“Value-at-risk (VAR) estimation and backtesting during COVID-19: Empirical analysis based on BRICS and US stock markets”

AUTHORS	Muneer Shaik Lakshmi Padmakumari

ARTICLE INFO	Muneer Shaik and Lakshmi Padmakumari (2022). Value-at-risk (VAR) estimation and backtesting during COVID-19: Empirical analysis based on BRICS and US stock markets. *Investment Management and Financial Innovations*, *19*(1), 51-63. doi:10.21511/imfi.19(1).2022.04
DOI	http://dx.doi.org/10.21511/imfi.19(1).2022.04
RELEASED ON	Monday, 31 January 2022
RECEIVED ON	Thursday, 11 November 2021
ACCEPTED ON	Monday, 24 January 2022

JOURNAL	"Investment Management and Financial Innovations"
ISSN PRINT	1810-4967
ISSN ONLINE	1812-9358
PUBLISHER	LLC “Consulting Publishing Company “Business Perspectives”
FOUNDER	LLC “Consulting Publishing Company “Business Perspectives”

NUMBER OF REFERENCES	NUMBER OF FIGURES	NUMBER OF TABLES
34	3	3

BUSINESS PERSPECTIVES

LLC "CPC "Business Perspectives"
Hryhorii Skovoroda lane, 10,
Sumy, 40022, Ukraine

www.businessperspectives.org

Muneer Shaik (India), **Lakshmi Padmakumari** (India)

VALUE-AT-RISK (VaR) ESTIMATION AND BACKTESTING DURING COVID-19: EMPIRICAL ANALYSIS BASED ON BRICS AND US STOCK MARKETS

Abstract

Value-at-risk (VaR) is the most common and widely used risk measure that enterprises, particularly major banking corporations and investment bank firms employ in their risk mitigation processes. The purpose of this study is to investigate the value-at-risk (VaR) estimation models and their predictive performance by applying a series of backtesting methods on BRICS (Brazil, Russia, India, China, South Africa) and US stock market indices. The study employs three different VaR estimation models, namely normal (N), historical (HS), exponential weighted moving average (EMWA) procedures, and eight backtesting models. The empirical analysis is conducted during three different periods: overall period (2006–2021), global financial crisis (GFC) period (2008–2009), and COVID-19 period (2020–2021). The results show that the EMWA model performs better compared to N and HS estimation models for all the six stock market indices during overall and crisis sample periods. The results found that VaR models perform poorly during crisis periods like GFC and COVID-19 compared to the overall sample period. Furthermore, the study result shows that the predictive accuracy of VaR methods is weak during the COVID-19 era when compared to the GFC period.

Keywords value-at-risk (VaR), estimation, backtesting, global financial crisis (GFC), COVID-19, BRICS stock markets

JEL Classification C30, C53, C58, G10

Received on: 11th of November, 2021
Accepted on: 24th of January, 2022
Published on: 31st of January, 2022

Muneer Shaik, Assistant Professor, School of Management, Mahindra University, India. (Corresponding author)

Lakshmi Padmakumari, Assistant Professor, IFMR Graduate School of Business, Krea University, India.

Conflict of interest statement:
Author(s) reported no conflict of interest

INTRODUCTION

Financial markets have become increasingly global and sophisticated in the current stage of the world economy. The stock market has never been more pessimistic than it has been in response to the COVID-19 outbreak. The COVID-19 disaster has had grave undesirable consequences on a global scale, hurting multiple economies and deteriorating their conditions, perhaps leading to a catastrophic recession akin to that seen during the Global Recession of 2008–2009.

The goal of this paper is to assess and backtest market risk during times of crisis. First, the study computes three popular VaR estimation techniques (namely, historical simulation model (HS), normal distribution method (N), and exponential weighted moving average (EWMA) model) and compares their predictive performance using eight different backtesting approaches (namely, Traffic Light (TL), Binomial (Bin), proportion of failure (POF), time till first failure (TUFF), conditional coverage independence (CCI), conditional coverage (CC), time between failures independence (TBFI), and time between failures (TBF) tests). Second, the study aims to investigate how well the VaR estimation models perform during crisis periods like the global financial crisis (2008–2009) and COVID-19 crisis (2020–2021) compared to the

overall sample period (2006–2021) of this study. Third, the study conducts empirical analysis on emerging BRICS nations and the developed US stock market indices to understand the outcomes of VaR estimation models on different market economies.

1. LITERATURE REVIEW

The value-at-risk (VaR) has been endorsed widely in measuring market risks (Jorion, 2011). Multiple research has been undertaken on VaR in emerging Southeast Asian countries (Cheong et al., 2011), the European Union nations (Iglesias, 2015), the Latin American countries (Ozun & Cifter, 2007), Nordic markets (Jobayed, 2017), the South African market (Mabitsela et al., 2015; Naradh et al., 2021), and others. For the period 2000–2012, Iglesias (2015) researched VaR based on major equities asset market indices of 11 EU nations. Glosten et al. (1993) used a GJR-GARCH model and estimated the returns of those 11 market indices. Ozun and Cifter (2007) associated the dynamic conditional symmetrized Copula to the EWMA method on the Latin American stock market. Subbiah and Fabozzi (2016) examined the prediction effectiveness of VaR methods for Asian nations. Choi and Min (2011) analyzed the effectiveness of the implicit and the unconditional methods, utilizing multiple VaR techniques.

There is a dearth of research on VaR estimation and predictive performance based on BRICS countries, which are the world's five most attractive and strongest emerging markets (Mukta & Muneer, 2020a). Aside from economic growth, the BRICS stock indices provide greater and more reliable market proceeds (Jiang et al., 2019; Mukta & Muneer, 2020b). Muteba Mwamba and Beytell (2015) deployed a risk model based on the values of the BRICS asset markets to provide point estimations for VaR and also the Expected Shortfall (ES). Bonga-Bonga and Nleya (2016) assessed portfolio market risk for BRICS nations using multivariate GARCH models. Song et al. (2019) employed two novel multivariate copulas to compute and relate financial risks for three groupings, namely the BRICS, G7, and G20 nations, using VaR and projected shortfall metrics. It was discovered that the BRICS have the highest danger, whereas the G20 have the lowest risk of the three groupings. Desheng and Chatpailin (2019) investigated historical and delta normal VaR estimation models for BRICS nations and performed backtesting using Kupiec's POF test and Christoffersen's test. Burdorf and van Vuuren (2018) examined stock portfolios from the banking and retail sectors in developed (UK) and emerging (South Africa) markets, finding that industries and periods affected risk measure accuracy, but not economies.

The present study fills the gap and contributes to the existing literature by analyzing the three popular VaR estimation models along with a comprehensive eight backtesting methods to understand the predictive performance of BRICS nations along with developed US stock market indices.

Degiannakis et al. (2012) looked at the performance of three different VaR models to come up with estimates that might be used to measure and anticipate market risk. During times of financial crisis, they discovered evidence that generally acknowledged methodologies provide solid VaR estimates and projections. Miletic and Miletic (2015) study the performance of value-at-risk (VaR) models in selected Central and Eastern European (CEE) developing capital markets during the global financial crisis. In comparison to GARCH-type models with normal distribution, historical simulations, and RiskMetrics approaches, backtesting analysis for the crisis shows that GARCH-type models with t-distribution of residuals produce higher VaR estimates. At a 95 percent confidence level, Su et al. (2010) found that the Historical Simulation VaR estimate model significantly outperforms the GARCH (1,1) – AR(1) model. For the nations of the United States, the United Kingdom, France, Germany, Italy, Japan, China, Spain, and Portugal, Ramalho (2020) estimate VaR using Historical Simulation, GARCH(1,1), and Dynamic EVT-POT with a time horizon of January 1, 2007, to August 31, 2020. It was discovered that as the number of deaths grew throughout the COVID-19 era, so did the volatility in these markets; the bulk of exceedances occur during crisis moments rather than normal ones. Using conditional extreme value theory, Omari et al. (2020) predicted the VaR of financial markets throughout the

http://dx.doi.org/10.21511/imfi.19(1).2022.04

COVID-19 timeframe. This study contributes to the existing literature by comparing the predictive performance of VaR during two major crisis periods GFC (2008–2009) and the COVID-19 (2020–2021) period along with the overall sample period (2006–2021) for emerging BRICS and the developed US stock markets.

2. METHODOLOGY

2.1. VaR estimation using normal distribution

This method presumes that the returns are regularly distributed when using the normal distribution approach. The normal distribution approach has the virtue of being simple and is also known as parametric VaR.

2.2. VaR estimation using historical simulation (H) method

The H approach, unlike the normal distribution method, is nonparametric. It makes no assumptions about the distribution of asset returns. The historical simulation calculates risk by presuming that the allocation of profits and losses in the prior period of returns will be used as the allocation of profits and losses in the subsequent period of returns. The VaR "today" is calculated by taking pth-quantile of previous results. The profile of the historical simulation curve is piecewise constant. This one is since quantiles need not alter for many days until important incidents happen. As a result, the H technique reacts slowly to increases in volatility.

2.3. VaR estimation using the exponential weighted moving average (EWMA) method

Prior algorithms assume that all previous returns have equal weight. EWMA approach allocates weights that are not equal, especially weights that decrease exponentially. The most recent returns have larger weights since they have a greater effect on "today's" return compared to returns from further back in time. The EWMA variance across an estimating window of size W_E is calculated by:

$$\hat{\sigma}_t^2 = \frac{1}{c}\sum_{i=1}^{W_E}\lambda^{i-1}y_{t-i}^2, \quad (1)$$

where c is a normalizing constant:

$$c = \sum_{i=1}^{W_E}\lambda^{i-1} = \frac{1-\lambda^{W_E}}{1-\lambda} \to \frac{1}{1-\lambda} as\, W_E \to \infty . \quad (2)$$

The decay factor is commonly used in practice at 0.94 (Danielsson, 2012).

2.4. VaR backtesting

2.4.1. *Binomial test (Bin)*

The binomial test (Jorion, 2011) compares the experiential amount of exemptions to the anticipated amount of exemptions. The outcome of the test is:

$$Z_{bin} = \frac{x - Np}{\sqrt{Np(1-p)}}, \quad (3)$$

where x, N denotes the number of failures, and amount of observations respectively; $p = 1 - VaR$ level is a likelihood of detecting a failure if VaR is right.

2.4.2. *Traffic Light Test (TL)*

For a certain list of exemptions, x, the traffic light test as described in the Basel committee report[1] can calculate the cumulative likelihood up to x. That is, $F(x|N, p)$. It is also known as the three zones test and denoted as:

- Green: $F(x|N, p) \leq 0.95$. If one has a small number of failures in the VaR model, they will be placed in the green zone.
- Yellow: $0.95 \leq F(x|N, p) \leq 0.9999$. Even though the failures are high, the violation count is not exceedingly high and hence they fall in the yellow zone.
- Red: $0.9999 < F(x|N, p)$. A proper VaR model is unlikely to provide too many exceptions. So, in case of too many failures, it falls in the red zone.

1 Basel Committee on Banking Supervision (1996).

2.4.3. Kupiec's POF and TUFF tests

Kupiec (1995) developed a proportion of failures (POF) test that accommodates a binomial distribution and employs a probability ratio to determine if the likelihood of exceptions is matched with the likelihood p given by the VaR level of confidence. The VaR model is rejected if the data advocates that the chance of exceptions is larger than p. The POF test statistic is:

$$LR_{POF} = -2\log\left[\frac{(1-p)^{N-x}\,p^{x}}{\left(1-\frac{x}{N}\right)^{N-x}\left(\frac{x}{N}\right)^{x}}\right], \quad (4)$$

where x, N denote the number of failures, and a number of observations respectively; $p = 1 - VaR$ level is the likelihood of detecting a failure if the VaR model is right.

Kupiec (1995) presented an alternative test called the time until first failure (TUFF) test which is likewise based on a likelihood ratio test but has a geometric distribution as the underlying distribution.

$$LR_{TUFF} = -2\log\left[\frac{p(1-p)^{n-1}}{\left(\frac{1}{n}\right)\left(1-\frac{1}{n}\right)^{n-1}}\right]. \quad (5)$$

Both statistics have an asymptotic distribution as a chi-square variable with one degree of freedom. If the likelihood ratio reaches a crucial value determined by the test confidence level, the VaR model fails the test.

2.4.4. Conditional coverage mixed test

Christoffersen (1998) suggested a measure to determine if the likelihood of detecting an exemption at a given time is affected by whether an exemption happened. In Christoffersen's interval forecast technique, the test statistic for independence is provided by:

$$LR_{CCI} = \\ = -2\log\left(\frac{(1-\pi)^{n00+n10}\,\pi^{n01+n11}}{(1-\pi_0)^{n00}\,\pi_0^{n01}\,(1-\pi_1)^{n10}\,\pi_1^{n11}}\right), \quad (6)$$

where $n00$: the number of failure-free periods followed by a failure-free period, $n10$: the number of periods with failures followed by a failure-free period, $n01$: the number of failure-free periods followed by a failed period, $n11$: the number of failed periods followed by a failed period.

$$\pi_0 = \frac{n01}{(n00+n01)};$$
$$\pi_1 = \frac{n11}{(n10+n11)}; \quad (7)$$
$$\pi = \frac{n01+n11}{(n00+n01+n10+n11)}.$$

Asymptotically, this statistic is distributed as a chi-square with one degree of freedom. This statistic may be used with the frequency POF test to create a conditional coverage (CC) mixed test:

$$LR_{CC} = LR_{POF} + LR_{CCI}. \quad (8)$$

As a chi-square variable with two degrees of freedom, this test is asymptotically distributed.

2.4.5. Time between failures (TBF) or mixed Kupiec's test

Haas (2001) modified Kupiec's TUFF test by including the time data among all failures in the experiment, thus defining the time between failures (TBF) test statistic by:

$$LR_{TBFI} = -2\sum_{i=1}^{x}\log\left(\frac{p(1-p)^{n_i-1}}{\left(\frac{1}{n_i}\right)\left(1-\frac{1}{n_i}\right)^{n_i-1}}\right). \quad (9)$$

In this statistic, $p = 1 - VaR$ level and n_i is the number of days between failures $i - 1$ and i (or until the first exception for $i = 1$). Asymptotically, this statistic is distributed as a chi-square variable with x degrees of freedom, where x is the rate of failure.

This test, like Christoffersen's test, may be combined with the frequency POF test to produce a TBF mixed test, also known as Haas' mixed Kupiec's test:

$$LR_{TBF} = LR_{POF} + LR_{TBFI}. \quad (10)$$

http://dx.doi.org/10.21511/imfi.19(1).2022.04

As a chi-square variable with $x + 1$ degrees of freedom, this test is asymptotically distributed.

3. DATA

The study uses daily data of stock market indices of BRICS nations and the US index for the period from 2003 to 2021. The daily price series data are downloaded from Bloomberg. The stock prices in this study include IBOVESPA (Brazil), IMOEX (Russia), NIFTY (India), SSE (China), JTOPI (South Africa), and DJIA (USA) index. The price series are converted to return series computed as $\log\left(P_t / P_{t-1}\right)\cdot 100$, to make the series stationary.

Table 1 shows the descriptive statistics of the returns. The number of daily observations of the stock indices returns is in the range between 4568 and 4736. The highest mean (0.00072) is for the Russian stock index and the lowest mean (0.00033) is observed for the Chinese index. The Russian index is observed to have the highest standard deviation (0.0188) and the US stock index to have the lowest standard deviation (0.0115). Further, all the stock indices are observed to have negative skewness except the Russian index. All the stock index prices are found to have high kurtosis values and are not normally distributed based on the p-values of the Jarque-Bera test. The return series are observed to be stationary in nature as per Augmented Dickey-Fuller (ADF) test.

For VaR estimation and backtesting, the data are divided into two window samples namely the estimation window and the testing window. The estimation window starts from 2003 to the end of 2005 and the estimation window size is 250 trading days. The testing window starts from 2006 to 2021. For VaR estimation and backtesting methods, the testing window sample is used and is further divided into three sub-periods: the overall period (2006–2021), GFC period (2008–2009), and the COVID-19 period (2020–2021).

4. RESULTS AND DISCUSSION

This section first estimates the VaR using three different models like normal (N), historical simulation (HS), and exponential weighted moving average (EWMA) methods. Secondly, the VaR backtesting using eight different models like Bin, POF, TL, TUFF, CC, TBF, CCI, and TBFI tests were applied. The analysis is conducted for three different periods: overall period (2006–2021), global financial crisis (GFC) period (2008–2009), and COVID-19 period (2020–2021) for BRICS and the US stock market indices.

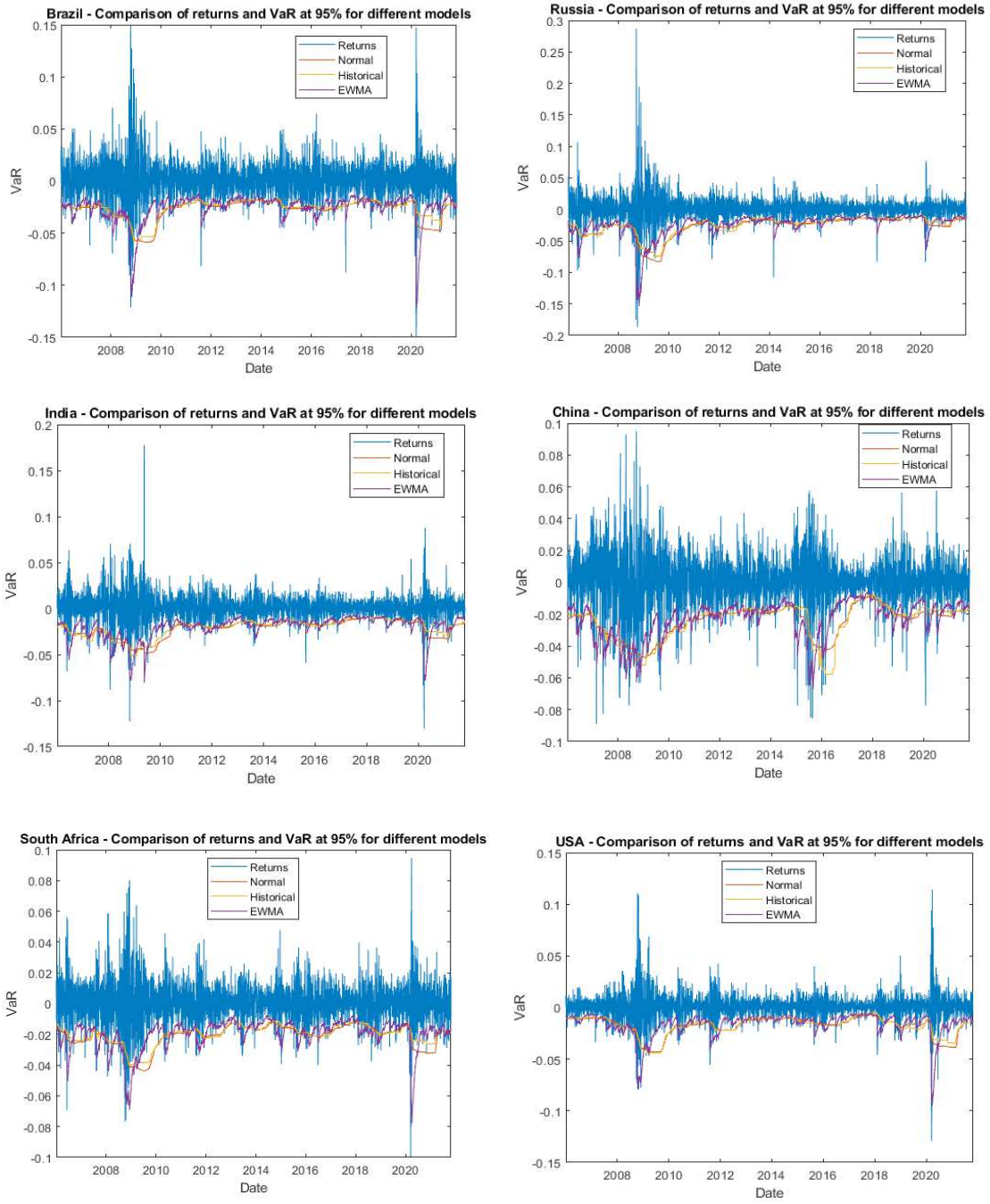
Figure 1 compares the returns of the indices and VaR estimation methods at a 95% level of confidence for the overall testing period from 2006 to 2021. The plot shows that EMWA performs better compared to HS and N estimation methods for all the six stock indices. Figure 2 displays the plots of returns of the indices and VaR estimation methods at a 95% confidence level for the GFC period from 2008 to 2009. Figure 3 shows the plots of returns of the indices and VaR estimation methods at a 95% confidence level for the COVID-19 period from 2020 to 2021. Similar findings are observed in both the GFC and COVID-19 periods, that the EMWA model performs better compared to HS and N models.

All three Figures 1, 2, and 3 exhibit the degree to which the three VaR estimation models N, HS, and EWMA are successful to evaluate the index returns related to the actual index returns. It is observed that actual index returns in all three periods exhibit stylized facts like volatility cluster-

Table 1. Descriptive statistics of the return series

Country Name	Count	Mean	Median	Std	Skewness	Kurtosis	JBTest	ADFTest
Brazil	4654	0.00067	0.0010	0.0173	–0.1377	12.0801	0.001	0.001
Russia	4711	0.00072	0.0009	0.0188	0.4038	27.9151	0.001	0.001
India	4676	0.00069	0.0011	0.0141	–0.1914	15.6844	0.001	0.001
China	4568	0.00033	0.0006	0.0154	–0.3678	7.6099	0.001	0.001
South Africa	4705	0.00049	0.0009	0.0130	–0.0784	8.0130	0.001	0.001
USA	4736	0.00036	0.0006	0.0115	–0.1616	19.3540	0.001	0.001

Note: Std denotes the standard deviations. JBTest and ADFTest denote the p-values of Jarque-Bera test statistic and Augmented Dickey-Fuller (ADF) test statistic. Count denotes the number of observations of the return series.



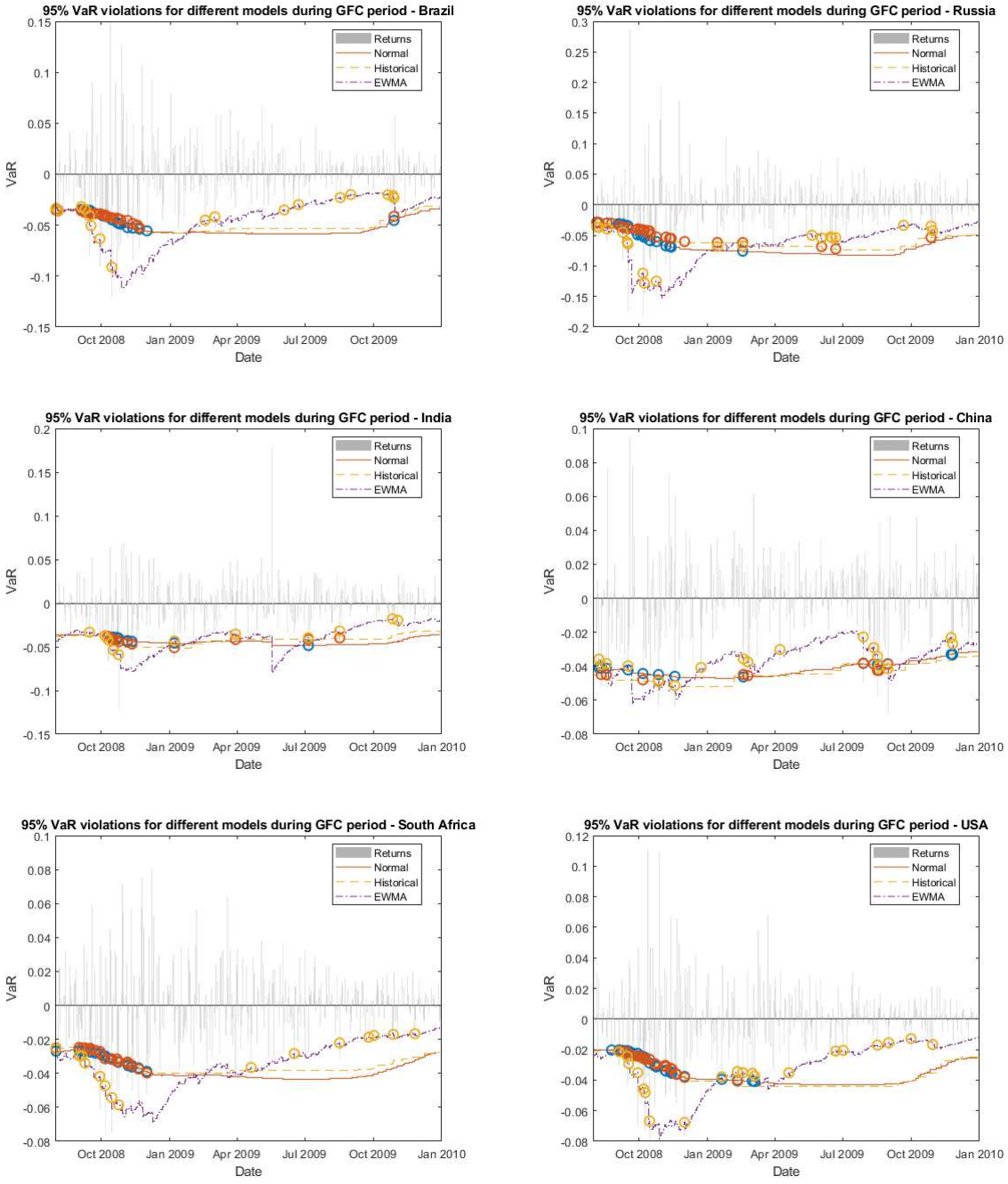
Note: The plot displays the returns and estimates the VaR using different models like normal, historical simulation, and EWMA method at a 95% level of confidence for the overall period from Jan 2006 to Oct 2021 for six stock market indices, i.e., Brazil, Russia, India, China, South Africa, and the USA.

Figure 1. Comparison of returns and VaR at 95% for different models (2006–2021)

ing. In Figure 1, the returns for all the stock indices are observed to be more volatile during extreme events like the GFC period (2008–2009) and COVID-19 period (2020–2021) and less volatile during other periods. All the stock market index returns exhibit similar patterns, suggesting that extreme episodes in stock markets have parallel effects in BRICS and US nations. Only in the case of the Chinese stock index, the returns are observed to be volatile even during 2015–2016 apart from GFC and COVID-19 periods.

The VaR estimation models, N, HS, and EWMA exhibit similar patterns for all the stock indices under study. Vague generalizations show that the EWMA VaR estimate methodology is substan-

http://dx.doi.org/10.21511/imfi.19(1).2022.04

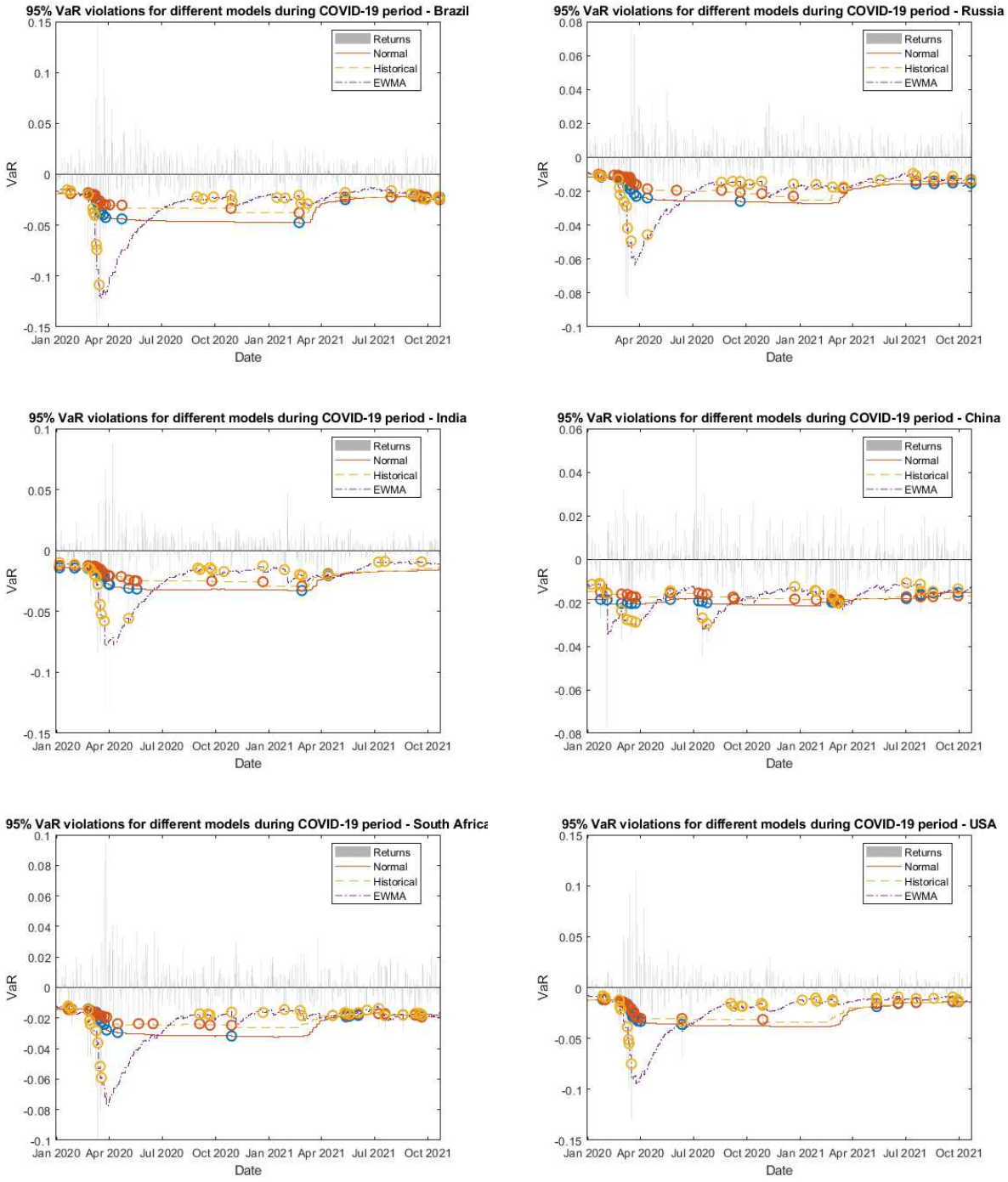


Note: The plot display the returns and estimate the VaR using different models like normal, historical simulation, and EWMA method at a 95% level of confidence for the GFC period from Jan 2008 to Dec 2009 for six stock market indices, i.e., Brazil, Russia, India, China, South Africa, and the USA.

Figure 2. Comparison of returns and VaR at 95% for different models (GFC, 2008–2009)

tially more effective in predicting risk exposures for all three periods under study. On the other hand, N and HS estimation methods are less fruitful in projecting the market risks and their correlation of predictive accuracy is seen to be relatively near to one across all periods. By looking at Figures 2 and 3, it is observed that during extreme event periods like GFC and COVID-19, both the models N and HS underestimate the risks whenever the moment of recession uncovers, and overestimate the risks whenever the market starts to stabilize again.

It is known that computing VaR is critical for enterprises and institutional investors to make wise financial decisions but it is also vital to guar-



Note: The plots display the returns and estimate the VaR using different models like normal, historical simulation, and EWMA method at a 95% level of confidence for the COVID-19 period from Jan 2020 to Oct 2021 for six stock market indices, i.e., Brazil, Russia, India, China, South Africa, and the USA.

Figure 3. Comparison of returns and VaR at 95% for different models (COVID-19, 2020–2021)

antee that the calculation is computed with the least estimation error and in the most accurate manner. Therefore, the estimation models in this study are evaluated through eight different back-testing methods to understand their predictive performance.

In Tables 2 and 3, the headings TL, Bin, POF, TUFF, CC, CCI, TBF, and TBFI denote traffic light, binomial, proportion of failure, time until first failure, conditional coverage, conditional coverage independence, time between failures, and time between failures independence test re-

http://dx.doi.org/10.21511/imfi.19(1).2022.04

Table 2. VaR model evaluation through different backtesting methods for 2006–2021

		Overall period (2006–2021)							
		TL	**Bin**	**POF**	**TUFF**	**CC**	**CCI**	**TBF**	**TBFI**
		VaR at a 95% level of confidence							
Brazil	Normal	Ğ	√	√	√	√	×	×	×
	Historical	Ğ	√	√	√	×	×	×	×
	EWMA	Ğ	√	√	√	√	√	×	×
Russia	Normal	Ğ	√	×	√	×	×	×	×
	Historical	Ğ	√	√	√	×	×	×	×
	EWMA	Ğ	√	√	√	√	√	×	×
India	Normal	Ğ	√	√	√	×	×	×	×
	Historical	Ğ	√	√	√	×	×	×	×
	EWMA	Ğ	√	√	√	√	√	×	×
China	Normal	Ğ	√	√	√	×	×	×	×
	Historical	Ğ	√	√	√	×	×	×	×
	EWMA	Ğ	√	√	√	√	√	×	×
SA	Normal	Ğ	√	√	√	×	×	×	×
	Historical	Ŷ	√	√	√	×	×	×	×
	EWMA	Ŷ	×	×	√	×	√	×	×
US	Normal	Ŷ	√	√	√	×	×	×	×
	Historical	Ŷ	√	√	√	×	×	×	×
	EWMA	Ŷ	×	×	√	√	√	×	×
		VaR at a 99% level of confidence							
Brazil	Normal	Ŷ	×	×	√	×	×	×	×
	Historical	Ŷ	×	×	√	×	×	×	×
	EWMA	Ŷ	×	×	√	×	×	×	×
Russia	Normal	Ř	×	×	√	×	×	×	×
	Historical	Ř	×	×	√	×	×	×	×
	EWMA	Ř	×	×	√	×	√	×	×
India	Normal	Ř	×	×	√	×	×	×	×
	Historical	Ŷ	×	×	√	×	×	×	×
	EWMA	Ř	×	×	√	×	√	×	×
China	Normal	Ř	×	×	√	×	×	×	×
	Historical	Ŷ	√	√	√	×	×	×	×
	EWMA	Ř	×	×	√	×	√	×	×
South Africa	Normal	Ř	×	×	√	×	√	×	×
	Historical	Ŷ	×	×	√	×	√	×	×
	EWMA	Ř	×	×	√	×	√	×	×
US	Normal	Ř	×	×	√	×	×	×	×
	Historical	Ŷ	×	×	√	×	×	×	×
	EWMA	Ř	×	×	√	×	√	×	×

Note: Ř, G, Ŷ, √, and × refer to Red, Green, Yellow, Accept, and Reject respectively. EWMA represents the exponential weighted moving average method. The tests TL, Bin, POF, TUFF, CC, CCI, TBF, and TBFI represents the traffic light test, Binomial test, Kupiec's proportion of failure test, time until first failure test, conditional coverage test, conditional coverage independent test, time between failure test, and time between failure independent test respectively.

spectively. The following criteria are used to assess the outcomes. In a TL test, G stands for the green zone, Y stands for the yellow zone, and Rd stands for the red zone. Under the prospective backtesting approach, A signifies acceptance and R signifies rejection of the VaR model for all other tests. The VaR backtesting is applied to all the three VaR estimation models at both the 95% and 99% levels of confidence.

The first column of Table 2 displays the outcome of the Traffic Light (TL) test at both 95% and 99% confidence levels. TL test is an efficient backtesting approach, particularly for the banks and monetary organizations that are required to keep stringent liquidity as part of their capital adequacy. Because the Basel Committee has only suggested categories for 0.01 coverage rate, this study presents the results of TL test data only at a 99 percent con-

fidence level[2]. The study discovers that all of the findings fall into the yellow or red zone. It shows that the predicted accuracy of all the VaR methods is more likely to be wrong during that period. All VaR models are rejected by the TL test at a 99 percent level of confidence because not all of the entities that comprise the stock market indices are banks or monetary institutions that are required to properly preserve their capital necessities.

The second column of Table 2 displays the results of the Binomial test (Bin) at both 95% and 99% confidence levels. Bin test examines the occurrence of failures of VaR methods. At a 95% confidence level, it is observed that all the three VaR estimation models for all the six stock indices get accepted based on the binomial test except for Russia (normal estimation), South Africa (EWMA estimation), and the US (EWMA estimation) that get rejected. At a 99 percent level of confidence, the study finds that all the VaR methods are rejected for all the stock indices, except for China (historical simulation) which is accepted. Similar findings are observed for Kupiec's POF test that is mentioned in the third column of Table 2, which checks if the experimentally observed probability equates to the theoretically determined probability. The rejections increase at a 99% confidence level for frequency-based backtesting methods like TL, Bin, and POF since under normality assumption the VaR is underestimated at higher confidence levels (Lin & Shen, 2006). In the fourth column of Table 2, the TUFF test merely specifies the moment when the very first fault occurs in any particular testing frame, providing very little data about the trend of subsequent failures. As a result, the TUFF test accepts all VaR models across the whole timeframe.

The independence tests, on contrary, such as CCI and TBFI, mostly reject all VaR methods for all stock market indexes at both the 95 and 99 percent levels of confidence. It means that the VaR methods overlook failure clustering in the testing window on an aggregate basis. In certain circumstances, the failures are not independent of evidence from previous failures. Furthermore, all VaR models are rejected at both confidence levels using joint tests such as CC and TBF tests. It means that VaR methods fail to fulfill the standards established by tests like CCI or TBFI tests together with the POF test.

In Table 3, the study compares the predictive performance during the global financial crisis period (2008–2009), and uncertainty induced due to COVID-19 (2020–2021). Based on the TL test, it is observed that at a 99 percent level of confidence, all the findings fall into the yellow or red zone except three instances during the GFC period and one instance in the COVID-19 period during which it falls into the green zone. Based on the frequency-based tests like Bin, POF, and TUFF tests, it is observed that at a 99 percent level of confidence, the VaR methods are rejected majorly during the GFC period compared to the COVID-19 period. Based on the independence tests like CCI, it is observed that at a 99 percent level of confidence, all the VaR methods are accepted except China (EWMA) is rejected during the GFC period, whereas Brazil (N, HS), China (H), and US (H) get rejected during the COVID-19 period. Based on the TBFI approach, the VaR methods are majorly rejected during the GFC period and all the methods for all the indices are rejected during the COVID-19 period. The study further looks at the outcomes of more stringent joint tests like CC and TBF approaches during crisis periods. Based on the CC method, it is observed that at a 99 percent level of confidence, VaR estimation models are rejected majorly during the COVID-19 period compared to the GFC period. By looking at the outcomes of stringent joint tests like TBF at a 99 percent level of confidence, it is observed that the VaR estimation method based on the EWMA method get accepted for the US, South Africa, and Brazil stock markets, and the rest are rejected during the GFC period. However, during the COVID-19 period, all the VaR estimation models are observed to be rejected based on the TBF backtesting model at a 99 percent level of confidence. Overall, the empirical results show that VaR estimation methods are found to have poor predictive accuracy during the COVID-19 period compared to the GFC period as all the VaR estimation methods get rejected for all the stock market indices considered in this study.

2 Results at a 95% confidence level are also shown in Table 2.

http://dx.doi.org/10.21511/imfi.19(1).2022.04

Table 3. VaR model evaluation through different backtesting methods for the GFC (2008–2009) and COVID-19 (2020–2021) periods

		GFC period (2008–2009)								COVID-19 period (2020–2021)							
		TL	Bin	POF	TUFF	CC	CCI	TBF	TBFI	TL	Bin	POF	TUFF	CC	CCI	TBF	TBFI
VaR at a 95% level of confidence																	
Brazil	Normal	Ř	×	×	√	×	√	×	×	Ğ	√	√	√	×	×	×	×
	Historical	Ř	×	×	√	×	√	×	×	Ğ	√	√	√	×	×	×	×
	EWMA	Ŷ	×	×	√	√	√	√	√	Ğ	√	√	√	√	×	×	×
Russia	Normal	Ř	×	×	√	×	√	×	×	Ğ	√	√	√	√	√	×	×
	Historical	Ř	×	×	√	×	√	×	×	Ŷ	×	×	√	×	×	×	×
	EWMA	Ŷ	√	√	√	√	√	√	√	Ğ	√	√	√	√	√	×	×
India	Normal	Ŷ	×	×	√	×	√	×	×	Ğ	√	√	√	×	×	×	×
	Historical	Ŷ	×	×	√	√	√	×	×	Ŷ	×	×	√	×	√	×	×
	EWMA	Ŷ	×	×	√	√	√	×	×	Ğ	√	√	√	√	√	×	×
China	Normal	Ŷ	×	×	√	×	√	×	√	Ğ	√	√	√	√	√	√	√
	Historical	Ğ	√	√	√	√	√	√	√	Ğ	√	√	√	√	√	√	×
	EWMA	Ŷ	×	√	√	√	√	×	×	Ğ	√	√	√	√	√	×	×
SA	Normal	Ŷ	×	×	√	×	√	×	×	Ğ	√	√	√	√	√	×	×
	Historical	Ŷ	×	×	√	×	√	×	×	Ŷ	√	√	√	√	√	×	×
	EWMA	Ğ	√	√	√	√	√	√	√	Ğ	√	√	√	√	√	×	×
US	Normal	Ř	×	×	×	×	√	×	×	Ğ	√	√	√	√	√	×	×
	Historical	Ř	×	×	√	×	√	×	×	Ğ	√	√	√	√	√	×	×
	EWMA	Ŷ	×	×	√	×	√	×	×	Ğ	√	√	√	√	√	×	×
VaR at a 99% level of confidence																	
Brazil	Normal	Ř	×	×	√	×	√	×	×	Ŷ	×	×	√	×	×	×	×
	Historical	Ŷ	×	×	√	×	√	×	×	Ŷ	×	×	√	×	×	×	×
	EWMA	Ğ	√	√	√	√	√	√	√	Ŷ	√	√	√	√	√	×	×
Russia	Normal	Ř	×	×	×	×	√	×	×	Ř	×	×	√	×	√	×	×
	Historical	Ř	×	×	×	×	√	×	×	Ř	×	×	×	×	√	×	×
	EWMA	Ř	×	×	×	×	√	×	×	Ř	×	×	×	×	√	×	×
India	Normal	Ř	×	×	√	×	√	×	×	Ř	×	×	√	×	√	×	×
	Historical	Ŷ	×	×	√	×	√	×	×	Ř	×	×	×	×	√	×	×
	EWMA	Ŷ	×	√	√	×	√	×	×	Ř	×	×	×	×	√	×	×
China	Normal	Ŷ	×	×	√	×	√	×	×	Ŷ	×	×	√	×	√	×	×
	Historical	Ğ	√	√	√	√	√	√	√	Ğ	√	√	√	√	×	×	×
	EWMA	Ğ	√	√	√	√	×	×	×	Ŷ	√	√	√	√	√	×	×
South Africa	Normal	Ř	×	×	√	×	√	×	×	Ř	×	×	√	×	√	×	×
	Historical	Ŷ	×	√	√	√	√	×	×	Ŷ	×	×	√	×	√	×	×
	EWMA	Ŷ	√	√	√	√	√	√	√	Ŷ	×	×	√	×	√	×	×
US	Normal	Ř	×	×	√	×	√	×	×	Ř	×	×	√	×	√	×	×
	Historical	Ř	×	×	√	×	√	×	×	Ŷ	×	×	√	×	×	×	×
	EWMA	Ŷ	√	√	√	√	√	√	√	Ř	×	×	√	×	√	×	×

Note: Ř, Ğ, Ŷ, √, and × refer to Red, Green, Yellow, Accept, and Reject respectively. EWMA represents the exponential weighted moving average method. The tests TL, Bin, POF, TUFF, CC, CCI, TBF, and TBFI represents the traffic light test, Binomial test, Kupiec's proportion of failure test, time until first failure test, conditional coverage test, conditional coverage independent test, time between failure test, and time between failure independent test respectively.

CONCLUSION

The primary objective of this study is to evaluate the three VaR estimation methods and understand their predictive performance for different periods based on eight backtesting methods on emerging BRICS and developed US stock markets. The empirical results show that during the overall period (2006–2021) EWMA VaR estimation method performs superior compared to historical and normal simulation methods. The study finds that during extreme event periods like GFC and COVID-19, both

the models N and HS underestimate the risks whenever the moment of recession uncovers, and overestimate the risks whenever the market starts to stabilize again. Amongst the three estimation methods considered in this study, the EWMA method performs better. Furthermore, it is observed that VaR estimation models have poor predictive accuracy, especially during the COVID-19 period compared to the global financial crisis.

Overall, the study rates the VaR models as EWMA > historical > normal based on their predictive performance. It is obvious that dynamic VaR models such as EWMA outperform static methods such as historical and conventional simulation approaches. During a crisis, however, the predictive performance of the VaR models fails catastrophically. This raises a crucial concern regarding the VaR models' usefulness during extreme event times.

As for recommendations for future study, it will be interesting to test parametric GARCH family models alongside alternative semi-parametric approaches such as CAViaR and EVT methods, which may bring novel views on this line of research. Because VaR has distinct limitations as a risk measure, it would be thought provoking to examine market risks using anticipated shortfall (ES) models in addition to VaR models in future studies.

AUTHOR CONTRIBUTIONS

Conceptualization: Muneer Shaik, Lakshmi Padmakumari.
Formal analysis: Muneer Shaik.
Investigation: Lakshmi Padmakumari.
Methodology: Muneer Shaik.
Software: Muneer Shaik.
Supervision: Muneer Shaik, Lakshmi Padmakumari.
Validation: Lakshmi Padmakumari.
Writing – original draft: Muneer Shaik.
Writing – review & editing: Muneer Shaik, Lakshmi Padmakumari.

REFERENCES

1. Basel Committee on Banking Supervision. (1996). *Supervisory Framework for the Use of 'Backtesting' in Conjunction with the Internal Models Approach to Market Risk Capital Requirements.* Retrieved from https://www.bis.org/publ/bcbs22.htm
2. Bonga-Bonga, L., & Nleya, L. (2016). *Assessing portfolio market risk in the BRICS economies: use of multivariate GARCH models* (MPRA Paper 75809). University Library of Munich, Germany. Retrieved from https://mpra.ub.uni-muenchen.de/75809/
3. Burdorf, T., & van Vuuren, G. (2018). An evaluation and comparison of Value at Risk and Expected Shortfall. *Investment Management and Financial Innovations, 15*(4), 17-34. https://doi.org/10.21511/imfi.15(4).2018.02
4. Cheong, C. W., Isa, Z., & Nor, A. S. M. (2011). Cross market value-at-risk evaluations in emerging markets. *African Journal of Business Management, 5*(22), 9385-9400. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.892.5978&rep=rep1&type=pdf
5. Choi, P., & Min, I. (2011). A comparison of conditional and unconditional approaches in value-at-risk estimation. *Japanese Economic Review, 62*(1), 99-115. https://doi.org/10.1111/j.1468-5876.2010.00456.x
6. Christoffersen, P. (1998). Evaluating Interval Forecasts. *International Economic Review, 39*(4), 841-862. https://doi.org/10.2307/2527341
7. Cogneau, P. (2015). Backtesting Value-at-Risk: how good is the model? In *Intelligent Risk: knowledge for the PRMIA community* (pp. 028-034). Retrieved from https://issuu.com/prmia/docs/07212015_-_intelligent_risk_-_july_
8. Danielsson, J. (2012). *Financial Risk Forecasting: The Theory and Practice of Forecasting Market Risk, with Implementation in R and MATLAB*. Wiley Finance.
9. Degiannakis, S., Floros, C., & Livada, A. (2012). Evaluating value-at-risk models before and after the financial crisis of 2008: International evidence. *Managerial*

http://dx.doi.org/10.21511/imfi.19(1).2022.04

Finance, 38(4), 436-452. https://doi.org/10.1108/03074351211207563

10. Desheng, W., & Chatpailin, T. (2019). Value at Risk Performance in BRICS Countries. *IEEE International Conference on Cybernetics*.
11. Edwards, D. W. (2014). *Risk management in trading. Techniques to drive profitability of hedge funds and trading desks*. New Jersey: John Wiley & Sons.
12. Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance, 48*(5), 1779-1801. https://doi.org/10.1111/j.1540-6261.1993.tb05128.x
13. Haas, M. (2001). *New Methods in Backtesting*. Bonn: Financial Engineering Research Center Caesar. Retrieved from https://www.ime.usp.br/~rvicente/risco/haas.pdf
14. Iglesias, E. M. (2015). Value at risk of the main stock market indexes in the European Union (2000–2012). *Journal of Policy Modeling, 37*(1), 1-13. http://dx.doi.org/10.1016%2Fj.jpolmod.2015.01.006
15. Jiang, Y., Fu, Y., & Ruan, W. (2019). Risk spillovers and portfolio management between precious metal and BRICS stock markets. *Physica A: Statistical Mechanics and its Applications, 534,* 120993. http://dx.doi.org/10.1016/j.physa.2019.04.229
16. Jobayed, A. (2017). *Evaluating the predictive performance of Value-at-Risk (VaR) models on Nordic market indices* (Master's Thesis). Hansen School of Economics. https://doi.org/10.13140/RG.2.2.33136.97284
17. Jorion, P. (2011). *Financial Risk Manager Handbook* (6th ed.). Wiley Finance.
18. Kupiec, P. (1995). Techniques for Verifying the Accuracy of Risk Management Models. *Journal of Derivatives, 3,* 73-84.
19. Lin, C. H., & Shen, S. S. (2006). Can the student-*t* distribution provide accurate value at risk? *The Journal of Risk Finance, 7*(3), 292-300. https://doi.org/10.1108/15265940610664960
20. Mabitsela, L., Maré, E., & Kufakunesu, R. (2015). Quantification of VaR: A Note on VaR Valuation in the South African Equity Market. *Journal of Risk and Financial Management, 8*(1), 103-126. https://doi.org/10.3390/jrfm8010103
21. Mathworks. (2017). *Overview of VaR backtesting*. Retrieved March 14, 2017, from https://se.mathworks.com/help/risk/overview-of-var-backtesting.html
22. McNeil, A., Frey, R., & Embrechts, P. (2005). *Quantitative Risk Management*. Princeton University Press.
23. Miletic, M., & Miletic, S. (2015). Performance of Value at Risk models in the midst of the global financial crisis in selected CEE emerging capital markets. *Economic research – Ekonomska istraživanja, 28*(1), 132-166. https://doi.org/10.1080/1331677X.2015.1028243
24. Mukta, K., & Muneer, S. (2020a). Are BRICS stock market indices mean reverting? Evidence based on expected lifetime range ratio. *International Journal of Business and Economics, 19*(2), 169-186. Retrieved from https://ideas.repec.org/a/ijb/journl/v19y2020i2p169-186.html
25. Mukta, K., & Muneer, S. (2020b). Testing asymmetry in mean reversion based on high and low prices: Evidence from BRICS markets. *Journal of Public Affairs,* e2443. https://doi.org/10.1002/pa.2443
26. Muteba Mwamba, J., & Beytell, D. (2015). Value At Risk, Minimum Capital Requirement and the use of Extreme Value Distributions: An application to BRICS Markets. *International Business & Economics Research Journal, 14*(1), 135-144. https://doi.org/10.19030/iber.v14i1.9037
27. Naradh, K., Chinhamu, K., & Chifurira, R. (2021). Estimating the value-at-risk of JSE indices and South African exchange rate with Generalized Pareto and stable distributions. *Investment Management and Financial Innovations, 18*(3), 151-165. http://dx.doi.org/10.21511/imfi.18(3).2021.14
28. Nieppola, O. (2009). *Backtesting Value-at-Risk Models*. Aalto University. Retrieved from https://aaltodoc.aalto.fi/handle/123456789/181
29. Omari, C., Mundia, S., & Ngina, I. (2020). Forecasting Value-at-Risk of Financial Markets under the Global Pandemic of COVID-19 Using Conditional Extreme Value Theory. *Journal of Mathematical Finance, 10*(4), 569-597. https://doi.org/10.4236/jmf.2020.104034
30. Ozun, A., & Cifter, A. (2007). Portfolio value-at-risk with time-varying copula: Evidence from Latin America. *Journal of Applied Sciences, 7*(14), 1916-1923. https://doi.org/10.3923/jas.2007.1916.1923
31. Ramalho, D. R. V. (2020). *Predictive performance of value-at-risk models: Covid-19 "Pandemonium"* (Master's Thesis). Universidade de Lisboa. Retrieved from https://www.repository.utl.pt/handle/10400.5/20949
32. Song, Q., Liu, J., & Sriboonchitta, S. (2019). Risk Measurement of Stock Markets in BRICS, G7, and G20: Vine Copulas versus Factor Copulas. *Mathematics, 7*(3), 274. https://doi.org/10.3390/math7030274
33. Su, Y-C., Huang, H-C., & Wang, S. (2010). Modeling value at risk of financial holding company: time varying vs. traditional models. *Banks and Bank Systems, 5*(1). Retrieved from https://www.businessperspectives.org/images/pdf/applications/publishing/templates/article/assets/3039/BBS_en_2010_1_Su.pdf
34. Subbiah, M., & Fabozzi, F. J. (2016). Hedge fund allocation: Evaluating parametric and nonparametric forecasts using alternative portfolio construction techniques. *International Review of Financial Analysis, 45*(C), 189-201. Retrieved from https://ideas.repec.org/a/eee/finana/v45y2016icp189-201.html