THE UNIVERSITY of York

Discussion Papers in Economics

No. 2003/08

Accurate Measures of Value at Risk Fitting Fat Tails

by

J Samuel Baixauli and Susana Alvarez

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

ACCURATE MEASURES OF VALUE AT RISK FITTING FAT TAILS*

J. SAMUEL BAIXAULI
Department of Management and Finance
University of Murcia (Spain)

SUSANA ALVAREZ
Department of Quantitative Methods for the Economy
University of Murcia (Spain)

ABSTRACT

The calculus of both VaR and CVaR involves dealing with the confidence level, the time horizon and the true underlying conditional distribution function of asset returns. In this paper, we prove that using a distribution function that fits well to the data of the low tail of the observed distribution assure to obtain accurate VaR estimates. In so doing, we design bootstrap procedures to carry out nonparametric goodness-of-fit tests based on the Cramér-von Mises test statistic, constructed with the standardized residuals of the model and allowing for the postulated null distribution can depend on some unknown parameters to gain generality. Some functional forms characterized by capturing the thickness of the tails of high frequency data better than the normal distribution have been considered in this analysis.

***Keywords*:** Value at Risk, Conditional Value at Risk, fitted distribution, Cramér-von Mises test statistic, parametric bootstrap.

***Classification JEL*:** C15, G10

***** Adresses for correspondence: J.Samuel Baixauli, Departamento de Organizacion de Empresas y Finanzas, Facultad de Economia y Empresa, 30100 Campus de Espinardo, Murcia (Spain). Tel 00349687819, Fax: 0034968363792, e-mail: sbaixaul@um.es. SusanaAlvarez, Departamento de Metodos Cuantitativos para la Economia, Facultad de Economia y Empresa, 30100 Campus de Espinardo, Murcia (Spain). Tel: 0034968367902, Fax: 0034968367905, e-mail: salvarez@um.es.

ACCURATE MEASURES OF VALUE AT RISK FITTING FAT TAILS*

1. INTRODUCTION

Value-at-Risk (hereafter, VaR) is a popular measure of market risk, employed in the financial industry for both the internal control and regulatory reporting. For example, since 1998, U.S. banks and bank holding companies with significant amounts of trading activity are subject to market risk requirements. They have been required to hold capital against their defined market risk exposures, and, the capital charges are a function of banks' own VaR estimates. Hence, given the relevance of VaR estimates, the evaluation of the accuracy of the models underlying them is very important.

VaR associates the maximum amount that can be lost during a period to a determined statistic likelihood level. However, VaR does not quantify the losses beyond VaR and is not a sub-additive risk measure; i.e., the risk of a portfolio can be larger than the sum of the stand-alone risk of its components when measures by VaR (Acerbi *et al*., 2001). Hence, some alternative risk measures have been proposed, as Conditonal Value at Risk (CVaR). Szegö (2002) discusses the conditions under which the classical measures of risk can be used as well as describes in detail the main recently proposed risk measures. At a given confidence level, CVaR is defined as the expected loss given that the loss is greater than or equal to the VaR. The fundamental properties of CVaR are discussed in Acerbi and Tasche (2002) and Rockafellar and Uryasev (2002). The calculus of both VaR and CVaR involves dealing with the confidence level, the time horizon and the true underlying conditional distribution function of asset returns. The critical point is that usually risk management systems suppose that asset returns are normally distributed. This hypothesis has been widely tested in financial literature due to the true underlying distribution of high frequency data is more peaked and has fatter tails than the normal distribution (Mittnik *et al*., 2000). Some recent studies focused on the treatment of fat tails and VaR estimation have concluded that estimates of VaR based on normality assumption can underestimate the true value of VaR for big levels of probability and overestimate for small confidence levels (Hull and White, 1998,

* Financial support from DGES(PB98-0979) is gratefully acknowledged. We thank suggestions and useful comments received at the "XXX Meeting of the Euro Working Group for Financial Modelling"

Rockafellar and Uryasev, 2002). For these reasons, some alternative specifications have been proposed for calculating VaR (Bauer, 2000; Lucas, 2000; Vlaar, 2000).

The main objective of this paper is to prove that using a specific distribution function that fits well to the data of the low tail of the observed distribution of asset returns assure to obtain accurate VaR estimates. It is proved evaluating VaR estimates based on normality hypothesis as well as VaR estimates calculated once the low tail of the observed distribution of asset returns is modelled. The evaluation methods we use in this paper are the following (Lopez, 1999a, Lopez 2001): (i) Evaluation of VaR estimates based on the binomial distribution; (ii) Evaluation of VaR interval forecasts; (iii) Evaluation of VaR probability forecasts; (iv) Calibration tests of probability forecasts.

In order to achieve our objective, previously to calculate VaR estimates, we propose to carry out nonparametric goodness-of-fit tests based on the Cramér-von Mises test statistic, using the standardized residuals of the model and allowing for the distribution function postulated under the null hypothesis can depend on a vector of unknown parameters to gain generality. We consider as possible null specifications the Student's *t* distribution, the logistic distribution and the Edgeworth-Sargan distribution (Sargan, 1976), characterized by capturing the thickness of the tails of high frequency data better than the normal distribution. We compare the performance of these specifications using bootstrap procedures to approximate the distribution of Cramér-von Mises test statistic. Bootstrap methodology is required to implement this type of goodness-of-fit tests because the tabulated critical values can not be used as they have been deduced for the case in which the postulated null distribution is totally known and the observations are independent and identically observable random variables (Shorack and Wellner, 1986). However, the limiting distribution of Cramér-von Mises test statistic when it is constructed with standardized residuals and with maximum likelihood estimates of the unknown parameters depends on the postulated null distribution and, in general, on the unknown parameters (Koul, 1991).

(Capri, Italy 2002), at "The 22nd International Symposium on Forecasting" (Dublin, Ireland 2002), and at the "2nd International Conference of Finance" (Hammamet,Tunisia 2003).

The daily series used to carry out our analysis are the following indices: SP500, NYSE, NIKKEI, FOOTSIE, CAC40 and IGBM. The data cover the period from January 1994 to December 2002. In accordance with the regulatory framework, the accuracy of VaR estimates is assessed with respect to their one-step-ahead forecasts and 99% coverage levels.

The structure of the paper is as follows. Section 2 provides a review of VaR and CVaR concepts as well as illustrates the importance of accurate measures. Moreover, the bootstrap procedure designed to implement the goodness-of-fit tests is briefly described in the last Subsection of Section 2. Section 3 shows the data set and gives summary statistics. The empirical results obtained for the six stock-exchange indices are presented in Section 4. Finally, conclusions are summarized in Section 5.

2. METHODOLOGY

2.1. VaR and CVaR definitions

Value-at-Risk measures the level of loss that a single financial asset or a portfolio of financial assets could lose, with a given probability α, over a given time horizon Δt. Analytically, it can be formulated as follows:

$$\Pr(\Delta P_{\Delta t} \leq VaR) = \alpha , \qquad [1]$$

where, $\Delta P_{\Delta t}$ is a change in the market value of portfolio *P* over the time horizon Δt with probability α. Equation (1) states that the probability of losing more than VaR is α. Alternatively, VaR can be written in terms of the conditional distribution function of the asset returns, denoted $F_{\Omega_{t-1}}(.,\theta)$, where Ω_{t-1} is the information set and θ is the vector of parameters appearing in *F*. Then, VaR is the solution to:

$$\int_{-\infty}^{VaR(\Delta t,\alpha)} f_{t+\Delta t,\Omega_{t-1}}(x,\theta)dx = \alpha , \qquad [2]$$

where, $f_{\Omega_{t-1}}(.,\theta)$ denotes the conditional density function. Equivalently, if $F_{\Omega_{t-1}}^{-1}(.,\theta)$ denotes the inverse transformation of the conditional distribution function, VaR can be expressed as:

$$VaR(\Delta t,\alpha) = F_{\Omega_{t-1}}^{-1}(1-\alpha,\theta). \qquad [3]$$

At a given confidence level, CVaR is the expected loss given that the loss is greater than or equal to the VaR. CVaR can be defined as the mean of the α-tail distribution. That is,

$$CVaR(\Delta t,\alpha)=\alpha^{-1}\int_{-\infty}^{VaR(\Delta t,\alpha)} x f_{t+\Delta t,\Omega_{t-1}}(x,\theta)dx\,. \qquad [4]$$

Equations above show that the calculus of both VaR and CVaR involves dealing with the confidence level, the time horizon and the true underlying conditional distribution function of the asset returns. In the practise, the confidence level $(1-\alpha)$ is typically chosen to be at least 95% (and very often, as high as 99%) and, the time horizon varies with the use made of VaR by management and asset liquidity. It must be noted that both confidence level and time horizon depend on the risk aversion of the manager. Once the distribution function in the model is defined, VaR may be calculated using the Variance-Covariance method as:

$$VaR(\Delta t,\alpha)=c\times\hat{\sigma}_{t+\Delta t}\,, \qquad [5]$$

where, c is the $(1-\alpha)$ quantile of the defined distribution function and $\hat{\sigma}_{t+\Delta t}$ is the estimated standard deviation for time $t+\Delta t$.

2.2. Evaluation of VaR estimates

VaR is a popular measure of market risk, employed in the financial industry for both the internal control and regulatory reporting. For example, since 1998, U.S. banks and bank holding companies with significant amounts of trading activity are subject to market risk requirements. These requirements ("Market Risk Amendment", MRA) are based on the 1988 Basle Accord adopted by the Basle Committee on Banking Supervision. In particular, the U.S. standards apply to banks and bank holding companies with trading account positions exceeding $ 1 billion or 10% of total assets. They have been required to hold capital against their defined market risk exposures, and, the capital charges are a function of banks' own VaR estimates. The capital charge is calculated using the so-called "internal models" approach. Under this approach, capital charges are based on VaR estimates generated by banks' internal, risk management models using the standardizing parameters of a ten-day holding period and 99 percent coverage. In other words, a bank's market risk capital charge is based on its own estimate of the potential maximum loss that would not be exceeded with one-percent probability over the subsequent two-week period. For further discussion of this point, see Hendricks and Hirtle (1997), and, Lopez and Saidenberg (2000).

Hence, given the relevance of VaR estimates (to bank and to their regulators), the evaluation of the accuracy of the models underlying them is very important. Firstly, the proportion of failures (times that VaR estimates are exceeded by losses) can be calculated $(\hat{\alpha})$ and compared with the significance level (α). In the practise, VaR estimates can be evaluated using different methods. In this paper, we focus on: (i) Evaluation of VaR estimates based on the binomial distribution (currently embodied in the MRA); (ii) Evaluation of VaR interval forecasts; (iii) Evaluation of VaR probability forecasts; (iv) Calibration tests of probability forecasts. Methods (i) and (ii) use hypothesis tests to determine whether the VaR estimates exhibit a specified property that is a characteristic of accurate VaR estimates. However, as Diebold and Lopez (1996) noted, it is unlikely that forecasts from a model will exhibit all the properties of accurate forecasts. Also, methods (i) and (ii) can misclassify forecasts from inaccurate models as acceptably accurate given the low power exhibited by their corresponding hypothesis tests. Method (iii) is based on determining how well VaR estimates minimize a proposed regulatory loss function. Thus, by directly incorporating regulatory loss functions into the forecast evaluations, this method provides useful information on the performance of VaR models with respect to regulatory criteria as opposed to the purely statistical criteria implied by methods (i) and (ii). Finally, method (iv) is based on the degree of equivalence between the observed frequencies of the event "*losses lie under VaR estimate*" and the predicted frequencies of occurrence. Next, methods (i)-(iv) are briefly described:

(i) <u>Evaluation of VaR estimates based on the binomial distribution</u> (LR_{PF}): This method is based on the assumption that the VaR estimates are accurate, so the observations of the events "*losses lie under VaR*" and "*losses do not lie under VaR*" can be modeled as draws from an independent binomial random variable with a probability of occurrence equal to a specified α percent. The null hypothesis "the empirical size of the test ($\hat{\alpha}$) is equal to the nominal size (α)" is tested versus the alternative "$\hat{\alpha} \neq \alpha$", using the following likelihood ratio test statistic based on the binomial distribution:

$$LR_{PF} = 2\left[\ln\left\{\hat{\alpha}^{z}(1-\hat{\alpha})^{T-z}\right\} - \ln\left\{\alpha^{z}(1-\alpha)^{T-z}\right\}\right], \qquad [6]$$

where, z denotes the number of times the loss lies under the estimated VaR. Under the null hypothesis, LR_{PF} is asymptotically distributed as $\chi^2(1)$ (Kupiec, 1995). As Kupiec (1995) and Lopez (1999b) noted, when small only sample size is available, the finite sample distribution may be quite different from $\chi^2(1)$ distribution and the asymptotic critical values may be inappropriate.

(ii) Evaluation of VaR interval forecasts (LR_{CC}): VaR estimates can be viewed as interval forecasts of the lower 1% tail of the one-step-ahead return distribution and, these forecasts can be evaluated conditionally or unconditionally. Given the presence of variance dynamics when high-frequency data are modelled, testing for conditional accuracy is important. LR_{CC} was proposed by Christoffersen (1998) as a test of the conditional coverage level in contrast to LR_{PF}, that ignores the presence of the time-dependence. The LR_{CC} test is a joint test of both correct unconditional coverage and serial independence. The test statistic is defined as:

$$LR_{CC} = LR_{PF} + LR_{ind}, \qquad [7]$$

and, under the null hypothesis of correct conditional coverage level, is asymptotically distributed as $\chi^2(2)$. The LR_{ind} statistic is a likelihood ratio statistic of the null hypothesis of serial independence against the alternative of first-order Markov dependence. The likelihood function under this alternative hypothesis is $L_A = (1-\pi_{01})^{T_{00}}\pi_{01}^{T_{01}}(1-\pi_{11})^{T_{100}}\pi_{11}^{T_{11}}$, where the T_{ij} notation denotes the number of observations in state j after having been in state i the period before, $\pi_{01} = T_{01}/(T_{00}+T_{01})$ and $\pi_{11} = T_{11}/(T_{10}+T_{11})$. Under the null hypothesis of independence, $\pi_{01} = \pi_{11} = \pi$ and the likelihood function is $L_0 = (1-\pi)^{T_{00}+T_{10}}\pi^{T_{01}+T_{11}}$, where $\pi = (T_{01}+T_{11})/T$. Hence, $LR_{ind} = 2[\log L_A - \log L_0]$, which is asymptotically distributed as $\chi^2(1)$. Hence, LR_{CC} is used to test the null hypothesis "$\hat{\alpha} = \alpha$" versus the alternative "$\hat{\alpha} \neq \alpha$", taking into account the presence of time-dependence, often found in financial time series. However, the finite distribution, for a specific α and sample size T, may differ from a $\chi^2(2)$ distribution.

(iii) <u>Evaluation of VaR probability forecasts</u> (*QPS*): The accuracy of VaR estimates is gauged by how well they minimize a loss function that represents the evaluator's concerns. It was proposed by Lopez (1999a), who examined the statistical power for this evaluation model with simulation experiments and concluded that the degree of model misclassification generally mirrors that of models (i) and (ii). Different loss functions can be consider, as for example, the quadratic probability score (QPS), developed by Brier (1950), that assign a quadratic numerical score when a VaR estimate is exceeded by its corresponding loss. In this case and given a sample of size *T*, QPS is defined as:

$$QPS = \frac{1}{T}\sum_{t=1}^{T} 2\left(P_t - Q_{t+1}\right)^2 , \qquad [8]$$

where $P_t = \int_{-\infty}^{CV(\alpha,\hat{F})} f_{t+1}(x)dx$, $\hat{F}$ is the fitted distribution function, $CV(\alpha,\hat{F}) = \hat{F}^{-1}(\alpha)$ is the unconditional quantile of interest and, Q_{t+1} is an indicator variable that equals to one if the VaR estimate is exceeded by its corresponding asset or portfolio loss and, zero otherwise. It must be highlight that QPS$\in[0,2]$ and smaller values of QPS indicate more accurate forecasts. That is, if the QPS for a model *A* is closer to zero than the QPS for a model *B*, then the forecasts from model *A* are more accurate than those for model *B*. However, one could be interested in determining if the difference between the QPS for model *A* and the QPS for model *B* is statistically significant. In doing this, a generalization to probability forecasts of a test proposed by Diebold and Mariano (1995) can be used since QPS can be viewed as the analog of mean square error for probability forecasts. Diebold and Mariano (1995) proposed a test for determining, under a general loss function, whether the expected losses induced by two sets of *point* forecasts are statistically different. In particular, when the assumed loss function is a quadratic loss function (QPS), denoting

$$d_t = 2\left(P_{At} - Q_{t+1}\right)^2 - 2\left(P_{Bt} - Q_{t+1}\right)^2 ,\ t = 1,\ldots,T \qquad [9]$$

the Diebold and Mariano (1995) test statistic is expressed as:

$$S = \frac{\bar{d}}{\hat{\sigma}_d / \sqrt{T}} , \qquad [10]$$

where $\bar{d}$ is the sample mean of *d*, $\hat{\sigma}_d$ represents the standard deviation of *d*. Under the null hypothesis that "the expected losses under the model *A* and model *B* are equal (QPS_A=QPS_B)", the test statistic *S* is asymptotic distributed as *N*(0,1).

(iv) Calibration tests of probability forecasts (Seillier-Moiseiwitsch and Dawid, 1993): It is based on the degree of equivalence between the observed frequencies of the event "*VaR estimate is exceeded by its corresponding loss*" and the predicted frequencies of occurrence; then, if those frequencies are similar, the forecasts are well calibrated. To analyze if a model is well calibrated, a forecaster does the following: for the event "*VaR estimate is exceeded by its corresponding loss*" and given a corresponding sequence of forecast probabilities, the forecaster creates *J* mutually exclusive subsets of these forecast probabilities. The forecaster denotes the midpoint of each range π_j and the number of observed losses belonging to the set *j* Q_j, *j=1,...,J*. If the model is well calibrated, for every set *j*, Q_j is similar to $e_j = T_j \pi_j$, where T_j is the number of forecast probabilities in set *j*. Hence, for each set *j*, the test statistic is defined as:

$$Z_j = \frac{Q_j - e_j}{\{e_j (1-\pi_j)\}^{1/2}}, \qquad [11]$$

The overall calibration test statistic is expressed as

$$Z_0 = \sum_{j=1}^{J} Z_j . \qquad [12]$$

Under the null hypothesis " the observed frequencies and the predicted frequencies are equal", the test statistic Z_0 is asymptotically distributed as *N*(0,1).

2.3 Bootstrap procedure for fitting fat tails

As it was said in Subsection 2.1, the calculus of both VaR and CVaR involves also dealing with the true underlying conditional distribution function of the asset returns. The critical point is that usually risk management systems suppose that asset returns are normally distributed. Some recent studies focused on the treatment of fat tails and VaR estimation have concluded that normality assumption is not reasonable (Hull and White, 1998, Rockafellar and Uryasev, 2002). Hence, some alternative specifications have been proposed for calculating VaR (Bauer, 2000; Lucas, 2000; Vlaar, 2000). As it does not seem a convenient hypothesis to assume normality, we consider relevant to analyze the behaviour of the distribution of the asset returns previously to calculate VaR and CVaR estimates. It is expected that choosing a distribution function that fits well to the data of the low tail of the observed distribution of assets allows obtaining obtain accurate risk measures.

In order to select a distribution function that captures well the low tail of the observed distribution of asset returns, we carry out nonparametric goodness-of-fit tests based on the Cramer-von Mises test statistic, constructed with the standardized residuals and substituting the unknown vector of parameters appearing in the postulated null distribution $F(.,\theta)$ by an estimate. The standardized residuals of the model are defined as $e_t = (R_t - \hat{\mu}_t)/\hat{\sigma}_t$, where R_t is the asset returns at time t and, $\hat{\mu}_t$ and $\hat{\sigma}_t$ are the maximum likelihood estimates of μ_t and σ_t, respectively. In this case, the expression of the Cramer-von Mises test statistic is:

$$\hat{W}_T^2 = \sum_{i=1}^{T} \left\{ \hat{F}_T(e_i) - F(e_i, \hat{\theta}) \right\}^2, \qquad [13]$$

where, we define:

$$\hat{F}_T(x) = \frac{1}{T} \sum_{i=1}^{T} I(e_i \leq x), \qquad [14]$$

for $x \in \Re$, $\hat{\theta}$ is the maximum likelihood estimate of the vector of unknown parameters appearing in $F(.,\theta)$ and $I(.)$ is the indicator function.

As we are interesting in fitting the data of the low tail of the observed distribution, we construct the test statistic $\hat{W}_T^2$ using only the standardized residuals of the first quartile of the distribution. We use parametric bootstrap to implement the test (Gine and Zinn, 1990). The stages of the bootstrap procedures are:

Stage 1: Let $\{R_t\}_t^T$ be a sequence of returns of an asset, following the equation $R_t = \mu_t + \sigma_t u_t$, where u_t are independent and identically distributed errors from a particular $F(.,\theta)$, with mean zero, unitary variance and depending on a vector of unknown parameters $\theta \in \Re^s$. Select the model that better describes the conditional behaviour of the asset returns.

Stage 2: Considering the distribution postulated under the null hypothesis, estimate by maximum likelihood μ_t, σ_t^2 and θ. In this way, the standardized residuals $e_t = (R_t - \hat{\mu}_t)/\hat{\sigma}_t$ and the estimate of the null distribution $F(.,\hat{\theta})$ are obtained. Draw the empirical distribution of the standardized residuals and use the data of the first quartile to evaluate $\hat{W}_T^2$.

Stage 3: Repeat B=500 times the following:

- Generate a sample of random variables from $F(.,\hat{\theta})$, $u_1^*, u_2^*, ..., u_T^*$. Using $\hat{\mu}_t$, $\hat{\sigma}_t$ and u_t^*, $R_t^* = \hat{\mu}_t + \hat{\sigma}_t u_t^*$ is obtained, *t=1,...,T*. Calculate new maximum likelihood estimates $\hat{u}_t^*, \hat{\sigma}_t^{2*}, \hat{\theta}^*$ and construct $v_t^* = (R_t^* - \hat{\mu}_t^*)/\hat{\sigma}_t^*$.

- Evaluate the test statistic using $v_1^* \leq ... \leq v_{(0.25 \cdot T)}^*$ and $F(.,\hat{\theta}^*)$. It is denoted $\hat{W}_T^{2*}$.

Stage 4: In this way, a sample of *B* independent (conditionally on the original sample) observations of $\hat{W}_T^2$, say $\hat{W}_{T_1}^{2*}, ..., \hat{W}_{T_B}^{2*}$, is obtained.

Stage 5: Let $\hat{W}_{T_{(1-\alpha)B}}^{2*}$ the (1-α)*B*-th order statistic of the sample $\hat{W}_{T_1}^{2*}, ..., \hat{W}_{T_B}^{2*}$, given a significance level α. Reject the null hypothesis at the significance level α if $\hat{W}_T^2 > \hat{W}_{T_{(1-\alpha)B}}^{2*}$.

Stage 6: Compute the bootstrap p-value as $p_B = card(\hat{W}_{T_b}^{2*} \geq \hat{W}_T^2)/B, b = 1, ..., B.$

Given a significance level α, we compute the exact critical value *c* of the chosen $F(.,\hat{\theta})$ as its (1-α) quantile. Therefore, an asset VaR is just the product of this critical value and the estimated conditional deviation for such model. That is, $VaR(\Delta t, \alpha) = c \cdot \hat{\sigma}_{t+\Delta t}$.

In our analysis, we select various functional forms as possible null distributions, all of them characterized by capturing the thickness of the tails of high frequency financial data better than the normal distribution. When more than one of these possible distribution functions can be accepted, we select that whose bootstrap p-value is the biggest. The functional forms considered are: Student's *t* distribution, logistic distribution and Edgeworth-Sargan (E-S) distribution. If μ_t and σ_t^2 represent the mean and the variance of the asset returns at time *t* respectively, then the corresponding density functions are summarized in Table 1.

Table 1. Conditional density functions

Distribution	Conditional density functions
Student's t	$f_{\Omega_{t-1}}(R_t;\mu_t,\sigma_t^2,\theta=g)=\dfrac{\Gamma\left(\frac{g+1}{2}\right)}{\Gamma\left(\frac{g}{2}\right)\Gamma\left(\frac{1}{2}\right)\sqrt{\sigma_t^2(g-2)}}\left\{1+\dfrac{(R_t-\mu_t)^2}{\sigma_t^2(g-2)}\right\}^{-\frac{(g+1)}{2}}$
Logistic	$f_{\Omega_{t-1}}(R_t;\mu_t,\sigma_t^2)=\dfrac{\pi}{\sqrt{3\sigma_t^2}}\dfrac{1}{\left\{1+\exp\left(\frac{\pi(R_t-\mu_t)}{\sqrt{3\sigma_t^2}}\right)\right\}^2}\exp\left(\dfrac{\pi(R_t-\mu_t)}{\sqrt{3\sigma_t^2}}\right)$
Edgeworth-Sargan	$f_{\Omega_{t-1}}(R_t;\mu_t,\sigma_t^2,\theta)=\dfrac{1}{\sqrt{2\pi\sigma_t^2}}\exp\left\{-\dfrac{(R_t-\mu_t)^2}{2\sigma_t^2}\right\}\cdot$ $\cdot\left\{1+d_2H_2\left(\frac{R_t-\mu_t}{\sigma_t}\right)+d_3H_3\left(\frac{R_t-\mu_t}{\sigma_t}\right)+d_4H_4\left(\frac{R_t-\mu_t}{\sigma_t}\right)\right\},$ $\theta=(d_2,d_3,d_4)',$ $H_2\left(\frac{R_t-\mu_t}{\sigma_t}\right)=\left(\frac{R_t-\mu_t}{\sigma_t}\right)^2-1$ $H_3\left(\frac{R_t-\mu_t}{\sigma_t}\right)=\left(\frac{R_t-\mu_t}{\sigma_t}\right)^3-3\left(\frac{R_t-\mu_t}{\sigma_t}\right)$ $H_4\left(\frac{R_t-\mu_t}{\sigma_t}\right)=\left(\frac{R_t-\mu_t}{\sigma_t}\right)^4-6\left(\frac{R_t-\mu_t}{\sigma_t}\right)^2+3$

Note: This reparametrization of Student's *t* avoids the influence of *g* in the variance, *g* denotes the degrees of freedom and d_2, d_4, d_6 are parameters appearing in the Edgeworth-Sargan distribution.

3. DATA ANALYSIS

The daily series used in the empirical analysis are six stock-exchange indices from different countries: SP500 and NYSE, from American Market; NIKKEI, from Japan; FOOTSIE, from UK; CAC40, from France; and, IGBM, from Spain. The daily return index is computed as $R_t = 100\cdot\log(S_t/S_{t-1})$, where S_t is the closing price at day *t*.

The data cover the period from January 1994 to December 2002. To implement the procedure described in the previous Section, we divide this time period into two smaller periods: the first, from January 1994 to May 2001; the second, from June 2001 to December 2002. Period 1 is used to estimate the model, choose the functional form that fits to the data of the low tail of the observed distribution and compare "in-sample" predictive power of VaR. Period 2 is used to compare "out-sample" predictive power of VaR.

Table 2 gives summary statistics and two normality tests: Jarque-Bera normality test and Kolmogorov-Smirnov normality test with Lilliefors correction (Lilliefors, 1967).

Table 2. Descriptive measures (3/1/1994-31/5/2001)

Index	Skewness	Kurtosis	Jarque-Bera	KS Lilliefors	Q(5)	Q(20)	Q*(5)	Q*(20)
SP500	-0.2755	7.7207	1750.6 (0.000)	0.068 (0.000)	11.754 (0.038)	29.862 (0.072)	216.64 (0.000)	436.01 (0.000)
NYSE	-0.3860	9.6101	3412.2 (0.000)	0.072 (0.000)	14.763 (0.011)	38.643 (0.007)	194.93 (0.000)	332.00 (0.000)
NIKKEI	0.1760	6.1886	783.42 (0.000)	0.060 (0.000)	13.114 (0.022)	41.232 (0.003)	113.53 (0.000)	226.68 (0.000)
FOOTSIE	-0.0963	4.5809	194.14 (0.000)	0.044 (0.000)	19.128 (0.002)	31.602 (0.048)	234.18 (0.000)	644.16 (0.000)
CAC40	-0.0929	4.4608	162.55 (0.000)	0.033 (0.000)	5.726 (0.334)	17.713 (0.606)	106.69 (0.000)	359.42 (0.000)
IGBM	-0.5053	6.3187	922.76 (0.000)	0.047 (0.000)	24.322 (0.000)	57.724 (0.000)	193.49 (0.000)	723.34 (0.000)

Note: In this table, we compute the descriptive measures: skewness, kurtosis and the test of normality Bera-Jarque test statistic and Kolmogorov-Smirnov test with Lilliefors correction. Q(5) and Q(20) are Box-Ljung statistic for the serie with 5 and 20 sample autocorrelations; $Q^*(5)$ and $Q^*(20)$ are Box-Ljung statistic for the squared serie with 5 and 20 sample autocorrelations. Between parentheses is the p-value of the test.

It can be observed that leptokurtosis is apparent in all cases. Moreover, all the series fail to pass both Jarque-Bera normality test and Kolmogorov-Smirnov normality test with Lilliefors correction. To check if the data present dynamic structure in the mean and in the variance, we compute Box-Ljung statistic for the serie, Q(5) and Q(20), and for the squared serie, Q*(5) and Q*(20), constructed with 5 and 20 sample autocorrelations. The results with the series show smooth dynamic in the mean while, as expected, squared series show strong evidence of autocorrelation. Therefore, we have to examine the performance of alternative models for conditional heteroskedastic time series with the observations R_t, described above. The models analyzed consist of two equations: one of them specifies a decomposition of R_t to capture the mean dependence of the serie and, the other one is the volatility equation. In order to capture the dynamic structure in the mean, we have fitted various autorregresive moving-average (ARMA) models[1]. We have selected different GARCH specifications[2] to analyze the behaviour of the conditional variance. We have estimated all models using the maximum likelihood procedure, with constant term and without constant term. All the estimated models are compared with the Schwarz Information Criterion (SIC); therefore, the preferred model is the model with highest SIC.

Table 3 summarizes the equations for the model used as well as reports the Box-Ljung test statistics for the series of standardized maximum likelihood residuals, $Q_r(5)$ and $Q_r(20)$, and for the series of standardized squared serie, $Q_r^*(5)$ and $Q_r^*(20)$.

Table 3. Selected models and diagnosis

Index	Selected model	$Q_r(5)$	$Q_r(20)$	$Q_r^*(5)$	$Q_r^*(20)$
SP500	$R_t = \phi_0 + h_t^{1/2}\eta_t$	7.166	27.499	8.972	19.239
	$h_t = \omega + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1}$	(0.209)	(0.122)	(0.110)	(0.506)
NYSE	$R_t = \phi_0 + h_t^{1/2}\eta_t$	7.702	25.079	5.688	11.321
	$h_t = \omega + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1}$	(0.173)	(0.198)	(0.338)	(0.937)
NIKKEI	$R_t = \phi_1 R_{t-1} + h_t^{1/2}\eta_t$	0.737	18.234	2.066	14.951
	$h_t = \omega + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1}$	(0.981)	(0.572)	(0.840)	(0.779)
FOOTSIE	$R_t = \phi_0 + \phi_1 R_{t-1} + \phi_2 R_{t-2} + h_t^{1/2}\eta_t$	9.933	23.964	5.611	15.098
	$h_t = \omega + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1}$	(0.077)	(0.244)	(0.346)	(0.771)
CAC40	$R_t = h_t^{1/2}\eta_t$	2.926	20.694	1.295	20.914
	$h_t = \omega + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1}$	(0.711)	(0.415)	(0.935)	(0.402)
IGBM	$R_t = \phi_0 + \phi_1 R_{t-1} + h_t^{1/2}\eta_t$	6.723	19.956	3.213	13.298
	$h_t = \omega + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1}$	(0.242)	(0.461)	(0.667)	(0.864)

Note: *where* $\varepsilon_t \equiv h_t^{1/2}\eta_t$ η_t i.i.d; $E[\eta_t] = 0, E[\eta_t^2] = 1$; $Q_r(5)$ and $Q_r(20)$ are Box-Ljung statistic for the serie of standardized residuals; $Q_r^*(5)$ and $Q_r^*(20)$ are Box-Ljung statistic for the squared serie of standardized residuals. Between parentheses is the p-value of the test.

It is worth noting that the dynamic structure in the mean is modelled with AR(1) in NIKKEI and IGBM cases, AR(2) in FOOTSIE case while in SP500, NYSE and

[1] We have considered AR(1), AR(2), MA(1), MA(2), ARMA(1,1), ARMA(1,2), ARMA(2,1) models.

CAC40 cases a model is not needed. The GARCH(1,1) model is the preferred model to capture the dynamic structure in the variance. As it can be seen in Table 3, the corresponding values of $Q_r(5)$, $Q_r(20)$ and $Q_r^*(5)$ and $Q_r^*(20)$ show that both dynamics are well-captured.

4. EMPIRICAL RESULTS

In this section, we present the results for the six indices introduced before. Given the regulatory framework, all the VaR estimates are calculated one-step-ahead and using 99% coverage level. Firstly, following the classical approach, we assume that asset returns are normally distributed and we calculate VaR estimate for each index. Secondly, as normality is not a reasonable hypothesis, we consider that the true underlying distribution function of each index can be different from normal distribution. In particular, we propose the following alternative specifications: Student's *t* distribution, logistic distribution and Edgeworth-Sargan distribution. Hence, assuming each of these specific distribution functions, the maximum likelihood estimation of AR(p)-GARCH(p,q) models is carried out again and new VaR estimates are calculated. Results of 1% VaR estimates assuming normality and the proposed alternative specifications of each index are evaluated using the methods described in Subsection 2.2. Table 4 reports the "in-sample" evaluation results.

[2] We have considered ARCH(1), ARCH(2), ARCH(3), ARCH(4), GARCH(1,1), GARCH(1,2), GARCH(2,1) models.

Table 4. Accuracy of VaR from different distributions forms (3/1/1994-31/5/2001)

Panel (1): percentage of the exceedings $(\hat{\alpha})$

Index	*Normal*	*Student t*	*Logistic*	*E-S*
SP500	0.0172	0.0118	0.0145	0.0134
NYSE	0.0189	0.0129	0.0140	0.0135
NIKKEI	0.0142	0.00986	0.0115	0.0126
FOOTSIE	0.0179	0.0136	0.0109	0.0141
CAC40	0.0150	0.0116	0.0100	0.0127
IGBM	0.0152	0.0125	0.0119	0.0125

Panel (2): Evaluation of VaR estimates based on the binomial distribution (LR_{PF})

Index	*Normal*	*Student t*	*Logistic*	*E-S*
SP500	8.0371 (0.0045)	0.5967 (0.4398)	3.3719 (0.0663)	2.0149 (0.1558)
NYSE	11.815 (0.0005)	1.5221 (0.2173)	2.7441 (0.0976)	2.0926 (0.1480)
NIKKEI	2.9377 (0.0865)	0.0034 (0.9535)	0.3991 (0.5276)	1.1536 (0.2829)
FOOTSIE	9.5686 (0.0019)	2.1940 (0.1389)	0.1474 (0.701)	2.8609 (0.0908)
CAC40	3.9609 (0.0465)	0.4861 (0.4857)	0.0000 (0.9962)	1.3009 (0.2542)
IGBM	4.3836 (0.0381)	1.0863 (0.2986)	0.6774 (0.4106)	1.0863 (0.2973)

Panel (3): Evaluation of VaR interval forecasts (LR_{CC})

Index	*Normal*	*Student t*	*Logistic*	*E-S*
SP500	9.1581 (0.0102)	1.1237 (0.5703)	4.1678 (0.1244)	2.6965 (0.2597)
NYSE	13.167 (0.0013)	2.1537 (0.3407)	3.4862 (0.1755)	3.4862 (0.1755)
NIKKEI	3.6962 (0.1575)	1.8932 (0.3880)	1.7794 (0.4107)	2.2552 (0.3238)
FOOTSIE	10.778 (0.0045)	2.8850 (0.2363)	0.5884 (0.7451)	3.6088 (0.1652)
CAC40	4.6043 (0.1000)	0.9825 (0.6120)	0.3640 (0.8336)	2.3808 (0.3042)
IGBM	5.2500 (0.0724)	1.6693 (0.4340)	1.2105 (0.5459)	1.6693 (0.4340)

Panel (4): Evaluation of VaR probability forecasts (QPS)

Index	*Normal*	*Student t*	*Logistic*	*E-S*
SP500	0.0397	0.0264	0.0307	0.0300
NYSE	0.0390	0.0266	0.0289	0.0277
NIKKEI	0.0287	0.0198	0.0231	0.0254
FOOTSIE	0.0220	0.0220	0.0220	0.0265
CAC40	0.0299	0.0233	0.0200	0.0254
IGBM	0.0306	0.0251	0.0240	0.0251

Panel (5): Comparisons of accuracy of probability forecasts. Z_0

Index	*Normal*	*Student t*	*Logistic*	*E-S*
SP500	-	-3.590 (0.0003)	-2.985 (0.0028)	-3.020 (0.0025)
NYSE	-	-3.483 (0.0004)	-3.136 (0.0017)	-3.325 (0.0008)
NIKKEI	-	-2.891 (0.0038)	-2.306 (0.0211)	-1.351 (0.1766)
FOOTSIE	-	-0.343 (0.7315)	1.025 (0.3053)	-1.268 (0.2047)
CAC40	-	-2.474 (0.0133)	-3.022 (0.0025)	-2.049 (0.0404)
IGBM	-	-2.260 (0.0238)	-2.477 (0.0132)	-2.258 (0.0239)

Panel (6): Calibration tests of probability forecasts (Seillier-Moiseiwitsch and Dawid, 1993)

Index	*Normal*	*Student t*	*Logistic*	*E-S*
SP500	2.8030 (0.0051)	0.1066 (0.9155)	1.8812 (0.0599)	-0.5414 (0.5882)
NYSE	3.5459 (0.0003)	0.2314 (0.8173)	0.9072 (0.3643)	0.0812 (0.9352)
NIKKEI	3.6208 (0.0002)	0.1906 (0.8493)	1.3055 (0.1917)	0.9221 (0.3564)
FOOTSIE	2.2740 (0.0229)	1.2257 (0.2224)	0.5953 (0.5516)	0.3520 (0.7248)
CAC40	6.3930 (0.0000)	1.6716 (0.0949)	0.0896 (0.9290)	1.7678 (0.0771)
IGBM	4.3034 (0.0000)	1.8993 (0.0587)	1.1301 (0.2584)	1.3412 (0.1798)

Note: The accuracy of VaR estimates is assessed with respect to their one-step-ahead forecasts and 99% coverage level. Between parentheses is the p-value of the test.

Panel 1 shows the percentage of the exceedings $(\hat{\alpha})$ for the normal, Student's *t*, logistic and Edgeworth-Sargan cases. It can be observed that differences between $\hat{\alpha}$ and α under normality are higher than under the other specifications for the six indices. Panel 2 and 3 report the evaluation of VaR estimates based on LR_{PF} and LR_{CC} test statistics. Given a 5% significance level, results for LR_{PF} show that the null hypothesis of equality between the proportion of the exceedings and the significance level is rejected in almost all cases when normality is assuming. The null hypothesis is only accepted for NIKKEI. In contrast to these results, the null hypothesis is accepted when one of the alternative specifications is assumed. Similar results can be observed using LR_{CC} test statistic. However, due to fact that both the finite sample size and the power characteristics of these tests have been discussed (Diebold and Lopez, 1996), we evaluated VaR estimates using two additional evaluation methods, the method based on determining how well VaR estimates minimize a quadratic loss function (QPS) and the method based on comparing the forecasted probabilities to observed relative frequencies (Calibration test). Panel 4 shows the value of QPS under each distributional assumption. It can be observed that the maximum QPS value for each index is the associated to the normality assumption. The results of testing if the differences of QPS under normality and QPS under Student's *t*, logistic and Edgeworth-Sargan distribution are statististical significant are reported in Panel 5. In this panel, it can be observed that such differences (normality=model *A* in equation [9] and Student's *t*, logistic or Edgeworth-Sargan=model *B* in equation [9]) are statistically different from zero, exceed by FOOTSIE. Finally, calibration tests are carried out, since a straightforward matter to evaluate probability forecasts is to compare the forecasted probabilities to observed relative frequencies. The results are shown in Panel 6: given a significance level equals to 5%, the model is well calibrated under each of the specific alternative specifications while it is not well calibrated when normality is assumed.

The results reported in Table 4 permit to conclude that VaR estimates calculated under Student's *t*, logistic and the Edgeworth-Sargan distributions are accurate in contrast to VaR estimates computed under the normality assumption for the six indices analyzed. This conclusion is relevant, specially if we observe Table 5 simultaneously. In this table, results of the nonparametric goodness-of-fit tests are reported, using the bootstrap procedure described in Subsection 2.3. As we are dealing with calculating accurate VaR estimates, our main aim is to fit the low tail of the observed distribution of each index returns. Panel 1 shows the p-values of these goodness-of-fit tests. It must be highlighted that, given a significance level equals to 5%, Student's *t* distribution, logistic distribution and Edgeworth-Sargan distribution are accepted to model the low tail of the observed distribution of each index in opposite to the normal distribution. Note that the three distribution functions considered are able to capture the thickness of the tails of high frequency data better than the normal distribution. FOOTSIE case is the only case in which the normality assumption can be accepted. This case is also the case in which the differences in QPS (Panel 5, Table 4) are not statistically different from zero and the model is well calibrated under normality assumption for a 1% significance level unlike other indices (Panel 6, Table 4). In Panel 2 of Table 5 we have reported the p-values of the goodness-of-fit tests when overall observed distribution is considered. Note that the p-values when the logistic distribution is tested are smaller than the 5% and 10% significance level for FOOTSIE and CAC40 respectively. That is, this specification is able to capture the fat tails of the observed distribution of FOOTSIE and CAC40 but not their overall shape.

Table 5. Results of goodness-of-fit tests

		Panel (1): p-value (low tail)		
Index	*Normal*	*Student t*	*Logistic*	*E-S*
SP500	0.000	0.280	0.310	0.695
NYSE	0.000	0.395	0.345	0.745
NIKKEI	0.065	0.650	0.810	0.980
FOOTSIE	0.165	0.295	0.505	0.550
CAC40	0.000	0.130	0.570	0.220
IGBM	0.035	0.465	0.140	0.515

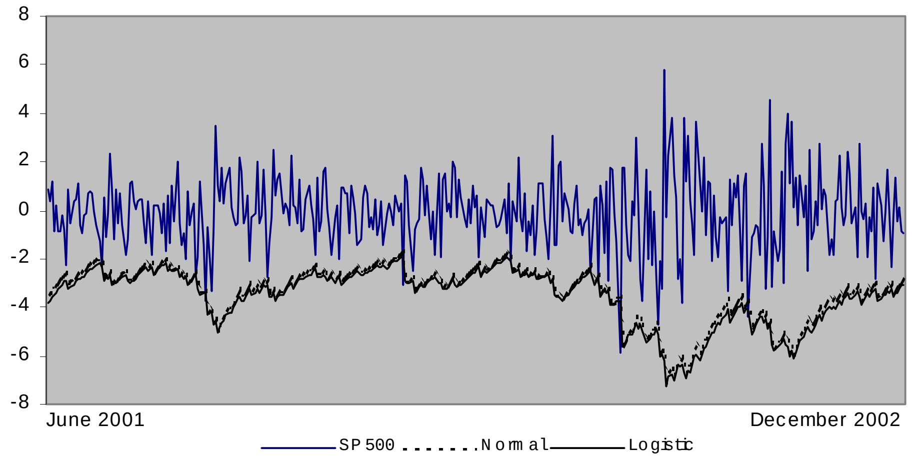
		Panel (2): p-value (overall distribution)		
Index	*Normal*	*Student t*	*Logistic*	*E-S*
SP500	0.000	0.370	0.375	0.255
NYSE	0.000	0.260	0.300	0.060
NIKKEI	0.005	0.265	0.320	0.130
FOOTSIE	0.095	0.440	0.035	0.465
CAC40	0.025	0.115	0.075	0.340
IGBM	0.000	0.440	0.310	0.770

Note: This table summarizes the results of the goodness-of-fit tests carried out using the Cramér-von Mises test statistic $\hat{W}_n^2$ to select the distribution functions that capture well or the low tail of the observed distribution of asset returns or the overall distribution.. Panel 1 reports the p-values obtained for testing the null hypothesis "$F(.,\theta)$ captures well the low tail". Panel 2 reports the p-values obtained for testing the null hypothesis "$F(.,\theta)$ captures well the overall observed distribution".

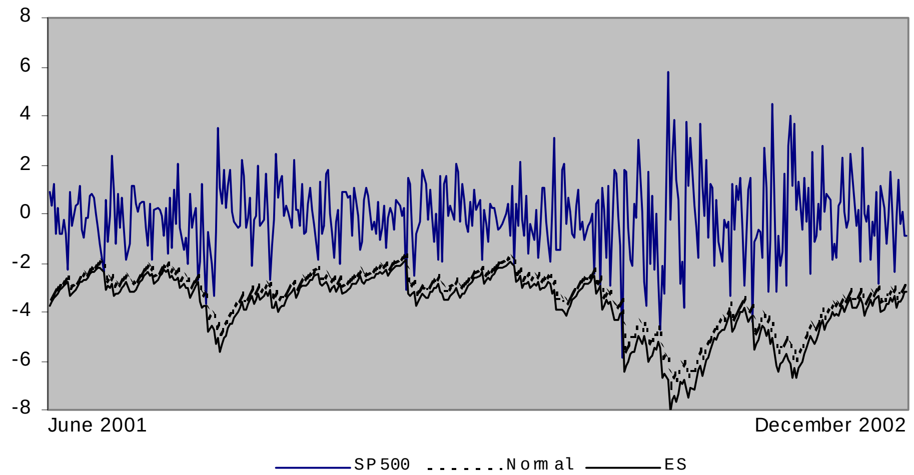
Overall, given Table 4 and Table 5 (Panel 1), it can be concluded that fitting the conditional behaviour of the low tail of the observed distribution of asset returns gives to assure to obtain accurate VaR estimates.

Table 6. Out-sample accuracy of VaR for SP500 (1/6/2001-31/12/2002)

Panel 1



Panel 2



Summary of evaluation method results

Evaluation Method	*Normal*	*Logistic*	*E-S*
$\hat{\alpha}$	0.0226	0.0175	0.0125
LR_{PF}	4.7112 (0.0411)	1.8880 (0.1694)	0.02442 (0.6211)
LR_{CC}	5.1276 (0.0770)	2.1386 (0.3432)	0.3714 (0.8305)
QPS	0.0455	0.0353	0.0256
Z_0	-	-1.4397 (0.1499)	-1.9934 (0.0462)
S-M&D(1993)	4.2482 (0.0000)	2.7518 (0.0059)	0.3119 (0.7551)

Note: For SP500, the accuracy of VaR estimates is assessed with respect to their one-step-ahead forecasts and 99% coverage level using the normal distribution, the logistic distribution (selected functional form to capture the overall observed distribution of SP500) and the Edgeworth-Sargan distribution, E-S (the selected distribution to fit the low tail of SP500). $\hat{\alpha}$: percentage of the exceedings, LR_{PF} : evaluation of VaR estimates based on the binomial distribution, LR_{CC} : evaluation of VaR interval forecasts, *QPS*: evaluation of VaR probability forecasts, Z_0: comparison of accuracy of probability forecasts; *S-M&D(1993)*: calibration test of probability forecasts. Between parentheses is the p-value of the test.

The empirical analysis concludes with a comparison of “out-sample” predictive power of VaR, using the data from June 2001 to December 2002. Table 6 summarizes the results obtained for SP500[3]. This table reports the results of using the evaluation methods when normality is assumed as well as when the conditional behaviour of this index is captured. Given the bootstrap p-values of the nonparametric goodness-of fit tests reported in Table 5, the logistic distribution is the selected functional form to capture the overall observed distribution of SP500 and the Edgeworth-Sargan distribution is the specification that better fits to the data of the low tail.

It can be observed that the worst results are those associated to the normality assumption: the biggest percentage of exceedings, rejection of the null hypothesis $\hat{\alpha} = \alpha$ using LR_{PF} and LR_{CC} at 5% significance level, the biggest QPS and reject that the model is well calibrated. Quite good results are obtained when the logistic distribution is assumed, not even it is concluded that the model is not well calibrated at 5% significance level. As it was expected, the best results are those obtained when the Edgeworth-Sargan distribution is assumed because fits well to the data of the low tail of the observed distribution of SP500. Assuming Edgeworth-Sargan distribution, the percentage of exceedings is very similar to 1% significance level, the null hypothesis $\hat{\alpha} = \alpha$ is accepted using both LR_{PF} and LR_{CC}, the value of QPS is the smallest and, it is accepted that the differences between QPS under normality and QPS assuming the Edgeworth-Sargan distribution are statistically significant and that the model is well calibrated. Panel 1 and Panel 2 in Table 6 show graphical comparisons of “out-sample” predictive power of VaR estimates, when the normal, logistic and Edgeworth-Sargan distributions are used. In Panel 1, VaR estimates under the normal and the logistic distributions is compared and it can be observed that VaR estimates using the logistic distribution are slightly better than when normality is considered. However, as Panel 2 reveals, the improvement in VaR estimates is higher when the Edgeworth-Sargan distribution is assumed. This graphical comparison is analytically supported by the evaluation method results summarized in Table 6.

[3] Similar results have been obtained for NYSE, NIKKEI, FOOTSIE, CAC40 and IGBM.

5. CONCLUSIONS

In this paper, it has been showed that risk measures, as VaR and CVaR, can be estimated accurately if the conditional behavior of the low tail of the observed distribution is fitted previously using the bootstrap procedure designed. Three alternative specifications to the normal distribution have been considered since they have fatter tails than the normal: Student's *t*, logistic and Edgeworth-Sargan distributions. VaR estimates have been calculated assuming each of these specifications and their accuracy has been compared using different evaluating methods: the percentage of exceedings, evaluation of VaR estimates based on the binomial distribution, evaluation of VaR interval forecasts, evaluation of VaR probability forecasts and calibration tests of probability forecasts. Bootstrap implementation was required to carry out the nonparametric goodness-of-fit tests based on Cramér-von Mises test statistic in order to compare the performance of the different proposed specifications for fitting to the data of the low tail of the observed distribution of the asset returns. It is worth noting that we have allowed for the postulated null distribution could depend on some unknown parameters (θ) in order to gain generality. We have provided empirical evidence using data from six stock-exchange indices: SP500, NYSE, NIKKEI, FOOTSIE, CAC40 and IGBM We have found that the specifications that fit well to the data of the low tail of the observed distribution are those distributions under which VaR estimates are accurate. In our analysis, both “in-sample” and “out-sample” predictive power comparisons of VaR estimates have been included. Overall, we consider relevant to include in the regulatory framework the necessity of carrying out goodness-of-fit tests for fitting the conditional behavior of the low tail of the observed distribution of asset returns previously to calculate VaR estimates. The accuracy of VaR is also important to calculate CVaR since it is the expected loss given that the loss is greater than or equal to the VaR.

REFERENCES

Acerbi, C., Nordio, C., Sirtori, C., 2001. Expected shortfall as a tool for financial risk management. Working paper. Available from <http://www.gloriamundi.org/>.

Acerbi, C., Tasche, D., 2002. On the coherence of expected shortfall. Journal of Banking and Finance 26, 1487-1503.

Bauer, C., 2000. Value at Risk using Hyperbolic Distributions. Journal of Economics and Business 52, 455-467.

Brier, G.W., 1950. Verification of forecasts expressed in terms of probability. Monthly Weather Review 75, 1-3.

Christoffersen, P.F., 1998. Evaluating interval forecasts. International Economic Review 39, 841-862.

Diebold, F.X., Lopez, J.A., 1996. Forecast evaluation and combination, in: Maddala, G.S., Rao, C.R. (Eds), Handbook of Statistics: Statistical Methods in Finance, Vol. 14. North-Holland, Amsterdam, pp. 241-268.

Diebold, F.X., Mariano, R., 1995. Comparing predictive accuracy. Journal of Bussiness and Economic Statistics 13, 253-264.

Gine, E., Zinn, J., 1990. Bootstrap general empirical measures. Annals of Probability 18, 851-869.

Hendricks, D., Hirtle, B., 1997. Bank capital requirements for market risk: The internal models approach. Federal Research Bank of New York, Economic Policy Review 3, 1-12.

Hull, J., White, A., 1998. Value at Risk when daily changes in market variables are not normally distributed. Journal of Derivatives 5, 9-19.

Koul, H., 1991. Weighted Empiricals and Linear Models. IMS Lecture Notes Monographs 21.

Kupiec, P.H., 1995. Techniques for verifying the accuracy of risk measurement models. Journal of Derivatives 2, 73-84.

Lilliefors, W.H., 1967. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. Journal of American Statistical Association 62, 399-402.

Lopez, J.A., 1999a. Regulatory evaluation of Value-at-Risk models. Journal of Risk 1, 37-64.

Lopez, J.A., 1999b. Methods for evaluating Value-at-Risk estimates. Federal Research Bank of New York, Economic Policy Review 2, 3-17.

Lopez, J.A., Saidenberg, M.R., 2000. Evaluating credit risk models. Journal of Banking and Finance 24, 151-167.

Lopez, J.A., 2001. Evaluating the Predictive Accuracy of volatility models. Journal of Forecasting 20, 87-109.

Lucas, A. 2000. A note on optimal estimation from a risk-management perspective under possibly misspecified tail behavior. Journal of Business and Economic Statistics 18, 31-39.

Mittnik, S., Paolella, M., Rachev, S.T., 2000. Diagnosis and treating the fat tails in financial returns data. Journal of Empirical Finance 7, 389-416.

Rockafellar, R.T., Uryasev, S., 2002. Conditional value-at-risk for general loss distributions. Journal of Banking and Finance 26, 1443-1471.

Sargan, J.D., 1976. Econometric estimators and the Edgeworth approximation. Econometrica 44, 421-448.

Seillier-Moiseiwitsch, F., Dawid, A.P., 1993. On testing the validity of sequential probability forecasts. Journal of the American Statistical Association 88, 355-359.

Shorack, G.R., Wellner, J.A., 1986. Empirical Processes with Applications to Statistics. Wiley, New York.

Szegö, G., 2002. Measures of Risk. Journal of Banking and Finance 26, 1253-1272.

Vlaar, P.J.G., 2000. Models for Dutch Bond Portfolios. Journal of Banking and Finance 24, 1131-1154.