Dependent bootstrapping for value-at-risk and expected shortfall

Ian Laker[1] · Chun-Kai Huang[1] · Allan Ernest Clark[1]

Abstract Estimation in extreme financial risk is often faced with challenges such as the need for adequate distributional assumptions, considerations for data dependencies, and the lack of tail information. Bootstrapping provides an alternative that overcomes some of these challenges. It does not assume a distributional form and asymptotically replicates the empirical density for resampled data. Moreover, advanced bootstrapping can cater for dependencies and stationarity in the data. In this paper, we evaluate the use of dependent bootstrapping, both for the original financial time series and for its GARCH innovations (under the Gaussian and Student t noise assumptions), in forecasting value-at-risk and expected shortfall. We also assess the effect of using different window sizes for these procedures. The two datasets used are daily returns of the S&P500 from NYSE and the ALSI from JSE.

✉ Chun-Kai Huang
chun-kai.huang@uct.ac.za

Ian Laker
ian.laker@alumni.uct.ac.za

Allan Ernest Clark
allan.clark@uct.ac.za

[1] Department of Statistical Sciences, University of Cape Town, Rondebosch, Cape Town 7701, South Africa

Introduction

The financial world is marred by levels of uncertainty and there is an ever increasing need for financial institutions to estimate the risk of various positions. In particular, the ability to quantify such risk enables risk managers to continuously re-evaluate the adequacy of their risk capitalization against extreme losses. Value-at-risk (VaR) is a financial risk measure made popular by J. P. Morgan, which is simply a chosen quantile of the loss distribution. More precisely, Holton (2014) defined VaR to be the threshold value of return or capital that a company's or portfolio's losses would not exceed, given a specific probability level and time horizon. While VaR is widely used in practice, it is not a coherent risk measure. Artzner et al. (1999) defined a coherent measure as one that is monotonic, translation invariant, homogeneous, and sub-additive. VaR satisfies the first three properties, but is not always sub-additive. Hence, an alternative risk measure called expected shortfall (ES), sometimes also referred to as conditional VaR, is introduced. It is defined as the expected loss of a portfolio or asset return, conditional on a loss exceeding the corresponding VaR. ES is a coherent risk measure and is more sensitive to the tail loss distribution. While VaR is still commonly used by practitioners, ES has also gained more popularity in recent years due to the above finding.

Parametric approaches for estimating VaR and ES typically depend on an accurate underlying distributional assumption of the financial data. Traditionally, such risk measures were based on the assumption that asset returns are independently and identically distributed (IID) according to the Gaussian distribution. It is well-known in practice that such assumptions are often inappropriate. The resulting model also needs to capture various stylized facts of the financial time series, such as volatility clustering, heavy tails, asymmetry, and long-range dependency. Moreover, such approaches can be hindered by the lack of extreme observations (which are, trivially, rare). Another common approach for modeling asset returns is using the generalized autoregressive conditional heteroskedastic (GARCH) model. While this model caters for the continually changing variance, assumptions also need to be made regarding its innovations, or residuals. Typically, the innovations are assumed to be IID and Gaussian. However, these assumptions are not always suitable (for example, see Byström 2004; and, Huang et al. 2016). For a comprehensive review on VaR estimation and backtesting, see Abad et al. (2014).

Bootstrapping procedures provide alternatives to overcome the challenges mentioned above. In general, they do not rely on distributional assumptions of the data and asymptotically replicate the empirical density for resampled data. Moreover, block bootstrapping and stationary bootstrapping can also cater for dependencies and stationarity in the data (Sunesson 2011). In this paper, we examine the daily rolling one-day-ahead forecasts of VaR and ES using bootstrapping procedures. More precisely, we apply ordinary, block, and stationary bootstrapping to both the original financial series and its GARCH innovations (for two different underlying noise assumptions). At the same time, we also assess the effects of using different window sizes of historical data for these procedures. The two datasets used are daily returns of the S&P500 from New York Stock Exchange (NYSE) and the ALSI from Johannesburg Stock Exchange (JSE).

The contribution to the literature is threefold. Firstly, the analyses should point out whether dependent bootstrapping can improve performances, in terms of VaR and ES estimation. In particular, we also examine whether the various bootstrap procedures are more effective with or without GARCH implementations. These are extensions to previous work on historical simulation and filtered historical simulation, where only ordinary bootstrapping is used (see, for example, Zenti and Pallotta 2000; Brandolini et al. 2001; Barone-Adesi et al. 2002; Lin et al. 2006; Hartz et al. 2006; Brandolini and Colucci 2012; Cecarone and Colucci 2016). Secondly, we test the approaches against different window sizes of historical data and aim to conclude on the effects of window size selection in bootstrap VaR and ES estimation. Lastly, we compare the performances of these methods across two vastly different markets.

Methodology

The procedures implemented in this paper are generalized forms of historical simulation and filtered historical simulation (Pritsker 2006). In other words, we infer on a forecast of VaR or ES depending on information provided by past data. Hence, one needs to decide on which and how much historical data to include. In our analyses, we use the moving window technique (Richardson et al. 1997) for the re-calculation of VaR and ES on a daily basis. The window sizes chosen are 250, 500, and 1000 days, equivalent to approximately 1 year, 2 years, and 4 years of daily observations. For example, using a window size of 250, we utilize the observations from day 1 to day 250 to estimate VaR for day 251 and, accordingly, use observations from day 2 to day 251 to estimate VaR for day 252, etc. This method is implemented for the following bootstrapping procedures.

Ordinary bootstrapping

Bootstrapping is a resampling procedure first introduced by Efron (1979). Typically, this involves repeated resampling, with replacement, from a given dataset, say $X_1, X_2, \ldots, X_n$, to produce numerous samples of size n. Then, a statistic of interest is estimated from each sample, creating a series of estimates which can be used to approximate the sampling distribution of the statistic. One drawback of the ordinary bootstrapping procedure is that it assumes IID data while most real-world data are rarely so rigid. Hence, to cater for dependent data, generalizations of the ordinary bootstrapping have been proposed.

Block bootstrapping

Block bootstrapping, or moving block bootstrapping, is suited for resampling from a dependent dataset. The method was introduced by Künsch (1989) and can be used to replicate the correlation in the dataset by resampling blocks of data. It involves dividing a dataset of n observations into $n - b + 1$ overlapping blocks of fixed length b. For example, the block beginning with the ith observation is of the form

$$B_i^b = \{X_i, X_{i+1}, \ldots, X_{i+b-1}\}.$$

Subsequently, n/b blocks are randomly selected, with replacement, from the $n - b + 1$ overlapping blocks and concatenated together. The concatenated observations are then used to estimate the statistic of interest. Again, the process is repeated several times to produce a series of estimates. In addition, we shall follow Hall et al. (1995) in using the optimal block length $b = n^{1/4}$ since we only focus on the negative side of the returns distribution for our estimation of VaR and ES.

Stationary bootstrapping

A possible problem in using block bootstrapping is that the procedure eliminates stationarity in the dataset. In other words, if the original dataset was stationary, the block bootstrapped samples may not be. This can be overcome by allowing the block length to be random (Politis and Romano 1994). The procedure is as follows. A value $p \in (0, 1]$ is predefined, which is optimally[1] taken to be $c^{-1}n^{-1/3}$. When deciding whether an observation should be included in a block, a number u is randomly drawn from the $UNIF(0, 1)$ distribution. If u is less than $1 - p$, we include the observation into the current block. If u is greater than $1 - p$, then a new block construction is started. This algorithm is continued until all the observations have been designated into blocks. Hence, the block length is a random variable following a geometric distribution with parameter p. It also consequently renders the number of blocks as a random variable. This modification preserves the stationarity property of any data series.

GARCH

In addition to applying bootstrapping directly on financial returns datasets, we also explore an alternative approach in bootstrapping the innovations, after the datasets have been fitted with a GARCH(1,1) model. GARCH(1,1) was first introduced by Engle (1982) and it has the general assumptions in the form of

$$R_t = \mu_t + \sigma_t Z_t,$$

where R_t is the return at time t, μ_t is the mean function of the overall series, σ_t^2 is the conditional variance, and Z_t is the residual or innovation at time t. It is further defined that

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

where $\varepsilon_t = \sigma_t Z_t$ and α_0, α_1, and β_1 are non-negative constants. It is typically assumed that the innovation series is IID and follows the Gaussian distribution.[2]

---

[1] The value for c is set to be 3.15 in our analysis, which is obtained by Monte Carlo simulation using a Gaussian AR(1) process..

[2] The process can also be estimated using a quasi-maximum likelihood estimation procedure (McNeil and Frey 2000). In our case, this yielded almost identical results as the Gaussian assumption. Hence, we only present the Gaussian results.

Consequently, the normal quantiles can be back transformed to estimate the VaR or ES of the original series. The innovations can also be extracted under the assumption that they follow a Student's t distribution and the respective quantiles utilized as above. However, we shall implement the bootstrapping procedures on the innovations instead and contrast our results against these approaches.

Validation Tests for VaR and ES

Mathematically, VaR can be defined as

$$VaR_\alpha(L) = \inf\{m \in \mathbb{R} : P(L \leq m) \geq \alpha\}$$

for some probability level α and loss function L. Additionally, ES is defined as

$$ES = \mathbb{E}\{L|L > \text{VaR}_\alpha(L)\}.$$

To compare the performances of our VaR and ES estimates, we implement various backtesting procedures for the two risk measures. In particular, we utilize the Kupiec likelihood ratio test (Kupiec 1995) and the VaR duration test (Christoffersen and Pelletier 2004) for the VaR estimates. The backtesting procedure in McNeil and Frey (2000) is implemented for ES estimates. Generally, a higher p value for each test implies a better estimate. However, care must be taken when interpreting the results as a whole. The Kupiec test checks for the correct number of exceedances (i.e., unconditional coverage) and the VaR duration test checks to see whether the VaR violations are IID. The second test is strongly sensitive to the number of violations, which is more problematic for very extreme levels of VaR. Moreover, the ES test is by definition strongly dependent on the corresponding VaR estimates. As a result, we take the Kupiec test as our primary first check for a suitable model (identifying the highest p value), and utilize the other two tests as secondary checks for desirable model properties (suitably high p values).

Empirical results

In this paper, we examine the daily returns obtained from two indices, namely S&P500 and ALSI, both extracted from McGregor BFA, and are dated from 03/01/2000 to 11/06/2015 (giving a total of 4029 data points[3]). For each dataset, we calculate the daily returns as

$$R_t = \ln\left(\frac{C_t}{C_{t-1}}\right),$$

where C_t is the closing stock price on day t. Table 1 presents the excess kurtosis and skewness for both return series. Both series are leptokurtic and asymmetric, as commonly observed in financial data. Hence, they deviate from the Gaussian

---

[3] This excludes weekends. However, values for public holidays are taken as replicates of last available prices by McGregor BFA. This is also in accordance with Cecarone and Colucci (2016)

Table 1 Excess kurtosis and skewness for S&P500 and ALSI

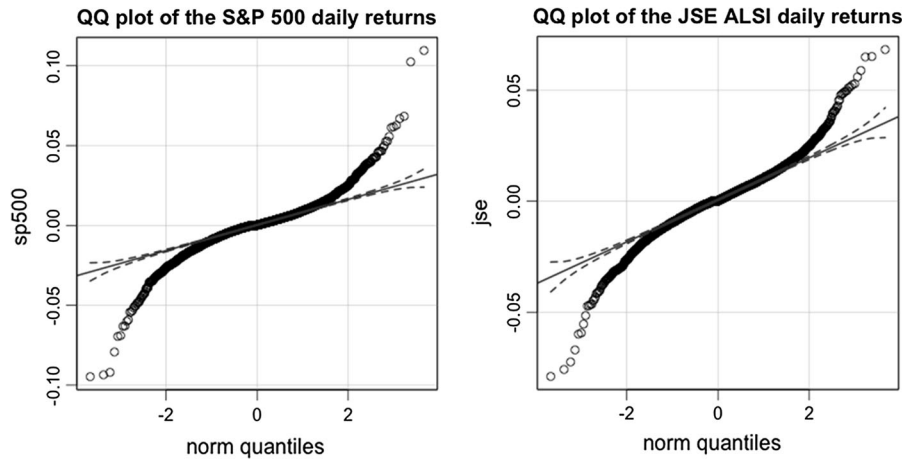| Index	Excess kurtosis	Skewness
S&P500	8.5335	−0.1869
ALSI	3.9215	−0.1781

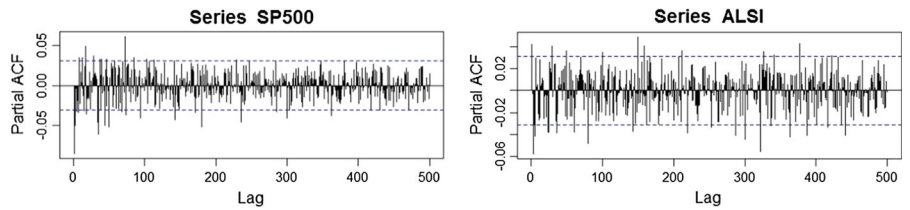**Fig. 1** Normal QQ plots of S&P500 (*left*) and ALSI (*right*)

Fig. 2 PACF plots for S&P500 (*left*) and ALSI (*right*)

distribution. This is also highlighted by the QQ plots in Fig. 1, which indicate both data series have tails significantly departing from the Gaussian distribution.

The partial autocorrelation function (PACF) plots in Fig. 2 show a number of lags being marginally significant for both datasets, while the time series plots of the returns (shown in Fig. 3) also exhibit volatility clustering. These are evidence that, at least, some weak dependencies exist in the datasets. Hence, it justifies the use of dependent bootstrapping.

After implementing a GARCH filter (for either case of Gaussian and Student t noises) for both data series, we extract the corresponding residuals. The skewness and excess kurtosis of these residual series are presented in Table 2. It is evidenced that there are significant excess kurtosis (albeit lesser compared to the original data series) and S&P500 innovations possess higher kurtosis than ALSI. The skewness is also persistent in the residuals.
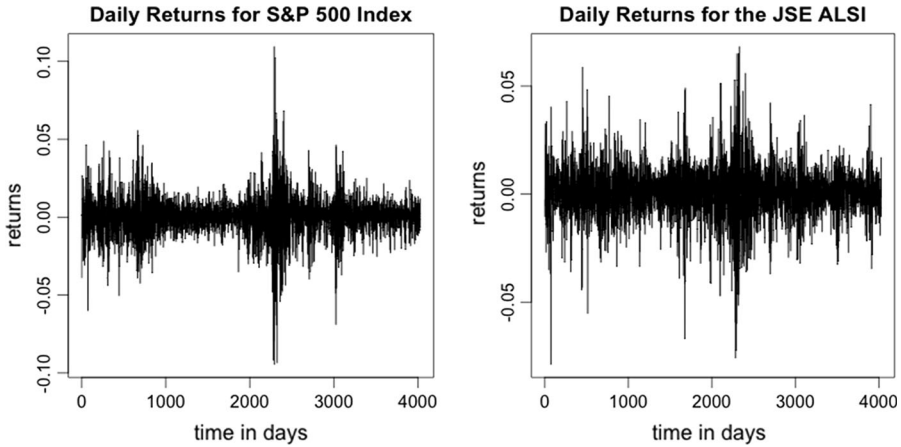
Fig. 3 Time series plots for S&P500 (*left*) and ALSI (*right*)

Table 2 Excess kurtosis and skewness for GARCH innovations of S&P500 and ALSI

Noise assumptions	S&P500		ALSI	
	Normal	Student t	Normal	Student t
Skewness	−0.3904	−0.4076	−0.2161	−0.2207
Excess kurtosis	1.3775	1.4741	0.7811	0.7836

Figures 4 and 5 present the PACF plots for the GARCH residuals of S&P500 and ALSI returns, respectively. All plots indicate very slow (if any) decay of autocorrelation, indicating long-range dependencies in the GARCH innovations. This is evidence again that dependent bootstrapping may be required, instead of the usual IID bootstrap.

We now estimate VaR on a daily basis by utilizing the various bootstrapping methods (without GARCH filters) and using a moving window of 250, 500, or 1000 trading days. VaR is calculated at 0.1, 1, 5, and 10% levels for both S&P500 and ALSI returns. The resulting series of VaR estimates are presented by Figs. 6, 7, and 8 (using original return series) in Appendix 1. The results are also benchmarked against corresponding VaR estimates obtained from the Gaussian distribution (Fig. 9). The formal backtesting results of VaR and ES estimates are displayed in Tables 5, 6, 7, and 8 in Appendix 1.

We observe that the bootstrap methods do perform better than the Gaussian distribution, in terms of VaR estimation. This is apparent as the Gaussian VaR estimates often underestimate the magnitude of risk, generating excess amount of VaR violations. We also see that the selection of the moving window size plays an essential role on how quickly the VaR estimates can return to normal levels after a shock to the market. In particular, a smaller window size tends to accelerate the return of VaR to normal levels, while a large window size can cause a lag of large VaR estimates. These large VaR estimates can drastically cause an overestimate of the actual magnitude of risk. Practically, this corresponds to a financial institute setting
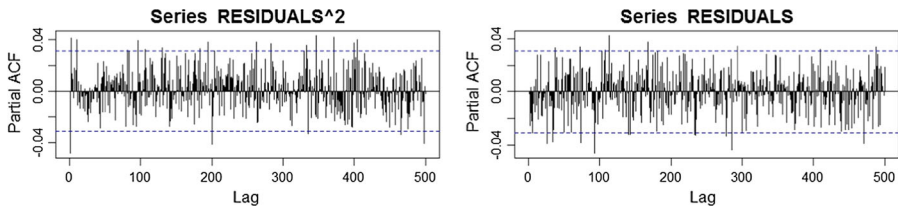
Fig. 4 PACF plots for S&P500 innovations (squared) under Gaussian noise (*left*) and innovations under Student *t* noise (*right*)
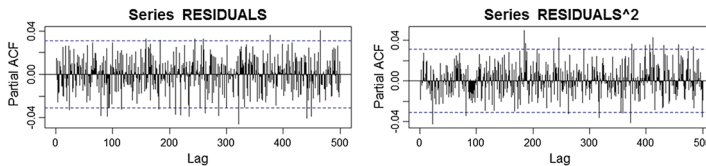

Fig. 5 PACF plots for ALSI innovations under Gaussian noise (*left*) and innovations (*squared*) under Student *t* noise (*right*)

aside a larger-than-required risk capital even after a crisis has passed, which could have been invested elsewhere. However, a small window allows for very little information for extreme tails of the data. Moreover, as often depicted in the literature (also evidenced from the figures in Appendix 1), direct bootstrapping on the returns is in general slow in reacting to changes in the markets (even for the smallest window size 250), as the model does not react quickly enough to the changing volatility. Hence, we next look at bootstrapping in a filtered historical simulation procedure.

The return series are now filtered by GARCH processes, with either Gaussian or Student *t* noise. The innovations are extracted and bootstrapped by the three methods as before. Again, the three different window sizes are used and the risk measures are estimated at the 4 different VaR levels. The backtesting results are given in Tables 3 and 4, while the plots of VaR estimates from the various models are given in Appendix 2.

It is clearly evidenced that the added GARCH effect is better at capturing the changing variance and produced significantly more adequate models, relative to the earlier results in Appendix 1. In Tables 3 and 4, we have also highlighted (in bold) situations where dependent bootstrapping has produced equal or better results (for the Kupiec test) than their IID counterpart. Although the performances between the different bootstrapping methods are quite similar, when restricted to a common window size and noise assumption, there is evidence of possible improvements from the IID approach.

It may be interesting to note as well that the window size selection do have an effect on the performance of the different models for the two indices, at various VaR levels. For 10% VaR, window size of 500 seems to be the best choice. However, window size 1000 is likely to be more adequate for more extreme VaR levels. In fact, for ALSI returns, models with window size 1000 and Student *t* noise are recorded as the only cases to produce satisfactory results for 0.1% VaR.

Table 3 Backtesting of daily one-day-ahead VaR and ES estimates for S&P500 with GARCH filters

GARCH	Win. size	VaR level Cond. quantile	10% (Exp. Vio. = 302)				5% (Exp. Vio. = 151)				1% (Exp. Vio. = 30)				0.1% (Exp. Vio. = 3)			
			Act. Vio.	Kupiec test	Dur. test	ES test	Act. Vio.	Kupiec test	Dur. test	ES test	Act. Vio.	Kupiec test	Dur. test	ES test	Act. Vio.	Kupiec test	Dur. test	ES test
Normal	250	Gaussian	335	0.055	0.041	<0.001	208	<0.001	0.661	<0.001	85	<0.001	0.546	0.001	26	<0.001	0.083	0.078
		BS	324	0.204	0.029	0.423	170	0.128	0.801	0.211	59	<0.001	0.413	0.201	26	<0.001	0.201	0.004
		BBS	325	0.183	0.076	0.471	170	**0.128**	0.858	0.211	58	<0.001	0.449	0.140	27	<0.001	0.079	0.003
		SBS	326	0.165	0.037	0.445	172	0.092	0.877	0.295	59	<0.001	0.413	0.208	25	<0.001	0.138	0.005
	500	Gaussian	321	0.274	0.067	<0.001	191	0.001	0.846	<0.001	77	<0.001	0.279	0.018	18	<0.001	0.202	0.077
		BS	304	0.942	0.089	0.607	169	0.149	0.609	0.718	44	0.019	0.102	0.704	9	0.006	0.936	0.034
		BBS	305	0.894	0.102	0.609	169	**0.149**	0.612	0.654	44	**0.019**	0.102	0.669	10	0.002	0.188	0.022
		SBS	302	**0.961**	0.073	0.589	168	**0.173**	0.549	0.637	44	**0.019**	0.102	0.699	9	**0.006**	0.936	0.053
	1000	Gaussian	311	0.621	0.115	<0.001	180	0.020	0.418	<0.001	72	<0.001	0.275	0.004	18	<0.001	0.143	0.081
		BS	293	0.551	0.224	0.537	155	0.765	0.977	0.491	40	0.091	0.707	0.607	4	0.594	0.119	0.062
		BBS	293	**0.551**	0.261	0.542	154	**0.829**	0.973	0.470	41	0.063	0.449	0.655	4	**0.594**	0.119	0.020
		SBS	294	**0.592**	0.233	0.553	155	**0.765**	0.977	0.481	40	**0.091**	0.707	0.601	4	**0.594**	0.119	0.071
Student t	250	Student t	379	<0.001	0.003	0.545	203	<0.001	0.948	0.690	46	0.008	0.568	0.966	1	0.175	1.000	NA
		BS	319	0.330	0.035	0.596	166	0.230	0.990	0.382	53	<0.001	0.723	0.140	25	<0.001	0.403	0.005
		BBS	317	**0.393**	0.093	0.605	168	0.173	0.812	0.461	52	<0.001	0.961	0.078	26	<0.001	0.188	0.004
		SBS	319	**0.330**	0.036	0.592	167	0.200	0.935	0.407	53	<0.001	0.723	0.150	25	<0.001	0.403	0.006
	500	Student t	358	0.001	0.012	0.566	199	<0.001	0.391	0.844	39	0.127	0.924	0.992	1	0.175	1.000	NA
		BS	302	0.961	0.097	0.838	163	0.339	0.511	0.857	41	0.063	0.265	0.876	8	0.018	0.158	0.022
		BBS	302	**0.961**	0.181	0.862	164	0.300	0.521	0.863	41	**0.063**	0.265	0.859	11	<0.001	0.093	0.047
		SBS	303	**0.990**	0.075	0.863	162	**0.382**	0.454	0.835	42	0.043	0.214	0.901	8	**0.018**	0.158	0.023

Table 3 continued

GARCH	Win. size	VaR level Cond. quantile	10% (Exp. Vio. = 302)				5% (Exp. Vio. = 151)				1% (Exp. Vio. = 30)				0.1% (Exp. Vio. = 3)			
			Act. Vio.	Kupiec test	Dur. test	ES test	Act. Vio.	Kupiec test	Dur. test	ES test	Act. Vio.	Kupiec test	Dur. test	ES test	Act. Vio.	Kupiec test	Dur. test	ES test
1000		Student t	365	<0.001	0.035	0.522	204	<0.001	0.412	0.850	39	0.127	0.298	0.982	1	0.175	1.000	NA
		BS	282	0.203	0.197	0.476	155	0.765	0.456	0.707	39	0.127	0.491	0.781	5	0.300	0.119	0.095
		BBS	282	**0.203**	0.196	0.484	154	**0.829**	0.536	0.679	38	**0.175**	0.333	0.689	6	0.132	0.076	0.075
		SBS	282	**0.203**	0.197	0.472	155	**0.765**	0.551	0.680	39	**0.127**	0.491	0.782	5	**0.300**	0.119	0.103

BS standard bootstrap, *BBS* block bootstrap, *SBS* stationary bootstrap, *Act. Vio.* actual number of VaR violations, *Exp. Vio.* expected number of VaR violations, *Dur Test* VaR duration test

Table 4 Backtesting of daily one-day-ahead VaR and ES estimates for ALSI with GARCH filters

GARCH	Win. Size	VaR level Cond. quantile	10% (Exp. Vio. = 302)				5% (Exp. Vio. = 151)				1% (Exp. Vio. = 30)				0.1% (Exp. Vio. = 3)			
			Act. Vio.	Kupiec test	Dur. test	ES Test	Act. Vio.	Kupiec test	Dur. test	ES Test	Act. Vio.	Kupiec test	Dur. test	ES test	Act. Vio.	Kupiec test	Dur. test	ES test
Normal	250	Gaussian	321	0.274	0.213	<0.001	196	<0.001	0.267	<0.001	65	<0.001	0.754	0.001	22	<0.001	0.287	0.276
		BS	311	0.621	0.231	0.187	184	0.008	0.442	0.304	44	0.019	0.413	0.125	21	<0.001	0.207	<0.001
		BBS	312	0.579	0.272	0.205	184	**0.008**	0.399	0.309	44	**0.019**	0.599	0.116	21	<0.001	0.207	<0.001
		SBS	308	**0.753**	0.208	0.145	184	**0.008**	0.335	0.289	45	0.012	0.376	0.173	21	<0.001	0.207	<0.001
	500	Gaussian	313	0.539	0.399	<0.001	195	<0.001	0.414	0.007	59	<0.001	0.706	0.007	17	<0.001	0.040	0.125
		BS	304	0.942	0.295	0.433	166	0.230	0.234	0.594	35	0.400	0.347	0.466	11	<0.001	0.266	0.005
		BBS	302	**0.961**	0.330	0.404	166	**0.230**	0.234	0.592	34	**0.505**	0.195	0.355	11	<0.001	0.266	0.005
		SBS	302	**0.961**	0.295	0.379	166	**0.230**	0.234	0.583	35	**0.400**	0.221	0.386	10	**0.002**	0.434	0.003
	1000	Gaussian	310	0.664	0.380	0.003	182	0.013	0.249	0.017	52	<0.001	0.531	0.006	19	<0.001	0.184	0.301
		BS	299	0.818	0.436	0.599	144	0.534	0.889	0.436	33	0.624	0.302	0.553	7	0.051	0.267	0.372
		BBS	300	**0.865**	0.451	0.612	144	**0.534**	0.889	0.466	33	**0.624**	0.302	0.562	7	**0.051**	0.267	0.345
		SBS	300	**0.865**	0.451	0.611	146	**0.651**	0.720	0.492	33	**0.624**	0.302	0.568	7	**0.051**	0.267	0.382
Student t	250	Student t	391	<0.001	0.036	0.658	216	<0.001	0.423	0.910	40	0.091	0.807	0.999	1	0.175	1.000	NA
		BS	313	0.539	0.201	0.215	182	0.013	0.363	0.342	43	0.029	0.417	0.183	17	<0.001	0.109	<0.001
		BBS	313	**0.539**	0.335	0.219	183	0.011	0.333	0.395	44	0.019	0.735	0.209	17	<0.001	0.109	<0.001
		SBS	312	**0.579**	0.196	0.181	182	**0.013**	0.363	0.321	43	**0.029**	0.417	0.225	18	<0.001	0.053	<0.001
	500	Student t	378	<0.001	0.066	0.571	221	<0.001	0.835	0.982	35	0.400	0.426	0.994	1	0.175	1.000	NA
		BS	301	0.913	0.156	0.388	167	0.200	0.366	0.661	35	0.400	0.333	0.442	12	<0.001	0.307	0.007
		BBS	299	0.818	0.224	0.361	166	**0.230**	0.389	0.645	35	**0.400**	0.333	0.440	12	<0.001	0.307	0.008
		SBS	302	**0.961**	0.178	0.401	167	**0.200**	0.366	0.653	35	**0.400**	0.333	0.443	12	<0.001	0.307	0.011

Table 4 continued

GARCH Win. Size	VaR level Cond. quantile	10% (Exp. Vio. = 302)				5% (Exp. Vio. = 151)				1% (Exp. Vio. = 30)				0.1% (Exp. Vio. = 3)			
		Act. Vio.	Kupiec test	Dur. test	ES Test	Act. Vio.	Kupiec test	Dur. test	ES Test	Act. Vio.	Kupiec test	Dur. test	ES test	Act. Vio.	Kupiec test	Dur. test	ES test
1000	Student t	360	0.001	0.206	0.603	206	<0.001	0.459	0.987	34	0.505	0.079	0.999	0	NA	1.000	NA
	BS	300	0.865	0.425	0.601	144	0.534	0.804	0.451	32	0.756	0.300	0.506	6	0.132	0.467	0.231
	BBS	298	0.771	0.472	0.547	145	**0.591**	0.688	0.482	32	**0.756**	0.300	0.515	6	**0.132**	0.467	0.235
	SBS	300	**0.865**	0.425	0.601	144	**0.534**	0.804	0.448	32	**0.756**	0.300	0.518	6	**0.132**	0.467	0.248

BS standard bootstrap, *BBS* block bootstrap, *SBS* stationary bootstrap, *Act. Vio.* actual number of VaR violations, *Exp. Vio.* expected number of VaR violations, *Dur Test* VaR duration test

The comparisons across the two indices are also peculiar. In general, the models seem to perform better for ALSI at 10 and 1% VaR levels, but are less effective (relatively) for 5 and 0.1% VaR levels. This may be attributed to the vastly different structures of the two different indices and how they react to sudden changes to the market. It may also be interesting to note that the difference between better performing models and the less adequate models is much larger for S&P500, as compared to cases for ALSI. This is largely caused by the choice of window size, making this factor a vital one for risk prediction in S&P500 returns.

In general, a GARCH filter should be applied and there seems to be some evidence that dependent bootstrapping can improve VaR and ES estimation. However, the models in general perform quite similarly, when restricted to similar conditions. At the same time, window size selection does have an impact on model performances, and more significantly so for S&P500 returns. The optimal model for the two indices varies at different VaR levels. Our recommendation, thus far, is to employ window size 1000 for ALSI returns (possibly for all VaR levels as window size 500 is only marginally better at 10% VaR level) and window size 500 for S&P500 at 10% VaR. As for 5, 1, and 0.1% VaR levels of S&P500, the better choice is window size 1000. We also recommend using both block bootstrapping and stationary bootstrapping to check which gives a better result. They tend to produce similar results to ordinary bootstrapping, but one of them can at times give improved estimates. It may be highly beneficial to explore a model switching procedure to optimize the performance of these models.

Conclusion

In this paper, we have examined the use of dependent bootstrapping in VaR and ES estimation for daily returns of S&P500 and ALSI. The bootstrap methods included the ordinary bootstrap, block bootstrap, and stationary bootstrap. Furthermore, these approaches were implemented with and without GARCH filters and are also contrasted against the corresponding parametric approaches. As expected, the bootstrap methods produced improved VaR results when compared to the parametric counterparts. Furthermore, GARCH models brought further improved VaR and ES estimates.

Overall, we suggest a GARCH filter to be applied for both VaR and ES estimations. Window size 1000 seems to be quite satisfactory for ALSI returns. However, for S&P500, window size 500 can significantly improve the estimates at 10% VaR level but must be reverted back to window size 1000 for more extreme VaR levels.

We suggest as further work to implement a model switching approach with varying window size selections to fully capture the changing results across the VaR levels. At the same time, this model should be able to adjust the bootstrapping method accordingly to the particular window at hand. This is also left for further investigation. Another avenue for exploration is to compare our results with other backtesting procedures. For example, one may consider the new backtesting method for ES proposed by Acerbi and Szekely (2017).
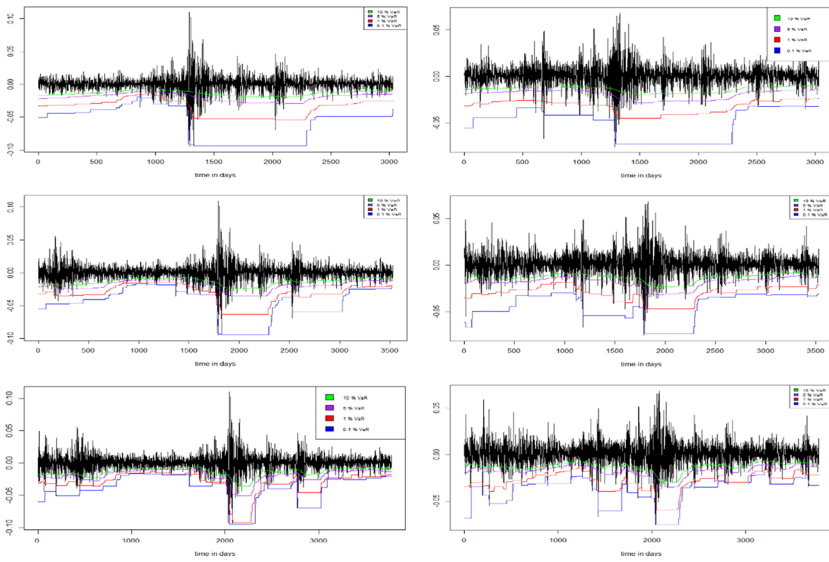
Appendix 1

Figures 6, 7, 8, and 9.

Fig. 6 Estimated VaR values for S&P500 (*left*) and ALSI (*right*) using *ordinary bootstrapping*, with moving window sizes 1000, 500, and 250, respectively (*top* to *bottom*)
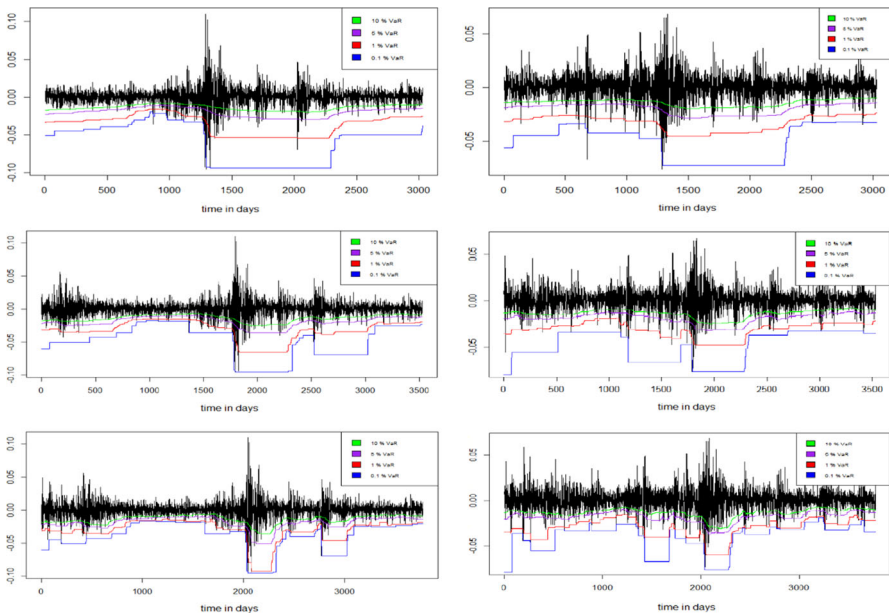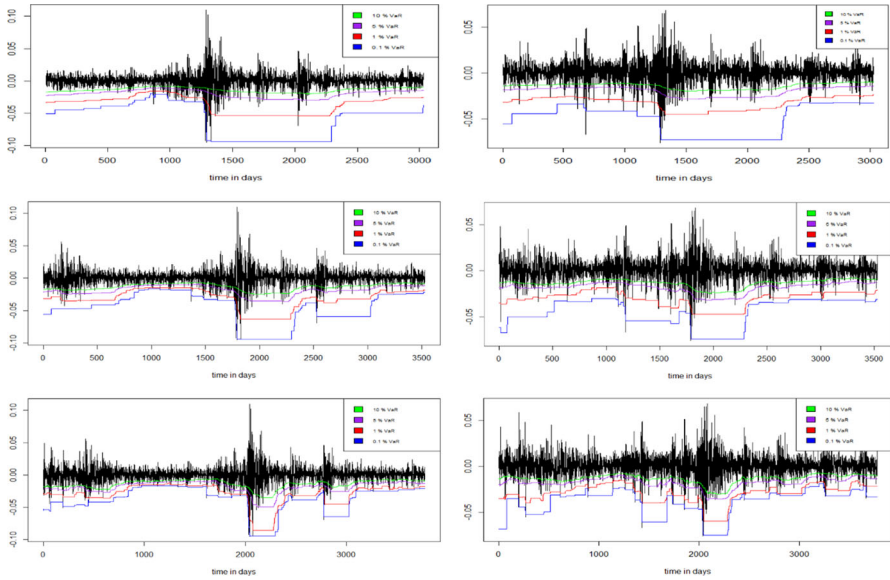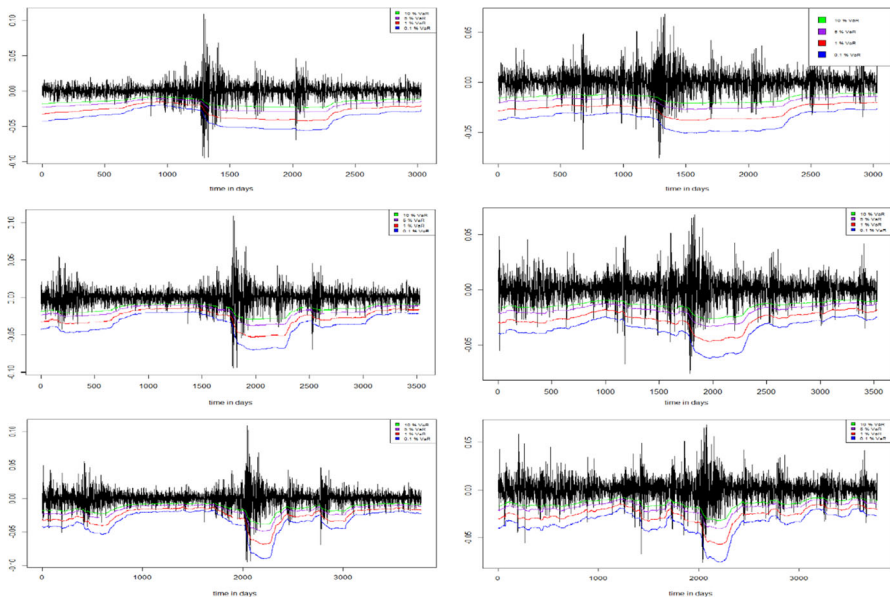

Fig. 7 Estimated VaR values for S&P500 (*left*) and ALSI (*right*) using *block bootstrapping*, with moving window sizes 1000, 500, and 250, respectively (*top* to *bottom*)

Fig. 8 Estimated VaR values for S&P500 (*left*) and ALSI (*right*) using *stationary bootstrapping*, with moving window sizes 1000, 500, and 250, respectively (*top* to *bottom*)

Fig. 9 Estimated VaR values for S&P500 (*left*) and ALSI (*right*) using *Gaussian distribution* with moving window sizes 1000, 500, and 250, respectively (*top* to *bottom*)

Tables 5, 6, 7, and 8.

Table 5 Backtesting of daily one-day-ahead VaR estimates for S&P500 without GARCH filter

Win. size	VaR level (%)	Exp. Vio.	Gaussian		Ordinary bootstrapping		Block bootstrapping		Stationary bootstrapping	
			Act. Vio.	Kupiec	Act. Vio.	Kupiec	Act. Vio.	Kupiec	Act. Vio.	Kupiec
1000	10	302	205	<0.001	248	0.001	248	0.001	248	0.001
	5	151	140	0.336	145	0.591	143	0.480	145	0.591
	1	30	65	<0.001	43	0.029	43	0.029	43	0.029
	0.1	3	40	<0.001	13	<0.001	13	<0.001	13	<0.001
500	10	352	302	0.004	331	0.217	329	0.177	331	0.217
	5	176	188	0.375	181	0.723	177	0.963	181	0.723
	1	35	79	<0.001	53	0.005	52	0.008	53	0.005
	0.1	3	34	<0.001	12	<0.001	8	0.041	12	<0.001
250	10	377	361	0.359	384	0.737	382	0.8200	384	0.737
	5	188	219	0.028	200	0.412	199	0.4550	200	0.412
	1	37	90	<0.001	59	0.001	54	0.0130	59	0.001
	0.1	3	33	<0.001	16	<0.001	16	<0.001	16	<0.001

Win. Size size of the moving windows, *Exp. Vio.* expected number of VaR violations, *Act. Vio.* actual number of VaR violations, *Kupiec p* value of the Kupiec LR test

Table 6 Backtesting of daily one-day-ahead VaR estimates for ALSI without GARCH filter

Win. size	VaR level (%)	Exp. Vio.	Gaussian		Ordinary bootstrapping		Block bootstrapping		Stationary bootstrapping	
			Act. Vio.	Kupiec	Act. Vio.	Kupiec	Act. Vio.	Kupiec	Act. Vio.	Kupiec
1000	10	302	237	<0.001	281	0.182	280	0.162	281	0.182
	5	151	146	0.651	144	0.534	144	0.534	144	0.534
	1	30	60	<0.001	35	0.400	34	0.505	35	0.400
	0.1	3	27	<0.001	10	0.002	10	0.002	10	0.002
500	10	352	306	0.007	367	0.428	366	0.461	367	0.428
	5	176	170	0.619	181	0.723	180	0.782	181	0.723
	1	35	69	<0.001	41	0.345	41	0.345	41	0.345
	0.1	3	25	<0.001	11	0.002	7	0.103	11	0.002
250	10	377	336	0.021	396	0.327	391	0.476	396	0.327
	5	188	189	0.994	210	0.122	201	0.371	210	0.122
	1	37	76	<0.001	55	0.008	49	0.079	55	0.008
	0.1	3	36	<0.001	27	<0.001	21	<0.001	27	<0.001

Win. Size size of the moving windows, *Exp. Vio.* expected number of VaR violations, *Act. Vio.* actual number of VaR violations, *Kupiec p* value of the Kupiec LR test

Table 7 Backtesting of daily one-day-ahead ES estimates for S&P500 without GARCH filter

Win. size	VaR level (%)	Gaussian	Ordinary bootstrapping	Block bootstrapping	Stationary bootstrapping
1000	10	0.053	0.013	0.016	0.018
	5	0.031	0.038	0.041	0.051
	1	0.018	0.022	0.023	0.025
	0.1	0.114	0.521	0.546	0.537
500	10	0.023	0.033	0.049	0.050
	5	0.041	0.044	0.056	0.053
	1	0.043	0.154	0.161	0.155
	0.1	0.015	0.233	0.203	0.217
250	10	0.018	0.041	0.045	0.046
	5	0.058	0.024	0.037	0.038
	1	0.410	0.289	0.135	0.134
	0.1	0.004	0.004	0.018	0.002

Table 8 Backtesting of daily one-day-ahead ES estimates for ALSI without GARCH filter

| Win. Size	VaR level (%)	Gaussian	Ordinary bootstrapping	Block bootstrapping	Stationary bootstrapping
1000	10	0.125	0.151	0.161	0.167
	5	0.260	0.157	0.172	0.166
	1	0.149	0.180	0.162	0.192
	0.1	0.229	0.387	0.417	0.417
500	10	0.405	0.528	0.576	0.605
	5	0.293	0.347	0.369	0.403
	1	0.214	0.285	0.315	0.316
	0.1	0.078	0.022	0.031	0.028
250	10	0.111	0.111	0.205	0.303
	5	0.158	0.158	0.271	0.238
	1	0.015	0.015	0.035	0.002
	0.1	<0.001	<0.001	<0.001	<0. 0 01

Appendix 2
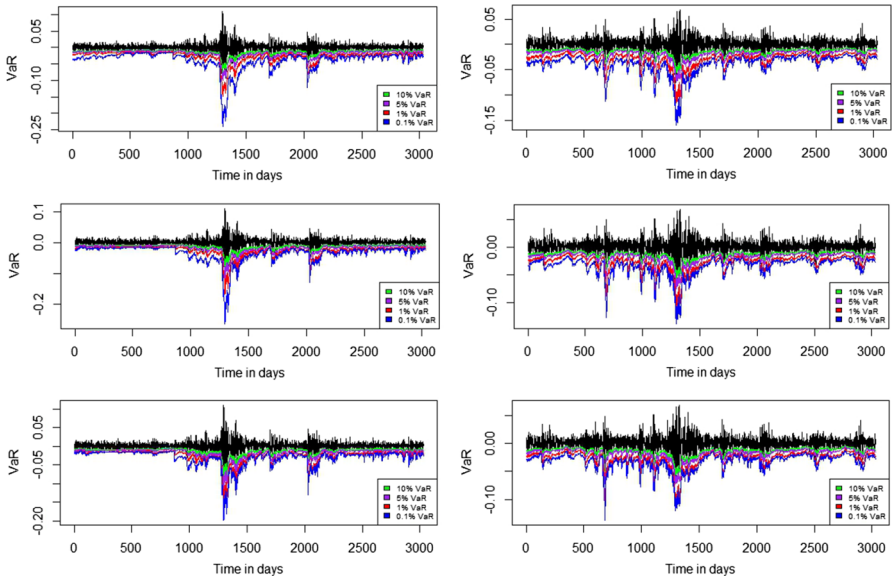
Fig. 10 Estimated VaR values for S&P500 (*left*) and ALSI (*right*) using *GARCH(1,1)-Normal and ordinary bootstrapping* with moving window sizes 1000, 500, and 250, respectively (*top* to *bottom*)
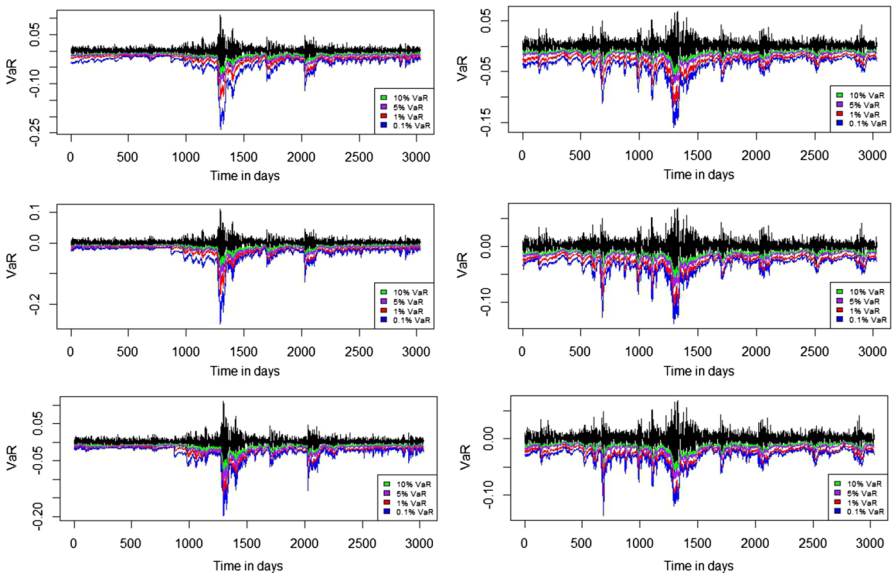

Fig. 11 Estimated VaR values for S&P500 (*left*) and ALSI (*right*) using *GARCH(1,1)-Normal and block bootstrapping* with moving window sizes 1000, 500, and 250, respectively (*top* to *bottom*)
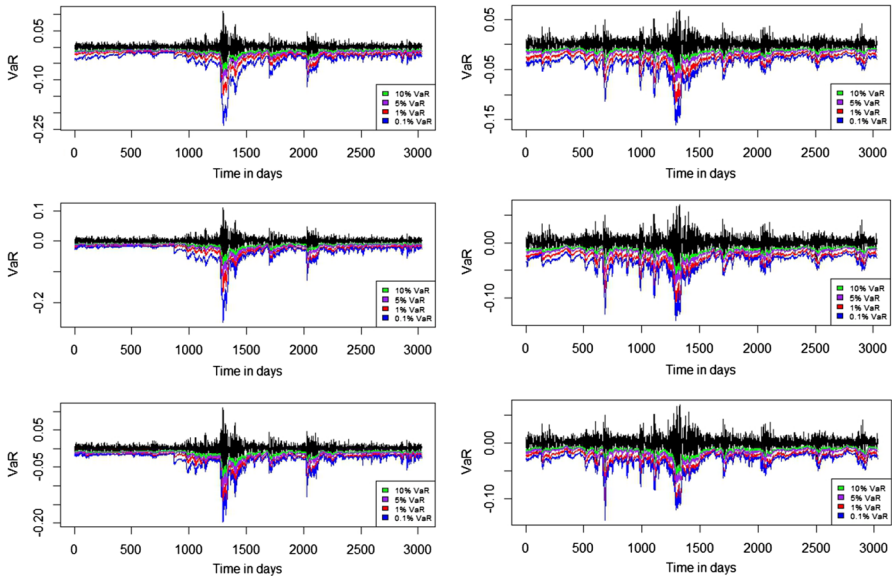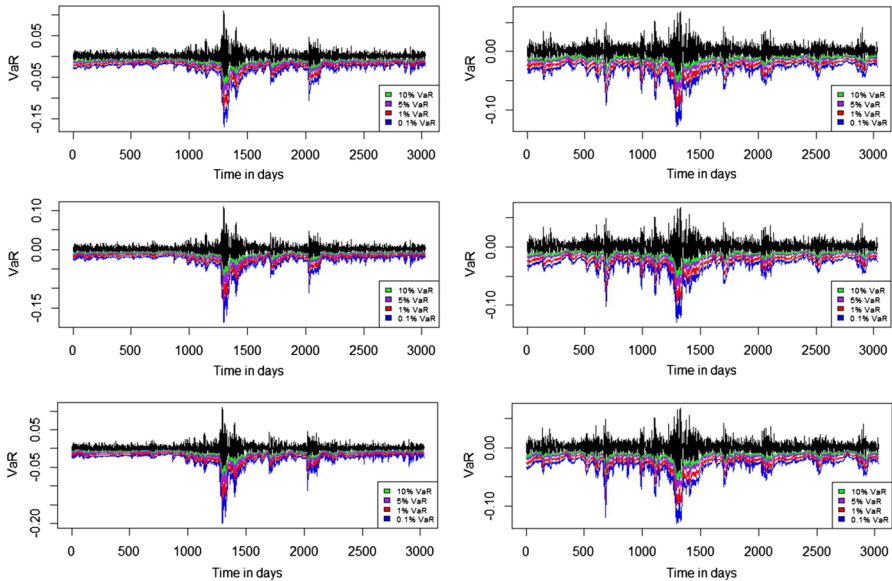
Fig. 12 Estimated VaR values for S&P500 (*left*) and ALSI (*right*) using *GARCH(1,1)-Normal and stationary bootstrapping* with moving window sizes 1000, 500, and 250, respectively (*top* to *bottom*)

Fig. 13 Estimated VaR values for S&P500 (*left*) and ALSI (*right*) using *GARCH(1,1)-Normal and Gaussian residuals* with moving window sizes 1000, 500, and 250, respectively (*top* to *bottom*)
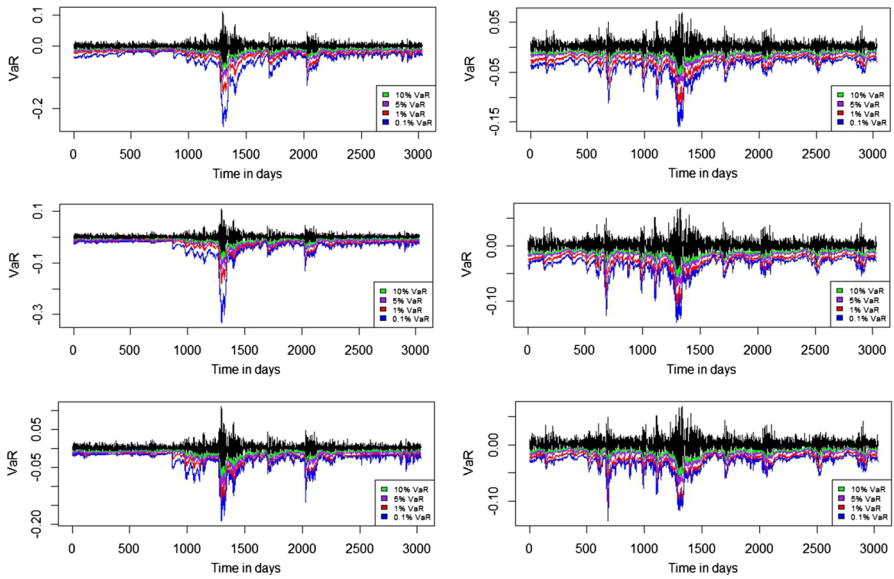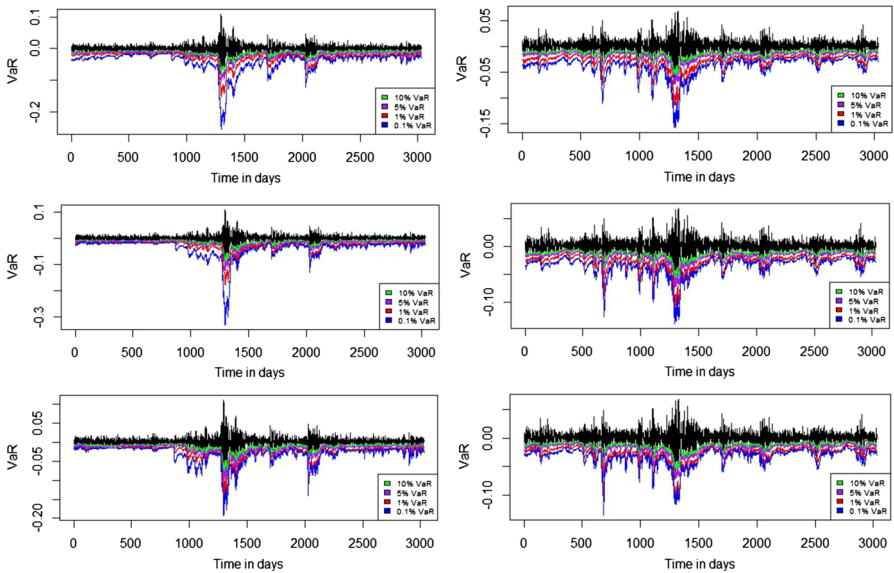
Fig. 14 Estimated VaR values for S&P500 (*left*) and ALSI (*right*) using *GARCH(1,1)-t and ordinary bootstrapping* with moving window sizes 1000, 500, and 250, respectively (*top* to *bottom*)

Fig. 15 Estimated VaR values for S&P500 (*left*) and ALSI (*right*) using *GARCH(1,1)-t and block bootstrapping* with moving window sizes 1000, 500, and 250, respectively (*top* to *bottom*)
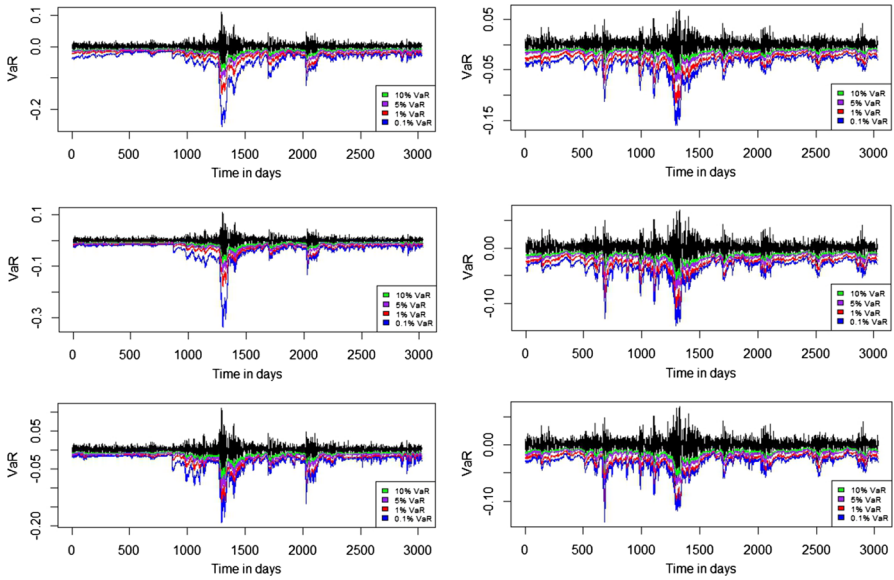
Fig. 16 Estimated VaR values for S&P500 (*left*) and ALSI (*right*) using *GARCH(1,1)-t and stationary bootstrapping* with moving window sizes 1000, 500, and 250, respectively (*top* to *bottom*)
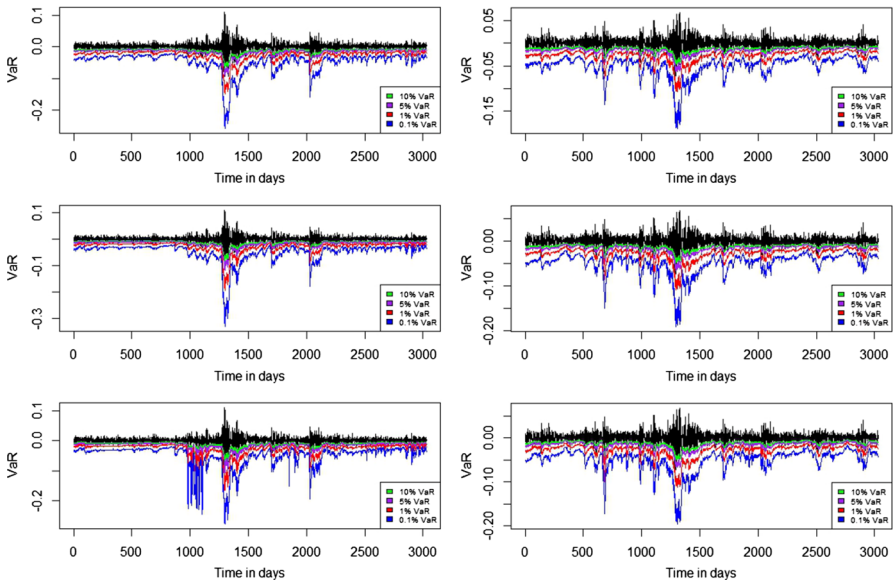

Fig. 17 Estimated VaR values for S&P500 (*left*) and ALSI (*right*) using *GARCH(1,1)-t and Student t residuals* with moving window sizes 1000, 500, and 250, respectively (*top* to *bottom*)

References

Abad, P., S. Benito, and C. López. 2014. A comprehensive review of value at risk methodologies. *The Spanish Review of Financial Economics* 12 (1): 15–32.

Acerbi, C., and B. Szekely. 2017. General properties of backtestable statistics. Available at SSRN: https://ssrn.com/abstract=2905109.

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath. 1999. Coherent measures of risk. *Mathematical Finance* 9 (3): 203–228.

Barone-Adesi, G., K. Giannopoulos, and L. Vosper. 2002. Backtesting derivative portfolios with filtered historical simulation (FHS). *European Financial Management* 8 (1): 31–58.

Brandolini, D., and S. Colucci. 2012. Backtesting value-at-risk: a comparison between filtered bootstrap and historical simulation. *Journal of Risk Model Validation* 6 (4): 3–16.

Brandolini, D., M. Pallotta, and R. Zenti. 2001. Risk management in an asset management company: a practical case. In *EFMA 2001 Lugano Meetings*. Available at SSRN: https://ssrn.com/abstract=252294.

Byström, H.N.E. 2004. Managing extreme risks in tranquil and volatile markets using conditional extreme value theory. *International Review of Financial Analysis* 13 (2): 133–152.

Cecarone, F., and S. Colucci. 2016. A quick tool to forecast value-at-risk using implied and realized volatilities. *Journal of Risk Model Validation* 10 (4): 71–101.

Christoffersen, P., and D. Pelletier. 2004. Backtesting value-at-risk: a duration-based approach. *Journal of Financial Econometrics* 2 (1): 84–108.

Efron, B. 1979. Bootstrap methods: another look at the jackknife. *Annals of Statistics* 7 (1): 1–26.

Engle, R.F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50 (4): 987–1007.

Hall, P., J.L. Horowitz, and B.-Y. Jing. 1995. On blocking rules for the bootstrap with dependent data. *Biometrika* 82 (3): 561–574.

Hartz, C., S. Mittnik, and M. Paolella. 2006. Accurate value-at-risk forecasting based on the normal-GARCH model. *Computational Statistics & Data Analysis* 51 (4): 2295–2312.

Holton, G.A. 2014. *Value-at-risk: theory and practice*, Second Edition, e-book available at http://value-at-risk.net.

Huang, C.-K., D. North, and T. Zewotir. 2016. A conditional extreme value approach to value-at-risk estimation with exchangeable innovations. In *Annual proceedings of the South African statistical association conference*, pp. 33–40.

Künsch, H.R. 1989. The Jackknife and the bootstrap for general stationary observations. *Annals of Statistics* 17 (3): 1217–1241.

Kupiec, P.H. 1995. Techniques for verifying the accuracy of risk management models. *Journal of Derivatives* 3 (2): 73–84.

Lin, S.-K., R.-H. Wang, and C.-D. Fu. 2006. Risk management for linear and non-linear assets: a bootstrap method with importance resampling to evaluate value-at-risk. *Asia-Pacific Financial Markets* 13 (3): 261–295.

McNeil, A.J., and R. Frey. 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance* 7 (3): 271–300.

Politis, D.N., and J.P. Romano. 1994. The stationary bootstrap. *Journal of the American Statistical Association* 89 (428): 1303–1313.

Pritsker, M. 2006. The hidden dangers of historical simulation. *Journal of Banking & Finance* 30 (2): 561–582.

Richardson, M.P., J. Boudoukh, and R. Whitelaw. 1997. The best of both worlds: a hybrid approach to calculating value at risk. *Risk* 11: 64–67.

Sunesson, A. 2011. Bootstrap of dependent data in finance. Master's thesis, Chalmers University of Technology and Göteborg University, Sweden.

Zenti, R., and M. Pallotta. 2000. Risk analysis for asset managers: historical simulation, the bootstrap approach and Value at Risk calculation. In *EFMA 2001 Lugano Meetings*. Available at SSRN: https://ssrn.com/abstract=251669 or http://dx.doi.org/10.2139/ssrn.251669.