International VaR approach: Backtesting for different capital markets

Marília Cordeiro Pinheiro¹

https://orcid.org/0000-0002-0168-530X
Email: marilia.cordeiro90@hotmail.com

Bruno Vinícius Ramos Fernandes²

https://orcid.org/0000-0003-0860-563X
Email: brunoramos@unb.br

¹ Universidade de Brasília, Faculdade de Economia, Administração, Contabilidade e Gestão Pública, Programa de Pós-Graduação em Contabilidade, Brasília, DF, Brazil

² Universidade de Brasília, Faculdade de Economia, Administração, Contabilidade e Gestão Pública, Departamento de Contabilidade e Atuária, Brasília, DF, Brazil

ABSTRACT

This article aims to compare distinct metrics of the value at risk (VaR), differing from prior studies with respect about compare three asset categories belonging to seven countries. Since VaR inception, several approaches were developed to improve the loss estimation accuracy. However, there is hardly a universal consensus on which approach is the most appropriate, since VaR depends on statistical properties of the target asset and the market in which it is traded. It is relevant to compare the results obtained not only among the assets, but also among the markets in which they are traded, considering their specifics properties to verify if there is any pattern of the methods for the data. Considering the three asset categories, the semiparametric and non-parametric models obtained the lowest rejections number. It was also found that the models tested were not effective for the estimation of exchange rate VaR, which may be due to more relevant risks than the market in it asset price formation. Five models belonging to the parametric, semiparametric, and non-parametric approaches were tested. The analyses were divided in two, aiming to test the VaRs performances in distinct economic cycles; the first analyses considered a 1,000 days estimation window, while the second one considered a 252 days estimation window. To validated the results statistically, were applied the Kupiec and the Christoffersen tests. The results show that the conditional VaR and historical simulation have the best performance to estimate VaR. Comparing the markets, Chinese assets were the ones with the highest average number of tests rejections, which can be a consequence of its closed economy. Finally, it was found that shorter estimation window tends to perform better for high volatility assets, while longer window tends for lower volatility assets.

Keywords: VaR, parametric models, semiparametric models, non-parametric models, backtesting.

Correspondence address

Marília Cordeiro Pinheiro

Universidade de Brasília, Faculdade de Economia, Administração, Contabilidade e Gestão de Políticas Públicas, Programa de Pós-Graduação em Contabilidade

Campus Universitário Darcy Ribeiro, Bloco A-2 – CEP 70910-900

Asa Norte – Brasília – DF – Brazil

R. Cont. Fin. – USP, São Paulo, v. 31, n. 83, p. 318-331, May/Aug. 2020

318

1. INTRODUCTION

The value at risk (VaR) was created in the 1980's by JP Morgan, being disseminated by The Basel Committee in April 1995. At the end of the same year, the Securities and Exchange Commission (SEC) defined VaR as one of the three risk metrics that traded companies must use. VaR consists on an econometric tool to predict the worst loss over a target horizon within a given confidence interval (Jorion, 2007). Since its inception, several approaches were developed with the objective to improve the loss estimation accuracy and as an answer for the financial crises that have occurred over the years.

Although different, these methods have similar structures; from the assets daily returns, it made an inference of the distributions of these returns to estimate the desired VaR. The main divergence lies in the premise of the returns distribution, since there is a division between parametric models, which assumes a linear distribution, and non-parametric models, which defends a non-linear distribution (Engle & Manganelli, 2004). Considering that VaR is computed from specific assets statics properties and the markets which they are traded, difficultly will have a consensus about which of these two approaches is more adequate, as the financial instruments form heterogeneous classes, with different theoretical foundation of price formation and, consequently, the levels exposure risk.

Therefore, is necessary to apply the backtesting that aims to test the VaR accuracy based on the historical data, making it possible to analyses if a certain model had a good or bad performance and, consequently, verify if it is suitable for the target asset (Adams & Füss, 2009). Given the importance of VaR, a range of models was developed with the purpose of testing its accuracy. The Kupiec (1995) test focused on the measurement of the proportion of unconditional losses of the VaR model; if the proportion of failures occurs above the established p-percentile, it is an indication that the VaR tested underestimates the maximum asset loss. Another commonly used backtesting is Christoffersen (1998), which identifies whether violations cluster, that is, whether violations are independent from each other. The null hypothesis rejection is an indication of the model's delay in absorbing market oscillations in the evaluation of asset loss (Campbell, 2006).

In this paper, five VaR models are tested for three distinct asset classes of seven markets: Brazil, China, Germany, Japan, South Africa, United Kingdom, and United States of America. The article is relevant for comparing the results obtained not only among the assets, but also among the different markets in which they are traded, considering their specifics statistics properties, and verify if there is any pattern of the VaR models for the data. Additionally, is also tested the influence of the size of estimation window on VaR accuracy.

The paper is organized as follows: the section 2 contains the related literature, the section 3 describes the methodology, the section 4 summarizes the empirical results, and the section 5 presents the conclusion.

2. RELATED LITERATURE

The VaR analysis is more complex than the traditional forms of risk estimation, due to the dependence of multivariate distribution of risk factors and their dynamic, as in portfolio risk mapping. Although VaR can be accurately measured, it is limited to a specific time horizon and to the established probability interval. Additionally, VaR estimation is obtained from specific statistical characteristics of the asset and the market which it is traded. Considering all these factors, several extensions of its calculation have been developed, seeking to improve its predictive capacity. The main difference between these metrics is the premise of the returns distribution, since the characteristic of non-linearity on financial series is predominant, which puts the accuracy of parametric models in questioning.

In view of the scope of VaR metrics, previous literature has already addressed the comparison among the models performance. One of the most explored thematic is the lack of VaR subadditivity, which means that the portfolio risk can be larger than the sum of isolated risks of its components when estimated by VaR. In response to this, a popular alternative for the subadditivity violation is the expected shortfall model (ES), also knowns as conditional VaR (CVaR), proposed by Acerbi, Nordio, and Sirtori (2001). The method allows the risk factors decomposition by using its optimization portfolio property. Also, CVaR

focuses on the information contained in the tail and not on the entire distribution, giving the conditional expected value beyond the VaR level. In contrast to the defenders of CVaR, Danielson et al. (2005) explored the subadditivity violations focusing on heavy tailed assets and using a bivariate generalized autoregressive conditional heteroscedastic (GARCH) model to estimate the loss. For the most of sample, VaR is subadditive in the tail at probabilities that are most relevant for practical applications. The author reexamined the subadditivity question in 2013 and concludes that VaR is subadditive in the relevant tail region if asset returns are multivariate regularly varying and that VaR estimated by historical simulation model may violate subadditivity.

Historical simulation (HS) is widely used because it does not assume the normality premise of asset returns distribution, representing the segment of non-parametric models. Considering this, the main advantage of the HS is its coverage, since its application is not restricted only to linear portfolios, which makes it one of the most popular risk management methods. However, due its entirely dependence on the information contained in the historical data, it is subject to distortions of extreme events that occurred in a distant past and are no longer relevant in the loss estimation. Pritsker (2006) finds that risk estimates using the method delays to changes in conditional volatility and that it reacts asymmetrically, since the risk forecasting after larges losses, but not after large gains. Barone-Adesi and Giannopoulos (2001) find that HS fails to condition forecasts on the current state of the market, because it makes interval forecasts that are static, taking no notice of the last trading dates risk level.

Monte Carlo (MC) is another common simulation method used for VaR estimation, which is similar to HS, differentiating by the movements of the risk variables that are generated by the outline of some probability. In fact, this is one of the main fragilities of MC, since it is necessary to make assumptions about the process and to understand the sensitivity of this (Jorion, 2003). In a review of MC risk management, Hong, Hu, and Liu (2014) pointed out two important features about the model; first, the result is limited to the quality of the VaR model, which can cause distortions in loss distribution and, consequently, the risk may hide in the tail distribution. Second, in practice is difficult to make a realistic inference about the distribution precisely, considering that it is necessary a sufficiently large sample to arrive at a number that is approximately equal to the mathematical expectation of risk.

Although MC is a widely used method, the most popular VaR estimators are those derived from the autoregressive moving-average (ARMA) and GARCH models. Angelidis, Benos, and Degiannakis (2004) evaluated the performance of GARCH family for stock indices from United States of America, France, Germany, Japan, and United Kingdom. First, they detect that leptokurtic distributions can produce better VaR forecast. Second, that the ARCH structure producing the most accurate forecasts is different and specific for each stock market. So and Philip (2006) extended the test for 12 different market indices and four foreign exchange rates. The results show that among the models, risk metrics tends to be more robust of having less variation in the sample coverage, and that VaR estimation for exchange rates is less relied on the volatility models than stock market data.

As can be seen, the review of related literature reveals divergent results, which is expected considering that the financial market is composed by heterogeneous assets classes, with their own statistical properties and particularities. This provides a motivation to reinvestigate the accuracy VaR metrics for distinct assets and economies.

3. METHODOLOGY

To estimate VaR and compare the performance of each method among the assets, six different models are tested; exponentially weighted moving-average (EWMA), GARCH, and MC representing the parametric approach, HS representing the non-parametric approach, and CVaR representing the semiparametric approach.

3.1 Exponentially Weighted Moving-Average (EWMA)

The EWMA consists on an improvement of moving-average methods, especially for having the advantage of putting more weight in the most recent observations,

considering that it has the most relevant information about the asset risk.

The EWMA estimates the returns volatility σ for date t over a window from date t-k to date t-1:

$$\sigma^2 = (1-\lambda)\sum_{i=1}^{\infty}\lambda^{i-1}r_{t-i}^2 \qquad \boxed{1}$$

where λ denotes the decay factor, which $0 < \lambda < 1$, $\lambda^t \to 0$, $k \to \infty$, and so, as returns move further into the past, they will have less influence on σ estimation.

Usually, empirical studies show that $\lambda = 0.94$ permits a nice risk forecasting for market assets. The EWMA represents a linear model, assuming the normal distribution of returns. The estimation of the EWMA VaR of the 100% h-day is:

$$EWMA\ VaR = \Phi^{-1}(1-\alpha)\sigma\sqrt{h} \qquad \boxed{2}$$

where h corresponds to the target horizon estimation and $\Phi^{-1}(1-\alpha)$ is the distribution function of the quantile $1 - \alpha$.

3.2 Generalized Autoregressive Conditional Heteroscedastic (GARCH)

The GARCH model considers that the conditional volatility σ is a function of continuous change its previous values squares which generates the volatility clusters. The model is autoregressive since the return Y_t depends on the values of Y_{t-1}, which suggests the heterosledasticity observed over different periods can be autocorrelated.

Let $r_t = ln(S_t / S_{t-1})$ be the continuously compound rate of return from time $t - 1$ to t, where S_t is the asset price at t moment. It is assumed that the time series of interest r_t is decomposed into two parts: the predictable

and unpredictable component $y_t = E(y_t|I_{t-1})+\varepsilon_t$, where I_{t-1} is the information time at $t - 1$, E is the mean operator, and ε_t is the unpredictable part that can be expressed as an ARCH process:

$$\varepsilon_t = z_t\sigma_t \qquad \boxed{3}$$

where z_t is a sequence of independently and identically distributed random variables with 0 mean and unit variance. The conditional variance of ε_t is σ_t^2, a time varying, positive and measurable function of the information set at time t-1 (Angelidis et al., 2004).

Engle and Manganelli (2004) developed the ARCH$_{(q)}$ model and expressed the conditional variance as a linear function of the past q squared innovations:

$$\sigma_t^2 = a_0 + \sum_{i=1}^{q}a_i\varepsilon_{t-1}^2 \qquad \boxed{4}$$

The GARCH(p,q) model is a generalization of the ARCH model, proposed by Bollersev (1994). For the conditional variance to be positive, the parameters must satisfy $a_0 > 0$ and $a_i \geq 0$ for $i = 1, \ldots, q$. Based on these restrictions, the GARCH model is expressed by:

$$\sigma_t^2 = a_0 + \sum_{i=1}^{q}a_i\varepsilon_{t-1}^2 + \sum_{i=1}^{q}b_i\sigma_{t-1}^2 \qquad \boxed{5}$$

The parameters are estimated by maximum likelihood under the assumption that the returns are normally distributed $Y_t \sim N(0,\sigma_t^2)$ and $D(z_t;v)$, being their density function, the log-likelihood function of $\{y_t(\theta)\}$ for a sample of T observations is given by:

$$L_t(\{y_t\};\theta) = \sum_{t=1}^{T}\left[\ln\left[D(z_t(\theta);v)\right]-\frac{1}{2}\ln(\sigma^2(\theta))\right] \qquad \boxed{6}$$

In summary, the one-step ahead conditional variance forecast $\sigma_{t+1|t}^2$ for the GARCH(p,q) model equals:

$$\sigma_{t+1|t}^2 = a_0 + \sum_{i=1}^{q}a_1\varepsilon_{t-i+1}^2 + \sum_{i=1}^{q}b_j\sigma_{t-j+1}^2 \qquad \boxed{7}$$

Therefore, the one-step VaR forecasts under all distributional assumptions and for zero mean observations calculated by

$$VaR_{t+1|t} = F(\alpha)\sigma_{t+1|t} \qquad \boxed{8}$$

where $F(\alpha)$ is the corresponding quantile (95th or 99th) of the assumed distribution and $\sigma_{t+1|t}^2 = a_0$ is the forecast of the conditional standard deviation at time $t+1$ given the information at time t.

3.3 Monte Carlo (MC)

The process for MC's estimation is based on risk factor mapping. It is assumed that the portfolio mapping is based on returns rather than on changes in the equity risk factors. Hence, the VaR will be estimated as a percentage

R. Cont. Fin. – USP, São Paulo, v. 31, n. 82, p. 318-331, May/Aug. 2020

321

of the portfolio value. The basic algorithm for generating correlated simulation on k risk factors returns is based on k-dimensional, *i.i.d* normal process. Therefore, the marginal distribution of the risk factor's return is $N(\mu_i, \sigma_i^2)$ for $i = 1,...k$, and the risk factor correlations are represented in a $k \times k$ matrix C. The algorithm begins with k independent simulations on standard uniform variables, transforms these into independent standard normal simulation, and then uses the Cholesky matrix of the risk factor returns covariance to transform these into correlate zero-mean simulations with the appropriate variance. Therefore, the mean excess return is added to each variable (Alexander, 2009).

Given this approach, the covariance matrix is written as:

$$\Omega = DCD \qquad \boxed{9}$$

where $D = diag(\sigma_1...\sigma_n)$. The Cholesky matrix is a lower triangular $k \times k$ matrix Q such $\Omega = QQ$. The expected returns in a vector is written as $\mu = (\mu_1,...\mu_k)$, then the $k \times 1$ multivariate normal vector x are generated by simulating $k \times 1$ independent standard vector z and setting $x = Qz + \mu$. It is simulated a very large number of such vectors x and apply the portfolio mapping to each simulation, producing N simulations on the portfolio returns. Next, it is simulated $N - h - day$ portfolio excess returns with the purpose to find their empirical distribution, to find the α quantile of this distribution, and to multiply this by -1, which is the $h - day$ VaR estimate (Alexander, 2009). In this work, the MC model was calculated using *VaR*, using 10 million simulations.

3.4 Historical Simulation (HS)

The HS is a non-parametric model which assumes that all possible future oscillations have been experienced in the past and that historically simulated distribution is identical to the returns' distribution over the forward target risk horizon. Historical scenarios in recent movements in risk factors are used to simulate many possible portfolio values in h days' time (Alexander, 2009). The VaR obtained by the HS is estimated from the construction of hypothetical values from a current observation given by:

$$f_i^k = f_{i,t} + \Delta f_i^k \qquad \boxed{10}$$

where $f_{i,t}$ is the risk factor of the portfolio P_t. These hypothetical values are used to construct the hypothetical portfolio p^k, considering the new scenario from the equation:

$$P^k = P\left[f_1^k, f_2^k ... f_N^k \right] \qquad \boxed{11}$$

The oscillations of portfolios values $P^k = (P^k - P_t)/P_t$ are obtained with the equations above. The returns t are ordered and then are chosen those that correspond to *quantil cth Rp* (c). The VaR is obtained by the difference between the mean and the *quantil*:

$$VaR = AVE\left[R_p\right] - R_p(c) \qquad \boxed{12}$$

3.5 Conditional Value at Risk (CVaR)

The last model tested is the CVaR, also known as ES, which is a differentiated model, since it is concentrated on the information that is associated on the quantile below the probability p. The math function on the estimated loss of the asset Y is given by:

$$E(Y) = \int_{-\infty}^{\infty} yf(y)dy \qquad \boxed{13}$$

However, to VaR estimation, the function must be changed, considering that the expectative does not range from $-\infty$ to ∞, but $-\infty$ to $-VaR(p)$. The area below $f_q(.)$ on the interval $[-\infty, -VaR(p)]$ is smaller than one, which implicates that $f_q(.)$ is not an adequate function for this context. Thus, the new density function $f_{VaR}(.)$ is defined by the positive adjust of $f_q(.)$, so that the area above this value becomes unitary (Danielsson, 2011). To identify the correct density distribution, it is applied:

$$E(Y) = \int_{-\infty}^{-VaR(p)} yf(y)dy \qquad \boxed{14}$$

Therefore, the density of the $f_{VaR}(.)$ tail is:

$$1 = \int_{-\infty}^{-VaR(p)} fVaR(y)dy = \frac{1}{p} \int_{-\infty}^{-VaR(q)} f_q(y)dy \qquad \boxed{15}$$

The CVaR is obtained by the reason between the profit and the loss over the density of $f_q(.)$ tail:

$$CVaR = \int_{-\infty}^{-VaR(p)} fVaR(y)dy \qquad \boxed{16}$$

3.6 Statistical Tests

To verify and compare the performance among the models is used the methodology proposed by Danielsson (2011). Firstly, is calculated the violation ratio (VR) that has the purpose of measures whether the current return of a specific day exceeds the VaR obtained based on the estimation window. Considering the violations equal to n_t, it is assumed that when the violation occurs, $n_t = 1$ and

322

R. Cont. Fin. – USP, São Paulo, v. 31, n. 82, p. 318-331, May/Aug. 2020

$n_t = 0$ otherwise. The number of violations is incorporated on the variable v_1, while v_0 corresponds to the number without violations.

$$n_t = \begin{cases} 1 \, if \, y_t \leq -VaR \\ 0 \, if \, y_t > -VaR \end{cases}$$

$$v_1 = \sum n_t$$ [17]

$$v_0 = testing \, window \, size - v_1$$

The VR is:

$$VR = \frac{observed \, number \, of \, violations}{expected \, number \, of \, violations} = \frac{v_1}{p \times testing \, window \, size}$$ [18]

Danielson (2011), based on Basel III accords, used the rule of thumb that if VR $\in [0.8, 1.2]$ it is a good forecast and, if VR < 0.5 or > 1.5 the model, respectively, underestimates and overestimated the risk. To validate statically the VR values, Kupiec (1995) and Christoffersen (1998) tests are applied. The first one considers only the frequency of the violations and not the time in which they occurred. Thus, the Christoffersen test (1998) is applied so that there is no error in rejecting a model that produced clustered violations.

For Kupiec (1995), the null hypothesis for VaR violation is:

$$H_0 : \eta \sim B(p)$$ [19]

with B representing the Bernoulli distribution. The Bernoulli density is given by:

$$\hat{p} = \frac{v_1}{W_t}$$ [20]

under H_0, $p = \hat{p}$ /, therefore the restrict maximum likelihood function is:

$$l_R(\hat{p}) = \prod_{t=W_e}^{T} (1-p)^{1-\eta_t} (p)^{\eta_t} = (1-p)^{v_0} (p)^{v_1}$$ [21]

As said before, the Christoffersen test (1998) has as advantage to identify whether violations cluster, considering that, theoretically, they should be independent. If the null hypothesis is rejected, it is an indication that the model delays in absorbing the oscillations that occur in the market for the asset tested. It is needed to calculate the probabilities of two consecutives violations and the probability of a violation if there was no violation on the previous day:

$$p_{ij} = \Pr(\eta_t = i | \eta_{t-1} = j |$$ [22]

The statistical test is given by:

$$LR = 2(logl_U(\Pi_0) - logl_R(l_R(\Pi_1)) \sim \chi^2$$ [23]

where Π_0 is the estimated transition matrix and Π_1 is the transition matrix. Under the null hypothesis of no violations cluster, the probability of a violation tomorrow does not depend on a today violation; therefore, $p_{01} = p_{11} = p$. The test of independence is asymptotically distributed as a χ^2.

The CVaR backtesting differs from the other models since what is being tested is a loss beyond VaR. Danielsson (2011) presents a methodology to backtesting CVaR that is analogous of the VR. When VaR is violated, normalized shortfall NS is calculated as:

$$NS_t = \frac{y_t}{ES_t}$$ [24]

with ES being the observed ES on day t . Then the expected y_t for a violated VaR is:

$$\frac{E[Y_t | Y_y] < -VaR_t]}{ES_t} = 1$$ [25]

Given that, the null hypothesis defines that average NS should be equal to one:

$$H_0 = \overline{NS} = 1$$ [26]

3.7 Data

The sample is composed of three categories of assets belonging to seven countries with different economic status. Countries were selected based on liquidity criteria

R. Cont. Fin. – USP, São Paulo, v. 31, n. 82, p. 318-331, May/Aug. 2020

323

and market representativeness: representing developing economies, the assets of South Africa, Brazil, and China were selected, 17th, 20th, and 5th largest capital markets respectively. Representing the developed countries, the assets of the United States of America, Germany, Japan, and United Kingdom, 1st, 10th, 3rd, and 4th largest capital markets, respectively, were chosen (based on the stock to trade major stock exchange in the world infographic, retrieved from https://stockstotrade.com/major-stock-exchanges-in-the-world-infographic/).

The equity market is represented by the New York Stock Exchange (NYSE), the Shanghai Stock Exchange (SSEC), the London Stock Exchange (LSE), the Bovespa Index (B3), the Nikkei 225, the Johannesburg Stock Exchange (JSE), and the German Stock Index (DAX). The bond market is composed by United States Treasury bond (U.S. T-Bond), Chinese government bond, Brazilian government bond (NTN-B), Japanese government bond, South African government bond, and Deutsche government bond. The exchange rate, last asset class tested, is represented by *yuan* (CNY), British *pound* (GPB), *real* (BRL), *yen* (JPN), South African *rand* (ZAR), and *euro* (EUR). An important observation is that the dollar is not used on the exchange rate, since it is the greater exchange representativeness, being used as parity for the other currencies. It covers the period from January 2 until December 31. The liquidity and the size criteria were used to select the indices for the sample composition.

4. EMPIRICAL ANALYSIS

The empirical analysis is structured as follows. It starts with the descriptive statistics, which is a fundamental topic considering that VaR uses the statistical properties to estimate the losses. The next subsection presents the values of VR with the purpose of analyses the performance of VaR models and simultaneously verify if there is a predominance of a model for a given type of market or asset. Finally, the results obtained are validated based on the Kupiec (1995) and Christoffersen (1998) tests.

4.1 Descriptive Statistics

The descriptive statistics provide an insight into the investment properties of different assets and markets. Table 1 summarizes the descriptive statistics for the data.

Table 1
Descriptive statistics of the raw data

Country	Assets	Annual mean (%)	Annual volatility (%)	Skewness	Kurtosis	JB	p-value JB
United States of America	NYSE	0.06	0.33	-0.35	9.23	6,076.90	0.00
	U.S. T-bond	0.01	0.06	-0.11	5.14	3,103.90	0.00
China	SSEC	-0.02	0.28	-0.61	4.08	1,959.10	0.00
	Chinese government bond	-0.02	0.31	-0.31	7.84	6,191.20	0.00
	CNY/USD	0.02	0.02	-0.56	17.35	36,117.00	0.00
United Kingdom	LSE UK	0.07	0.37	0.18	6.76	5,205.60	0.00
	United Kingdom government bond	-0.04	0.10	-1.00	13.31	2,164.00	0.00
	GPB/USD	-0.04	0.10	-1.00	13.30	21,617.00	0.00
Brazil	B3	0.01	0.28	-0.04	6.02	4,100.00	0.00
	NTN-B	-0.04	0.29	0.32	8.97	7,669.10	0.00
	BRL/USD	-0.05	0.17	-0.28	6.43	4,976.90	0.00

Table 1
Cont.

Country	Assets	Annual mean (%)	Annual volatility (%)	Skewness	Kurtosis	JB	p-value JB
Japan	Nikkei 225	-0.01	0.25	-0.50	7.98	7,337.10	0.00
	Japanese government bond	0.00	1.56	2.99	10.21	1,155.90	0.00
	JPY/USD	-0.02	0.11	-0.10	2.81	875.60	0.00
South Africa	JSE	0.06	0.21	-0.09	3.74	1,608.20	0.00
	South Africa government bond	-0.01	0.15	0.15	13.13	20,999.00	0.00
	ZAR/USD	-0.06	0.18	-1.28	17.28	36,447.00	0.00
Germany	DAX	0.03	0.22	-0.02	6.01	4,204.20	0.00
	Deutsche government bond	-0.66	1.46	-31.51	17.46	2,134.36	0.00
	EUR/USD	-0.01	0.10	0.09	2.29	632.50	0.00

Note: *The data summarizes statistics for the equity markets, government bond markets, and exchange rates from January 2008 until December 2017.*

B3 = Bovespa Index; BRL = real; CNY = yuan; DAX = German Stock Index; EUR = euro; GPB = British pound; JPY = Japanese yen; JSE = Johannesburg Stock Exchange; LSE UK = London Stock Exchange; NTN-B = Brazilian government bond; NYSE = New York Stock Exchange; SSEC = Shanghai Stock Exchange; U.S. T-bond = United States Treasury bond; USD = dollar; ZAR = South African rand.

Source: *Elaborated by the authors.*

As can be seen, the null hypothesis of normality is rejected at any significance level for all markets, which is a violation for the parametric model's main premise. Additionally, all assets have positive excess kurtosis, a characteristic of leptokurtic returns distribution with fat tails and exposed to extreme events. The equity market with the highest return is the LSE, which is also the most volatile and the only one with positive asymmetry. On the other hand, the SSEC has the lowest return, being the third most volatile. In general, the stock portfolios are the assets with the highest average return and greater volatility. For the government bond market, German bond record the lowest return. This may be a consequence of the policy adopted in 2016, which were issued bonds with a negative yield. An important point to add is that Japan also implemented this policy and, according to Table 1,

although the annual average is not negative, the Japanese bond is the asset with the highest volatility. By treating these two bonds as outliers, government bonds have the second highest average volatile of the data, while equity indices have the highest one. The exchange rate is the asset class with the lowest average volatility, and the *yuan* is the only one with a positive annual return. The South African and Brazilian currencies presented the greatest volatility and average devaluation.

4.2 Analysis for the Entire Period

Firstly, is applied the test for the whole period (2007-2017), with a 1,000 days estimation window. Table 2 shows the VRs and their statistical significance of Kupiec (1995) and Christoffersen (1998) tests with the purpose to validate the results.

Table 2
Backtesting for 99% value at risk (VaR) estimation (2007-2017)

Countries	Assets	EWMA	HS	GARCH	CVaR	MC
United States of America	NYSE	1.43	0.14*	1.28	1.87	0.43
	U.S. T-bond	2.21*	0.39*	1.71*	1.03	1.16
China	SSEC	2.52*	1.26	1.89*	1.14*	2.01*
	Chinese government bond	2.43***	0.07	1.43	1.28*	0.43*
	CNY/USD	1.71*	1.71***	1.87*	1.80	2.94***

R. Cont. Fin. – USP, São Paulo, v. 31, n. 82, p. 318-331, May/Aug. 2020

325

Table 2
Cont.

Countries	Assets	EWMA	HS	GARCH	CVaR	MC
United Kingdom	LSE UK	1.16*	0.52***	0.87**	1.16	0.93
	United Kingdom government bond	1.77*	0.91	1.71*	1.03	1.16**
	GPB/USD	1.93*	1.93	1.44	1.17	2.89*
Brazil	B3	1.28	0.99*	1.05	1.19	1.22
	NTN-B	1.33	0.39	1.02	1.01	1.33
	BRL/USD	2.14*	0.96	1.55*	0.91	1.71*
Japan	Nikkei 225	2.09***	1.05**	1.98*	1.22**	1.63***
	Japanese government bond	1.70**	2.86	1.46	1.10	4.20
	JPY/USD	1.93*	1.15	1.15	0.77	1.15
South Africa	JSE	1.83*	0.69	1.43	0.97	1.20
	South Africa government bond	1.74*	1.20**	1.20	1.10	1.25**
	ZAR/USD	2.30*	1.02**	1.98	1.13*	1.71***
Germany	DAX	2.11*	0.61	1.56*	1.00	1.34***
	Deutsche government bond	1.43**	3.93***	1.59***	1.11	4.20***
	EUR/USD	1.27	0.48*	1.22	1.16	1.38

B3 = Bovespa Index; BRL = real; CNY = yuan; CVaR = conditional VaR; DAX = German Stock Index; EUR = euro; EWMA = exponentially weighted moving-average; GARCH = generalized autoregressive conditional heteroscedastic; GPB = British pound; HS = historical simulation; JPY = Japanese yen; JSE = Johannesburg Stock Exchange; LSE UK = London Stock Exchange; MC = Monte Carlo; NTN-B = Brazilian government bond; NYSE = New York Stock Exchange; SSEC = Shanghai Stock Exchange; U.S. T-bond = United States Treasury bond; USD = dollar; ZAR = South African rand.

* = significant at 5% level for the Kupiec test (1995); ** = significant at 5% level for the Christoffersen test (1998); *** = significant at 5% level for the Kupiec (1995) and the Christoffersen (1998) tests.

Source: *Elaborated by the authors.*

Relying on the premise that a VR ∈ [0.8,1.2] is a good forecast, based on the data of Table 2, United Kingdom has the highest percentage of adequate VR (53%), followed by Brazil and South Africa (both with 47%), while China has the lowest value (7%). The Chinese market is also the one with the highest number of rejections of Kupiec (1995) and Christoffersen (1998) tests. These results may be due to intrinsic particularities of Chinese financial market, considering not only its closed economy, but also the intervention of the government to keep the currency and the interest rates at a low level.

Comparing the assets, the percentage of appropriate VR is close, with stocks and government bonds having the best performance (35%) and exchange rate having the worst (30%). This can be a consequence of the complex formation of currency pricing, given that it depends of internal and external policies, especially of the United States of America. The Chinese currency is the only one that did not get any adequate VR, while the Japanese has the largest number (3). Among the stock indices, LSE has the highest number of appropriate VR (3), while NYSE did not have any one. For government bonds, United Kingdom bond

has the highest number (3), while Chinese bond the lowest (0). While CVaR has the highest number of accurate VR for all asset categories, EWMA has the lowest number. HS has the second best performance for equity indices, MC has the second best performance for government bond indices, except for CVaR and EWMA; all other models have only one adequate VR for exchange rate.

Considering the results of VaR estimation models, CVaR presents the highest number of accurate VR (15), however, it must be emphasized a restriction on the comparability of this method on the others, since CVaR is distinguished by concentrating on the information contained in the left tail and estimates the losses below VaR quantile. The second model with best performance is HS (6), characterized by not assuming the normality premise. The EWMA is the model with the lowest number of appropriate VR. Compared to other methods, EWMA is the simplest one in the aspect that is a GARCH model with only one parameter with $\alpha = 0$.

The results of the Kupiec (1995) statistical test at 5% of significance show that EWMA has the highest number of rejections for the null hypothesis (14), while

326

R. Cont. Fin. – USP, São Paulo, v. 31, n. 82, p. 318-331, May/Aug. 2020

CVaR has the lowest (3). For the Christoffersen test (1998), MC presents the highest number of rejection for the null hypothesis (7), indicating a delay to absorb markets movements information. GARCH and CVaR have the lowest number (2). The good performance of GARCH model for the test may be a consequence of the relevance of heteroscedasticity on the risk estimation, a typical characteristic on financial data. HS has the second highest number of rejections for both statistical tests, especially for those VR, which are not on the appropriate interval. Among the asset categories, stocks have the lowest rejections percentage, while exchange rates have the highest.

4.3 Analysis for Subperiods

Next, is tested a smaller window and analysis the performance of VaR models over subperiods. The estimation window has been reduced for 252 days, equivalent to one year of trading. The data is segregated in the following subperiods: 2007-2010, 2011-2014, and 2015-2017.

Table 3
Subperiod 2007-2010 backtesting for 99% value at risk (VaR)

Countries	Assets	EWMA	HS	GARCH	CVaR	MC
United States of America	NYSE	1.03	0.11	1.03	1.00	0.57
	U.S. T-bond	1.04	1.24	0.89	1.10	0.48*
China	SSEC	1.94	0.51	2.06	1.23	2.19
	Chinese government bond	1.68	1.00	1.51	1.16	2.18*
	CNY/USD	1.97	0.00***	1.43	1.37***	1.28***
United Kingdom	LSE UK	2.27	0.75	2.27	1.01	2.14*
	United Kingdom government bond	2.13	1.06	1.78	1.09	3.08*
	GPB/USD	2.12	0.95	1.78	1.09	3.08*
Brazil	B3	2.16	1.14	1.91	1.06	2.16
	NTN-B	0.00	0.56	0.00	0.00	1.12
	BRL/USD	1.70	0.85	1.46	1.13	2.19*
Japan	Nikkei 225	1.81	1.16	1.29	1.08	3.10***
	Japanese government bond	0.90	0.72	0.72	1.03	1.09
	JPY/USD	1.47	0.73	1.47	1.27	1.47
South Africa	JSE	1.63	0.99	1.49	1.03	1.99
	South Africa government Bond	1.17	0.99	1.49	1.03	1.99**
	ZAR/USD	2.21	1.35	1.72	1.11	1.72
Germany	DAX	2.07	0.61	1.59	1.09	2.07
	Deutsche government bond	1.99	1.24	2.24	0.99	2.62*
	EUR/USD	1.42	2.13	1.89	1.09	2.96

B3 = Bovespa Index; BRL = real; CNY = yuan; CVaR = conditional VaR; DAX = German Stock Index; EUR = euro; EWMA = exponentially weighted moving-average; GARCH = generalized autoregressive conditional heteroscedastic; GPB = British pound; HS = historical simulation; JPY = Japanese yen; JSE = Johannesburg Stock Exchange; LSE UK = London Stock Exchange; MC = Monte Carlo; NTN-B = Brazilian government bond; NYSE = New York Stock Exchange; SSEC = Shanghai Stock Exchange; U.S. T-bond = United States Treasury bond; USD = dollar; ZAR = South African rand.
***** *= significant at 5% level for the Kupiec test (1995); ** = significant at 5% level for the Christoffersen test (1998); *** = significant at 5% level for the Kupiec (1995) and the Christoffersen test (1998) tests.*
Source: *Elaborated by the authors.*

R. Cont. Fin. – USP, São Paulo, v. 31, n. 82, p. 318-331, May/Aug. 2020

327

Table 4

Subperiod 2011-2014 backtesting for 99% value at risk (VaR)

Countries	Assets	EWMA	HS	GARCH	CVaR	MC
United States of America	NYSE	0.84	0.17	0.68	1.01	0.12***
	U.S. T-bond	0.91	0.91	0.91	1.21	0.34
China	SSEC	1.95	0.91	1.30	1.10	1.43
	Chinese government bond	2.84	0.59	2.55	1.36	1.35
	CNY/USD	1.55	0.88	1.33	1.17	1.32
United Kingdom	LSE UK	1.39	0.76	1.39	1.15	0.88
	United Kingdom government bond	2.00	1.06	2.01	1.09	1.90
	GPB/USD	2.03	1.08	2.03	1.09	1.92
Brazil	B3	1.39	1.14	1.14	1.05	1.51
	NTN-B	0.55	0.00***	0.00***	1.90***	0.55
	BRL/USD	2.57	1.47***	2.08***	1.05***	2.81
Japan	Nikkei 225	1.66	0.77	1.28	1.12	1.79***
	Japanese government bond	2.17	1.81	1.99	1.09	2.53
	JPY/USD	0.67	0.67	0.67	1.17	0.67
South Africa	JSE	1.50	1.25	1.87	1.12	2.38
	South Africa government bond	1.71	1.45	1.45	1.18	1.18*
	ZAR/USD	2.13	0.59	1.78	1.14	1.66
Germany	DAX	2.07	0.61	1.59	1.09	2.07
	Deutsche government bond	1.99	1.24	2.24	0.99	2.62
	EUR/USD	1.42	2.13	1.89	1.09	2.96

B3 = Bovespa Index; BRL = real; CNY = yuan; CVaR = conditional VaR; DAX = German Stock Index; EUR = euro; EWMA = exponentially weighted moving-average; GARCH = generalized autoregressive conditional heteroscedastic; GPB = British pound; HS = historical simulation; JPY = Japanese yen; JSE = Johannesburg Stock Exchange; LSE UK = London Stock Exchange; MC = Monte Carlo; NTN-B = Brazilian government bond; NYSE = New York Stock Exchange; SSEC = Shanghai Stock Exchange; U.S. T-bond = United States Treasury bond; USD = dollar; ZAR = South African rand.

* = significant at 5% level for the Kupiec test (1995); ** = significant at 5% level for the Christoffersen test (1998); *** = significant at 5% level for the Kupiec (1995) and the Christoffersen test (1998) tests.

Source: *Elaborated by the authors.*

Table 5

Subperiod 2015-2017 backtesting for 99% value at risk (VaR)

Countries	Assets	EWMA	HS	GARCH	CVaR	MC
United States of America	NYSE	0.44	0.22	0.22	0.22	1.00*
	U.S. T-bond	1.53	0.34	1.53	1.18	1.02
China	SSEC	2.90	0.22	2.45	1.19	1.34
	Chinese government bond	1.67	0.37	1.30	1.11	1.11***
	CNY/USD	1.89	0.95	2.01	2.31	2.49
United Kingdom	LSE UK	1.11	1.11	1.29	1.15	1.11
	United Kingdom government bond	1.72	0.86	1.37	1.33	1.20
	GPB/USD	1.39	0.76	1.39	1.33	1.20
Brazil	B3	1.39	1.14	1.14	1.05	1.52
	NTN-B	0.00*	0.39	0.39	0.00*	0.39
	BRL/USD	1.43**	0.71	1.25	1.44**	0.71
Japan	Nikkei 225	1.76	0.88	2.12	1.25	1.24
	Japanese government bond	1.13	0.99	0.99	1.05	2.27
	JPY/USD	0.64	0.64	0.74	1.17	0.64

Tabela 5
Cont.

Countries	Assets	EWMA	HS	GARCH	CVaR	MC
	JSE	2.00	0.54	1.27	0.98	0.91**
South Africa	South Africa government bond	1.52	0.76	0.95	1.02	1.90**
	ZAR/USD	1.88	1.69	2.44	1.12	1.69
	DAX	1.42	0.71	1.07	1.11	0.89
Germany	Deutsche government bond	1.03	3.11	1.55	1.01	2.93***
	EUR/USD	1.00	1.17	0.84	1.47	1.00

Notes: *The table presents the percentage of appropriate violation ratio and the percentage of rejection for Kupiec (1995) and Christoffersen (1998) tests at 5% significance level for each asset.*
B3 = Bovespa Index; BRL = real; CNY = yuan; CVaR = conditional VaR; DAX = German Stock Index; EUR = euro; EWMA = exponentially weighted moving-average; GARCH = generalized autoregressive conditional heteroscedastic; GPB = British pound; HS = historical simulation; JPY = Japanese yen; JSE = Johannesburg Stock Exchange; LSE UK = London Stock Exchange; MC = Monte Carlo; NTN-B = Brazilian government bond; NYSE = New York Stock Exchange; SSEC = Shanghai Stock Exchange; U.S. T-bond = United States Treasury bond; USD = dollar; ZAR = South African rand.
* = significant at 5% level for the Kupiec test (1995); ** = significant at 5% level for the Christoffersen test (1998); *** = significant at 5% level for the Kupiec (1995) and the Christoffersen test (1998) tests.
Source: *Elaborated by the authors.*

Comparing the VRs and the Kupiec (1995) and Christoffersen (1998) rejection number among the subperiods, 2007-2010 has the highest percentage of tests rejections. Although the subprime crisis (2007-2008) started in the United States of America, the country has the highest percentage of adequate VRs, with a good performance for EWMA and GARCH, characterized by considering the volatility clusters on the estimation loss and CVaR. China has the highest percentage of rejections, largely concentrated on its currency. From 2007 to 2009, the Chinese central bank implemented a series of measures to depreciate the *yuan* against the dollar and to contain the effects of the financial crisis. Considering that the models tested are focused on market risk, it is expected that they are not able to incorporate governmental actions on the asset's values. Another important observation is that, with exception of China, for the first subperiod, all the rejections among the countries occurred for MC method. One of the fragilities of this model is that the prespecified model is not correct, which is a risk that can increase during financial crisis, considering the high volatility of the assets. By the opposite, based on the percentage of appropriate VR, the CVaR has the best performance, which was expected since the model estimate the risk based on the tail of the loss distribution, being more conservative than the other models.

For the subperiod of 2011-2014, Brazil was the country with the highest percentage of rejection, which is concentrated on NTN-B and in BRL/USD. These can be a result of two associate factors: for the NTN-B, the government implemented a monetary policy to raise

interest rates, causing the basic interest rate to rise as from 2013. Considering the Brazilian currency, during these years, the country was affected by the commodity crisis, the main export good of Brazilian economy, which decreased the amount dollar in the country, leading to the *real* (Brazilian currency) depreciation. An additional observation is that, unlike the subprime crisis of the first period, the CVaR does not present a good performance for the Brazilian crisis, considering that the model was rejected for Kupiec (1995) and Christoffersen (1998) tests. However, despite the Brazilian case, again CVaR have the best performance, followed by HS, a non-parametric model.

For the period of 2015-2017, one more time the Brazilian assets have the highest rejection number concentrated in NTN-B. The main risk to which the bonds are exposed is the raise of interest yield, considering that its increase has a negative impact on its value. As well as the Chinese assets on the first subperiod, again it can be seen that the models tested are not efficient in incorporate the political risk, since during 2015-2017, the government established a political of increase yield. Based on the VR, Germany has the highest number of appropriate ratios, followed by United Kingdom. During 2015-2017, it was implemented an investment package on European Union, which, according to the World Bank Union Europe Annual Report from 2018 (retrieved from https://publications.europa.eu/en/publication-detail/-/publication/e977293e-8743-11e9-9f05-01aa75ed71a1/language-en/format-PDF), promoted a modest economic recovery started in 2014. Assets tend to be less volatile

R. Cont. Fin. – USP, São Paulo, v. 31, n. 82, p. 318-331, May/Aug. 2020

329

during more stable economic periods, which may have improved the accuracy of the models for these countries (Mei & Guo, 2004; Shwert, 2011). The MC was the method with the highest rejection percentage, concentrated on the government bonds.

Comparing the assets, the equity indices are the one with the lowest average rejection percentage for Kupiec (1995) and Christoffersen (1998) tests (14 and 1%, respectively), government bonds have the highest percentage for Kupiec (1995) (24%), and exchange rate for Christoffersen (1998) (12%). It must be emphasized that, despite the Kupiec (1995) results for government bonds, Japanese and Germany bonds, the most volatilities assets of the data, present a forecast improvement based on the statistical tests. These results corroborate those of Harmatiz, Miao, and Chien (2006), who conclude that more volatile assets tend to have their best forecast loss in lower horizon estimation windows, while less volatile assets tend to have it in longer horizon windows. Again, CVaR has the highest percentage of VR (78%); however, as mentioned earlier, the method differs from the others because it concentrates on values that exceed VaR. Among the traditional metrics, HS has the highest percentage of VR (20%), which reinforces the hypothesis that the model has a better predictive capacity due to the non-use of the normality premise of the assets. These two models also present the lowest percentage of rejection for the Kupiec test (1995).

Despite these results, is observed that HS and MC have the highest percentage of Christoffersen (1998) rejection. As both correspond to simulation methods, this factor can be indicative of the delay in adjusting these metrics to fluctuations in asset prices. GARCH is the model with the lowest percentage of appropriate VR, which may indicate a worsening of this method due to the reduction of the estimation window, considering that it uses the information contained in the past volatility to forecast losses. For Kupiec test (1995), MC has the highest percentage of rejections. Considering that one of the assumptions of the MC method is the need for a relevant number to perform risk factors simulation, it was expected that the reduction of the estimation window could weaken the model predictive capacity.

In summary, based on the percentage of null hypothesis rejection for both statistical tests, MC method has the weakest performance for all asset classes: 26% for equity indices, 31% for government bonds, and 27% for exchange ratios. CVaR has the best performance for equity indices (12%) and exchange rate (6%), and HS is the best result for the government bond (17%). Among the markets, Chinese assets have the highest average rejection percentage, while Japanese assets have the lowest percentage.

5. CONCLUSION

This paper tests the performance of five VaR methods and differs from prior studies with respect about compare distinct asset categories belonging to different economies. Is also tested the influence of the estimation window horizon on the models' forecast capacity. Therefore, two analyses are made: the first for the entire data period with a 1,000-days estimation window, and the second for subperiods of the data with a 252-days estimation window.

For both analyses, considering the percentage of VR, CVaR is the model that presents the best performance, followed by HS. Both have especial properties; the first consists on a semiparametric model with focus on left tail information for risk forecasting, and the second is a non-parametric model, which estimates the risk factors behavior based directly on the historical observations. The EWMA has the weakest performance in the first analysis. In the second one, MC has not only the weakest performance, but also the highest rejection number for the Kupiec (1995) and the Christoffersen (1998) tests, which

indicates the need for a larger horizon estimate window as expected. It is concluded that a smaller estimation window is better for more volatile assets, while a larger estimation window is better for less volatile assets. Among the markets, the Chinese present the highest average percentage of rejection for the both analysis, the British has the lowest average for the first analysis, while the Japanese has the second one.

The main limitation lies in the data, since indices are used as assets proxies, which generates two fragilities: firstly, because of the stock portfolios composed of different economic sectors, so that, not necessarily a model that performed well for a portfolio will perform well for a stock individually. Secondly, because of commonly the investment strategies consists of diversified portfolios, containing several classes of assets. Therefore, it is suggested that further studies carry out these tests for distinct industrial niches and portfolios composed of more than one asset category.

REFERENCES

Acerbi, C., Nordio, C., & Sirtori, C. (2001). Expected shortfall as a tool for financial risk management. *Working paper*. Retrieved from http://www.gloriamundi.org/var/wps.html.

Adams, Z., & Füss, R. (2009). *VaR performance criterion (VPC): A performance measure for evaluating value-at-risk models*. Maidenhead: McGraw-Hill.

Alexander, C. (2009). *Market risk analysis. Value at risk models* (Vol. 4). West Sussex: John Wiley & Sons.

Angelidis, T., Benos, A., & Degiannakis, S. (2004). The use of GARCH models in VaR estimation. *Statistical Methodology*, *1*(1-2), 105-128.

Barone-Adesi, G., & Giannopoulos, K. (2001). Non parametric VaR techniques. Myths and realities. *Economic Notes*, *30*(2), 167-181.

Bollersev (1994).

Campbell, S. D. (2006). A review of backtesting and backtesting procedures. *Journal of Risk*, *9*(2), 1-17.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, *39*(4), 841-862.

Danielsson, J. (2011). *Financial risk forecasting: The theory and practice of forecasting market risk with implementation in R and Matlab*. West Sussex: John Wiley & Sons.

Danielsson, J., Jorgensen, B. N., Mandira, S., Samorodnitsky, G., & De Vries, C. G. (2005). Subadditivity re-examined: the case for Value-at-Risk. *Discussion paper 549*. Financial Markets Group, London School of Economics and Political Science, London, UK.

Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, *22*(4), 367-381.

Harmantzis, F. C., Miao, L., & Chien, Y. (2006). Empirical study of value-at-risk and expected shortfall models with heavy tails. *The Journal of Risk Finance*, *7*(2), 117-135.

Hong, L. J., Hu, Z., & Liu, G. (2014). Monte Carlo methods for value-at-risk and conditional value-at-risk: a review. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, *24*(4), 22.

Jorion, P. (2003). *Financial risk manager handbook* (Vol. 241). Hoboken (NJ): John Wiley & Sons.

Jorion, P. (2007). *Financial risk manager handbook* (Vol. 406). Hoboken (NJ): John Wiley & Sons.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, *3*(2), 73-84.

Mei, J., & Guo, L. (2004). Political uncertainty, financial crisis and market volatility. *European Financial Management*, *10*(4), 639-657.

Pritsker, M. (2006). The hidden dangers of historical simulation. *Journal of Banking & Finance*, *30*(2), 561-582.

So, M. K., & Philip, L. H. (2006). Empirical analysis of GARCH models in value at risk estimation. *Journal of International Financial Markets, Institutions and Money*, *16*(2), 180-197.

R. Cont. Fin. – USP, São Paulo, v. 31, n. 82, p. 318-331, May/Aug. 2020

331