Risk Management and VaR: Comparison of the accuracy of risk measurement for different assets

Pinheiro, Marilia Cordeiro; Fernandes, Bruno Vinícius Ramos
Risk Management and VaR: Comparison of the accuracy of risk measurement for different assets
Base Revista de Administração e Contabilidade da UNISINOS, vol. 17, no. 4, 2020
Universidade do Vale do Rio dos Sinos, Brasil
Available in: https://www.redalyc.org/articulo.oa?id=337265148007

Risk Management and VaR: Comparison of the accuracy of risk measurement for different assets

Gestão de risco e modelos de VaR: Comparação de poder preditivo para diferentes classes de ativos

Marilia Cordeiro Pinheiro marilia.cordeiro90@hotmail.com
Universidade de Brasília – UNB, Brasil
Bruno Vinícius Ramos Fernandes brunoramos@unb.br
Universidade de Brasília – UNB, Brasil

Abstract: This paper investigates the performance of VaR models for seven categories of assets traded in Brazilian market. Six different VaR methodologies are tested: Normal Delta, EWMA, GARCH, Historical Simulation (HS), Monte Carlo Simulation (MC) and CVaR, which have as main differences the treatment given to volatility and the inference about the returns distribution. For the statistical results validation, are applied the Kupiec test, to evaluate the proportion of violations, and the Christoffersen test, to verify the adjustment speed of the model against market oscillations. Two analyses are made; the first one considerate an estimation window of 1000 days and the second one 252 days. For both, GARCH and CVaR have the highest number of accurately violation ratio (VR) having the good performance validated by backtesting tests. Among the assets, IFIX and IMA-B have the best performance for first analyse and Ibov for the second one. The models have low accurately loss forecast for private bond and commodities indices, which indicates that methodologies focused on market risk are not appropriate for these assets categories. The results also suggest that a smaller estimation window tends to favour the estimation of loss for high volatility assets.

Keywords: VaR, Parametric models, Semi-parametric models, Non-parametric models, Backtesting.

Resumo: O presente artigo analisa a performance do VaR para sete categorias de ativos negociados no mercado brasileiro. Foram testadas seis metodologias diferentes de VaR: Delta normal, EWMA, GARCH, Simulação Histórica (HS), Simulação de Monte Carlo e o CVaR, que têm como principal diferença o tratamento dado para volatilidade e a inferência sobre a distribuição dos retornos. Para validação dos testes estatísticos, foram aplicados os testes de Kupiec, para mensurar a proporção de violações, e o teste de Christoffersen, para verificar a velocidade de ajuste frente às oscilações de mercado. Duas análises foram feitas: A primeira considera uma janela de estimação de mil dias e a segunda de 252 dias. Para ambas, GARCH e CVaR possuíram o número mais alto de índices de violações (IV) precisos, tendo a boa performance validada pelos backtetings. Dentre os aivos, IFIX e IMA-B tiveram melhor performance para primeira anáilise enquanto que Ibov para a segunda. Os modelos tiveram pouca acurácia preditiva para as debêntures e índice de commodities, o que indica que metodologias focadas no risco de mercado não são apropriadas para essa categoria de ativos. Os resultados também sugerem que uma janela de estimação menor tende a favorecer a estimação de perda para ativos de volatilidade elevada.

Palavras-chave: VaR, Modelos paramétricos, Modelos semi.paramétricos, Modelos não paramétricos, Backtesting.

Introduction

The quantitative risk management is a science that uses mathematical, probabilistic and statistical language to forecast, control, eliminate or reduce risk exposure. In the financial market context, this tool is essential for investment management, considering that is an objective method for analysing the consequences of economic oscillations that can generate losses. Generally, risk estimation is done by performance models, complemented by some stress test, which are applied as a way of verifying the accuracy of the applied method. (Crouhy, Galay & Mark, 2006).

One of the most traditional performance models used by risk managers is the Value at Risk (VaR), an econometric tool defined as a value such that there is a probability of exhibiting the maximum loss over the next days, and where and are predetermined by the risk manager (Christoffersen, 2009). The VaR was created at the end of the 80's by J.P. Morgan, being disseminated by the Basel Committee in April 1995 when the organisation established that the capital adequacy framework of financial institutions should be based on the model. The VaR is obtained from the inference of the returns distribution, using the statistical properties of the asset to loss estimation.

Since its inception, several models were created with the purpose of improving its predictive capacity. The main divergence among the VaR methods lies in the inference of return distribution; Parametric models assumes that the density function of risk factors of asset returns must conform to normal distribution; Non- parametric models does not require any statistical assumption beyond stationary of the return distribution, since the normality premise does not reflect the market reality (Barone-Adesi & Giannopoulos, 2000). It will be difficult to reach a consensus on which approach is more appropriate, since financial instrument consist of heterogeneous assets classes, with different fundamentals in prices formation, and consequently, distinct levels of risk exposure.

Despite the VaR widespread diffusion, its methodology has some limitations, generating a debate about the real effectiveness of its adoption. Jorion (2007) presents three of the main VaR models fragilities: (1) VaR is not able to provides the worst loss. (2) VaR is not appropriate to calculate the loss in extreme events. (3) VaR is measured with error, since the values estimated for different time scales, but for the same data, tend to be different. In addition, it is important to highlight that any financial model is a simple representation of the economic world and the way agents manage investments (Gibson, Lhabitant & Talay, 2010). Therefore, is necessary to apply the backtesting which aims to test VaR accuracy based on historical data, making possible to validate good or bad performance of the model.

The Brazilian financial market is characterized by high volatility patterns, with long memory volatility and a slight departure from normality (Beran & Ocker, 2012). These are typical emerging markets characteristics, with unstable economic and financial policies, being

susceptible to internal crises and vulnerable to external crises (Gençay & Selçuk, 2004). Based on this, emerging countries should have robust risk management tools as a means not only of preventing turbulence in their economies, but also of gaining credibility and attracting investors.

The present study aims to test the accuracy of six VaR metrics for distinct asset classes traded in the Brazilian financial market. The main objectives are to analyse whether market risk models can accurately estimate the losses occurring for the data during 1997-2017 and to identify if there is a more appropriate or not model for the domestic market. The sample is composed of seven assets classes: Equity portfolio, government bond portfolio, commodity portfolio, private bond portfolio, real estate portfolio, multimarket fund and the exchange rate. The study is relevant for analysing the predictive capacity of the risk management models and consequently identifying the most efficient form of risk forecasting for each asset in the Brazilian environment.

Based on the percentage of null hypothesis rejection, the results indicate that GARCH and CVaR have the best performing for the data period while Delta and Monte Carlo have the worst performance. It is also observed that estimation windows with greater number of days favour the losses forecast of more volatile assets, whereas smaller estimation windows favour the losses forecast of less volatile assets.

Literature Review

The capital markets volatility induced regulators, managers and academics to develop sophisticated risk measurement tools. With the portfolio diversification theory, volatility (standard deviation) and correlation have become traditional methods of risk measurement, however such statistical concepts are considered limited since they can only capture accurately the risk for returns with a multivariate normal distribution (Alexander, 2009). In response to financial crises that occurred prior to the 1990s, Value at Risk (VaR) emerged, being a mathematical model underlying the theory of risk diversification, but created with a focus on market risk and adverse movements effects on an investment portfolio (Damodaran, 2008).

VaR distinguishes itself from other metrics by its aim to provide a probability statement about possible changes in the portfolio value. It is an aggregate measure of risk across all risk factors, giving a nice representation of investors "risk appetite" considering that it estimates the worst loss at a given level of confidence within a time horizon (Crouhy, Galai & Mark, 2006).

VaR presents three crucial elements in its application; The specific loss value level, the fixed period at which the risk is measured and the confidence interval. Although initially created to market risk measurement, several VaR models have been developed, making it a universal metric, used for a variety of financial and non-financial institutions exposed to risk. (Alexander, 2009). Its advantage lies mainly

in the creation of a common denominator that allows the comparison of different risk activities in a variety of economic markets (Jorion, 2007).

The development of different approaches of VaR calculation is important for investment allocation considering that its traditional model is not sub-additive, which can result in inefficient diversification and hamper the implementation of portfolio optimization algorithms. The lack of subadditivity means that the portfolio risk can be larger than the sum of isolated risk of its components when estimated by VaR (Danielsson et al., 2005). Another important factor is that VaR is limited to a specific horizon and to the established probability level, which makes it not appropriate as the official capital value required to support risk exposure (Duffie & Pan, 1997). New metrics were developed in response to these weaknesses also to improve the maximum loss estimation accuracy. The main difference among VaR models lies in the portfolio returns distribution premise. The parametric approach assumes the assets normality; however non-linearity is a predominant feature in the financial series, making the capacity of adjustment of this methodology to the returns shocks questionable (Füss; Adams & Kaiser, 2009). Based on these failures, the non-parametric approach emerged, with the advantage that no assumption about the returns distribution is required. Models based on Historical Simulation (HS) are the main representatives of this VaR approach, which scenarios are simulated from the empirical returns distribution. In view of the particularities of the distinct statistical approaches, several studies have been developed with the objective of analysing the models performance for different assets classes. Danielson et al. (2005) explores the subadditivity violations for GARCH model for heavy tailed assets. The authors find that for the most of assets there is no sub additivity violation, except for those with a high degree of asymmetry and kurtosis. Zymler, Kuhn and Rustem (2012) test the performance of a VaR approximation for a derivative portfolio in European options market. The authors developed two estimators; one which is suitable for long positions expiring at the end of the investment horizon whereas the second is suitable for portfolios containing long or short position expiring beyond the investment horizon. Dimitrakopoulos, Kavussanos and Spyrou (2010) investigates the performance of VaR models and Extreme Value Theory (EVT) for emerging and developed market equity portfolios. The results indicate that despite the differences among the economies, the most successful VaR methods are common for both markets. Additionally, for emerging equity portfolios, most VaR models turn out to yield conservative risk forecasts the opposite of developed markets, where models underestimate the realised VaR.

Methodology

To forecast the VaR and compare the accuracy of each metric among the assets, six different methods are tested; Delta normal, Exponentially Weighted Moving-average (EWMA), Generalized Autoregressive Conditional Heteroskedastic (GARCH) and Monte Carlo Simulation

PDF generated from XML JATS4R by Redalyc
Project academic non-profit, developed under the open access initiative
667

(MC) representing the parametric approach; Historical Simulation (HS) representing the non-parametric approach and Conditional Value at Risk (CVaR) representing the semi-parametric approach.

Delta normal

The Delta Normal is one of the simplest calculation forms, being the starting point for the development of other forecast models, since it is necessary to understand the advantages and disadvantages of linear delta to understand these same factors in more complex models (Choudhry, 2013).

The model is based on the presumption that assets returns are i.i.d., which assumes that the portfolio value is a linear function of it risk factors. The correlation represents the dependencies among the risk factors, which have their variances grouped with the intention of forming the daily covariance matrix, used to provide the probability of portfolio distribution (Alexander, 2008). The delta normal VaR is estimated by the normal standard deviation corresponding to the confidence interval :

$$VaR = \alpha\sigma(R_{p,t+1})$$

(1)

Exponentially weighted moving-average (EWMA)

The EWMA consists on an improvement of moving average methods especially due to the advantage of putting more weight in the most recent data, considering that it has the most relevant information about the returns behaviour. As the window moves forward, the extreme returns will be further apart, having less significant weight in the risk estimation. This fact allows that the model reacts faster to market shocks, since the volatility decreases exponentially as the shock observation moves away from the present (Longerstaey, 1996; Alexander, 2008).

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

$$= \sqrt{(1-\lambda)\sum_{s=t=k}^{t=1} \lambda^{(t=s=1)}r_s^2} = \sqrt{\lambda\sigma_{t=1}^2 + (1-\lambda)r_{t=1}^2}$$

(2)

Where λ denotes the decay factor, which demonstrates how the influence of past observations while estimating σ_{t+1}. Empirical studies show that λ =0.94 allows a nice risk forecasting for market assets. The EWMA assumes the normal distribution of returns. The estimation of the EWMA VaR of the 100% h-day is:

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

GARCH

The GARCH model, derived from the ARCH, is taken as a complete risk metric in the aspect of considering the property that the conditional volatility σ is a function of continuous change of the squares of its previous values, which generates the volatility clusters. The model it is auto regressive, since the value of the return depends Y_t on the values of Y_{t-1}, which suggests the heterosledasticity observed over different periods can be autocorrelated (Alexander, 2008). The GARCH can capture a range of financial series properties by taking three of their main characteristics: heavy tails, volatility clusters and nonlinear dependence. To obtain the VaR from the GARCH method, initially, it's necessary to model the conditional volatility:

$$\sigma_t^2 = \omega + \sum_{i=1}^{L_1} \alpha_i\ Y_{t-i}^2 + \sum_{j=1}^{L_2} \beta_j \sigma_{t-j}^2$$

(4)

Here, L represents the number of lags while $\omega, \alpha_i, \beta_j$ are the parameters equation, estimated by the maximum likelihood. The model also assumes the returns normality.

$$Z_t \sim N(0,1)$$

(5)

Therefore, the assets returns are obtained by $Y_t \sim N(0, \sigma_t^2)$. This model is known as normal GARCH. After the volatility estimation by GARCH (1,1), VaR is calculated by the product of the estimated conditional volatility and the percentile of the normal distribution, according to the following expression:

$$VaR = Z_p \times \sigma_t$$

(6)

Monte Carlo

The Monte Carlo VaR estimation is divided into four stages: Firstly, and the most important, consists in the choice of the stochastic process and the parameters. The Brownian motion is commonly used because it

assumes that price-related innovations of the assets are uncorrelated over time and that small movements can be expressed by:

$$dS_t = \mu_t S_t dt + \sigma_t S_t dz$$

(7)

Where *dz* is a random variable with normal distribution, zero mean and *dt* variance. Integrating *dS/s* over a finite interval of time, we have approximately:

$$\Delta S_t = S_{t-1}(\mu \Delta t + \sigma \epsilon \sqrt{\Delta t}) =$$

(8)

To simulate the fluctuation of the *S* price, starting from S_{t+1} is generated a sequence of the random variable ϵ's for $i = 1,2,3 \dots n$. So S_{t+1} defined as $S_{t+1} = S_t + S_t(\mu \Delta + \sigma \epsilon_1 \sqrt{\Delta t}}$. This process is repeated until the target horizon is reached.

For VaR forecast, portfolio value is *Ft + n = FT* calculated under the sequence prices *St* within the target horizon. This process results on the $F_T^1, \dots, F_T^B$. The data are ordered and tabulated to obtain the expected return $E(F_t)$ and the quantiles $Q(F_t, c)$ which correspond to the surplus value in times for replications. VaR forecast is obtained by:

$$VaR(c, T) = E(F_t) - Q(F_{t,}c)$$

(9)

Historical Simulation

The VaR obtained by the historical simulation is estimated from the construction of hypothetical values from a current observation given by:

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

(10)

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

$$P^k = P[f_1^k, f_2^k \ldots f_N^k]$$

(11)

The variations of portfolios prices $P^k = (P^k - P_t)/P_t$ are obtained with the equations above. The returns are ordered and then are chosen those that correspond to ####### ### ## (#). The VAR is obtained by the difference between the mean and the #######:

$$VaR = AVE[R_p] - R_p(c)$$

(12)

Conditional Value at Risk

The last model tested is the CVaR. It is as a differentiated method, since it aims to estimate the loss that exceeds the forecast estimation of VaR. To identify the correct distribution, is applied:

$$p = \int_{-\infty}^{-VaR(p)} yf(y)dy$$

(13)

The tail density of $f_{VaR}(\cdot)$ is given by:

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

(14)

The CVaR is estimated by the negative value of the profit/loss ratio on the tail density of $f_{VaR}(\cdot)$:

$$CVaR = \int_{-\infty}^{-VaR(p)} f_{VaR}(y)dy$$

Statistical tests

As presented by the Basel Committee (2010), the VaR backtesting is applied with the purpose to test the number of violations that go beyond the 99th percentile. If this number is statistically significant, the hypothesis that the estimation model is adequate can be rejected.

Danielsson (2011) divides the methodology of backtesting into three procedures; Initially, the estimation and the test windows are defined. The first one represents the number of observations used to loss forecast, and the second on consists of the part of the sample after loss, i.e. the day on which VaR is calculated. The test window moves each day by adding *# + 1* and removing *#-#*. In this study, two estimation windows are tested; the first with n = 1000 days and the second with n = 250 days, with the objective to analysing if the horizon of the estimation window impacts on the model performance.

In sequence, the violation ratio (VR) is obtained which aims to 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 $n_t = 0$ otherwise. The number of violations it's incorporated on the variable v_1, while v_0 correspond to the number without violations.

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

$$v_1 = \sum n_t$$

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

(16)

The violation ratio is:

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

(17)

Using the general rule, 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, is applied the Kuppiec test (1995) and the Christoffersen (1998). The first test is used to analyses the statistical

significance of the violations proportion. The null hypothesis for VaR violation is:

$$H_0: \eta \sim B(p)$$

(18)

With *B* representing the Bernoulli distribution. The Bernoulli density is obtained by:

$$\hat{p} = \frac{v_1}{W_t}$$

(19)

Under $_{H_0, p = \hat{p}}$ so the restrict maximum likelihood function is:

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

(20)

The Christoffersen test is the second one applied. The test has as advantage identifying whether violations cluster, considering that, theoretically, they should be independent. If the null hypothesis is rejected, is an indication that the model delays in absorbing the oscillations that occur in the market for the asset tested. It's needed to calculate the probabilities of two consecutives violations and the probability of a violation if there was no violation on the previous day:

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

(21)

The statistical test is given by:

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

(22)

Where $_{\Pi_0}$ is the estimated transition matrix and $_{\Pi_1}$ 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; so $_{p_{01} = p_{11} = p}$. The test of independence is asymptotically distributed as a $_{\chi^2}$.

The CVaR backtesting differs from the other models since what is being tested is a loss beyond VaR. Danielsson (2001) presents a methodology to backtesting CVaR that is analogous of the violation ratio. When VaR is violated, normalized shortfall *NS* is calculated as:

$$NS_t = \frac{y_t}{ES_t}$$

(23)

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

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

(24)

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

$$H_0 = \overline{NS} = 1$$

(25)

Data

The dataset correspond to daily logarithmic returns of seven assets classes: Stocks, private bonds, government bonds, exchange rate, commodities, real estate fund and multimarket investment fund. The data covers the time period from January 1997 until January 2017, totalling 20 years. For assets that started trading after January 1997, the first trading day is considered as the beginning of the sample period. The criteria used to compose the sample are those of liquidity and representativeness in the Brazilian financial market. Table 1 summarises the data composition:

Table 1.

Data description

Asset	Period	Description	Source
Ibovespa	1997-2017	Theoretical equity portfolio with stocks that accounted 80% of the volume traded in the last 12 months.	Economatica
Exchange Rate	1997-2017	USD to BRL Exchange Rate	Economatica
ANBIMA Bond Index (IDA)	2009-2016	Theoretical bond portfolio at market price, used as benchmark in the segment.	Economatica
ANBIMA Market Index	1997-2016	Theoretical government bonds portfolio, divided into four sub-indices: Fixed yield, indexed to IPCA and IGP-M and indexed to SELIC	Economatica
Commodities Index	1997-2017	Theoretical commodities portfolio at market price	Reuters
Asset	Period	Description	Source
Real estate fund Index (IFIX)	2011-2017	Theoretical portfolio of shares of exchange-traded and OTC-traded real estate funds authorized for trading by B3.	Economatica
LP 200	1997-2017	Multimarket fund of shares of investment funds composed vy government bonds, private bonds, derivatives and stocks.	Economatica

Source: Authors' own elaboration (2018).

Empirical Results

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

Descriptive Statistic

According to Hair et al. (2010), Cronbach's Alpha (α) coefficient is the most used measure to assess the reliability of collection instruments used in scientific researches. Therefore, such measure was chosen to assess the consistency of the scales used in the research questionnaire.

The coefficient α calculated for the collection instrument was 0.601. Maroco and Garcia-Marques (2006) affirm that an average coefficient α of 0.60 can be acceptable in scientific researches. Hair et al. (2010) also

suggest that the minimum acceptable coefficient α is 0.60. Therefore, the collection instrument was reliable pursuant to the minimum acceptable levels of reliability.

Descriptive statistics

The descriptive statistics provide an insight into the investment features. It must be emphasised that the comparison among the statistical results of assets are limited, since IFIX and IDA had their beginning trading after the other assets. Table 2 presents the descriptive statistics for the data:

Table 2.
Data statistical results

	Ibov (1997-2017)	Dollar (1997-2017)	IDA (2009-2017)	IMA-B (2006-2017)	ICB (2004-2017)	IFIX (2011-2017)	LP 200 (1998-2017)
Observations	4970	5039	2175	2915	3237	1505	4786
Minimum	-17.21%	-9.36%	-1.35%	-3%	-7%	-4.98%	-1.87%
Quartile 1	-1.02%	0.42%	0,03%	0.2%	0.13%	0.15%	-0.94%
Median	0.09%	0.02%	0,05%	0.05%	0%	0.04%	0.05%
Annual Average	0.04%	0.02%	0.05%	0.05%	0.03%	0.04%	0.04%
Quartile 3	1.17%	0.41%	0.07%	0.13%	0.52%	0.26%	0.10%
Maximum	28.83%	10.53%	0.84%	0.14%	8.65%	1.77%	1.78%
Annual Standard-deviation	2.08%	0.97%	0.10%	0.18%	0.94%	0.42%	0.17%
Skewness	0.33	0.34	-1.72	-1.51	0.46	-1.28	-1.10
Kurtosis	12.93	14.76	28.32	27.41	7.44	15.89	21.30
JB- Value	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Source: Authors' own elaboration (2018).

Initially, it is observed that all assets have positive average returns for the period; IDA and IMA-B show the highest returns, with approximately the same value (0.05%), while the Exchange Rate has the lowest (0.02%) which may have been occasioned by the strong asset devaluation during the years of 2008 and 2009. In relation to volatility, represented by the standard deviation, the Ibov is the most volatile, followed by the Exchange Rate. The least volatile assets are IDA (0.10%) and LP 200 (0.17%).

Analysing the maximum values of the financial series, it should be noted that the Ibov and the dollar reached their peak on the same day, 15/01/1999. In fact, this factor did not occur in reason of an economic boom, but due to a sharp drop in the price of the two assets on the previous day generated by the BACEN exchange rate policy in which the maximum dollar price was established, leading to the country's foreign capital flight. LP 200 and IDA have the lowest values; however, because they have a lower volatility, it is in line with expectations. When the minimum points are analysed, the Ibov has the lowest value in the second half of 1999 followed by the Exchange Rate in the second half of 2002.

Based on the p-values of JB test, the null hypothesis of normality is rejected at any significance level for all markets, which is a violation for

the parametric models main premise. In addition, all assets have positive kurtosis, a characteristic of leptokurtic returns distribution with fat tails and exposed to extreme events. Ibov, Exchange Rate and ICB present positive skewness while the other assets, negative.

Analysis for the entire period

Firstly, is applied the test for the whole period (1997-2017), with a 1000 days estimation window. As mentioned previously, it is considered that a VR provides a good forecast. Table 3 shows the VRs values and their respective statistical significance for Kupiec and Christoffersen tests.

Table 3.
Backtesting for 99% VaR estimation (1997-2017).

Assets	Delta	EWMA	HS	GARCH	CVaR	MC
Ibov	1.36***	1.76*	0.98**	1.31	1.41	0.71**
Exchange Rate	1.56***	1.14	1.31**	1.06	1.61**	1.58***
IDA	4.08*	1.37	4.25***	0.85	4.17***	1.49*
IMA-B	1.20*	1.56***	1.15**	0.99*	1.62	1.60***
ICB	0.45**	1.43	0.45**	0.71	0.49	1.23***
IFIX	0.4	1.19	0.4	0.8	0.59	0.8
LP 200	3.38***	1.74*	2.29***	1.40*	3.38	1.20***

Notes: * significant at 5% level for the Kupiec test ; ** significant at 5% level for the Christoffersen test; *** significant at 5% level for the Kupiec and the Christoffersen test

with the best estimation results are the Exchange Rate, IFIX and IMA-B, whereas the ICB has the worst performance, presenting no adequate VRs. Among the models, the GARCH has the best performance (4), while the Normal Delta and MC have no appropriate VRs.

HS is the model with the best performance for Ibov, which may be a consequence of the asset high volatility, considering that the method does not make any assumption about the distribution, being able to incorporate the non-linearity of returns. The EWMA and GARCH have the best performance for the Exchange Rate. Both models consider volatility clusters to risk forecast, which is a feature of the exchange rates, considering that, according to previous studies, financial foreign market volatility occurs in waves, especially because it is an asset exposed to external crises (Kearney & Patton, 2000; Baillie & Bollerslev, 2012). The IMA-B presents three appropriate VRs; EWMA, GARCH and HS. Diebold and Yilmaz (2012) point out that, since the government bonds depend mainly on the yield curve, these assets tend to have auto correlated volatility during economic shocks, which explains the performance of EWMA and GARCH.

The GARCH is the only model that presents adequate VR for IDA. One possible explanation for the poor performance of the models tested for IDA is that credit risk is the main risk of a private bond, so that the models tested in the present work, which estimate the market risk, have not been able to estimate the exposure to default. The models also do not

perform satisfactorily for ICB, which has the worst performance. This result might be caused by the distinct characteristics that commodities have when compared to other traditional assets. Commodity prices are influenced by factors such as demand and supply shocks, as well as natural disasters, which suggest that the VaR forecast models applied in this paper are not appropriate to capture these risks. Lastly, only CVaR obtained adequate VR for the LP 200. The model is characterized by being more conservative, and, considering that the LP 200 contains greater risk assets in its composition, such as the derivatives, it is possible that the CVaR can estimate higher losses than the other assets.

Based on the statistical tests, is verified that the total number of rejection for Kupiec test is 19, which indicates that 40% of the models do not forecast the VaR accurately. The number of rejection corroborates with the VRs results; GARCH and CVaR have the lowest rejection number (2), while Delta and MC have the largest one (6).; IDA has the highest rejection number (5), whereas IFIX has no rejection. For the Christoffersen test, the number of rejections is also equal to 19. This means that 40% of the models have a delay in absorbing market oscillations for some data asset. The HS presents the highest number of rejections for the test (6) and that GARCH model the lowest one (0). For assets perspective, IMA-B and exchange rate presents present the highest number of rejection (6) and IFIX has no rejection.

Analysis for sub-periods

Next, is tested a smaller window with the objective to analyse the predictive power of VaR methods over distinct economic cycles and to compare the impact of the reduction of the estimation window horizon on the models accuracy. The estimation window is reduced for 252 days, equivalent to one-year trading. The sample is divided into groups of three years, from 2005 until 2016. IFIX, IDA and IMA-B have a smaller number analysis than the other assets due to the start of their trading on the financial market. Table 4 and Table 5 present the percentage of appropriate VR and the percentage of rejection for Kuppiec and Christoffersen tests at 5% significance level for each asset and for each model, respectively.

Table 4.

Sub-periods backtesting for 99% VaR per asset

Asset	Percentage of adequate VR	Percentage of Kupiec rejections	Percentage of Christoffersen rejections	Period with highest number of Kupiec rejections	Period with highest number of Christoffersen rejections
Ibov	38%	25%	0%	2005-2007	-
Exchange Rate	25%	13%	8%	2005-2007	2005-2007
IDA	25%	17%	8%	2011-2013	2011-2013
IMA-B	33%	21%	8%	2009-2011	2006-2008
ICB	21%	13%	4%	2008-2010	2008-2010
IFIX	33%	8%	8%	2011-2013	2011-2013
LP 200	13%	17%	13%	2005-2007	2011-2013

Source: Autor's own elaboration (2019)

Table 5.

Sub-periods backtesting for 99% VaR per model

Asset	Percentage of adequate VR	Percentage of Kupiec rejections	Percentage of Christoffersen rejections	Period with highest number of Kupiec rejections	Period with highest number of Christoffersen rejections
Delta Normal	17%	29%	4%	2005-2007	2005-2007 2011-2013
EWMA	21%	17%	8%	2005-2007	2011-2013
HS	21%	13%	8%	2008-2010 2011-2013	2008-2010
GARCH	29%	4%	0%	2011-2013 2014-2016	2011-2013
CVaR	63%	4%	8%	2011-2013	2005-2007
MC	17%	33%	13%	2005-2007 2011-2012	2011-2013

Source: Author own's elaboration (2019)

Based on VRs values, the results of this second analysis are similar to the first one. A total of 144 VRs are estimated, of which 26% are accurate and 74% inaccurate. Ibov presents the highest percentage of VRs (38%), while LP 200 has the lowest percentage (13%). It is also observed the improvement of the loss forecast for ICB, which present a greater number of adequate VRs in relation to the larger estimation window. The results corroborate those found in previous studies, in which larger size windows tend to favour the loss forecast for high volatility assets, while those with smaller size favour more stable assets (Hendrick, 1996; Harmantzis, Miao & Chien, 2006; Dimitrakopoulos, Kavussanos & Spyrou, 2010). Larger estimation windows contain extreme events which overestimate VaR forecasting. This factor occurs due to the estimation that is carried during subsequent periods, regardless of their relevance to the asset's future behaviour.

For the Kupiec test, the period with the highest rejection number is 2005-2007, which is expected, since in 2007-2008 the financial crisis occurred, which increase the question about the use of VaR for risk estimation for extreme events. Ibov is the asset with the highest percentage of rejections, with 70% of these being concentrated in the financial crisis period. HS and CVaR are the only models that did not have H0 refuted for this period. Again, IFIX has the lowest rejection percentage, presenting only one for the EWMA model in the period 2011-2013. Comparing the models, as occurred in the first analysis, Delta Normal and MC have the worst performance for Kupiec test. However, it is important to highlight that the reduction of the estimation window decreases the rejection percentage for both models. GARCH and CVaR have the best performance, but an observation must be made; CVaR is a differentiated model, considering that it estimates the loss beyond the VaR, being more conservative than the other methods.

Considering the Christoffersen test, the period of 2011-2013 has the highest rejection number. This factor may be a consequence of the Brazilian economic recession, which began in 2012 (IPEA, 2015), indicating a delay to absorb the market changes for this context. Different from the Kupiec test, Ibov has the lowest rejection percentage for the Christoffersen. So, although the violations proportion for this asset is significant, is not found evidence of dependence among them. LP 200 has the highest rejection percentage, with all of them concentrated in 2011-2013, which also may be a consequence of the Brazilian recession, since it is a multimarket fund. MC has the worst performance for the test, whereas GARCH has the best.

In summary, considering both estimation windows analysis and the statistical tests, GARCH has the best performance in the Brazilian market, being the best model for the exchange rate, IMA-B an LP 200. For Ibov, the non-parametric model (HS) and semi-parametric (CVaR) model have the best performance, which can be consequence of the high volatility and non-linearity of the asset. Delta normal and MC have the weakest performance; the first is the simplest metric of the data and the second is limited to the quality of the VaR applied, being necessary the inference about the stochastic process. Based on the reasonable VRs interval and considering the inaccurately values, CVaR and HS are the only models which overestimated VaR while the other underestimated. Finally, the results also suggest that the performance of VaR models during crises deteriorates, except for the CVaR, which is a differentiated metric since it concentrate on the percentile above the distribution.

Conclusion

This paper tests the performance of six VaR methods during a long data period, which includes crises and post-crises. It differs from prior studies with respect about compare the accuracy of the models among distinct asset categories. 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 1000-days estimation window and the second for sub-periods of the data with a 252-days estimation window.

For both analyses, considering the percentage of VR, GARCH is the model that presents the best performance, followed by CVaR. Both have especial properties; the first consists on an auto-regressive model which considers that assets present volatility clusters and the second model is semi-parametric, with focus on left tail information for risk forecasting. Delta normal and MC have the weakest performance for both analyses. Among the assets, for the first estimation window, IFIX has the highest number of VRs and the lowest number of rejection for the statistical tests, while for the second one, Ibov has the highest number of VRs, but also presented the weakest performance for the Kupiec test which is probably a consequence of the subprime crisis, since most of the rejections are concentrated in this period. In addition, the models tested in this study show a weak performance for IDA, although the VRs number and the statistical tests have improved in the second analysis. This factor indicates that methods for market risk are not the most appropriate for forecasting losses of private bonds.

The main limitation lies in the data, considering that market indices are used as proxies of assets, which generates two fragilities; firstly, because of the stock portfolios composed of different economic sectors, so that, not necessarily a model that estimated accurately for a portfolio will have the same performance for a stock individually. Secondly, because commonly the investment strategies consist 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

Alexander, C. (2008). *Market risk analysis, practical financial econometrics* (Vol. 2). John Wiley & Sons.

Alexander, C. (2009). *Market risk analysis, value at risk models* (Vol. 4). John Wiley & Sons.

Baillie, R. T., & Bollerslev, T. (2002). The message in daily exchange rates: a conditional-variance tale. *Journal of Business & Economic Statistics, 20*(1), 60-68. https://doi.org/10.1198/073500102753410390

Barone - Adesi, G., & Giannopoulos, K. (2001). Non parametric var techniques. myths and realities. *Economic Notes, 30*(2), 167-181. https://doi.org/10.1111/j.0391-5026.2001.00052.x

Basel Committee on Banking Supervision (2010). *Sound practices for backtesing counterparty credit risk models*. Bank for International Settlements.

Beran, J., & Ocker, D. (2001). Volatility of stock-market indexes—an analysis based on SEMIFAR models. *Journal of Business & Economic Statistics, 19*(1), 103-116. https://doi.org/10.1198/07350010152472661

Choudhry, M. (2013). *An introduction to value-at-risk.* John Wiley & Sons.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International economic review*, 841-862. https://doi.org/10.2307/2527341

Christoffersen, P. (2009). Value–at–risk models. In Handbook of Financial Time Series (pp. 753-766). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-71297-8_33

Crouhy, M., Galai, D., & Mark, R. (2006). *The essentials of risk management* (Vol. 1). McGraw-Hill.

Damodaran, A. (2007). *Strategic risk taking: a framework for risk management.* Pearson Prentice Hall.

Danielsson, J., Jorgensen, B. N., Mandira, S., Samorodnitsky, G., & De Vries, C. G. (2005). *Subadditivity re-examined: the case for Value-at-Risk.* Cornell University Operations Research and Industrial Engineering.

Danielsson, J. (2011). *Financial risk forecasting: the theory and practice of forecasting market risk with implementation in R and Matlab* (Vol. 588). John Wiley & Sons.

Diebold, F. X., & Yilmaz, K. (2012). Better to give than to receive: Predictive directional measurement of volatility spillovers. *International Journal of Forecasting, 28*(1), 57-66. https://doi.org/10.1016/j.ijforecast.2011.02.006

Dimitrakopoulos, D. N., Kavussanos, M. G., & Spyrou, S. I. (2010). Value at risk models for volatile emerging markets equity portfolios. *The Quarterly Review of Economics and Finance, 50*(4), 515-526. https://doi.org/10.1016/j.qref.2010.06.006

Duffie, D., & Pan, J. (1997). An overview of value at risk. *Journal of derivatives, 4*(3), 7-49. https://doi.org/10.3905/jod.1997.407971

Füss, R., Adams, Z., & Kaiser, D. G. (2010). The predictive power of value-at-risk models in commodity futures markets. *Journal of Asset Management, 11*(4), 261-285. https://doi.org/10.1057/jam.2009.21

Gençay, R., & Selçuk, F. (2004). Extreme value theory and Value-at-Risk: Relative performance in emerging markets. *International Journal of forecasting, 20*(2), 287-303. https://doi.org/10.1016/j.ijforecast.2003.09.005

Gibson, R., Lhabitant, F. S., & Talay, D. (2010). *Modeling the term structure of interest rates: a review of the literature.* Now Publishers Inc. https://doi.org/10.1561/0500000032

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2010). *Multivariate Data Analysis.* Seventh Edition. Prentice Hall.

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. https://doi.org/10.1108/15265940610648571

Hendricks, D. (1996). Evaluation of value-at-risk models using historical data. *Economic policy review, 2*(1). https://doi.org/10.2139/ssrn.1028807

IPEA - Instituto de Pesquisa Econômica Aplicada (2015). *Carta de Conjuntura n. 29.* Retrieved from: http://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=26918&Itemid=3

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

Kearney, C., & Patton, A. J. (2000). Multivariate GARCH modeling of exchange rate volatility transmission in the European monetary system.

Financial Review, 35(1), 29-48. https://doi.org/10.1111/j.1540-6288.20 00.tb01405.x

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *The J. of Derivatives, 3*(2). https://doi.org/10.3905/jod.1995.40 7942

Longerstaey, J., & Spencer, M. (1996). Riskmetricstm—technical document. *Morgan Guaranty Trust Company of New York: New York, 51*, 54.

Maroco, J., & Garcia-Marques, T. (2006). Qual a fiabilidade do alfa de Cronbach? Questões antigas e soluções modernas?. *Laboratório de psicologia*, 65-90.

Zymler, S., Kuhn, D., & Rustem, B. (2013). Worst-case value at risk of nonlinear portfolios. *Management Science, 59*(1), 172-188. https://doi.org/10.1287 /mnsc.1120.1615