Business Analyst, ISSN 0973-211X, 38(1), 36-57, ©SRCC

BACKTESTING VAR MODELS: THE CASE OF COMMODITIES

Devesh Shankar[*], Prateek Bedi[**], Shalini Agnihotri[#] and Jappanjyot Kaur Kalra[Φ]

Abstract

One of the most widely used methods to quantify risk is 'Value at Risk'. VaR models are useful only if they predict future risks accurately. This paper focuses on a comparative evaluation of three broad approaches to calculate VaR for nine commodities traded on Multi Commodity Exchange of India. The primary objective of the study is to identify the most accurate VaR model for each commodity in particular and commodity asset class in general. VaR is calculated using five different methods (two methods each of parametric & non-parametric approaches and one method of semi-parametric approach) for all nine commodities for a period of nine years starting October 2006 till October 2015. To identify the better performing VaR methods accurately, the analysis is performed in two phases, Pre-Crisis (October 2006 to December 2009) and Post Crisis (January 2010 to October 2015). Results suggest Volatility Weighted Historical Simulation (VWHS) VaR method has outperformed other methods in both parts of the analysis exhibiting a success ratio of 100% each time. We also conclude that the selection of similar or contrasting data periods in terms of market conditions for VaR calculation and VaR backtesting affects the performance of VaR methods in general. These findings are relevant for retail and institutional investors who hold commodities in their portfolios and traders who need to calculate VaR for their commodity portfolios.

Keywords: Value at Risk, Backtesting, Historical VaR, Bootstrapping, Volatility Weighted Normal VaR, GARCH (1,1) VaR, Kupiec's Test

1. Introduction

Over the past few decades, risk management has evolved to a point where it is considered

[*] Research Scholar, Faculty of Management Studies, University of Delhi; Email: devesh.shankar_phd@fms.edu

[**] Research Scholar, Department of Financial Studies, University of Delhi New Delhi; Email: prateekbedi.du@gmail.com (Corresponding Author)

[#] Faculty Associate, Lal Bahadur Shastri Institute of Management, Delhi; Email: agnihotri123shalini@gmail.com

[Φ] Research Scholar, Faculty of Management Studies, University of Delhi; Email: jappanjyot@yahoo.com

to be a distinct sub-field in the theory of finance. The growth of risk management industry traces back to the increased volatility of financial markets in 1970's. Risk is an important determinant of investment decision. Measurement of risk has evolved over time since the advent of Harry Markowitz's idea of using standard deviation as a proxy for risk in 1952. Risk has been defined and measured in different ways by researchers, academicians and practitioners. One of the most widely used methods to quantify risk is 'Value At Risk' (commonly known as VaR). During the past decade, VaR has become one of the most popular risk measurement techniques in finance. VaR aims to capture the market risk of assets. According to Bank for International Settlements (BIS), market risk is defined as the risk of losses in on and off balance sheet positions arising from movements in market prices. Put formally, VaR measures the maximum loss in value for a portfolio over a predetermined time period for a given level of confidence.

Investment in commodities is seen as a balancing effect on the portfolio and acts as a price risk management tool to avoid prices of all the assets in a portfolio from going down or to hedge against inflation. Commodities, besides being a unique hedging instrument, also provide for efficient portfolio management due to diversification benefits. These benefits result in improved returns to domestic as well as international investors. Commodity price risks should incorporate risk factors corresponding to each of the commodity markets in which long/short positions exist. Producers hedge price risk by assuming short positions in futures contracts on the commodity that they produce.

Commodities exhibit certain risk characteristics that are different from traditional assets like stocks and bonds. A good value-at-risk (VaR) model must capture those characteristics.

First, many commodities, such as agricultural products, are not storable. Others, such as livestock and energy products like electricity, can only be stored at very high costs. As a result, supply or demand changes are translated immediately into price changes, which lead to higher volatility of commodity investments compared to traditional assets. Second, supply and demand shocks occur more frequently and on a larger scale. Drought or frost can lead to an unexpected decrease in the supply of agricultural products and a subsequent sharp increase in prices. Natural disasters can also affect commodity prices. Political instability in oil exporting countries accounts for the additional variation in oil prices. The third source of commodity price variation is the lack of governmental control. While central banks can influence stock and bond markets they cannot compensate for supply shocks or changes in commodity prices.

The risk characteristics mentioned above are reflected in the return-generating process, which makes commodities the ideal time series for testing value-at-risk models: Long tranquil periods alternate with sudden volatility spikes or long periods of high volatility. Some commodity series do not show mean-reverting behaviour, but rather exhibit shifts in volatility. Others switch between positive and negative skewness depending on the time period under investigation. These sudden changes in the return distribution pose a challenge to every VaR model.

Throughout the past decade, the commodity futures market in India and abroad has witnessed a significant growth in terms of both network and volume. There are currently 19 commodity derivatives exchanges in India. In terms of total number of contracts traded, MCX has become the world's largest commodity futures exchange in gold and silver, second largest in natural gas, and third largest in crude oil. As per the annual report of chief regulator of commodity futures markets in India, Forwards Market Commission, the total size of commodity futures market was INR 101,447 billion in the financial year 2013-14. The monthly turnover in Indian commodity exchanges is next only to USA and China. There are currently 19 commodity derivatives exchanges in India. However, the bulk of trading (~99%) is concentrated in the national-level commodity exchanges namely Multi Commodity Exchange of India (MCX), National Commodity and Derivatives Exchange of India (NCDEX), National Multi Commodity Exchange (NMCE), Indian Commodity Exchange (ICEX), ACE Derivatives & Commodity Exchange Limited and Universal Commodity Exchange Limited.

Although there has been substantial research on evaluation of VaR models across various asset classes and geographical markets in the past two decades, the literature for commodity asset class in emerging economies is relatively thin. As commodities form a significant portion of trading portfolios of financial institutions, hedgers, speculators and retail investors, it becomes pertinent to find out an accurate VaR model for commodities. Keeping in view the distinct characteristics of commodities and dearth of literature on evaluation of VaR models for this asset class in India, this study aims to identify an accurate VaR model for nine commodities traded on MCX in India through backtesting.

The paper is structured as follows. The concept of VaR and backtesting are discussed next. In the following section, we review the existing literature on VaR methodologies and backtesting. In the third section we discuss the data used in the study. In the fourth section we describe different VaR models and procedures for measuring VaR adequacy used in the study. Fifth section shows the results of analysis and its interpretation. Last section presents the main conclusions of the study.

1.1 Value at Risk (VaR)

The mathematical roots of VaR were developed in the context of portfolio theory by Harry Markowitz and others in 1950's. Financial institutions began to construct their own risk management models in 1970's and 1980's, but it was not until the pioneering work from J.P. Morgan and their publication of Risk Metrics system in 1994 that made VaR the industry-wide standard (Dowd, 1998; Jorion, 2001). The Basel Capital Accord of 1996 played a significant role in the growth of VaR as it allowed banks to use their internal VaR models for computation of their regulatory capital requirements (Linsmeier & Pearson, 1996). Since then, VaR has been one of the most used measures of market risk. The concept of Value at Risk (VaR) has quickly become the standard for measurement of risk in the financial sector for both financial institutions and financial regulators (Engle & Manganelli, 1999) for its ease of use and clarity of meaning.

According to (Dowd, 1998), market risks can be subdivided into four classes: interest rate risks, equity price risks, exchange rate risks and commodity price risks. For this study, we intend to measure market risk through VaR for a portfolio consisting of commodities in India. By market risk, we specifically refer to exposure of the portfolio to losses due to changes in the prices of the constituent commodities. These changes may be caused by a variety of exogenous factors like variation in the demand and supply of commodities, trade barriers, input costs, tax rates etc. (Linsmeier and Pearson, 1996) give the following formal definition for VaR.

"Using a probability of x per cent and a holding period of t days, an entity's value at risk is the loss that is expected to be exceeded with a probability of only x per cent during the next t-day period."

Put simply, VaR measures the maximum loss for a portfolio over a predetermined time period for a given confidence interval. The VaR measure is oriented to act as a proxy for extreme downside risk. The basic idea behind VaR is straightforward since it gives a simple quantitative measure of portfolio's downside risk. VaR has two important and appealing characteristics. First, it provides a common consistent measure of risk for different positions and instrument types. Second, it takes into account the correlation between different risk factors. This property is essential whenever computing risk figures for a portfolio of more than one instrument (Dowd, 1998).

1.2 Backtesting

Despite the wide use and common acceptance of VaR as a risk management tool, the method has frequently been criticized for being incapable to produce reliable risk estimates. When implementing VaR systems, there will always be numerous simplifications and assumptions involved. Moreover, every VaR model attempts to provide a forward looking estimate of risk using historical data which does not necessarily reflect the market environment in the future. Thus, VaR models are useful only if they predict future risks accurately. In order to verify that the results acquired from VaR calculations are consistent and reliable, the models should always be backtested with appropriate statistical methods. Backtesting is a procedure where actual profits and losses are compared to projected VaR estimates. (Jorion, 2001) refers to these tests aptly as 'reality checks'. If the VaR estimates are not accurate, the models should be re-examined for incorrect assumptions, wrong parameters or inaccurate modelling. Selection of an appropriate VaR methodology to capture market risk accurately is of prime importance with regard to relevance and utility of VaR as a risk measurement and management tool.

A variety of different testing methods have been proposed for backtesting purposes. Basic tests, such as (Kupiec, 1995) POF-test, examine the proportion of losses in excess of VaR. This so called failure rate should be in line with the selected confidence level. For instance, if daily VaR estimates are computed at 99% confidence for one year (250 trading days), we would expect on average 2.5 VaR violations, or exceptions, to occur during this period. In the POF-test we then examine whether the observed amount of exceptions is reasonable compared to the expected probability of exceptions i.e. level of significance. This is the technique of backtesting that is used in our study to compare VaR models.

Backtesting is, or at least it should be, an integral part of VaR reporting in today's risk management. Without proper model validation one can never be sure that the VaR system yields accurate risk estimates. On the other hand, VaR is known to have severe problems in estimating losses at times of turbulent markets. As a matter of fact, by definition, VaR measures the expected loss only under normal market conditions (e.g. Jorion, 2001). This limitation is one of the major drawbacks of VaR and it makes the backtesting procedures very interesting and challenging, as will be shown later in the study.

2. Review of Literature

(Linsmeier & Pearson, 1996) provide introduction to VaR concept and discuss basic methodology. One of the earliest studies on VaR was made by (Allen, 1994) who compared the performance of historical simulation (HS) and variance-covariance approaches (under normal distribution). (Crnkovic and Drachman, 1995) have also compared the standard variance-covariance method and HS approaches with a metric developed by them. (Beder, 1995) applied eight common VaR methodologies and compared their performances. (Hendricks, 1996) and (Dave and Stahl, 1997) have conducted comparative evaluations of standard VaR models. (Jamshidian and Zhu, 1996) and (Jamshidian and Zhu, 1996) studied the efficiency of Monte Carlo methods compared to variance-covariance approach mainly for non-linear positions (such as options). (Danielsson and Morimoto, 2000) have examined the forecasting ability of a VaR model based on extreme value theory (EVT) in capturing the Japanese market risk.

Value at Risk by (Jorion, 2007) covers the introduction to the concept of VaR and overall picture of risk and risk measurement. (Christoffersen, 1998), (Christoffersen & Diebold, 2000) and (Christoffersen & Pelletier, 2004) in their work employ conditional coverage tests whereas (Lopez, 1998) introduces loss function. In 1996, Basel Committee implemented VaR approach for regulatory purpose and introduced fundamentals for a system of backtesting. (Bao, Lee and Saltoglu, 2006) investigate the predictive performance of various classes of VaR models in several dimensions and conclude that forecasting performance of the VaR models considered varies over the three periods before, during and after the crisis. (Ñíguez, 2008) assesses the forecasting performance of a wide range of models for predicting VaR in the Madrid Stock Exchange and finds that the Student's t FIAPARCH outperforms other models. (Abad and Benito, 2013) investigate the performance of different models of value at risk for several international indices for stable and volatile periods. Results show that parametric models can obtain successful VaR measures if conditional variance is estimated properly. (Bhat, 2015) compared the performance of alternative models for estimating Value at Risk (VaR) of four different currencies against the Indian rupee and found that VaR models based on an estimate of time-varying volatility performed better than traditional models during turbulent times

To the best of our knowledge, most empirical studies dealing with VaR calculation focused on market risk in equity and foreign exchange markets (Brooks and Persand, 2002); (Giot and Laurent, 2003b); (Giot and Laurent, 2004); (Huang and Lin, 2004) and (Chiu et al., 2005). In contrast, relatively few studies have attempted to evaluate VaR models for commodities. (Füss, Adams and Kaiser, 2010) examine the in and out-of-sample performance of various VaR approaches for investment in commodity futures. Results suggest that dynamic VaR models such as the CAViaR and the GARCH-type VaR models generally outperform traditional models.

3. Objectives

The purpose of this paper is to evaluate performance of different approaches and methods of VaR estimation by backtesting. Simply put, the objective of this study is to determine the accuracy of VaR models. How can we assess the accuracy and performance of a VaR model? To answer this question, we first need to define what we mean by "accuracy". By accuracy, we mean: How well does the model measure a particular percentile of the entire profit-and-loss distribution? This implies that if a VaR model provides a result of 5% as the maximum loss incurred on a portfolio at a confidence level of 99% on a daily basis, this loss of 5% should then be exceeded only once in a data set of 100 daily returns i.e. only 1% of the times (as the level of significance is 1%).

We use data of nine commodities to calculate VaR and perform backtesting. The results of this study may well be generalized for Indian commodity markets in general. The identification of suitable models to calculate VaR for commodities is useful for traders and investors who deal in commodities in India.

Another aim of this study is to verify whether the selection of similar or contrasting data periods in terms of market conditions for VaR calculation and VaR backtesting lead to a difference in the performance of VaR models. It shall also be interesting to know if this difference leads to preference of different models for VaR estimation.

The research hypothesis of the study is that VaR models differ in terms of their accuracy of estimation and this accuracy depends on the extent of congruence in the market conditions (and thus the returns of underlying asset) during VaR calculation and VaR backtesting periods.

4. Data

The data for this study has been taken from Bloomberg database for nine commodities, all of which are traded on the Multi Commodity Exchange (MCX SX) in India. Daily Closing Prices for all commodities were collected for a period of 9 years starting 25th October 2006 till 23rd October 2015. The data points for all nine selected commodities were made consistent by ensuring the inclusion of same dates for the nine year period before dividing it into pre-crisis and post-crisis sets.

VaR calculation and backtesting is performed on daily returns in the spot market for Aluminium, Zinc, Natural Gas, Nickel, Crude Oil, Gold, Silver, Sugar and Copper. This diversified set of commodities spread across categories of agriculture, precious metals, energy etc. is taken as a representative of commodities as an asset class in India.

Many theoretical financial models assume the financial return series to be normally distributed. This is a strong assumption because empirically, financial return series across assets have been found to exhibit negative skewness and excess kurtosis. Therefore it is pertinent to discuss the descriptive statistics of all nine commodities under study.

For the first part of analysis i.e. Pre-Crisis Evaluation, we calculate VaR over a 20 month period i.e. 25th October 2006 to 30th June 2008. The mean daily return ranges from -0.19% (Zinc) to 0.19% (Crude Oil) across all nine commodities for the 20 month period. The standard deviation ranges from 1.1% (Gold) to 3% (Natural Gas) and skewness ranges from -0.62 (Silver) to 0.23 (Sugar) across all 9 commodities for the 20 month period. Eight commodities have negatively skewed returns. The Kurtosis ranges from 3.02 (Zinc) to 54.06 (Sugar) across all nine commodities for the 20 month period. All kurtosis values are greater than 3 and indicate a leptokurtic distribution for return series of all commodities. A leptokurtic distribution has fatter tails as compared to a normal distribution. This means that there is a higher possibility of generating extreme returns as compared to a normal distribution. Fat tails in our return data are evident of the fact that there is higher possibility of generating extreme daily returns (negative or positive) in comparison to a normal distribution. The null hypothesis of normal distribution of returns is rejected for 7 commodities as Jarque-Bera test's p-value is close to 0 for these commodities. However, Zinc and Nickel have normally distributed returns.

For the second part of analysis i.e. Post-Crisis Evaluation, we calculate VaR over a four year period i.e. 1st January 2010 till 31st December 2013. The mean daily return ranges from 0% (Sugar) to 0% (Gold) across all nine commodities for the four year period. The standard deviation ranges from 1.02% (Gold) to 2.6% (Natural Gas) and skewness ranges from -1.15 (Silver) to 0.24 (Natural Gas) across all 9 commodities for the four year period. All commodities have negatively skewed returns except Crude Oil and Natural Gas. The Kurtosis ranges from 4.08 (Aluminium) to 49.8 (Sugar). The null hypothesis of normal distribution of returns is rejected for all nine commodities as Jarque-Bera test's p-value is close to 0 for all commodities.

In order to run GARCH (1,1) volatility estimation model, stationarity of parameters has been checked for all nine commodities for pre-crisis and post-crisis analysis. The null hypothesis of unit root (i.e. non-stationarity) in the Augmented Dickey Fuller test is rejected for all nine commodities for pre-crisis and post-crisis VaR calculations as the p value is less than 5% level of significance. Hence, all nine commodities have a stationary return distribution.

The details of descriptive statistics for pre-crisis and post-crisis VaR calculations are presented in Table 3 and Table 4 of Annexure 1 respectively.

5. Research Methodology

VaR is calculated using historical data. This implies that if VaR is calculated in a relatively tranquil period with positive returns on an average, it shall produce an estimate which is not suitable for a crisis situation having large negative returns and vice-versa. This indicates that VaR is heavily dependent on historical data and assumes that the returns in near future shall be consistent with the returns in the past. Keeping this in view, we decided to divide our nine years of daily return data into two parts: Pre-Crisis Evaluation and Post-Crisis Evaluation. Our total data starting in October 2006 and ending in October 2015 has a phase of severely distressed returns during the period of financial crisis of 2008-2009. This global meltdown owing to the sub-prime mortgage crisis in the US had its fair share of impact on India too.

According to National Bureau of Economic Research (NBER), the US business cycle reached its peak in December 2007 and trough in June 2009. To mark the crisis period accordingly, we have taken a period of 18 months starting 1st July 2008 till 31st December 2009 to represent this phase of financial distress. Although the financial crisis had started showing its symptoms in early 2008, we decided to mark its beginning in July 2008 as the crisis had a delayed and relatively weaker effect on the Indian markets. Similarly, although the trough might have been reached in June 2009, since the economy takes time to recover and equity returns stabilize gradually, we decided to add another 6 months and stretch the period of distress to 31st December 2009 making it a 18 month long phase of distressed returns.

VaR backtesting has been applied by adopting the intuitive out-of-sample approach. This means that VaR is first calculated over a period of time and then backtesting is applied ahead of this time to test whether VaR estimates are able to stand the test of time by predicting probability of violations accurately or not. This out-of-sample approach is appropriate to compare VaR methods because it examines the relevance of a VaR estimate as a forward looking risk measure. It helps us to identify which methods are able to produce future-oriented estimates on the basis of historical data. If a method sustains the out-of-sample backtesting approach and produces lesser number of violations than the significance level, it is an appropriate way of calculating VaR.

For the first part of analysis i.e. Pre-Crisis Evaluation, we calculate VaR over a 20 month period i.e. 25th October 2006 till 30th June 2008 and backtest the estimates over a 18 month period of financial crisis i.e. 1st July 2008 to 31st December 2009. This implies that we try to find out which methods produce accurate estimates of VaR (on the basis of historical data during period of relatively stable returns) which are able to withstand backtesting during the turbulent phase of financial crisis.

For the second part of analysis i.e. Post-Crisis Evaluation, we calculate VaR over a 4 year period i.e. 1st January 2010 till 31st December 2013 and backtest the estimates over a 22 month period i.e. 1st January 2014 to 25th October 2015. This implies that we try to find out if there is a difference between quality of VaR estimates produced by different methods if the VaR calculation period and backtesting period are largely similar in terms of economic conditions.

After dividing the data into relevant periods for a two-fold analysis, we calculate VaR by using following methods:

5.1Parametric VaR Methods

a. Normal VaR: VaR is calculated with the assumption that returns follow a normal probability distribution. The historical volatility is calculated as standard deviation of the return series and plugged in the formula for VaR. The one tailed standard normal variate corresponding to the chosen confidence level is multiplied by the volatility to calculate the extent of deviation from the mean to arrive at VaR figure.

$$VaR(\alpha) = F^{-1}(\alpha) = \mu_t + \sigma_t(z_\alpha)$$

where, μ represents mean, σ represents historical standard deviation, α denotes confidence level, z_α denotes one tailed z score corresponding to the chosen confidence level and t denotes time period considered.

b. GARCH (1,1) VaR: The only difference between Normal VaR and GARCH (1,1) VaR is the difference in calculation of volatility to be inserted in the VaR calculation. This methodology produces a forward looking estimate of volatility and hence can predict VaR more accurately if there is volatility clustering in the return series. In this method, σ is calculated using GARCH (1,1). GARCH (1,1) is estimated using following equations:

$$R_t = \mu_t + \varepsilon_t$$

$$\sigma_t^2 = \omega + \alpha_1[\varepsilon_{t-1}^2] + \beta_1\sigma_{t-1}^2$$

where, R_t represents return in period t, μ_t represents mean return and ε_t is error term in mean equation; σ_t is volatility, coefficient ω is constant, $\alpha_1$represents ARCH term and

β_1 represents GARCH term.

The selection of GARCH (1,1) among the family of GARCH models is somewhat arbitrary to an extent that we did not do any analysis as to which specification of GARCH would perform better for VaR forecasting. However, we believe that GARCH (1,1) is simpler, parsimonious and decently powerful in volatility estimation as it gives a good approximation to the observed temporal dependencies in daily data as documented by (Andersen and Bollerslev, 1998) and (Hansen and Lunde, 2005). Also, (Javed and Mantalos, 2013) claim that the first lag is sufficient to capture the movements of the volatility.

5.2 Non-parametric VaR Method

a. Historical VaR

In this model, VaR is estimated, without taking making any assumption about underlying return distribution. It is based on the assumption that history repeats itself. It is the simplest method of estimating VaR. The justification for this methodology is that if returns are stationary then empirical distribution is a consistent estimator of the unobserved future distribution function. This method is defined as follows.

Consider a sample of past ω returns. The historical VaR at α level of significance for period t+1 is given by: $VaR_{\alpha} = -Q_{1-\alpha}(r_t, r_{t-1,\ldots,}\, r_{t-\omega+1})$

where, r_t is return of the asset under consideration at time t and Q is the relevant quantile function at α level of significance.

b. Historical Simulation using Bootstrapping

Bootstrap historical simulation approach is an extension of traditional historical VaR model. It is a simple and intuitive estimation procedure. The bootstrap technique draws a sample from the data set, records the VaR from that particular sample and "returns" the data. This procedure is repeated over and over and records multiple sample VaRs. Since the data is always returned to the data set, this procedure is like sampling with replacement. The best VaR estimate from the full data set is the average of all sample VaRs.

5.3 Semi-Parametric Methods

The process combines the traditional simulation model with conditional volatility models like GARCH (1,1) which makes it attractive in dealing with volatility dynamics. In this category, following method is used for this study.

a. Volatility-Weighted Historical Simulation (VWHS)

This method was proposed by (Hull & White, 1998). It combines the benefits of Historical VaR with volatility updating. The basic premise of this approach is to update return information with recent changes in volatility. According to this approach, VaR (α) is the α quantile of the distribution of the volatility adjusted returns where α is the confidence level. In this methodology all past returns are scaled by volatility adjustment factor. Volatility adjusted returns are calculated as follows.

$$R_{t,i} = r_{t,i} * \frac{\sigma_{T,i}}{\sigma_{t,i}}$$

where, $r_{t,i}$ is actual return for asset *i* on day *t*; $\sigma_{T,i}$ is current forecast of volatility for asset *i*; $\sigma_{t,i}$ is volatility forecast for asset *i* on day *t* (made at the end of day *t*-1).

5.4 Backtesting of VaR Models

The most common backtesting method for a VaR model is to count the number of VaR exceptions, i.e. days (or holding periods of other length) when portfolio losses exceed VaR estimates. This method is known as the Kupiec's Test as it was suggested by (Kupiec, 1995). If the number of exceptions is less than what the selected confidence level would indicate, the system overestimates risk. On the contrary, too many exceptions signal underestimation of risk. Naturally, it is rarely the case that we observe the exact number of exceptions as suggested by the confidence level. It therefore comes down to statistical analysis to study whether the number of exceptions is reasonable or not, i.e. will the model be accepted or rejected.

Denoting the number of exceptions as x and the total number of observations as T, we may define the failure rate as x/T. In an ideal situation, this rate would be equal to the level of significance. For instance, if a confidence level of 99 % is used, we have a null hypothesis that the probability of tail losses is equal to p = (1 – c) = 1 − .99 = 1%. Assuming that the model is accurate, the observed failure rate x/T should act as an unbiased measure of p, and thus converge to 1% as sample size is increased. (Jorion, 2001)

Each trading outcome either produces a VaR violation exception or not. This sequence of 'successes and failures' is commonly known as Bernoulli trial. The number of exceptions x follows a binomial probability distribution:

$$f(x) = \frac{n!}{x!\,(n-x)!}\, p^{x}(1-p)^{n-x}$$

where *n* is number of trials; *x* is number of exceptions; *p* is probability of successes

As the number of observations increase, the binomial distribution can be approximated with a normal distribution:

$$z = \frac{x - pn}{\sqrt{p(1-p)n}} \approx N(0,1)$$

where *pn* is the expected number of exceptions and *p(1-p)n* is the variance of exceptions. (Jorion, 2001)

By utilizing this binomial distribution, we can examine the accuracy of the VaR model. However, when conducting a statistical backtest that either accepts or rejects a null hypothesis (of the model being 'good'), there is a trade-off between two types of errors. Type 1 error refers to the possibility of rejecting a correct model and type 2 error refers to the possibility of not rejecting an incorrect model. A statistically powerful test would efficiently minimize both of these probabilities. (Jorion, 2001)

Hence, the only information required to implement a POF-test is the number of observations (n), number of exceptions (x) and the confidence level (c). (Dowd, 2006)

6. Analysis and Interpretation

We have divided our analysis into separate two heads viz. Pre Crisis Evaluation and Post Crisis Evaluation. We shall start by studying the results of Pre Crisis Evaluation. The key phenomenon to keep in consideration is that for Pre Crisis Evaluation, we calculate VaR over a 20 month period i.e. 25th October 2006 till 30th June 2008 and backtest the estimates over a 18 month period of financial crisis i.e. 1st July 2008 to 31st December 2009. The VaR calculation and backtesting periods are contrasting in terms of their respective broad market conditions. Following is a table showing results of the Pre-Crisis Evaluation.

Table 1: Pre-Crisis Evaluation of VaR Models

Category of VaR Method	Non Parametric		Parametric		Semi Parametric
Null Hypothesis: x/T=Significance Level	**Historical VaR**	**Bootstrapped Historical VaR**	**Normal VaR**	**GARCH (1,1) VaR**	**VWHS VaR**
Hypotheses 'Not Rejected' out of 9	3	2	2	3	9
Success Ratio	33.33%	22.22%	22.22%	33.33%	100%

The null hypothesis being tested in the Kupiec test is that proportion of failure i.e. x/T (x is the number of violations and T is the total number of observations) is equal to level of significance. If this null hypothesis is not rejected, the VaR method passes the Kupiec Test. This is because the total number of violations observed in the testing period is not statistically different from the level of significance. The level of significance has been fixed at 5% for all hypotheses tests conducted in this study. For each VaR method, the success ratio has been defined as ratio of number of commodities for which the null hypotheses is not rejected to the total number of commodities i.e. nine.

As can be observed from the above table, out of the total 9 commodities for which VaR has been calculated and backtested, Volatility Weighted Historical Simulation (VWHS) VaR has been able to accurately predict VaR figures for all nine commodities i.e. the success ratio of VWHS VaR is 100%. This means that there is no difference between the proportion of failure observed in the backtesting period and the significance level of 5%. This shows that standardizing the daily log returns by GARCH volatility estimates produces a return series in such a way that it gives accurate VaR estimates.

Parametric VaR models viz. Normal VaR and GARCH (1,1) VaR have performed poorly in the backtest with a success ratio of 22.22% and 33.33% respectively. This is mainly due to the use of standard normal variate to determine VaR quantile whereas the return distribution is far from normal.

Non-Parametric methods viz. Historical VaR and Bootstrapped Historical VaR have also performed poorly in the backtesting analysis with a success ratio of 22.22% each.

The detailed table consisting of VaR estimates and result of hypothesis test for each pair of commodity and VaR method pertaining to pre-crisis evaluation is provided in Annexure II.

We now move on to the analysis of results of Post-Crisis Evaluation. In this case, we calculate VaR over a four year period i.e. 1st January 2010 till 31st December 2013 and backtest the estimates over a 22 month period i.e. 1st January 2014 to 25th October 2015.

The period of VaR calculation and VaR backtesting are broadly similar in the sense that none of the periods had experienced a situation of distressed returns. It is interesting to know whether different VaR methods perform similarly if there is no significant difference between the VaR calculation and VaR backtesting period in terms of return dynamics. Following is a table showing results of the Post-Crisis Evaluation.

Table 2: Post-Crisis Evaluation of VaR Models

Category of VaR Method	**Non Parametric**		**Parametric**		**Semi Parametric**
Null Hypothesis: x/T=Significance Level	**Historical VaR**	**Bootstrapped Historical VaR**	**Normal VaR**	**GARCH (1,1) VaR**	**VWHS VaR**
Hypotheses 'Not Rejected' out of 9	7	7	6	3	9
Success Ratio	77.78%	77.78%	66.67%	33.34%	100%

We observe that for post crisis evaluation, VWHS VaR has again produced a success ratio of 100%. In fact, Historical VaR and Bootstrapped Historical VaR also give highly accurate results with a success ratio of 77.78% each. Normal VaR exhibits a success ratio of 66.67%.

All methods except GARCH (1,1) VaR perform broadly well in this scenario where the data used for VaR calculation is quite similar in its behaviour to the data used for VaR backtesting. It is because of this similarity in results that our first analysis of pre-crisis evaluation becomes even more important to differentiate among VaR methods

However, VWHS VaR has emerged as a clear winner in both parts of the analysis. It has accurately predicted VaR estimates in pre-crisis and post-crisis evaluation delivering 100% success ratio each time. Hence, we conclude that VWHS VaR is a superior methodology out of the five different methods compared in this study for measuring VaR of a portfolio of commodities in India.

The detailed table consisting of VaR estimates and result of hypothesis test for each pair of commodity and VaR method pertaining to post-crisis evaluation is provided in Annexure III.

The results give an insight into the relative precision of the three different approaches of VaR modelling for commodities in India. However, the findings should be considered in light of the limitations of the study. Firstly, the study carries out a comparison of only five VaR models out of the umpteen diverse specifications available. There might be much more accurate models than VWHS VaR for VaR calculation for commodities. Also, more number of models can be tested simultaneously for each of the three approaches of VaR modelling. Secondly, there are more advanced techniques for backtesting than Kupiec's POF test like conditional coverage tests, loss functions etc. Each technique has its own advantages as it caters to different aspects of return distributions. The superiority of any model may not be sustained across all backtesting criteria. Thirdly, the cues for dates to divide the data period into different business cycles were taken from the information given on website of NBER which corresponds to business cycles in the US. The idea behind taking dates as per NBER was the emergence of global financial crisis in the US. The methodology of identifying structural breaks according to statistical break point test was not used to maintain consistency in the analysis of all nine commodities. A different period for crisis may lead to a change in the performance accuracy of VaR models. Lastly, VWHS VaR model has its own share of limitations as it augments the volatility trends in arriving at a VaR figure. A higher volatility forecast would lead to a higher VaR forecast and vice versa. This may lead to over-estimation or under-estimation of VaR if the volatility trend does not sustain in future as suggested by VWHS VaR estimate.

7. Conclusion

The variants of VaR are evident of its suitability and universal appeal as a standard measure of risk. It is a measure that needs to be reported by commercial banks as a regulatory compliance according to the Basel Norms. But the flexibility of VaR is also one of its limitations. Different approaches to calculate VaR are based on different assumptions. These assumptions relate to the underlying distribution of returns and volatility measurement. It is thus important to identify the most appropriate method of calculating VaR for different assets and markets. In this study we identified a relatively suitable VaR methodology for nine commodities traded on Multi Commodity Exchange of India. According to Kupiec's POF test of backtesting, VWHS VaR has outperformed other methods namely Historical VaR, Bootstrapped Historical VaR, Normal VaR and GARCH (1,1) VaR exhibiting a success ratio of 100% each time over a two-fold analysis separated by the financial crisis of 2008-09. In our study, Non-Parametric methods performed poorly in Pre-Crisis analysis but relatively well in Post-Crisis period. However, Parametric methods have performed poorly in accurately estimating VaR in both parts of the analysis. Hence, we conclude that VWHS VaR is a comparatively superior methodology out of the five different methods compared in this study for measuring VaR of investments in commodities in India. We also conclude that the selection of similar or contrasting data periods in terms of market conditions for VaR calculation and VaR backtesting affects the performance of VaR methods in general. However, it did not lead to a change in supremacy of VWHS VaR across the two periods of analysis in our study.

Bibliography

Abad Romero, P. and Benito Muela, S. (2013). "A Detailed Comparison of Value at Risk Estimates," *Mathematics and Computers in Simulation*, Vol. 94, pp. 258–276.

Berkowitz, J. and O'Brien, J. (2002). "How Accurate Are Value-at-Risk Models at Commercial Banks?" *Journal of Finance*, Vol. 57 No. 3, pp. 1093–1111.

Christoffersen, P. and Pelletier, D. (2004). "Backtesting Value-at-Risk: A Duration-Based Approach", *Journal of Financial Econometrics*, Vol. 2 No. 1, pp. 84–108.

Christoffersen, P.F. (1998). "Evaluating Interval Forecasts," *International Economic Review*, Vol. 39 No. 4, pp. 841–862.

Christoffersen, P.F. and Diebold, F.X. (2000). "How Relevant is Volatility Forecasting for Financial Risk Management?" *Review of Economics and Statistics*, Vol. 82 No. 1, pp. 12–22.

Engle, R.F. and Manganelli, S. (2004). "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles," *Journal of Business & Economic Statistics*, Vol. 22 No. 4, pp. 367–381.

Giot, P. and Laurent, S. (2004). "Modelling daily Value-at-Risk Using Realized Volatility and Arch Type Models," *Journal of Empirical Finance*, Vol. 11 No. 3, pp. 379–398.

Goorbergh, R. Van Den, Vlaar, P. and Bank, D.N. (1999). "Value-at-Risk Analysis of Stock Returns Historical Simulation , Variance Techniques or Tail Index Estimation ? ∗", *Techniques*, No. 2.

Huang, Y.C. and Lin, B.J. (2004). "Value-at-risk Analysis for Taiwan Stock Index Futures: Fat Tails and Conditional Asymmetries in Return Innovations," *Review of Quantitative Finance and Accounting*, Vol. 22 No. 2, pp. 79–95.

Kupiec, P.H. (1995). "Techniques for Verifying the Accuracy of Risk Measurement Models", *The Journal of Derivatives*, Vol. 3 No. 2, pp. 73–84.

Linsmeier, T.J. and Pearson, N.D. (1996). "Risk Measurement : An Introduction to Value at Risk," *Working Paper*, Vol. 6153 No. July, pp. 1–45.

Ñíguez, T.M. (2008). "Volatility and VaR forecasting in the Madrid Stock Exchange", *Spanish*

Economic Review, Vol. 10 No. 3, pp. 169–196.

Nolan, J.P. (2003). “Modeling Financial Data with Stable Distributions”, *Handbook of Heavy Tailed Distributions in Finance: Handbooks in Finance*, Vol. 1 No. May 2002, pp. 105–129.

Brooks, C. and Persand, G. (2002). "Model Choice and Value-at-Risk Performance," Financial Analysts Journal, 58(5), pp.87-97.

Giot, P. and Laurent, S. (2003). "Value-at-risk for Long and Short Trading Positions," Journal of Applied Econometrics, 18(6), pp.641-663.

Giot, P. and Laurent, S. (2004). "Modelling Daily Value-at-Risk Using Realized Volatility and ARCH Type Models," Journal of Empirical Finance, 11(3), pp.379-398.

Huang, Y. and Lin, B. (2004). Value-at-Risk Analysis for Taiwan Stock Index Futures: Fat Tails and Conditional Asymmetries in Return Innovations. Review of Quantitative Finance and Accounting, 22(2), pp.79-95.

Chiu, C., Lee, M. and Hung, J. (2005). Estimation of Value-at-Risk under jump dynamics and asymmetric information. Applied Financial Economics, 15(15), pp.1095-1106.

Allen, M. (1994). "Building a Role Model," Risk, 7, pp. 73-80.

Crnkovic, C. and Drachman, J. (1995). "A Universal Tool to Discriminate Among Risk Management Techniques," Corporate Risk Management Group, J.P. Morgan.

Hendricks, D. (1996), "Evaluation of Value- at-Risk Models Using Historical Data," Federal Reserve Bank of New York Policy Review, pp. 39-69.

Dave, R. and Stahl, G. (1997). "On the Accuracy of VaR Estimates Based on the Variance-Covariance Approach," SSRN Electronic Journal.

Jamshidian F. and Zhu, Y. (1996), "Scenario Simulation Model for Risk Management," Capital Market Strategies, 12, 26-30.

Jamshidian, F. and Zhu, Y. (1996). "Scenario Simulation: Theory and Methodology," Finance and Stochastics, 1(1), pp.43-67.

Danielsson, J. and Morimoto, J. (2000). “Forecasting Extreme Financial Risk: A Critical Analysis of Practical Methods for the Japanese Market,” IMES Discussion Paper Series-2000-E-8

Füss, R., Adams, Z. and Kaiser, D. (2010). "The predictive power of value-at-risk models in commodity futures markets," Journal of Asset Management, 11(4), pp.261-285.

Engel, J., Gizycki, M., & Authority, A. P. R. (1999). *Value at Risk: On the Stability and Forecasting of the Variance-covariance Matrix*. Economic Research Department, Reserve Bank of Australia

Frain, J. C. (2008). *Value at Risk (VaR) and the alpha-stable distribution* (No. tep0308).Trinity College Dublin, Department of Economics

Goorbergh, R. W. Van den, & Vlaar, P. J. (1999). *Value-at-Risk analysis of stock returns historical simulation, variance techniques or tail index estimation?*. De Nederlandsche Bank NV.

Angelidis, T. and Benos, A. (2006). “Liquidity Adjusted Value-At-Risk Based on the Components of the Bid-Ask Spread,” Applied Financial Economics, 16(11), pp.835-851.

Dowd, K. (1998). Beyond value at risk. Chichester: Wiley.

Engle, R. and Manganelli, S. (2004). “CAViaR,” Journal of Business & Economic Statistics, 22(4), pp.367-381.

Giot, P. and Laurent, S. (2004). “Modelling Daily Value-at-Risk Using Realized Volatility And ARCH Type Models,” Journal of Empirical Finance, 11(3), pp.379-398.

Jorion, P. (2001). Financial risk manager handbook 2001-2002. New York: John Wiley & Sons.

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

Lopez, J. (n.d.). “Methods for Evaluating Value-at-Risk Estimates,” SSRN Electronic Journal.

Bao, Y., Lee, T. and Saltoglu, B. (2006). “Evaluating Predictive Performance of Value-at-Risk Models in Emerging Markets: A Reality Check,” Journal of Forecasting, 25(2), pp.101-128.

Bhat, A. (2015). “A Test of Alternative Value-at-Risk Models during Volatile Periods,” Indian Journal of Finance, 9(8), p.19.

Andersen, T. and Bollerslev, T. (1998). “Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts,” International Economic Review, 39(4), p.885.

Hansen, P. and Lunde, A. (2005). “A Forecast Comparison Of Volatility Models: Does Anything Beat a GARCH(1,1)?” Journal of Applied Econometrics, 20(7), pp.873-889.

Javed, F. & Mantalos, P. (2013). “GARCH-Type Models and Performance of Information Criteria,” *Communications in statistics - Simulation and Computation*. 42(8), pp. 1917-1933.

Hull, J. and White, A. (1998). “Incorporating Volatility Updating Into the Historical Simulation Method for Value-At-Risk,” The Journal of Risk, 1(1), pp.5-19.

Beder, T. (1995). VAR: Seductive but Dangerous. Financial Analysts Journal, 51(5), pp.12-24.

Annexure I

Table 3: Descriptive Statistics: Pre-Crisis Analysis

	Aluminium	Zinc	Nat Gas	Nickel	Crude Oil	Gold	Silver	Sugar	Copper
Mean	0.000162	-0.00193	0.001065	-0.00115	0.001989	0.000921	0.000715	-0.0004	0.000163
Median	0	-0.00148	0.001244	0	0.001722	0.001146	0.002043	-0.00068	0
Maximum	0.059189	0.069733	0.150331	0.084943	3.801019	0.036604	0.063067	0.149812	0.060615

Minimum	-0.05428	-0.07702	-0.14885	-0.09421	-3.7991	-0.05336	-0.06549	-0.15059	-0.07817
Std. Dev.	0.015238	0.025049	0.030125	0.027153	0.319216	0.011137	0.015465	0.017508	0.02048
Skewness	-0.04656	-0.09857	-0.12903	-0.10444	-0.01016	-0.23718	-0.62743	0.238854	-0.26626
Kurtosis	4.118058	3.020456	6.064447	3.425164	117.6244	4.709059	6.345122	54.06256	4.258258
Jarque-Bera	22.13262	0.690677	166.2931	3.94561	231023.2	55.31555	224.4425	45850.53	32.82426
p-value	1.56E-05	0.707981	0	0.139066	0	9.74E-13	0	0	7.45E-08
Observations	422	422	422	422	422	422	422	422	422

Table 4: Descriptive Statistics: Post-Crisis Analysis

	Aluminium	**Zinc**	**Nat Gas**	**Nickel**	**Crude Oil**	**Gold**	**Silver**	**Sugar**	**Copper**
Mean	5.90E-05	7.38E-05	2.87E-05	2.34E-06	0.000512	0.000561	0.000479	-0.00027	0.000297
Median	0	0	0	0.000213	0	0.000281	0.000209	-0.00067	0
Maximum	0.052432	0.06908	0.131891	0.055417	0.10194	0.043873	0.079697	0.119553	0.069446
Minimum	-0.06417	-0.07102	-0.08894	-0.12118	-0.09084	-0.08793	-0.16449	-0.13803	-0.08647
Std. Dev.	0.013161	0.016074	0.026913	0.017081	0.018063	0.010265	0.019413	0.010886	0.016235
Skewness	-0.23743	-0.30612	0.244747	-0.5874	0.140535	-1.08827	-1.15023	-0.33158	-0.12134
Kurtosis	4.089999	4.158842	4.303992	6.021621	6.06696	12.39169	13.87831	49.80391	4.98099
Jarque-Bera	58.36932	70.9292	80.1058	433.9905	391.6616	3837.693	5104.879	90471.95	164.4736
p-value	2.11E-13	4.44E-16	0	0	0	0	0	0	0
Observations	991	991	991	991	991	991	991	991	991

Annexure II
Results of VaR Backtesting: Pre-Crisis Analysis

		Historical VaR	**Normal VaR**	**Historical Simulation VaR using bootstrap**	**GARCH (1,1) VaR**	**VWHS VaR**
Aluminium	VaR Value	0.039649206	0.03528631	0.03961610	0.03664894	0.04015181
	Decision	Rejected	Rejected	Rejected	Rejected	Accepted
Zinc	VaR Value	0.061452779	0.06020445	0.06032271	0.05916738	0.05633088
	Decision	Rejected	Rejected	Rejected	Rejected	Accepted

Natural Gas	VaR Value	0.077436654	0.06901550	0.07823509	0.05448986	0.0881792
	Decision	Rejected	Rejected	Rejected	Rejected	Accepted
Nickel	VaR Value	0.06928471	0.06431264	0.06878288	0.06103846	0.06446742
	Decision	Rejected	Rejected	Rejected	Rejected	Accepted
Crude Oil	VaR Value	0.050269268	0.74061953	0.04964276	0.12239652	0.37505467
	Decision	Rejected	Accepted	Rejected	Accepted	Accepted
Gold	VaR Value	0.02825	0.02498710	0.02756968	0.03644088	0.02616895
	Decision	Rejected	Rejected	Rejected	Accepted	Accepted
Silver	VaR Value	0.053810449	0.03526270	0.05269958	0.03793999	0.04812322
	Decision	Accepted	Rejected	Accepted	Rejected	Accepted
Sugar	VaR Value	0.026233337	0.04112591	0.04230372	0.02460733	0.04883971
	Decision	Accepted	Accepted	Accepted	Accepted	Accepted
Copper	VaR Value	0.064004284	0.04748062	0.06115966	0.04188975	0.08056324
	Decision	Accepted	Rejected	Rejected	Rejected	Accepted
Total number of Hypotheses 'Not Rejected'		**3**	**2**	**2**	**3**	**9**

Annexure III
Results of VaR Backtesting: Post-Crisis Analysis

		Historical VaR	**Normal VaR**	**Historical Simulation VaR using bootstrap**	**GARCH (1,1) VaR**	**VWHS VaR**
Aluminium	VaR Value	0.035303	0.030559	0.034199	0.025368	0.037855
	Decision	Accepted	Accepted	Accepted	Accepted	Accepted
Zinc	VaR Value	0.04353	0.037319	0.043557	0.025318	0.048607
	Decision	Accepted	Accepted	Accepted	Rejected	Accepted
Natural Gas	VaR Value	0.06732	0.06258	0.0651	0.045157	0.062674
	Decision	Rejected	Rejected	Rejected	Rejected	Accepted
Nickel	VaR Value	0.04398	0.039734	0.044098	0.02665	0.039712
	Decision	Accepted	Rejected	Accepted	Rejected	Accepted
Crude Oil	VaR Value	0.050034	0.041508	0.048497	0.028048	0.065038
	Decision	Rejected	Rejected	Rejected	Rejected	Accepted
Gold	VaR Value	0.031203	0.023318	0.030659	0.017384	0.033703
	Decision	Accepted	Accepted	Accepted	Accepted	Accepted
Silver	VaR Value	0.058474	0.044681	0.058811	0.038073	0.070489
	Decision	Accepted	Accepted	Accepted	Accepted	Accepted

Sugar	VaR Value	0.027175	0.025592	0.029269	0.006752	0.027743
	Decision	Accepted	Accepted	Accepted	Rejected	Accepted
Copper	VaR Value	0.044321	0.037473	0.043287	0.022661	0.056463
	Decision	Accepted	Accepted	Accepted	Rejected	Accepted
Total number of Hypotheses 'Not Rejected'		**7**	**6**	**7**	**3**	**9**