Theory article

VaR calculation by binary response models

Kasra Pourkermani*

Faculty of Economics and Management, Khorramshahr University of Marine Science and Technology, Khorramshahr, 64199, Iran

* **Correspondence:** Email: pourkermani@kmsu.ac.ir.

Abstract: The original Risk-Metrics method is underpinned by the assumption that daily asset returns are conditional Gaussian independently identically distributed (iid) random variables with a mean of zero. In this paper, a new method to calculate Value at Risk (VaR) was suggested to overcome the shortcoming of Risk-Metrics by employing binary response models to compute probability forecasts of the portfolio return by exceeding a grid of candidate quantile values. From those values, the VaR quantile value was selected. The proposed model was called BRV (Binary Response VaR method). Consistent application of BRV to the Dow Jones Industrial Average (INDEXDJX: DJI) and Dow Jones U.S. Marine Transportation Index (DJUSMT) time series proved that it was more accurate than the Risk-Metric system. This method not only worked similar to quantile regression but had the advantage that conventional maximum likelihood methods could be used for parameter estimation and inference. The BRV method was the best performing method for computing the daily VaR at both the 95% and 99% confidence levels over the period 02/01/06–31/12/08. The BRV and the QR (quantile regression) methods performed similarly, but the BRV method had the practical advantage that conventional maximum likelihood (ML) technique could be used for parameter estimation and robust inference.

1. Introduction

The Value at Risk (VaR) system falls into three major classes: The parametric method introduced by the Morgan Risk-Metric system, the non-parametric method based on historical simulation, and the semi-parametric method based on Extreme Value Theory (EVT) tail distribution. These methods are based on the good approximation of probability distributions extracted from an asset market price. The regulatory capital requirement for market risk is typically determined using the VaR for aggregate trading portfolio over a ten-day horizon and with a 99% confidence interval, and by the performance of the banks VaR models in backtesting exercises (Zumbach, 2007). Backtest is the best way to check the risk model performance. In backtesting, the estimated VaR is compared with the actual return over the same period. The VaR exceedance occurs when the return is more negative than the VaR. In order to backtest the accuracy for the estimated VaRs, we compute the empirical failure rates. By definition, the failure rate is the number of times returns exceed the forecasted VaR. If the model is correctly specified, the failure rate should be equal to the specified VaR level. In this paper, the backtesting VaR relies on the Christoffersen (2006) and Kupiec (1995) proposed system. Dumitrescu et al. (2012) have also proposed a backtest based on the non-linear regression model. Backtesting is a formal statistical framework that consists of verifying if actual trading losses are in line with model generated VaR forecasts and relies on testing over VaR violations. A violation is said to occur when the realized trading loss exceeds the VaR forecast.

The Risk Metrics method to compute VaR is set out in Zumbach (2006) and Jorge and Jerry (2001). The original Risk Metrics method is underpinned by the assumption that daily asset returns are conditional, Gaussian independently and identically distributed (iid) random variables with a mean of zero. Under this assumption, VaR at the one-day horizon can be computed by multiplying the relevant quantile of the standard normal distribution by the one day ahead forecast of the conditional standard deviation of the portfolio return and multiplying the result by the market-to-market value of the portfolio. A convenient consequence of the iid assumption is that VaR for longer horizons can be computed by multiplying the daily VaR by the square root of the time horizon on days. Other methods to compute VaR are the historical simulation method (HS) where relevant quantile from the empirical distribution of simulated returns is applied (see Jorion, 2007 for details on both methods). Emenogu et al. (2020) discovered that the persistence of the GARCH models is robust, with the exception of a few cases where IGARCH and EGARCH were unstable. The SGARCH and GJRGARCH models also failed to converge for t-student innovation, and the mean reverting number of days for returns varied between models. Altun (2020) also found that GARCH models listed under the TSLx innovation distribution produce more accurate VaR forecasts than other competing models. Slim et al. (2017) claimed that in developed markets, the related models show signs of long memories, suggesting that the FIGARCH model is preferable to the GARCH and GJR models. In frontier and emerging markets, the GJR and GARCH are the most important specifications for capturing risk. This means that when analyzing frontier markets, risk managers should favor models that account for asymmetry.

In practice, the Risk Metrics and HS methods to computing VaR have been the most popular methods within the industry. Note, however, that the HS method is highly restrictive since it ignores any additional explanatory information that might be thought to have an impact on the relevant return quantile (such as financial or macroeconomics information). This information can be incorporated into the VaR computed using the Risk Metrics method but only through the conditional mean or the conditional variance of return. The other weakness of the original Risk Metrics method is that it

assumes normality of the returns. In practice, empirical evidence suggests that for many financial prices, the conditional distributions is fat-tailed.

VaR can also be computed using quantile regression (QR) (Jorge and Jerry, 2001) (see for example (Taylor 1999, 2018)). When QR is used, the relevant return quantile can be modelled as a function of contemporaneous or lagged explanatory variables. In addition to this, QR based methods to VaR are more flexible than the Risk Metrics method since they do not require returns to be conditionally Gaussian. However, QR has several drawbacks. For example, the linearity assumption is highly questionable when data is heteroscedastic, (see Kupiec 1995, Peracchi 2002) and while nonlinear QR estimation techniques have been proposed, the asymptotic theory is not well developed. Furthermore, with both linear and nonlinear QR, the presence of heteroscedasticity can lead to estimated quantile that cross each other, and in both cases, robust inference typically requires bootstrapping, which increases the computational cost of the method. Tabasi et al. (2019) used GARCH models to model the volatility-clustering feature and found that using the t-student distribution function instead of the Normal distribution function improved model parameter estimation. Nieto and Ruiz (2016) compared the forecasting potential of various GARCH-based VaR models to their alternatives in an updated report. Surprisingly, they found that forecasting outcomes are affected by the number of out-of-sample observations as well as the time span being studied. They concluded that no single model outperforms another. Furthermore, only the asymmetric EGARCH-based model with skewed Student's-t distribution can be approved under the various model tests. Thavaneswaran et al. (2020) have introduced a volatility estimator applying an estimating function approach.

I propose an alternative parametric method to computing VaR. This alternative method allows the practitioner to utilize additional explanatory information to forecast the relevant quantile. However, relative to QR-based methods, the method proposed here has the significant practical advantage that the conditional maximum likelihood (ML) technique can be used for parameter estimation and robust inference.

The proposed method exploits the inverse relationship between the conditional quantile function (QF) and the conditional cumulative distribution (CDF), utilizing a technique for estimating the conditional CDF developed by Foresi and Peracchi (1995). I assume the market-to-market value of the relevant portfolio is one; hence, the daily VaR depends only on a one day ahead forecast of the relevant quantile for portfolio returns. Rather than directly forecast the VaR quantile, here it is proposed that the forecast is obtained indirectly using binary response models to compute probability forecasts over a grid of candidate quantile values. The candidate quantile value with an associated probability forecast closest to the desired probability (e.g. $p = 0.01$ for VaR at the 99% confidence level) is used as the VaR. This method is equivalent to forecasting points on the left tail of the conditional CDF and then inverting at the required VaR probability (Jelito and Pitera 2021).

Binary response models have previously been shown to be useful for forecasting the direction of asset returns. For example, Christoffersen and Diebold (2006) use a Logit model that conditions volatility as a predictor to forecast the direction of a time series index. Using binary response models to estimate points on the conditional CDF for stock returns has not previously been used for computing VaR. For brevity, we will refer to the method proposed here as the BRV (Binary Response VaR method). Ugurlu (2023) has also proposed a coherent multivariate average Var to quantify the total risk.

I compare the empirical performance of the BRV method with the orthodox Risk Metrics and HS method and a QR method in Monte Carlo simulations and an empirical application. The empirical application involves recursively computing daily VaR for two stock market indices, Dow Jones Industrial Average (INDEXDJX: DJI), and Dow Jones U.S. Marine Transportation Index (DJUSMT)

over a three-year period 02/01/06–31/12/18 using the previous five years of daily data for parameter estimation at each day. The results are analyzed using tests for correct unconditional coverage and independence of the VaR exceedances. Computing VaR over this period is challenging as the period begins with benign market conditions and ends with extremely volatile conditions associated with the global financial crisis that began in 2007. I found that the BRV method and QR method clearly dominate the Risk Metrics and HS method over this period. In particular, it appears that the Risk Metrics and HS method consistently underestimate the population VaR over this period since the proportion of VaR exceedances is too large for the given confidence levels. In contrast, exceedances when the BRV method is used are much closer to the expected number. Underestimating the population VaR can lead to serious penalties for banks operating in countries where the Basel ll Capital Accord has been implemented. Hence, this result is of practical importance.

2. Materials and methods

2.1. BRV method

Define the probability of the log portfolio returns R_t exceeding a threshold r_i conditional on a known k×1 vector of predictors X_{t-1} as,

$$P_{i,t} = P_t (R_t \leq r_i \mid X_{t-1}) = E (Y_{i,t} \mid X_{t-1}) \tag{1}$$

where $-\infty < r_i < \infty$ (i = 1, 2 ...., N) and $Y_{i,t}$ is a binary indicator,

$$Y_{i,t} = \begin{matrix} 1, & \text{if } R_t \leq r_i \\ 0, & \text{if } R_t > r_i \end{matrix} \tag{2}$$

Note that: $P_{i,t}$ can be interpreted as the value of the conditional CDF for R_t evaluated at r_i.

The VaR return quantile at the confidence level (1 − p) ×100%, 0 < p < 1, is the value r_i.
Such that $P_t(R_i \leq r_i \mid X_{t-1}) = p$. Let this be denoted by $Q_i (p)$. In practice the aim when computing VaR is to forecast the future value $Q_{T+h} (p)$. Throughout this paper we focus on computing daily VaR, so T denotes the current day and h = 1.
The BRV method to forecasting $Q_{T+1} (p)$. proposed here has three clear steps:
(i) Estimate multiple binary response models with the binary indicator (2) as the dependent variable over a range of candidate VaR return quantile, r_i , using conventional ML (the form of the link function in the binary response model and the location and number of values for r_i will be discussed below);
(ii) project the estimated binary response models forward to compute a one step ahead forecast of the probability of exceeding

$$r_i, \hat{p}_{i,T+1} = P_{T+1} (R_{T+1} \leq r_i \mid X_T); \tag{3}$$

(iii) as a forecast of the VaR quantile $Q_{T+1} (p)$ use the threshold r_i from (i) that minimizes the distance between the probability forecast $\hat{p}_{i,T+1}$ and the desired VaR probability p,

$$\hat{r}_i = \arg\min\{\hat{p}_{i,T+1} - p\} \tag{4}$$

The use of binary response models to estimate the conditional CDF for stock returns was proposed by Foresi and Peracchi (1995). Let $-\infty < r_1 < r_2 < \cdots < r_N < \infty$ be N feasible values of the return over the conditional CDF for R_t. Foresi and Peracchi (1995) show that points on the conditional CDF correspondence to $r_i (i = 1, 2, \ldots, N)$ can be estimated using a functional form that best approximates the population conditional CDF, F_t, as the link function in separate binary response models with the binary indicator (2) as the dependent variable. The "best approximation" is formalized as the approximation that minimizes the Kullback–Leibler divergence. Under weak regularity conditions, the parameters of the binary respons model can be consistently estimated by ML giving the estimated point $\hat{F}_{i,t}$.

Clearly, the conditional CDF should satisfy the standard condition,

$$0 < F_{i,t} < 1 \tag{5}$$

An attractive approach of the Foresi and Peracchi (1995) technique is that since it involves modelling the log-odds in $\left[F_{i,t} / (1 - F_{i,t}) \right]$ rather than $F_{i,t}$ directly, (5) is automatically satisfied. Foresi and Peracchi (1995) use a semi-parametric Logit model in their empirical work. In principle, any twice-differentiable CDF can be used. Note that monotonicity of the estimated CDF,

$$0 < \hat{F}_{i,t} < \cdots < \hat{F}_{N,t} < 1 \tag{6}$$

will not necessarily be satisfied if the ML estimation is unrestricted. Whether monotonicity is satisfied or not depends on numerous factors, including the sample size and the spread of the r_i values. In practice, even if monotonicity is violated, this might not have a serious detrimental impact on the practical performance of the technique. Monotonicity can be incorporated into the estimation algorithm if in practice it is a serious problem.

Foresi and Peracchi (1995) focus on estimating points on the conditional CDF using binary response models. Here, estimated binary response models are used to produce forecasts of the probability of exceeding candidate quantiles in the left tail of the conditional CDF (i.e., forecasts of points on the left tail of the conditional CDF). In the simulation and empirical work here, for simplicity, the cumulative normal and logistic CDFs are used. For example, when the logistic is used,

$$p_{i,t} = \frac{1}{1 + \exp(- \hat{X}_{t-1} \beta_i)} \tag{7}$$

Where β_i is a k×1 vector of parameters and X_{t-1} is the vector of predictors in (1). The conditional CDF $F_{i,t}$ can then be estimated by replacing B in (7) with the ML estimator $\hat{\beta}_i$.[1]

Step (ii) of the BRV method is to use the estimated parameters from (i) to compute a one-step ahead probability forecast for each candidate threshold r_i. Therefore, when the logistic CDF is used,

$$\hat{p}_{i,T+1} = \frac{1}{1 + \exp(- \hat{X}_T \hat{\beta}_i)} \tag{8}$$

where X_T is the vector of predictors at time T. Step (iii) of the BRV method involves finding the relevant VaR $\hat{r}_i = \arg\min\{\hat{p}_{i,T+1} - p\}$. This can be done using a linear computer grid search. To implement this method, the practitioner needs to decide on a functional form for the link function and on the total number of thresholds r_i to use the size of N) and on their location and spacing. In the empirical application below, for simplicity, the logistic CDF is used as a link function. Similar link functions can also be used (e.g., normal CDF, Student t CDF, Generalized Extreme Value CDF, semi-parametric link functions, etc.) and, in practice, backtesting over a historical sample period could be

---

[1] The asymptotic properties of the ML estimator for a Logit model are well-known and they are omitted.

employed to select the best performing link function from a set of candidate functions. In simulations and empirical applications discussed below, I found a grid of r_i values in backtesting, starting with the third value of the order statistics for the historical returns followed by the 1st, 3rd, 5th, 10th and 15th percentiles, which produce good results (thus N = 6). Cubic spline interpolation is then used to increase the number of forecasts and thresholds.

Clearly, if the link function has exactly the same form as the population conditional CDF for returns (e.g., normal-normal or logistic-logistic) and appropriate regressors are employed, then as $T \to \infty$, the Foresi and Peracchi (1995) method provides a consistent estimator of points on the left tail of the conditional CDF, providing that the regularity conditions required for ML to be a consistent estimator in this instance, are satisfied. To illustrate this in action, assume the following Data Generating Process (DGP) for log returns,

$$R_t = \gamma_1 X_{1,t} + \gamma_2 X_{2,t} + \varepsilon_t, \tag{9}$$

$$\varepsilon_t \sim L(0,1) \tag{10}$$
$$X_{1,t} = \theta_1 X_{1,t-1} + V_{1,t}, \qquad V_{1,t} \sim N(0,1) \tag{11}$$

$$X_{2,t} = \theta_2 X_{2,t-1} + V_{2,t}, \qquad V_{2,t} \sim N(0,1) \tag{12}$$

where L(0,1) and N(0,1) denote a logistic and normal distribution with mean of zero and a variance of one. Therefore, conditional on the X variables generated by AR(1) models, returns have a logistic distribution. We simulate representative series of returns from this general DGP. For one set the following parameter values are used; $\gamma_1 = 0.50$, $\gamma_2 = 0$, $\theta_1 = 0.30$ For the other set; $\gamma_1 = 0.50, \gamma_2 = 0.50, \theta_1 = 0.30, \theta_2 = 0.30$. Therefore in the first set of simulations the model contains a single stationary regressor, while in the second set the model contains two stationary regressors. Observations from (9)–(12) are simulated for the following sample sizes, T = 100, 200, 500, 1000, 10000. For each series, we then estimate the left tail of the conditional CDF using the method of Foresi and Peracchi (1995) employing the logistic CDF as the link function over a grid of r_i values starting with the third value of the order statistics for the historical returns followed by the 1st, 3rd, 5th, 10th and 15th percentile values.[2] In both cases, a constant and the correct explanatory variables are used in the link function.

2.2. Monte carlo simulation results

Table 1. Simulated $\hat{p}$ distribution 95% confidence.

Percentile	5th	25th	50th	75th	95th
Gaussian-GARCH DGP					
1-BRV	0.021	0.030	0.042	0.063	0.074
t-GARCH DGP					
2-BRV	0.014	0.036	0.049	0.079	0.128
AR-Gaussian- GARCH DGP					
3-BRV	0.019	0.028	0.062	0.064	0.090
AR-t GARCH DGP					
4-BRV	0.022	0.023	0.042	0.082	0.122

---

[2] Note that here the model is not predictive since the explanatory variables are current values which we make no attempt to forecast. In the Monte Carlo simulations and empirical application discussed below, which involves forecasting, lagged values of the explanatory variables are used.

For each replication VaR is computed using QR, along with the original Risk Metrics and HS methods, the true volatility is used when computing the Risk Metrics VaR.

To assess the finite-sample performance of each method, for each replication the estimated unconditional coverage is computed $\hat{p}$. The 5th, 25th, 50th, 75th and 95th percentiles of the empirical distribution of $\hat{p}$ are reported in Table 1 for the 95% confidence level and Table 2 for the 99% confidence levels. In Table 1 and 2, these are reported in four rows: The first row is Gaussian GARCH data generating process (DGP), the second row is threshold GARCH data generating process, the third row contains autoregressive Gaussian GARCH data generating process, and the fourth-row reports autoregressive threshold GARCH date generation process. The selection of a confidence level for an interval determines the probability that the confidence interval produced will contain the true parameter value. Common choices for the confidence level are 0.95, and 0.99. These levels correspond to percentages of the area of the normal density curve. For example, a 95% confidence interval covers 95% of the normal curve-the probability of observing a value outside of this area is less than 0.05. Because the normal curve is symmetric, half of the area is in the left tail of the curve, and the other half of the area is in the right tail of the curve.

Table 2. Simulated $\hat{p}$ distribution 99% confidence.

Percentile	5th	25th	50th	75th	95th
Gaussian-GARCH DGP					
BRV	0.000	0.007	0.011	0.014	0.026
t-GARCH DGP					
BRV	0.000	0.003	0.013	0.016	0.030
AR-Gaussian-GARCH DGP					
BRV	0.000	0.005	0.013	0.016	0.024
AR-t GARCH DGP					
BRV	0.000	0.004	0.013	0.021	0.040

The second-fourth simulation experiments are the same as the first but with different DGPs for the returns, allowing for serial correlation and conditionally non-Gaussian returns. The various DGPs for all the experiments are given below:

DGP 1. Gaussian-GARCH

$$R_t = \varepsilon_t h_t^{1/2} \, , \; h_t = 0.1 + 0.1\, R_{t-1}^2 + 0.8\, h_{t-1}\, , \varepsilon_t \sim N\,(0,1) \tag{13}$$

DGP 2. t-GARCH.

$$R_t = v_t h_t^{1/2} \, , \; h_t = 0.1 + 0.1\, R_{t-1}^2 + 0.8\, h_{t-1}\, , v_t \sim t\,(5) \tag{14}$$

DGP 3. AR-Gaussian-GARCH

$$R_t = 0.3\, R_{t-1} + u_t\, , u_t = \varepsilon_t\, R_t^{1/2}\, , \varepsilon_t \sim N\,(0,1) \tag{15}$$

$$h_t = 0.1 + 0.1\, u_{t-1}^2 + 0.8\, h_{t-1} \tag{16}$$

DGP 4. AR – t – GARCH

$$R_t = 0.3\, R_{t-1} + u_t\, , u_t = v_t\, R_t^{1/2}\, , v_t \sim t\,(5) \tag{17}$$

$$h_t = 0.1 + 0.1\, u_{t-1}^2 + 0.8\, h_{t-1} \tag{18}$$

For DGPs 1 and 3, a normal CDF is used as the link function when using the BRV method (*i.e.* Probit models are estimated). For DGPs 2 and 4, a logistic CDF is used as the link function (Logit models are estimated). For DGPs 1 and 2, just a constant is included as a predictor in the relevant link function, and for DGPs 3 and 4, the link function also contains a lag of returns (hence, the estimated models are correctly specified).

For the 95% confidence level, it can be seen in Table 1 that the empirical distribution of $\hat{p}$ for both the BRV and QR methodes are virtually identical. They both have good levels of unconditional coverage given the small size of the backtesting period (250 observations). The distribution of $\hat{p}$ is centered close to the population value of p=0.05, irrespective of the DGP. In both cases, the performance of these methods is similar, irrespective of whether returns are conditionally Gaussian or non-Gaussian. It can be shown that the logistic CDF closely approximates a student t CDF with 9 degrees of freedom (see Mudholkar and George (1978). Hence, the BVR method using Logit models is clearly well-suited to computing VaR if the conditional distribution for returns is thought to be fat tailed. The results in Table 2 show that at the 99% confidence level, the BRV and QR methods also produce very similar results. Again, the distributions of $\hat{p}$ are centered close to the population value of *p= 0.01* and both have a similar variance.

In contrast, however, it can be seen that at both confidence levels, the Risk Metrics method significantly underestimates the population VaR for DGPs 2 and 4 (non-Gaussian returns). For DGP 3, (Gaussian returns but with serial correlation), the Risk Metrics method, which ignores serial correlation, slightly underestimates the population VaR. The HS method gives similar results to the BRV and QR methods for all the DGPs at both confidence levels, and there is no distinguishable difference in the performance of the HS method for the DGPs with and without serial correlation.

3. Results

In this section, I discuss an empirical application involving DJI and DJUSMT series. The application involves recursively computing the daily VaR at the 95% and 99% confidence levels using the BRV, QR, Risk Metrics, and HS methods for every trading day over the three-year period 02/01/06–31/12/08 (755 days), using a five-year window of historical data for parameter estimation (approximately 1250 observations). For example, VaR on 02/01/06 is computed using data from 31/12/00–31/12/05. Note that the parameters of the BRV and QR models are re-estimated each day. The conditional standard deviation $h_{t-1}^{1/2}$ is chosen as an important predictor following the evidence in Christoffersen and Diebold (2006) on its ability to forecast the direction of stock returns, and a positive relationship with $p_{i,t}$ is expected. TB_{t-1} is included to allow for present value effects, and V_{t-1} is included to capture market sentiment, and again for both, a positive relationship with $p_{i,t}$ is expected. Conventional ML is used for parameter estimation in the BRV method and the interior point algorithm of Koenker and Park (1978) is used for parameter estimation in the QR method. In the Risk Metrics method, an EWMA volatility forecast with a weight of 0.94 is used, which is the default choice for the Risk Metrics method applied to daily data.

Prior to discussing the backtesting results, as an example, the estimated Logit model parameters and robust t-statistics computed using Huber-White robust standard errors are given in Table 3 for each index and stock at three points over the backtesting period (the points are 29/12/06, 31/12/07, and 30/12/08). In each case, the 1st percentile of the order statistics is used to define the threshold. On the basis of the robust t – statistics, I found clear evidence that all three explanatory variables are statistically significant at one or more of these points, and that when statistically significant, the signs of the estimated parameters are consistent with our expectations. Note that in Table 3, $\hat{\beta}_0$ is the

estimated constant, which are -13.162 for DJI and -10.708 for DJUSMT, and $\hat{\beta}_1$ is the estimated parameter on TB_{t-1}, which are 0.518 for DJI and -0.018 for DJUSMT. $\hat{\beta}_2$ is the estimated parameter on V_{t-1}, which are 0.119 for DJI and 0.107 for DJUSMT. $\hat{\beta}_3$ is the estimated parameter on $h_{t-1}^{1/2}$, which are 2.617 for DJI and 2.551 for DJUSMT. Robust t-statistics computed using Huber-White standard errors are in parentheses.

Table 3. Logit model estimation results.

Sample end	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
DJI				
1-31/12/08	-13.162	0.518	0.119	2.617
DJUSMT				
2-31/12/08	-10.708	-0.018	0.107	2.551

As one might expect, the exact statistical significance of the estimated parameters varies depending on the index or stock and sample period, but there are some patterns. For example, for each series, I found that $h_{t-1}^{1/2}$ is strongly statistically significant, but that the statistical significance of TB_{t-1} and V_{t-1} varies over the sample. Note that I do not eliminate statistically insignificant predictors prior to computing VaR using the BRV method; however, this method could be taken in future research to allow for structural change.

4. Backtesting results

To summarize the backtesting results, for each method and at each confidence level, the estimated unconditional coverage $\hat{p}$ is reported for comparison with the population unconditional coverage p. Christoffersen (1998) proposes a complete methodology for evaluating the number of exceedances and their independence. The independence test rationale dictates that, if the violations are dependent, then the transition probabilities would not be equal. Finally, Christoffersen (1998) proposes a joint test that combines both hypotheses (Conditional Coverage CC hypothesis). Correct unconditional coverage LR_{uc} and independence of the VaR exceedances LR_{ind} are proposed by Christoffersen (1998). The LR_{uc} and LR_{ind} tests utilize the fact that if the VaR method is perfect, then VaR exceedances should be unpredictable and so a binary indicator of exceedances (the hit indicator),

$$H_{t+1} = \begin{array}{ll} 1, & if \quad R_{t+1} < \hat{Q}\,(p)_{t+1} \\ 0, & if \quad R_{t+1} \geq \hat{Q}\,(p)_{t+1} \end{array} \tag{19}$$

where $\hat{Q}\,(p)_{t+1}$ is the forecast of the relevant return quantile, should be an independent Bernoulli random variable. The LR_{uc} and LR_{ind} tests are straightforward to compute and have a x^2 (1) asymptotic distribution (see Christoffersen, 1998, for further details).

The backtesting results for the DJI at the 95% and 99% confidence levels, respectively, are given in Table 4. The results for the method that is optimal on the basis of the estimated unconditional coverage $\hat{p}$ relative to the population value p are bolded. For DJI, at the 95% confidence level, the BRV result is $\hat{p} = 0.05$, suggesting that the population VaR over this period is estimated extremely well ($p=0.05$). In contrast, the Risk Metrics, HS, and QR results are $\hat{p} = 0.061$, $\hat{p} = 0.110$, and $\hat{p} = 0.052$, respectively, suggesting that the population VaR is underestimated by each of these methods. The LR_{ac} test rejects the null hypothesis of correct unconditional coverage for the Risk Metrics and

HS methodes at either the 5% or 1% significance levels. For DJI at the 99% confidence level, the Risk Metrics result is $\hat{p}$ = 0.029 while the HS result is $\hat{p}$ = 0.045, and the QR result is $\hat{p}$ = 0.034.

Table 4. Backtesting results: DJI.

	$\hat{p}$	LR_{uc}	LR_{ind}
95% confidence			
Risk Metrics	0.061	5.092 **	1.086
BRV	0.050	0.002	0.681
QR	0.052	1.037**	0.272
HS	0.110	30.091 **	0.469
99% confidence			
Risk Metrics Metrics	0.029	18.439 *****	1.382
BRV	0.010	0.024	0.634
QR	0.034	5.095 **	0.561
HS	0.045	28.333 **	1.104

Note that: *, **, and *** indicate a rejection of the null hypothesis at the 10%, 5%, and 1% levels, respectively.

On the basis of the estimated unconditional coverage $\hat{p}$, in both cases, the BRV method is superior to any of the other methodes considered. For DJUSMT, at the 95% confidence level, the BRV result is $\hat{p}$ = 0.05, suggesting that the population VaR over this period is estimated extremely well ($p=0.05$). In contrast, the Risk Metrics, HS, and QR results are $\hat{p}$ = 0.065, $\hat{p}$ = 0.088, and $\hat{p}$ = 0.032, respectively, suggesting that the population VaR is underestimated by each of these methods. The LR_{ac} test rejects the null hypothesis of correct unconditional coverage for the Risk Metrics and HS methodes at either the 5% or 1% significance levels. For DJUSMT at the 99% confidence level, the Risk Metrics result is $\hat{p}$ = 0.024 while the HS result is $\hat{p}$ = 0.053, and the QR result is $\hat{p}$ = 0.19. Therefore, again, in all three cases, these results suggest that the population VaR is underestimated. Furthermore, in all three cases, LR_{uc} rejects the null hypothesis of correct unconditional coverage at conventional significance levels. The BRV result is $\hat{p}$ = 0.015, which is much closer to the desired level of coverage.

Table 5. Backtesting results: DJUSMT.

	$\hat{p}$	LR_{uc}	LR_{ind}
95% confidence			
Risk Metrics Metrics	0.065	3.239 *	0.705
BRV	0.021	0.543	5.111 **
QR	0.032	0.067	0.265
HS	0.088	21.81 ***	7.002 **
99% confidence			
Risk Metrics Metrics	0.023	8.817 ***	0.830
BRV	0.015	0.894	0.365
QR	0.19	6.655***	0.574
HS	0.053	36.339***	9.55**

The results for the DJI and DJUSMT indices in Table 5 show that at the 95% confidence level, the QR method is preferred on the basis of the estimated unconditional coverage with the BRV method being the next most accurate. At the 99% confidence level, the BRV method is preferred. Again, for the other method considered $\hat{p} > p$ and the rejections obtained from LR_{uc}, it is suggested that the population VaR is underestimated. In this case, rejections are also obtained from LR_{ind} for the BRV, QR, and HS methods at the 95% confidence level, suggesting mis-specified models.

The optimal method on the basis of the estimated unconditional coverage $\hat{p}$ are either the BRV method (optimal in four out of six cases) or the QR method (optimal in the remaining two cases).

5. Discussion

Underestimating the population VaR can lead to serious penalties for banks in countries where the Basel II Capital Accord has been implemented and is a well-known weakness of the Risk Metrics method if the population conditional distribution is fat-tailed.

6. Conclusions

I propose an alternative parametric method to computing the widely used financial risk measure VaR. The BRV method involves using binary response models to compute probability forecasts of the portfolio return exceeding a grid of candidate quantile values. The candidate quantile value associated with a probability forecast closest to the desired VaR probability is chosen as the VaR. The performance of the BRV method is impressive relative to the orthodox Risk Metrics and HS methods and a QR-based method, both in Monte Carlo simulation experiments and an empirical application involving DJI and DJUSMT indexes. In the empirical application, the BRV method is the most accurate method in most cases on the basis of the estimated unconditional coverage. Note in particular that the BRV method is the best performing method for computing the daily VaR at both the 95% and 99% confidence levels over the turbulent period 02/01/06–31/12/08. The BRV and QR methods perform similarly, but relative to QR, the BRV method has the practical advantage that conventional ML methods can be used for parameter estimation and robust inference.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declare no conflict of interest.

References

Altun M (2020) A new approach to Value-at-Risk: GARCH-TSLx model with inference. *Comm Stat Simu Comput* 49: 3134–3151. https://doi.org/10.1080/03610918.2018.1535069

Christoffersen PF (1998) Evaluating Interval Forecasts. *Int Econ Review* 39: 841–862. https://doi.org/10.2307/2527341

Christoffersen PF, Diebold FX (2006) Financial Asset Returns, Direction-of-Change Forecasting and Volatility Dynamics. *Manag Sci* 52: 1273–1287. https://doi.org/10.1287/mnsc.1060.0520

Dumitrescu EI, Hurlin C, Pham V (2012) Backtesting Value-at-Risk: From Dynamic Quantile to Dynamic Binary Tests. *Halshs*. https://shs.hal.science/halshs-00671658

Emenogu NG, Adenomon MO, Nweze NO (2020) On the volatility of daily stock returns of Total Nigeria Plc: evidence from GARCH models, value-at-risk and backtesting. *Financ Innov* 6: 18. https://doi.org/10.1186/s40854-020-00178-1

Engle RF, Manganelli S (2004) CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *J Bus Econ Stat* 22: 367–381. http://www.jstor.org/stable/1392044

Foresi S, Peracchi F (1995) The Conditional Distribution of Excess Returns: An Empirical Analysis. *J Amer Statist Assoc* 90: 451–466. https://doi.org/10.1080/01621459.1995.10476537

Huber PJ (1967) The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* 1: 221–233.

Jelito D, Pitera M (2021) New fat-tail normality test based on conditional second moments with applications to finance. *Stat Pap* 62: 2083–2108. https://doi.org/10.1007/s00362-020-01176-2

Jorge M, Jerry X (2001) Return to Risk Metrics: The Evolution of a Standard. Risk Metrics Group Inc.

Jorion P (2007) Value at Risk: The New Benchmark for Managing Financial Risk. 3. McGraw-Hill, New York.

Koenker R, Bassett G (1978) Regression Quantiles. *Econometrica* 46: 33–50. http://dx.doi.org/10.2307/1913643

Koenker R, Park BJ (1966) An interior point algorithm for nonlinear quantile regression. *J Econometrics* 71: 265–283. https://doi.org/10.1016/0304-4076(96)84507-6

Kupiec P (1995) Techniques for Verifying the Accuracy of Risk Measurement Models. *J Deriv* 3: 73–84. http://dx.doi.org/10.3905/jod.1995.407942

Mudhokar GS, George O (1978) A remark on the shape of the logistic distribution. *Biometrika* 65: 667–668. https://doi.org/10.1093/biomet/65.3.667

Nieto MR, Ruiz E (2016) Frontiers in VaR forecasting and backtesting. *Intl J Forecasting* 32: 475–501. https://doi.org/10.1016/j.ijforecast.2015.08.003

Peracchi F (2002) On estimating conditional quantiles and distribution functions. *Comput Statis Data Anal* 38: 433–447. https://doi.org/10.1016/S0167-9473(01)00070-6

Slim S, Koubaa Y, BenSaida A (2017) Value-at-Risk under Lévy GARCH models: Evidence from global stock markets. *J Int Financ Mark I* 46: 30–53 https://doi.org/10.1016/j.intfin.2016.08.008

Tabasi H, Yousefi V, Tamošaitienė J, et al. (2019) Estimating Conditional Value at Risk in the Tehran Stock Exchange Based on the Extreme Value Theory Using GARCH Models. *Adm Sci* 9: 40. https://doi.org/10.3390/admsci9020040

Taylor JW (1999) A quantile regression method to estimating the distribution of multi period returns. *J Deriv* 7: 64–78. https://doi.org/10.3905/jod.1999.319106

Taylor JW (2008) Using Exponentially Weighted Quantile Regression to Estimate Value at Risk and Expected Shortfall. *J Financ Economet* 6: 382–406. https://doi.org/10.1093/jjfinec/nbn007

Thavaneswaran A, Paseka A, Frank J (2020) Generalized value at risk forecasting. *Commun Stat-Theor M* 49: 4988–4995. https://doi.org/10.1080/03610926.2019.1610443

Ugurlu K (2023) A new coherent multivariate average-value-at-risk. *Optimization* 72: 493–519. https://doi.org/10.1080/02331934.2021.1970755

White H (1980) A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica* 48: 817–838. https://doi.org/10.2307/1912934

Zumbach G (2006) The Risk Metrics 2006 Methodology. *Available at SSRN*. http://dx.doi.org/10.2139/ssrn.1420185