 - \alpha \right\}. \tag{18}$$

The critical value $\hat{q}_N^k$ is bounded away from zero in probability. Since η can be chosen to be arbitrarily small, we can assume that $\eta < \hat{q}_N^k$, which implies that $\hat{c}_{\eta,N}^k = \max\{\tilde{c}_{,N}^k, \eta\} \leq \hat{q}_N^k$ given

that $\tilde{c}^k_{\eta,N} \leq \hat{q}^k_N$. Thus, Donald and Hsu (2013) are able to show that, given Assumptions 4.1, 4.2, and $\alpha < 0.5$:

$$P(\hat{S} > \hat{q}^k_N) \leq P(\hat{S} > \hat{c}^k_{\eta,N}), \text{ for } k = \text{mp, bs and bc.}$$

This would imply that the powers and sizes of these tests are never smaller than those of Barrett and Donald (2003).

4.4 Linton, Maasoumi and Whang's (2005) Subsampling Test

Linton, Maasoumi and Whang (2005) estimate the critical value by the subsampling method proposed in Politis and Romano (1992). Donald and Hsu (2013) introduce LMW's test with a modification that allows for different sample sizes. For $s \geq 1$, let X_s denote the collection of all of the subsets of size s of $\{X_1, ..., X_N\}$:

$$X_s \equiv \left\{ \{X_{r_1}, \ldots, X_{r_s}\} \,|\, \{r_1, \ldots, r_s\} \subseteq \{1, \ldots, N\} \right\}.$$

A random draw denoted by $\{X^b_1, \ldots, X^b_s\}$ from X_s would be a random sample of size s without replacement from the original data. Let $\hat{F}^b_{X,s}$ be the empirical CDF based on the random draw, $\{X^b_1, \ldots, X^b_s\}$. Define $\hat{F}^b_{Y,s}$ similarly. Let s_N and s_M denote the subsampling sizes for the X and Y samples, respectively. The subsampling critical value $\hat{c}^s_N$ is given by:

$$\hat{c}^s_N = \sup\left\{ c \,|\, P^u\left(\sqrt{\frac{s_N s_M}{s_N + s_M}} \sup_{z \in Z} (\hat{F}^b_{Y,s_M}(z) - \hat{F}^b_{Y,s_N}(z)) \leq c \right) \leq 1 - \alpha \right\}.$$

Assume that:

1. $s_N \to \infty$, $s_M \to \infty$, $s_N/N \to 0$ *and* $s_M/M \to 0$ *as* $N \to \infty$.
2. $s_N/(s_N + s_M) \to \lambda$, *where* λ *is defined in Assumption 4.2.*

The first assumption is standard for the subsampling method. The second assumption requires that the subsample sizes from the two samples grow at the same rate, and that the limit of the ratio of the subsample sizes be the same as that of the original samples. This condition is important if, for example, $s_N/(s_N + s_M) \to \lambda_s \neq \lambda$, so that, under the null hypothesis:

$$\sqrt{\frac{s_N s_M}{s_N + s_M}} \sup_{z\in Z^*}(\hat{G}^b_{s_M}(z) - \hat{F}^b_{Y;s_N}(z)) \xrightarrow{D} \sup_{z\in Z^*}\sqrt{\lambda_s}\, B_G \circ G(z) - \sqrt{1-\lambda_s} B_F \circ F(z), \quad (19)$$

conditional on the sample(s) with probability one (denoted as $D \to p$). When $\lambda_s \neq \lambda$, in the limit, the left-hand side of (19) may not be a good approximation to the limiting null distribution of the original test statistic.

The limiting distribution theory in Linton, Maasoumi and Whang (2005)covers weakly stationary, dependent samples, with certain mixing conditions, such as in our applications. In addition, they allow for the prospects that are ranked to be estimated functions, rather than the original series described above. If the estimators involved in these functions permit certain expansions, as described in Linton, Maasoumi and Whang (2005), Assumption 2, Section 3.1, the limiting distribution theory will be preserved with re-centering.

The DCC functions being ranked here are certainly estimated functions of the data. The conditional means and variance estimators in the VaR forecasts are generally consistent, and admit the type of expansions assumed in Linton, Maasoumi and Whang (2005).However, DCC is a maximal function of two functions of past VaRs and, to the extent that this may lead to discontinuities, or non-smoothness, in the estimated empirical CDFs, the accuracy of the underlying limiting distributions may be affected. It is known that the subsampling method is valid for many non-standard cases of this kind, but the same may not be true of bootstrap methods. A more detailed technical examination of this issue is beyond the scope of this paper.

5. Data and Implementation of Tests

5.1. Data description

The data used for estimation and forecasting are the closing daily prices for Standard and Poor's Composite 500 Index (S&P500), which were obtained from the Thomson Reuters-Datastream database for the period 1 January 1999 to 26 June 2014, giving 4040 observations.

The returns at time $t(R_t)$ are defined as:

$$R_t = \log\left(P_t / P_{t-1}\right), \tag{20}$$

where P_t is the market price.

[Insert Figure 1 here]

Figure 1 shows the S&P500 returns. The extremely high positive and negative returns are evident from September 2008 onward, after the Lehmann Brothers bankruptcy, and have continued well into 2009. Then, in spring, 2010, the European debt crisis starts, with the European Union along with the International Monetary Fund providing 110 million Euros to Greece that became unable to borrow from the market. Greece required a second bailout in mid-2011, and thereafter Ireland and Portugal also received bailouts in November 2010 and May 2011, respectively. Higher volatility in the S&P500 returns is observed during these periods. Regarding the descriptive statistics, the median (0.022) is above the mean (0.012) and the range is between 11% and -9.5%, with a standard deviation of 1.27. S&P500 returns show negative skewness (-0.17) and high kurtosis (10.99), which would seem to indicate the existence of extreme observations.

Figure 2 shows several graphs that provide valuable information for identifying the returns probability distribution. Panel A displays the empirical histogram, together with the density function of the Gaussian distribution and a kernel density estimate of the distribution that show fatter tails than normal and some slight asymmetry. Panels B and C exhibit two theoretical quantile-quantile plots (QQ-plot) comparing the quantiles of the S&P500 returns with the quantiles of both a fitted normal, Panel B, and Student-t, Panel C, distributions. For the Gaussian case, the QQ-plot does not lie on a straight line, overall on the tails, supporting the non-normality of returns. According to the QQ-plots, the Student-t distribution seems to fit the observed data better than does the Gaussian distribution. Finally, Panel D displays a boxplot that summarizes the returns distributions showing the extreme observation mentioned above.

[Insert Figure 2 here]

Figures 1 and 2 show that stock markets have been working under stress during the last seven years. Traditional risk measurement, specifically VaR, might not work properly under these extreme price fluctuations. The BIS Committee on the Global Financial System discussed the

shortcomings of VaR for measuring and monitoring market risk when many such events are taking place in the tails of the distributions. VaR that suffers tail risk only measures the distribution quantile, and disregards the extreme loss beyond the VaR level, ignoring important information regarding the tails of the distribution. Expected shortfall might be a more appropriate tool for risk monitoring under stress circumstances. As Yamai and Yoshiba (2005) state, expected shortfall has no tail risk under more lenient conditions than VaR.

Fat tails might be explained by clusters of volatility that seem to appear in Figure 1. A closer examination of the volatility of returns using a measure proposed in Franses and van Dijk (1999) is given as:

$$V_t = \left(R_t - E\left(R_t \mid F_{t-1}\right)\right)^2, \tag{21}$$

where F_{t-1} is the information set at time *t*-1. This measure highlights the volatility clustering (see Figure 3).

[Insert Figure 3 here]

5.2. Block bootstrapping and subsampling

In order to test for SD between the 99%-VaR and 97.5%-ES risk measures using different conditional models for forecasting volatility, we implement the Circular Block Bootstrapping (CBB) method developed in Politis and Romano (1992) for resampling theS&P500 through the MFE toolbox of Sheppard (2013). The block bootstrap is widely used for implementing the bootstrap with time series data. It consists of dividing the data into blocks of observations and sampling the blocks randomly, with replacement.

In the CBB, let the data consist of observations $\{X_i : i = 1,...,n\}$, and let $l \in \{1,...,n\}$ and $b \geq 1$ denote the length and the number of blocks, respectively, such that $lx\, b \leq n$. Let *n* and *m* be the initial data size and the bootstrap sample size, $m \leq n$ and *k* the number of blocks chosen. CBB consists of dividing the time series into *b* blocks of consecutive observations denoted by:

$$B_i = \left(X_{(i-1)l+1},...,X_{il}\right),\ i = 1,...,n.$$

A random sample of k blocks, $k \geq 1$, $B_1^*, ..., B_k^*$ is selected with replacement from $B_1^*, ..., B_k^*$. Joining the k blocks with $m = k \times l$ observations, the bootstrap sample is given as:

$$\left(X_1^*, ..., X_l^*, ..., X_{(k-1)l+1}^* ..., X_l^*\right).$$

The CBB procedure is based on wrapping the data around a circle and forming additional blocks using the "circularly defined" observations. For $i \geq n$, it is defined that $X_1 = X_{i_n}$, where $i_n = i \bmod n$ and $X_0 = X_n$. The CBB method resamples overlapping and periodically extended blocks of length *l*. Notice that each X_i appears exactly *l* times in the collection of blocks and, as the CBB resamples the blocks from this collection with equal probability, each of the original observations X_1, ..., X_n receives equal weight under the CBB. This property distinguishes the CBB from previous methods, such as the non-overlapping block bootstrap of Carlstein (1992)**.** Note that and 97.5%-ES and99%-VaR are estimated for each drawn sample, thereby generating the bootstrap (subsample) distribution of the test statistics.

5.3.VaR and ES evaluation framework: Stochastic Dominance

Stochastic dominance establishes a partial ordering of probability distributions and is suitable for addressing the question of whether there is any significant statistical difference between 97.5%-ES and 99%-VaR. Therefore stochastic dominance will allow us to rank 97.5%-ES and 99%-VaR by analyzing the entire distribution of both statistics. The main point of the analysis is to help risk managers and regulators to choose between alternative risk measures.

We want to test if the two risk measurements are strictly ordered against the alternative that they are not. The way that SD may be used to compare both measures of risk is as follows. For notational simplicity, we write $Risk\,1 \overset{FSD}{\succeq} Risk\,2$ and $Risk\,1 \overset{SSD}{\succeq} Risk\,2$ whenever Risk1 (Risk measure #1) dominates Risk 2 (Risk measure #2), according to FSD and SSD, respectively. FSD implies SSD. Let Risk1 and Risk2 be the 97.5%-ES and 99%-VaR, respectively, produced using the same conditional volatility model and the same error distribution. Based on the previous definition, if Risk1 first-order stochastically dominates Risk2, then Risk1 will involve higher risk than Risk2, in the sense that it has a higher probability of producing higher forecast risk.

Graphically, FSD exists when the cumulative distribution functions do not intersect; if they do cross, then FSD is statistically indeterminate. Therefore, neither Risk1 nor Rik2 is dominant. Accordingly, FSD might be a weak result and using a stronger constraint, SSD, to select the best risk model provides valuable information. SSD implies that the difference between the ICDF (area under CDFs) Risk2 and ICDF Risk1 is always positive for every level of probability. This may be interpreted as "uniformly less down-side risk at every level of probability" for Risk1 model.

As an example, consider the probability density functions from two alternative distributions Gaussian and Chi-squared, based on an illustrative example using 1000 simulated observations, In the left panel of Figure 4, the bold line represents Model 1, which has a Gaussian distribution function with mean 4.5 and standard deviation 1. The dotted line is for Model 2, and depicts the density of a Chi-squared with 4 degrees of freedom. SD analysis requires computing the corresponding empirical CDFs, $\hat{F}_X$ and $\hat{F}_Y$. For any two variables, *X* and *Y*, such empirical CDFs are computed as follows: (i) a vector *z* containing all the distinct *X* and *Y* values is created, and is sorted in ascending order; and (ii) each point j in $\hat{F}_X\left(\hat{F}_Y\right)$ is calculated as the number of observations less than or equal to observation *j* in vector *z*, divided by the total number of observations of X (Y). The number of points on the horizontal axes in Figures 4-5, as determined by the number of distinct values of X and Y, show the numbers of observations for $\hat{F}_X$ and $\hat{F}_Y$.

In the right panel of Figure 4, which depicts the CDF for Models 1 and 2, we see that the values of Model 2 are more uncertain, with a greater probability of either small (the CDF for Model 2 is initially above that for Model 1) or large (the CDF for Model 2 lies below that of Model 1 for high values) values. As these CDF functions cross, first-order stochastic dominance cannot be established.

Regarding second-order dominance, the test is based on the cumulative areas under the CDFs (ICDF), as shown in the right panel of Figure 5. As the difference in ICDF Model 2 - ICDF Model 1) is always positive, this implies that model 1 second-order dominates Model 2, and will be preferred by a risk averse risk manager, whatever their cardinal (concave) loss function. Choosing Model 2 would imply a greater risk during periods of turmoil and greater uncertainty.

6. Empirical Results

In this section we analyze whether 97.5%-ES is statistically different from 99%-VaR, as well as its implications for financial regulation. We use conditional volatility models for estimating the conditional variances and two different conditional distributions for the S&P500 returns, namely Gaussian and Student-t. The focus is on analyzing whether the one-step-ahead forecast of 97.5%-ES for each conditional volatility model stochastically dominates the corresponding one-step-ahead forecast of 99%-VaR. If stochastic dominance is found, it would imply that the two measures are stochastically different, and using either ES or VaR would produce both different expected risk forecast measures and down-side risk.

As neither ES nor VaR is observed, they have to be estimated. We proceed as follows:

(1) We use a long 3000-observation rolling window (from 1 January 1999 to 1 July 2010, which amounts to 75% of the total number of available observations) for smoothing out spikes of volatility during the Global Financial Crisis, while estimating the conditional volatility models for producing one-step-ahead 97.5%-ES and 99%-VaR forecasts.

(2) In order to obtain the empirical distribution of both risk measures, block bootstrapping is used for simulating 500 time series of the S&P500 returns for the 3000-observation rolling window chosen in step 1 that will be used for producing a total of 500 one-step-ahead 97.5%-ES and 99%-VaR forecasts.

(3) Steps 1 and 2 are then repeated for the 1040 days remaining in the total sample (from 2 July 2010 until 26 June 2014), each time leaving out the first observation and adding a new observation at the end. This procedure yields a 500×1040 matrix for computing the cumulative distribution functions needed for testing SD.

As ES and VaR are both affected by errors, this raises the question as to whether the statistical differences to be found might be a consequence of large and significant estimation errors. Yamai and Yoshiba (2002) analyze the advantages and disadvantages of ES and VaR as risk measures. They find that increasing the sample size reduces the estimation error. In order to place the ES and VaR estimation errors found in this paper into perspective, we refer to Yamai

and Yoshiba's (2002) exhaustive analysis based on Monte Carlo simulations for different sample sizes and several stable distributions. Following Yamai and Yoshiba (2002), we evaluate the ES and VaR estimation errors using the relative standard deviation, computed as the standard deviation divided by the average of the ES and VaR forecasts:

1. First, calculate the daily mean and standard deviation for the 500 estimates of 97.5%-ES and 99%-VaR to obtain means and standard deviations for these estimates.
2. Repeat Step 1 above for 1040 days, which provides 1040 relative standard deviations for both 97.5%-ES and 99%-VaR.

Figures 6A-6C show three boxplots for the 1040 relative standard deviations for the 97.5%-ES and 99%-VaR estimates using the three conditional volatility models, when Gaussian and Student-t distributions are assumed. Boxplots are a convenient technique for visually representing the ES and VaR estimation errors. They allow us to identify outliers and compare distributions. Each of Figures 6A-6C contains four boxplots: the first relies on the distribution of 97.5%-ES estimate errors assuming Gaussian (97.5%-ES_G), the second relies on the distribution of 99%-VaR estimate errors assuming Gaussian (99%-VaR_G), the third relies on the distribution of 97.5%-ES estimate errors assuming Student-t (97.5%-ES_t), and the fourth relies on the distribution of 99%-VaR estimation errors assuming Student-t (99%-VaR_t).

If we focus on each boxplot, the bottom of the box in the middle is the 25th percentile, the top is the 75th percentile and the line in the middle is the 50th percentile. "Whiskers" above and below each box are added to provide additional information about the spread of the data. Whiskers are the vertical lines that end in a horizontal stroke. Whiskers are drawn from the 75th and 25th percentiles to the most extreme values the algorithm considers not to be outliers. Additional marks beyond the whiskers are outliers. The four boxplots for the estimation errors under GARCH are shown in Figure 6A. Figures 6B and 6C show boxplots for EGARCH and GJR, respectively.

[Insert Figures 6A-6C here]

First, note that, regardless of the conditional volatility model, the estimation errors of ES and VaR are extremely close when the Gaussian distribution is assumed, as shown in the first two boxplots that turn out to be practically the same. The estimation error of 97.5%-ES seems to be

larger than that of 99%-VaR when the underlying returns distribution is Student-t. This can be seen from the different values for the 25th, 50th and 75th percentiles of the last two boxplots. These results are consistent with those of Yamai and Yoshiba (2002), namely as the underlying returns distribution becomes fat-tailed, the relative standard deviation of the ES estimates becomes much larger than that of VaR.

Yamai and Yoshiba's (2002) analysis is based on 10000 sets of Monte Carlo simulations with a sample size of 1000, assuming several underlying distributions, ranging from Gaussian to Cauchy for both the 95% and 99% quantiles. Using those results as a benchmark, we observe that a relative standard deviation around 0.30 for the ES and VaR estimate errors (50th percentile of the first two boxplots in Figure 6A) under the Gaussian distribution are rather large when compared with a value of 0.05 found in Table 2 for the 99% confidence level of Yamai and Yoshiba's(2002) results under a Gaussian distribution.

Regarding relative standard deviations of ES and VaR when the Student-t is assumed, they are 0.36 (50th percentile of the third boxplot in Figure 6A) and 0.38 (50th percentile of the fourth boxplot in Figure 6A), respectively, for GARCH. These values go to 0.38 (50th percentile of the third boxplot in Figure 6C) and 0.41 (50th percentile of the fourth boxplot in Figure 6C) for VaR and ES, respectively, for GJR.Yamai and Yoshiba (2002) compare the ES and VaR estimate errors by simulating random variables with stable distributions as defined in Feller (1969) and Shiryaev (1999). For these distributions the index of stability, which ranges from 2 (Gaussian) to 1 (Cauchy), defines the distribution function's shape: the heavier is the tail, the smaller is the index of stability of the distribution and the higher is the relative standard deviation, as can be seen for 95%-ES[2]. It is equal to 0.04under a distribution with index of stability of 2 (Gaussian), 0.33 when the index of stability is 1.9, 0.63 if the index is 1.8, and 1.09 when the index of stability is 1.7, which is a common estimate of the stability index for daily rates of returns (see Fama and Roll, 1968)).

The previous comparison focused on assessing the numerical accuracy of the ES and VaR estimates. We note that our estimates are not free of estimation error, although we conclude that the approach adopted yields VaR and ES estimation errors that are close to those of Yamai and Yoshiba (2002) for similar distributions.

[2] 97.5% is not provided in Yamai and Yoshiba (2002).

[Insert Tables 1 and 2 here]

Table 1 presents rejection rates from three different tests, namely Donald and Hsu (2013) (BB), Barrett and Donald (2003)(BD), and Linton, Maasoumi and Whang (2005) (LMW), for the null hypothesis: H_0: riskES FSD riskVaR, where RiskES and riskVaR denote the ES and VaR risk measurements, respectively. Forecast risk measures are produced using two probability distributions, Gaussian (left panel of the table) and Student-t, and three conditional volatility models given in the first row of each table. Table 2shows rejection rates for SSD tests of BB, BD and LMW for the same distributions and volatility models. The *p*-values for the blockwise bootstraps method are approximated based on 200 replications, and the *p*-values for the subsampling method are approximated based on 176 possible subsamples. The significance level is set to 5%. For example, in Table 1, under the Gaussian distribution and using GARCH to produce 97.5%-ES and 99%-VaR forecasts, the BB test obtained a 0% rejection rate for the null hypothesis that *riskES* FSD *riskVaR*. Following Donald and Hsu (2013), when implementing the blockwise bootstrap, the block sizes are set to 12 and the subsample size is set to 25.

In particular, the main results are as follows:

(1) BB, BD and LMW tests in Table 1 show that riskES measure, assuming both Gaussian and Student-t distributions and for every conditional volatility model, first-order stochastically dominates the riskVaR measure. According to these results, 97.5%-ES implies higher risk forecasts than using the 99%-VaR prediction, regardless of the model and distribution used for producing such forecasts.

(2) As FSD implies SSD, then 97.5%-ES SSD 99%-VaR. Therefore, 97.5%-ES shows "uniformly less down-side risk at every level of probability" than 99%-VaR. This result is quite interesting based on the well-known fact that VaR is only the amount which is at risk with a particular probability. It does not suggest how much is at risk at twice that probability, or at half that probability, as it only tells part of the risk condition. Following the analysis in this paper, using the complete cumulative probability distribution of ES and VaR, it can be seen how much is at risk at every level of probability for both risk measures. Therefore, using SD tests, it seems that, contrary to the previous analysis by Danielsson (2013) based only on a comparison of

standard deviations, VaR turns out to be a stochastically different measure of tail-risk, as compared with ES.

(3) The fact that Student-t has heavier tails than the Gaussian distribution explains why CDFs and ICDFs are very close under Gaussianity, which is not the case when the Student-t is used. This can be seen in Figure 7 assuming the Gaussian distribution, and Figure 8 under Student-t. The outcomes of the SD tests for the Gaussian distribution, showing that riskES FSD riskVaR, are illustrative in the light of previous results (see Danielsson, 2013), in which 97.5%-ES and 99%-VaR had similar statistical properties (namely mean and standard deviation), thereby making it difficult to uncover any empirical differences between them. Figures 7 and 8 display 97.5%-ES and 99%-VaR Cumulative Distribution Functions (CDF) (left panel) and the integrated CDFs (ICDF) (right panel) for observation 3288, 8 August 2011 when the S&P500 shows a negative return of 6.89%, which is the maximum loss during the forecast period. These figures represent 97.5%-ES and 99%-VaR when GARCH is used.

[Insert Figures 7 and 8 here]

ES has the advantage that it reveals information about the magnitude of losses in the case of extreme events, besides being a coherent risk measure, that is, it encourages diversification to reduce overall risk for a portfolio. On the one hand, FSD implies that using 97.5%-ES would result in a higher probability of larger estimates of the daily risk forecast. On the other hand, SSD has revealed that ES provides more accurate estimates of the magnitude of losses under extreme events than does VaR. This difference is due to the fact that ES is the average of the daily losses that may occur in the tails, whereas VaR is simply the α percentile for a return distribution. VaR does not factor in the magnitude of losses above such a percentile, whereas ES does.

7. Conclusions

This paper focused on the recent proposal of the Basel committee on Banking Supervision to substitute the quantitative risk measure Value-at-Risk (VaR) for expected shortfall (ES). While VaR is still in great use, the ES may be a preferable risk measure (Dowd, 2005). By definition, the calculation of ES will almost always involve greater computation than VaR. In this paper, we have presented easily-computed expressions for evaluating the ES for two distributions that

are commonly used for modelling asset returns. For other cases, obtaining ES might be more challenging.

For comparing the 97.5%-ES and 99%-VaR, as Basel III recommends moving from 99%-VaR to 97.5%-ES, we used the SD approach, which allows an analysis of not only a limited amount of moments, but rather the whole distributions of ES and VaR. We bootstrapped empirical distributions functions of 97.5%-ES and 99%-VaR from S&P500 returns. We used three conditional volatility models, together with Gaussian and Student-t distributions, in order to simulate the one-step-ahead 97.5%-ES and 99%-VaR forecasts

It was found that 97.5%-ES stochastically dominated 99%-VaR, which suggests that using ES as a risk measure would imply a higher probability of larger one-step-ahead risk values. This risk measure would also entail a lower down-side risk. Although providing worthwhile information about the ES and VaR statistical properties, our empirical findings suggest that there is not a unique winner in terms of risk management. On the one hand, lacking the tail risk problem, the information provided by ES would not mislead investors but, on the other, ES implies a greater chance of larger daily capital charges (DCC) than using VaR. It might be the case that ES is more appropriate in terms of minimizing DCC in times of high financial turmoil, while VaR might be more adequate during more tranquil times.

Contrary to previous research, our results show that there is a stochastic difference between 97.5% ES and 99% VaR. Being quite close in the Gaussian case, we also find that they are not the same. On the one hand, 97.5%-ES would imply higher risk forecast but, on the other, it would provide a smaller down-side risk than using 99%-VaR. This difference among risk measures is even more easily perceived for fat-tailed conditional distributions. The results of this paper support the Basel Committee on Banking Supervision (2013) proposal of moving from 99%-VaR to 97.5%-ES. ES not only accounts for the tail-risk, but it also provides a more stable measurement of down-side risk, and is often less sensitive to extreme observations.

Finally, while these results are obtained using a single asset and a limited range of models and distributions, they are likely to hold in a more general setting. The next step in this research project is to use a variety of assets, both from the USA and other countries, and also alternative univariate and multivaritate risk models to ascertain the validity of these empirical results.

References

Alexander, C. (2009), Market Risk Analysis: Value-at-Risk Models, Volume 4, Wiley, New York.

Artzner, P., F. Delbaen, J.M.Eber, and D. Heath (1997), Thinking coherently, Risk 10(11), 68–71.

Barrett, G., and S. Donald (2003), Consistent tests for stochastic dominance, Econometrica, 71, 71-104.

Basel Committee on Banking Supervision (2013), Consultative Document, Fundamental Review of the Trading Book: A revised Market Risk framework, BIS, Basel, Switzerland, http://www.bis.org/publ/bcbs265.pdf

Black, F. (1976), Studies of stock market volatility changes, in 1976 Proceedings of the American Statistical Association, Business & Economic Statistics Section, pp. 177-181.

Bollerslev, T. (1986), Generalised autoregressive conditional heteroskedasticity, Journal of Econometrics, 31, 307-327.

Caporin, M. and M. McAleer (2012),Model selection and testing of conditional and stochastic volatility models, in L. Bauwens, C. Hafner and S. Laurent (eds.), Handbook on Financial Engineering and Econometrics: Volatility Models and Their Applications, Wiley, New York, pp. 199-222.

Carlstein, E. (1992). Resampling Techniques for Stationary Time-Series: Some Recent Developments. New Directions in time series Analysis, 75–85.

Chang, C.-L., J-A. Jimenez-Martin, E. Maasoumi, and T. Pérez-Amaral (2015), A stochastic dominance approach to financial risk management strategies, to appear in Journal of Econometrics (http://www.sciencedirect.com/science/article/pii/S0304407615000573).

Chang, C.-L., J-A. Jimenez-Martin, M. McAleer, and T. Pérez-Amaral (2011), Risk management of risk under the Basel Accord: Forecasting value-at-risk of VIX futures, Managerial Finance, 37, 1088-1106.

Danielsson, J. (2013), "The new market regulations", VoxEU.org, 28 November 2013, http://www.voxeu.org/article/new-market-risk-regulations.

Donald, S. and Y. Hsu (2013), Improving the power of tests of stochastic dominance. Mimeo. http://yuchinhsu.yolasite.com/resources/papers/powerful_SD.pdf

Dowd, K. (2005), Measuring Market Risk. Wiley, New York, 2nd edition.

Engle, R.F. (1982), Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, Econometrica, 50, 987-1007.

Fama, E. F. and R. Roll (1968), Some properties of Symmetric Stable Distributions," Journal of the American Statistical Association, 63, 817-36.

Fama, E. F. and R. Roll (1971), Parameter estimates for symmetric stable distributions, Journal of the American Statistical Asssociation, 66, 331-38.

Feller, W. (1969), An Introduction to Probability Theory and Its Applications, Volume 2, John Wiley and Sons.

Franses, P.H. and D. van Dijk (1999), Nonlinear Time Series Models in Empirical Finance, Cambridge, Cambridge University Press.

Glosten, L., R. Jagannathan and D. Runkle (1992), On the relation between the expected value and volatility of nominal excess return on stocks, Journal of Finance, 46, 1779-1801.

Hansen, P.R. (2005), A test for superior predictive ability, Journal of Business and Economic Statistics, 23, 365-380.

Li, W.K., S. Ling and M. McAleer (2002), Recent theoretical results for time series models with GARCH errors, Journal of Economic Surveys, 16, 245-269. Reprinted in M. McAleer and L. Oxley (eds.), Contributions to Financial Econometrics: Theoretical and Practical Issues, Blackwell, Oxford, 2002,9-33.

Ling, S. and M. McAleer (2002a),Stationariety and the existence of moments of a family of GARCH processes, Journal of Econometrics, 106, 109-117.

Ling, S. and M. McAleer (2002b), Necessary and sufficient moment conditions for the GARCH(r,s) and asymmetric power GARCH(r,s) models, Econometric Theory, 18, 722-729.

Ling, S. and M. McAleer (2003a), Asymptotic theory for a vector ARMA-GARCH model, Econometric Theory, 19, 278-308.

Ling, S. and M. McAleer (2003b), On adaptive estimation in nonstationary ARMA models with GARCH errors, Annals of Statistics, 31, 642-674.

Linton, O., E. Maasoumi and Y.J. Whang (2005). Consistent testing for stochastic dominance under general sampling schemes. Review of Economic Studies, 72, 735-765.

McAleer, M. (2005), Automated inference and learning in modelling financial volatility, Econometric Theory, 21, 232-261.

McAleer, M. (2014), Asymmentry and leverage in conditional volatility models, Econometrics, 2(3), 145-150.

McAleer, M., F. Chan and D. Marinova (2007), An econometric analysis of asymmetric volatility: theory and application to patents, Journal of Econometrics, 139, 259-284.

McAleer, M. and C. Hafner (2014), A one line derivation of EGARCH, Econometrics, 2(2), 92-97.

Nelson, D.B. (1991), Conditional heteroskedasticity in asset returns: a new approach, Econometrica, 59, 347-370.

Politis, D.N. and J.P. Romano (1992), A Circular Block-resampling Procedure for Stationary Data,in R. Lepage and L. Billard (eds.), Exploring the Limits of Bootstrap, Wiley, New York, 263–270.

Sheppard, K. (2013), MFE Toolbox, https://www.kevinsheppard.com/MFE_Toolbox, 7 June, 2013.

Shiryaev, A.N. (1999), Essentials of Stochastic Finance, Facts, Models, Theory, World Scientific.

Tsay, R.S. (1987), Conditional heteroskedastic time series models, Journal of the American Statistical Association, 82, 590-604.

Yamai, Y. and T. Yoshiba (2002), Comparative analyses of expected shortfall and Value-at-Risk: Their estimation error, descomposition and optimization, Monetary and Economic Studies, 20, 2, Bank of Japan.

Yamai, Y. and T. Yoshiba (2005), Value-at-Risk versus Expected Shortfall: A practical perspective, Jounal of Banking and Finance, 29(4), 997-1015.

Figure 1

S&P500 Returns

1 January 1999 – 24 June 2014

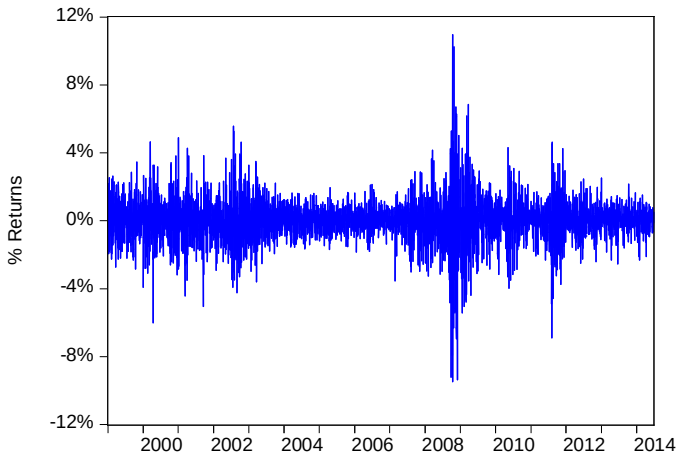
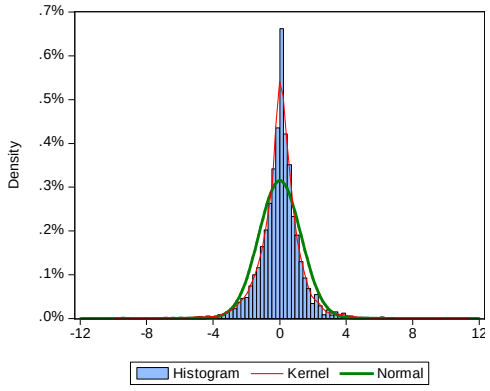


Figure 2

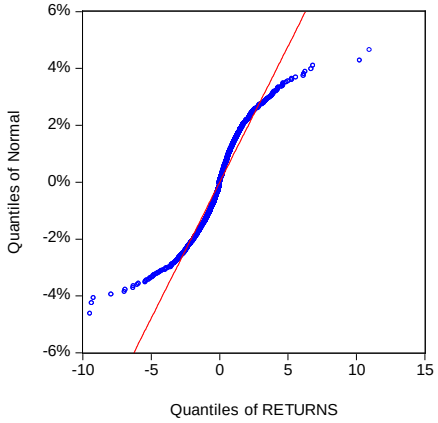
SP500 Returns distribution analysis

1 January 1999 – 24 June 2014

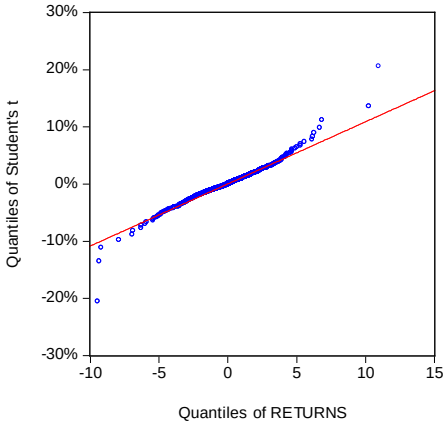
Panel A. Returns histogram, kernel density estimation and Gaussian theoretical density



Panel B. Q-Q plot Gaussian vs Returns



Panel C. Q-Q plot Gaussian vs Returns



Panel D. Q-Q plot Gaussian vs Returns

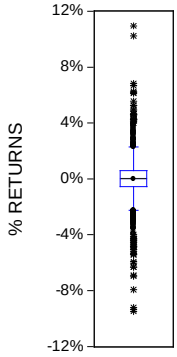


Figure 3

Volatility of S&P500 Returns

1 January 1999 – 24 June 2014

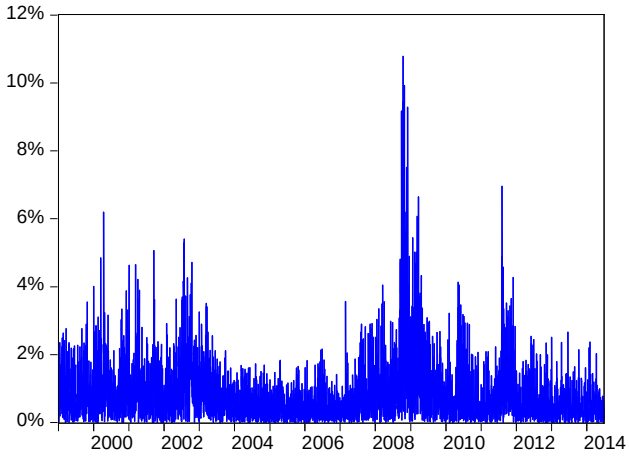
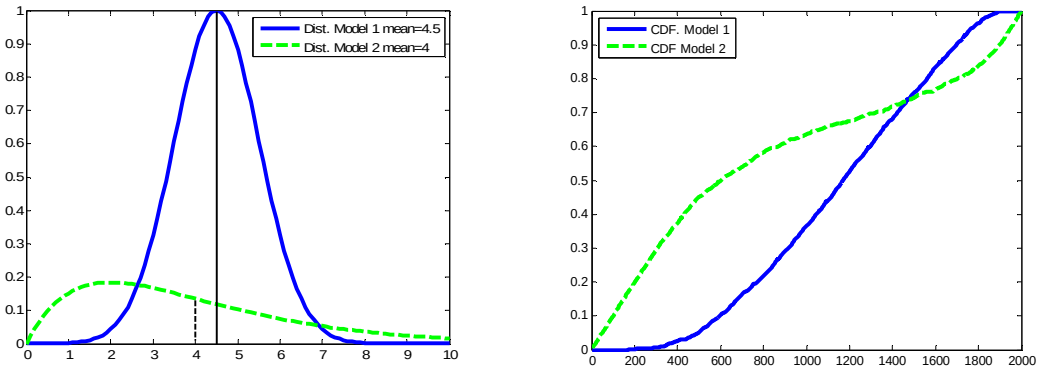


Figure 4

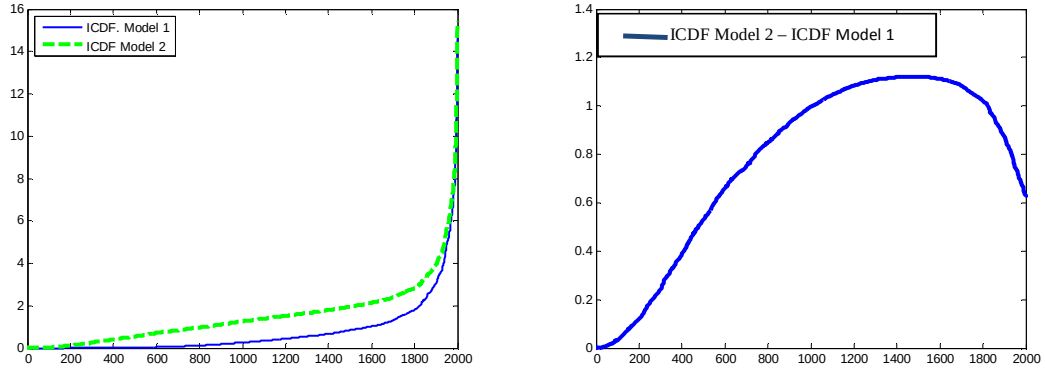
Probability density functions and CDFs of two distribution functions



Note: In the left panel, the solid line is for Model 1, a Gaussian density with mean 4.5 and standard deviation 1. The dashed line for Model 2 depicts a Chi-squared density with 4 degrees of freedom. The right panel depicts the CDFs for Models 1 (Gaussian) and 2 (Chi-square).

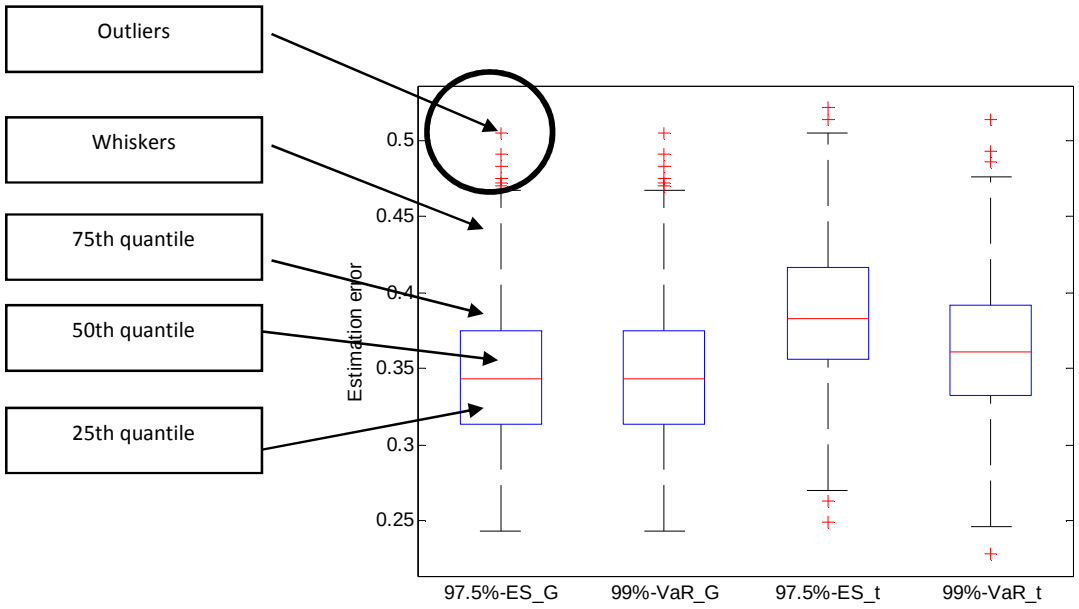
Figure 5

ICDF for Models1 and 2and their differences



Note: In the left panel, the solid line represents the integrated cumulative distribution function (ICDF) for a Gaussian variable with mean 4.5 and standard deviation1. The dashed line for model 2 depicts the ICDF of a 4-degrees of freedom Chi-squared variable. The right panel depicts the differences in the ICDFs given in the left panel.

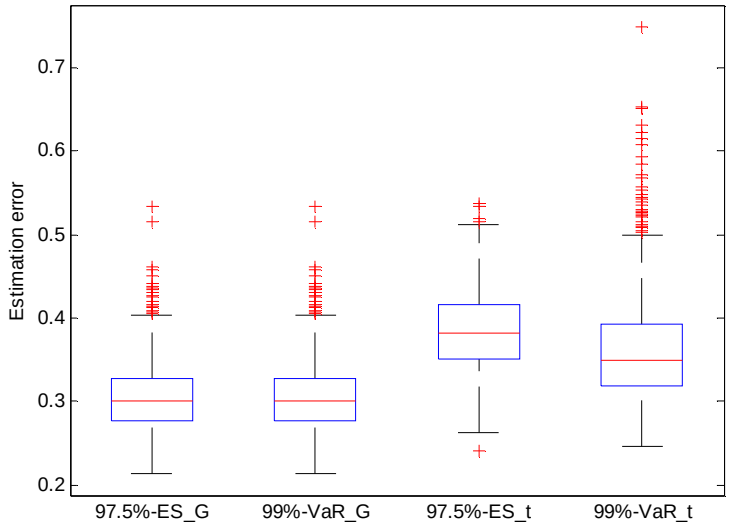
Figure 6A
99%-VaR and 97.5%-ES estimation error



VaR and ES for GARCH

Note: Figure 6A contains 4 boxplots: the first relies on the distribution of 97.5%-ES estimate errors assuming Gaussian (97.5%-ES_G), the second relies on the distribution of 99%-VaR estimate errors assuming Gaussian (99%-VaR_G), the third relies on the distribution of 97.5%-ES estimate errors assuming Student-t (97.5%-ES_G), and the fourth relies on the distribution of 99%-VaR estimate errors assuming Student-t (99%-VaR_t). The bottom of each boxplot is the 25th percentile, the top is the 75th percentile, and the line in the middle is the 50th percentile. "Whiskers" above and below each box are added to provide additional information about the spread of the data. Whiskers are the vertical lines that end in a horizontal stroke. Whiskers are drawn from the 75th and 25th percentiles to the most extreme values the algorithm considers not to be outliers. Additional marks beyond the whiskers are outliers.

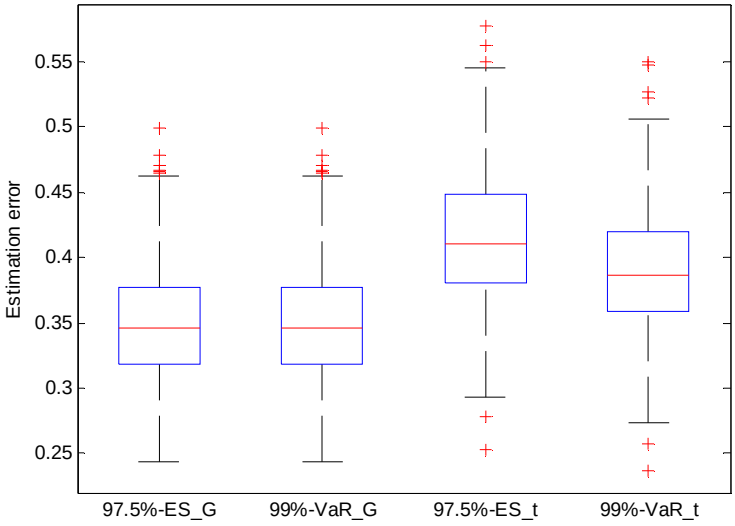
Figure 6B
99%-VaR and 97.5%-ES estimation error



VaR and ES for EGARCH

Note: Figure 6B contains 4 boxplots: the first relies on the distribution of 97.5%-ES estimate errors assuming Gaussian (97.5%-ES_G), the second relies on the distribution of 99%-VaR estimate errors assuming Gaussian (99%-VaR_G), the third relies on the distribution of 97.5%-ES estimate errors assuming Student-t (97.5%-ES_G), and the fourth relies on the distribution of 99%-VaR estimate errors assuming Student-t (99%-VaR_t). The bottom of each boxplot is the 25th percentile, the top is the 75th percentile, and the line in the middle is the 50th percentile. "Whiskers" above and below each box are added to provide additional information about the spread of the data. Whiskers are the vertical lines that end in a horizontal stroke. Whiskers are drawn from the 75th and 25th percentiles to the most extreme values the algorithm considers not to be outliers. Additional marks beyond the whiskers are outliers.

Figure 6C
99%-VaR and 97.5%-ES estimation error

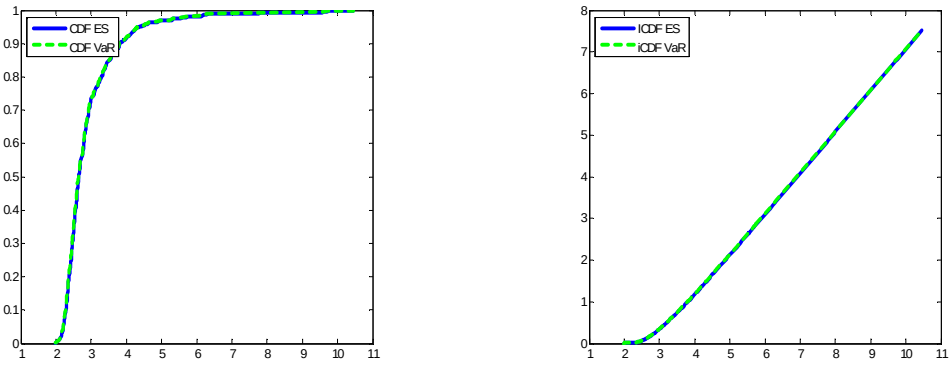


VaR and ES for GJR

Note: Figure 6C contains 4 boxplots: the first relies on the distribution of 97.5%-ES estimate errors assuming Gaussian (97.5%-ES_G), the second relies on the distribution of 99%-VaR estimate errors assuming Gaussian (99%-VaR_G), the third relies on the distribution of 97.5%-ES estimate errors assuming Student-t (97.5%-ES_G), and the fourth relies on the distribution of 99%-VaR estimate errors assuming Student-t (99%-VaR_t). The bottom of each boxplot is the 25th percentile, the top is the 75th percentile, and the line in the middle is the 50th percentile. "Whiskers" above and below each box are added to provide additional information about the spread of the data. Whiskers are the vertical lines that end in a horizontal stroke. Whiskers are drawn from the 75th and 25th percentiles to the most extreme values the algorithm considers not to be outliers. Additional marks beyond the whiskers are outliers.

Figure 7

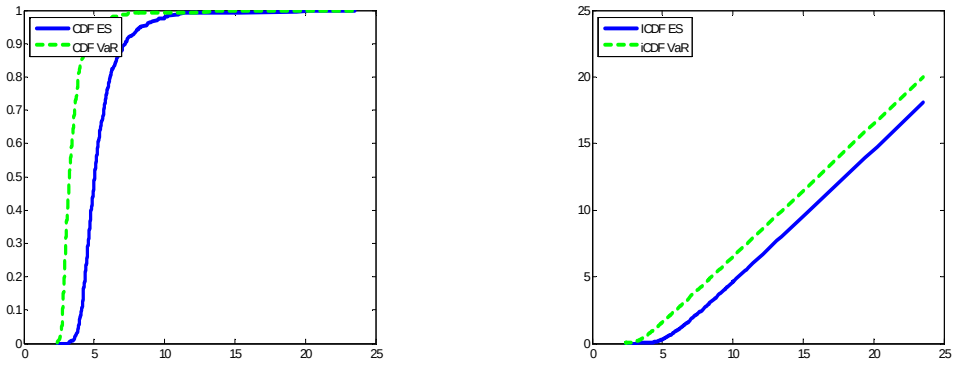
CDF and ICDF for 97.5%-ES and99%-VaR



Note: In the left panel, the solid line is the CDF of 97.5%-ES and the dashed line is the CDF of 99%-VaR produced by a GARCH model assuming a Gaussian distribution of S&P500 returns. In the right panel, the solid and dashed lines depict the integrated cumulative distribution functions (ICDF) of the CDFs shown in the left panel. These are empirical distributions of one-step-ahead 99%-ES and 99%-VaR forecasts of observation 3288 (8 August, 2011), corresponding to a negative return of S&P500 of 6.89%.

Figure 8

CDF and ICDF for 97.5%-ES and99%-VaR



Note: In the left panel, the solid line is the CDF of 97.5%-ES and the dashed line is the CDF of 99%-VaR produced by a GARCH model assuming a Student-t distribution of S&P500 returns. In the right panel, the solid and dashed lines depict the integrated cumulative distribution functions (ICDF) of the CDFs shown in the left panel. These are empirical distributions of one-step-ahead 99%-ES and 99%-VaR forecasts of observation 3288 (8 August, 2011), corresponding to a negative return of S&P500 of 6.89%.

Table 1
Rejection Rates forFirst-order SD Tests

	Gaussian			Student-t		
Design	**GARCH**	**EGARCH**	**GJR**	**GARCH**	**EGARCH**	**GJR**
BB	0.00	0.00	0.00	0.00	0.00	0.00
BD	0.00	0.00	0.00	0.00	0.00	0.00
LMW	0.00	0.00	0.00	0.00	0.00	0.00

Note. Rejection rates are from three different tests, namely Donald and Hsu (2013) (BB), Barrett and Donald (2003) (BD), and Linton, Maasoumi and Whang (2005) (LMW) for the null hypothesis: H_0: riskES FSD riskVaR, where RiskES and riskVaR denote the ES and VaR risk measurements, respectively. Forecast risk measures are produced using two probability distributions, Gaussian (left panel of the table) and Student t, and three conditional volatility models, stated in the first row of each table.

Table 2
Rejection Rates forSecond-order SD Tests

	Gaussian			Student-t		
Design	**GARCH**	**EGARCH**	**GJR**	**GARCH**	**EGARCH**	**GJR**
BB	0.00	0.00	0.00	0.00	0.00	0.00
BD	0.00	0.00	0.00	0.00	0.00	0.00
LMW	0.00	0.00	0.00	0.00	0.00	0.00

Note. Rejection rates are from three different tests, namely Donald and Hsu (2013) (BB), Barrett and Donald (2003) (BD), and Linton, Maasoumi and Whang (2005) (LMW) for the null hypothesis: H_0: riskES FSD riskVaR, where RiskES and riskVaR denote the ES and VaR risk measurements, respectively. Forecast risk measures are produced using two probability distributions, Gaussian (left panel of the table) and Student t, and three conditional volatility models, stated in the first row of each table.