ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

A thesis presented for the bachelor degree of
Econometrics and Operations Research with a major in Quantitative Finance

Extended Geometric-VaR Backtesting

Gert van Dasler

433958

Prof. dr. D.J.C. van Dijk

S.H.L.G.C. Vermeulen

July 7, 2018

Abstract This paper extends on the methods provided by Pelletier & Wei (2015). In the first part of the paper, we replicate the simulation-driven results where we extend with empirical research in the second part. Opposed to the Historical Simulation method used in Pelletier & Wei (2015), we create a set of risk specifications derived from Wong et al. (2016). These specifications are exploited using two distinct methods of innovations inspired by the findings of Bao et al. (2007). We empirically show that the Geometric-VaR test possesses high power against alternatives within the framework for our empirical specifications. Additionally, we show that parametrized GARCH specifications lead to better specified Value-at-Risk estimations than the Historical Simulation approach does. Lastly, we show that GARCH specifications with skewed innovations generally lead to better-specified Value-at-Risk estimations than specifications that use a Filtered Historical Simulation instead.

---

1 Introduction

In managing risk on daily returns it is useful to introduce certain terminology as a widely accepted measurement of risk. In light of the 1996 Risk Amendment in Basel, VaR (Value-at-Risk) has become the general used toolkit to cope with measuring risk. It is defined as the maximum loss that will not be exceeded over a period of time with a certain probability. Mathematically it can thus be treated as a quantile of the conditional daily returns. For example, given a VaR of 5% on returns for the next day, the probability that the realized daily return of the next day will fall below this VaR value is equal to 5%.

These Value-at-Risk forecasts can be computed using a various set of estimation methods. In back-testing these specifications, the ex-ante Value-at-Risk are compared with the ex-post daily returns for each observation. If an actual realized daily return falls below the corresponding Value-at-Risk forecast, it is denoted as a violation. The Value-at-Risk forecasts with coverage rate p are said to be correctly specified if and only if the sequence of violations consist of independent and identically distributed Bernoulli random variables that share the same coverage rate.

In formally testing whether these assumptions are harmed by the computed Value-at-Risk estimates, Pelletier & Wei (2015) provides us with a framework of tests on these properties and assumptions. Within the framework, these assumptions can be tested both individually and jointly. This testing framework is inspired by econometric literature from Christoffersen (1998), Haas (2005), Candelon et al. (2010) and Berkowitz et al. (2011). Pelletier & Wei (2015) extends to existing literature with the Geometric-VaR test, that tests both the Geometric- and the VaR test jointly.

To calculate Value-at-Risk forecasts over a certain period, various methods have been provided by common literature. The method that is most often used is Historical Simulation, as has been shown in Pérignon & Smith (2010) where 73% of European Banks utilize Historical Simulation as their method to calculate Value-at-Risk estimates. Besides the wide applicability, the method of Historical Simulation is non-parametric and thus easy to implement. For these reasons, this method is utilized by Pelletier & Wei (2015) in conducting their results. However, the method of Historical Simulation implicitly assumes independence and identical distribution among returns. These are assumptions that often do not hold.

We extend to other methods apart from the Historical Simulation that require parameter estimation. Wong et al. (2016) provides a selection of risk specifications in order to obtain better-specified Value-at-Risk estimates. We provide a selection of models that are inspired by these risk specifications. The innovations of the parametrized risk models follow the skewed t-distribution introduced by Hansen (1994).

In addition to the specifications that possess skewed innovations, we utilize the same specifications with another method that utilizes the Filtered Historical Simulation approach. This method is inspired by Bao et al. (2007), that states that models with skewed innovations generally perform better than models with non-skewed innovations. This finding is exploited in creating two distinct methods. In addition to the method that utilizes the skewed t-distribution, we also propose a method that utilizes Filtered Historical Simulation.

We obtain data from the Realized Library of the Oxford Man Institute of Quantitative Finance. For the Swiss SMI, the Japanese NIKKEI and the American S&P500 we obtain daily returns dating from the beginning of 2005 until the last daily observation of 2017.

With these data, we show that the Geometric-VaR test possesses higher power against a set of various alternatives for our parametrized specifications. The parametrized risk specifications lead to much better-specified Value-at-Risk estimation than the Historical Simulation method, which is shown both graphically and using the methods provided by Dufour (2006). In addition to these findings, we show that the Value-at-Risk estimation with skewed innovations lead to better-specified Value-at-Risk estimation than the models with non-skewed innovations as in the Filtered Historical Simulation approach.

The paper is built as follows. Section two walks through the methodology that is used in Pelletier & Wei (2015) and can be read as a general walk-through. The third section further specifies the testing framework and how the log likelihood is constructed for each test. In the fourth chapter we provide information on the Simulation Studies. It provides an overview on how the data is generated and how the Value-at-Risk forecasts are estimated in the replication of Pelletier & Wei (2015). The fifth chapter gives a brief overview on the most important simulation-based replications from the paper. The sixth section provides our empirical extension to the paper, with corresponding graphs and tables in the appendix. The last chapter draws a conclusion out of both the replication and our empirical addition to Pelletier & Wei (2015). It also provides some recommendations for further research.

2 Methodology

2.1 Testing Framework

In this research we solely focus on one-step-ahead forecasts in estimating the Value-at-Risk. To be more thorough an one-day-ahead Value-At-Risk with coverage rate p denotes a quantile for which the actual realized daily return of the respective day is equal or less than this threshold value with probability p. One-day-ahead Value-at-Risk forecasts are then denoted as follows.

$$\mathbb{P}[-\infty \leq r_t \leq \text{VaR}_t(p)|\mathcal{I}_{t-1}] = p \tag{1}$$

Here $\mathcal{I}_{t-1}$ denotes the information set that captures all available information up until time $t-1$. If equation (1) holds, the VaR forecast $\text{VaR}_t(p)$ is said to be efficient. After calculating the Value-at-Risk estimates, we want to check for so-called violations. Violations occur when the actual daily return is lower than the corresponding Value-at-Risk forecast for that respective observation. Stated differently, we denote the violations such that:

$$I_t = \begin{cases} 1, & \text{if } r_t \in (-\infty, \text{VaR}_t(p)) \\ 0, & \text{if } r_t \notin (-\infty, \text{VaR}_t(p)) \end{cases} \tag{2}$$

To extend the idea of efficiency, a Value-at-Risk forecast is efficient if and only if the violations are independent and identically distributed Bernoulli random variables with coverage rate p. Pelletier & Wei (2015) suggests to obtain durations such that the i-th duration is given as $D_i = t_i - t_{i-1}$, where t_i denotes the time on which the i-th violation occurs. Under the null hypothesis that the Value-at-Risk forecast is correctly specified, it follows that a sum of Bernoulli distributed variables follows a geometric distribution.

$$\mathbb{P}[D_i = d] = p(1-p)^{d-1} \tag{3}$$

Furthermore, Pelletier & Wei (2015) utilizes hazard functions for their testing framework. An hazard function is denoted as:

$$\lambda_d^i = \frac{f^i(d)}{S^i(d)} \tag{4}$$

Hence λ_d^i denotes the hazard function, $f^i(d)$ denotes the probability density function and $S^i(d)$ denotes the survival function. For our binary results of violations we can write that:

$$f^i(d) = \mathbb{P}[D_i = d] = \lambda_d^i(1-\lambda_{d-1}^i)...(1-\lambda_1^i) \tag{5}$$

$$S^i(d) = \mathbb{P}[D_i \geq d] = (1-\lambda_{d-1}^i)...(1-\lambda_1^i) \tag{6}$$

Combining both expressions for the probability density- and the survival function we obtain the following expression for the hazard function.

$$\lambda_d^i = \mathbb{P}(I_{t_i+d} = 1|I_{t_i+d-1} = 0, ..., I_{t_i+1} = 0, \mathcal{I}_{t_i+d-1}) \tag{7}$$

The hazard function is constant and equal to coverage rate p, if and only if both the Value-at-Risk forecast is efficient and it has correct unconditional coverage.

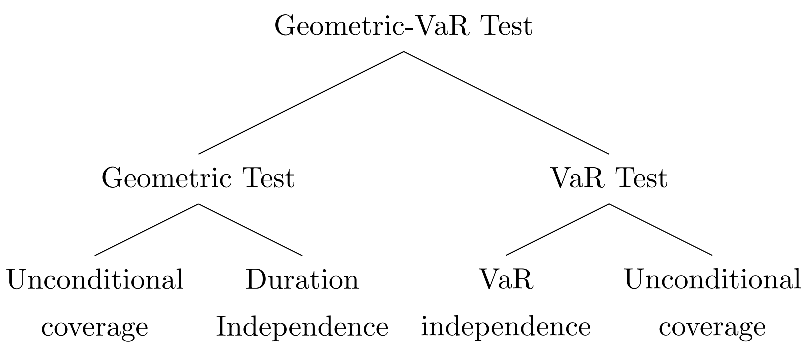
The framework in Pelletier & Wei (2015) consists of the Correct Unconditional Coverage-, Duration Independence-, VaR Independence-, Geometric-, VaR-, and the Geometric-VaR test. The Geometric-VaR test is the extension of the paper to common literature. It utilizes the following hazard function.

$$\lambda_d^i = ad^{b-1}e^{-c\text{VaR}_{t_i+d}} \tag{8}$$

Note that the Value-at-Risk estimates in the hazard function are positive. This hazard function is sequentially implemented in the methodology after which we test on Correct Unconditional Coverage, Duration Independence and VaR independence jointly. These tests can be evaluated solely as well. The bounds on the parameters are given by $0 \leq a < 1$, $0 \leq b \leq 1$ and $c \geq 0$. For some tests, we restrict certain parameters such that the hazard function will change. These derivations are given later on in Section 3.

As the name already suggests, The Geometric-VaR is built on the Geometric- and the VaR test. If the Geometric-VaR test is rejected, we can sequentially investigate whether its components hold or not. If one of these two are rejected, we take a step further in the tree given in Figure 1 and see which of the properties do not hold. This results in an iterative framework.

Figure 1: Graphical insight in iterative testing procedure



Note: This tree can be interpreted as an iterative testing procedure. The Geometric-VaR tests is built out of components such as the Geometric Test and the VaR test. Sequentially, if the Geometric-VaR test is rejected we can iteratively check which of its components are rejected. Each component consists of other components in a next, deeper layer until the third one.

3 Test evaluation

It seems likely that the hit sequence does not start with a violation. Ignoring the first duration is not the optimal choice, since this duration is informative in terms of the survival function and the sample size is relatively small. If the sequence does not start with a violation, we know that the first duration consists of at least the amount of observations until the first hit. Similarly, if the the hit sequence does not end with a violation the last duration consists of at least the amount of observations between the last violation and the end of the hit sequence. These durations are said to be left- and right censored respectively.

Assume that we have N durations. The sequence of durations is denoted by $\{D_i\}_{i=1}^N$. Apart from the durations we conduct a sequence of censored durations $\{C_i\}_{i=1}^N$. These censored durations are binary and equal to one if a duration is indeed censored. Only the first and the last observation can be censored. We ultimately get the following likelihood function:

$$\begin{aligned}\log L(D|\Theta) &= C_1 \log S^1(D_1) + (1 - C_1) \log f^1(D_1) \\ &+ \sum_{i=2}^{N-1} \log f^i(D_i) + C_N \log S^N(D_N) + (1 - C_N) \log f^N(D_N)\end{aligned} \tag{9}$$

Here, both the probability density function and the survival function are filled in using equations (5) and (6) respectively. Θ contains all parameters $[a, b, c, d]$ out of equation (8) or a subset of those.

Given the null hypothesis of each test, we calculate the likelihood both under the null- and the alternative hypothesis. A general form of the likelihood is then given by $\mathrm{LR} = -2(\log L(D|\hat{\Theta}_R) - \log L(D|\hat{\Theta}_{UR}))$. Θ_R denotes set of constrained estimated parameters and Θ_{UR} denotes the set of unconstrained estimated parameters.

These LR tests are assumed to asymptotically follow chi-squared distributions. Pelletier & Wei (2015) however, only uses relatively small samples. For a sample consisting of 500 observations with Value-at-Risk coverage rate equal to one percent, we expect to only obtain five violations under the null hypothesis of correct specification. Thus we expect to get such a small amount of durations, that asymptotic properties do not seem to hold.

To cope with this difficulty, Pelletier & Wei (2015) opts to use the Monte Carlo simulation techniques as introduced in Dufour (2006). We generate a large amount of data replications under the null hypothesis, to create simulated distributions for each test. Critical values are sequentially deducted from these simulated distributions. Thus we calculate p-values according to Dufour (2006), given by:

$$\hat{p}_N(\mathrm{LR}_0) = \frac{N\hat{G}(\mathrm{LR}_0) + 1}{N + 1} \tag{10}$$

$\hat{G}_N(\mathrm{LR}_0)$ is given by the following formula:

$$\hat{G}_N(\mathrm{LR}_0) = 1 - \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(\mathrm{LR}_i \leq \mathrm{LR}_0) + \sum_{i=1}^{N} \mathbf{1}(\mathrm{LR}_i = \mathrm{LR}_0)\mathbf{1}(U_i \geq U_0) \tag{11}$$

LR_i denotes the test statistics of the i-th replication under the null hypothesis, LR_0 denotes the test statistics from the sample and $\mathbf{1}(*)$ denotes an indicator function. If the test statistic of the i-th replication is equal to the sample test statistic, we draw an uniform number U_i and compare it to the uniform number drawn at the sample test statistic U_0. This contribution is denoted in the last part of the formula in equation (11).

3.1 Brief Overview

In the following section we provide a summary on how the log-likelihood is evaluated in terms of the hazard function for each test. These hazard functions are derived from equation (4).

3.1.1 Correct Unconditional Coverage

With the Correct Unconditional Coverage test, we test $H_0 : a = p$ vs. $H_a : a \neq p$. We assume that $b = 1$ and $c = 0$, such that the hazard function simplifies to $\lambda = a$. We drop the subscripts i and d here. The log-likelihood is given by $\mathrm{LR_{UC}} = -2(\log L(D|a = p,\ b = 1,\ c = 0) - \log L(D|\hat{a},\ b = 1,\ c = 0))$.

3.1.2 Duration Independence

To test for Duration Independence, we set $H_0 : b = 1$ vs. $H_a : b < 1$. We assume that $c = 0$. The hazard function simplifies to $\lambda_d = ad^{b-1}$ such that log-likelihood can be computed with $\mathrm{LR_{Dind}} = -2(\log L(D|\hat{a},\ b = 1,\ c = 0) - \log L(D|\hat{a},\ \hat{b},\ c = 0))$.

3.1.3 VaR Independence

For VaR-Independence we test $H_0 : c = 0$ vs. $H_a : c > 0$. We make no further assumptions about the parameters, such that the hazard function is evaluated such as in equation (8). The test statistic is sequentially evaluated as $\mathrm{LR_{Vind}} = -2(\log L(D|\hat{a},\ \hat{b},\ c = 0) - \log L(D|\hat{a},\ \hat{b},\ \hat{c}))$.

3.1.4 Geometric Test

For the Geometric test, the null hypothesis $H_0 : a = p$ AND $b = 1$ is tested against the alternative hypothesis of $H_a : a \neq p$ OR $b < 1$. The corresponding test statistic is given by $\mathrm{LR_G} = -2(\log L(D|a = p,\ b = 1,\ c = 0) - \log L(D|\hat{a},\ \hat{b},\ c = 0))$. This corresponds to the sum of its third layer components, such that $\mathrm{LR_{GV}} = \mathrm{LR_{UC}} + \mathrm{LR_{Dind}} + \mathrm{LR_{Vind}}$.

3.1.5 VaR test

For the VaR test, we assume Duration Independence amongst the durations such that $b = 0$. Specifically, the hypothesis $H_0 : a = p$ AND $c = 0$ is tested against the alternative of $H_a : a \neq p$ OR $c > 0$. The corresponding test statistic is obtained by $\mathrm{LR_V} = -2(\log L(D|a = p,\ b = 1,\ c = 0) - \log L(D|\hat{a},\ b = 1,\ \hat{c}))$.

3.1.6 Geometric-VaR test

The final test within the framework of Pelletier & Wei (2015) is used to evaluate the Geometric- and VaR test jointly. We test the null hypothesis $H_0 : a = p$ AND $b = 1$ AND $c = 0$ against the alternative hypothesis such that $H_0 : a \neq p$ OR $b < 1$ OR $c > 0$. Thus the test statistic can be evaluated such that $\mathrm{LR_{GV}} = -2(\log L(D|a = p,\ b = 1,\ c = 0) - \log L(D|\hat{a},\ \hat{b},\ \hat{c}))$. This corresponds to the sum of its three components, namely $\mathrm{LR_{GV}} = \mathrm{LR_{UC}} + \mathrm{LR_{Dind}} + \mathrm{LR_{Vind}}$.

4 Simulation Studies

4.1 Data Generating Process

Pelletier & Wei (2015) utilizes NGARCH-t(d) specifications according to past literature, on which it executes its testing framework. The NGARCH-t(d) process allows for the leverage effect, volatility clustering and more extreme returns by using a t-distribution instead of a standard normal function. To be more precise, the data is generated using the following set of equations:

$$R_{t+1} = \sigma_{t+1}\Big(\frac{d-2}{d}\Big)^{1/2} z_{t+1} \tag{12}$$

$$\sigma_{t+1}^2 = \omega + \alpha\sigma_t^2\left(\Big(\frac{d-2}{d}\Big)^{1/2} z_t - \theta\right)^2 + \beta\sigma_t^2 \tag{13}$$

z_t follows the t(d)-distribution. The unconditional variance is given by $\omega(1-\alpha(1+\theta^2)-\beta)^{-1}$, while $\alpha(1+\theta)^2+\beta$ denotes the volatility persistence. In other words, the volatility persistence is interpreted as the memory in volatility. High persistence results in future volatility that strongly depends on previous values of volatility. Pelletier & Wei (2015) distinguishes four business lines with their corresponding parameters. They are given in the table below:

Table 1: Business Lines provided by Pelletier & Wei (2015)

	Business Line 1	Business Line 2	Business Line 3	Business Line 4
d	3.808	3.318	6.912	4.702
θ	-0.245	0.503	-0.962	0.093
β	0.749	0.928	0.873	0.915
α	0.155	0.052	0.026	0.072
ω	0.550	0.215	0.213	1.653

Note: In this table each column represents the parameters of each theoretical business line. d denotes the degrees of freedom. θ is used to implement the leverage effect. If θ is positive, future volatility is stronger caused by negative returns than it is by positive ones.

4.2 Value-at-Risk Estimation

Unless stated differently, the method of Value-at-Risk estimation on which the tests are backtested is the Historical Simulation approach. As mentioned before, Historical Simulation has the advantage of being non-parametric and thus easy to implement. In calculating one-step-ahead Value-at-Risk forecasts, we calculate the value of the p-th quantile over a rolling window. Thus, the estimate for the one-step-ahead forecast is given by the following formula. T_e denotes the length of the rolling window.

$$\mathrm{VaR}_{t+1}^{HS} = \mathrm{quantile}(\{R_i\}_{i=t-Te+1}^{t}, p) \tag{14}$$

5 Replication

5.1 Size of the Duration-based Tests

This section in Pelletier & Wei (2015) shows why the simulation method in Dufour (2006) is desired to control the size of the duration-based tests. We generate data following a NGARCH-t(d) process with parameters equal to $[d, \theta, \beta, \alpha, \omega] = [10, 0, 0.93, 0.05, 0.21]$. Using these estimated returns and variances, we estimate Value-at-Risk estimates using the conditional distribution of the returns. Thus, the Value-at-Risk estimates are obtained using the following equation.

$$\text{VaR}_t = \sigma_t \left(\sqrt{\frac{d-2}{d}} \right) z_t \tag{15}$$

σ_t denotes the conditional standard deviation of the NGARCH-t(d) process, $\sqrt{\frac{d-2}{d}}$ is used to scale the estimates to the corresponding degrees of freedom of the t(d)-distribution. After obtaining the Value-at-Risk estimates and the generated returns, we compute the corresponding violations using framework 2.

The results are given in Table 2, where we test on 10% Duration-based tests. The table provides two sets of critical values. The asymptotic chi-squared critical values are given by $\chi^2_{(1-\alpha),\nu}$. ν denotes the respective degrees of freedom per test and α is equal 0.10 to obtain the desired quantile. The first three tests asymptotically have one degree of freedom, the Geometric and the VaR test have two degrees of freedom and the Geometric-VaR test has three degrees of freedom.

The simulated critical values are obtained somewhat differently and sequentially show the need for Dufour (2006). For $T = 50,000$ we generate a sequence of independent and identically distributed Bernoulli random variables with parameter p being equal to the Value-at-Risk coverage rate. Additionally, we generate Value-at-Risk estimates out of an independent NGARCH-t(d) process with the parameters as described above. The asymptotic chi-squared critical value is calculated as the 90% percentile out of 10,000 replications.

Table 2: Size of 10% Duration-based tests applied to 5% VaR

Sample Size	UC	Duration ind.	VaR ind.	Geometric	VaR	GV
			Using chi-squared asymptotic critical values			
250	0.146	0.027	0.047	0.063	0.083	0.048
500	0.118	0.033	0.052	0.071	0.086	0.050
750	0.127	0.037	0.049	0.059	0.076	0.053
1000	0.101	0.035	0.053	0.062	0.074	0.042
1250	0.105	0.035	0.051	0.062	0.078	0.046
1500	0.102	0.039	0.052	0.059	0.069	0.051
			Using simulated test statistics $T = 50,000$			
250	0.146	0.066	0.217	0.105	0.157	0.144
500	0.118	0.072	0.195	0.109	0.152	0.130
750	0.100	0.075	0.181	0.095	0.137	0.136
1000	0.101	0.073	0.168	0.110	0.134	0.128
1250	0.105	0.790	0.164	0.104	0.129	0.122
1500	0.102	0.081	0.156	0.097	0.125	0.120
			Simulated critical value			
	2.716	1.601	1.662	3.763	3.820	4.774
			Asymptotic chi-squared critical value			
	2.706	2.706	2.706	4.605	4.605	6.251

Note : We generate Bernoulli i.i.d. and an independent NGARCH process to calculate Value-at-Risk estimates. The entries in the table correspond to the rejection frequency of each test out of 10,000 replications. The first part uses the asymptotic chi-squared value as critical value. The second part uses the simulated critical value as the critical value of the rejections.

As can be concluded from the results in Table 2, the test statistics are not chi-squared under the null. The respective critical values differ from the chi-squared distribution and this problem can be solved by using the Monte Carlo method of Dufour (2006).

5.2 Finite sample power of each test

In this respective section of Pelletier & Wei (2015), the paper mimics the finite sample power of each test when the Value-at-Risk is misspecified using Historical Simulation. This procedure is executed for each theoretical business line provided in Table 1 and for each sample size, where the sample size varies from 250 to 1500. In using Historical Simulation we utilize a rolling window consisting of $T_e = 250$ observations.

For each entry, we generate $T + T_e$ returns using the parameters of the corresponding business line. The Value-at-Risk estimates are obtained using Historical Simulation. We obtain test statistics LR_0 out of the sample and add the CaViaR test of Engle & Manganelli (2004) with a logistic distribution for the error terms. Its respective regressors are given by a constant, the lagged hit and the lagged Value-at-Risk estimate.

We generate independent and identically distributed Bernoulli random variables and Value-at-Risk estimates of an independent NGARCH-t(d) process with the same parameters such that we obtain the test statistics $\{\mathrm{LR}_i\}_{i=1}^{9999}$. These are to be held constant. Additionally, we calculate sample test-statistics LR_0 using Historical Simulation. We calculate p-values in line with equation 10, such that the entries in the table correspond to the rejection frequency of p-values that fall below 10% significance on 5000 replications.

The results are given in 6 and 7 respectively. On the 10% Duration Tests on 5% Value-at-Risk estimates we get fairly similar results and the same conclusions hold as in Pelletier & Wei (2015). For the 1% Value-at-Risk estimates on 10% Duration Tests we get more volatile results. For larger sample sizes, the obtained values are close. For smaller sample sizes, the values can differ. This could be due to the limited amount of simulations that we do. Whereas Pelletier & Wei (2015) does 5000×9999 simulations, we execute $5000 + 9999$ simulations.

6 Extension

Besides the theoretical findings of Pelletier & Wei (2015) we want to give a more tangible, empirical extension on real-life data. As Pérignon & Smith (2010) states that Historical Simulation only shows very little information about future volatility, we would like to investigate whether the Geometric-VaR test shows high power for other Value-At-Risk specifications. If so, can we sequentially choose better-specified Value-at-Risk estimation methods than the approach of Historical Simulation. In selecting the best specified model, we use the framework provided by Pelletier & Wei (2015).

6.1 Models

Besides the Historical Simulation approach, we provide a selection of models out of common econometric literature. These models are specified according to the following framework.

1. $r_t = \varepsilon_t$
2. $\sigma_t^2 = \sigma^2(\mathcal{V}|\mathcal{I}_{t-1})$
3. $\varepsilon_t = \sigma_t z_t, \quad z_t \sim \mathrm{skT}(0, 1; \xi, \nu)$

$\mathcal{V}$ represents the corresponding set of parameters conditional on the information set on time

$t-1$. The standardized innovations z_t follow the skewed t-distribution first proposed by Hansen (1994). The probability density function of the skewed t-distribution is as follows.

$$f(z_t;\xi,\nu) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi(\nu-2)}}\Big(\frac{2s}{\xi+\xi^{-1}}\Big)\Big(1+\frac{sz_t+m}{\nu-2}\xi^{-d_t}\Big)^{-(\nu+1)/2} \tag{16}$$

Here ξ denotes the asymmetry parameter, ν denotes the tail thickness and $\Gamma(*)$ represents the gamma function. Note that in equation (16) $d_t = 1$ if $z_t \geq -m/s$ and $d_t = -1$ otherwise. m and s respectively represent the mean and the variance of the distribution where the corresponding formulas are given below.

$$m = \frac{\Gamma((\nu+1)/2)\sqrt{\nu-2}}{\Gamma(\nu/2)\sqrt{\pi}}(\xi-\xi^{-1}), \quad s = \sqrt{\xi^2+\xi^{-2}-1-m^2} \tag{17}$$

We implement two methods on the the parametrized risk specifications. First we execute the following procedure. The first step is, given our rolling window $\{r_{t-i}\}_{i=t-Te+1}^{t}$, to approximate the parameters $(\mathcal{V}^{(t)},\xi^{(t)},\nu^{(t)})$. These are used to obtain the one-step-ahead forecast of the conditional variance $\hat{\sigma}^2_{t+1|t}$. We then compute the Value-at-Risk forecast using the following formula.

$$\text{VaR}_{t|t-1}^{(1-p)} = f_p(z_t;\xi^{(t)},\nu^{(t)})\hat{\sigma}_{t+1|t} \tag{18}$$

$\hat{\sigma}_{t+1|t}$ denotes the estimated conditional standard deviation and $f_p(z_t;\xi^{(t)},\nu^{(t)})$ corresponds to the p-th quantile of the skewed t-distribution with estimated parameters equal to $\nu^{(t)}$ and $\xi^{(t)}$.

Secondly we also introduce another method to calculate Value-at-Risk estimates. We utilize a Filtered Historical Simulation method on the parametrized GARCH specifications, which are to be described in Section 6.1.1. For each rolling window, we utilize the standardized residuals. We then compute the one-step-ahead forecast of the Value-at-Risk using the following equation.

$$\text{VaR}_{t|t-1}^{(1-p)} = \sigma_{t|t-1}\text{quantile}(\{\hat{z}_i\}_{i=t-Te+1}^{t},p), \quad \hat{z}_i = \frac{r_i}{\hat{\sigma}_i} \tag{19}$$

6.1.1 Specifications of Conditional Variance

The one-step-ahead forecast of the conditional variance in (18) is estimated using a diverse set of methods inspired by the models proposed in Wong et al. (2016). We denote the following specifications for the conditional variance in Table 3.

Table 3: GARCH Specifications of Conditional Variance

Model	Specification of σ_t^2
RiskMetrics	$\sigma_t^2 = 0.06\varepsilon_{t-1}^2 + 0.94\sigma_{t-1}^2$
IGARCH(1,1)	$\sigma_t^2 = \omega + \alpha\varepsilon_t^2 + (1-\alpha)\sigma_{t-1}^2$
GARCH(1,1)	$\sigma_t^2 = \omega + \alpha\varepsilon_t^2 + \beta\sigma_{t-1}^2$
EGARCH(1,0,1)	$\sigma_t^2 = \exp\{\omega + \alpha z_{t-1} + \gamma(\lvert z_t\rvert - E[z_t]) + \beta\ln(\sigma_{t-1}^2)\}$
TGARCH(1,1,1)	$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \gamma\mathbf{1}(\varepsilon_{t-1} < 0)\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$

Note : This table contains all specifications for the conditional variance used in the Extension. Further details on each model are provided in the text.

The RiskMetrics approach has been introduced by Morgan et al. (1996), with fixed short-memory parameters $[\alpha, \beta] = [0.06, 0.94]$. Its conditional variance can be obtained recursively by formula $\hat{\sigma}_t^2 = \lambda\hat{\sigma}_{t-1}^2 + (1-\lambda)(r_{t-1} - \bar{r}_t)^2$. The first part of the formula denotes the persistence in volatility and the second part denotes the update in the first factor and the reaction in the second factor. $\bar{r}_t = \frac{1}{T_e}\sum_{i=1}^{T_e} r_{t-i}$ denotes the mean return over the rolling window.

The other models are all derivations of the GARCH model as proposed by Bollerslev (1986). The sum of the short-memory parameters of the GARCH model is less than 1, such that the process is covariance stationary. The IGARCH model is defined such that the sum of the short-memory parameters is equal to 1, but they are able to move freely within this constraint. The RiskMetrics approach falls within the scope of the IGARCH models. The EGARCH(1,0,1) proposed by Nelson (1991) automatically takes the leverage effect into account. The model allows for asymmetric effects for both negative and positive shocks due to the logarithmic transformation. Besides, the EGARCH model has the advantage of not needing further restrictions in obtaining positive estimates of the conditional variance. Lastly, the TGARCH(1,1,1) model extends to the GARCH model with taking the leverage effect into account as an extra dependent variable $\mathbf{1}(\varepsilon_{t-1} < 0)\varepsilon_{t-1}^2$. The model is covariance stationary if the $(\alpha + \gamma)/2 + \beta < 1$.

6.2 Empirical Data

In showing whether the Geometric-VaR test shows more power and to see which method performs best, we analyze daily returns on the Japanese NIKKEI-225 index, the American S&P-500 index and the Swiss SMI index. The daily data is obtained out of the Realized Library of the Oxford Man Institute of Quantitative Finance. Out of the obtained data from the Realized Library we compute the daily returns using the following formula. Note that P_t corresponds to the closing price of day t.

$$r_t = 100 \times \ln\Big(\frac{P_t}{P_{t-1}}\Big) \tag{20}$$

The first in-sample date for the rolling window starts at the first available observation in 2005. The size of the rolling window we obtain is given by $T_e = 1000$ for the GARCH specifications, which is compared to the Historical Simulation method with a rolling window consisting of 250 observations. The out-of-sample period roughly starts at the end of 2008, until the last available observation of 2017. This out-of-sample period is divided in a 2008-2013 sample, which we refer to as the crisis period, and a 2013-2017 sample. This latter period is referred to as the post-crisis period. Lastly, we take the whole sample altogether. Table 4 contains the summary statistics for the three indices.

Table 4: Summary statistics on daily returns

	Mean	Standard Deviation	Skewness	Kurtosis	Jarque-Bera	Observations
				2008-2017		
SMI	0.023	1.004	-0.598	8.774	3286.119	2268
NIKKEI	0.046	1.394	-0.433	7.626	2018.772	2187
S&P500	0.049	1.041	-0.295	8.235	2626.063	2271
				2008-2013		
SMI	0.028	1.076	-0.232	5.752	368.116	1134
NIKKEI	0.052	1.463	-0.617	7.506	995.114	1094
S&P500	0.054	1.271	-0.250	6.694	657.695	1136
				2013-2017		
SMI	0.019	0.926	-1.163	13.723	5688.659	1134
NIKKEI	0.041	1.322	-0.182	7.623	980.751	1093
S&P500	0.045	0.743	-0.405	6.155	501.908	1135

Note: This table contains the summary statistics the Swiss SMI index, the Japanese NIKKEI-225 index and the American S&P500 on the range of data given in the text. Stylized facts of returns seem to hold, since we have negative skewness and excess kurtosis. Normality among the daily returns does not hold.

6.3 Results

To provide empirical insights in the section of data, we follow the methods provided in the Extension section. For the SMI index we obtain the following summary statistics on the Value-at-Risk estimates in Table 5.

Table 5: Summary statistics 5% Value-at-Risk estimates

	UWMA	EWMA	IGARCH (1,1)	GARCH (1,1)	EGARCH (1,0,1)	TGARCH (1,1,1)
SMI			Skewed t-distribution			
Mean	-1.797	-1.673	-1.672	-1.628	-1.622	-1.616
St. Dev.	0.726	0.698	0.696	0.640	0.613	0.736
Violations	102	109	108	118	113	106
Coverage rate	0.045	0.048	0.048	0.052	0.050	0.047
			Filtered Historical Simulation			
Mean	..	-1.676	-1.675	-1.671	-1.657	-1.622
St. Dev.	..	0.710	0.709	0.659	0.637	0.751
Violations	..	110	110	105	107	103
Coverage rate	..	0.049	0.049	0.046	0.047	0.045

Note: This table denotes the mean, standard deviation, the amount of violations and the coverage rate for each model proposed in the Models section and for each dataset as has been proposed in the Data section for the 5% Value-at-Risk estimates. The first part of the table denotes the results for the methods that involve the skewed t-distribution. The second part denotes the results for the methods that utilize the Filtered historical simulation method. UWMA denotes the Historical Simulation method and EWMA denotes the RiskMetrics approach.

Summary statistics on both the NIKKEI and the S&P500 index can be found in the appendix at Table 9 and 10 respectively. Besides the summary statistics on each index, we also provide plots on the Value-at-Risk methods and plots of the mean coverage rate in the appendix. For the Value-at-Risk plots, the first subplot corresponds to the Historical Simulation method. The blue line corresponds to the daily returns and the red line corresponds to the Value-at-Risk estimated by Historical Simulation. For the other five subplots per index, the blue line corresponds to the daily returns, the red line corresponds to the respective method that involves the skewed t-distribution and the yellow line corresponds to the corresponding method that utilizes the Filtered Historical Simulation method.

We can clearly see how the Value-at-Risk methods correspond to the movement in daily returns. The lines of the skewed t-distribution and the Filtered Historical Simulation slightly coincide, since they roughly follow the same track. It is clear that the parametrized GARCH specifications seem to better capture volatility than the Historical Simulation method does. They thoroughly follow the silhouette of the daily returns. The Historical Simulation method however, fails to follow the movement on the daily returns and therefore seems to fail in capturing the daily volatility properly.

This finding is further illustrated with the provided plots of the mean coverage rate, given in the appendix. All subplots on the left side show the mean coverage rate of the skewed t-distribution methods. The subplots on the right side show the mean coverage rate of the filtered Historical Simulation methods. The results of the Historical Simulation approach correspond to the navy line in each graph, where the other lines correspond to the respective parametrized GARCH specifications. We can clearly see that the Historical Simulation method provides the least stable mean coverage rate, and only slowly converges to values close to 5% coverage.

Apart from these findings, we conduct the method provided in Dufour (2006). We analyze for 5% Value-at-Risk estimates, where we follow a similar procedure as in the section of Finite Sample power. To emulate independent Value-at-Risk estimates, we utilize a bootstrap method on the Value-at-Risk estimates from the sample. Each bootstrap has a length equal to five observations, which corresponds to one week of trading. We opt to bootstrap sequential observations instead of a singular bootstrap method to allow for volatility clustering.

The results of the Historical Simulation Dufour (2006) p-value approach are provided in Table 11. The results of the parametrized models are given in Table 12 until 17 for both the skewed-t distributed- and Filtered Historical Simulation models. The entries behind a 'I' in these tables correspond to the respective sample test statistics, where the entries behind a 'II' correspond to the respective Dufour (2006) p-values of the corresponding method.

For the Historical Simulation approach, we see that the Correct Unconditional Coverage test is only rejected twice at a 5% significance level. This is caused by the slow convergence of the empirical mean coverage rate to the desired rate. However, the method scores poorly on the other tests within the framework. The Duration Independence- and the Geometric test are rejected in all cases, where the Value-at-Risk Independence and the VaR test hold only three and two times respectively. The Geometric-VaR test shows most power of the tests within the framework, with all p-values below 5%. As a result of this finding, we can conclude that the Historical Simulation leads to misspecified Value-at-Risk estimates for all indices. The CaViaR test holds twice, which only contradicts the conclusion of misspecified Value-at-Risk in two occasions.

First we analyze the SMI index. The Correct Unconditional Coverage Test is never rejected, neither for the skewed-t distributed methods nor for the Filtered Historical Simulation methods. The same holds for the Duration Independence test. Testing both properties jointly on the Geometric Test also leads to zero rejections. The VaR Independence test is only rejected for a few samples where the RiskMetrics approach (EWMA) is executed. The same scenario holds for the VaR Test. The Geometric-VaR does not reject for any model specification that has not been rejected yet by one of the other tests of the framework in Pelletier & Wei (2015). It even holds for all RiskMetrics specifications in the Crisis and Post-Crisis period, which were rejected on the VaR Independence test or the VaR Test. Hence, the Geometric-VaR test does not uniformly show most power in all samples. The CaViaR p-values are lower in all cases, resulting in the CaViaR test showing more power than the Geometric-VaR test. A similar analysis can be made for both the NIKKEI and the S&P500 index.

Secondly, the NIKKEI index. In contrast to the results on the SMI index, most Dufour (2006) p-values for the skewed t-distributed methods are higher than the ones of the respective methods that utilize Filtered Historical Simulation. The Geometric-VaR shows most power in most cases, but occasionally generates higher p-values in some cases where it is undermined by the Duration Independence Test and the VaR Independence test. The Geometric-VaR test is rejected for all methods over the whole sample. The Geometric-VaR test fails to reject when tests in the third layer in Figure 1 are rejected for the small samples. However, the Geometric-VaR test never holds if one of its components in the second layer in Figure 1 are rejected. In contrast to the SMI index, the CaViaR Test uniformly generates higher p-values than the Geometric-VaR test and therefore conducts less power.

Lastly the S&P500 index. As with the NIKKEI index, the p-values for the skewed t-distributed methods are in general higher than the ones of the respective Filtered Historical Simulation methods. The property of correct unconditional coverage is only rejected once. For the skewed t-distributed methods, the Geometric-VaR Test is only rejected during the Post-Crisis period for the RiskMetrics approach. In contrast to the NIKKEI, we find a case where the Geometric-VaR tests holds even though one of its second-layer components is rejected. This is for the GARCH model in the Post-Crisis period. This is also the case with the Filtered Historical Simulation TGARCH model in the Crisis period. Furthermore, the Geometric-VaR test is rejected for most methods that utilize the Filtered Historical Simulation approach. The Geometric-VaR test uniformly shows more power than the CaViaR test, with all p-values falling below the p-values of the respective CaViaR test.

7 Conclusion and Recommendations

Value-at-Risk is the common used measurement of risk by practitioners, institutions and academics. The back-testing framework however, was limited. In relatively old literature by Christoffersen (1998), we could only tests for Duration Independence and Unconditional Coverage. This testing framework was rapidly extended. Pelletier & Wei (2015) ultimately provides a framework where properties could be tested jointly in combination with the property of Value-at-Risk independence.

In the first part of this paper we replicated the Monte Carlo methods Pelletier & Wei (2015) which utilizes the framework provided by Dufour (2006). In most cases the Geometric-VaR test seems to conduct more power than other tests within the framework. It even shows more power than tests such as the CaViaR test proposed in Engle & Manganelli (2004).

In addition to the above, we conducted an empirical research as extension to the paper. We analyzed the Swiss SMI, the Japanese NIKKEI and the American S&P500 index. With this research we showed that the Geometric-VaR test indeed possesses high power against various alternatives. In general, it shows most power in back-testing the Historical Simulation method, which scores abysmal on the tests within the framework. This can also be observed by a lacking ability to follow the silhouette of the daily returns and a generally slowly convergence of the mean coverage rate.

We further extended on several other methods to see whether the Geometric-VaR test would still conduct high power or not. We introduced parametrized models inspired by the risk framework in Wong et al. (2016), where the innovations follow a skewed t-distribution. The methods were further extended with a Filtered Historical Simulation approach, inspired by the findings in Bao et al. (2007). This paper states that methods with skewed innovations in general score better than methods that do not possess skewed innovations.

In terms of containing most power, the Geometric-VaR test shows high power when one of its second-layer components of Figure 1 is rejected. When either the VaR Test or the Geometric Test is rejected, the Geometric-VaR Test is rejected as well in most cases. The Geometric-VaR Test however fails to reject sometimes, when one of its third-layer components is rejected. However, the Geometric-VaR test generally possesses high power and the results are in line with both the theoretical- and empirical findings of Pelletier & Wei (2015).

For the Japanese NIKKEI and the American S&P500, we can conclude that the methods that utilize skewed innovations indeed perform better in general. For the SMI we were not able to draw such a strong conclusion about which innovation-driven set of methods scored better. In choosing an optimal model for obtaining the Value-at-Risk, Historical Simulation does not seem adequate. Within the set of parametrized methods, the RiskMetrics approach scores the worst in general. Among the other methods we can not find an inferior method.

As recommendations for further research, one could investigate whether the implementation of intra-day data leads to better-specified Value-at-Risk estimates than the model specifications we exploited do. One could also investigate other methods of estimating Value-at-Risk, where normality or a general t-distribution holds amongst the standardized innovations.

As other recommendations to our research, one could use another set of explanatory variables in fulfilling the CaViaR test and see whether this provides different results or not. Another suggestion would be to include other explanatory variables in the hazard function of each test.

We could also differ the approach of how we look at Value-at-Risk. For example, we only look at violations where the actual daily return falls below the Value-at-Risk estimation. However, the magnitude of exceedance can also be informative of how Value-at-Risk specifications perform.

Lopez (1997), Berkowitz (2001) and Colletaz et al. (2013) propose research on this topic which could also be used for further research.

Appendix

Table 6: Power of 10% Duration Test applied to 1% VaR estimates

Sample Size	UC	Duration ind.	VaR ind.	Geometric	VaR	GV.	CaViaR
			Business Line 1				
250	0.103	0.368	0.232	0.217	0.273	0.238	0.433
500	0.039	0.360	0.511	0.214	0.311	0.415	0.524
750	0.033	0.496	0.674	0.255	0.409	0.564	0.632
1000	0.086	0.560	0.768	0.367	0.549	0.739	0.713
1250	0.109	0.700	0.828	0.476	0.674	0.826	0.795
1500	0.128	0.738	0.892	0.547	0.763	0.882	0.840
			Business Line 2				
250	0.180	0.309	0.201	0.257	0.239	0.284	0.477
500	0.094	0.487	0.440	0.294	0.301	0.441	0.529
750	0.103	0.609	0.603	0.431	0.424	0.615	0.606
1000	0.186	0.669	0.664	0.486	0.561	0.737	0.663
1250	0.230	0.745	0.715	0.559	0.616	0.808	0.722
1500	0.277	0.791	0.774	0.651	0.711	0.874	0.792
			Business Line 3				
250	0.051	0.157	0.376	0.070	0.052	0.086	0.344
500	0.013	0.152	0.654	0.038	0.138	0.158	0.399
750	0.008	0.158	0.802	0.049	0.396	0.346	0.502
1000	0.025	0.151	0.868	0.045	0.537	0.519	0.604
1250	0.031	0.148	0.922	0.042	0.684	0.679	0.704
1500	0.041	0.151	0.952	0.058	0.796	0.773	0.773
			Business Line 4				
250	0.178	0.333	0.197	0.164	0.243	0.272	0.522
500	0.100	0.556	0.455	0.350	0.322	0.479	0.555
750	0.124	0.679	0.588	0.498	0.433	0.658	0.631
1000	0.223	0.752	0.686	0.585	0.582	0.791	0.721
1250	0.283	0.821	0.746	0.648	0.701	0.873	0.773
1500	0.296	0.843	0.789	0.716	0.738	0.881	0.820

Note: This table provides the results of the Duration tests of Pelletier & Wei (2015) for 1% Value-at-Risk estimates out of Historical Simulation. How each entry is determined is clarified in the text.

Table 7: Power of 10% Duration Test applied to 5% VaR estimates

Sample Size	UC	Duration ind.	VaR ind.	Geometric	VaR	GV.	CaViaR
			Business Line 1				
250	0.155	0.475	0.313	0.349	0.267	0.423	0.423
500	0.062	0.679	0.646	0.472	0.462	0.693	0.510
750	0.037	0.783	0.796	0.612	0.625	0.842	0.598
1000	0.024	0.860	0.856	0.703	0.716	0.913	0.692
1250	0.019	0.908	0.911	0.790	0.794	0.954	0.771
1500	0.016	0.938	0.943	0.825	0.854	0.978	0.859
			Business Line 2				
250	0.351	0.452	0.358	0.506	0.506	0.579	0.578
500	0.227	0.706	0.614	0.622	0.626	0.776	0.651
750	0.182	0.828	0.698	0.746	0.700	0.876	0.667
1000	0.170	0.895	0.753	0.809	0.751	0.934	0.720
1250	0.158	0.933	0.800	0.860	0.797	0.963	0.767
1500	0.153	0.962	0.813	0.907	0.836	0.983	0.824
			Business Line 3				
250	0.059	0.133	0.399	0.075	0.135	0.138	0.289
500	0.015	0.147	0.709	0.050	0.295	0.271	0.374
750	0.005	0.152	0.857	0.056	0.475	0.447	0.443
1000	0.004	0.155	0.922	0.045	0.674	0.638	0.531
1250	0.002	0.164	0.957	0.048	0.778	0.726	0.617
1500	0.002	0.178	0.979	0.051	0.861	0.819	0.832
			Business Line 4				
250	0.343	0.504	0.361	0.519	0.488	0.595	0.579
500	0.216	0.760	0.638	0.660	0.643	0.814	0.627
750	0.176	0.865	0.719	0.782	0.713	0.901	0.698
1000	0.159	0.928	0.774	0.864	0.769	0.953	0.737
1250	0.155	0.961	0.835	0.905	0.831	0.976	0.801
1500	0.145	0.977	0.850	0.943	0.877	0.987	0.845

Note: This table provides the results of the Duration tests of Pelletier & Wei (2015) for 5% Value-at-Risk estimates of Historical Simulation. How each entry is determined is clarified in the text.

Table 8: Feasibility proportion for each test

VaR Coverage Rate	Sample Size	Geometric Test	Geometric-VaR Test
		Business Line 1	
1%	250	0.8054	0.6144
1%	500	0.9932	0.9756
1%	750	0.9998	0.9996
5%	250	0.9996	0.9986
		Business Line 2	
1%	250	0.7386	0.5768
1%	500	0.9864	0.9554
1%	750	0.9998	0.9982
5%	250	0.9890	0.9770
		Business Line 3	
1%	250	0.7952	0.6106
1%	500	0.9984	0.9890
1%	750	1.0000	1.0000
5%	250	1.0000	1.0000
		Business Line 4	
1%	250	0.7604	0.5858
1%	500	0.9880	0.9574
1%	750	0.9992	0.9982
5%	250	0.9932	0.9832

Note: This table denotes the fraction of feasibility for each test. The selection criteria that have been chosen are more than two durations for the Geometric test and more than three durations for the Geometric-VaR test.

Table 9: Summary statistics 5% Value-at-Risk estimates NIKKEI

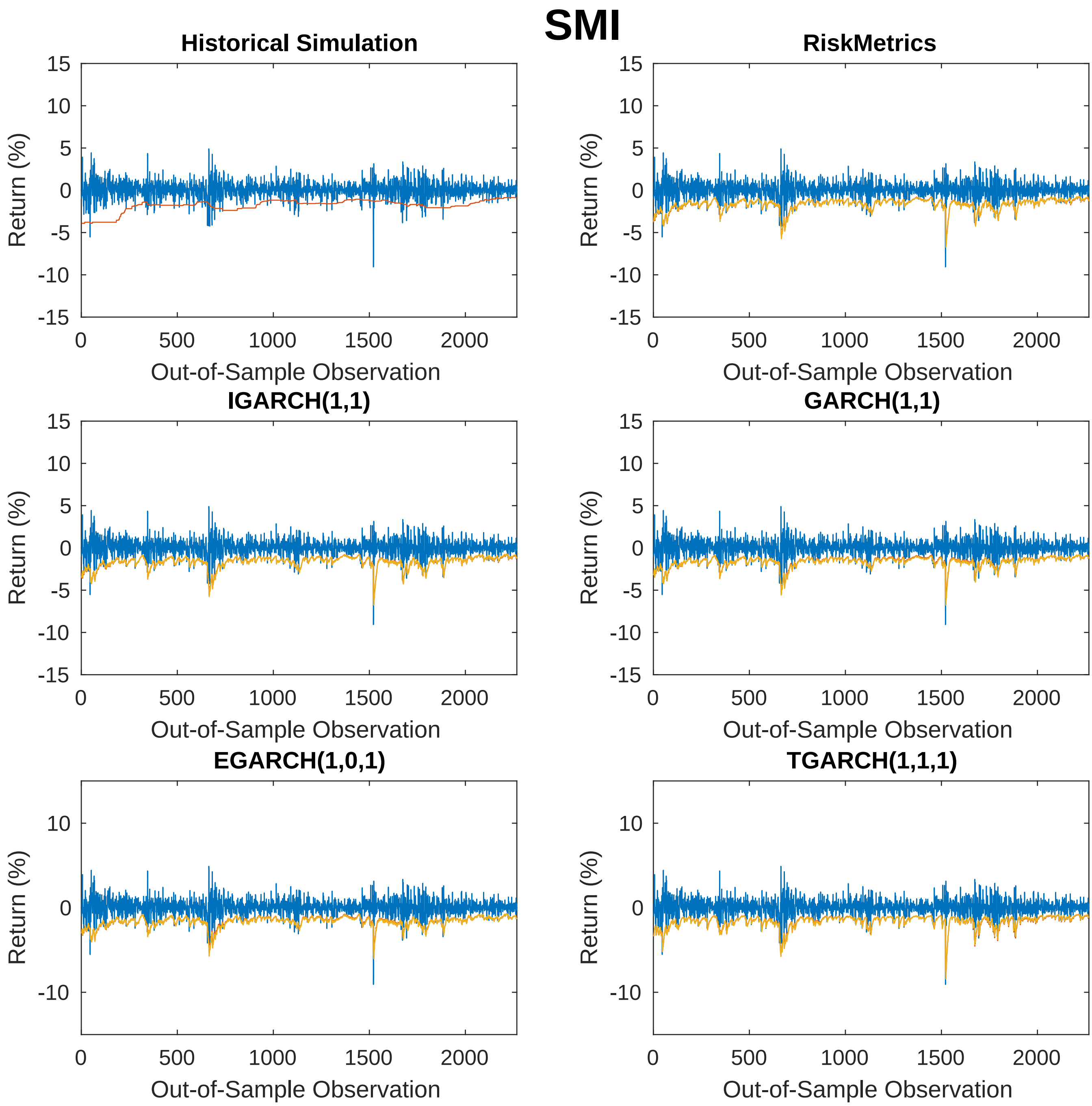
	Historical Simulation	RiskMetrics	GARCH (1,1)	EGARCH (1,0,1)	TGARCH (1,1,1)	IGARCH (1,1)
NIKKEI			Skewed t-distribution			
Mean	-2.475	-2.386	-2.385	-2.317	-2.299	-2.277
St. Dev.	0.919	0.871	0.870	0.771	0.680	0.839
Violations	89	97	102	100	101	96
Coverage rate	0.041	0.044	0.047	0.046	0.046	0.044
			Filtered Historical Simulation			
Mean	..	-2.398	-2.396	-2.373	-2.353	-2.255
St. Dev.	..	0.883	0.882	0.810	0.723	0.846
Violations	..	94	98	94	93	99
Coverage rate	..	0.043	0.045	0.043	0.043	0.045

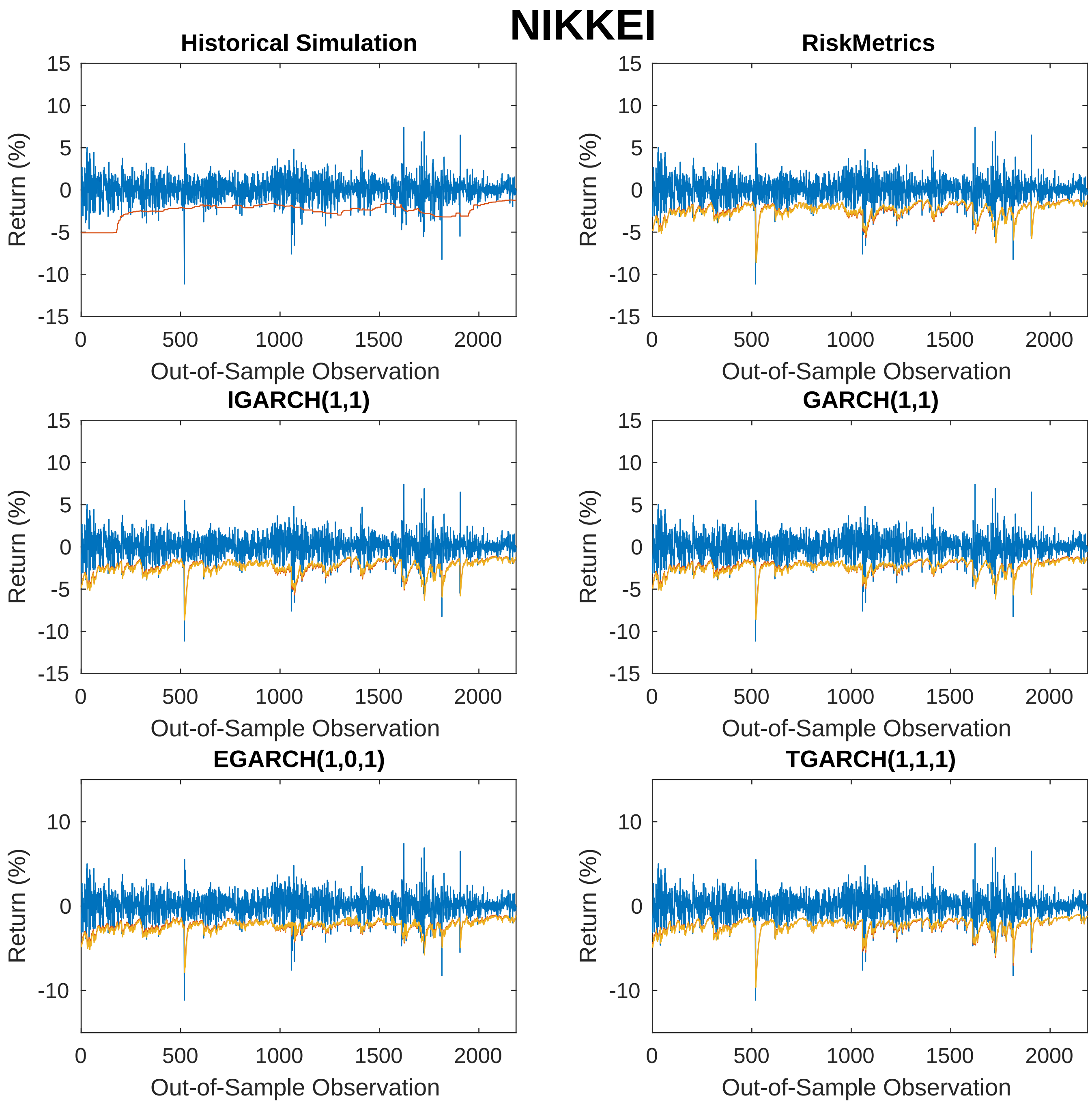
Note: This table denotes the mean, standard deviation, the amount of violations and the coverage rate for each model proposed in the Models section and for each dataset as has been proposed in the Data section for the 5% Value-at-Risk estimates.

Table 10: Summary statistics 5% Value-at-Risk estimates S&P500

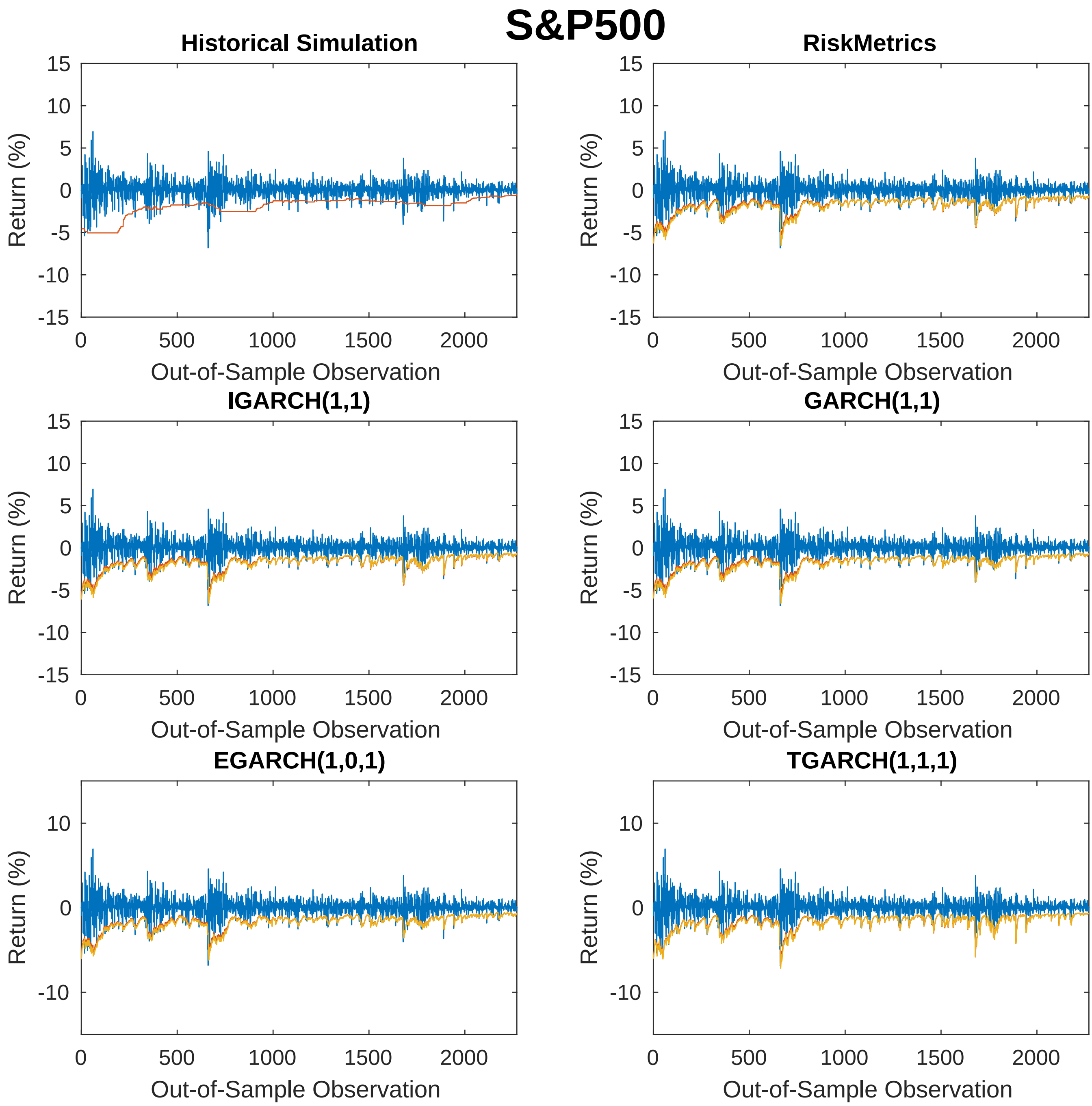
	Historical Simulation	RiskMetrics	IGARCH (1,1)	GARCH (1,1)	EGARCH (1,0,1)	TGARCH (1,1,1)
S&P500			Skewed t-distribution			
Mean	-1.905	-1.700	-1.697	-1.669	-1.674	-1.679
St. Dev.	1.114	0.873	0.869	0.849	0.862	0.945
Violations	94	99	101	102	105	102
Coverage rate	0.041	0.044	0.045	0.045	0.046	0.045
			Filtered Historical Simulation			
Mean	..	-1.774	-1.772	-1.770	-1.762	-1.737
St. Dev.	..	0.986	0.982	0.962	0.951	1.002
Violations	..	96	96	89	92	95
Coverage rate	..	0.042	0.042	0.039	0.041	0.042

Note: This table denotes the mean, standard deviation, the amount of violations and the coverage rate for each model proposed in the Models section and for each dataset as has been proposed in the Data section for the 5% Value-at-Risk estimates.

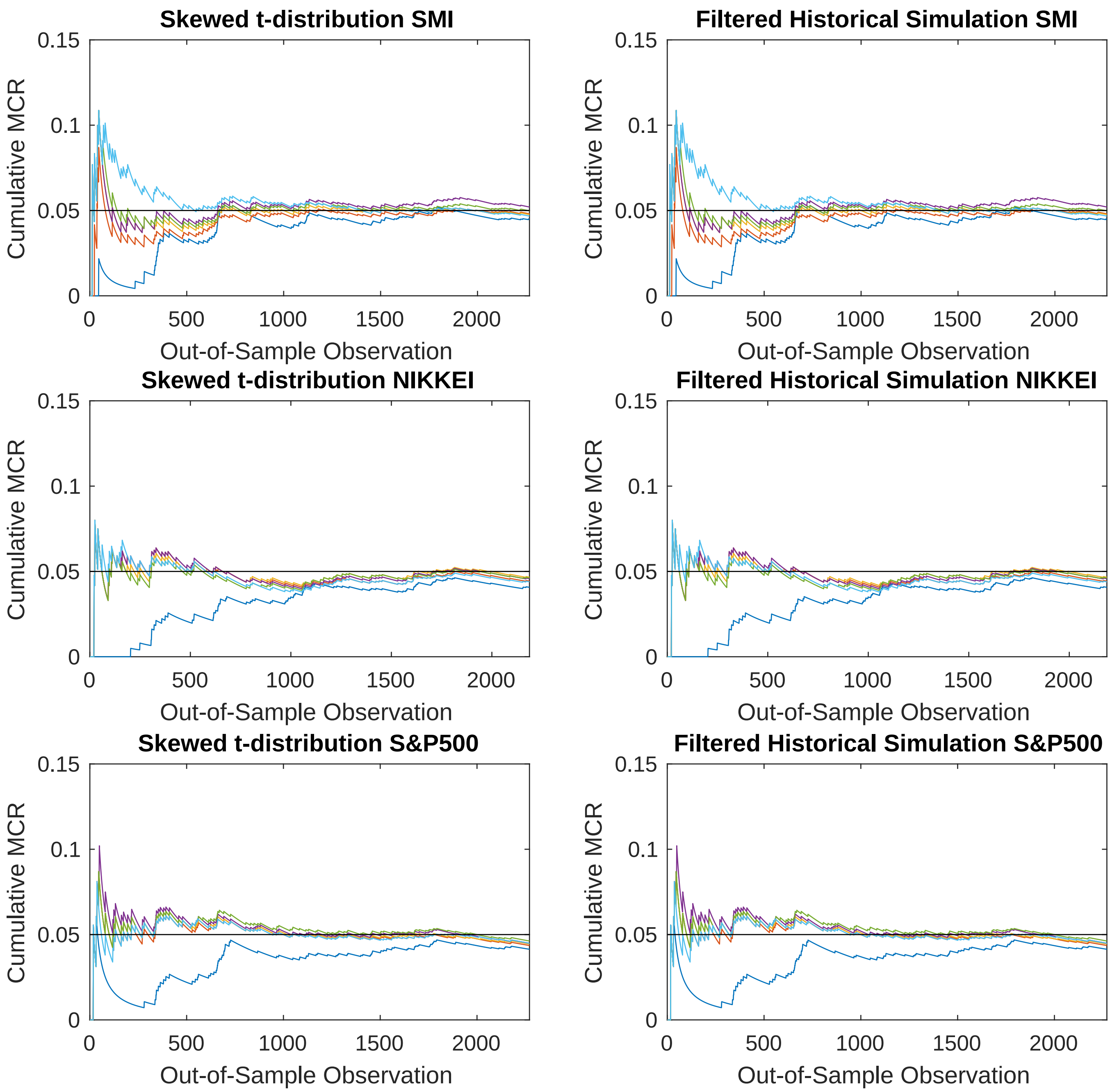

SMI
Historical Simulation
RiskMetrics
IGARCH(1,1)
GARCH(1,1)
EGARCH(1,0,1)
TGARCH(1,1,1)
Return (%)
Out-of-Sample Observation
15
10
5
0
-5
-10
-15
500
1000
1500
2000



NIKKEI
Historical Simulation
RiskMetrics
IGARCH(1,1)
GARCH(1,1)
EGARCH(1,0,1)
TGARCH(1,1,1)
Return (%)
Out-of-Sample Observation
15
10
5
0
-5
-10
-15
500
1000
1500
2000



S&P500
Historical Simulation
RiskMetrics
IGARCH(1,1)
GARCH(1,1)
EGARCH(1,0,1)
TGARCH(1,1,1)
Return (%)
Out-of-Sample Observation
15
10
5
0
-5
-10
-15
500
1000
1500
2000



Skewed t-distribution SMI
Filtered Historical Simulation SMI
Skewed t-distribution NIKKEI
Filtered Historical Simulation NIKKEI
Skewed t-distribution S&P500
Filtered Historical Simulation S&P500
Cumulative MCR
Out-of-Sample Observation
0.15
0.1
0.05
0
500
1000
1500
2000

Table 11: Empirical Results: Historical Simulation on three indices

			Uncond. Cov.	Duration Ind.	VaR Ind.	Geom.	VaR Test	Geom. VaR	CaViaR
SMI									
	2008-2017	I	1.457	50.893	16.650	52.349	32.218	68.990	39.1000
		II	0.217	0.000	0.000	0.000	0.000	0.000	0.000
	2008-2013	I	0.128	34.929	22.509	35.056	40.850	57.565	38.952
		II	0.729	0.000	0.000	0.000	0.000	0.000	0.000
	2013-2017	I	2.226	17.296	1.741	19.522	6.011	21.263	11.020
		II	0.148	0.000	0.092	0.000	0.038	0.000	0.005
NIKKEI									
	2008-2017	I	4.648	29.406	12.177	34.054	19.316	46.230	23.900
		II	0.035	0.000	0.000	0.000	0.000	0.000	0.000
	2008-2013	I	2.790	21.968	14.847	24.758	27.182	39.604	26.992
		II	0.099	0.000	0.000	0.000	0.000	0.000	0.000
	2013-2017	I	2.293	8.505	0.737	10.798	2.341	11.535	4.105
		II	0.138	0.000	0.199	0.003	0.234	0.005	0.129
S&P500									
	2008-2017	I	4.124	36.572	7.471	40.695	17.836	48.166	20.005
		II	0.045	0.000	0.003	0.000	0.000	0.000	0.000
	2008-2013	I	3.788	21.069	11.683	24.857	24.232	26.540	21.575
		II	0.053	0.000	0.000	0.000	0.000	0.000	0.000
	2013-2017	I	1.137	15.383	0.029	16.520	1.137	16.549	3.808
		II	0.311	0.000	0.439	0.000	0.445	0.000	0.297

Note: This table denotes the Monte Carlo p-values of the Historical Simulation method out of Dufour (2006) for three indices and three distinct periods. For further description of the table, see the corresponding results section.

Table 12: Empirical results skewed t-distribution SMI

			Uncond. Cov.	Duration Ind.	VaR Ind.	Geom.	VaR Test	Geom. VaR	CaViaR
SMI			2008-2017						
	EWMA	I	0.265	0.806	6.372	1.071	5.220	7.442	8.987
		II	0.609	0.160	0.006	0.419	0.050	0.029	0.010
	IGARCH	I	0.375	0.122	1.264	0.497	1.248	1.761	1.779
		II	0.556	0.322	0.135	0.625	0.429	0.423	0.098
	GARCH	I	0.126	2.349	1.351	2.474	0.408	3.825	2.117
		II	0.710	0.054	0.129	0.196	0.726	0.152	0.086
	EGARCH	I	0.016	0.280	0.824	0.296	0.442	1.120	0.959
		II	0.887	0.267	0.197	0.715	0.712	0.566	0.106
	TGARCH	I	0.655	0.001	0.160	0.655	0.815	0.815	0.708
		II	0.415	0.774	0.604	0.560	0.555	0.657	0.111
			2008-2013						
	EWMA	I	0.007	0.032	4.320	0.039	3.935	4.359	5.323
		II	0.920	0.375	0.019	0.903	0.101	0.120	0.061
	IGARCH	I	0.106	0.453	1.474	0.558	0.978	2.032	0.543
		II	0.765	0.208	0.127	0.593	0.526	0.381	0.211
	GARCH	I	0.736	1.187	2.235	1.923	1.944	4.158	1.479
		II	0.407	0.115	0.071	0.254	0.303	0.136	0.177
	EGARCH	I	0.352	0.111	0.475	0.463	0.629	0.937	0.604
		II	0.572	0.323	0.285	0.626	0.645	0.636	0.206
	TGARCH	I	0.352	0.001	0.268	0.352	0.620	0.620	0.699
		II	0.578	0.655	0.480	0.679	0.646	0.737	0.205
			2013-2017						
	EWMA	I	0.602	1.524	2.637	2.127	2.140	4.764	11.901
		II	0.460	0.082	0.056	0.233	0.004	0.098	0.010
	IGARCH	I	1.815	0.001	0.386	1.815	2.201	2.201	4.471
		II	0.181	0.910	0.326	0.282	0.264	0.345	0.075
	GARCH	I	0.246	1.258	0.078	1.504	0.246	1.581	2.802
		II	0.628	0.101	0.817	0.325	0.894	0.467	0.111
	EGARCH	I	0.842	0.193	0.916	1.035	1.368	1.951	4.343
		II	0.373	0.291	0.183	0.432	0.407	0.388	0.071
	TGARCH	I	3.738	0.001	0.172	3.738	3.910	3.910	3.953
		II	0.051	0.771	0.720	0.094	0.111	0.156	0.086

Note: This table denotes the Dufour (2006) p-values and test statistics for the methods with skewed innovations on the Swiss SMI index.

Table 13: Empirical results Filtered Historical Simulation SMI

		Uncond. Cov.	Duration Ind.	VaR Ind.	Geom.	VaR Test	Geom. VaR	CaViaR
SMI				2008-2017				
EWMA	I	0.174	0.224	6.713	0.398	5.900	7.111	9.122
	II	0.670	0.277	0.006	0.652	0.037	0.034	0.009
IGARCH	I	0.174	0.035	0.923	0.208	0.870	1.131	1.375
	II	0.676	0.388	0.175	0.757	0.531	0.567	0.100
GARCH	I	0.825	0.472	0.251	1.297	0.860	1.548	1.935
	II	0.366	0.217	0.363	0.374	0.537	0.462	0.100
EGARCH	I	0.506	0.277	1.738	0.783	1.631	2.521	2.023
	II	0.493	0.267	0.098	0.519	0.349	0.288	0.093
TGARCH	I	1.226	0.001	1.303	1.226	2.528	2.528	2.656
	II	0.267	0.719	0.133	0.394	0.210	0.289	0.086
				2008-2013				
EWMA	I	0.003	0.001	4.547	0.003	4.402	4.550	5.293
	II	0.981	0.888	0.017	0.990	0.087	0.116	0.063
IGARCH	I	0.527	0.255	1.342	0.781	1.397	2.123	1.147
	II	0.464	0.263	0.140	0.509	0.410	0.366	0.193
GARCH	I	0.036	0.381	0.875	0.417	0.460	1.292	0.215
	II	0.881	0.221	0.201	0.629	0.738	0.534	0.224
EGARCH	I	0.036	0.195	1.356	0.231	0.973	1.587	0.548
	II	0.854	0.279	0.131	0.751	0.535	0.470	0.211
TGARCH	I	0.128	0.001	1.208	0.128	1.336	1.336	1.248
	II	0.729	0.805	0.149	0.809	0.420	0.522	0.188
				2013-2017				
EWMA	I	0.602	0.844	2.800	1.446	2.537	4.246	11.523
	II	0.448	0.150	0.046	0.350	0.212	0.125	0.003
IGARCH	I	2.226	0.001	0.302	2.226	2.528	2.528	4.764
	II	0.154	0.772	0.427	0.228	0.222	0.297	0.068
GARCH	I	2.684	0.112	0.001	2.795	0.268	2.795	5.279
	II	0.104	0.314	0.964	0.157	0.204	0.261	0.055
EGARCH	I	1.815	0.093	0.932	1.908	2.432	2.840	4.542
	II	0.173	0.324	0.185	0.248	0.227	0.249	0.067
TGARCH	I	1.815	0.001	0.454	1.815	2.269	2.269	2.263
	II	0.200	0.555	0.310	0.271	0.259	0.339	0.115

Note: This table denotes the Dufour (2006) p-values and test statistics for the methods following Filtered Historical Simulation on the Swiss SMI index.

Table 14: Empirical results skewed t-distribution NIKKEI

			Uncond. Cov.	Duration Ind.	VaR Ind.	Geom.	VaR Test	Geom. VaR	CaViaR
NIKKEI						2008-2017			
	EWMA	I	1.761	4.297	2.870	6.058	2.403	8.928	5.193
		II	0.187	0.016	0.049	0.031	0.221	0.016	0.088
	IGARCH	I	0.672	1.499	4.212	2.171	2.467	6.383	2.145
		II	0.436	0.097	0.020	0.234	0.208	0.042	0.181
	GARCH	I	1.044	3.731	4.842	4.775	2.306	9.617	2.130
		II	0.321	0.021	0.014	0.053	0.228	0.009	0.175
	EGARCH	I	0.847	3.517	5.051	4.365	2.399	9.416	1.351
		II	0.372	0.025	0.013	0.067	0.215	0.009	0.241
	TGARCH	I	2.043	0.743	4.027	2.786	3.506	6.812	3.478
		II	0.150	0.164	0.022	0.160	0.122	0.035	0.120
						2008-2013			
	EWMA	I	1.885	0.870	3.408	2.756	3.860	6.163	5.274
		II	0.170	0.144	0.038	0.170	0.114	0.052	0.080
	IGARCH	I	1.503	0.023	3.099	1.526	4.007	4.615	4.170
		II	0.230	0.382	0.043	0.324	0.106	0.112	0.116
	GARCH	I	2.314	0.246	2.739	2.560	3.990	5.299	4.217
		II	0.147	0.260	0.050	0.177	0.101	0.075	0.112
	EGARCH	I	1.503	1.656	5.653	3.159	4.431	8.812	3.995
		II	0.228	0.078	0.010	0.137	0.086	0.015	0.124
	TGARCH	I	3.314	0.006	1.962	3.320	4.590	5.282	4.172
		II	0.070	0.411	0.091	0.119	0.081	0.081	0.113
						2013-2017			
	EWMA	I	0.411	3.855	0.359	4.266	0.411	4.625	1.285
		II	0.531	0.016	0.301	0.065	0.723	0.097	0.291
	IGARCH	I	0.006	2.195	1.269	2.201	0.071	3.470	0.025
		II	0.904	0.054	0.133	0.227	0.963	0.179	0.389
	GARCH	I	0.006	4.347	1.583	4.353	0.015	5.936	0.046
		II	0.948	0.013	0.104	0.067	0.984	0.056	0.385
	EGARCH	I	0.047	1.778	0.488	1.825	0.047	2.313	0.793
		II	0.837	0.072	0.256	0.274	0.971	0.317	0.320
	TGARCH	I	0.127	1.033	1.392	1.161	0.283	2.553	0.343
		II	0.735	0.126	0.127	0.395	0.802	0.289	0.362

Note: This table denotes the Dufour (2006) p-values and test statistics for the methods with skewed innovations on the Japanese NIKKEI index.

Table 15: Empirical results Filtered Historical Simulation NIKKEI

			Uncond. Cov.	Duration Ind.	VaR Ind.	Geom.	VaR Test	Geom. VaR	CaViaR
NIKKEI						2008-2017			
	EWMA	I	2.674	3.775	4.840	6.449	4.668	11.289	5.297
		II	0.010	0.022	0.013	0.023	0.066	0.004	0.083
	IGARCH	I	1.500	4.454	7.044	5.954	3.933	12.998	3.099
		II	0.218	0.012	0.004	0.028	0.091	0.002	0.139
	GARCH	I	2.674	5.561	5.756	8.235	3.950	13.991	3.003
		II	0.104	0.007	0.008	0.009	0.096	0.001	0.146
	EGARCH	I	3.023	4.440	7.606	7.463	6.033	16.069	3.670
		II	0.086	0.015	0.003	0.014	0.034	0.001	0.120
	TGARCH	I	1.261	2.106	3.752	3.367	2.046	7.092	1.740
		II	0.264	0.058	0.025	0.118	0.271	0.032	0.202
						2008-2013			
	EWMA	I	6.702	0.026	2.259	6.728	8.489	8.987	7.944
		II	0.101	0.382	0.074	0.016	0.010	0.010	0.028
	IGARCH	I	5.189	0.256	2.403	5.441	6.481	7.847	5.998
		II	0.025	0.264	0.063	0.038	0.031	0.024	0.062
	GARCH	I	8.440	0.160	0.486	8.600	8.584	9.086	7.658
		II	0.005	0.304	0.281	0.007	0.010	0.012	0.028
	EGARCH	I	7.524	0.847	4.135	8.389	9.692	12.524	8.279
		II	0.008	0.146	0.026	0.011	0.007	0.003	0.023
	TGARCH	I	5.918	0.041	1.677	5.959	6.846	7.636	6.145
		II	0.016	0.368	0.103	0.028	0.024	0.025	0.054
						2013-2017			
	EWMA	I	0.004	4.532	1.120	4.356	0.032	5.656	0.314
		II	0.957	0.012	0.150	0.062	0.984	0.062	0.365
	IGARCH	I	0.114	4.362	2.538	4.477	0.374	7.015	0.289
		II	0.754	0.011	0.060	0.067	0.752	0.033	0.367
	GARCH	I	0.114	5.351	2.326	5.465	0.215	7.791	0.230
		II	0.768	0.007	0.069	0.039	0.856	0.025	0.366
	EGARCH	I	0.004	2.695	1.344	2.699	0.064	4.043	0.286
		II	0.993	0.039	0.131	0.177	0.966	0.143	0.373
	TGARCH	I	0.373	1.784	0.321	2.157	0.373	2.478	0.693
		II	0.570	0.072	0.329	0.233	0.750	0.300	0.315

Note: This table denotes the Dufour (2006) p-values and test statistics for the methods following Filtered Historical Simulation on the Japanese NIKKEI index.

Table 16: Empirical results skewed t-distribution S&P500

			Uncond. Cov.	Duration Ind.	VaR Ind.	Geom.	VaR Test	Geom. VaR	CaViaR
S&P500					2008-2017				
	EWMA	I	2.317	1.062	1.648	3.379	3.132	5.027	2.921
		II	0.139	0.134	0.100	0.118	0.152	0.087	0.410
	IGARCH	I	1.745	0.053	2.035	1.798	3.344	3.833	2.936
		II	0.188	0.368	0.080	0.276	0.132	0.150	0.415
	GARCH	I	1.491	1.849	1.040	3.340	1.683	3.379	1.948
		II	0.234	0.075	0.155	0.120	0.319	0.117	0.574
	EGARCH	I	0.851	0.025	0.579	0.877	1.287	1,456	1.609
		II	0.346	0.398	0.213	0.470	0.387	0.451	0.665
	TGARCH	I	1.491	0.001	1.656	1.491	3.147	3.147	3.836
		II	0.226	0.495	0.100	0.332	0.147	0.206	0.289
					2008-2013				
	EWMA	I	0.009	0.101	1.992	0.110	1.684	2.102	1.653
		II	0.920	0.327	0.083	0.826	0.331	0.343	0.651
	IGARCH	I	0.009	0.181	3.219	0.190	2.481	3.409	2.885
		II	0.930	0.285	0.037	0.780	0.209	0.177	0.421
	GARCH	I	0.002	0.304	2.167	0.305	1.479	2.472	2.439
		II	0.967	0.242	0.072	0.703	0.372	0.279	0.493
	EGARCH	I	0.097	0.001	2.101	0.097	2.198	2.198	3.275
		II	0.748	0.689	0.077	0.838	0.243	0.318	0.352
	TGARCH	I	0.054	0.001	2.852	0.054	2.906	2.906	2.992
		II	0.808	0.668	0.042	0.882	0.164	0.223	0.401
					2013-2017				
	EWMA	I	5.014	0.973	2.836	5.987	6.299	8.823	8.206
		II	0.027	0.132	0.046	0.029	0.034	0.014	0.045
	IGARCH	I	3.763	0.001	1.701	3.763	5.319	5.464	6.076
		II	0.051	0.785	0.104	0.092	0.050	0.070	0.129
	GARCH	I	3.763	1.543	0.766	5.306	3.767	6.072	3.484
		II	0.055	0.083	0.200	0.044	0.109	0.054	0.339
	EGARCH	I	3.210	0.295	0.530	3.505	3.319	4.035	2.927
		II	0.069	0.249	0.239	0.110	0.134	0.136	0.406
	TGARCH	I	2.705	0.001	0.389	2.705	3.094	3.094	3.061
		II	0.097	0.612	0.281	0.164	0.154	0.213	0.399

Note: This table denotes the Dufour (2006) p-values and test statistics for the methods with skewed innovations on the American S&P500 index.

Table 17: Empirical results Filtered Historical Simulation S&P500

			Uncond. Cov.	Duration Ind.	VaR Ind.	Geom.	VaR Test	Geom. VaR	CaViaR
S&P500					2008-2017				
	EWMA	I	3.335	0.806	5.218	4.141	7.607	9.359	7.747
		II	0.068	0.158	0.012	0.076	0.013	0.010	0.052
	IGARCH	I	3.718	0.142	3.934	3.861	7.000	7.795	5.773
		II	0.059	0.321	0.025	0.092	0.021	0.022	0.129
	GARCH	I	6.491	0.607	2.505	7.098	8.101	9.603	6.924
		II	0.013	0.195	0.059	0.016	0.012	0.009	0.076
	EGARCH	I	5.002	0.112	2.331	5.113	6.881	7.444	5.578
		II	0.027	0.333	0.064	0.043	0.021	0.025	0.141
	TGARCH	I	3.718	0.001	3.606	3.718	7.324	7.324	6.688
		II	0.051	0.568	0.030	0.099	0.014	0.027	0.086
					2008-2013				
	EWMA	I	3.234	0.338	2.216	3.572	4.928	5.788	4.583
		II	0.069	0.237	0.069	0.098	0.058	0.057	0.230
	IGARCH	I	2.266	0.123	3.123	2.389	4.794	5.512	3.617
		II	0.138	0.323	0.039	0.197	0.065	0.068	0.310
	GARCH	I	2.727	0.006	2.519	2.732	4.959	5.251	3.775
		II	0.110	0.411	0.055	0.161	0.062	0.074	0.302
	EGARCH	I	2.727	0.001	2.320	2.727	4.951	5.047	3.796
		II	0.104	0.790	0.065	0.171	0.063	0.089	0.303
	TGARCH	I	1.151	0.001	4.722	1.151	5.873	5.873	4.564
		II	0.297	0.574	0.014	0.407	0.034	0.052	0.235
					2013-2017				
	EWMA	I	0.854	0.456	4.468	1.311	3.844	5.779	7.427
		II	0.359	0.205	0.019	0.361	0.111	0.066	0.069
	IGARCH	I	1.832	0.055	2.130	1.887	3.273	4.017	4.051
		II	0.183	0.355	0.078	0.258	0.137	0.138	0.277
	GARCH	I	4.364	1.102	1.569	5.466	4.681	7.036	4.908
		II	0.036	0.117	0.112	0.036	0.068	0.028	0.208
	EGARCH	I	2.705	0.377	0.601	3.082	2.824	3.683	2.443
		II	0.103	0.226	0.227	0.141	0.177	0.165	0.496
	TGARCH	I	3.210	0.001	0.707	3.210	3.917	3.917	3.831
		II	0.066	0.527	0.215	0.119	0.103	0.146	0.298

Note: This table denotes the Dufour (2006) p-values and test statistics for the methods following Filtered Historical Simulation on the American S&P500 index.

References

Bao, Y., Lee, T.-H., & Saltoğlu, B. (2007). Comparing density forecast models. *Journal of Forecasting*, *26*(3), 203–225.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, *19*(4), 465–474.

Berkowitz, J., Christoffersen, P., & Pelletier, D. (2011). Evaluating value-at-risk models with desk-level data. *Management Science*, *57*(12), 2213–2227.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *31*(3), 307–327.

Candelon, B., Colletaz, G., Hurlin, C., & Tokpavi, S. (2010). Backtesting value-at-risk: a gmm duration-based test. *Journal of Financial Econometrics*, *9*(2), 314–343.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International economic review*, 841–862.

Colletaz, G., Hurlin, C., & Pérignon, C. (2013). The risk map: A new tool for validating risk models. *Journal of Banking & Finance*, *37*(10), 3843–3854.

Dufour, J.-M. (2006). Monte carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics. *Journal of Econometrics*, *133*(2), 443–477.

Engle, R. F., & Manganelli, S. (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, *22*(4), 367–381.

Haas, M. (2005). Improved duration-based backtesting of value-at-risk. *The Journal of Risk*, *8*(2), 17.

Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, 705–730.

Lopez, J. A. (1997). Regulatory evaluation of value-at-risk models.

Morgan, J., et al. (1996). Riskmetrics technical document.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347–370.

Pelletier, D., & Wei, W. (2015). The geometric-var backtesting method. *Journal of financial econometrics*, *14*(4), 725–745.

Pérignon, C., & Smith, D. R. (2010). The level and quality of value-at-risk disclosure by commercial banks. *Journal of Banking & Finance*, *34*(2), 362–377.

Wong, Z. Y., Chin, W. C., & Tan, S. H. (2016). Daily value-at-risk modeling and forecast evaluation: The realized volatility approach. *The Journal of Finance and Data Science*, *2*(3), 171–187.