ERASMUS UNIVERSITEIT ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS

Bachelor Thesis Econometrics and Operations Research: Major Quantitative Finance

The Geometric-VaR Backtesting Method on Cryptocurrencies

Anish Nathie
408606

Supervisor: prof. dr. D.J.C. van Dijk
First reader: S.H.L.C.G. Vermeulen

8 July 2018

Abstract

This paper investigates a recent value at risk (VaR) backtest that uses both the duration between two consecutive VaR violations and the value of the VaR: the geometric-VaR test from Pelletier & Wei (2016). Investigation is done by replicating two tables, reporting the size and power of the test, with Monte Carlo simulation. This test is then used to evaluate VaR estimates from five cryptocurrencies. Four methods are used to estimate VaRs with RiskMetrics providing the 'best' VaR estimates for three cryptocurrencies.

1 Introduction

Financial market participants are subject to uncertainty about future net returns, also known as risk. More specifically, the uncertainty about future returns due to changes in market conditions is called market (or systematic) risk and cannot be eliminated through diversification. Participants are therefore also subject to risk management. An important concept of financial risk management, in particular for market risk, is value at risk (VaR). This is defined as the maximum potential loss in value of a portfolio of financial instruments with a given probability over a certain horizon (Manganelli & Engle 2001). Thus, VaR summarizes market risk with a number. With a given probability, the loss in the next period exceeds VaR, making it a quantile of the conditional distribution of returns (Pelletier & Wei 2016).

Since the 1996 Market Risk Amendment to the Basel Accord, financial institutions should compute VaR daily for the purpose of calculating their capital charge. Manganelli & Engle (2001) state that non-accurate VaR estimates may lead to a sub-optimal capital allocation with consequences on the profitability or the financial stability of the institutions. A widely used method to calculate VaR is Historical Simulation (HS). It is popular because it is nonparametric and easy to implement. However, this method also has its drawbacks: it does not take time-varying characteristics of returns into account.

Comparing VaRs with realized returns is called backtesting. An example of a VaR backtesting method is the Likelihood Ratio (LR) test for correct conditional coverage from Christoffersen (1998). This test consists of two individual tests: one for correct unconditional coverage and one for VaR independence. If VaR independence is rejected, then the probability of getting a VaR violation tomorrow would depend on information known today. Another backtest is the duration-based test from Christoffersen & Pelletier (2004). This test is also a joint test: the first test tests for correct unconditional coverage and the second test tests whether it for every day holds that the expected duration until the next VaR violation is equal to $1/p$ where p is the coverage rate. A recent backtest comes from Pelletier & Wei (2016). They propose the geometric-VaR test, which is a combination of the geometric test and the VaR test. The geometric test is a joint test for correct unconditional coverage and for duration independence. If duration independence is rejected, the probability of getting a VaR violation tomorrow would be higher when the last VaR violation happened recently than when the last VaR violation was a longer time ago. The VaR test tests jointly for correct unconditional coverage and for VaR independence. The properties of the geometric-VaR test are investigated and replicated using simulation.

As an extension, the geometric-VaR test is used to evaluate VaR estimates for five cryptocurrencies (Bitcoin, Dogecoin, Litecoin, Monero, Ripple) computed from four different methods: Historical Simulation, nonlinear asymmetric GARCH, RiskMetrics and Filtered Historical Simulation. RiskMetrics can be used to get VaR estimates for Dogecoin, Litecoin and Ripple. Nonlinear asymmetric GARCH is appropriate for Monero. The geometric-VaR test reports for Bitcoin that Historical Simulation is the best method to use.

The structure of this paper is as follows: Section 2 gives more information about the geometric-VaR test and how to implement it. Furthermore, it introduces the methods used to get the VaR estimates. Section 3 describes the used data in more detail. Section 4 discusses the results from both the replication and the extension. Section 5 concludes.

2 Methodology

This paper makes use of the geometric-VaR test introduced by Pelletier & Wei (2016). Sections 2.1 and 2.2 provide a short overview about the test and how to implement it. Sections 2.3 and 2.4 provide volatility models that are used to simulate returns.

2.1 Duration-based backtesting

For any $0 < p < 1$, the one-day $\text{VaR}_{t+1}(p)$ for period $t + 1$ is the value such that the return in period $t + 1$ is expected to be exceeded with probability p. In other words, it is the $100 \times p$-th quantile of the conditional distribution of the return r_{t+1}:

$$P(r_{t+1} < -\text{VaR}_{t+1}(p)|\Omega_t) = p \qquad\qquad t = 0, ..., T - 1$$

Note that in this notation, VaR is reported as a positive number. Ω_t is the set of all information known at time t. When the realized return r_{t+1} is known, the hit sequence $\{I_{t+1}\}$ can be obtained by comparing VaR_{t+1} with r_{t+1}, where $I_{t+1} = 1$ if r_{t+1} exceeds $\text{VaR}_{t+1}(p)$ and 0 else:

$$I_{t+1} = \begin{cases} 1 & \text{if } r_{t+1} < -\text{VaR}_{t+1}(p), \\ 0 & \text{else.} \end{cases}$$

If $\text{VaR}_{t+1}(p)$ is efficient with respect to Ω_t, then the distribution of I_{t+1} should be Bernoulli with parameter p for all t.

The backtesting is based on the duration between two violations. Define t_i as the day of the i-th violation. Then, the no-hit duration D_i is defined as $t_i - t_{i-1}$. If the hit sequence I_{t+1} is i.i.d. Bernoulli (which is the case under the null hypothesis of correct VaR specification), D_i is geometrically distributed with parameter p:

$$f^i(d) = P(D_i = d) = p(1 - p)^{d-1}$$

The survival function $S^i(d)$ is defined as

$$S^i(d) = P(D_i \geq d) = (1 - p)^{d-1}$$

The hazard function λ_d^i is the ratio between the probability density function (pdf) and the survival function:

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

The geometric distribution has a flat hazard function, making it the only discrete distribution with the memoryless property. For any discrete distribution, it holds that

$$P(D_i = d) = \lambda^i_d(1 - \lambda^i_{d-1})...(1 - \lambda^i_1)$$

$$P(D_i \geq d) = (1 - \lambda^i_{d-1})...(1 - \lambda^i_1)$$

It is seen that λ^i_d measures the probability of getting a hit on day d, given that the i-th no-hit duration has survived for $d - 1$ days. In other notation:

$$\lambda^i_d = P(I_{t_i+d} = 1 | I_{t_i+d-1} = 0, ..., I_{t_i+1} = 0, I_{t_i} = 1, \Omega_{t_i+d-1})$$

The geometric-VaR test defines the hazard function as follows:

$$\lambda^i_d = ad^{b-1}e^{-c\text{VaR}_{t_i+d}} \tag{1}$$

with $0 < a < 1$, measuring unconditional coverage. $0 \leq b \leq 1$, measuring duration dependence of VaR violations and $c \geq 0$, measuring the percentual change caused by a unit change in VaR_{t+1}. Under the null hypothesis of correct VaR specification, it holds that $a = p$, $b = 1$ and $c = 0$. The hypotheses can also be tested individually, introducing six tests that are explained in the next section.

2.2 Test implementation

The hypotheses are tested using LR tests. The log-likelihood has to be computed to perform the LR test. The tests are based on durations. It can be that the data is left- and/or right-censored. That is, the hit sequence does not start and/or end with a violation. Therefore, define variables C_1 and C_N. These variables are equal to 0 when the hit sequence respectively starts ($I_1 = 1$) and ends ($I_N = 1$) with a VaR violation and 1 else. The log-likelihood function is given by

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

where Θ is estimated using maximum likelihood. The formula for the test statistic is then

$$\text{LR} = -2\Big(\log(L(D|\Theta_0)) - \log(L(D|\Theta_1))\Big)$$

where Θ_0 is the maximum likelihood estimate under the null hypothesis and Θ_1 is the maximum likelihood estimate under the alternative hypothesis. The test statistics for all six tests are computed through this LR test. The tests are given below.

1. The unconditional coverage (UC) test tests whether the percentage of VaR violations equals a certain rate p. It is assumed that $b = 1$ and $c = 0$:

$$\text{LR}^{\text{UC}} = -2\Big(\log(L(D|a = p, b = 1, c = 0)) - \log(L(D|\hat{a}, b = 1, c = 0))\Big) \tag{3}$$

2. The duration independence (Dind) test checks whether VaR violations are clustered together by testing whether $b = 1$. Under the alternative hypothesis ($b < 1$), the probability of getting a VaR violation would decrease as the no-hit duration gets longer: the hazard function would have 'memory'. The true coverage rate p does not matter for this test. For this reason, $\bar{a}$ is allowed to differ from $\hat{a}$. It is assumed that $c = 0$:

$$\text{LR}^{\text{Dind}} = -2\Big(\log(L(D|\bar{a}, b = 1, c = 0)) - \log(L(D|\hat{a}, \hat{b}, c = 0))\Big) \tag{4}$$

3. The VaR independence (Vind) test tests whether the probability of a VaR violation depends on past information. That is, whether $c = 0$. It does not depend on the true coverage rate p and time-dependence of VaR violations is allowed:

$$\text{LR}^{\text{Vind}} = -2\Big(\log(L(D|\bar{a}, \bar{b}, c = 0)) - \log(L(D|\hat{a}, \hat{b}, \hat{c}))\Big) \tag{5}$$

4. The geometric (G) test tests jointly whether $a = p$ and $b = 1$. $c = 0$ is assumed:

$$\text{LR}^{\text{G}} = -2\Big(\log(L(D|a = p, b = 1, c = 0)) - \log(L(D|\hat{a}, \hat{b}, c = 0))\Big) \tag{6}$$

5. The VaR test is also a joint test: it tests for $a = p$ and $c = 0$. $b = 1$ is assumed:

$$\text{LR}^{\text{VaR}} = -2\Big(\log(L(D|a = p, b = 1, c = 0)) - \log(L(D|\hat{a}, b = 1, \hat{c}))\Big) \tag{7}$$

6. Lastly, the geometric-VaR (GV) test tests for all three parameters:

$$\text{LR}^{\text{GV}} = -2\Big(\log(L(D|a = p, b = 1, c = 0)) - \log(L(D|\hat{a}, \hat{b}, \hat{c}))\Big) \tag{8}$$

Also, it holds that

$$\text{LR}^{\text{G}} = \text{LR}^{\text{UC}} + \text{LR}^{\text{Dind}}$$

$$\text{LR}^{\text{VaR}} = \text{LR}^{\text{UC}} + \text{LR}^{\text{Vind}}$$

$$\text{LR}^{\text{GV}} = \text{LR}^{\text{UC}} + \text{LR}^{\text{Dind}} + \text{LR}^{\text{Vind}}$$

The advantage of these tests is that they do not only check whether VaR forecasts are misspecified, but also how they are misspecified by checking which individual hypothesis is rejected.

The data is binary, so the distribution of the LR statistics is not continuous. Also, parameter values are tested at different boundaries of the parameter space. This may give different asymptotic distributions for different null hypotheses. However, the biggest complication is that a minimum number of violations has to be observed to run the backtest: the Vind and the GV tests require estimation of three parameters. Therefore, the hit sequence needs at least three hits. Three hits are sufficient to run the backtest if the hit sequence is only left- or right-censored, or not censored at all. The same holds for the Dind, G and VaR tests which require estimation of two parameters and therefore also require at least two hits in the hit sequence. The Monte Carlo method from Dufour (2006) is useful to get the simulated distribution under the null hypothesis and to get reliable p-values in finite samples. In the Monte Carlo method, N realizations of data under the null hypothesis are generated. Then, the test statistics $\text{LR}_i, i = 1, ..., N$ are computed. Those statistics form the simulated distribution under the null hypothesis. If LR_0 is the test statistic from the sample, it can be compared with LR_i. The distribution of the test statistics is discrete, so LR_0 can be equal to LR_i. Therefore, U_i is drawn from the uniform distribution on [0,1] for each test statistic. The Monte Carlo p-value is then given by

$$\hat{p}_N(LR_0) = \frac{N\hat{G}(LR_0) + 1}{N + 1} \tag{9}$$

with

$$\hat{G}(LR_0) = 1 - \frac{1}{N}\sum_{i=1}^{N} I(LR_i \leq LR_0) + \frac{1}{N}\sum_{i=1}^{N} I(LR_i = LR_0)I(U_i \geq U_0)$$

2.3 (Filtered) Historical Simulation and GARCH-modeling

A widely used method for computing VaR is Historical Simulation (HS). As explained in Section 2.1, VaR is a quantile of the conditional distribution of returns. That is

$$\text{VaR}_{t+1}(p) = F_{t+1}^{-1}(p)$$

Note that in this notation, VaR can be negative. F_{t+1} is the cumulative distribution function (cdf) of r_{t+1}. HS is widely used because it is nonparametric and easy to implement. It estimates F_{t+1} using realized returns from the past year(s), with a moving window in this case:

$$\text{VaR}_{t+1}^{HS}(p) = \text{percentile}(\{R_s\}_{s=t-T_e+1}^{t}, p)$$

where T_e is the size of the moving window. In practice, T_e is chosen to be 250, the number of trading days in a year. The crypto market operates continuously, so it would be convenient to choose $T_e = 365$. For simplicity, T_e remains 250.

HS also has its disadvantages: it assumes that the returns within the window are all i.i.d. and, apart from the fact that some characteristics do change through the moving window, it does not take time-varying characteristics of returns into account. Also, outliers can have impact on the VaR forecasts as long as they are in the moving window.

Some characteristics of asset returns are fat tails, volatility not being constant over time (volatility clustering) and that negative returns have a bigger impact on volatility than positive returns (leverage effect). GARCH models are often used to capture those characteristics. In this research, the nonlinear asymmetric GARCH (NAGARCH) model as described in Engle & Ng (1993) is used. This model takes leverage effects into account with parameter θ. Large negative returns have a bigger impact on volatility than large positive returns when $\theta > 0$. The data generating process for the returns and volatility is the same as in Pelletier & Wei (2016):

$$\begin{aligned} R_{t+1} &= \sigma_{t+1}((d-2)/d)^{1/2} z_{t+1} \\ \sigma_{t+1}^2 &= \omega + \alpha\sigma_t^2\left(\left(\frac{d-2}{d}\right)^{1/2} z_t - \theta\right)^2 + \beta\sigma_t^2 \end{aligned} \tag{10}$$

with $\omega, \alpha > 0$, $\beta \geq 0$ and $-1 \leq \theta \leq 1$. z_t is drawn from the $t(d)$-distribution. The process is stationary when $\alpha(1+\theta^2)+\beta < 1$. The unconditional variance of the returns is given by $\omega/(1-\alpha(1+\theta^2)-\beta)$.

When combining HS with GARCH, we can get the best of both worlds: a HS method that takes time-varying properties into account. Barone-Adesi et al. (1999) introduce this method and call it Filtered Historical Simulation (FHS). FHS still has the nonparametric properties of HS when looking at returns but it makes assumptions in the GARCH process about the distribution of the standardized error terms z_t, which makes this method semi-parametric. With FHS, one can simulate returns to compute VaRs. In this research, the $t(d)$ distribution is used for z_t and returns using FHS are generated with the process in equation (10).

2.4 RiskMetrics

Another method to estimate volatility is RiskMetrics, introduced by J. P. Morgan/Reuters (1996). Volatility is estimated using an exponentially weighted moving average:

$$\sigma_{t+1}^2 = \lambda \sigma_t^2 + (1-\lambda)(r_t - \bar{r})^2 \tag{11}$$

where r_t is the return at time t and $\bar{r} = \frac{1}{t}\sum_{i=1}^{t} r_i$, the mean of the returns. λ is called the decay factor. For daily returns, it is optimal to choose $\lambda = 0.94$.

There are four methods introduced to generate VaR estimates: HS, GARCH, RiskMetrics and FHS. For the first three methods, VaRs are computed with the inverse cdf: $\text{VaR}_{t+1}(p) = F_{t+1}^{-1}(p)$ where $F_{t+1}(p)$ is the cdf of the $t(d)$ distribution with variance equal to σ_{t+1}^2. With FHS, returns are generated. The VaR estimate is equal to the 5% percentile of the past T_e returns. For all methods, the hit sequence is obtained by comparing realized returns with VaRs.

3 Data

Pelletier & Wei (2016) apply the test on profits and losses (P/Ls) from four business lines of a large commercial bank. This data is not available. However, they use the data to estimate the parameters of NAGARCH-$t(d)$ process in equation (10) such that the data can be simulated. In the first part of the replication, the size of the six tests is reported. The values in the first column of Table 1 are used to estimate the conditional variance with the NAGARCH-$t(d)$ process such that returns and VaRs can be generated. The returns and VaRs are used to assess the size properties for each test. In the second part, the values in the remaining four columns of Table 1 are used to estimate the conditional variance and eventually to generate returns and VaRs. These returns and VaRs are used to assess the power of the tests for each of the four business lines.

Table 1: Parameter values for the replication part

	Size	Business line 1	Business line 2	Business line 3	Business line 4
d	10	3.808	3.318	6.912	4.702
θ	0	−0.245	0.503	−0.962	0.093
β	0.93	0.749	0.928	0.873	0.915
α	0.05	0.155	0.052	0.026	0.072
ω	0.21	0.550	0.215	0.213	1.653

Note: These parameters are used to simulate returns to test the size and power of the six tests. The parameter values in the first column are used assess the size of the tests. The parameter values in the other columns are used to assess the power of the tests.

For the extension part, daily prices (in US dollar) from five cryptocurrencies are available from CoinMetrics. The five cryptocurrencies are Bitcoin, Dogecoin, Litecoin, Monero and Ripple. With the prices, one can compute the log-returns:

$$r_{t+1} = \ln\left(\frac{P_{t+1}}{P_t}\right)$$

where P_t is the price level at time t. The estimation sample for the returns is from 25 May 2014 to 17 February 2017, exactly 1000 observations. Returns are available until 20 June 2018. The sample from 18 February 2017 to 20 June 2018 (488 observations) is used as backtesting sample. Note that the prices of several cryptocurrencies made a rapid increase (and decrease) in the backtesting sample period. The descriptive statistics are given in Table 2.

Table 2: Descriptive statistics for the cryptocurrencies

	Bitcoin	Dogecoin	Litecoin	Monero	Ripple
Mean	0.0678	−0.0688	−0.1087	0.1499	0.0578
Std. dev.	3.1211	5.0384	4.7941	7.5830	4.5778
Maximum	16.7085	52.1150	36.1613	56.7670	32.7988
Minimum	−23.5570	−27.0973	−51.9272	−31.0968	−22.1142
Skewness	−0.9048	1.6511	−0.6745	0.7954	1.0357
Kurtosis	11.7397	19.3787	25.7117	9.1846	11.4323

Note: This table shows some descriptive statistics for the daily returns of the cryptocurrencies. The mean, maximum and minimum are in %.

4 Results

All performed tests have a significance level of 10%. First, the size of the tests is assessed through simulation. Table 2 from Pelletier & Wei (2016) reports the values for size for 5% VaR. The sample size T varies from 250 to 1500. The VaR independence, VaR and geometric-VaR tests also require VaR estimates as input. For the size, these estimates are computed with the inverse cdf of the $t(d)$ distribution with variance σ^2_{t+1}. The coverage rate p is equal to 5%.

Table 3 reports the size of the six tests. The size is the rejection frequency of 10000 replications. In the first part, the size is assessed using critical values of the chi-squared distribution. The hit sequence is obtained from T Bernoulli trials. All tests are undersized except for the test for unconditional coverage, which is oversized but may converge to 10% as the sample size gets larger. The conclusions differ from Pelletier & Wei (2016) when looking at the VaR independence and the VaR tests as they found that those tests are oversized for the smaller sample sizes. Furthermore, the size for the VaR independence, the VaR and the geometric-VaR tests is substantially lower than the size reported by the developers of the test. Since these tests require VaR estimates as input, it could be that the implementation and/or computation of the log-likelihood with the VaR estimates is done in a slightly different way in this research.
In the second part, the size is computed using simulated critical values. The NAGARCH-$t(d)$ process from equation (10) is used to generate returns. The hit sequence is obtained by comparing returns with VaRs. The critical values are taken as the 90% percentile of 10000 simulations with sample size 50000. The values are different from the chi-squared critical values. Apart from the unconditional coverage test, the size of all tests increases when using the simulated critical values. The geometric and VaR tests go from being undersized in the first part to having the correct size. The duration independence, VaR independence and geometric-VaR test are still undersized.

The most important conclusions are the same: Table 3 shows that the test statistics are not chi-squared distributed under the null hypothesis and the finite sample properties differ from the asymptotic ones. This gives a good reason to use the Monte Carlo method from Dufour (2006) (see Section 2.2).

Table 3: Size of 10% Duration-based tests applied to 5% VaR

Sample Size	UC	Dind	Vind	G	VaR	GV
	Testing with chi-squared critical values					
250	0.1411	0.0287	0.0123	0.0606	0.0573	0.0347
500	0.1183	0.0289	0.0155	0.0647	0.0553	0.0340
750	0.1236	0.0345	0.0236	0.0642	0.0566	0.0343
1000	0.1196	0.0343	0.0231	0.0619	0.0525	0.0336
1250	0.1151	0.0350	0.0246	0.0607	0.0555	0.0390
1500	0.1088	0.0381	0.0266	0.0582	0.0540	0.0397
	Chi-squared critical values					
	2.7055	2.7055	2.7055	4.6052	4.6052	6.2514
	Testing with simulated critical values					
250	0.1415	0.0666	0.0375	0.1072	0.1010	0.0751
500	0.1183	0.0671	0.0517	0.1061	0.1009	0.0721
750	0.0984	0.0674	0.0592	0.0935	0.0960	0.0773
1000	0.0855	0.0781	0.0705	0.0989	0.1000	0.0821
1250	0.0904	0.0765	0.0746	0.1015	0.0947	0.0814
1500	0.1031	0.0784	0.0729	0.0945	0.0999	0.0823
	Simulated critical values					
	2.5840	1.6253	2.0185	3.7593	3.7967	4.9496

Note (1): The size is computed as the rejection frequency from 10000 replications.

Note (2): The size from the tests using the chi-squared critical values is computed by generating i.i.d. Bernoulli hit sequences and VaRs that are independent of the simulated hit sequence. The critical values are from the chi-squared distribution.

Note (3): The size from the tests using simulated critical values is computed by generating returns with the NAGARCH-$t(d)$ process (see Table 1) and obtaining the hit sequence by comparing returns to VaRs. The critical values are computed from 10000 simulations with sample size 50000.

Second, the power the tests is assessed. Table 4 of Pelletier & Wei (2016) reports the values for the power for 5% VaR. Returns are generated with the NAGARCH-$t(d)$ process in equation (10) with sample size $T_e + T$ where $T_e = 250$ and T varies from 250 to 1500. The VaR forecasts are obtained from HS with the size of the moving window equal to T_e. The sample hit sequence is obtained by comparing returns to VaRs. In the Monte Carlo method, $N = 9999$ Bernoulli hit sequences of sample size T are generated. VaR estimates are computed with the inverse cdf of the $t(d)$ distribution with variance σ^2_{t+1}: $\text{VaR}_{t+1}(p) = F^{-1}_{t+1}(p)$. With the hit sequences and the VaRs, the test statistics LR_i, $i = 1, ..., N$ form the simulated distribution under the null hypothesis. The p-value is computed as in equation (9). A test rejects the null hypothesis if the p-value is smaller than the significance level, which is 10% in this case. The power is the rejection frequency from 5000 replications. Results are in Table 4.

Table 4: Power of 10% Duration-based tests applied to 5% VaR

Sample Size	UC	Dind	Vind	G	VaR	GV
			Business Line 1			
250	0.1538	0.5058	0.5278	0.3828	0.3204	0.5156
500	0.0606	0.7076	0.4912	0.5180	0.3966	0.6642
750	0.0369	0.8208	0.7106	0.6588	0.5520	0.8312
1000	0.0254	0.8808	0.8284	0.7280	0.6826	0.9162
1250	0.0246	0.9260	0.8990	0.8132	0.7578	0.9602
1500	0.0194	0.9554	0.9312	0.8796	0.8526	0.9814
			Business Line 2			
250	0.2298	0.3790	0.4930	0.3590	0.3714	0.4816
500	0.1266	0.5624	0.6348	0.4132	0.5068	0.6430
750	0.0912	0.6728	0.7840	0.5276	0.6396	0.7830
1000	0.0638	0.7588	0.8574	0.5940	0.7426	0.8632
1250	0.0728	0.8268	0.8946	0.6660	0.8110	0.9234
1500	0.0620	0.8604	0.9246	0.7294	0.8554	0.9522
			Business Line 3			
250	0.0710	0.1660	0.4870	0.1038	0.1874	0.1904
500	0.0140	0.2212	0.6114	0.0702	0.2958	0.3012
750	0.0092	0.2338	0.8770	0.0988	0.6080	0.5812
1000	0.0054	0.2468	0.9188	0.0852	0.6838	0.6804
1250	0.0024	0.2734	0.9630	0.1052	0.8014	0.7974
1500	0.0028	0.2958	0.9764	0.1190	0.8580	0.8504
			Business Line 4			
250	0.3340	0.4904	0.5344	0.5092	0.4958	0.6078
500	0.2232	0.7364	0.6686	0.6356	0.6374	0.7948
750	0.1942	0.8620	0.7968	0.7698	0.7472	0.9038
1000	0.1754	0.9200	0.8580	0.8378	0.8118	0.9482
1250	0.1716	0.9532	0.8482	0.8924	0.8268	0.9718
1500	0.1670	0.9768	0.8730	0.9328	0.8830	0.9878

Note: For each business line, returns are generated with the corresponding parameter values (see Table 1). VaRs are computed using Historical Simulation with a moving window of size 250. The power is the rejection frequency from 5000 replications.

The power of the unconditional coverage test is low for all business lines. This is because this test only checks whether the probability of getting a hit is equal to the coverage rate $p = 5\%$. The duration independence test has high power, except for business line 3, which has a large negative θ. This means that large negative returns do not have a great impact on volatility compared to large positive returns: volatility will be low after a large negative return and therefore, volatility clustering is less likely to happen. The VaR independence test has the best power in this business line. The power of the geometric and the VaR tests are lower because their test statics are equal to the sum of the test statistics of two individual tests. One of the two being the test statistic for unconditional coverage. The geometric-VaR test has the highest power in the other business lines. To summarize, the conclusions are in line with the conclusions of the developers.

The parameters of the NAGARCH-$t(d)$ model are also estimated for the cryptocurrencies. A little side note is that the process for all cryptocurrencies is non stationary without imposing the corresponding restriction. The estimates in Table 5 are obtained after imposing the stationarity restriction.

Table 5: Parameter estimates for cryptocurrencies

	Bitcoin	Dogecoin	Litecoin	Monero	Ripple
d	3.3082	3.1207	2.8752	3.2633	3.0468
θ	0.0287	0.0411	0.0100	0.1052	0.1408
β	0.6868	0.3196	0.7646	0.0000	0.1515
α	0.3128	0.6792	0.2353	0.9890	0.8320
ω	1.4891	10.4612	1.8081	57.6695	11.0743

Note: This table shows the parameter estimates of the NAGARCH-$t(d)$ model for each cryptocurrency.

Figures 1-4 in the Appendix show the returns together with the estimated VaRs and the magnitude of the violations for all four methods. Tables 6-9 report the number of VaR violations and the value of the LR test statistics with the Monte Carlo p-value between parentheses below for each test and for all four methods. All methods provide independent VaR estimates, as the p-values for the VaR independence test are all higher than 0.6 for each cryptocurrency in each method. In cases where the unconditional coverage test rejects, it is because the number of hits is too high when compared with the number of hits expected under the null hypothesis (except for one case). An explanation for this is that cryptocurrencies are more volatile than traditional assets: large positive and negative returns occur more often.

The geometric-VaR test rejects the HS approach only for Dogecoin, where all other hypotheses (except for VaR independence) are also rejected. Under the null hypothesis, approximately 25 violations are expected. The unconditional coverage test rejects Bitcoin at 10% because there are more violations than expected under the null hypothesis. The geometric test also rejects Bitcoin mainly because the unconditional coverage is not correct. Note that none of the tests would be rejected when testing with a significance level of 5%. The duration independence test rejects

Table 6: Backtesting 5% VaRs obtained from Historical Simulation

	No. of violations	UC	Dind	Vind	G	VaR	GV
Bitcoin	38	4.0906 (0.0546)	0.2820 (0.2223)	0.0000 (0.9516)	4.3726 (0.0620)	4.0906 (0.1356)	4.3726 (0.1466)
Dogecoin	40	5.5546 (0.0195)	8.3126 (0.0019)	0.0000 (0.9517)	13.8672 (0.0008)	5.5546 (0.0697)	13.8672 (0.0028)
Litecoin	34	1.7782 (0.2014)	0.3574 (0.2048)	0.0000 (0.9608)	2.1356 (0.2220)	1.7782 (0.3873)	2.1356 (0.3866)
Monero	28	0.0579 (0.8393)	3.4745 (0.0192)	0.0000 (0.9516)	3.5324 (0.1010)	0.0579 (0.9245)	3.5324 (0.2111)
Ripple	36	2.8277 (0.1001)	0.9708 (0.1145)	0.0000 (0.9517)	3.7985 (0.0871)	2.8277 (0.2342)	3.7985 (0.1858)

Note: This table shows the number of violations and the value of the LR statistics when the VaR estimates are obtained using Historical Simulation.

Monero. It is seen in Figure 1 that VaR violations come in clusters. However, the geometric test does not reject Monero because there is correct unconditional coverage.

The VaR estimates obtained from GARCH perform well in the duration independence test as the test does not reject the null hypothesis for all coins except for Bitcoin. However, the unconditional coverage test and therefore also the geometric-VaR test heavily reject all coins except for Monero. This method may have given VaR estimates that are too low[1]. This is seen in Figure 2, especially for Bitcoin and Litecoin.

Table 7: Backtesting 5% VaRs obtained from GARCH

	No. of violations	UC	Dind	Vind	G	VaR	GV
Bitcoin	62	32.3912 (0.0001)	1.6219 (0.0692)	0.0000 (0.9606)	34.0131 (0.0001)	32.3912 (0.0001)	34.0131 (0.0001)
Dogecoin	67	40.7118 (0.0001)	0.0000 (0.8363)	0.0000 (0.9616)	40.7118 (0.0001)	40.7118 (0.0001)	40.7118 (0.0001)
Litecoin	73	51.5739 (0.0001)	0.3268 (0.2113)	0.0000 (0.9606)	51.9007 (0.0001)	51.5739 (0.0001)	51.9007 (0.0001)
Monero	20	1.8527 (0.2014)	0.2291 (0.2375)	0.0000 (0.7898)	2.0818 (0.2263)	1.8527 (0.3776)	2.0818 (0.3935)
Ripple	65	37.3002 (0.0001)	0.0000 (0.8295)	0.0000 (0.9496)	37.3002 (0.0001)	37.3002 (0.0001)	37.3002 (0.0001)

Note: This table shows the number of violations and the value of the LR statistics when the VaR estimates are obtained using the NAGARCH-t(d) process.

The geometric-VaR test rejects the RiskMetrics approach only for Bitcoin. This is because he unconditional coverage test is rejected. The duration and VaR independence tests do not reject Bitcoin, but the geometric and VaR tests do reject due to the rejection of the unconditional coverage test. The unconditional coverage test also rejects Monero, but all other tests do not.

[1] VaR is reported as a positive number.

No test is rejected for the other coins. The duration independence test does not reject for all coins. This suggests that RiskMetrics might be a good approach to get VaR estimates from.

Table 8: Backtesting 5% VaRs obtained from RiskMetrics

	No. of violations	UC	Dind	Vind	G	VaR	GV
Bitcoin	41	6.3585 (0.0182)	0.6071 (0.1589)	0.0000 (0.7137)	6.9656 (0.0196)	6.3585 (0.0515)	6.9656 (0.0491)
Dogecoin	33	1.3378 (0.2267)	0.0000 (0.8293)	0.0000 (0.9616)	1.3378 (0.3414)	1.3378 (0.4561)	1.3378 (0.5328)
Litecoin	25	0.1141 (0.7533)	0.0002 (0.3907)	0.0000 (0.9510)	0.1143 (0.8290)	0.1141 (0.8783)	0.1143 (0.9168)
Monero	37	3.4333 (0.0628)	0.0000 (0.8363)	0.0000 (0.6328)	3.4333 (0.1054)	3.4333 (0.1769)	3.4333 (0.2219)
Ripple	28	0.0579 (0.8393)	0.0000 (0.9883)	0.0000 (0.9430)	0.0579 (0.8931)	0.0579 (0.9245)	0.0579 (0.9512)

Note: This table shows the number of violations and the value of the LR statistics when the VaR estimates are obtained using RiskMetrics.

The geometric-VaR test rejects VaRs obtained from FHS for all coins. The test is rejected for Bitcoin because violations come in clusters as seen in Figure 4 and therefore, the duration independence test is rejected. All other methods reject correct unconditional coverage for Bitcoin VaR estimates, but FHS does not. The unconditional coverage test rejects Monero because the amount of VaR violations is too low. This is due to the high ω that is estimated in the NAGARCH process (see Table 5). For the other coins, the geometric, VaR and geometric-VaR tests reject because the unconditional coverage test heavily rejects. The NAGARCH-$t(d)$ model might not be appropriate to generate returns for cryptocurrencies.

Table 9: Backtesting 5% VaRs obtained from Filtered Historical Simulation

	No. of violations	UC	Dind	Vind	G	VaR	GV
Bitcoin	34	1.7782 (0.2014)	9.5202 (0.0005)	0.0000 (0.7108)	11.2984 (0.0033)	1.7782 (0.3873)	11.2984 (0.0088)
Dogecoin	63	33.9991 (0.0001)	0.0038 (0.3741)	0.0000 (0.9613)	34.0029 (0.0001)	33.9991 (0.0001)	34.0029 (0.0001)
Litecoin	72	49.7007 (0.0001)	0.0000 (0.8363)	0.0000 (0.9619)	49.7007 (0.0001)	49.7007 (0.0001)	49.7007 (0.0001)
Monero	9	16.1596 (0.0003)	0.0000 (0.9981)	0.0000 (0.6216)	16.1596 (0.0003)	16.1596 (0.0006)	16.1596 (0.0006)
Ripple	66	38.9924 (0.0001)	0.0000 (0.9623)	0.0000 (0.9606)	38.9924 (0.0001)	38.9924 (0.0001)	38.9924 (0.0001)

Note: This table shows the number of violations and the value of the LR statistics when the VaR estimates are obtained with Filtered Historical Simulation. Volatility is estimated with the NAGARCH-$t(d)$ process.

5 Conclusion and Further Research

This research uses four methods to estimate 5% VaRs for five cryptocurrencies: Bitcoin, Dogecoin, Litecoin, Monero and Ripple. The four methods are Historical Simulation, GARCH, RiskMetrics and Filtered Historical Simulation. The geometric-VaR test from Pelletier & Wei (2016) evaluates those estimates. This test tests jointly for correct unconditional coverage, duration independence and VaR independence. It can detect how VaRs are misspecified by checking which individual hypothesis is rejected. All methods have no problem in estimating independent VaRs for cryptocurrencies. RiskMetrics is the only method that also estimates time-independent VaRs for all coins. This is also the best approach from all four methods to get VaR estimates, only rejecting the geometric-VaR test for Bitcoin. The GARCH method is more appropriate for Monero than RiskMetrics since no tests are rejected. For Bitcoin, either Historical Simulation or RiskMetrics can be used. However, the unconditional coverage test rejects for these methods. Historical Simulation is preferred because the geometric-VaR test does not reject.

This paper uses a handful of methods to backtest VaRs from cryptocurrencies. For the GARCH and FHS methods, only the NAGARCH model is used. This paper does not focus on the ability of the used models to forecast returns for cryptocurrencies. On the other hand, the cryptocurrencies made a rapid increase (and decrease) in the backtesting sample. The models will possibly have better forecasting power when this sample is taken into the estimation sample. Chu et al. (2017) fit different GARCH models for cryptocurrencies. A topic for further research would be how VaRs (and generated returns) from their best models would perform in the geometric-VaR test. Another risk measure is expected shortfall, which uses VaRs to estimate the loss in the next period. If VaRs are correctly specified, the next step would be to compute the possible loss in the next period (see e.g. Chan et al. (2017)) and eventually develop a new backtest. This is left as a topic for further research.

Bibliography

Barone-Adesi, G., Giannopoulos, K. & Vosper, L. (1999), 'VaR Without Correlations for Portfolios of Derivative Securities.', *Journal of Futures Markets* **19**(5), 583–602.

Chan, S., Chu, J., Nadarajah, S. & Osterrieder, J. (2017), 'A Statistical Analysis of Cryptocurrencies.', *Journal of Risk and Financial Management* **10**(2), 1–23.

Christoffersen, P. F. (1998), 'Evaluating Interval Forecasts.', *International Economic Review* **39**(4), 841–862.

Christoffersen, P. F. & Pelletier, D. (2004), 'Backtesting Value-at-Risk: A Duration-Based Approach.', *Journal of Financial Econometrics* **2**(1), 84–108.

Chu, J., Chan, S., Nadarajah, S. & Osterrieder, J. (2017), 'GARCH Modelling of Cryptocurrencies.', *Journal of Risk and Financial Management* **10**(4), 1–15.

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. & Ng, V. K. (1993), 'Measuring and Testing the Impact of News on Volatility.', *Journal of Finance* **48**(5), 1749–78.

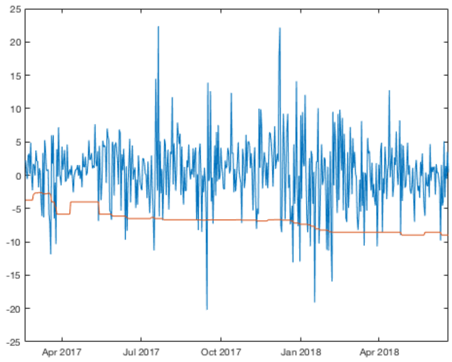
J. P. Morgan/Reuters (1996), RiskMetrics$^{\mathrm{TM}}$—Technical Document, 4th Edition, New York. **URL:** *https://www.msci.com/documents/10199/5915b101-4206-4ba0-aee2-3449d5c7e95a*

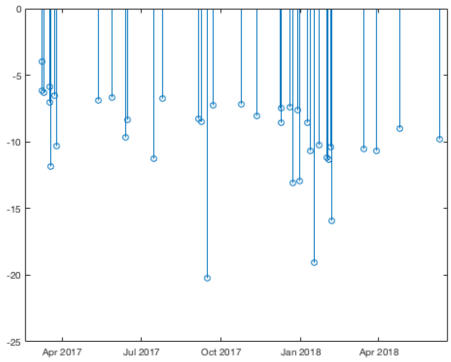
Manganelli, S. & Engle, R. F. (2001), Value at Risk Models in Finance. ECB Working Paper Series, Working Paper No. 75.

Pelletier, D. & Wei, W. (2016), 'The Geometric-VaR Backtesting Method.', *Journal of Financial Econometrics* **14**(4), 725–745.

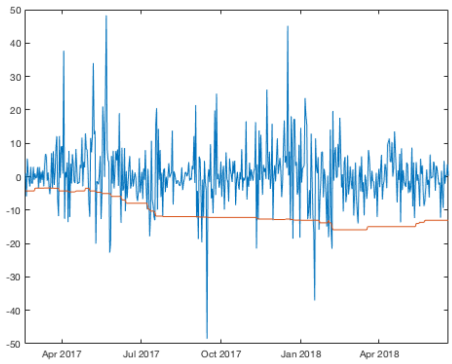
Appendix A Returns and VaRs from Historical Simulation

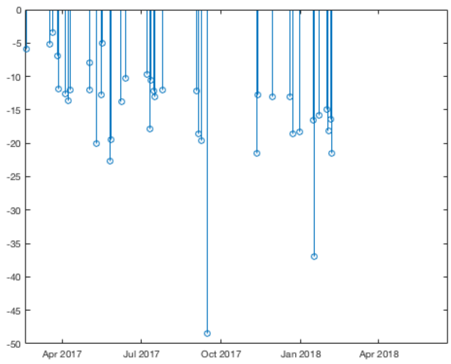
Bitcoin



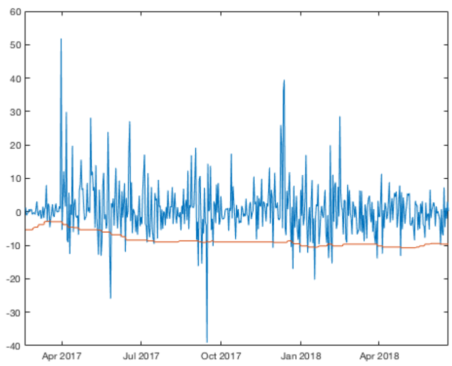


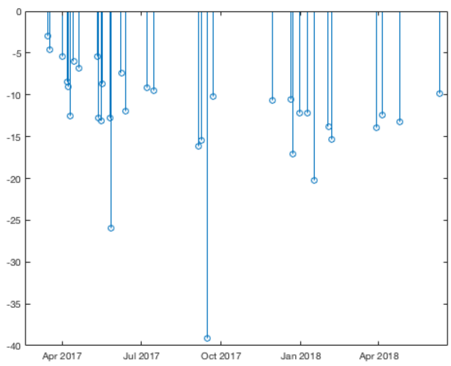
Dogecoin





Litecoin





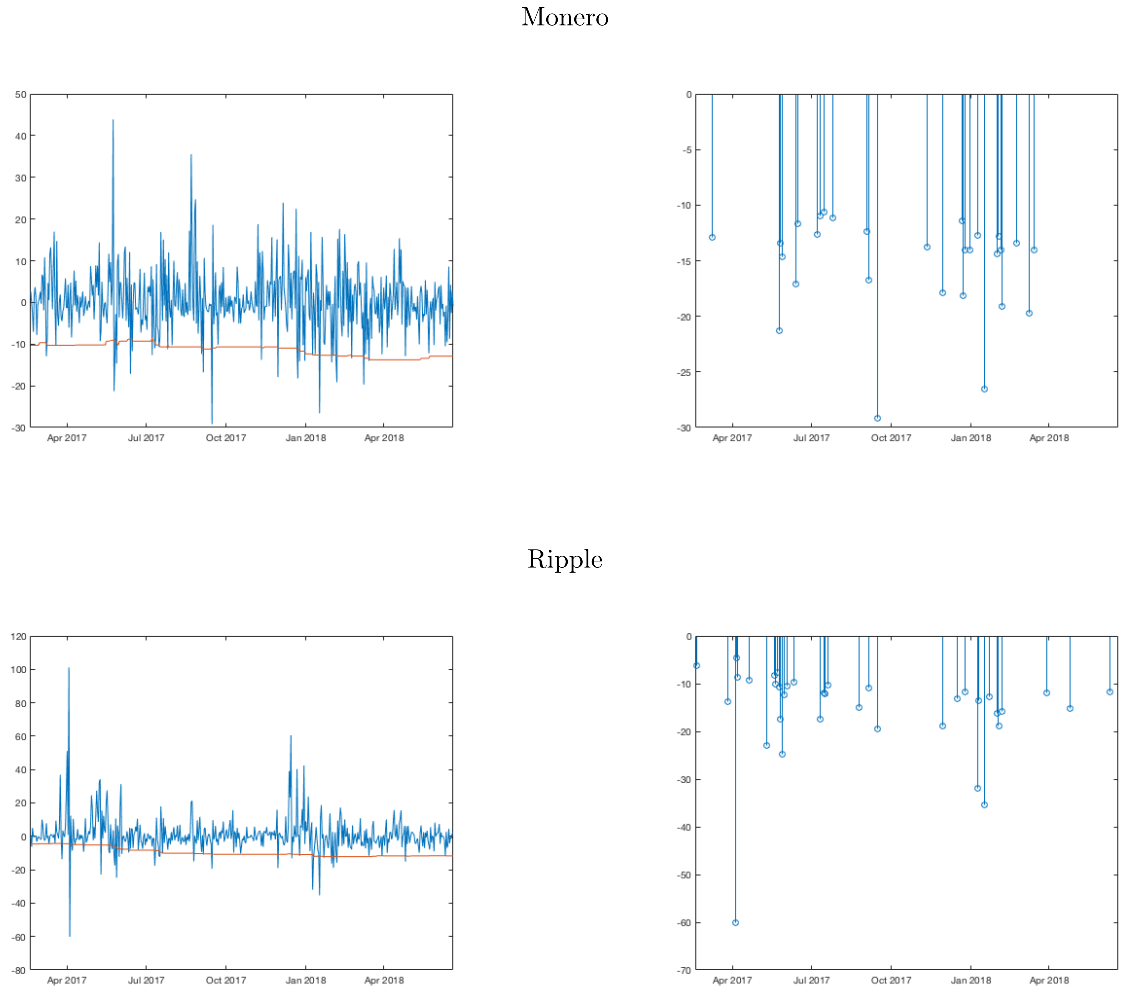
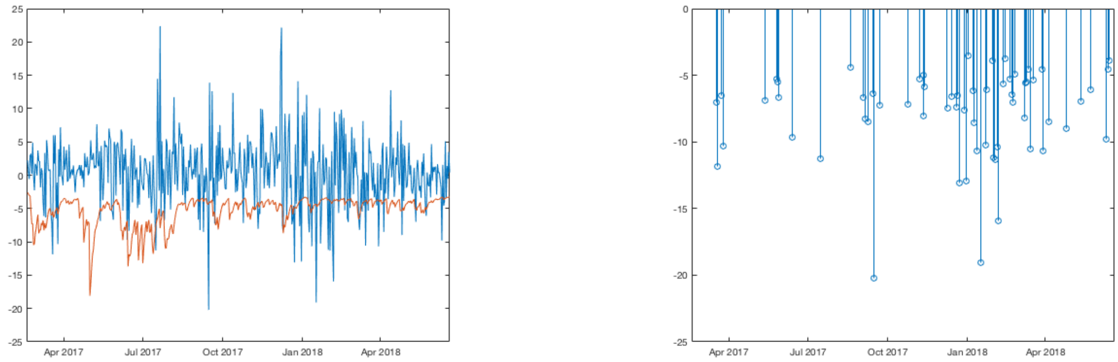


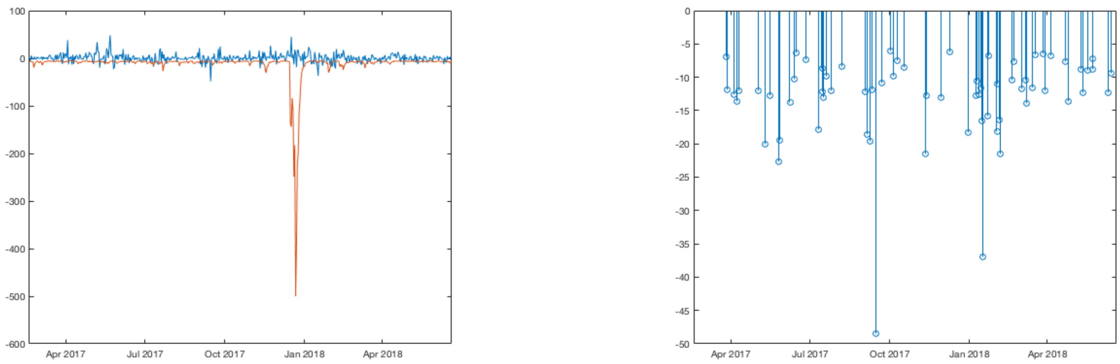
Figure 1: Time series of returns, VaRs and VaR violations for the cryptocurrencies. The left panel displays returns (blue) and one-day 5% VaRs obtained from HS (red). The right panel displays violations. The magnitude is the return on the day of the violation.

Appendix B Returns and VaRs from GARCH

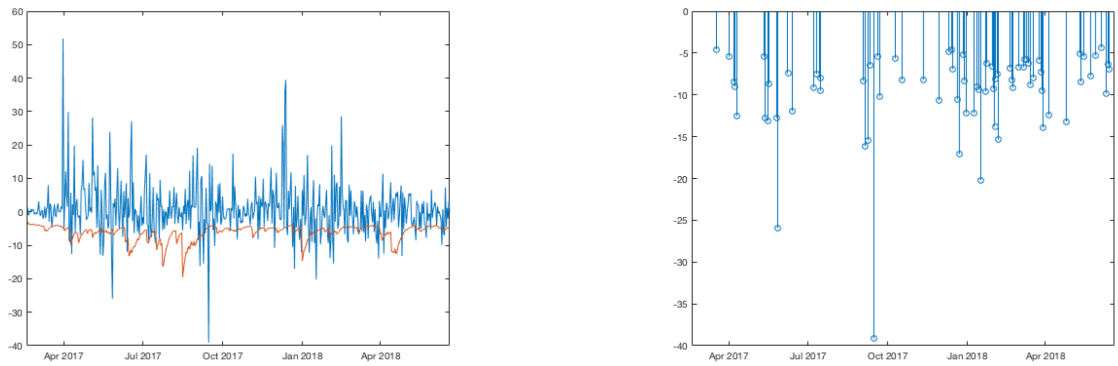
Bitcoin



Dogecoin



Litecoin



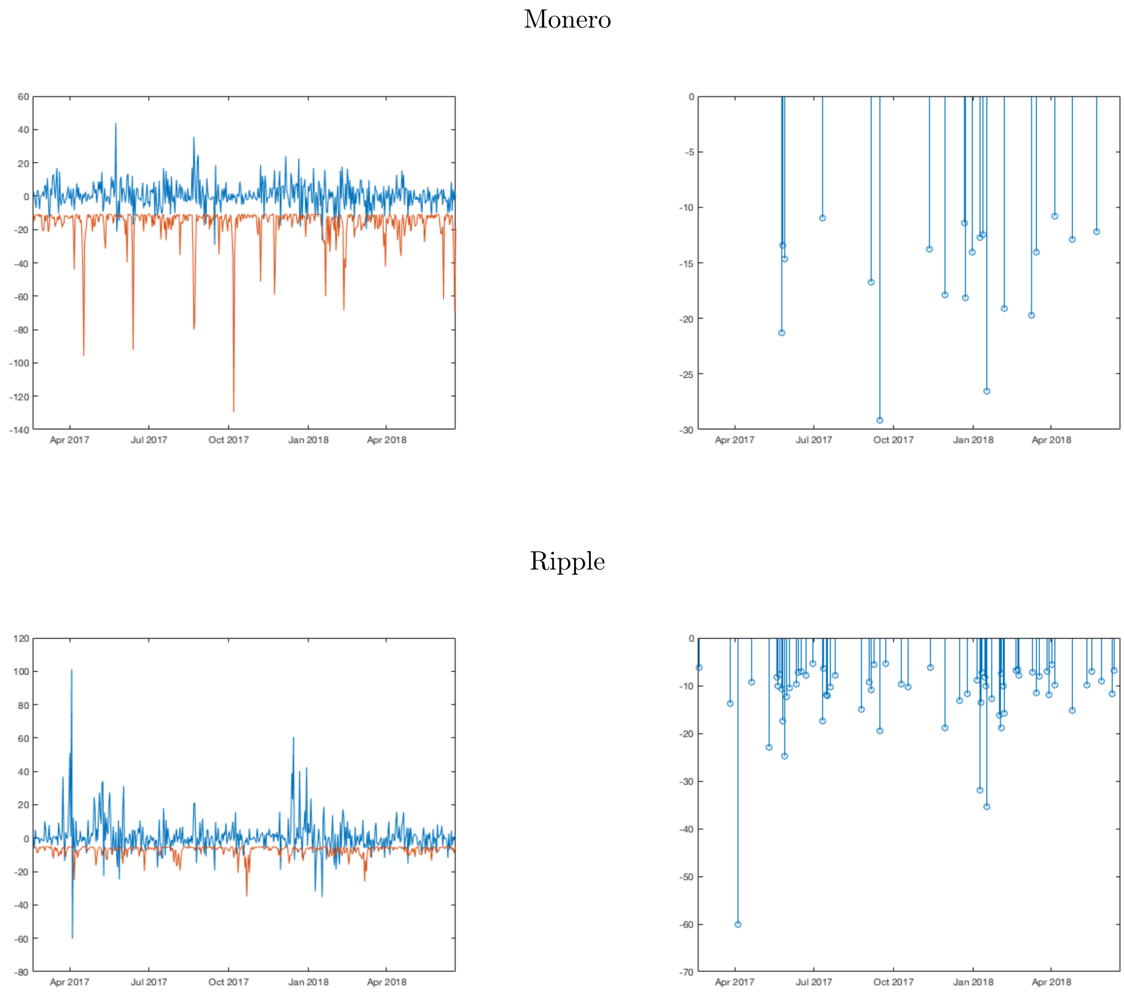
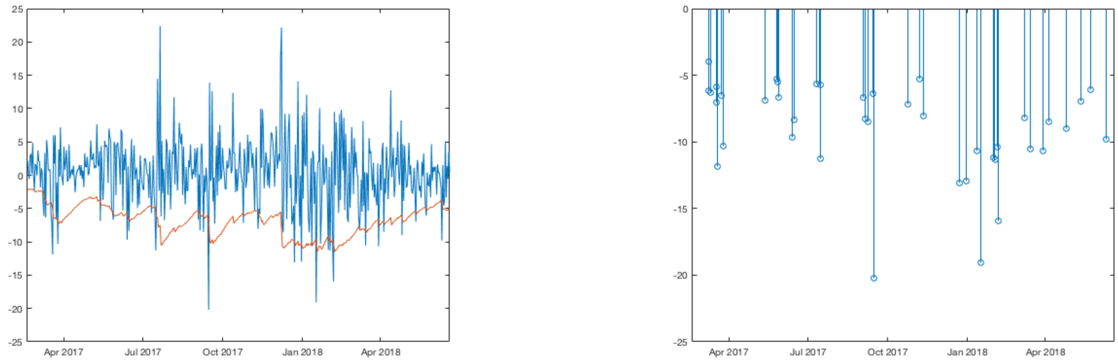


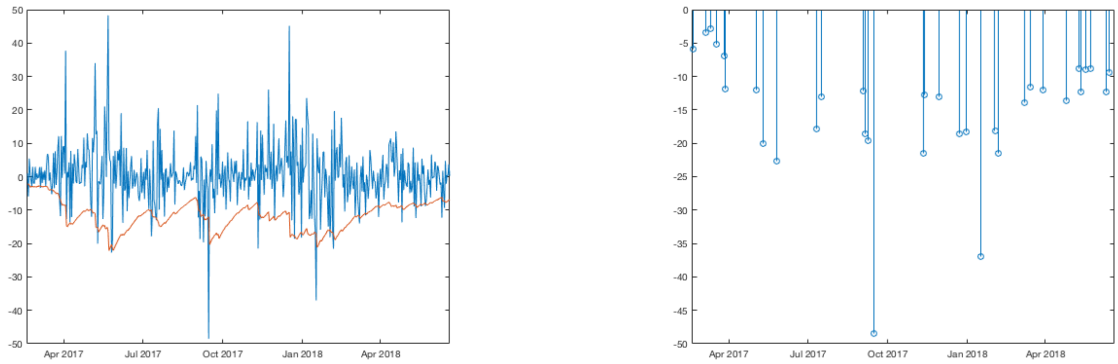
Figure 2: Time series of returns, VaRs and VaR violations for the cryptocurrencies. The left panel displays returns (blue) and one-day 5% VaRs obtained from GARCH (red). The right panel displays violations. The magnitude is the return on the day of the violation.

Appendix C Returns and VaRs from RiskMetrics

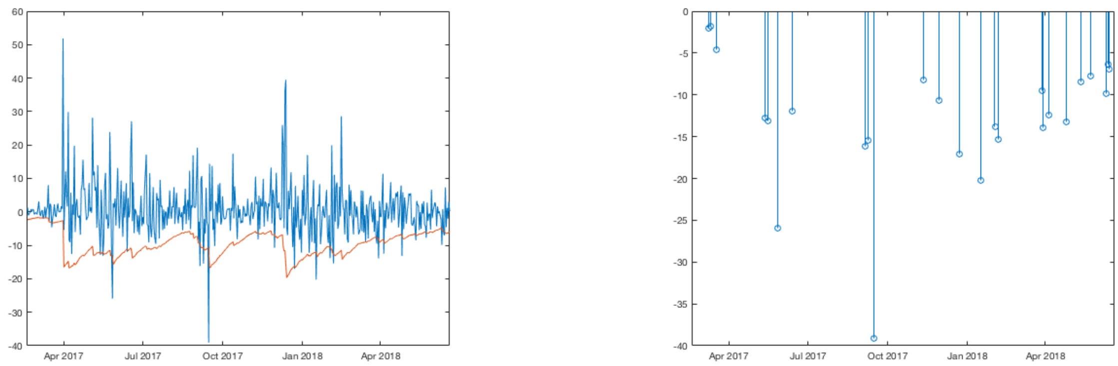
Bitcoin



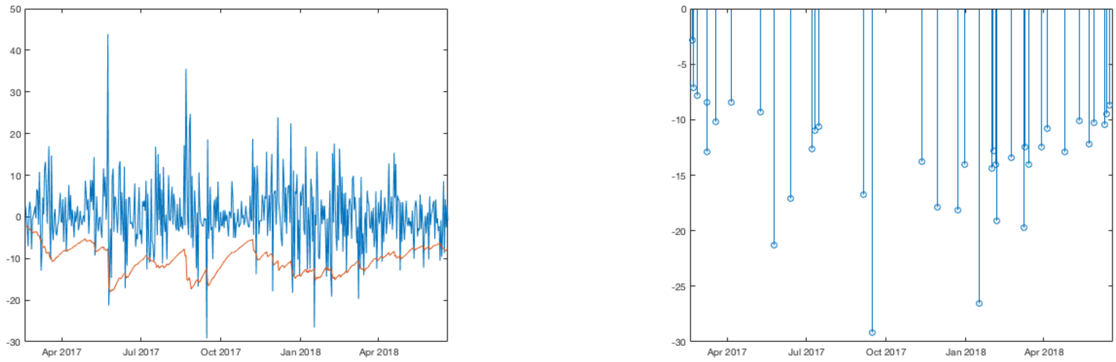
Dogecoin



Litecoin



Monero



Ripple

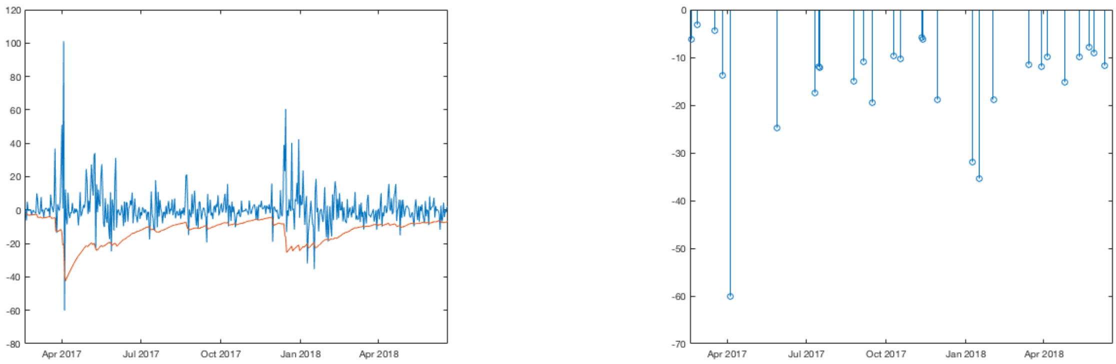
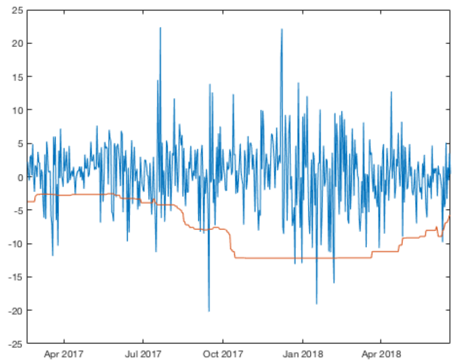
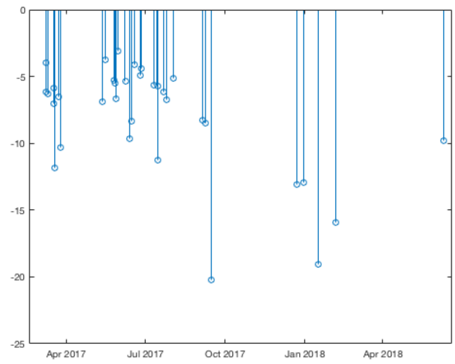


Figure 3: Time series of returns, VaRs and VaR violations for the cryptocurrencies. The left panel displays returns (blue) and one-day 5% VaRs obtained from RiskMetrics (red). The right panel displays violations. The magnitude is the return on the day of the violation.

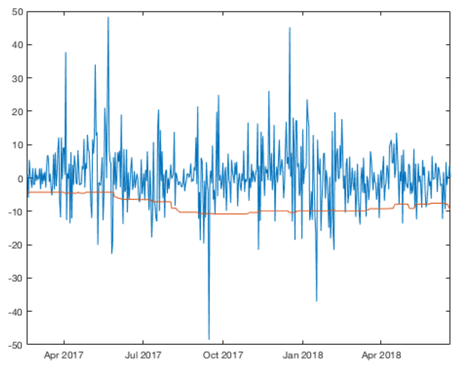
Appendix D Returns and VaRs from Filtered Historical Simulation

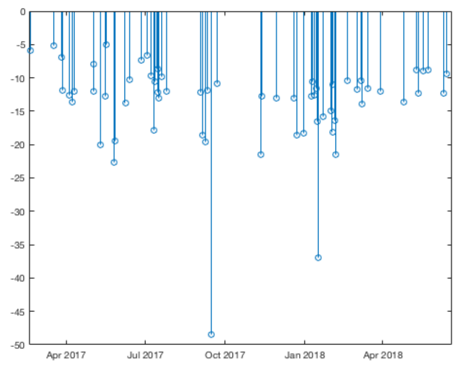
Bitcoin



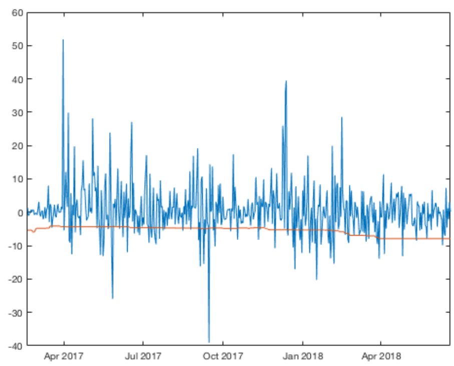


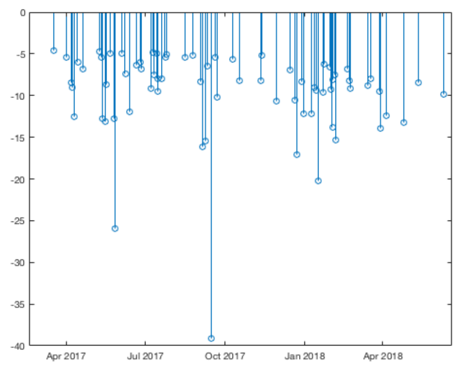
Dogecoin



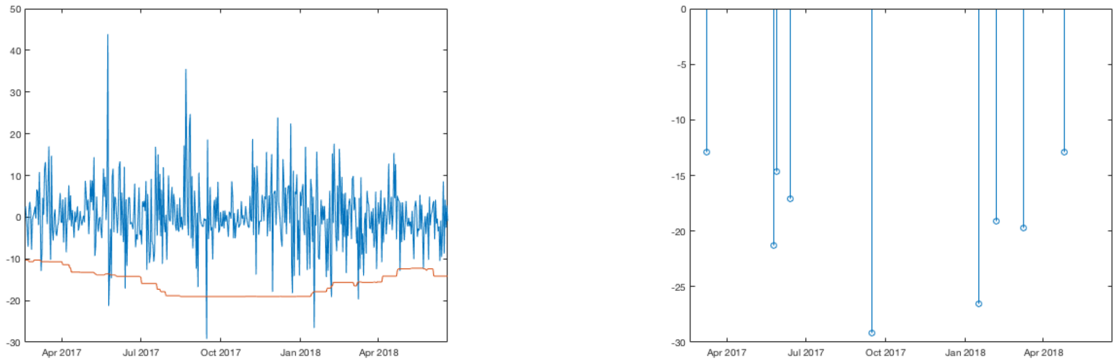


Litecoin





Monero



Ripple

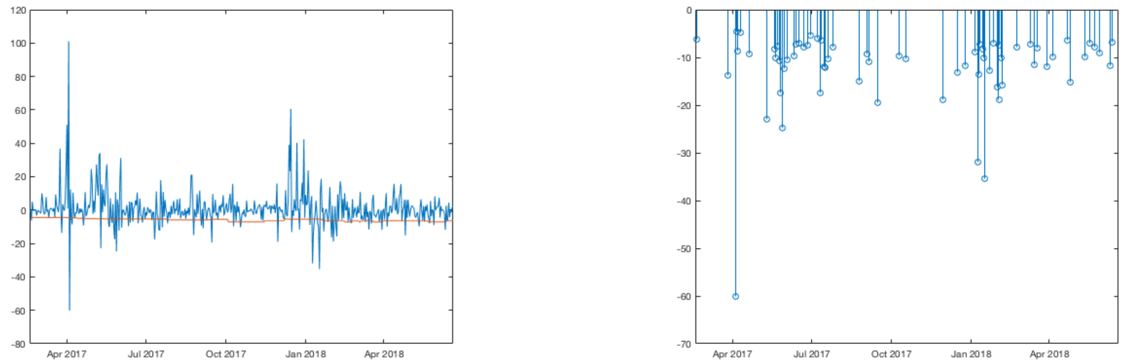


Figure 4: Time series of returns, VaRs and VaR violations for the cryptocurrencies. The left panel displays returns (blue) and one-day 5% VaRs obtained from FHS (red). The right panel displays violations. The magnitude is the return on the day of the violation.