Goodness-of-fit Tests Focus on Value-at-Risk Estimation*

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

Abstract

A common statistical problem in finance is measuring the goodness-of-fit of a given distribution to real world data. This can be done using distances to measure how close an empirical distribution is from a theoretical distribution. The tails of the distribution should receive special importance if the focus is on Value-at-Risk (VaR) calculations. This paper analyzes the use of distances to test the goodness-of-fit of estimated distributions for VaR calculation purposes. The Crnkovic and Drachman (1996) distance and a new distance are used to perform goodness-of-fit tests. The critical values of the tests are obtained using Monte Carlo simulation, and goodness-of-fit tests are performed based on the distances. The power of the tests is assessed through Monte Carlo experiments, showing good results for sample sizes greater than 250. The US Dollar/Brazilian Real exchange rate and the Ibovespa index are used as examples of practical applications of how to test the hypothesis that an empirical distribution is equal to an estimated one. The estimated distributions considered are the Generalized Hyperbolic (GH), the NIG (Normal Inverse Gaussian) and Normal. The test results rejected the null hypothesis for the Normal distribution, but did not reject it for the Generalized Hyperbolic and NIG, both at a 1% significance level.

Keywords: Value-at-Risk, Distance, Goodness-of-fit Test.

JEL Codes: C12, C13, C15, C16.

*Submitted in August 2004. Revised in August 2006. The views expressed in this work are those of the authors and do not reflect those of the Banco Central do Brasil or its members. We thank comments from two anonymous referees and Antonio Francisco from Gepom-Bacen for helpful discussions. José Fajardo thanks CNPq and PRONEX E-26/171.193/2003 for financial support.

**Ibmec Business School. E-mail: pepe@ibmecrj.br
***Università Luigi Bocconi and Banco Central do Brasil.
****Universidade de Brasília and Banco Central do Brasil.

1. Introduction

A frequent problem in statistics and finance is measuring the goodness-of-fit of a theoretical[1] distribution to real world data. Knowing the distribution of empirical data, one can use the distribution properties to perform several useful calculations in finance, such as Value-at-Risk (VaR), Option Pricing, Credit Risk, etc.

Usual goodness-of-fit tests, such as the Kolmogorov-Smirnov and Kuiper tests, analyze whether all parts of the empirical distribution have a specific distribution. However, in a risk management environment, the focus of analysis is on the tails of the distributions, since risk managers are concerned about the extreme returns of financial assets. As pointed out by Berkowitz (2002), "*...risk managers are often exclusively interested in an accurate description of large losses or tail behavior. They do not want to reject a model that forecasts tail events well because of failure to match the small day-to-day moves that characterize the interior of the forecast distribution*". Therefore, in a risk management environment, the goal of a test should be whether the tails of the theoretical distribution are a good approximation to the tails of the empirical distribution, and then we would obtain better risk measures when using this theoretical distribution.

The main goal of this paper is to analyze the use of distances to measure how close an empirical distribution is from a theoretical distribution, giving special importance to the tails. These kinds of distances would be appropriate to measure the goodness-of-fit of distributions to estimate the risk of a portfolio, especially to calculate the VaR, the most used risk measure.

This paper analyzes two tail-focused distances: the CD (Crnkovic and Drachman, 1996) distance and a new one, proposed in this article. These distances are used as a criterion to perform goodness-of-fit tests to verify whether the theoretical distribution is equivalent to the empirical distribution. Tail-focused distances have been used to rank distributions in terms of goodness-of-fit to empirical data, but no hypothesis tests have been performed so far in the literature, except for the Crnkovic and Drachman (1996) paper.[2] Ranking the distributions only is not sufficient, since even the best ranked distribution may have a poor fitting to empirical data. If more than one distribution is not rejected by the test, we should choose the one that is best ranked according to the distances being considered. Therefore, the methodology proposed in this paper can be used to assess which distributions are adequate, and to choose the best one from a set of distributions.

[1]With parameters estimated based on empirical data. In this paper, we use the expression "theoretical distribution" as synonymous with "estimated distribution".

[2]Crnkovic and Drachman (1996) use a tail-focused distance to perform a backtest. This paper, however, focuses on a distributional test, instead of a backtest. We use past empirical data to choose the distribution that best fits past data.

To perform the tests using tail-focused distances, it is necessary to calculate the critical values, which is done in Section 4 of this paper, using Monte Carlo Simulation, to two kinds of distribution: Normal and Generalized Hyperbolic (GH). Section 5 assesses the power of the test to identify differences among different distributions.

In Section 6, data from the US Dollar/Brazilian Real (USD/BRL) exchange rate and from the São Paulo Stock Exchange Index (Ibovespa) are used to exemplify a goodness-of-fit test: first, the parameters of an unconditional GH (Generalized Hyperbolic) and NIG (Normal Inverse Gaussian) distributions[3] are estimated using maximum log-likelihood; second, the critical values for unconditional Normal, GH and NIG estimated are calculated for the sample sizes using Monte Carlo Simulation; third, the null hypothesis that the estimated distribution is equal to the empirical distribution using distance criteria is tested.

This paper is organized as follows: in Section 2, we motivate the use of VaR in risk management. In Section 3, we have a brief revision of goodness-of-fit tests and distances and the new distance is presented. In Section 4, the critical values are calculated. Section 5 analyzes the power of the proposed test. Section 6 provides a practical application. In the last sections, we have the conclusions and an appendix with the description of the GH distributions.

2. Value-at-Risk and Risk Management

A risk manager who is trying to measure the market risk needs to define many important issues, as for example, which is the most suitable measure of market risk? Which method should be used, a simulation or an analytical (also called parametric) method to compute this measure? Which is the adequate sample size from historical data to make the calculations?

The Basel Committee on Banking Supervision has suggested a single number that summarizes the total market risk in a portfolio of financial assets called Value-at-Risk (VaR). The VaR is defined as:

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

where R is the returns and α is the significance level at which the VaR is being calculated. It shows how often things can get bad in a given time horizon. The Basel Committee has also suggested the use of a 10-day time horizon and a 99% confidence level for measuring the bank's capital adequacy and requirement.

The VaR has been widely used by financial institutions and central bank regulators to measure risk exposures. A precise estimation of the VaR is therefore useful from many points of view: a downward bias VaR may lead to excessive risk to the institution, whereas an upward bias VaR may lead to excessive capital

[3] More details about GH and NIG distributions in the Appendix.

requirement by central banks. It is worth mentioning also that banks with inaccurate internal VaR models are penalized by central banks through an increase in capital requirements.

Then, once we have chosen VaR as the market risk measure, we turn to the other issues: analytical or simulation methods? Well, here we find a trade-off between simplicity, computational effort and forecasting (Dowd, 2002).

When analytical methods are used, a crucial decision is how to model each risk factor (for example, how to model a stock price or an exchange rate). One of the most important steps in a model is the distribution to be used. Many statistics based on distances have been developed to measure this goodness-of-fit of a distribution to the empirical data. In the next section we present the most important distances and statistics used to test this goodness-of-fit, and we introduce a new distance, which gives more weight to deviations occurring in the tails of the distribution.

3. Distances

To measure how close an empirical distribution is from a theoretical distribution, several distances have been proposed. Among them, we can cite three: Kolmogorov distance, Kuiper distance and Anderson-Darling distance.

The Kolmogorov distance (see, for example, Massey (1951)) is defined by the greatest distance between the empirical distribution and the theoretical distribution, for all possible values:

$$D_{Kol} = \max_{x \in \Re} |f_{Emp}(x) - f_{Theo}(x)| \qquad (1)$$

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

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

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

The Kuiper distance (see Kuiper (1962)) is similar to the Kolmogorov distance, but it considers the direction of the deviation, adding the greatest distances upwards and downwards:

$$D_{Kui} = \max_{x \in \Re} \{f_{Emp}(x) - f_{Theo}(x)\} + \max_{x \in \Re} \{f_{Theo}(x) - f_{Emp}(x)\} \qquad (2)$$

The Anderson and Darling (1952) paper proposes two distances, stated in equations 2.1 and 2.2 of that paper. Let us consider the second one in this general case:

$$D_{ADn} = \max_{x \in \Re} \sqrt{n} \; |f_{Emp}(x) - f_{Theo}(x)| \sqrt{\psi(x)} \quad (3)$$

where n is the sample size and $\psi(x)$ is a weight function.

But the AD distance most widely known in the literature is the one expressed in example 2 of the Anderson and Darling (1952) article, with the weight function $\psi(x)$ defined as follows:

$$\psi(x) = \frac{1}{\sqrt{f_{Theo}(x)\,(1 - f_{Theo}(x))}} \quad (4)$$

This function has the effect of weighting the tails heavily since this function is large near $f_{Theo} = 1$ and $f_{Theo} = 0$. A common simplification used is the suppression of $n^{1/2}$ from (3). It can be used if the sample sizes that are being compared are the same. So the AD distance considered in this paper will be the following:

$$D_{AD} = \max_{x \in \Re} \frac{|f_{Emp}(x) - f_{Teo}(x)|}{\sqrt{f_{Teo}(x)\,(1 - f_{Teo}(x))}} \quad (5)$$

The AD distance is especially interesting to perform VaR calculations, since it is more sensitive in the tails than in the middle range of the distribution. Prause (1999) uses the AD distance to assess which theoretical distribution fits better the data of German Stocks. The distributions assessed were Normal, GH, Hyperbolic and NIG. Nevertheless, a hypothesis test using AD distance was not performed by Prause.

Another distance that is appropriate for VaR calculations is the Crnkovic and Drachman (1996). It can be viewed as Kuiper distance with weights. The CD distance uses the worry about the direction of the deviation from Kuiper distance and incorporates a weight function to give special importance to the tails:

$$\begin{aligned} D_{CD} &= \max_{x \in \Re} -0.5\,(f_{Emp}(x) - f_{Theo}(x))\,(Log\,(f_{Theo}(x)\,(1 - f_{Theo}(x)))) \\ &+ \max_{x \in \Re} -0.5\,(f_{Teo}(x) - f_{Emp}(x))\,(Log\,(f_{Theo}(x)\,(1 - f_{Theo}(x)))) \quad (6) \end{aligned}$$

The distance proposed in this paper is similar to the CD distance, and is a combination of the Kuiper and AD distances. It uses the weight function of the AD distance and the worry about the direction of deviation as in the Kuiper distance. So, it captures the strengths of Kuiper and AD distances. The difference between the new distance and the CD distance is that the new distance gives even more emphasis to tails. Figure 1 plots the weight functions of the CD and of the new distance. The formula of the new distance is:

$$D_{New} = \max_{x \in \Re} \frac{f_{Emp}(x) - f_{Teo}(x)}{\sqrt{f_{Teo}(x)\,(1 - f_{Teo}(x))}} + \max_{x \in \Re} \frac{f_{Teo}(x) - f_{Emp}(x)}{\sqrt{f_{Teo}(x)\,(1 - f_{Teo}(x))}} \tag{7}$$

The expectation is that this new distance could be more appropriate for VaR calculations because it captures the tail discrepancies much better.

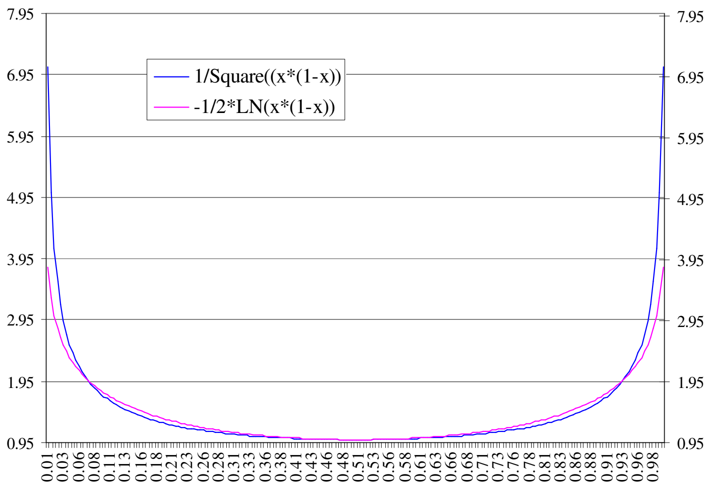

Figure 1

4. Critical Values

To perform the hypothesis tests, it is necessary to calculate critical values of the distances for some target significance levels and sample sizes. In this section we present examples of critical values for the new distance and for the CD distance, using two distributions: a Standard Normal and a Normal Inverse Gaussian (NIG).[4] We use 50, 100 and 200 for sample sizes. The critical values were obtained after 10,000 Monte Carlo (MC) Simulation runs. For each MC run, we do the following steps:

- Draw n independent random numbers U_i from a uniform distribution [0,1], where n is the sample size (50, 100 or 200);

[4] We are using a Symmetric Centered Normal Inverse Gaussian distribution with parameters alpha=1, beta=0, sigma=1, mi=0 and lambda=−0.5. The N.I.G. distribution is a special case of the Generalized Hyperbolic and it is closed under convolutions. See more details in the Appendix.

- Make $X_i = F^{-1}(U_i)$, where F is the theoretical cumulative distribution function (Normal or NIG);
- Calculate the distance (CD or the New one) between the sample generated in the previous steps and the theoretical distribution F.

At the end of the 10,000 runs, we will have 10,000 distances that will be ordered, so that we obtain a sequence Q of size 10,000. So, the critical value for the significance level S will be the $(1 - S)$ percentile of the ordered distance sequence Q.

The tables cover the significance levels of 1%, 5%, 10% and 20%, and the sample sizes of 50, 100 and 200. The null hypothesis (H_0) and the alternative hypothesis (H_1) are the following:

H_0: the empirical distribution is equal to the theoretical one (Normal or NIG);

H_1: the empirical distribution is different from the theoretical one (Normal or NIG).

The results are shown in Tables 1 and 2. The tables contain the critical values V's such that:

$$P[\text{Distance}>\boldsymbol{V}] = \alpha$$

Where α is the confidence level and D is the distance (the new one or the CD).

Table 1
Critical values – NIG

	New distance				**Distance CD**			
	Significance level				Significance level			
Sample size	1%	5%	10%	20%	1%	5%	10%	20%
50	1.7421	0.8954	0.7272	0.5768	0.2475	0.2073	0.1893	0.1691
100	1.3328	0.6835	0.5458	0.4386	0.1808	0.1534	0.1400	0.1258
250	0.9237	0.4465	0.3569	0.2933	0.1124	0.0984	0.0906	0.0825

Table 2
Critical values – Standard normal distribution

	New distance				**Distance CD**			
	Significance level				Significance level			
Sample size	1%	5%	10%	20%	1%	5%	10%	20%
50	1.4429	0.7592	0.5980	0.4817	0.1882	0.1658	0.1532	0.1396
100	1.2147	0.5990	0.4576	0.3684	0.1414	0.1223	0.1142	0.1043
250	0.9879	0.4380	0.3318	0.2617	0.0910	0.0813	0.0756	0.0694

Note that these critical values are purposes. For real financial market applications one would probably need to calculate the *minimum* significance level at which a hypothesis would be rejected (the so called p-value). And the sample sizes would be different from those of Tables 1 and 2. So, it would be necessary to perform MC simulation runs for that specific sample size, and also to calculate the minimum significance level at which the hypothesis is rejected.

5. Assessing 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 are considered the "true" empirical distribution (TED). For each sample generated, the distance between the standard Normal (the theoretical distribution) and the generated sample (the empirical distribution) is calculated and compared with the critical values of the Standard Normal for that distance, in order to evaluate the null hypothesis that both distributions are equal. So, for each sample, we have a binary result of the "reject" or "do not reject" type. Note that the two distributions are different by construction, so the desirable result is to reject the null hypothesis. Therefore, the higher the percentage of rejection, the more powerful the test. The percentage of "do not reject" may be viewed as the percentage of type II error[5] 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, see, for example, Lopez (1998) and Kerkhoff and Melenberg (2004). As this paper focuses on distributional tests, our approach is slightly different – our aim is to assess pairs of different distributions instead of pairs of distributions with the same probability function, but with different parameters.[6] Kerkhof and Melenberg (hereafter referred to as K&M) compare a Standard Normal distribution with a student-t and two Normal Inverse Gaussian (NIG) distributions – one symmetric and other with high asymmetry. The rationale behind this is that real world financial data possess two characteristics: fat tails and negative asymmetry (see, for example, Rydberg (1997)), but in general risk models, a Normal distribution is used to model data. So, by considering the Normal distribution as the theoretical distribution and empirical distributions with fat tails and asymmetry, we can assess whether the test is able to detect distributional differences that often occur in real world applications.

This paper will use three "true" empirical distributions that are very similar to those of K&M, and compare them with a Standard Normal distribution (the one we chose as theoretical). The three TED used are:

- A Scaled-t distribution (see the Appendix for details), 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 from the Standard Normal distribution is an excess kurtosis. K&M use a Student-t with 5 degrees of freedom, i.e., symmetric and centered, but with a variance larger

[5] Type II error is not to reject the null hypothesis when it is actually false.

[6] Lopez (1998) compares Normal and student-t distributions with different parameters to assess backtesting procedures. K&M and Lopez also use econometric models such as GARCH to model the volatility of the distribution. This paper analyzes only distributional discrepancies, rather than differences in the parameterization.

than 1. We decided to use a Scaled-t to use the same variance of he Standard Normal distribution;

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

- A negatively skewed NIG. This is the same distribution 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 distribution, and a large excess kurtosis.

Table 3 summarizes the characteristics of the distributions used, and Table 4 shows the parameters. Note that the Scaled-t is the closest to the Standard Normal distribution, and the Asymmetric NIG is the most different.

Table 3
Distribution characteristics

Panel A – True empirical distributions				
"True" empirical distribution	Expected value	Standard deviation	Symmetry	Relative kurtosis
Scaled-t	0	1	Symmetric	Small excess
Symmetric NIG	0	1	Symmetric	Moderate excess
Asymmetric NIG	0	1	Negatively skewed	Large excess
Panel B – Theoretical distribution				
Standard normal	0	1	Symetric	–

Table 4
Parameters

Panel A – Normal inverse Gaussian				
Parameters	α	β	δ	μ
Symmetric NIG	1	0	1	0
Asymmetric NIG	1.031	−0.250	0.941	0.235
Panel B – Scaled-t				
	μ	σ	DoF	
Scaled-t	0	1	5	

We use a Monte Carlo simulation approach to estimate the power of the test. In order to reduce the variance of the simulation we apply the Stratified Sampling technique.[7] We considered sample sizes of 125, 250, 500 and 1,000. Besides the new distance and the CD, we also tested the power of the Kolmogorov distance. The procedure was the following:

We performed 25 sets of 1,000 MC runs. Each MC run had the following steps:

[7]That can be viewed as the one-dimensional case of the Latin Hypercube.

- Draw n random numbers U_i from a uniform distribution [0,1] using the Stratified Sampling technique[8] with 20 strata, where n is the sample size (125, 250, 500 and 1,000);

- Make $X_i = F^{-1}(U_i)$, where F is the cumulative distribution function of the true empirical distributions considered (see Table 3 for the list of distributions and the Appendix for a description of the distributions, including the pdf's);

- Calculate the three distances (Kolmogorov, CD, the New one) between the sample generated in the previous steps and the Standard Normal distribution (our theoretical distribution);

- For each distance and each TED, calculate the p-value for the null hypothesis that the true empirical distribution is equal to the theoretical distribution (Standard Normal) using critical values calculated as described in Section 4;

- If the p-value is lower than the significance level of 5%, we reject the null hypothesis, otherwise we do not reject it.

For each set of 1,000 MC runs, we calculated the percentage of type II error (i.e. do not reject the null hypothesis when it is actually false) dividing the number of "do not reject" by 1,000. So we had 25 type II errors from which we took the average and the standard deviation. The results are shown in Table 5, in terms of type II error percentages.

Table 5
Power of the tests

Panel A – Type II error – Mean									
	Scaled-t			Symmetric NIG			Asymmetric NIG		
Sample Size	New	CD	Kolmog	New	CD	Kolmog	New	CD	Kolmog
125	42.07%	81.97%	100%	38.15%	74.52%	100%	20.59%	44.86%	100%
250	20.67%	55.15%	100%	16.46%	25.49%	100%	5.60%	1.39%	98.97%
500	5.28%	5.80%	99.99%	3.69%	0.02%	98.89%	0.70%	0.00%	0.00%
1000	0.26%	0.00%	46.87%	0.06%	0.00%	0.00%	0.00%	0.00%	0.00%
Panel B – Type II error – Standard deviation									
	Scaled-t			Symmetric NIG			Asymmetric NIG		
Sample Size	New	CD	Kolmog	New	CD	Kolmog	New	CD	Kolmog
125	1.17%	1.58%	0%	1.55%	1.02%	0%	1.16%	1.07%	0%
250	1.44%	1.78%	0%	1.17%	1.87%	0%	0.59%	0.49%	0.29%
500	0.76%	0.83%	0.03%	0.66%	0.05%	0.36%	0.29%	0.00%	0.00%
1000	0.19%	0.00%	1.68%	0.06%	0.00%	0.00%	0.00%	0.00%	0.00%

The new and CD distances produce much better results than the Kolmogorov distance. For the two tail-focused distances, the type II error amounts to approximately zero with a sample size of 500, when considering the NIGs, while for the Scaled-t the type II error is near 5% with a sample size of 500, and near zero

[8] See Dowd (2002:300) for details of the Stratified Sampling technique.

with a 1,000-sample size. For the sample sizes of 125 and 250, the new distance outperforms the CD distance, while for samples sizes of 500 and 1,000, the results are similar, with the CD distance yielding slightly better results.

Results for the Kolmogorov distance can be considered acceptable only with big sample sizes. In the case of the Scaled-t, even the sample of 1,000 observations does not produce very reliable results.

Concluding, the two tail-focused distributions have a good power for samples with 500 or more observations. Even for samples of 250 observations, these tests can detect a mismatch, if the two distributions are not very similar. On the other hand, the Kolmogorov distance sometimes requires more than 1,000 observations.

6. Parameter Estimation and Goodness-of-fit Tests

In this section, data from the US Dollar/Brazilian Real exchange rate and São Paulo Stock Exchange Index are used to exemplify a goodness-of-fit test. We aim to test whether unconditional Normal, Generalized Hyperbolic (GH) and NIG[9] (Normal Inverse Gaussian) distributions are statistically equal to the empirical distribution. The parameters were estimated by maximum log-likelihood.

6.1 Data description

We use two samples of data to apply the test. The first is the US Dollar/Brazilian Real exchange rate (USD/BRL) from 01/13/1999 to 08/29/2002. The initial date coincides with the beginning of the free-floating exchange rate regime in Brazil. The second is the Ibovespa Index, from the São Paulo Stock Exchange, Brazil, from 07/01/1994 to 12/13/2001. The initial date was chosen because it was the beginning of the low inflation period in Brazil (the so-called "Plano Real" or Real Plan), after several years of high inflation.

Note that the size of both samples is big enough to produce reliable results, according to the assessment made in Section 5.

The returns used were logarithmic. In Table 6 we have the main information on the samples.

Table 6
Power of the tests

	Ibovespa	USD/BRL
Return average	0.000699379	0.00100
Return standard deviation	0.0279509	0.01310
Asymmetry	0.6033660	0.23560
Kurtosis	14.74590	21.2117
Number of observations	1,843	945

6.2 Parameter estimation

The parameters of a NIG and GH are estimated using maximum log-likelihood. Blæsild and Sørensen (1992) use maximum log-likelihood estimation for hyperbolic

[9]See in the Appendix that the NIG is a special case of the GH.

distributions, and Fajardo and Farias (2004) also use log-likelihood, but to estimate the general case of the GH. As we need to estimate a GH and NIG distribution, we use the Fajardo and Farias' approach and programs. Estimated parameters are shown in Table 7:

Table 7
Power of the tests

	N.I.G.		Generalized hyperbolic	
Parameters	Ibovespa	Dollar/Real	Ibovespa	Dollar/Real
α	31.909	32.77	1.7102	20.412
β	−0.00348	3.413905	−1.66835	0.150185
δ	0.023296	0.00527	0.03574	0.006388
μ	0.0012222	0.0004294	0.00199	0.0006121
λ	−0.5	−0.5	−1.828	−0.727

6.3 Critical values

After parameter estimation, the critical values for both CD and the new distances can be calculated considering our sample sizes (945 and 1,843 observations), for the GH and NIG estimated in Subsection 6.2 and for the Normal distribution. That is done using 10,000 Monte Carlo runs. The null hypothesis (H_0) and the alternative hypothesis (H_1) of the test are the following:
H_0: the estimated distribution is equal to the empirical distribution;
H_1: the estimated distribution is different from the empirical distribution.
The critical values for some selected significance levels are shown in Tables 8 and 9.[10]

Table 8
Critical values – New distance

	Distribution					
	Normal		NIG		GH	
	Sample size		Sample size		Sample size	
	945	1843	945	1843	945	1843
1%	0.5937	0.4638	0.5318	0.319	0.6209	0.5458
5%	0.2616	0.203	0.2853	0.181	0.3213	0.3072
10%	0.1957	0.1479	0.2125	0.1442	0.237	0.2125
20%	0.1512	0.1135	0.1727	0.1187	0.1861	0.1507

Table 9
Critical values – Normal – CD

	Distribution					
	Normal		NIG		GH	
	Sample size		Sample size		Sample size	
	945	1843	945	1843	945	1843
1%	0.0484	0.035	0.0587	0.0428	0.0589	0.0427
5%	0.0431	0.0311	0.0511	0.0375	0.0509	0.0375
10%	0.0402	0.0292	0.0475	0.0347	0.0475	0.0349
20%	0.037	0.0268	0.0434	0.0318	0.0434	0.0316

[10] Again, the tables contain the critical values V's such that $P[\text{Distance}>V]$ = confidence level, where the distance could be the new one or the CD.

6.4 Hypothesis tests

Finally, the null hypothesis that the estimated distribution is equal to the empirical distribution can be tested using criteria based on these two distances. Also, hypothesis tests using the Kolmogorov and Kuiper distances are performed to compare the results.

The distances between the empirical and theoretical distributions are calculated and compared to the critical values at the 1% significance level. The results are shown in Table 10 for the USD/BRL and Table 11 for the Ibovespa.

Table 10
Dollar/Real hypothesis tests

	Distribution								
	Normal			NIG			GH		
	Distance value	*p*-value	Test result at 1%	Distance value	*p*-value	Test result at 1%	Distance value	*p*-value	Test result at 1%
New distance	182,980	0.000	Reject	0.14304	0.374	Not reject	0.1475	0.398	Not reject
CD	0.2141	0.000	Reject	0.03530	0.547	Not reject	0.0524	0.038	Not reject
Kolmogorov	0.1334	0.000	Reject	0.02989	0.362	Not reject	0.0312	0.312	Not reject
Kuiper	0.2620	0.000	Reject	0.05859	0.34	Not reject	0.0616	0.019	Not reject

Table 11
Ibovespa hypothesis tests

	Distribution								
	Normal			NIG			GH		
	Distance value	*p*-value	Test result at 1%	Distance value	*p*-value	Test result at 1%	Distance value	*p*-value	Test result at 1%
New distance	50,957	0.000	Reject	0.1385	0.112	Not reject	0.0737	0.915	Not reject
CD	0.1169	0.000	Reject	0.0336	0.13	Not reject	0.0222	0.794	Not reject
Kolmogorov	0.0661	0.000	Reject	0.0166	0.683	Not reject	0.0093	0.997	Not reject
Kuiper	0.1305	0.000	Reject	0.0253	0.698	Not reject	0.0172	0.992	Not reject

As can be seen in the tables above, the tests reject the hypothesis that the estimated Normal distribution is equal to the empirical distribution at the 1% significance level for all distances considered. Also, the tables show that the hypothesis that the empirical distribution is equal to the estimated NIG cannot be rejected at the 1% significance level for distance tests. The only *p*-value of the NIG that is somewhat near rejection is that of the Kuiper distance of the Dollar/Real exchange rate; all others are far from rejection.

For both assets, the hypothesis tests do not reject that the empirical data have GH and NIG distribution. Now a new question arises: which distribution should we use in a risk model? If we are concerned only with goodness-of-fit, we should choose the distribution with the lowest distances. For the USD/BRL exchange rate, the NIG has all distances lower than GH, whereas in the Ibovespa index the GH has lower distances. Therefore, when using a parametric model for calculating the VaR of portfolios, we should use a NIG for the exchange rate, and a GH for the index.

If we are also interested in the parsimony of the model, we should consider also the number of parameters of the distributions. As the NIG has one less parameter than the GH (it is a special case), for the USD/BRL exchange rate, we would choose the NIG. For the Ibovespa index, we have a trade-off between goodness-of-fit and parsimony.

Finally, it is worth mentioning that this empirical application uses unconditional distributions, whereas actual risk models use conditional distributions such as a Normal distribution with a volatility given by EWMA or GARCH models. The tests used here can be adapted for conditional distributions, and in this case, the theoretical distribution has to be built with the use of the volatility model.

7. Conclusion

This paper analyzed the use of distances to test the goodness-of-fit of estimated distributions for VaR purposes. Besides the Crnkovic and Drachman (1996) distance, a new distance was proposed to perform goodness-of-fit tests on financial assets return distributions. Critical values were calculated to perform such tests.

The US Dollar/Brazilian Real exchange rate and the Ibovespa Index were used as examples of a practical application of how to test the hypothesis that an empirical distribution is equal to an estimated one. These tests were done considering Normal, NIG and GH distributions, and four distance criteria (Kolmogorov, Kuiper, CD and the new distance). For the Normal distribution, the test results rejected the hypothesis that the empirical distributions are equal to the estimated one, but there was no rejection for the NIG and GH at the 1% significance level.

The test can be easily applied to other kinds of distributions and assets, including portfolios, since it requires only the returns series and the expression for the distribution's density.

As a suggestion for further research, these VaR-focused distances (CD and the new one) can be used to estimate the distribution parameters through minimization of the distance. That was done in Prause (1999), where the author used an estimation method to minimize the Anderson-Darling distance. So, with an estimation focused on the tails of the distribution, the VaR measure is expected to be more reliable. Another interesting improvement is to consider more efficient sampling techniques in the Monte Carlo simulation, as is done by Saliby and Araújo (2001).

References

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

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society London*, A:401–419.

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

Berkowitz, J. (2002). Testing assumptions. *Risk*, 15(5):77–80.

Blæsild, P. & Sørensen, M. (1992). 'Hip' a computer program for analyzing data by means of the hyperbolic distribution. Research report 248, Department of Theoretical Statistics, Aarhus University.

Crnkovic, C. & Drachman, J. (1996). Quality control. *Risk*, 9(7):138–143.

Dowd, K. (2002). *Measuring Market Risk*. John Wiley and Sons.

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

Eberlein, E. & Prause, K. (2000). Thegeneralized hyperbolic model: Financial derivatives and risk measures. Mathematical Finance – Bachelier Congress 2000, 245–268.

Fajardo, J. & Farias, A. (2004). Generalized hyperbolic distributions and Brazilian data. *Brazilian Review of Econometrics*, 2(24):249–272.

Jorion, P. (2000). *Value at Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill, New York, second edition.

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

Kuiper, N. H. (1962). Proceedings of the Koninklijke Nederlandse Akademie van Wetenchappen. Ser. A, vol. 63, p. 38–47.

Lopez, J. A. (1998). Methods for evaluation value-at-risk estimates. *Federal Reserve Bank of New York Economic Policy Review*, October:119–124.

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

Prause, K. (1999). *The Generalized Hyperbolic Model: Estimation, Financial Derivatives and Risk Measures.* PhD thesis, University of Freiburg.

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

Saliby, E. & Araújo, M. S. (2001). Cálculo do VaR através de simulação de Montecarlo: Uma avaliação do uso de métodos amostrais mais eficientes. Annals of ENANPAD.

Appendix

Generalized Hyperbolic

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

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

where K_x is the modified Bessel function of the third kind and

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

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

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

The parameter α 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.

Its log-density is hyperbolic while Gaussian distribution log-density is a parabola, for this reason it is called Generalized Hyperbolic. We can do a reparametrization of the distribution so that the new parameters are scale invariant. The new parameters are defined in the equations:

$$\begin{array}{cl} \text{Second parametrization:} & \zeta=\delta\sqrt{\alpha^2-\beta^2}, \psi=\beta\alpha \\ \text{Third parametrization:} & \xi=(1+\zeta)^{-1/2}, \chi=\xi\psi \\ \text{Fourth parametrization:} & \bar{\alpha}=\alpha\delta, \bar{\beta}=\beta\delta \end{array}$$

The GH has several subclasses, among them, the Hyperbolic and Normal Inverse Gaussian (NIG). By 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 $\delta\to\infty$ and $\delta/\alpha\to\sigma^2$.

Scaled-t

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

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

where v is the degrees of freedom parameter, μ is the location parameter and σ the dispersion parameter.

When $v\rightarrow\infty$ the Student t converges to the Normal Distribution.