Risky Business: Quantifying Value at Risk

Sierra Laine
University of North Carolina at Charlotte

Abstract

Value-at-Risk (VaR) is one of the best known and most heavily used measures of financial risk, as it denotes the dollar amount one stands to lose for a given investment with a certain probability. In this project, we demonstrate the performance of several commonly used methods for the estimation of VaR. We use five major market indices: the S&P 500 Index, the Nikkei 225 Index, the Dow Jones Industrial Average, the Russell 2000 Index, and the Bitcoin Price Index, for two significant time periods. The one is during the financial crisis, a period with more market irregularities (2004-2009), and the other is the period followed the financial crisis, a period with more financial stability (2011-2016). The purpose of this paper is to compare the performance of the different methods for those different periods.

1 Introduction

The financial crisis of 2007-2008 highlighted the importance of accurate estimation of financial risk. In this project we focus on the estimation of Value-at-Risk (VaR), which is one of the best known and most heavily used measures of financial risk. VaR is a statistic that quantifies the level of financial risk over a specific time frame and can be easily converted into the dollar amount that one stands to lose from a given investment. Specifically, let $\{r_t\}_{t=1}^n$ be a time series of returns. For $\tau \in (0,1)$, the τth VaR at time t, denoted by $\mathrm{VaR}_\tau(t)$, is the smallest number for which $P\{r_t < -\mathrm{VaR}_\tau(t)|\mathcal{F}_{t-1}\} = \tau$, where $\mathcal{F}_{t-1}$ denotes the information set at time $t-1$. The parameter τ is typically chosen to be a small number such as 0.01, 0.025, or 0.05. It is clear that $-\mathrm{VaR}_\tau(t)$ is the τth conditional quantile of r_t.

There is considerable literature on VaR estimation; see e.g. Duffie and Pan (1997) for an overview. Many methods are based on first modeling the time series of returns, and then using the conditional distribution of r_t to calculate VaR. In this case, the dependence structure is most commonly modeled using an ARCH model (Engle 1982) or one of its variants (GARCH, ARMA-GARCH, Bollerslev 1986; linear GARCH, Taylor 1986), and the conditional distribution is modeled by a parametric family, typically the normal distribution. This is the approach adopted, for instance, in the well-known framework developed by RiskMetrics (RiskMetrics Group 1996), which uses a particular type of GARCH(1,1) model with normal innovations. However, it is well-known that financial returns exhibit non-normal characteristics including negative skewness, excess kurtosis, and heavy tails. For this reason, more flexible parametric families are sometimes used. In another direction, interest has turned to quantile regression methods, as they are robust to extremes and make no distributional assumptions. Koenker and Xiao (2006) introduced the quantile autoregression (QAR) model, while, Engle and Manganelli (2004) introduced a more general model, named the conditional autoregressive value-at-risk (CAViaR).

In this project, we review and compare existing techniques for VaR estimation. More interestingly, we investigate how commonly used methods perform under unusual financial patterns and specifically, which methods were able to better 'predict' the financial crisis. In Section 2 we present the different methods considered in this project, while in Section 3 we present examples with real-world data. A brief discussion is given in Section 4.

2 Commonly Used Methods for VaR Estimation

The methods used in this project include the Historical method, the GARCH(1,1) model with different distributions for the innovations, the distribution free GARCH(1,1) model, and the quantile autoregression method (QAR). Below we explain the aforementioned methods in more detail.

2.1 Historical Method

The historical method estimates VaR using empirical quantiles. More specifically, the τth VaR at time t is given by the negative τth sample quantile of the returns prior to r_t. Its computational simplicity makes historical method a very attractive approach. However, the historical method is not very adaptive to changes on the market, which makes it a conservative method under extreme volatile markets.

2.2 GARCH Model

The GARCH(1,1) model assumes that

$$r_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \omega + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{1}$$

where ω, α_1, and β_1 are unknown parameters and $\epsilon_1, \epsilon_2, \ldots$ are independent and identically distributed random variables. Then, $\mathrm{VaR}_\tau(t) = -\sigma_t F^{-1}(\tau)$, where $F(\cdot)$ denotes the cumulative distribution function (cdf) of ϵ.

Modifications of the GARCH(1,1) model include using different distributions for the error term, with the most common one to be the standard normal distribution. In that case, $\mathrm{VaR}_\tau(t) = -\sigma_t \Phi^{-1}(\tau)$, where $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution. This paper also considers the skewed normal distribution.

2.3 DFGARCH Model

This distribution free version of the GARCH(1,1) model assumes that (1) holds, but no distributional assumptions on the error term are made. Instead, the $\mathrm{VaR}_\tau(t)$ is estimated by $\sigma_t Q_\tau$, where Q_τ is the τ empirical quantile of the standardized residuals.

2.4 QAR Model

Quantile Regression (QR) was first introduced by Koenker and Bassett (1978) and has been an attractive alternative when interest lies in certain conditional quantiles. For a univariate response Y and a d-dimensional vector of covariates $\mathbf{X}$, let

$$Q_\tau(Y|\mathbf{x}) \equiv Q_\tau(Y|\mathbf{X} = \mathbf{x}) = \inf\{y : P(Y \leq y|\mathbf{X} = \mathbf{x}) \geq \tau\}$$

denote the τth conditional quantile of Y given $\mathbf{X} = \mathbf{x}$. Koenker and Bassett (1978) considered the linear QR model $Q_\tau(Y|\mathbf{x}) = \boldsymbol{\beta}^\top \mathbf{x}$, where $\boldsymbol{\beta}$ is a d-dimensional vector of unknown parameters, and used the representation

$$Q_\tau(Y|\mathbf{x}) = \arg\min_{\mathbf{b} \in \mathbb{R}^d} E\{\rho_\tau(Y - \mathbf{b}^\top \mathbf{X})\},$$

where, for $0 < \tau < 1$, the function $\rho_\tau(u) = \{\tau - I(u < 0)\}u$ is the check function, to define the estimator $\widehat{\boldsymbol{\beta}}$.

For time series of returns $\{r_t\}_{t=1}^n$, the QAR(p) model assumes that

$$Q_\tau(r_t|\mathcal{F}_{t-1}) = \beta_{0,\tau} + \sum_{i=1}^{p} \beta_{i,\tau} r_{t-i},$$

where $\beta_{i,\tau}$, $i = 0, 1, \ldots, p$, are unknown parameters. Here we use $p = 2$, but other choices for p can be considered.

3 Real-World Data

In this section, we apply the different methods to several real-world datasets and compare their performance for two different time periods. The data was downloaded from https://finance.yahoo.com and consists of the daily returns from January 2004 to December 2009 and from January 2011 to December 2016 of four major market indices: the S&P

500 Index, the Nikkei 225 Index, the Dow Jones Industrial Average, and the Russell 2000 Index. Since the data for the 2004-2009 period for Bitcoin Price Index is not available, an application is presented only for the 2011-2016 period. Figure 1 gives histograms for the daily returns for the two time periods and the different data sets, which confirm the skewed distribution and non-normality of financial data. Note that the histograms that include the financial crisis data have longer tails.

For our analysis, we used the first observations as historical data and the last 500 observations for one-step ahead VaR forecasts. For each of the time points, t, of the 500 observations not included in the historical data, we estimated the model based on the data up to time $t-1$ and forecasted the one-step-ahead VaR. We then checked if the return is less than $-\widehat{\mathrm{VaR}}_\tau(t)$, i.e. if $r_t < -\widehat{\mathrm{VaR}}_\tau(t)$. If it is, then we consider time t to be an exception, otherwise we do not consider it to be an exception. Let E be the number of exceptions and let $E/500$ be the proportion of exceptions. If the forecast procedure works well, then we should have $E/500 \approx \tau$.

To check the performance of the various methods, we also performed a formal backtesting procedure. Specifically, we considered three likelihood ratio tests. The first is the classical test for unconditional coverage (uc) introduced in Kupiec (1995), the second is the test for independence (ind) introduced in Christoffersen (1998), and the third is the test for conditional coverage (cc) also introduced in Christoffersen (1998). The test for unconditional coverage tests if $E/500$ is close enough to τ, the test for independence tests if there is clustering of exceptions, and the test of conditional coverage tests both of these simultaneously.

We begin by considering the performance of the different methods for the 2004-2009 time period. Table 1 reports the proportions of exceptions and Table 2 reports the p-values resulting from the three backtesting procedures. We see that DFGARCH(1,1) has the best performance in terms of the proportion of exceptions for all values of τ for the S&P 500 and Nikkei 225 data sets, as well as, for the 1% and 2.5% VaR for Dow and Jones and 1% and 5% VaR for Russell 2000. Moreover, GARCH(1,1) with skewed normal innovations has a comparable performance for 1% VaR for Nikkei 225 and for all values of τ for Russell 2000. GARCH(1,1) with skewed normal innovations also has comparable performance. Note that all methods have a proportion of exceptions considerably higher than the true value of τ, since they are conservative for volatile periods.

Although there are many problems with predicting the correct amount of exceptions, where $\widehat{\tau}$ differs significantly from τ, there are few instances where clustering is a problem. This indicates that the models are operating accordingly under the assumption of independent returns. We note that the historical method is the only method that does not have any issues with independence. For all of the other methods, we reject the assumption of independence for at least one value of τ.

Table 1: Percentage of exceptions when forecasting VaR_τ by various methods for the four market indices and for the 2004-2009 period. For each τ, the value that is closest to τ is bolded. In the case of ties, all relevant values are bolded.

	GARCH(1,1)-$\mathcal{N}$	**GARCH(1,1)-$\mathcal{SN}$**	**Historical**	**DFGARCH(1,1)**	**QAR(2)**
τ =.01					
SP500	0.036	0.024	0.056	**0.022**	0.056
N225	0.020	**0.014**	0.046	**0.014**	0.048
DJI	0.026	0.020	0.062	**0.010**	0.060
RUT	0.022	**0.018**	0.056	**0.018**	0.056
τ =.025					
SP500	0.058	0.048	0.108	**0.040**	0.112
N225	0.052	0.040	0.076	**0.034**	0.084
DJI	0.050	0.042	0.108	**0.040**	0.114
RUT	0.044	**0.036**	0.106	0.038	0.104
τ =.05					
SP500	0.076	**0.070**	0.164	**0.070**	0.160
N225	0.086	0.074	0.140	**0.072**	0.140
DJI	0.086	**0.078**	0.158	0.080	0.158
RUT	0.080	**0.066**	0.158	**0.066**	0.156

Table 2: p-values for the backtesting methods when forecasting VaR_τ by various methods for the four market indices and for the 2004-2009 period. (*) denotes significant p-values according to the 0.05 significance level.

UC	/ $\tau = 0.01$				
	GARCH(1,1)-$\mathcal{N}$	**GARCH(1,1)-$\mathcal{SN}$**	**Historical**	**DFGARCH(1,1)**	**QAR(2)**
SP500	0.000(*)	0.008(*)	0.000(*)	0.020(*)	0.000(*)
N225	0.048(*)	0.397	0.000(*)	0.397	0.000(*)
DJI	0.003(*)	0.048(*)	0.000(*)	1.000	0.000(*)
RUT	0.020(*)	0.106	0.000(*)	0.106	0.000(*)
IND					
SP500	0.246	0.442	0.276	0.481	0.079
N225	0.186	0.655	0.389	0.655	0.449
DJI	0.404	0.522	0.441	0.750	0.000(*)
RUT	0.481	0.565	0.276	0.565	0.079
CC					
SP500	0.000(*)	0.021(*)	0.000(*)	0.052	0.000(*)
N225	0.059	0.632	0.000(*)	0.632	0.000(*)
DJI	0.008(*)	0.115	0.000(*)	0.951	0.000(*)
RUT	0.052	0.230	0.000(*)	0.230	0.000(*)
UC	/ $\tau = 0.025$				
	GARCH(1,1)-$\mathcal{N}$	**GARCH(1,1)-$\mathcal{SN}$**	**Historical**	**DFGARCH(1,1)**	**QAR(2)**
SP500	0.000(*)	0.003(*)	0.000(*)	0.048(*)	0.000(*)
N225	0.001(*)	0.048(*)	0.000(*)	0.221	0.000(*)
DJI	0.002(*)	0.026(*)	0.000(*)	0.048(*)	0.000(*)
RUT	0.014(*)	0.139	0.000(*)	0.083	0.000(*)
IND					
SP500	0.942	0.119	0.942	0.196	0.079
N225	0.737	0.824	0.217	0.601	0.751
DJI	0.104	0.174	0.337	0.196	0.025(*)
RUT	0.154	0.246	0.286	0.220	0.109
CC					
SP500	0.000(*)	0.052	0.000(*)	0.061	0.000(*)
N225	0.003(*)	0.138	0.000(*)	0.413	0.000(*)
DJI	0.002(*)	0.034(*)	0.000(*)	0.061	0.000(*)
RUT	0.018(*)	0.171	0.000(*)	0.105	0.000(*)
UC	/ $\tau = 0.05$				
	GARCH(1,1)-$\mathcal{N}$	**GARCH(1,1)-$\mathcal{SN}$**	**Historical**	**DFGARCH(1,1)**	**QAR(2)**
SP500	0.013(*)	0.052	0.000(*)	0.052	0.000(*)
N225	0.001(*)	0.021(*)	0.000(*)	0.034(*)	0.000(*)
DJI	0.001(*)	0.008(*)	0.000(*)	0.004(*)	0.000(*)
RUT	0.004(*)	0.117	0.000(*)	0.117	0.000(*)
IND					
SP500	0.012(*)	0.022(*)	0.626	0.022(*)	0.303
N225	0.075	0.199	0.759	0.230	0.134
DJI	0.082	0.159	0.608	0.136	0.412
RUT	0.008(*)	0.031(*)	0.620	0.031(*)	0.210
CC					
SP500	0.002(*)	0.011(*)	0.000(*)	0.011(*)	0.000(*)
N225	0.001(*)	0.031(*)	0.000(*)	0.051	0.000(*)
DJI	0.001(*)	0.011(*)	0.000(*)	0.006(*)	0.000(*)
RUT	0.001(*)	0.028(*)	0.000(*)	0.028(*)	0.000(*)

We now compare the performance of the different methods for the 2011-2016 time period. Table 3 reports the proportions of exceptions and Table 4 reports the p-values resulting from the three backtesting procedures. For the first four market indices, we see that Historical, DFGARCH(1,1), and QAR(2) are the methods with the best performance in terms of exceptions. Note that, the proportion of exceptions are now closer to the true values of τ, as the time period includes less volatile daily returns. DFGARCH(1,1) appears to be one of the leading methods and has issues with independence for only one case, i.e., for the 2.5% VaR for the S&P 500 index. However, the only method that has no issues with independence is the GARCH(1,1) with skewed normal innovations.

Tables 3 and 4 also include the results for the Bitcoin Price Index. Bitcoin is an interesting index since it is relatively new and also considers highly volatile by financial institutions. We therefore thought it would be interesting to apply commonly used methods to this index. However, data is not available for the 2004-2009 period, and therefore, we only consider the 2011-2016 time period. In contrast with the other indices, GARCH(1,1) with normal and skewed normal innovations perform the best in terms of proportions of exceptions and have no issues with independence.

Table 3: Percentage of exceptions when forecasting VaR_τ by various methods for the five market indices for the 2011-2016 period. For each τ, the value that is closest to τ is bolded. In the case of ties, all relevant values are bolded.

	GARCH(1,1)-$\mathcal{N}$	**GARCH(1,1)-$\mathcal{SN}$**	**Historical**	**DFGARCH(1,1)**	**QAR(2)**
τ =.01					
SP500	0.018	0.024	**0.008**	**0.012**	**0.012**
N225	0.030	0.026	**0.016**	0.022	0.020
DJI	0.016	0.014	**0.010**	0.012	**0.010**
RUT	0.018	0.014	**0.008**	0.014	0.006
BTC	**0.012**	**0.012**	0.000	0.004	0.000
τ =.025					
SP500	0.032	0.048	**0.022**	0.016	0.020
N225	0.044	0.036	0.046	**0.034**	0.046
DJI	0.038	0.036	0.022	0.016	**0.026**
RUT	0.034	0.030	0.018	**0.026**	0.018
BTC	**0.018**	**0.018**	0.006	0.012	0.004
τ =.05					
SP500	0.056	0.070	0.038	0.036	**0.046**
N225	0.068	**0.064**	0.072	0.066	0.068
DJI	0.058	0.064	**0.052**	0.046	**0.052**
RUT	0.058	0.054	0.038	**0.048**	0.036
BTC	**0.020**	**0.020**	0.014	**0.020**	0.014

To get a better understanding of how each method works, we plotted the time series of returns for each market index with the corresponding one-step ahead 1%, 2.5%, and 5% VaR forecasts overlaid for the different methods and the different time periods. However, since these plots convey similar results, and to save some space, we only present the plots for the S&P 500 Index in Figure 2. Note that the Historical method does not adapt easily to sudden changes on the returns. Moreover note that, for the 2004-2009 time period, all methods are very conservative during the crisis. This is the reason we ended up with estimated proportions of exceptions much higher than the true vales of τ.

4 Discussion

In this work we demonstrated the performance of commonly used methods for VaR estimation during two time periods: during the financial crisis (2004-2009), and the period followed the financial crisis (2011-2016). The results show that DFGARCH(1,1) is a competitive method for both time periods, producing proportions of exceptions closer to the true values of τ, and not having issues with independence. However, during the 2011-2016 period, QAR(2) and Historical methods also have comparable performance, while for the Bit-

Table 4: p-values for the backtesting methods when forecasting VaR_τ by various methods for the five market indices for the 2011-2016 period. (*) denotes significant p-values according to the 0.05 significance level.

UC	/ $\tau = 0.01$				
	GARCH(1,1)-$\mathcal{N}$	**GARCH(1,1)-$\mathcal{SN}$**	**Historical**	**DFGARCH(1,1)**	**QAR(2)**
SP500	0.106	0.397	0.641	0.663	0.663
N225	0.000(*)	0.003(*)	0.215	0.020(*)	0.048(*)
DJI	0.215	0.397	1.000	0.663	1.000
RUT	0.106	0.397	0.641	0.397	0.331
BTC	0.663	0.663	0.002(*)	0.125	0.002(*)
IND					
SP500	0.007(*)	0.079	0.019(*)	0.054	0.054
N225	0.463	0.339	0.109	0.232	0.522
DJI	0.109	0.079	0.034(*)	0.054	0.034(*)
RUT	0.565	0.655	0.019(*)	0.655	0.009(*)
BTC	0.702	0.702	1.000	0.899	1.000
CC					
SP500	0.007(*)	0.149	0.059	0.142	0.142
N225	0.001(*)	0.007(*)	0.128	0.033(*)	0.115
DJI	0.128	0.149	0.106	0.142	0.106
RUT	0.230	0.632	0.058	0.632	0.021(*)
BTC	0.845	0.845	0.007(*)	0.306	0.007(*)
UC	/ $\tau = 0.025$				
	GARCH(1,1)-$\mathcal{N}$	**GARCH(1,1)-$\mathcal{SN}$**	**Historical**	**DFGARCH(1,1)**	**QAR(2)**
SP500	0.336	0.673	0.661	0.168	0.458
N225	0.014(*)	0.139	0.007(*)	0.221	0.007(*)
DJI	0.083	0.139	0.661	0.168	0.887
RUT	0.221	0.487	0.292	0.887	0.292
BTC	0.292	0.292	0.001(*)	0.039(*)	0.000(*)
IND					
SP500	0.096	0.055	0.019(*)	0.004(*)	0.186
N225	0.975	0.674	0.099	0.601	0.389
DJI	0.192	0.155	0.019(*)	0.109	0.039
RUT	0.601	0.463	0.145	0.404	0.145
BTC	0.145	0.145	0.849	0.702	0.899
CC					
SP500	0.158	0.144	0.057	0.006(*)	0.316
N225	0.048(*)	0.306	0.007(*)	0.413	0.018(*)
DJI	0.095	0.122	0.057	0.107	0.119
RUT	0.413	0.601	0.198	0.699	0.198
BTC	0.198	0.198	0.005(*)	0.109	0.001(*)
UC	/ $\tau = 0.05$				
	GARCH(1,1)-$\mathcal{N}$	**GARCH(1,1)-$\mathcal{SN}$**	**Historical**	**DFGARCH(1,1)**	**QAR(2)**
SP500	0.158	0.144	0.057	0.006(*)	0.316
N225	0.048(*)	0.306	0.007(*)	0.413	0.018
DJI	0.095	0.122	0.057	0.107	0.119
RUT	0.413	0.601	0.198	0.699	0.198
BTC	0.000(*)	0.000(*)	0.000(*)	0.000(*)	0.000(*)
IND					
SP500	0.727	0.389	0.032(*)	0.155	0.018(*)
N225	0.820	0.969	0.046(*)	0.894	0.275
DJI	0.802	0.969	0.582	0.389	0.582
RUT	0.547	0.671	0.748	0.878	0.674
BTC	0.186	0.186	0.002(*)	0.186	0.002(*)
CC					
SP500	0.784	0.633	0.044(*)	0.117	0.056
N225	0.209	0.386	0.014(*)	0.117	0.118
DJI	0.703	0.386	0.842	0.633	0.842
RUT	0.605	0.842	0.417	0.967	0.293
BTC	0.001(*)	0.001(*)	0.000(*)	0.001(*)	0.000(*)

coin Price Index, GARCH(1,1) with normal and skewed normal innovations have a better performance.

All of the methods considered in this paper rely heavily on past information to forecast the next step. Thus, when markets are in a period of stability, these methods work well. However, when an unforeseen event happens, such as the 2007-2008 financial crisis, these methods fail not only to predict such an extreme event, but also to adapt their VaR forecasts. The proportions of exceptions for these methods during the 2004-2009 data become increasingly conservative. Specifically, the proportions tend to be bigger than the specified level of VaR, meaning that VaR is over-projected, which could have negative impacts for decision making for financial institutions and could discourage healthy investing.

It is interesting to see that DFGARCH(1,1) consistently has a good performance for both time periods, meaning that its good performance is maintained even in a highly volatile period. However, it is also advantageous to use the Historical and QAR(2) methods during relative financial stability. There is no answer for a best method under any circumstances. One needs to use judgment to choose a method.

For further research, a hybrid method could be developed that combines such methods. When a drastic event happens, DFGARCH(1,1) can be used to forecast VaR during the volatile times, and once relative stabilization has happened, either QAR or the Historical method can be used until another extreme event happens. Of course how one characterizes an extreme event, or re-stabilization is subjective and would need to be further defined.

Another route of potential interest would be to take the methods explored in this project and combine them with the estimation of another important quantity, the Expected Shortfall (ES). ES is, perhaps, the most prominent alternative to VaR and it defines the conditional expectation of the loss given that this loss exceeds the VaR.

References

[1] Bollerslev, T. (1986) Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics* 31, 307–327.

[2] Christoffersen, P. (1998) Evaluating Interval Forecasts. *International Economic Review* 39(4), 841–862.

[3] Duffie, D. and Pan, J. (1997) An overview of value at risk. *Journal of Derivatives* 4, 7–49.

[4] Engle, R.F. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* 50, 987–1008.

[5] Engle, R.F. and Manganelli, S. (2004) CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics* 22, 367–381.

[6] Koenker, R. and Bassett, G. (1978) Regression quantiles. *Econometrica* 46, 33–50.

[7] Koenker, R. and Xiao, Z. (2006) Quantile autoregression. *Journal of the American Statistical Association* 101(475), 980–1006.

[8] Kupiec, P. (1995) Techniques for Verifying the Accuracy of Risk Management Models. *Journal of Derivatives* 3(2), 73–84.

[9] RiskMetrics Group (1996) *RiskMetrics -Technical Document*, New York: J. P. Morgan/Reuters.

[10] Taylor, S. (1986) *Modelling Financial Time Series*, New York, Wiley.

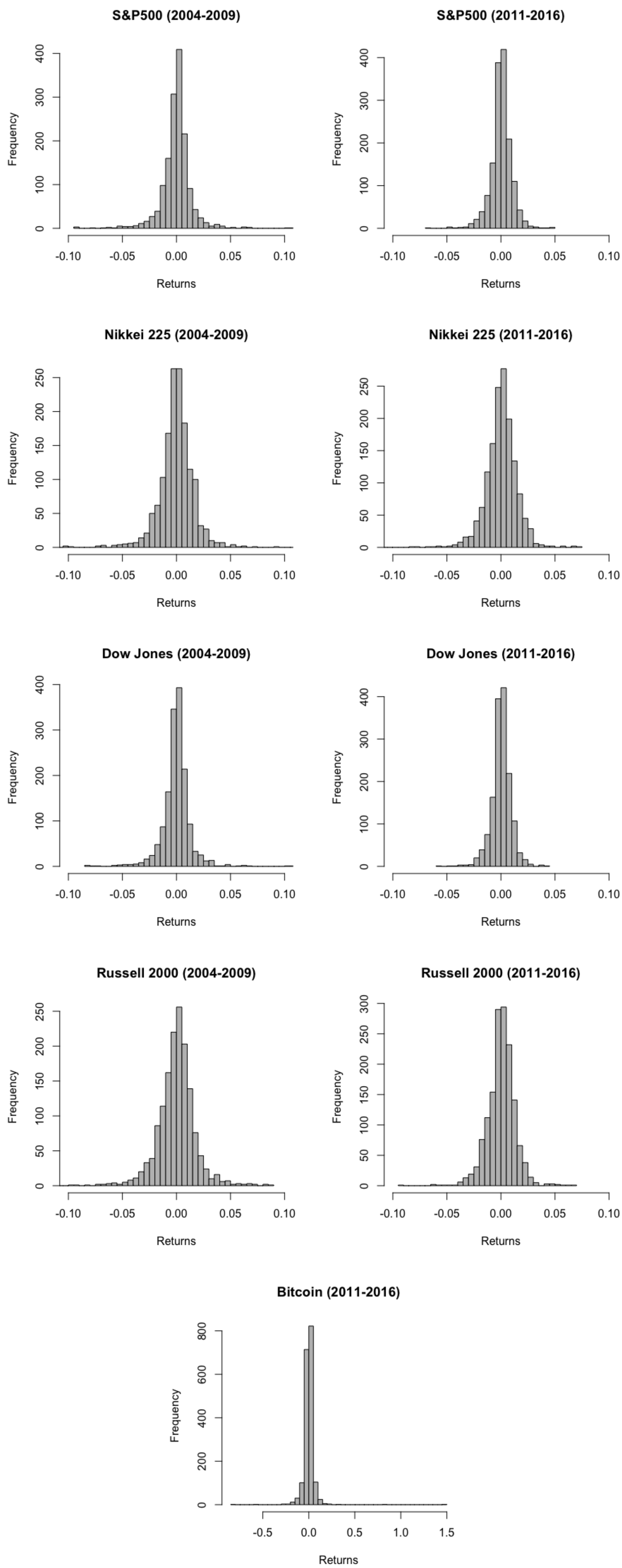

Figure 1: *Histograms for daily returns for the five data sets for the periods 2004-2009 and 2011-2016.*

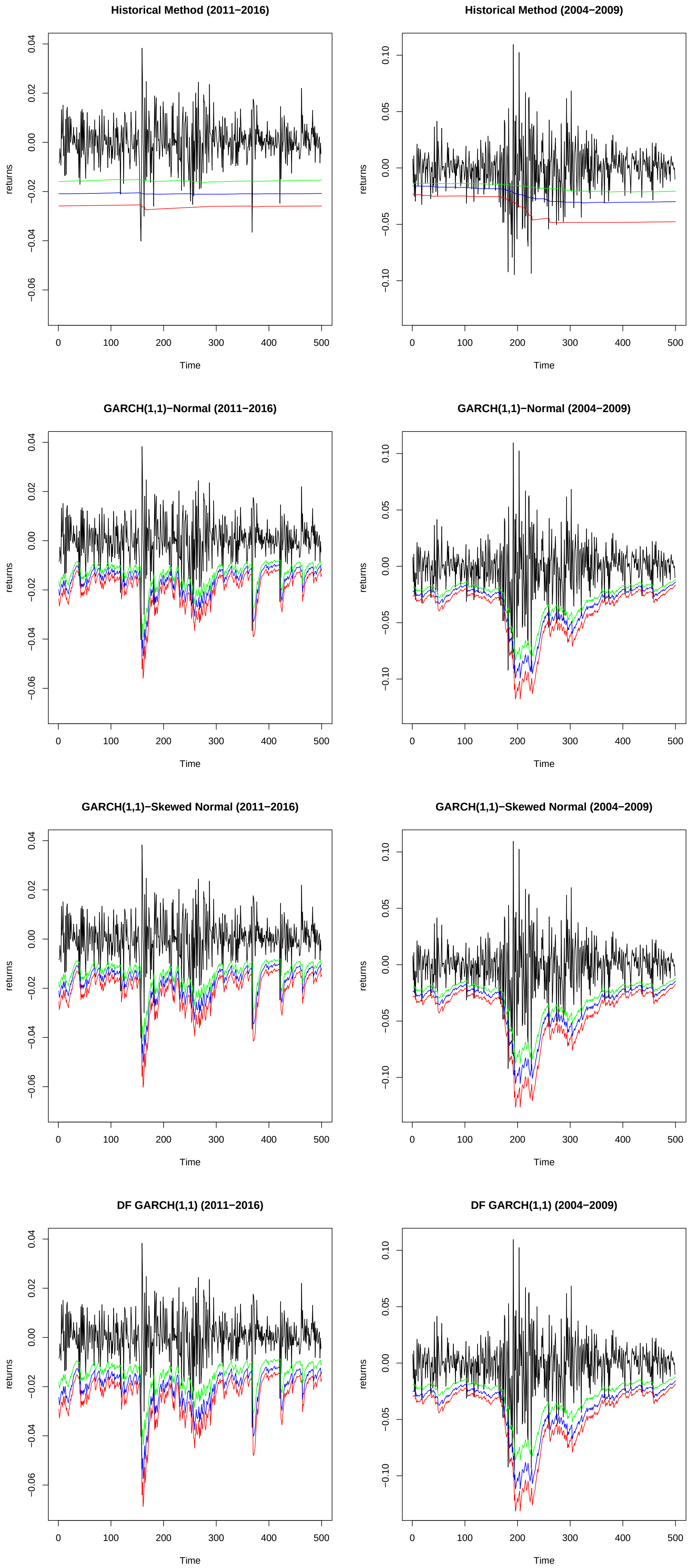

Figure 2: Time series of returns and one-step ahead VaR forecasts for the S&P500 Index using the different methods. Green, blue, and red represent the 1%, 2.5%, and 5% VaR forecasts, respectively.

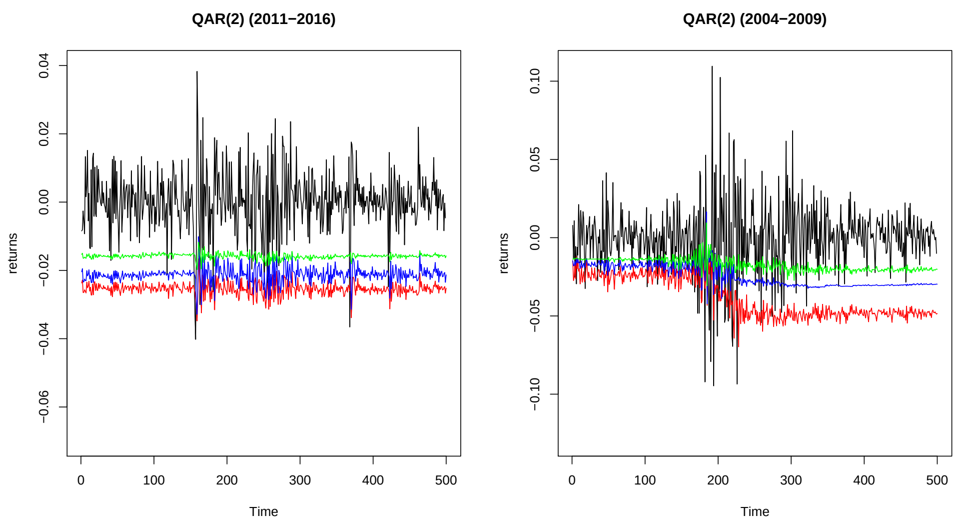

Figure 2: continued.