A Goodness-of-Fit Test with Focus on Conditional Value at Risk

José Santiago Fajardo Barbachan*
Aquiles Rocha de Farias**
José Renato Haas Ornelas***

Abstract

To verify whether an empirical distribution has a specific theoretical distribution, several tests have been used like the Kolmogorov-Smirnov and the Kuiper tests. These tests try to analyze if all parts of the empirical distribution has a specific theoretical shape. But, in a Risk Management framework, the focus of analysis should be on the tails of the distributions, since we are interested on the extreme returns of financial assets. This paper proposes a new goodness-of-fit hypothesis test with focus on the tails of the distribution. The new test is based on the Conditional Value at Risk measure. Then we use Monte Carlo Simulations to assess the power of the new test with different sample sizes, and then compare with the Crnkovic and Drachman, Kolmogorov-Smirnov and the Kuiper tests. Results showed that the new distance has a better performance than the other distances on small samples. We also performed hypothesis tests using financial data. We have tested the hypothesis that the empirical distribution has a Normal, Scaled Student-t, Generalized Hyperbolic, Normal Inverse Gaussian and Hyperbolic distributions, based on the new distance proposed on this paper.

***Keywords*:** conditional value at risk, goodness-of-fit, Monte Carlo simulation.

JEL Codes: C12; C16.

Resumo

Para verificar quando uma distribuição empírica tem uma determinada distribuição teórica, vários testes têm sido usados, como os testes de Kolmogorov-Smirnov e Kuiper. Estes testes tentam analisar se todas as partes da distribuição empírica têm uma determinada forma teórica. Porém, no contexto de Administração do Risco, o foco da análise deveria ser nas caudas das distribuições, já que estamos interessados nos retornos extremos dos ativos financeiros. O presente artigo propõe um novo teste de ajuste com foco nas caudas da distribuição. O novo teste é baseado na medida de Valor em Risco Condicionado. Logo, usamos simulação de Monte Carlo para analisar o poder do novo teste com diferentes tamanhos de amostra, e depois os comparamos com os testes de Crnkovic and Drachman, Kolmogorov-Smirnov e Kuiper. Os resultados mostram que a nova distância possui um melhor desempenho do que as outras distâncias para pequenas amostras. Também analisamos

Submitted in December 2007. Accepted in July 2008. Article was double blind referred and evaluated by the editors. Supervising editor: Marcelo Medeiros. J. Fajardo thanks CNPq for financial support. The views expressed in this paper are those of the authors and do not reflect those of the Banco Central do Brasil or its members.

*Ibmec Business School Rio de Janeiro. E-mail: pepe@ibmecrj.br
**Banco Central do Brasil. E-mail: aquiles.farias@bcb.gov.br
***Banco Central do Brasil. E-mail: jrenato.ornelas@bcb.gov.br

testes de hipóteses usando dados financeiros. Testamos as hipóteses de que a distribuição empírica seja uma Normal, t-Student, Hiperbólica Generalizada, Normal Inversa Gaussiana e Hiperbólica, usando a nova distância proposta neste artigo.

***Palavras-chave*:** valor em risco condicionado; ajuste; simulação de Monte Carlo.

1. Introduction

Normality of returns is an assumption widely used in risk models, like the Riskmetrics™(1995, 1996), considered a benchmark model on measuring market risk. Nevertheless, the Normality assumption was later refuted by several empirical researches, starting with Mandelbrot (1963) and Fama (1965). The actual distribution of financial assets has some differences from the normal distribution, as fat tails (positive excess of kurtosis) and asymmetry (see Rydberg (1997) and Hull and White (1998)). Then researches tried to find which other distributions could replace the Normal.

In the 1960's, Mandelbrot (1963) and Fama (1965) were the pioneers in proposing alternative distributions to model financial data. They proposed that the Stable Paretian distribution. In the 1970's, Praetz (1972) and Blattberg and Gonedes (1974), proposed the scaled Student-t distribution, and in the 1980's, Kon (1984) suggested the mixture of Normals as an alternative to model financial assets. In the 1990's several papers analyzed the use of the Generalized Hyperbolic distribution to model financial assets, starting with Eberlein and Keller (1995).

When choosing among different distributions to model financial assets, a common problem that arises is to verify whether an empirical distribution has a specific distribution or not. Berkowitz (2002) suggests the distributional test as one of the steps to assess a risk model. Several tests have been used, for example: Kolmogorov-Smirnov, Likelihood Ratio and Kuiper. These tests try to analyze if all parts of the empirical distribution has a Normal shape. But, in a Risk Management framework, the focus of analysis is on the tails of the distributions, since we are interested on the extreme returns of financial assets. So, in a Risk Management framework, the goal of a goodness-of-fit test should be whether the tails of the theoretical distribution are a good approximation to the tails of the empirical distribution.

In this way usual goodness-of-fit tests would assess the goodness-of-fit of a theoretical distribution based on mismatch of all distribution, and not only the tails. Therefore, the main goal of this paper is to propose a new goodness-of-fit hypothesis test with focus on the tails of the distribution, i.e., a tail-goodness-of-fit test with focus on Risk Management. First, we propose a statistical distance based on the Conditional Value at Risk (CVaR) of two distributions. Then we use Monte Carlo Simulations to assess the power of the new test and also to perform hypothesis tests using financial data. We test the hypothesis that the empirical distribution has a Normal, Scaled Student-t, Generalized Hyperbolic (GH), Normal Inverse Gaussian (NIG) and Hyperbolic distributions, based on the new distance proposed on this paper.

This paper is organized as follows: Section 2 gives a brief overview of market risk models. Section 3 reviews some goodness-of-fit tests. In Section 4 we introduce the new goodness-of-fit test proposed by this paper. Section 5 addresses the power of the test and in Section 6 we have an empirical application of the new test. Section 7 concludes the paper with the final remarks and suggestions for further research.

2. Market Risk Models

2.1 Market risk measures

In recent years, risk management became popular among researchers, market practitioners and regulators. The Value at Risk (VaR) emerged as one of the benchmark measure for market risk.

According to Basak and Shapiro (2001), "*evidence abounds that in practice VaR estimates not only serve as summary statistic for decision makers, but are also used as a tool to manage and control risk – where economic agents struggle to maintain the VaR of their market exposure at a prespecified level*".

The VaR allows the market risk to be expressed in one number: the loss one expected to suffer with a certain confidence level to a fixed holding period:

$$P[R < -VaR(\alpha)] = 1 - \alpha$$

where R is the random variable of the asset's returns, and α is the confidence level with which the VaR is being calculated.

The distribution of R can be an empirical non-continuous distribution, or a theoretical specified continuous distribution. The Value at Risk of an empirical distribution can also be viewed as the α quantile of the distribution.

Although the intense use of VaR, researchers have criticized this risk measure. One question not addressed by the concept of VaR is what is the magnitude of the loss when the VaR limit is exceeded. Another issue on VaR pointed out by the article of Artzner et al. (1999) is that it is not a "coherent" measure of risk. They say that a risk measure is a coherent measure of risk if it satisfies the following four properties:

- Translation invariance: for a constant k, $\rho(x + k) = \rho(x) + k$
- Subadditivity: for all X and Y, $\rho(X + Y) \leq \rho(X) + \rho(Y)$
- Positive homogeneity: for a positive constant k, $\rho(kx) = k\rho(x)$
- Monotonicity: For all $X \leq Y$ for each outcome, then $\rho(X) \leq \rho(Y)$

The VaR is not considered a coherent measure of risk, because it fails to hold the subadditivity property, i.e., the VaR of a two assets portfolio can be greater than the sum of the two individual VaR's.

Also, Prause (1999) argues that to avoid bankruptcy one should forecast the distribution of the maximum expected loss. From this point of view regulators should use other risk measures than VaR. A better incorporation of extreme events especially in view of nonlinear portfolios is desirable.

Therefore, a new measure of risk has emerged in the literature, the Conditional VaR[1], that is the expected value of the loss, given that it exceeds a certain level. The Conditional VaR (CVaR) can be written as:

$$CVaR(R, \alpha) = -E[R|R \leq -VaR(\alpha)]$$

Considering that distribution of R is known, we have:

$$CVaR(R, \alpha) = \frac{\int_{-\infty}^{-VaR(\alpha)} x f_R(x) dx}{1-\alpha}$$

where f_R is the probability density function of R.

This measure, differently from VaR, satisfies the subadditivity property above mentioned, and addresses a question not answered by VaR: “How bad is bad?”. The VaR informs only if the loss is above a certain level, but no information about the magnitude of the loss is given. So, the CVaR is used to answer this question.

2.2 The parametric approach

Several approaches can be used to estimate the market risk measures. The choice will depend on the kind of portfolio, computational resources available and time constraints. The main three approaches are: parametric, historical simulation and Monte-Carlo simulation. This paper focuses on the parametric approach.

The parametric approach assumes that the asset returns have a specific probability distribution, (for example, the Normal). The parameters of the distribution are estimated (for example, the volatility) and so the risk measure is calculated based on the estimated distribution. So the risk measure depends on the parameters used. The Normal distribution is the most used with this approach. The Risk-Metrics? model is the most popular market risk model, and used this approach, assuming that asset returns follow a Normal distribution, with mean equal to zero, and volatility estimated by the EWMA (Exponential Weighted Moving Average) method, using historical data. Various other methods to estimate the volatility have been tested to replace the EWMA, usually methods of the GARCH family. This approach is usually called Conditional Normal methods, since we use a Normal distribution with the variance being calculated by conditional methods.

As the Normal distribution has thinner tails than the empirical distributions, other distributions different from the Normal have also been tested, such as student-t, mixture of Normals, Hyperbolic, etc.

Hull and White (1998) used a mixture of two Normal distributions to assess the Value at Risk of 12 exchange rates from 1988 to 1997. The first half of data was

[1]This measure has been used with many other names such as Expected Shortfall and Tail VaR.

used to estimate the parameters and the second to assess the model. An EWMA was used to estimate the volatility. Results when each currency has a separate estimation show the single currency model could be rejected for only four of the currencies, using the chi-squared test with 95% confidence. When the same parameters are used for all currencies, the model cannot be rejected with 95% confidence.

Brauer (2000) used a symmetric Hyperbolic distribution to perform VaR calculations, and used data from German stocks and international stock indexes (DAX, Dow Jones and Nikkei) from 1987 to 1997. His results showed that the model with Hyperbolic distribution outperformed Riskmetrics model.

Prause (1999) used the Generalized Hyperbolic distribution and its subclasses to fit German stocks, U.S. stock indexes and exchange rates. He used several statistics to assess the goodness-of-fit such as the Kolmogorov and Anderson-Darling distances, with the Normal being always the worst one. Also, Prause performed VaR calculations with single assets and multi-assets portfolios.

Also Generalized Pareto distributions are good candidates to fit extreme event data, then we can use these distributions to calculate VaR, for more details see Embrechts et al. (1997).

The main advantage of the parametric approach is the speed of calculation. Also, when using some conditional volatility estimation like EWMA or GARCH (conditional parametric methods), this method captures better the market stresses situations. On the other hand, this method has limitations if applied to portfolios with non-linear instruments, such as options.

3. Goodness-of-fit Tests

To test whether an empirical distribution has a specific distribution (or not), several tests have been used. The most common is the Kolmogorov-Smirnov. The Kolmogorov distance (see, for example, Massey (1951)) is defined as the greatest distance between empirical distribution and theoretical distribution, for all possible values:

$$D_{Kol} = \max_{x \in \mathbb{R}} |F_{Emp}(x) - F_{Theo}(x)|$$

where F_{Emp} is the empirical cumulative function and F_{Theo} is the continuous and completely specified theoretical cumulative function. F_{Emp} can be defined by:

$$F_{Emp}(x) = (number of X_i' s \leq x)/n$$

where $X_i' s$ are the sample's elements and n is the sample number of elements.

The Kuiper (1962) distance is similar to the Kolmogorov distance, but considers both directions of the discrepancy adding the greatest distances upwards and downwards:

$$D_{Kui} = \max_{x\in\mathbb{R}}\{F_{Emp}(x) - F_{Theo}(x)\} + \max_{x\in\mathbb{R}}\{F_{Theo}(x) - F_{Emp}(x)\}$$

In particular, these tests can be used to analyze if all parts of the empirical distribution has a Normal shape. But, in a Risk Management framework, the analysis focus is on the tails of the distributions, since we are interested on the extreme returns of financial assets.

One approach to give emphasis on the tails of a distribution is to use the Anderson and Darling (AD) distance, proposed in a 1952's paper. They propose a distance that would be viewed as Kolmogorov distance with weight. Weighting can be defined giving special importance to tails, and so being especially relevant to risk measures. The formula of this distance with tail emphasis is:

$$D_{AD} = \max_{x\in\mathbb{R}} \frac{|F_{Emp}(x) - F_{Theo}(x)|}{\sqrt{F_{Theo}(x)[1 - F_{Theo}(x)]}}$$

Prause (1999) uses the AD distance to assess which theoretical distribution fits better the data of German Stocks. The distributions assessed were Normal, Generalized Hyperbolic, Hyperbolic and Normal Inverse Gaussian. Nevertheless, he did not perform a hypothesis test, he just compared the AD distance of the distributions, to find out which one is the best. According to Prause, the Normal distribution is the worst one to his set of data.

There are other kinds of distances and tests that give emphasis on the tail, like the ones proposed in Crnkovic and Drachman (1996), hereafter CD, and Fajardo et al. (2005), hereafter FOF. Both use distances similar to the Kuiper, but with greater weights on the tails. Crnkovic and Drachman have proposed the following formula:

$$\begin{aligned} D_{CD} = \max_{x\in\mathbb{R}} -\frac{1}{2}\left\{F_{Emp}(x) - F_{Theo}(x)\right\}\left(\log\left(F_{Theo}\left(x\right)\left(1 - F_{Theo}\left(x\right)\right)\right)\right) \\ + \max_{x\in\mathbb{R}} -\frac{1}{2}\{F_{Theo}(x) - F_{Emp}(x)\}\left(\log\left(F_{Theo}\left(x\right)\left(1 - F_{Theo}\left(x\right)\right)\right)\right) \end{aligned}$$

4. Tail-Goodness-of-fit Test considering CVaR

In this paper, we propose a new distance to test the goodness-of-fit of a theoretical distribution to an empirical distribution.

The distance proposed here gives emphasis on the tails of the distribution (like the AD, CD and FOF tests) and so is more adequate to risk measures than the Kolmogorov distance. Our distance gives special emphasis on a specific risk measure – the Conditional Value at Risk (CVaR), and it is the absolute difference between the empirical and the theoretical CVaR for a given significance level. Intuitively, we are calculating the error of the theoretical distribution in estimating the CVaR of the empirical distribution. Mathematically:

$$\begin{aligned} D_{CV}(X,\alpha) &= |CVaR_{Theo}(X,\alpha) - CVaR_{Emp}(X,\alpha)| \\ &= \frac{1}{1-\alpha}\left|\left(\int_{-\infty}^{VaR_{Emp}} x\, f_{Emp}(x)\,dx\right)\right. \\ &- \left.\left(\int_{-\infty}^{VaR_{Theo}} x\, f_{Theo}(x)\,dx\right)\right| \end{aligned}$$

where:
X is the random variable representing the returns of the asset;
VaR_{Emp} is the Value at Risk calculated using the Empirical distribution with a confidence level α;
VaR_{Theo} is the Value at Risk calculated using the Theoretical with a confidence level α.;
f_{emp} is the probability density function of the empirical distribution and
f_{theo} is the probability density function of the theoretical distribution.

Differently from the Kolmogorov, AD, CD, FOF and other distances analyzed on this paper, the distance proposed is not based on a Maximum operator over the cumulative probability distribution, but it is a sum over the density probability functions. This has the purpose to fit in a better way the CVaR measure, which is also a sum.

Note that the distance has a parameter α besides the two distributions (empirical and theoretical). This is the level of confidence of the VaR beyond which the distance works. So, if one is interested on both tails of the distribution,[2] it is necessary to define a bi-caudal distance BCV (Bi-caudal Conditional Value at Risk):

$$\begin{aligned} D_{BCV}(X,\alpha) &= |CVaR_{Theo}(X,\alpha) - CVaR_{Emp}(X,\alpha)| \\ &+ |CVaR_{Theo}(-X,\alpha) - CVaR_{Emp}(-X,\alpha)| \end{aligned}$$

Based on this distance, we can perform a hypothesis test with the null hypothesis that the empirical distribution is equal to the theoretical distribution. As our distance focus on the tails of the distribution, we can say this is a “tail-goodness-of-fit” test.

5. The Power of the Test

To evaluate the power of the proposed test, the following procedure is used: a standard Normal distribution is taken as the theoretical distribution, and several other distributions are used to generate a large number of samples, i.e., they

[2]This is important when short positions on the asset are frequent.

are considered as the "true" empirical distribution (TED). For each sample generated, the distance between the standard Normal (the theoretical distribution) and the sample (the empirical distribution) is calculated and compared with the critical values of the Standard Normal, in order to perform the test with the null hypothesis being that both distributions are equal. So for each sample we have a result "reject" or "don't reject". Note that, as the two distributions are actually different by construction, the desirable result is to reject the null hypothesis. Therefore, the higher the percentage of rejection, the more powerful is the test. The percentage of "don't reject" may be viewed as the percentage of type II error of the test, and the lower this number, the better the test. This approach has been used to assess statistical tests, including backtests of VaR models, for example see Lopez (1997) and Kerkhof and Melenberg (2004).

As this paper is concentrated on distributional tests, our approach is slightly different, we aim to assess pairs of different distributions, instead of pairs of the distribution with the same probability function, but with different parameters. Kerkhof and Melenberg (2004), hereafter K& M, compare a Standard Normal with a student-t and two Normal Inverse Gaussian (NIG) distributions, one symmetric and other with high asymmetry. The idea behind is that real world financial data possess two characteristics: fat tails and negative asymmetry (see for example Rydberg (1997)), but in general risk models use a Normal distribution to model data. The three distributions used by K&M have fat tails, and one also negative asymmetry.

We use three "true" empirical distributions very similar with those of K&M, and compare with a Standard Normal (the one we choose as theoretical). The three TED used are:

- A Scaled-t distribution, with scale parameter equal to one, location parameter equal to zero and 5 degrees of freedom. This is a symmetric distribution, with expected value equal to zero and standard deviation equal to one. The only difference to the Standard Normal is an excess of kurtosis. K&M use a Student-t with 5 degrees of freedom, i.e., symmetric and centered, but with a variance larger than 1.

- A symmetric NIG. This is exactly the same symmetric NIG used by K&M, and has a moderate excess kurtosis.

- A negative skewed NIG. This the same used by K&M, except for the location parameter that we adjusted in order to obtain an expected value equal to zero. Therefore, this distribution has the same expected value and variance of the Standard Normal, and a large excess kurtosis.

Table 1 summarizes the characteristics of the distributions used, and table 2 shows the parameters. Note that the Scaled-t is the closest to the Standard Normal, and the Asymmetric NIG the most different.

Table 1
Distribution characteristics "true" empirical distribution

Distribution	Scaled-t	Symmetric NIG	Asymmetric NIG
Expected Value	0	0	0
Standard Deviation	1	1	1
Symmetry	yes	yes	Negative Skewed
Kurtosis	small excess	Moderate Excess	Large Excess

Table 2
Parameters

Distribution	Scaled-t	Symmetric NIG	Asymmetric NIG
α	–	1	1.031
β	–	0	-0.250
δ	–	1	0.941
μ	0	0	0.235
σ	1	–	–
DF	5	–	–

We used a Monte Carlo (MC) simulation approach with stratified sampling. For comparison purposes, we estimated the power of three other tests: Kolmogorov-Smirnov, Kuiper and CD. The BCV distance used on this simulation was calculated with an alfa equal to 95%. The procedure to estimate the power of the tests is the following:

1. Using 10,000 MC runs, we calculate the Critical Values for a Standard Normal, considering the BCV, CD, Kolmogorov and Kuiper distances, for sample sizes of 50, 125, 250, 500 and 1000.

2. For each sample size and TED, generate 10,000 MC runs.

3. For each MC run of the previous step, calculate the distance between the sample generated by the MC run and a Standard Normal.

4. For each distance of the previous step, calculate the p-value using the critical values of the first step.

5. Comparing the p-value with the standard significance level of 5%, we get the test result: reject or don't reject. Calculate the percentage of type II error dividing the number of "don't reject" by the number of MC runs, i.e., 10,000.

The results for the BCV, CD, Kuiper and Kolmogorov distances are shown on table 3, in terms of type II error percentages. We see that for small samples, the BCV distance is in general better than the CD. For large samples the CD performs better. The Kolmogorov distance has the worst performance by far. The Kuiper distance has a good relative performance for samples over 500 observations, but a bad performance for samples under 250.

Table 3
Power of the tests in terms of type II error

	Type II Errors											
Sample	Scaled-t				Symmetric NIG				Asymmetric NIG			
Size	Kuiper	Kolmog	BCV-95	CD	Kuiper	Kolmog	BCV-95	CD	Kuiper	Kolmog	BCV-95	CD
50	100.00%	100.00%	78.43%	90.49%	100.00%	100.00%	75.88%	88.74%	100.00%	100.00%	51.30%	74.65%
125	100.00%	100.00%	72.40%	80.98%	100.00%	100.00%	66.81%	72.86%	99.90%	100.00%	26.29%	42.65%
250	99.84%	100.00%	49.20%	54.68%	77.21%	100.00%	12.06%	23.94%	3.99%	99.03%	2.11%	1.33%
500	0.06%	99.99%	4.73%	5.32%	0.00%	99.00%	0.16%	0.01%	0.00%	0.00%	0.07%	0.00%
1000	0.00%	46.49%	0.61%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Generally speaking, we can say that the BCV has the best performance on small samples. This is especially important when time period available is small or when data has low frequency (monthly or yearly). However, for sample sizes under 125 observations, none of the tests can adequatedely detect a mismatch, except, perhaps, for the case of BCV with Asymmetric NIG, which is very different from the Normal. For the sample size of 250 the two tail-focused distances have a reasonable performance, while for sample sizes over 250 only the Kolmogorov still performs poorly.

6. Empirical Application

In this section we exemplify the tail-goodness-of-fit test, proposed on section 4. We use daily data from three major exchange rates, two fixed income indices and two stock indices. The three exchange rates used are Japanese Yen (JPY) per U.S. Dollar, Swiss Franc (CHF) per U.S. Dollar and U.S. Dollar per Great Britain Pound (GBP). The fixed income indices are the U.S. Treasury 5 to 10 years (UST) and U.S. Corporate High Yield (HY) provided by Lehman Brothers. The stock indices are the Nikkei 225 (NIKKEI), which is a major Japanese index, and the FTSE Eurotop 100 which represents the performance of the 100 most highly capitalized blue chip companies in Europe. Exchange rates time series cover the period from January 1st 1987 to August 29th 2002, and Fixed Income and Equity indices cover the period from February 1st 2001 to June 6th 2008. We fit data from these series into five theoretical distributions: Normal, Scaled-t, GH, NIG and Hyperbolic. The estimation method used is the maximum log-likelihood. After estimating the parameters, the next step to perform the test is to calculate the critical values and the p-values for each time series, considering the specific sample size. The critical values were obtained after 10.000 Monte Carlo (MC) Simulations runs for the sample sizes. For each MC run, we do the following two steps:

1. Generate a sample with the size of the time series, using each distribution and parameters estimated;

2. Calculate the Bi-caudal CvaR (BCV) distance with alfa equal to 95% between the sample generated on step 1, with the theoretical distribution;

At the end of the 10.000 runs, we will have 10.000 BCV distances that will be ordered. So the critical value for the significance level S will be the percentile $(1 - S)$ of the ordered BCV distance sequence.

In order to obtain the p-value, we have just to find the significance level corresponding to the BCV distance calculated between the empirical and theoretical distribution. Then we can compare the p-value to a certain significance level (we choose 1%), to assess the null hypothesis that the empirical distribution is equal to the theoretical distribution. These results are shown on table 4:

Table 4
BCV Hypothesis Test Results

	UST distance	HY distance	EUROTOP distance	NIKKEI distance	JPY distance	GBP distance	CHF distance
Normal	8,45E-02	1,42E-01	6,22E-03	6,33E-03	3,08E-03	3,16E-03	2,73E-03
	[0,08%]	[0,00%]	[0,00%]	[0,00%]	[0,00%]	[0,00%]	[0,00%]
Scaled-t	1,01E-01	6,44E-02	1,85E-03	5,56E-04	2,53E-03	5,82E-04	1,10E-03
	[87,10%]	[99,92%]	[37,56%]	[91,38%]	[0,97%]	[69,97%]	[23,07%]
NIG	1,35E-02	1,70E-02	6,96E-04	7,41E-04	5,37E-04	3,32E-04	2,97E-04
	[99,99%]	[99,87%]	[79,32%]	[81,57%]	[68,14%]	[84,21%]	[87,74%]
Hyperbolic	1,16E-02	2,04E-01	1,86E-03	3,14E-04	2,90E-03	5,72E-04	4,77E-04
	[99,99%]	[0,00%]	[9,94%]	[95,33%]	[0,00%]	[50,30%]	[69,07%]
GH	1,01E-02	1,02E-02	6,74E-04	3,79E-04	3,16E-04	4,46E-04	5,32E-04
	[99,99%]	[99,91%]	[79,95%]	[94,20%]	[89,00%]	[71,69%]	[62,17%]

As expected, results from table 4 had rejected Normal distribution for all time series tested, corroborating previous empirical studies. For the CHF and GBP, all the others distribution have been not rejected on the BCV hypothesis test, but for the JPY, only the NIG and GH were not rejected with 1% confidence level, although the scaled-t is very near this confidence level. The best distribution according to the minimum BCV distance criteria is the NIG for GBP and CHF, and the GH for the JPY.

For the Fixed Income indices, besides the Normal, only the Hyperbolic distribution for the High Yield Time Series was rejected at 1%. The GH distribution has the minimum BCV distance for both Fixed Income indices. For the Stock indices, the Normal distribution was the only one rejected at 1%. The Hyperbolic was rejected at 10% for the FTSE Eurotop. However, for the Nikkei, the Hyperbolic was the best distribution using the BCV distance criteria. The GH was the best distribution for the FTSE Eurotop. To provide a comparison to classical tests, the Kolmogorov-Smirnov and Kuiper tests were also performed on the same data as can be seen on tables 5 and 6:

Table 5
Kolmogorov-Smirnov hypothesis test results

	UST distance	HY distance	EUROTOP distance	NIKKEI distance	JPY distance	GBP distance	CHF distance
Normal	3,72E-02	1,36E-01	5,97E-02	5,60E-02	5,60E-02	5,60E-02	4,40E-02
	[1,19%]	[0,00%]	[0,00%]	[0,00%]	[0,00%]	[0,00%]	[0,00%]
Scaled-t	1,43E-02	2,54E-02	1,09E-02	1,17E-02	1,28E-02	1,25E-02	1,19E-02
	[84,69%]	[16,60%]	[57,23%]	[40,29%]	[51,44%]	[54,76%]	[60,96%]
NIG	8,15E-03	1,12E-02	8,26E-03	7,63E-03	6,92E-03	9,13E-03	9,81E-03
	[99,95%]	[95,97%]	[87,17%]	[88,15%]	[98,95%]	[88,38%]	[82,48%]
Hyperbolic	8,69E-03	4,80E-02	1,04E-02	5,75E-03	1,15E-02	9,11E-03	9,06E-03
	[99,93%]	[0,04%]	[63,34%]	[98,86%]	[64,83%]	[88,60%]	[88,97%]
GH	8,75E-03	9,80E-03	8,27E-03	6,10E-03	8,58E-03	9,13E-03	8,85E-03
	[99,94%]	[98,43%]	[86,76%]	[98,13%]	[92,37%]	[88,45%]	[90,49%]

Table 6
Kuiper Hypothesis Test Results

	UST distance	HY distance	EUROTOP distance	NIKKEI distance	JPY distance	GBP distance	CHF distance
Normal	6,61E-02	2,56E-01	1,19E-01	1,12E-01	1,04E-01	1,11E-01	7,80E-02
	[0,00%]	[0,00%]	[0,00%]	[0,00%]	[0,00%]	[0,00%]	[0,00%]
Scaled-t	2,28E-02	3,94E-02	2,13E-02	2,32E-02	2,41E-02	2,12E-02	2,31E-02
	[90,94%]	[8,04%]	[17,09%]	[5,04%]	[14,63%]	[31,96%]	[19,59%]
NIG	1,59E-02	2,15E-02	1,44E-02	1,40E-02	1,32E-02	1,65E-02	1,76E-02
	[99,98%]	[86,53%]	[77,11%]	[71,85%]	[95,75%]	[74,45%]	[64,09%]
Hyperbolic	1,70E-02	8,51E-02	1,84E-02	1,15E-02	2,05E-02	1,76E-02	1,59E-02
	[99,74%]	[0,00%]	[36,82%]	[93,81%]	[37,15%]	[64,83%]	[79,73%]
GH	1,63E-02	1,80E-02	1,48E-02	1,06E-02	1,70E-02	1,58E-02	1,62E-02
	[99,94%]	[96,97%]	[73,21%]	[97,13%]	[70,31%]	[81,01%]	[77,36%]

The classical tests also rejected Normal distribution as was expected, with the all other distributions being not rejected. Note that for the CHF and GBP, test results are the same of the BCV, but for the JPY, the Scaled-t and the Hyperbolic that were rejected on the BCV test, now are being not rejected. So we can say that, for overall applications, we can use Scaled-t and Hyperbolic distributions for the JPY, but for Risk Management applications, they are not adequate. Results about the best distribution for these criteria are mixed: for the JPY currency, the NIG was the best distribution on both Kolmogorov and Kuiper distances. For the GBP and CHF, Hyperbolic and GH were the best distribution depending on the criteria (Kolmogorov or Kuiper).

7. Conclusions

In this paper we proposed a way to test the goodness-of-fit of theoretical distributions to empirical data with focus on the CVaR risk measure. The distance we propose showed a better performance on We used a Monte-Carlo Simulations with three types of distributions to assess the power of the new test and compare with tests commonly used in the literature. Results showed a better performance of the new test in general. A sample with three exchange rates, two stock indexes and two Bond indexes was used used to exemplify the use of the test proposed with real world data. The test proposed on this article can be easily applied to other kinds of distributions and assets, including portfolios, since it needs only the series of returns and the expression for the distribution's density. So, as suggestion for future research, other kinds of distributions maybe be also tested, together with other classes of assets. It is worth to mention that the test proposed here assesses the unconditional distribution of returns, but most of the risk management approaches use conditional distributions approaches. Anyway, as suggested by Berkowitz (2002, figure 1), one of the steps to check the validity of a risk model is to test the distribution used, and this would be done by the CVaR test proposed on this paper. Also the CVaR-focuses distance proposed can be used to estimate the parameters of distributions, through minimization of this distance. For example, Prause (1999) uses an estimation method that minimizes the Anderson-Darling distance. Then, with an estimation focused on the tails of the distribution, the

CVaR measure is expected to be more reliable. After the distance minimization estimation, a conditional volatility model may be used to re-scale the distribution to get the volatility forecasted by the model. Backtests would then assess the validity of this approach.

References

Anderson, T. W. & Darling, D. A. (1952). Asymptotic theory of certain goodness of fit criteria. *Annals of Mathematical Statistics*, 23(2):193–212.

Artzner, P., Delbaen, F., Eber, J., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9:203–228.

Barndorff-Nielsen, O. & Blæsild, P. (1981). Hyperbolic distributions and ramifications: Contributions to theory and application. *Statistical Distributions in Scientific Work*, 4:19–44.

Basak, S. & Shapiro, A. (2001). Value-at-risk-based risk management: Optimal policies and asset prices. *The Review of Financial Studies*, 14(2):371–405.

Berkowitz, J. (2002). Testing the distributional assumptions embedded in risk models. *Risk*. Available at http://www.uh.edu/ jberkowi/disttest.pdf (May 2002).

Blattberg, R. & Gonedes, N. (1974). A comparison of the stable and student distributions as statistical models for stock prices. *Journal of Business*, 47:244–280.

Brauer, C. (2000). Value at risk using hyperbolic distributions. *Journal of Economics and Business*, 52:455–467.

Crnkovic, C. & Drachman, J. (1996). Quality control. *Risk*, 9(9):138–143.

Eberlein, E. & Keller, U. (1995). Hyperbolic distributions in finance. *Bernoulli*, 1:281–299.

Embrechts, P., Klueppelberg, C., & Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, Berlin.

Fajardo, J., Ornelas, J., & Farias, A. (2005). Analyzing the use of generalized hyperbolic distributions to value at risk calculations. *Brazilian Journal of Applied Economics*, 9(1):25–38.

Fama, E. F. (1965). The behavior of stock market prices. *Journal of Business*, 38:34–105.

Hull, J. & White, A. (1998). Value at risk when daily changes in market variables are not normally distributed. *Journal of Derivatives*, 5(3):9–19.

Kerkhof, J. & Melenberg, B. (2004). Backtesting for risk-based regulatory capital. *Journal of Banking & Finance*, 28:1845–1865.

Kon, S. J. (1984). Models of stock returns – a comparison. *The Journal of Finance*, 39(1):147–165.

Kuiper, N. H. (1962). Proceedings of the koninklijke nederlandse akademie van wetenchappen. 63:38–47. Ser. A.

Lopez, J. (1997). Regulatory evaluation of value-at-risk models. *Research Paper*, 9710. Federal Reserve Bank of New York.

Mandelbrot, B. B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36:394–419.

Massey, F. J. (1951). The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46:68–78.

Praetz, P. (1972). The distribution of share price changes. *Journal of Business*, 45:49–55.

Prause, K. (1999). The generalized hyperbolic model: Estimation, financial derivatives, and risk measures. PhD Dissertation, University of Freiburg.

RiskMetrics (1995). Technical document. J.P.Morgan / Reuters.

RiskMetrics (1996). Technical document. J.P.Morgan / Reuters.

Rydberg, T. H. (1997). Why financial data are interesting to statisticians. Centre for Analytical Finance of the university of Aarhus. Working Paper, n. 5.

Appendix A

A.1 Stable Paretian distributions

Characteristic function of the Stable Paretian distribution is the following:

$$\phi_x(t) = \exp\left\{-\gamma^\alpha |t|^\alpha \left(1 - i\beta sign(t) tg(\pi\alpha/2)\right) + it\delta\right\} \text{ for } \alpha \neq 1$$

and

$$\phi_x(t) = \exp\left\{-\gamma |t| \left(1 + 2i\beta sign(t) \ln(|t|)\right)/\pi + it\delta\right\} \text{ for} \alpha = 1$$

where:
δ is the location parameter,
γ is the scale/dispersion parameter,
β is the asymmetry parameter and α is the stability index.

A.2 Scaled Student t distribution

The density function of the scaled Student t distribution is the following:

$$f(x;\mu,\sigma,\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\nu/2)\sqrt{\pi(\nu-2)\sigma^2}}\left[1 + \frac{(x-\mu)^2}{(\nu-2)\sigma^2}\right]^{-(\nu+1)/2}$$

where:
υ is the degrees of freedom parameter,
μ is the location parameter and
σ the dispersion parameter.
when $\upsilon \to \infty$ the Student t converges to the Normal Distribution.

A.3 Mixture of Normals

Density function of the returns is a weighted average of several Normal densities:

$$f(x;\mu_1,\ldots,\mu_n,\sigma_1,\ldots,\sigma_n,\lambda_1,\ldots,\lambda_{n-1}) = \sum_{i=1}^{n} \lambda_i N(\mu_i,\sigma_i)$$

where:
μ_i and σ_i are respectively the mean and standard deviation of each Normal distribution, and
λ is the weight of each Normal. As the sum of the weights must be one, the last λ is completely defined by the others.

A.4 Generalized Hyperbolic Distribution

The density probability function of one-dimensional GH distribution is defined by the following equation:

$$\begin{aligned} GH(x;\lambda,\alpha,\beta,\delta,\mu) &= a(\lambda,\alpha,\beta,\delta)\left(\delta^2+(x-\mu)^2\right)^{(\lambda-1/2)/2}\times \\ &\quad K_{\lambda-1/2}\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)e^{(\beta(x-\mu))} \end{aligned}$$

where K_x is the modified Bessel function of third kind and

$$a(\lambda,\alpha,\beta,\delta)=\frac{\left(\alpha^2-\beta^2\right)^{\lambda/2}}{\sqrt{2\pi}\alpha^{\lambda-0.5}\delta^{\lambda}K_{\lambda}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}$$

The parameters are real numbers with the following restrictions (see Prause (1999)):

$$\begin{array}{lll} \delta\geq 0, & |\beta|<\alpha & \text{if}\lambda>0 \\ \delta>0, & |\beta|<\alpha & \text{if}\lambda=0 \\ \delta>0, & |\beta|\leq\alpha & \text{if}\lambda<0 \end{array}$$

The parameter δ is a scale factor, compared to the σ of a Normal distribution, and μ is a location parameter. Parameters α and β determine the distribution shape and λ defines the subclasses of GH and is directly related to tail fatness (Barndorff-Nielsen and Blæsild, 1981). The function $a(.)$ is introduced to guarantee that the cumulative distribution has values between zero and one.

The GH has several subclasses, among them the Hyperbolic and Normal Inverse Gaussian(NIG). Setting $\lambda=-1/2$, we get the NIG, and with $\lambda=1$, we get the Hyperbolic distribution. The Gaussian is a limiting distribution of GH, when $\delta \quad \infty$ and $\delta/\alpha\sigma^2$.