universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master‘s Thesis

„Cointegrated portfolios and volatility modelling in the Cryptocurrency market“

verfasst von / submitted by

Stefan Gabriel, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2023 / Vienna 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 974
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Banking and Finance
Betreut von / Supervisor:	Univ.-Prof. Dipl.-Ing. Dr. Robert Kunst

Abstract

This master thesis examines two major topics in the field of Cryptocurrencies. On the one hand, I investigate possible long-run equilibrium relationships among ten different Cryptocurrencies by applying two different cointegration tests. This analysis aims to construct cointegrated portfolios that enable statistical arbitrage. Moreover, I found evidence for a connection between market volatility and the spread used for trading. The results of the trading strategies suggest that cointegrated portfolios based on the Johansen procedure generate the highest abnormal log-returns, both in-sample and out-of-sample. Furthermore, five out of six trading strategies generate a positive overall profit and outperform a passive investment approach out-of-sample. The second part of the econometric analysis builds on the granger causality results between volatility and the spread where I found evidence for granger causality. For this analysis, I implement two different types of models for forecasting the volatility of Bitcoin, where the GARCH family uses daily price data and the HAR family is based on 5-min high-frequency data. Moreover, in both categories I also use models that consider potential jumps in the price series as I found that price jumps play an important role when forecasting the volatility of Bitcoin. In addition, the findings indicate that the realized GARCH model is the only GARCH model that can compete against the HAR-RV model as it has a statistically significant smaller QLIKE when performing out-of-sample forecasting.

Keywords: cryptocurrencies, bitcoin volatility, realized variance, integrated variance, jump variation, cointegration, cointegrated portfolios, statistical arbitrage

Kurzfassung

Diese Masterarbeit untersucht zwei Hauptthemen im Bereich Kryptowährungen. Einerseits untersuche ich mögliche langfristige Gleichgewichtsbeziehungen zwischen zehn verschiedenen Kryptowährungen, indem ich zwei verschiedene Kointegrationstests anwende. Diese Analyse zielt darauf ab, kointegrierte Portfolios zu konstruieren, die statistische Arbitrage ermöglichen. Darüber hinaus fand ich Hinweise auf einen Zusammenhang zwischen der Marktvolatilität und dem für das Trading verwendeten Spread. Die Ergebnisse der Handelsstrategien deuten darauf hin, dass kointegrierte Portfolios, die auf dem Johansen-Verfahrens basieren, die höchsten anormalen Log-Renditen generieren, sowohl in-sample als auch out-of-sample. Außerdem erzielen fünf von sechs Handelsstrategien einen positiven Gesamtgewinn und übertreffen einen passiven Anlageansatz out-of-sample. Der zweite Teil der ökonometrischen Analyse baut auf den Granger-Kausalitätsergebnissen zwischen Volatilität und Spread auf, wo ich Hinweise auf Granger-Kausalität gefunden habe. Für diese Analyse implementiere ich zwei verschiedene Arten von Modellen zur Vorhersage der Volatilität von Bitcoin, wobei die GARCH-Familie tägliche Preisdaten verwendet und die HAR-Familie auf 5-Minuten-Hochfrequenzdaten basiert. Darüber hinaus verwende ich in beiden Kategorien auch Modelle, die potenzielle Sprünge in den Preisreihen berücksichtigen, da ich festgestellt habe, dass Preissprünge eine wichtige Rolle bei der Prognose der Volatilität von Bitcoin spielen. Des Weiteren weisen die Ergebnisse darauf hin, dass das realisierte GARCH-Modell das einzige GARCH-Modell ist, das mit dem HAR-RV-Modell konkurrieren kann, da es bei der Durchführung von out-of-sample Prognosen einen statistisch signifikant kleineren QLIKE aufweist.

Schlüsselwörter: Kryptowährungen, Bitcoin Volatilität, Realisierte Varianz, Integrierte Varianz, Sprungvariation, Kointegration, Kointegrierte Portfolios, Statistische Arbitrage

Table of content

List of Tables

List of Figures

1. Introduction

Cryptocurrencies, especially Bitcoin (BTC), have become very popular in the recent years. They are known to be highly volatile financial assets that carry a high risk of losing everything. Some econometricians claim that the variance in this market is infinite. However, many time series procedures are no longer useful if this is the case, so I assume still a finite variance. Nevertheless, the market has been growing enormously, and more and more different Cryptocurrencies have emerged. The total market capitalization has increased from USD 5.5 billion on 1st Jan 2015 to USD 2.2 trillion on 1st Jan 2022 which is almost a 40,000% increase (CoinMarketCap, 2022). BTC with a market cap of USD 902 billion on 1st Jan 2022 was the first Cryptocurrency on the market and was developed by the pseudonymous Satoshi Nakamoto. During recent years, the crypto market has become more and more popular for research. At the beginning of Bitcoin there was a lot of research whether Bitcoin can be considered as money. Mittal (2012) found that Bitcoin does not fulfil the classical three functions of money. Especially the store of value property is clearly not fulfilled because of the high volatility of Bitcoin. Kubát (2015) also investigated whether BTC can function as a store of value. He found that Bitcoin has a significantly higher volatility than Gold and the EUR-USD exchange rate, which is in line with the findings of Mittal (2012). The bottom line is that the volatility of Bitcoin is conspicuously higher than the price variation of many other typical assets and fiat currencies. Such highly volatile financial assets have also aroused interest of hedge funds and portfolio managers, which led to a strong incentive in modelling the volatility of BTC. Urquhart (2018) uses Google Trends data and found that the realized volatility of BTC together with the trading volume have a major impact on investor attention in the following day. Catania and Grassi (2017) use a robust score driven filter for modelling the volatility of 606 Cryptocurrencies. Their forecasting results suggest that the GAS-GHSKT model with a time-varying skewness component has the best forecasting performance.

In this master thesis, I use crypto assets to construct trading strategies that are superior to a "buy-and-hold" strategy and are also promising in an out-of-sample backtesting analysis. Profitable trading strategies that also work during a bear market and in times with high inflation rates are crucial in risk management and for hedging. Therefore, this master thesis makes a valuable contribution for banks and insurance companies since the presented statistical arbitrage strategies are based on Cryptocurrencies that share a common stochastic trend implying a long-run relationship and hence perform well over a longer time horizon. Furthermore, I examine a potential connection between market volatility and the presented trading strategies. Ni, Pan, and Poteshman (2008) found a significant link between investor's trading behavior and private volatility information in the option market. Likewise, the findings

of Omane-Adjepong, Alagidede, and Akosah (2019) suggest that three crypto markets violate the Efficient Market Hypothesis and that accurate volatility forecasting can enhance optimal portfolio hedging. The authors also emphasize the importance of taking the high volatility persistence into account. Hence, the second objective is to evaluate the forecasting performance of the volatility of BTC by estimating different types of GARCH models with different distributions and the simple Heterogeneous Autoregressive model of Realized Volatility (HAR-RV) by Corsi (2009) including several extensions. Additionally, I focus on the incorporation of price jumps as jumps can account for a substantial portion of the variation.
The goal of this analysis is to answer the question whether any GARCH model can compete with the HAR-RV type models, specifically the more recent realized GARCH model. Modelling and especially accurate forecasting of the volatility of an asset is also essential for many areas in business finance. Recently, the CME Group launched options on Micro Bitcoin futures on 28th Mar 2022 (CME Group, 2022). Therefore, the volatility modelling analysis, which is done in this thesis is a relevant topic because the volatility of an asset is an essential component for option pricing.

To compare and place this paper into the existing literature, I discuss some of the existing literature in the next section. Section 3 introduces the reader to the data by providing some descriptive statistics and presenting relevant plots. Section 4 covers the cointegration part including the construction of cointegrated portfolios that enable statistical arbitrage. In section 5, I perform the volatility forecasting analysis by estimating different GARCH models and several HAR-type models. In section 6, I conclude and summarize the most important results.

2. Literature Review

The first part of this literature review aims to discuss applications of cointegration in finance, which have become a very popular topic for researchers and econometricians. However, in this thesis I focus on exploiting cointegration for constructing mean-reverting trading strategies permitting statistical arbitrage. In contrast to many other researchers, Yan and Wong (2022) consider pairs trading from a game-theoretical perspective where the trading strategy is a Nash sub-game perfect equilibrium of a sequential game and estimate a vector error-correction model (VECM) in a continuous-time setting allowing for "delayed" cointegration in order to construct a statistical arbitrage strategy. This terminology used by the authors goes back to the construction of a discrete-time VECM. The regressors of a VECM are apart from the error-correction term lagged differences causing adjustment towards the equilibrium with a lag, thus "delayed" cointegration. In the continuous-time framework the limit of the VECM becomes a stochastic delay differential equation (SDDE) which has a unique solution using backward stochastic differential equations (BSDEs). For a robustness check, the new developed strategy is compared to a time-consistent dynamic pairs trading strategy suggested by Chiu and Wong (2015) based on the Markowitz mean–variance (MV) criterion. Following the results of Yan and Wong (2022), the non-Markovian strategy is able to outperform the MV pairs trading strategy by generating additional abnormal returns if the in-sample data suggest a high-order VAR.

The literature of identifying cointegrating relationships in the crypto market is relatively scarce compared to the stock market but due to the rapid market growth and the increasing popularity, researchers and practitioners have started investigating potential long-term relationships among cryptocurrencies. Sovbetov (2018) explores the main factors that impact the prices of five different Cryptocurrencies by estimating an error-correction model based on the more recent Autoregressive Distributed Lag (ARDL) method by Pesaran, Shin, and Smith (2001) in order to use I(0) and I(1) variables. His findings indicate that the cryptomarket beta, trading volume and volatility are the main factors that influence the market price, both in the long- and short-run. In addition, the error-correction terms are all highly significant and range from 23.68% for Bitcoin to 10.2% for Dash, suggesting a relatively fast adjustment speed. On the other hand, Adebola, Gil-Alana, and Madigu (2019) examine the long-term interdependencies of Cryptocurrencies and the gold price by testing for fractional cointegration. In their analysis, bivariate cointegrating relationships between the gold price and the price of the considered Cryptocurrencies are only rarely found. In a similar way, Tan, Huang, and Xiao (2021) test for fractional cointegration between the value at risk of altcoins and Bitcoin by estimating a more sophisticated fractionally Cointegrated Vector Autoregression (FCVAR) model. The authors split the sample into a pre-crash and post-crash period to identify potential differences. Their findings differ substantially between both periods since they only find evidence for a long-run fractional cointegration relationship between the value at risk of altcoins and Bitcoin for two out of eight Cryptocurrencies and no significant leverage effect before the crisis. However, when doing the same analysis after the crisis a positive risk-return trade-off and leverage effect appears, implying that investors demand a higher risk premium after the crash to compensate for taking the additional risk.

The second part of the literature review discusses some of the empirical work that has already been done on modelling the volatility of Cryptocurrencies. Several papers have already analyzed which model in the literature is most suitable in forecasting the volatility of a financial asset. However, the findings vary significantly across the existing literature. Cermak (2017) analyzes whether Bitcoin has the potential in replacing classical fiat currencies. Of course, this only works if there is less variation in the price. He estimates the volatility of BTC with a modified GARCH(1,1) by including relevant macroeconomic variables in the return and variance equation . He found that the level of volatility is constantly declining and expects that Bitcoin reaches the volatility level of fiat currencies in 2019-2020, which was certainly not the case. Hou et al. (2019) use a different approach by estimating a stochastic volatility with correlated jump (SVCJ) model to provide a possibility for pricing BTC options. They found that BTC is very prone to jumps and it is essential to incorporate these jumps when pricing Bitcoin derivatives. Assenting to this finding, Hung, Liu, and Yang (2020) emphasize the usage of a jump-robust realized measure when estimating the more recent realized GARCH model. Their results indicate that the realized GARCH model with a jump-robust realized measure outperforms the standard GARCH model out-of-sample. These results also provide evidence for the usefulness of intraday information based on high-frequency data. Bergsli et al. (2022) compare the forecasting performance of Bitcoin's volatility by using a number of GARCH models with different distributions and two HAR-RV models. They find that the two HAR models are superior to all other considered GARCH models. Their findings are also robust for different forecasting horizons, which suggests that HAR models, which are based on intraday data, are better in forecasting the volatility of BTC than GARCH models based on daily data. Shen, Urquhart, and Wang (2020) apply several extensions of the standard HAR-RV model estimating the so called HARQ-F-J for estimating the volatility of Bitcoin, which is a new specification of the authors based on Bollerslev, Patton, and Quaedvlieg (2016). After the analysis, the HARQ-F-J model has turned out as the best forecasting model. Furthermore, the authors found that HAR models with structural breaks are superior to HAR models without structural breaks. This result is also robust for three different forecasting horizons. Another approach that has become very favored in the last years is the usage of machine learning techniques to improve the forecasting accuracy. Bouri et al. (2021) use so-called random forests to analyze the effect of the US–China trade war when performing out-of-sample forecasts of the realized volatility of Bitcoin. Their results suggest that including a proxy for the US–China trade war based on Google-Trends improves the forecasting performance. Similarly, Aras (2021) applies a new meta-learning strategy on the basis of support vector machines (SVM) to perform volatility forecasts for BTC. He compared the forecasting performance of this new approach with 110 GARCH-type models and found that the meta-learning strategy outperforms all 110 GARCH-type models. Building on this approach, Tapia and Kristjanpoller (2022) apply more advanced machine learning techniques. The authors use long short-term memory (LSTM) networks with lagged errors as inputs based on an Asymmetric Multiplicative Error Model (AMEM). The motivation for applying such a framework is to account for the non-linear components in the volatility of BTC, which can hardly be captured with classical econometric models. Their results suggest that the LSTM-AMEM-MM model, which is a multiplicative hybrid model, outperforms all other considered models. For comparison, different econometric models, machine learning models, hybrid models and multiplicative hybrid models as well are considered. Their results are robust for different estimation window lengths and are significant when applying the Diebold–Mariano test and the model confidence set (MCS) procedure introduced by Hansen, Lunde, and Nason (2011). Moreover, the best LSTM-AMEM-MM model for example yields a 38% and 50% lower MSE than the best HAR-RV and GARCH model, respectively.

3. Data

This section presents interesting insights into the data that are used for analyzing cointegrated relationships amongst ten different Cryptocurrencies. Daily price data are downloaded from *Yahoo Finance*. The full dataset contains over two and half years of price data from 31st Dec 2019 until 31st Jul 2022. The training data that is used for estimating the cointegrating vectors and in-sample evaluation is based on the sample period from 31st Dec 2019 until 29th Apr 2022 consisting of 851 days and the test period for out-of-sample evaluation uses the last three months of the data set starting on 30th Apr 2022 until 31st Jul 2022. Table 1 presents some descriptive statistics of all used Cryptocurrencies. Here I use adjusted closing prices in levels instead of logs allowing for an easier interpretation. What stands out is the high standard deviation of all ten Cryptocurrencies. Apart from that BTC with a market capitalization of $734.59 billion is by far the dominating Cryptocurrency in the market followed by ETH with $339.51 billion. The total market capitalization of the ten considered Cryptocurrencies amounts to a total of $1.233,82B as of 29th Apr 2022, which is almost equivalent to the GDP of Spain in 2020 (Worldbank, 2022). This indicates that the crypto market is huge and continuously growing.

Table 1: Descriptive Statistics of all ten Cryptocurrencies (Training period)

Symbol	Name	Min	Max	Mean	Std	Market cap
BTC	Bitcoin	$4,970.79	$67,566.83	$30,919.50	$18,764.50	$734.59B
ETH	Ethereum	$110.61	$4,812.09	$1,740.47	$1,426.85	$339.51B
BNB	Binance Coin	$9.39	$675.68	$229.22	$213.02	$64.18B
ADA	Cardano	$0.02	$2.97	$0.83	$0.78	$27.16B
XRP	Ripple	$0.14	$1.84	$0.59	$0.38	$29.41B
DOGE	Dogecoin	$0.002	$0.68	$0.11	$0.13	$17.91B
LTC	Litecoin	$30.93	$386.45	$120.50	$69.17	$7.04B
BCH	Bitcoin Cash	$152.22	$1542.43	$422.95	$205.79	$5.61B
XLM	Stellar	$0.03	$0.73	$0.22	$0.15	$4.41B
XMR	Monero	$33.01	$483.58	$170.15	$90.31	$4.00B

Notes: Market capitalization as of 29th Apr 2022 obtained from coinmarketcap.com, "B" denoting 10^9(1 billion)

Figure 1 visualizes the daily logarithmic prices (log-prices) of the training period for all ten Cryptocurrencies.

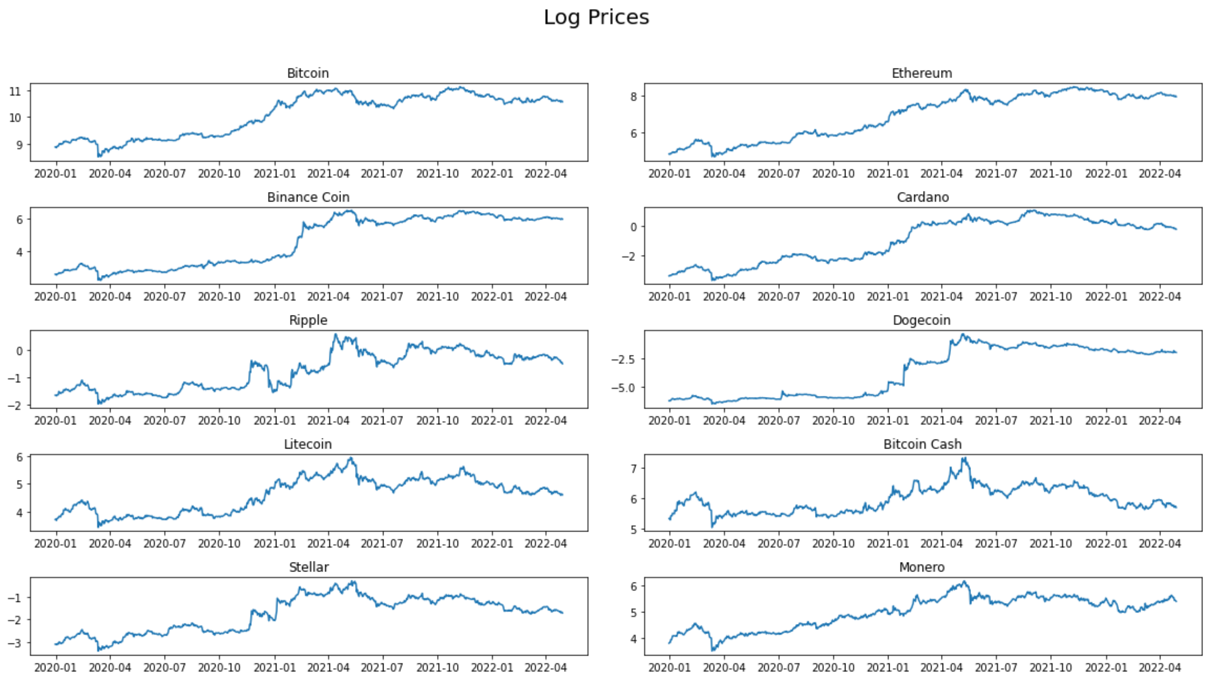


Figure 1: Log-prices (Training period)

As indicated by figure 1, all the time series follow a similar pattern suggesting the existence of a common stochastic trend and hence a possible long-term relationship between the Cryptocurrencies. To identify any potential cointegrating relationships two different cointegration tests are applied and discussed in section 5. Another interesting aspect that is worth to look into are return correlations. The level of correlation between returns of several stocks differs substantially from return correlations among Cryptocurrencies. Pollet and Wilson (2010) found an average correlation of just 23.7% between the 500 largest stock pairs in the market from 1963 to 2006. Noteworthy, figure 2 suggests that there are five crypto pairs with a return correlation of 0.8 or more, which highlights the attempt in creating mean-reverting portfolios in chapter 5. Moreover, I find that the average correlation of the log-returns is in the upper range at 0.634. It is also striking that the log-returns of Dogecoin pairs have the lowest correlation with less than 50%.

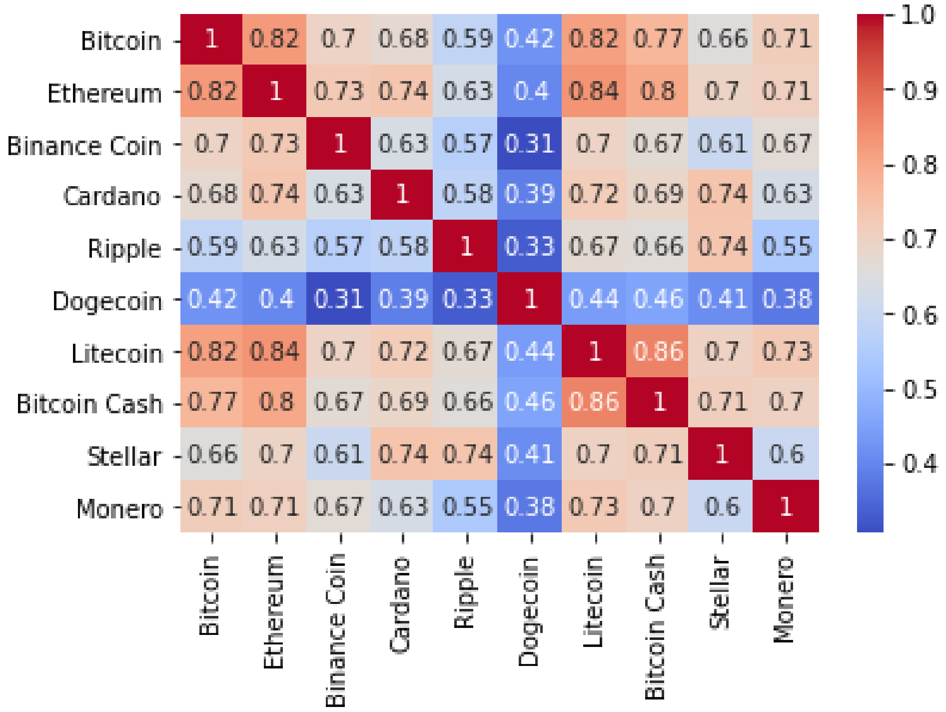
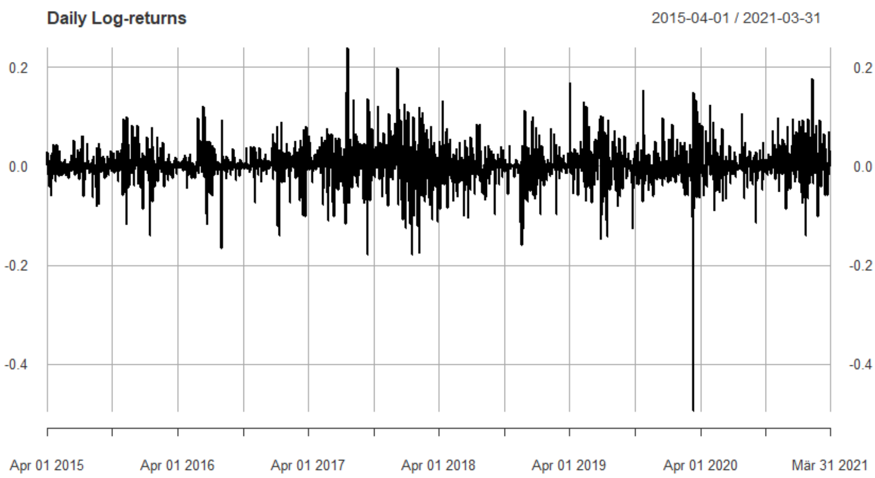


Figure 2: Correlation matrix (log-returns)

Furthermore, in the following the data that are used for modelling the volatility of BTC are introduced. Daily and intraday price data are downloaded from the cryptocurrency exchange *Bitstamp*. The full dataset for the volatility modeling part (chapter 5) contains six years of price data (1st Apr 2015 – 31st Mar 2021). The training data that is used for estimating the models and in-sample evaluation is based on the sample period from 1st Apr 2015 until 31st Mar 2020 and the test period for out-of-sample evaluation uses the last year of the data set starting on 1st Apr 2020 until 31st Mar 2021 (COVID-19 crisis period). The next plot visualizes the daily log-returns of BTC and the corresponding descriptive statistics for the whole sample.



	BTC Returns
count	2192.000000
mean	0.002502
std	0.039411
min	-0.493969
25%	-0.011496
50%	0.001997
75%	0.018233
max	0.238396

Figure 3: Daily Log-returns

Figure 3 suggests that there is a lot of variation in the returns, which provides evidence for volatility clustering. The high number of extreme values is also striking. While the maximum daily log-return is 0.24 the minimum is -0.49, which indicates that huge negative returns are more severe. The next plot shows a normal Q-Q-Plot for looking whether the returns are approximately normally distributed, which is hardly the case for asset returns though.

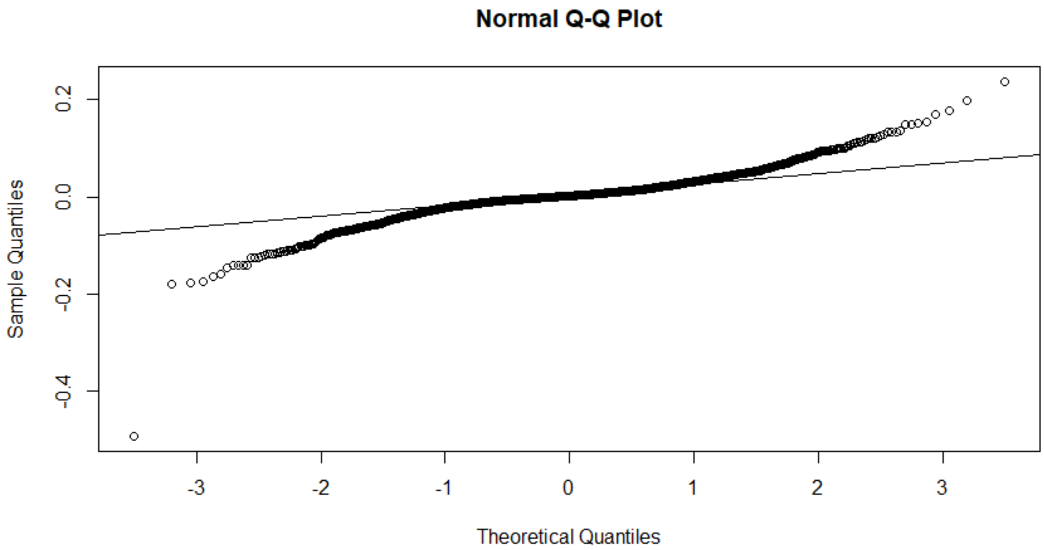


Figure 4: Q-Q Plot of daily log-returns

As can be seen by the plot the daily log-returns clearly do not follow a normal distribution since there is a lot of weight in the tails. These fat tails indicated by an excess kurtosis are a typical observation in empirical finance. This stylized fact about the returns could speak for the usage of a heavy tail distribution like the Student’s t-distribution when estimating the GARCH models. In the following, the realized variance based on 5-min high-frequency data and its stylized facts are analyzed.

Table 2: Descriptive Statistics of the Realized Variance

	Full sample	**Training period**	**Test period**
count	2192	1827	365
mean	0.0021	0.0022	0.0017
std	0.0043	0.0046	0.0026
min	0.00003	0.00003	0.00003
25%	0.0005	0.0005	0.0005
50%	0.0009	0.0009	0.0009
75%	0.0021	0.0022	0.0018
max	0.1066	0.1066	0.0297
skewness	10.4957	10.2992	5.4818
kurtosis	189.8533	177.8116	48.0258
Jarque-Bera (p-value)	<0.0001	<0.0001	<0.0001

Table 2 provides some descriptive statistics of the realized variance of BTC subdivided into the full sample, the training period and the test sample. The data suggest that the mean and the median of all subsamples are more or less equal. Moreover, not very surprisingly the realized variance for any subgroup is non-Gaussian, which is visible through the excess kurtosis (significantly greater than 3 for all subsamples) and the positive skewness. Taking also the very low p-value of the Jarque-Bera test into account one can conclude that the realized variance of Bitcoin follows a leptokurtic and right-skewed distribution.

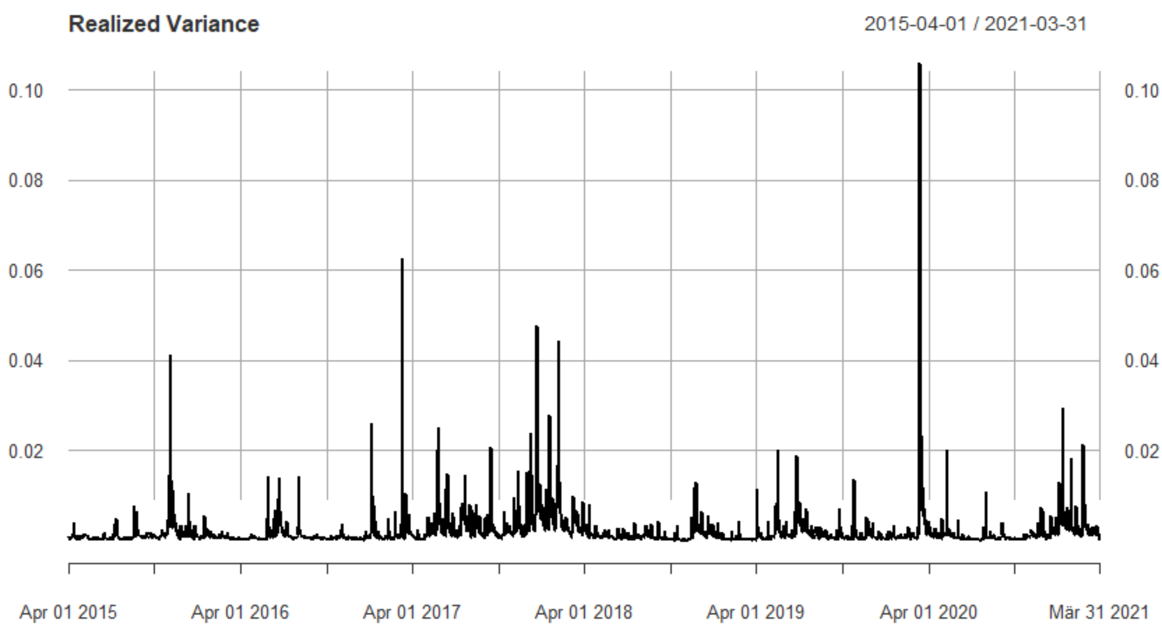


Figure 5: Realized Variance of BTC

Figure 5 indicates that there are huge peaks occasionally. These extreme values occur in particular at the beginning of the COVID-19 crisis, which led to considerable fear in the market and hence much volatility. An additional explanation for these extreme values could be the presence of price jumps so it makes sense to have a look at the jump variation too. The discontinuous variation of the realized variance can be estimated with any estimator that consistently estimates the integrated variance of the quadratic variation in the presence of jumps. In this master thesis, I use the MedRV estimator that can be calculated with equation (26). To ensure the positivity of the jump variation I define it as $JV = max\{RV - MedRV, 0\}$. More details of this estimator and the quadratic variation with its components are described in chapter 5.1.

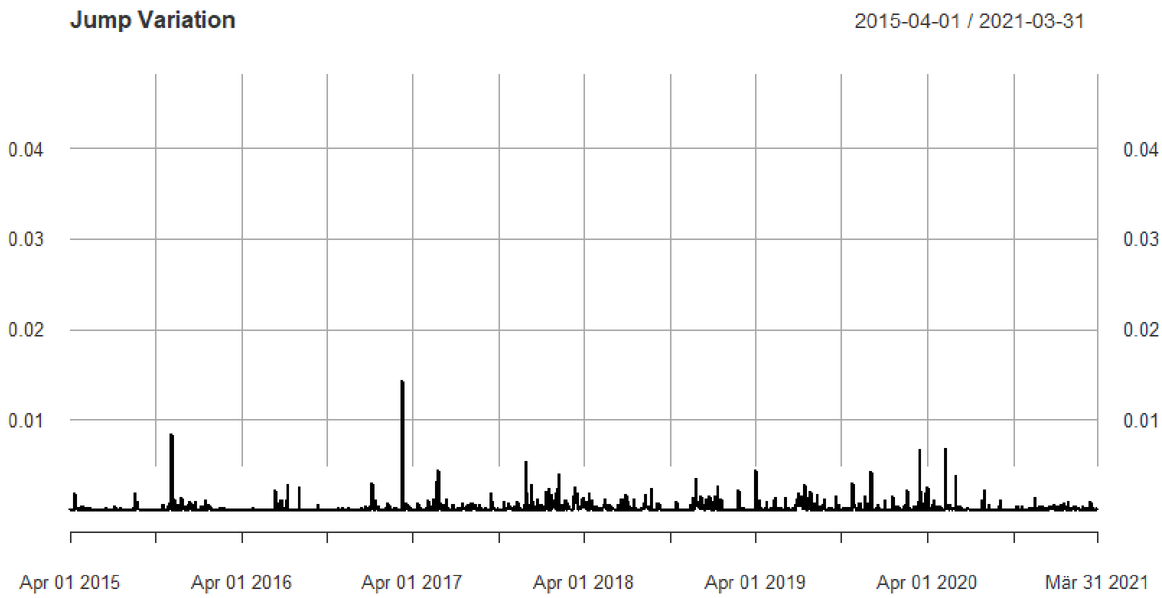


Figure 6: Jump Variation of BTC

Figure 6 implies that there are indeed many days with a relatively high jump variation. In order to detect significant jumps in the price process of Bitcoin the JO Jump test by Jiang and Oomen (2008) is performed. Further details of the test are provided in chapter 5.1.2.

Lastly, I want to investigate whether the realized variance of BTC follows a log-normal distribution like it is the case with many other assets. The next three plots present the Q-Q-Plot and the corresponding density of the logarithmic realized variance of the full sample, of the first three years of the sample (1st Apr 2015 – 31st Mar 2018) and of the last three years (1st Apr 2018 – 31st Mar 2021), respectively.

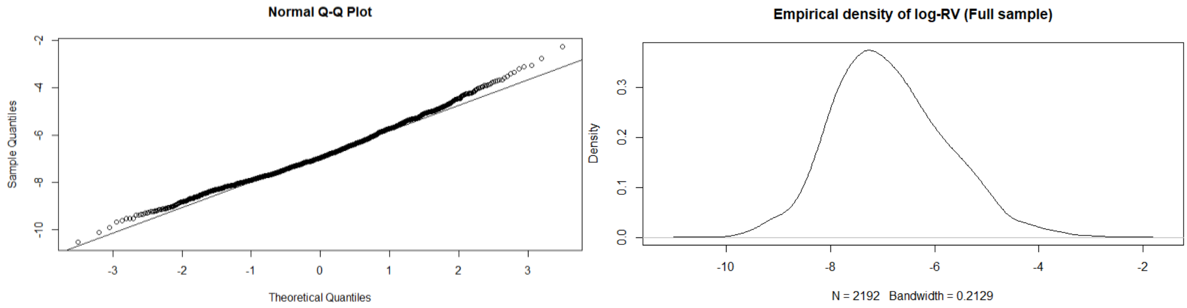


Figure 7: Q-Q Plot and empirical density of log-RV (2015-2021)

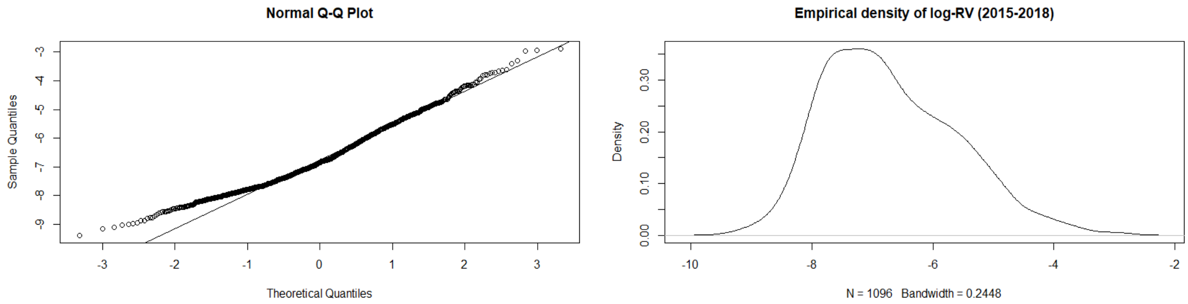


Figure 8: Q-Q Plot and empirical density of log-RV (2015-2018)

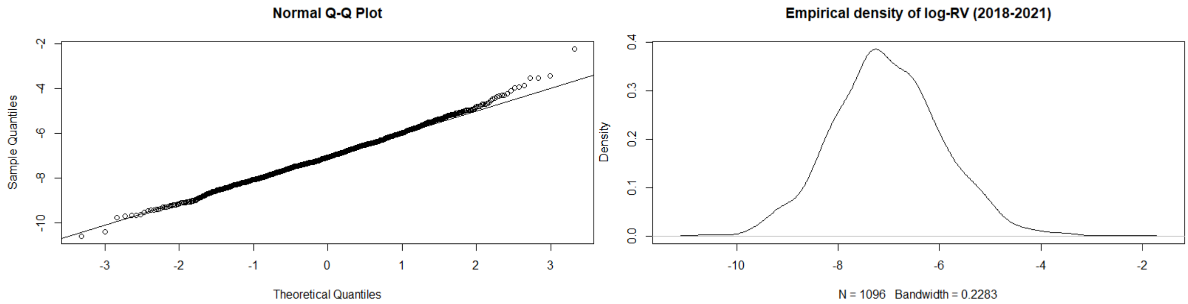


Figure 9: Q-Q Plot and empirical density of log-RV (2018-2021)

Figure 7 insinuates that the logarithm of the realized variance follows a Gaussian distribution approximately. However, there is still too much weight in the tails. In figure 8, the evidence against the log-normality property is even stronger. There is excessively more weight in the tails than it should be the case if it were Gaussian and the density looks relatively right-skewed. In contrast to figure 8, the log-normality property of the realized variance holds much better in figure 9. The density looks more symmetric and the tails are much thinner than it is the case in figure 8. Taking these things together, the results suggest that there are more extreme values of the realized variance in the first three years than in last three years, which implies that the realized variance is time-varying. This is also in line with figure 5 because the realized variance in the second sub-period contains less outliers except at the beginning of the COVID-19 crisis.

4. Cointegration

In time series analysis there is a distinction between stationary and non-stationary variables. In the literature stationary invertible processes are also called I(0). Non-stationary variables are often “integrated of order d” or abbreviated I(d), which means that after taking the difference d times the variables becomes covariance stationary or I(0). However, most financial macroeconomic variables, which are increasing over time, like GDP, consumption expenditure and almost all types of financial assets are first-order integrated or I(1) when taking the logarithm. The reason for this is that taking the logarithm and the first difference yields the growth rate, which is mostly stationary. A big problem that frequently arises in this context is that regressing a I(1) variable on another I(1) variable often provides an indication of a relationship that does not exist. This is a so-called spurious regression, which is visible through excessive significance of the coefficients, a relatively high R^2 and strong autocorrelation in the residuals. The residuals in this case are also I(1). The concept of spurious regression was intensively studied in the 1970s when Granger and Newbold (1974) showed that regression results in numerous high-quality published papers and several textbooks were subject to a very high R^2 and strong autocorrelation in the residuals, indicated by a very small Durbin-Watson statistic. As pointed out by the authors autocorrelated residuals are problematic since the estimates of the parameters are inefficient and the usual t-distribution for the coefficients is invalid. However, if regressing a I(1) variable on anther I(1) variable yields covariance stationary residuals and thus a covariance stationary linear combination then there is evidence of a long-run relationships between both variables (cointegrating regression). In this case the residuals are I(0) but usually not white noise and strongly autocorrelated. Granger (1981) developed the concept of cointegration, which describes a dynamic long-run equilibrium relationship between two I(1) variables. A few years later Engle and Granger (1987) proposed the first approach to test for the existence of a cointegrated relationship by applying the so-called Engle-Granger two-step procedure (EG-2) and established the term “Cointegration”. More details of this approach are described in the next chapter.

In this master thesis, I focus on the usage of cointegration in finance since cointegration corresponds to a useful property that can be exploited in finance when establishing a trading strategy. If two (or more) financial variables like stock prices are cointegrated, it is possible to identify common stochastic trends and investigate the long-run relationship between the variables. Alexander (1999) showed how cointegration is used for many financial applications. In risk management for example, cointegration can be used to develop hedging strategies that also perform well in the long run. Another application would be spot-futures arbitrage, which is possible if there exists a system of financial asset prices with a mean-reverting spread (Alexander, 1999). In this paper, I also present a simple trading strategy that exploits cointegrating relationships among Cryptocurrencies, allowing for statistical arbitrage. Before the methodology of the trading strategy is illustrated, as a next step two different tests for identifying potential cointegrating relationships are described.

4.1. Engle-Granger two-step procedure

The Engle-Granger two-step procedure introduced by Engle and Granger (1987) was the first approach to test whether two I(1) time series are cointegrated. The methodology of Engle-Granger cointegration test works as follows. Consider two first-order integrated variables X and Y.

1. Estimate the simple cointegrating regression $Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$ using OLS to get estimates of $\hat{\beta}_0,\ \hat{\beta}_1\ and\ \hat{\varepsilon}_t$
2. Run a unit-root test on the residuals $\hat{\varepsilon}_t$ and check whether $\hat{\varepsilon}_t \sim I(0)$

The commonly used unit-root test is the (augmented) Dickey-Fuller test by Dickey and Fuller (1979), which is also used in this paper. The augmented Dickey-Fuller test has the same properties as the simple Dickey-Fuller test. The null hypothesis of the Dickey-Fuller test is the existence of a unit root. Two different variants can be distinguished. The DF_μ test considers the model

$$X_t = \delta + \phi X_{t-1} + \varepsilon_t \tag{1}$$

and tests for $\phi = 1$. Its null is a random walk and assumes (without testing) that $\delta = 0$. Its alternative is the stable AR(1) model. This version is used for data without a deterministic trend.

The DF_τ test considers the model

$$X_t = \delta + \gamma t + \phi X_{t-1} + \varepsilon_t \tag{2}$$

and also tests for $\phi = 1$. Its null is a random walk with drift and assumes (without testing) that $\gamma = 0$. Its alternative is the trend-stationary model. This version is used for trending data.
The test statistic for both variants is given by

$$DF_\mu = DF_\tau = \frac{\hat{\phi} - 1}{s.e.(\hat{\phi})} \tag{3}$$

If the data generating process is AR(p) rather than AR(1) for $p \geq 2$ the model can be rewritten as

$$\Delta X_t = \delta + \pi X_{t-1} + c_1 \Delta X_{t-1} + \cdots + c_{p-1} \Delta X_{t-p+1} + \varepsilon_t \tag{4}$$

Equation 4 represents the test regression of the "augmented" Dickey-Fuller test to obtain the DF_μ test statistic. If there is a unit root, $\phi = 1$ in which case $\pi = 0$. Thus, testing for $\pi = 0$ makes sense. The (augmented) test has the same null distribution and other properties as the DF test in the AR(1) model. The lag order is chosen by minimizing the AIC.

If the data are trending and the data generating process is AR(p) for $p \geq 2$ the following test regression is used:

$$\Delta X_t = \delta + \gamma t + \pi X_{t-1} + c_1 \Delta X_{t-1} + \cdots + c_{p-1} \Delta X_{t-p+1} + \varepsilon_t \tag{5}$$

The t-value of π in equation (5) can then be used as DF_τ statistic.

Taking all these together the test regression of the unit-root test for the EG-2 is described by

$$\Delta \hat{\varepsilon}_t = \pi \hat{\varepsilon}_{t-1} + c_1 \Delta \hat{\varepsilon}_{t-1} + \cdots + c_{p-1} \Delta \hat{\varepsilon}_{t-p+1} + v_t \tag{6}$$

The reason for the difference to equation (4) and (5) is that the residuals of the cointegrating regression $\hat{\varepsilon}_t$ are by construction imposed to have a mean of zero and hence $\delta = 0$. There should also be no time trend because the residuals from a regression that includes a linear trend do not have a trend by construction, which implies $\gamma = 0$. However, as pointed out by Phillips and Ouliaris (1990) for this test, the original DF significance points are invalid and it is necessary to use correct significance points specified in Phillips and Ouliaris (1990). The correct significance points for one explanatory variable are presented in the following table:

Table 3: Significance Points EG-2 (N=1)

Significance level	10%	5%	1%
Asymptotic critical values	-3.07	-3.37	-3.96

Even though Clive Granger received the Nobel Prize for his paper, it has turned out that the EG–2 is inefficient and is not really used anymore in high-quality empirical works nowadays. The reason is that the Engle-Granger two-step procedure is not reliable for more than two I(1) variables and even if the residuals of the cointegrating regression are covariance stationary the estimator of the cointegrating vector $\hat{\beta}_1$ is inefficient. Another cointegration test that provides efficient estimates and can handle I(0) and I(1) variables and multiple cointegrating vectors is presented in the next section.

4.2. Johansen procedure

The Johansen cointegration test considers a multivariate VAR(p) rather than a univariate OLS regression. More specifically a VAR(p) containing an additional error-correction term (ECT) $\beta' X_{t-1}$, which is called a vector error-correction model (VECM). The VECM has a similar form like equation (4):

$$\Delta X_t = \delta + \alpha\beta' X_{t-1} + \Gamma_1 \Delta X_{t-1} + \cdots + \Gamma_p \Delta X_{t-p+1} + \varepsilon_t, \tag{7}$$

where $\alpha\beta' X_{t-1}$ is stationary and $\alpha\beta'$ is called the impact matrix Π which is an n x n-matrix with rank r. Here three different cases can be distinguished:

1. If r = 0 then Π is a zero matrix and there is no cointegration in the system which implies that the VAR(p) model is really a VAR(p-1) model for differences ΔX

2. If Π has full rank (r = n) and is non-singular then X is already stable

3. If the rank of Π is $0 < \mathrm{r} < \mathrm{n}$ then there are r linear independent cointegrating vectors $\beta_j, j = 1, \dots r,$ such that $\beta_j' X$ is stationary.

Equation (7) is balanced because all ΔX_{t-j} are stationary, and also $\beta' X_{t-1}$ is stationary. β contains dynamic equilibrium conditions, and the loading matrix α describes how the components of ΔX react to deviations from these conditions.
To identify the rank r of the impact matrix Π and hence the number cointegrating vectors the trace test of the Johansen procedure is used. The trace statistic for any r_0 with $0 \leq r_0 < n$ obtained by Johansen (1995) is defined by

$$\lambda_{tr}(r_0) = -T \sum_{j=r_0+1}^{n} \log\left(1 - \lambda_j\right), \tag{8}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ are eigenvalues of a canonical correlation eigenvalue problem. For further details see Johansen (1995). The following four steps summarize the Johansen procedure:
Let $X_t = (x_{1,t}, x_{2,t}, \dots, x_{N,t})'$ denote a (N×1) vector of N time series.

1. Run a unit-root test on all N components of X_t. I(0) and I(1) time series can be kept. I(d) for $d \geq 2$ variables must be dropped or differenced
2. Determine the VAR lag order p for X_t by multivariate versions of information criteria
3. Choose a significance level and run a sequence of trace tests for r_0 with $0 \leq r_0 < n$ to determine the smallest r_0 for which the trace test does not reject
4. Estimate the error-correction model given $r = r_0$. This is a so-called reduced rank regression and requires solving the canonical correlation eigenvalue problem. More details are provided in Johansen (1988; 1991; 1995)

4.3. Trading the spread

In this section, I shortly describe the implementation of trading strategies exploiting cointegration relationships among Cryptocurrencies. The presented strategies are very similar to a market neutral pairs trading strategy where the spread of two cointegrated stocks is traded and a trader short sells the overvalued stock and takes a long position in the undervalued stock. Pairs trading and my trading strategies as well of course only work if the spread is mean-reverting such that it is necessary to have a long run equilibrium. Gatev, Goetzmann, and Rouwenhorst (2006) adopted such an investment strategy for different stock pairs. Their results suggest that pairs trading yields average annualized excess returns of approximately 11 percent. Likewise, Tokat and Hayrullahoğlu (2022) apply pairs trading to a portfolio of 45 pairs. They find an average annual return of 15% with an average Sharpe ratio of 1.43 after considering transaction costs. The trading design of the investment strategy that is used in this master thesis works as follows. In line with Leung and Nguyen (2018) I also create a cointegrated portfolio of different Cryptocurrencies but in contrast to Leung and Nguyen (2018) I use ten Cryptocurrencies instead of just four. The first step is to find potential cointegrating vectors that guarantee a mean-reverting spread. For this purpose, the Johansen procedure and the Engle-Granger two-step procedure introduced previously are used. At first for each BTC pair and after that, I construct a portfolio of all ten Cryptocurrencies and apply both procedures. Even though the EG-2 is inefficient and has problems if there is more than one cointegrating vector I still use the EG-2 for two reasons. On the one hand, the trading strategy just requires a mean-reverting spread. If this is the case, the possible long-run relationship found by the EG-2 need not be true. Nevertheless, MacKinnon (2010) provides critical values for the EG-2 with more than five explanatory variables. The asymptotic critical values for a regression with nine regressors are stated in table 4.

Table 4: Significance Points EG-2 (N=9)

Significance level	10%	5%	1%
Asymptotic critical values	-5.19	-5.47	-6.00

On the other hand, I want to compare the performance between an investment strategy based on the EG-2 and the Johansen procedure, respectively. Alexander (1999) emphasized the importance of a thorough out-of-sample performance evaluation when applying statistical arbitrage investment strategies. Therefore, I split my sample into a training period and a test period. After finding potential cointegrating vectors, two (or more) different spreads are obtained depending on the rank of the impact matrix. $Spread_t^{EG}$ denotes the spread of the EG-2 and $Spread_t^{J,1}$ naming the spread as a result of the Johansen procedure for the first cointegrating vector, $Spread_t^{J,2}$ is based on the second vector and so on. For identifying statistically significant cointegrating vectors, I choose the 5% significance level. The next step is to find the optimal threshold in order to determine at what level to buy or sell the spread. There exists a lot of research how to find the optimal threshold for maximizing the overall profit. For a discussion I refer to Song and Zhang (2013), Ngo and Pham (2016) and Liu, Wu, and Zhang (2020) for instance.

In general, the trading strategy for in-sample evaluation in this thesis works as follows:

- Long the spread if $Spread_t^i \leq \mu - \tau \; for \; i \in \{EG, J1, J2, \ldots\}$
- Unwind a long position at the first date when $Spread_t^i \geq \mu \;\; for \; i \in \{EG, J1, J2, \ldots\}$
- Short the spread if $Spread_t^i \geq \mu + \tau \; for \; i \in \{EG, J1, J2, \ldots\}$
- Unwind a short position at the first date when $Spread_t^i \leq \mu \; for \; i \in \{EG, J1, J2, \ldots\}$
- If there is any open position until the end, the position is closed at the last trading day,

where μ is the mean of the spread and τ the threshold. “Long the spread” has the same meaning as buying one unit of the spread, which requires purchasing every Cryptocurrency with a positive sign and short selling the Cryptocurrencies with a negative sign as indicated by the corresponding cointegrating vector. In order to unwind a long position, it is necessary to sell the Cryptocurrencies with a positive sign and buy back all Cryptocurrencies with a negative sign. “Short the spread” just means short selling one unit of the spread where one need to short sell the Cryptocurrencies with a positive sign and buy the Cryptocurrencies with a negative sign. Unwinding a short position requires buying back the short Cryptocurrency with a positive sign and selling the Cryptocurrencies with a negative sign. It should be noted that short selling only happens if one long or short the spread, as defined before, but not when the position is closed since the Cryptocurrencies are already in the inventory. Furthermore, in this paper I use two different thresholds for the in-sample trading strategy, i.e. $\tau \in \{\sigma, \tau^*\}$, where σ

denotes the standard deviation of the spread and τ^* the optimal threshold based on a parametric approach. Generally, the total profit is simply the number of trades times the profit of each trade. The number of trades can be estimated if the distribution of $Spread_t^i \; for \; i \in \{EG, J1, J2, ...\}$ is known. For the case of $Spread_t^{EG}$ this should be a normal distribution since the spread are the residuals of an OLS regression which mostly approximately Gaussian. Nevertheless, the distribution of $Spread_t^{J,1}, Spread_t^{J,2}, ...$ is unknown but can be estimated by minimizing a certain loss function for example. For estimating the best fitting distribution and the corresponding optimal parameters, I use the "Fitter" package in Python, which minimizes the sum of the square errors between the data and the fitted distribution. After this analysis, it is possible to estimate the number of trades. Let $\Psi^i(.)$ denote the estimated CDF of $Spread_t^i \; for \; i \in \{EG, J1, J2, ...\}$. Then the number of trades is roughly given by

$$N * \left[Prob\left(Spread_t^i < \mu - \tau\right) + Prob\left(Spread_t^i > \mu + \tau\right)\right] =$$
$$851 * \left[\, \Psi^i(\tau) + \left(1 - \; \Psi^i(\tau)\right)\right] \; for \; i \in \{EG, J1, J2, ...\}, \tag{9}$$

where N is the number of trading days. Furthermore, the profit of each trade is approximately τ. As a result, the optimal threshold solves a simple maximization problem, formally

$$\tau^* = \arg\max_\tau \left\{851 * \tau * \left[\Psi^i(\tau) + \left(1 - \Psi^i(\tau)\right)\right]\right\} \; for \; i \in \{EG, J1, J2, ...\}. \tag{10}$$

One shortcoming of this approach is that (9) does not exactly describe the number of profit realizations since a profit is only realized if the spread crosses the mean after it reaches the upper or lower threshold. Nevertheless, if the spread is mean-reverting due to cointegrated time series it will always converge to equilibrium after it deviates from the mean by $|\tau|$. Therefore, (10) holds approximately for calculating the optimal threshold τ^*. For simplicity I also assume that it is only possible to long or short one unit of the spread. Finally, I compare the trading strategies with both thresholds when investing \$1,000 at the beginning with a passive trading strategy where \$1,000 are equally invested in all ten Cryptocurrencies (i.e. \$100 per Cryptocurrency) at the start then sold at the end of the period.

As already mentioned, it is crucial that the trading strategies are subject to an out-of-sample backtesting analysis. For backtesting, I use a time interval of three months where I introduce two different trading strategies. For both strategies I use the same cointegrating vectors estimated from the training data. If there really exists a long-term relationship I expect that the ECT is also stationary when using the same vector but the log-prices of the test period. Since the future and hence the price data of the test period are assumed to be unknown it is not possible to calculate the mean and the standard deviation of the test data. To overcome this issue, I employ a rolling window approach with two different window sizes: A long window with the same length as the test period, which reacts slowly to a sudden significant fall or increase of the spread, and a short window with a length of ten days, which is able to adjust fast to abrupt behavior of the spread. To accomplish this, I extend the out-of-sample data at the start with the last three months (ten days) of the training data for calculating the 90-day (10-day) moving average and the 90-day (10-day) rolling standard deviation. This approach allows for calculating the moving average and the rolling upper and lower threshold for each single day in the test period as time goes by. A difference to the in-sample evaluation is that the mean and both thresholds are not constant anymore but time-varying. This means in mathematical terms $\tau \in \{\sigma_t(10), \sigma_t(90)\}$ since with this backtesting approach it is not possible to estimate the distribution of the test sample like it is done in-sample. Again, all out-of-sample trading strategies are compared to a passive investment strategy in the same way as described for the in-sample evaluation.

4.4. Empirical results

4.4.1. Unit-root test results

Table 5: Dickey-Fuller test results (X_t)

Variable	Test statistic	Type	Critical value 1%	Critical value 5%
BTC	-1.416	DF_τ	-3.970	-3.416
ETH	-1.2534	DF_τ	-3.970	-3.416
BNB	-1.5050	DF_τ	-3.970	-3.416
ADA	-0.4915	DF_τ	-3.970	-3.416
XRP	-2.2439	DF_τ	-3.970	-3.416
DOGE	-1.1167	DF_τ	-3.970	-3.416
LTC	-1.5778	DF_τ	-3.970	-3.416
BCH	-1.4857	DF_τ	-3.970	-3.416
XLM	-1.3145	DF_τ	-3.970	-3.416
XMR	-2.1101	DF_τ	-3.970	-3.416

The results of table 5 indicate that no variable is I(0) implying that each time series contains a unit root. Furthermore, I use the DF_τ for all variables as I assume that all Cryptocurrencies are trending. Since there are no stationary variables it is crucial that all variables are I(1) to be able to apply both cointegration tests. Table 6 shows the Dickey-Fuller test results of the first differences of the non-stationary variables. As hoped, the null hypothesis is now rejected every time at the 1% significance level. This means that the first difference of the non-stationary variables has no unit root, which implies that they are integrated of order one. Here the DF_μ is used since the growth rate should be non-trending.

Table 6: Dickey-Fuller test results (ΔX_t)

Variable	Test statistic	Type	Critical value 1%	Critical value 5%
BTC	-13.5199	DF_μ	-3.438	-2.865
ETH	-8.6162	DF_μ	-3.438	-2.865
BNB	-7.4993	DF_μ	-3.438	-2.865
ADA	-13.1993	DF_μ	-3.438	-2.865
XRP	-29.9345	DF_μ	-3.438	-2.865
DOGE	-15.5711	DF_μ	-3.438	-2.865
LTC	-10.8248	DF_μ	-3.438	-2.865
BCH	-7.8711	DF_μ	-3.438	-2.865
XLM	-30.4498	DF_μ	-3.438	-2.865
XMR	-13.248	DF_μ	-3.438	-2.865

4.4.2. Engle-Granger two-step procedure results

Table 7: EG-2 Unit root test results (pairs)

Pair	Test statistic	Critical value 1%	Critical value 5%
BTC-ETH	-1.7321	-3.96	-3.37
BTC-BNB	-1.9864	-3.96	-3.37
BTC-ADA	-2.0167	-3.96	-3.37
BTC-XRP	-2.8475	-3.96	-3.37
BTC-DOGE	-1.9706	-3.96	-3.37
BTC-LTC	-2.1886	-3.96	-3.37
BTC-BCH	-1.5833	-3.96	-3.37
BTC-XLM	-2.9836	-3.96	-3.37
BTC-XMR	-2.834	-3.96	-3.37

The results of the EG-2 are very surprising in a sense that no pair is cointegrated at a reasonable significance level. However, this does not mean that a portfolio containing all Cryptocurrencies has no long-run relationship. Therefore, for creating such a portfolio with all ten Cryptocurrencies I run the following regression:

$$\mathrm{BTC}_t = \beta_0 + \beta_1 \mathrm{ETH}_t + \beta_2 \mathrm{BNB}_t + \beta_3 \mathrm{ADA}_t + \beta_4 \mathrm{XRP}_t + \beta_5 \mathrm{DOGE}_t + \beta_6 \mathrm{LTC}_t + \beta_7 \mathrm{BCH}_t + \beta_8 \mathrm{XLM}_t + \beta_9 \mathrm{XMR}_t + \varepsilon_t \tag{11}$$

The regression results are presented below:

Table 8: EG-2 OLS Regression

OLS Regression Results

Dep. Variable:	BTC-USD	R-squared:	0.991
Model:	OLS	Adj. R-squared:	0.991
Method:	Least Squares	F-statistic:	1.021e+04
Date:	Sun, 28 Aug 2022	Prob (F-statistic):	0.00
Time:	12:14:45	Log-Likelihood:	1016.7
No. Observations:	851	AIC:	-2013.
Df Residuals:	841	BIC:	-1966.
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	6.8297	0.190	35.927	0.000	6.457	7.203
ETH-USD	0.2034	0.016	12.581	0.000	0.172	0.235
BNB-USD	0.1130	0.012	9.277	0.000	0.089	0.137
ADA-USD	0.0684	0.013	5.160	0.000	0.042	0.094
XRP-USD	-0.1231	0.013	-9.285	0.000	-0.149	-0.097
DOGE-USD	-0.0667	0.008	-8.431	0.000	-0.082	-0.051
LTC-USD	0.8213	0.029	28.488	0.000	0.765	0.878
BCH-USD	-0.5001	0.026	-19.410	0.000	-0.551	-0.450
XLM-USD	0.1459	0.018	7.917	0.000	0.110	0.182
XMR-USD	0.1016	0.019	5.450	0.000	0.065	0.138

Omnibus:	0.734	Durbin-Watson:	0.150
Prob(Omnibus):	0.693	Jarque-Bera (JB):	0.785
Skew:	-0.069	Prob(JB):	0.675
Kurtosis:	2.943	Cond. No.	990.

Using the results of Table 8 enables calculating the spread that can be used for statistical arbitrage. The spread is then given by

$$Spread_t^{EG} = \mathrm{BTC}_t - 6.8297 - 0.2034\,\mathrm{ETH}_t - 0.1130\,\mathrm{BNB}_t - 0.0684\,\mathrm{ADA}_t + 0.1231\,\mathrm{XRP}_t + 0.0667\,\mathrm{DOGE}_t - 0.8213\,\mathrm{LTC}_t + 0.5001\,\mathrm{BCH}_t - 0.1459\,\mathrm{XLM}_t - 0.1016\,\mathrm{XMR}_t \tag{12}$$

If there is cointegration in the system, the spread as defined in equation (12) should be stationary. Applying the Dickey-Fuller test yields

Table 9: EG-2 Unit root test results (portfolio)

Variable	Test statistic	Critical value 1%	Critical value 5%
$Spread_t^{EG}$	-5.7634	-6.00	-5.47

The findings in table 9 are quite interesting. Even if the EG-2 does not find any cointegrating relationship for any Bitcoin pair the EG-2 suggest that the portfolio containing all ten Cryptocurrencies is cointegrated at the 5% significance level. However, this outcome should be taken with caution since the EG-2 has several problems with more than two I(1) variables. The next plot shows the time series of $Spread_t^{EG}$.

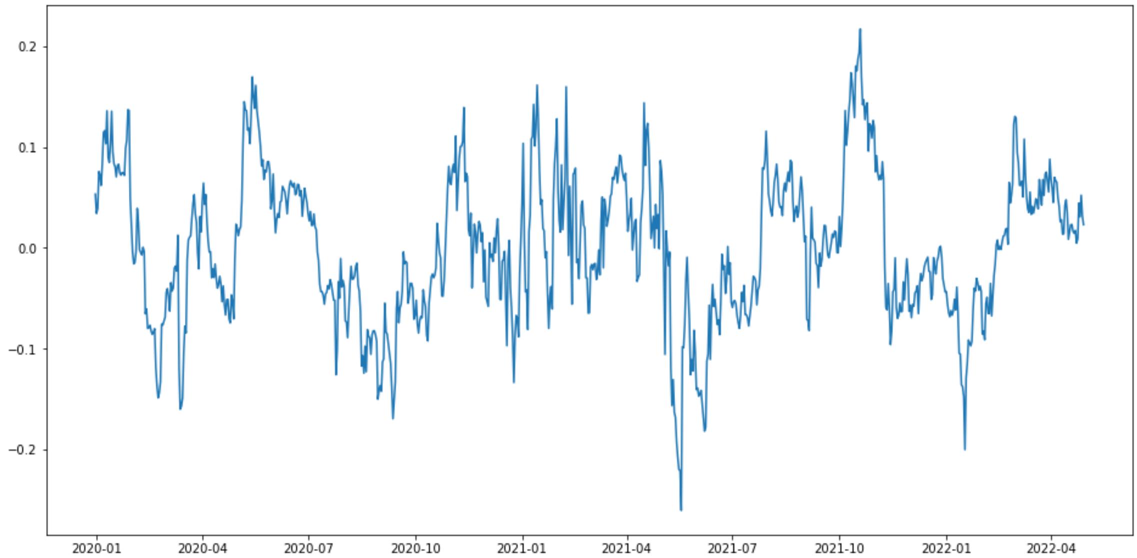


Figure 10: Time series of $Spread_t^{EG}$

Figure 10 Indicates that the residuals of (11) look indeed relatively stationary. Anyway, the portfolio vector and hence cointegrating vector (without intercept) takes the form

$$\beta'_{EG} = (\mathrm{BTC}_t, \mathrm{ETH}_t, \mathrm{BNB}_t, \mathrm{ADA}_t, \mathrm{XRP}_t, \mathrm{DOGE}_t, \mathrm{LTC}_t, \mathrm{BCH}_t, \mathrm{XLM}_t, \mathrm{XMR}_t)' =$$
$$(1, -0.2034, -0.1130, -0.0684, 0.1231, 0.0667, -0.8213, 0.5001, -0.1459, -0.1016) \quad (13)$$

In the following I use β'_{EG} for calculating the spread (i.e. without intercept) which yields

$$Spread_t^{EG} = \mathrm{BTC}_t - 0.2034\mathrm{ETH}_t - 0.1130\mathrm{BNB}_t - 0.0684\mathrm{ADA}_t + 0.1231\mathrm{XRP}_t + 0.0667\mathrm{DOGE}_t -$$
$$0.8213\mathrm{LTC}_t + 0.5001\mathrm{BCH}_t - 0.1459\mathrm{XLM}_t - 0.1016\mathrm{XMR}_t \quad (14)$$

This means in terms of content that in order to long one unit of the spread one need to buy 1, 0.1231, 0.0667 and 0.5001 units of BTC, XRP, DOGE and BCH, respectively. In addition, it is necessary to short sell 0.2034, 0.1130, 0.0684, 0.8213, 0.1459 and 0.1016 shares of ETH, BNB, ADA, LTC, XLM and XMR, respectively. Shorting one unit of the spread works with the same logic but exactly the other way around. In contrast to classical stocks, it is feasible to buy fractional shares up to eight decimal places of a crypto asset.

4.4.3. Johansen procedure results

Table 10: Johansen test results (pairs)

	H_0	Test statistic	Critical value 10%	Critical value 5%	Critical value 1%
BTC-ETH	$r \leq 1$	1.7348	2.7055	3.8415	6.6349
	$r = 0$	14.2598	13.4294	15.4943	19.9349
BTC-BNB	$r \leq 1$	5.1611	2.7055	3.8415	6.6349
	$r = 0$	20.9306	13.4294	15.4943	19.9349
BTC-ADA	$r \leq 1$	2.1820	2.7055	3.8415	6.6349
	$r = 0$	14.0817	13.4294	15.4943	19.9349
BTC-XRP	$r \leq 1$	1.8931	2.7055	3.8415	6.6349
	$r = 0$	20.009	13.4294	15.4943	19.9349
BTC-DOGE	$r \leq 1$	3.2619	2.7055	3.8415	6.6349
	$r = 0$	16.1165	13.4294	15.4943	19.9349
BTC-LTC	$r \leq 1$	0.969	2.7055	3.8415	6.6349
	$r = 0$	10.1242	13.4294	15.4943	19.9349
BTC-BCH	$r \leq 1$	0.6756	2.7055	3.8415	6.6349
	$r = 0$	12.8733	13.4294	15.4943	19.9349
BTC-XLM	$r \leq 1$	1.7671	2.7055	3.8415	6.6349
	$r = 0$	11.087	13.4294	15.4943	19.9349
BTC-XMR	$r \leq 1$	1.8913	2.7055	3.8415	6.6349
	$r = 0$	11.255	13.4294	15.4943	19.9349

Notes: The VAR lag order is chosen by minimizing the AIC for lags up to ten

The Johansen procedure for each pair provides remarkable insights. In contrast to the EG-2 there is cointegration occasionally. When considering the 5% significance level the null of no-cointegration is rejected three times but five times when using the 10% significance level. These results suggest that there are large differences to the Engle-Granger two-step procedure, where no cointegrated pairs were found. As stated in chapter 4.3, I choose the 5% significance level, whereupon I conclude that the three Cryptocurrencies pairs BTC-BNB, BTC-

XRP and BTC-DOGE are cointegrated. The BTC-DOGE relationship is rather surprising since figure 2 shows that the log-returns of Bitcoin have the lowest correlation with the log-returns of DOGE. Anyhow, the ECT of the three cointegrated pairs can be found in the appendix.

Table 11: Optimal lag order

AIC	BIC	FPE	HQIC
5	1	5	1

Table 11 presents the optimal lag order for the VAR for the whole portfolio when using four different information criteria for lags up to ten. Two information criteria suggest a lag order of five and the BIC and HQIC a lag order of one. Since the BIC tends to suggest a too small lag order I rely on the AIC in the following, which implies that a VAR(5) is estimated.

Table 12: Johansen test results (portfolio)

H_0	Test statistic	Critical value 10%	Critical value 5%	Critical value 1%
$r \leq 3$	113.224	120.367	125.618	135.982
$r \leq 2$	157.392	153.634	159.529	171.09
$r \leq 1$	207.712	190.871	197.377	210.037
$r = 0$	274.497	232.103	239.247	253.253

Notes: Critical values are based on J. MacKinnon, Haug, and Michelis (1996)

The results of the trace test suggest that $H_0: r = 0$ is rejected at the 1% significance level and $H_0: r \leq 1$ is rejected at the 5% significance level but $H_0: r \leq 2$ is only rejected at the 10% significance level, which I consider as insufficient evidence. Therefore, I conclude that the rank of the impact matrix is two, which means that as a next step, a VECM with a cointegrating rank of $r_0 = 2$ is estimated. The two significant cointegrating vectors are given by

$$\beta'_{J,1} = (\mathrm{BTC}_t, \mathrm{ETH}_t, \mathrm{BNB}_t, \mathrm{ADA}_t, \mathrm{XRP}_t, \mathrm{DOGE}_t, \mathrm{LTC}_t, \mathrm{BCH}_t, \mathrm{XLM}_t, \mathrm{XMR}_t)' = (1, 0, -0.5311, -0.5807, 0.027, 0.4241, -1.0504, 0.2079, 0.9163, -0.2590)' \quad (15)$$

$$\beta'_{J,2} = (\mathrm{BTC}_t, \mathrm{ETH}_t, \mathrm{BNB}_t, \mathrm{ADA}_t, \mathrm{XRP}_t, \mathrm{DOGE}_t, \mathrm{LTC}_t, \mathrm{BCH}_t, \mathrm{XLM}_t, \mathrm{XMR}_t)' = (0, 1, 1.8189, -0.959, 0.3268, -0.865, -5.4176, 3.899, 3.205, -2.3504)' \quad (16)$$

Table 13 again shows both significant cointegrating vectors of equation (15) and (16) but also includes the significance of their parameters. Notwithstanding that the t-values are not very reliably t-distributed, it is still a good approximation. When considering the vector $\beta_{J,1}$ there are only four significant parameters (apart from BTC and ETH), namely BNB, ADA, LTC and XLM, while ADA is only significant at the 10% level. $\beta_{J,2}$ has a more desired outcome since there are only two insignificant parameters (ADA and XRP) when leaving BTC and ETH out of consideration. LTC and BCH represent the largest entries (in absolute terms) of the second vector and are highly significant.

Table 13: Cointegrating vectors (Johansen procedure)

	$\boldsymbol{\beta_{J,1}}$	$\boldsymbol{\beta_{J,2}}$
BTC	1.0000	0.0000
ETH	0.0000	1.0000
BNB	-0.5311***	1.8189***
ADA	-0.5807*	-0.9590
XRP	0.0266	0.3268
DOGE	0.4241	-0.8650*
LTC	-1.0504***	-5.4176***
BCH	0.2079	3.8990***
XLM	0.9163***	3.2049***
XMR	-0.2590	-2.3504***

*Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level and *** denotes significance at the 1% level, the significance level of BTC and ETH is not reported since they are (1,0) and (0,1) respectively by construction and hence do not have a finite t-value*

The next two graphs show the time series of both error-correction terms, which should be stationary and represent $Spread_t^{J,i}$ for $i = 1,2$.

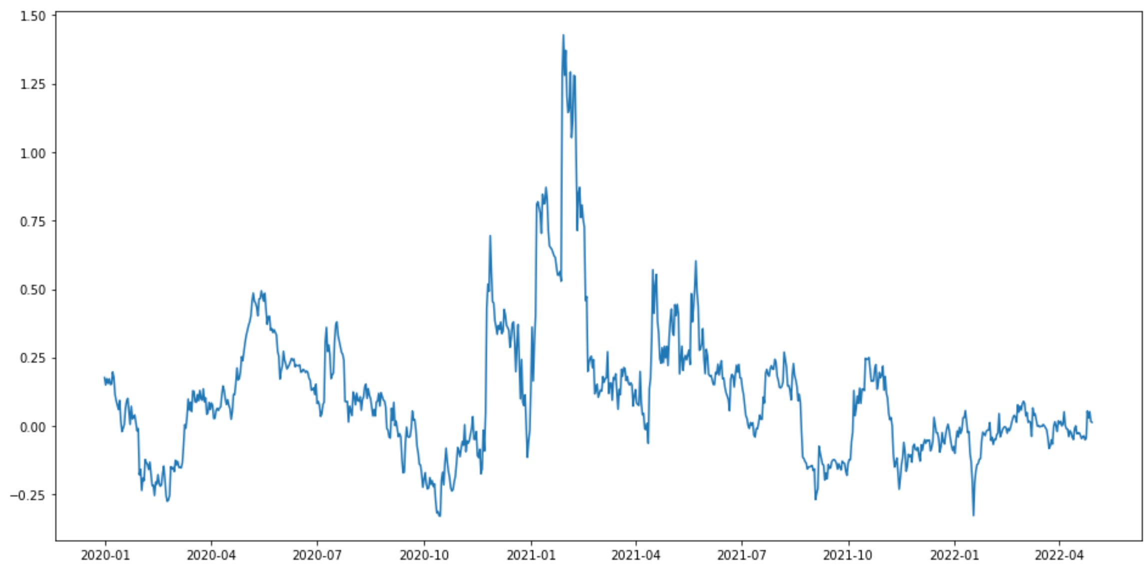


Figure 11: Time series of the ECT 1

The time series of the ECT 1 in Figure 11 looks indeed stationary but it is striking that there was a huge deviation from the long-run equilibrium at the beginning of the year 2021. This may have been caused by the fast increase of the BTC price, which peaked at around $50,000 in March 2021. This conjecture seems plausible since monetarily $\beta'_{J,1}$ puts the greatest weight on BTC with one unit. It is also remarkable that the start of the COVID-19 crisis has not resulted in a significant disequilibrium. $\beta'_{J,1}$ puts no weight on ETH because of the basis vectors of the unique cointegrating space the BTC-ETH interaction is by construction excluded. Other than that, however, the whole process looks pretty stationary.

Figure 12: Time series of the ECT 2

Figure 12 shows a similar picture since the time series of the ECT 2 fluctuates around zero. Again, at the start of the year 2021 the Cryptocurrencies were in disequilibrium. However, the spread converges back to its long-run equilibrium relatively fast, which indicates a long-term relationship between the variables. As expected $\beta'_{J,2}$ puts zero weight on BTC because of the routine as programmed. The exclusion of any BTC-ETH interaction may be also responsible for the better fit of the ECT based on the EG-2 for two reasons. First, the correlation matrix in section 3 insinuates that Bitcoin forms a group of closely related assets with Ethereum, Litecoin and Bitcoin Cash. This is why the cointegrating vectors may depend strongly on the BTC-ETH, BTC-LTC and BTC-BCH interaction. While $\beta_{J,1}$ and $\beta_{J,2}$ exclude the BTC-ETH relationship by definition, β_{EG} allows any interaction. Second, the Johansen trace test rejects the null of no-cointegration between BTC and ETH at least at the 10% significance level indicating the importance of the BTC-ETH interaction.

Before discussing the trading strategies in more detail table 14 provides the estimates of the loading matrix $\hat{\alpha}$.

Table 14: Loading matrix

	$\boldsymbol{ECT_1}$	$\boldsymbol{ECT_2}$
BTC	0.0133**	-0.0077***
ETH	0.0202***	-0.0088***
BNB	0.0401***	-0.0100***
ADA	0.0344	-0.0094**
XRP	0.0049	-0.0058
DOGE	0.0263**	-0.0044
LTC	0.0147*	-0.0051
BCH	0.0116	-0.0099***
XLM	0.0055	-0.0177***
XMR	0.0203***	0.0013

*Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level and *** denotes significance at the 1% level*

The outcome of the loading matrix is quite interesting in a sense that the adjustment parameters in ECT_1 are all positive while in ECT_2 the parameters are all negative (except for XMR). However, the coefficients are relatively small implying a slow speed of adjustment. Since ten variables are involved in the VECM, the interpretation of the parameters is complex. In such a complex system the signs of the parameters do not indicate whether there is better adjustment to return to the long-run equilibrium after a disequilibrium. Hence, the most important information in the loading is about significance not the sign. When considering the significance in ECT_1 I found that only BTC, ETH, BNB, DOGE, LTC and XMR are significant at least at the 10% significance level, which means that only these variables adjust to deviations from the long-run equilibrium. The significant loading parameters in ECT_2 include BTC, ETH, BNB, ADA, BCH and XLM, which are all significant at the 5% level meaning that the variables react to the error-correction term. The variables with a true alpha of zero are weakly exogenous for the cointegrating vector as defined by Engle, Hendry, and Richard (1983), which implies that Ripple (XRP) is the only exogenous variable.

4.4.4. Trading strategy results

4.4.4.1. In-sample results

Before discussing the results of each trading strategy with different thresholds, figure 13 to 15 at first show the estimated distribution of $Spread_t^i \; for \; i \in \{EG, J1, J2\}$.

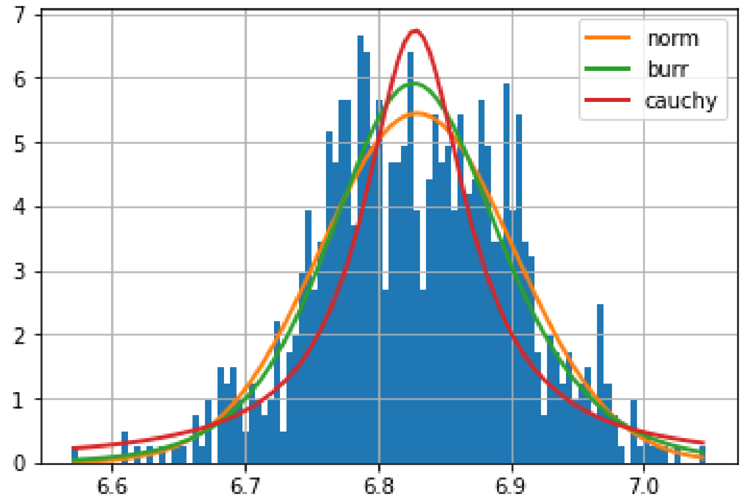


Figure 13: Estimated distribution of $Spread_t^{EG}$

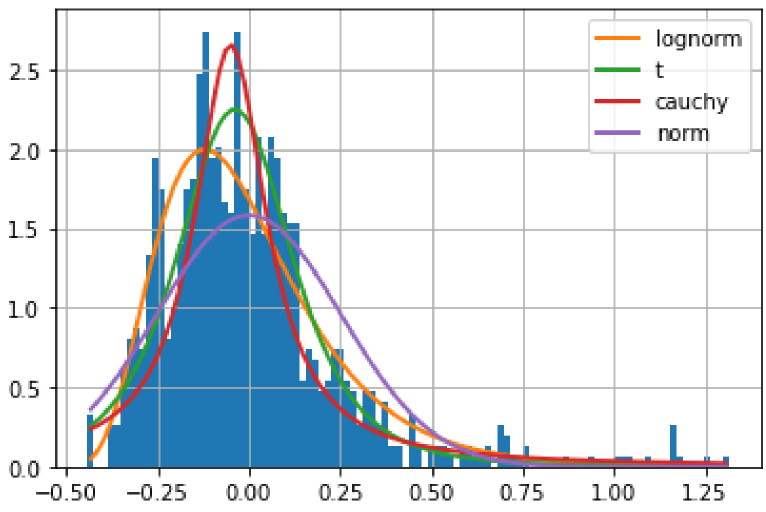


Figure 14: Estimated distribution of $Spread_t^{J1}$

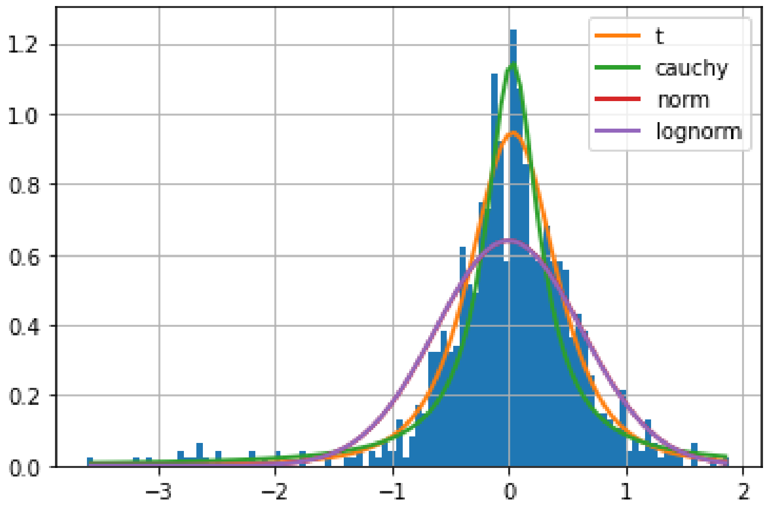


Figure 15: Estimated distribution of $Spread_t^{J,2}$

While figure 13 suggests that $Spread_t^{EG} \sim N(\mu = 6.8297, \sigma = \ 0.0733)$, figure 14 recommends a log-normal distribution as the best fitting distribution with $Spread_t^{J,1} \sim LN(\mu = 0.5561, \sigma = \ 0.3869)$. Furthermore, figure 15 indicates that the t-distribution has the smallest sum of the squared errors, which implies $Spread_t^{J,2} \sim t_{3.0768}(\mu = 0.0351, \sigma = \ 0.3889)$, where 3.0768 are the degrees of freedom. Now it is possible to solve equation (10) for each CDF in order to obtain the theoretical optimal threshold. The results are provided in the next table.

Table 15: Optimal thresholds

Spread	**Estimated CDF**	$\boldsymbol{\tau^*}$
$Spread_t^{EG}$	$\Psi^{EG}(.) = \ N(\mu = 6.8297, \sigma = \ 0.0733)$	0.055
$Spread_t^{J,1}$	$\Psi^{J,1}(.) = LN(\mu = 0.5561, \sigma = \ 0.3869)$	0.209
$Spread_t^{J,2}$	$\Psi^{J,2}(.) = \ t_{3.0768}(\mu = 0.0351, \sigma = \ 0.3889)$	0.37

For the further analysis, I use the following notation for the sake of simplicity:

- Trading strategy 1: Use $Spread_t^{EG}$ with $\tau = \sigma$
- Trading strategy 2: Use $Spread_t^{EG}$ with $\tau = \tau^*$
- Trading strategy 3: Use $Spread_t^{J,1}$ with $\tau = \sigma$
- Trading strategy 4: Use $Spread_t^{J,1}$ with $\tau = \tau^*$
- Trading strategy 5: Use $Spread_t^{J,2}$ with $\tau = \sigma$
- Trading strategy 6: Use $Spread_t^{J,2}$ with $\tau = \tau^*$
- Trading strategy 7: Passive investing approach

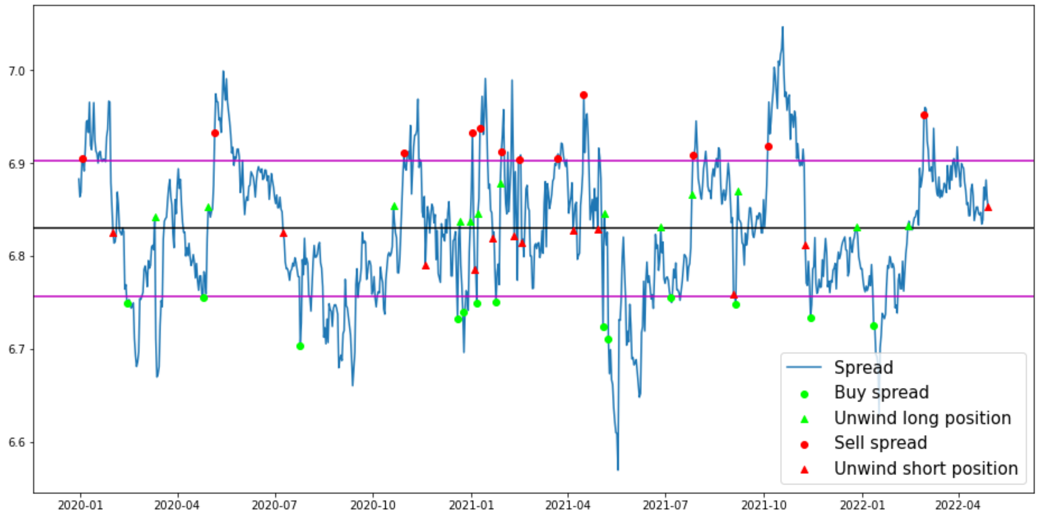


Figure 16: Trading strategy 1 (In-sample)

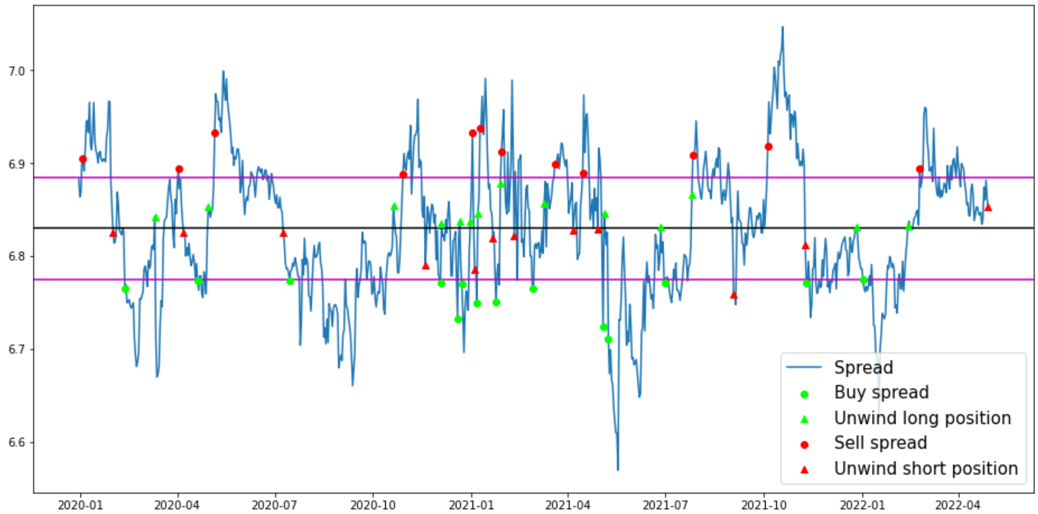


Figure 17: Trading strategy 2 (In-sample)

Figure 18: Trading strategy 3 (In-sample)

Figure 19: Trading strategy 4 (In-sample)

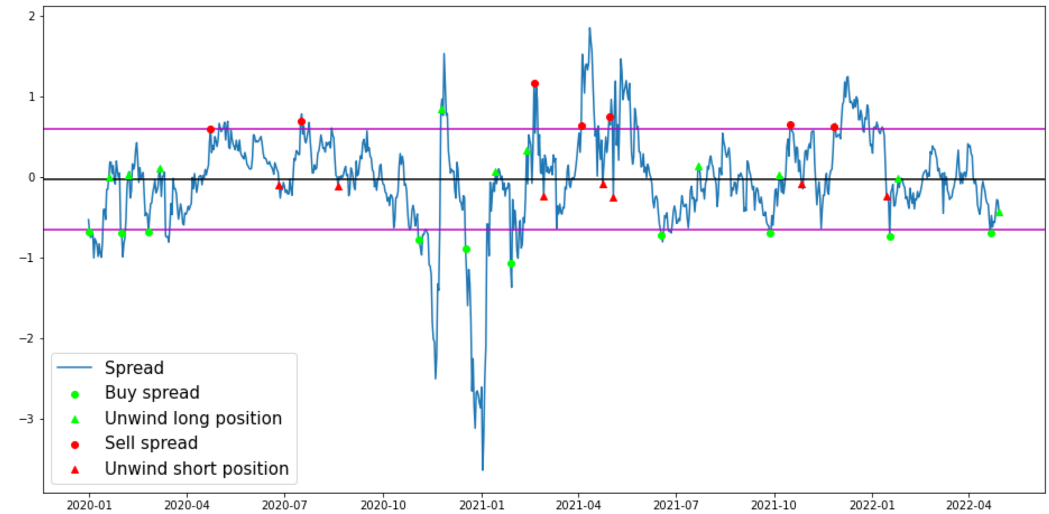


Figure 20: Trading strategy 5 (In-sample)

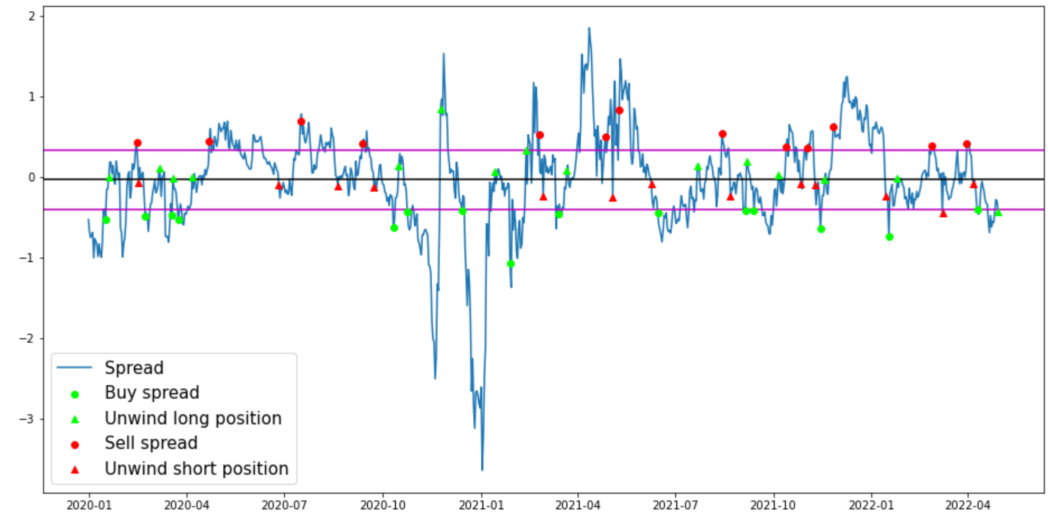


Figure 21: Trading strategy 6 (In-sample)

When comparing all strategies, it is striking that both spreads based on the Johansen procedure have a much higher standard deviation (in absolute terms) than the spread of the EG-2. Since the profit for two subsequent trades equals approximately τ it can be conjectured that $Spread_t^{J,1}$ and $Spread_t^{J,2}$ have a higher average return. Moreover, $Spread_t^{EG}$ and $Spread_t^{J,2}$ seem to have a better mean-reverting property than $Spread_t^{J,1}$ resulting in more mean-crossings and thus more trades. When looking at the optimal thresholds in table 22 and comparing them to the standard deviations of the spreads it is discernable that τ^* of all spreads is smaller than σ leading to more profit realizations. Table 16 contains a summary of the most important key data.

Table 16: Summary of all trading strategies (In-sample)

Strategy	**Trading strategy 1**	**Trading strategy 2**	**Trading strategy 3**	**Trading strategy 4**	**Trading strategy 5**	**Trading strategy 6**
Total return realizations	25	26	9	10	17	28
Total transactions	50	52	18	20	34	56
Largest log-return	15.02%	15.00%	60.22%	55.18 %	162.41%	139.54%
Lowest log-return	7.72%	4.18%	16.33%	11.50%	26.56%	-1.56%
Average log-return	11.11%	9.18%	32.81%	29.87%	87.90%	64.84%
Total log-return	277.83%	238.66%	295.31%	298.74%	1,494.37%	1,815.42%

As expected, trading strategy 3 and 4 result in relatively less return realizations due to the slow mean reversion. A return realization occurs after unwinding a long/short position that is why two transactions in total (long/short the spread and unwind long/short position) are needed. Nevertheless, the overall returns of both strategies are higher than the total return of trading strategy 1 and 2 according to $Spread_t^{EG}$, even though strategy 1 and 2 lead to almost three times as many transactions as strategy 3 and 4. Another outcome that stands out is the clear superiority of trading strategy 5 and 6, which generate a total return of 1,494.37% and 1,815.42%, respectively. The main driver are most likely the fast mean-reverting $Spread_t^{J,2}$ and its relatively high (absolute) standard deviation. Even though the average return of strategy 5 is around 20 percent points larger than of strategy 6, the significantly more transactions of strategy 6 predominate yielding a higher overall return. On the one hand, both trading strategies resting upon the Johansen approach with the theoretical optimal threshold outperform the strategies where σ is used as threshold. On the other hand, the mean-reverting portfolio constructed with $Spread_t^{EG}$ and $\tau = \tau^*$ has a weaker performance than strategy 1 where just the standard deviation is used. Because of this result the effect of the usage of τ^*

is ambiguous but because of the superiority in two out of three cases I presume that estimating the distribution of the spread still makes sense for calculating the optimal threshold. Finally, the next plot illustrates the time series of the cumulative returns of all six trading strategies and the passive investment approach.

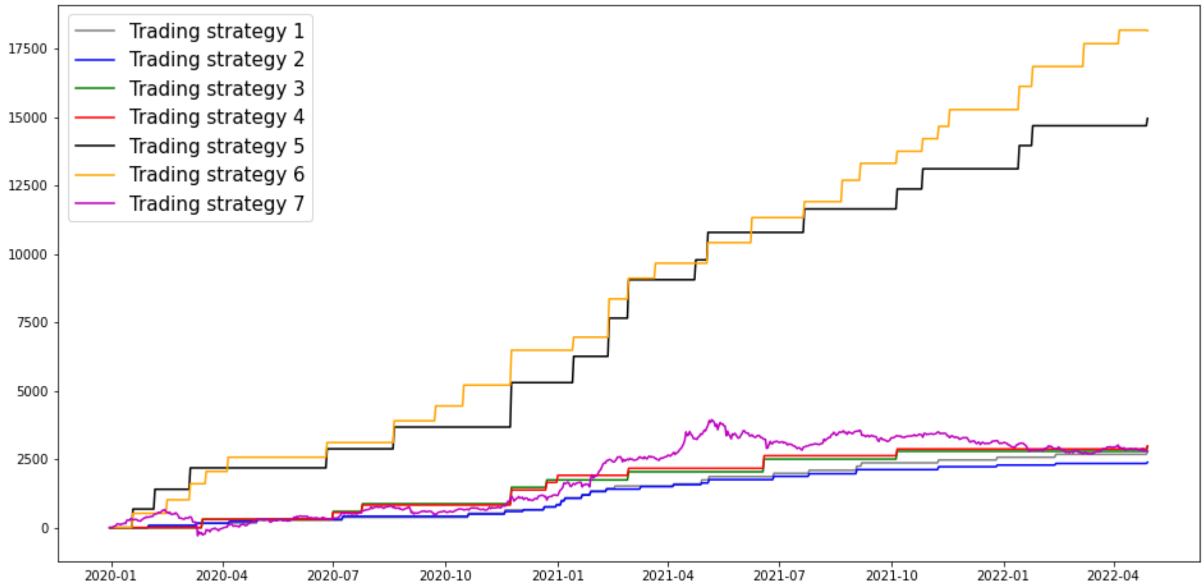


Figure 22: Cumulative returns – All trading strategies (In-sample)

It is eye-catching that trading strategy 5 and 6 are clearly superior to all other strategies. Moreover, I found that if all cryptocurrencies bought at the beginning were sold at almost any time point during 2021, the passive investment strategy would be superior to all trading strategies except both strategies generated by $Spread_t^{J,2}$. This is a strong result since it suggests that the risk associated with trading has no remarkable advantage over a long-term passive investing strategy. However, due to the crypto bear market starting at the beginning of 2022 strategy 7 becomes less attractive ex post and approaches a similar level as trading strategy 3 and 4. Because of the strong superiority of strategy 5 and 6 it is hard to see big differences between the other investment strategies, this is why the next figure shows the cumulative returns of the other strategies. Figure 23 indicates that trading strategy 3 and 4 outperform the passive investment strategy, while both strategies constructed by $Spread_t^{EG}$ generate less profit than strategy 7. Table 17 summarizes the final wealth of all strategies including the $1,000 investment at the beginning. Trading strategy 6 with a final wealth of $19,154.22 is second to none and yields a five-time higher final wealth than the passive investing strategy. Whether the superiority of the strategies generated by $Spread_t^{J,2}$ is also present out-of-sample, this is investigated in the next chapter.

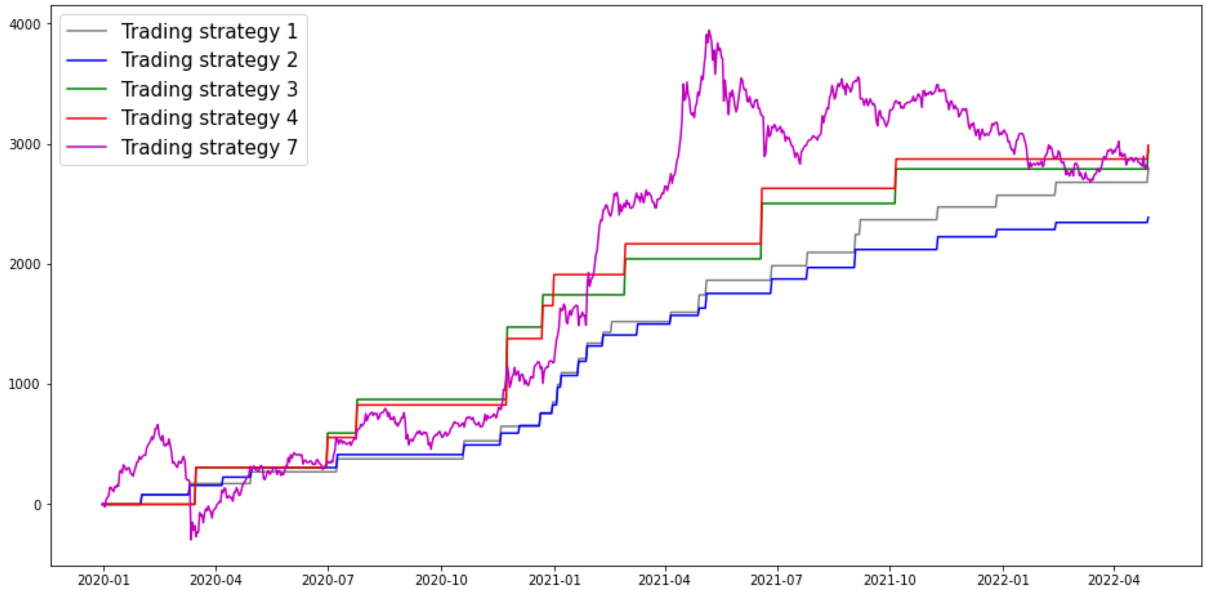


Figure 23: Cumulative returns – Excluding trading strategy 5 and 6 (In-sample)

Table 17: Final wealth of all trading strategies (In-sample)

Strategy	Final wealth
Trading strategy 1	\$3,778.31
Trading strategy 2	\$3,386.56
Trading strategy 3	\$3,953.14
Trading strategy 4	\$3,987.43
Trading strategy 5	\$15,943.67
Trading strategy 6	**\$19,154.22**
Trading strategy 7	\$3,789.56

4.4.4.2. Out-of-sample results

As before, for simplicity the further analysis considers the following notation:

- Trading strategy 1: Use $Spread_t^{EG}$ with $\tau = \sigma_t(90)$
- Trading strategy 2: Use $Spread_t^{EG}$ with $\tau = \sigma_t(10)$
- Trading strategy 3: Use $Spread_t^{J,1}$ with $\tau = \sigma_t(90)$
- Trading strategy 4: Use $Spread_t^{J,1}$ with $\tau = \sigma_t(10)$
- Trading strategy 5: Use $Spread_t^{J,2}$ with $\tau = \sigma_t(90)$
- Trading strategy 6: Use $Spread_t^{J,2}$ with $\tau = \sigma_t(10)$
- Trading strategy 7: Passive investing approach

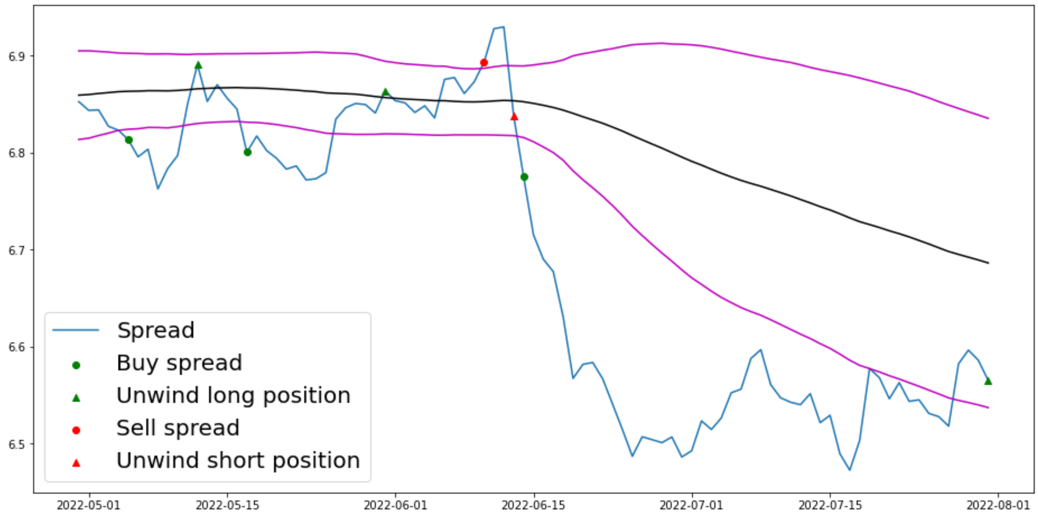


Figure 24: Trading strategy 1 (Out-of-sample)

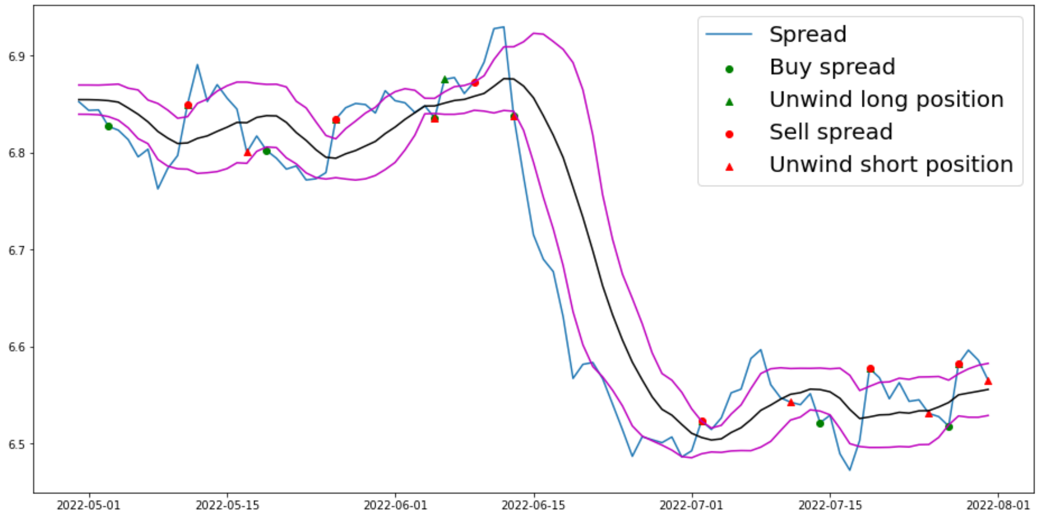


Figure 25: Trading strategy 2 (Out-of-sample)

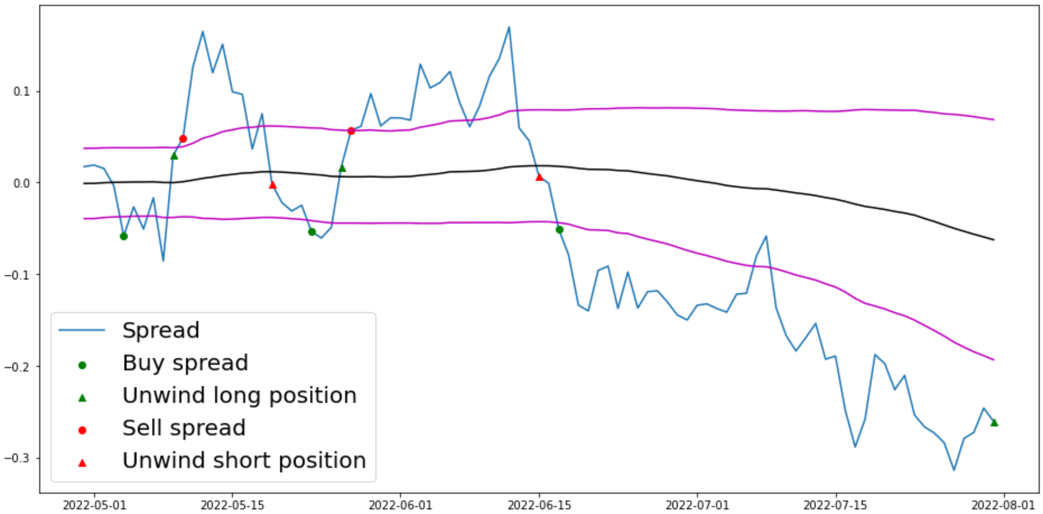


Figure 26: Trading strategy 3 (Out-of-sample)

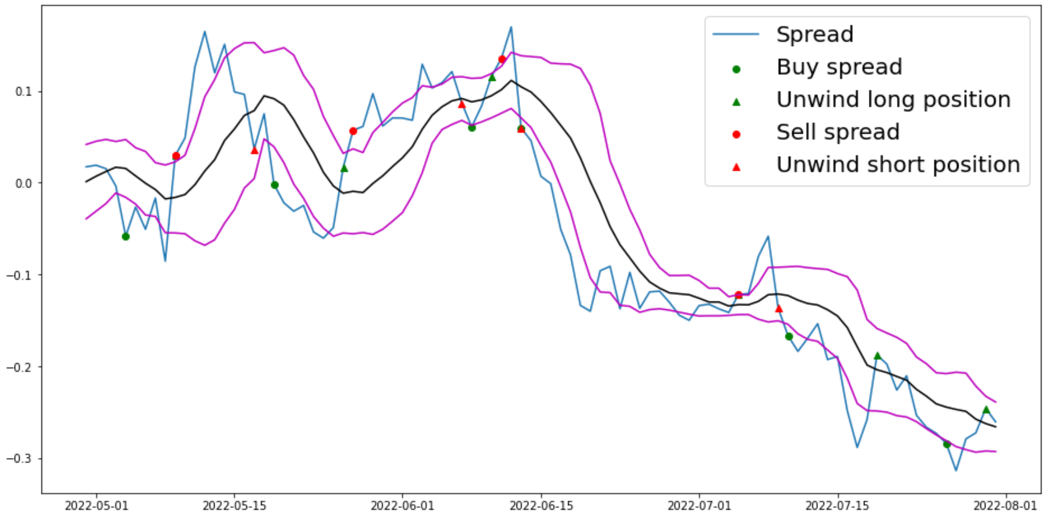


Figure 27: Trading strategy 4 (Out-of-sample)

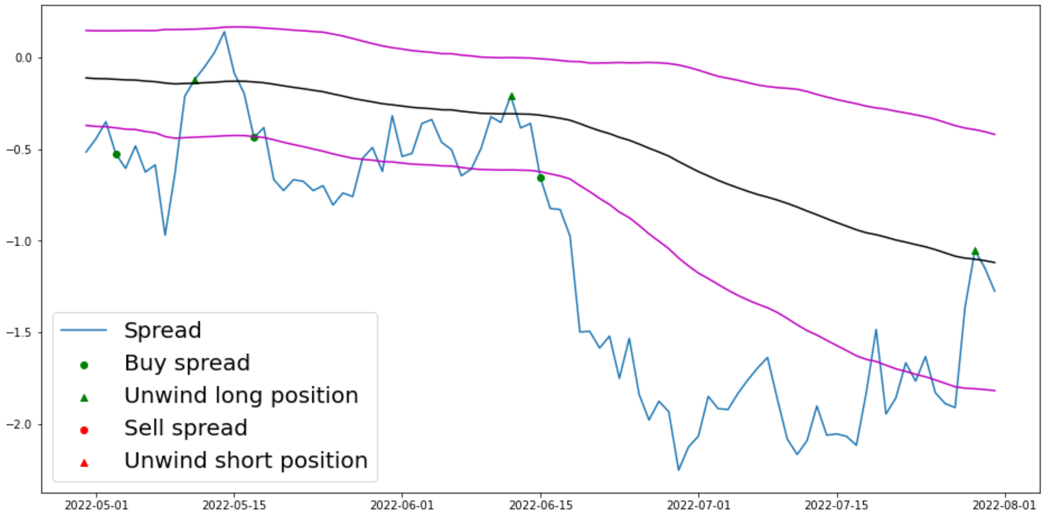


Figure 28: Trading strategy 5 (Out-of-sample)

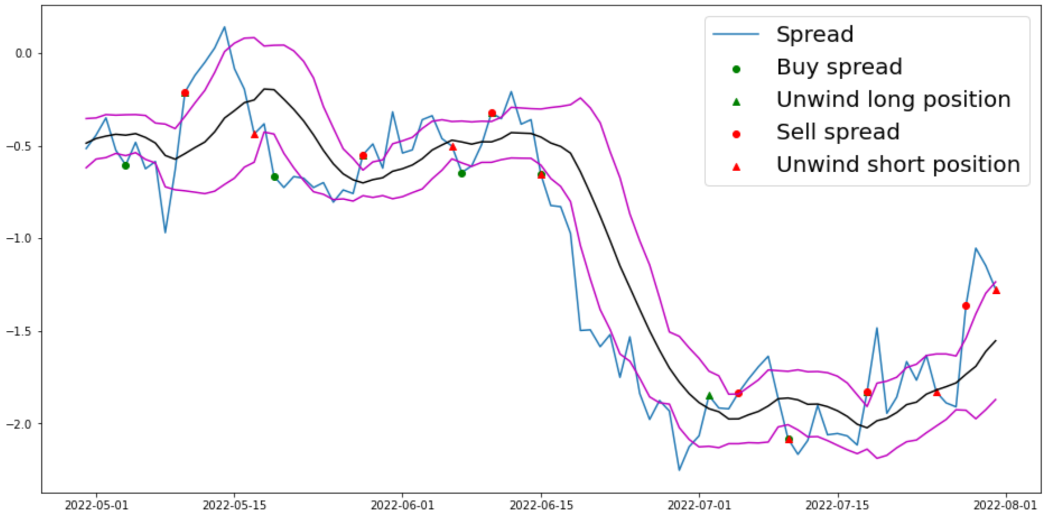


Figure 29: Trading strategy 6 (Out-of-sample)

First of all, it is immediately apparent that the process of all three spreads has the same pattern. Until the mid of June 2022, the spread is more or less mean-reverting followed by a huge downward movement. This suggests that during this time interval something happened causing a disequilibrium of the Cryptocurrencies. As it is the case in-sample the mean-reversion of $Spread_t^{J,1}$ is very slow leading to a continuing downward movement, while in July 2022 $Spread_t^{EG}$ fluctuates at a relatively constant level and $Spread_t^{J,2}$ even slowly converges back to its long-run equilibrium at the end of the period indicating the best mean-reversion. Another explanation for the better mean-reverting property of $Spread_t^{J,2}$ are the findings of table 14 since the negative loading parameters lead to an adjustment to return to the long-run equilibrium after the disequilibrium. Furthermore, all trading strategies with a 90-days rolling window generate a loss at the end of the test period by unwinding the last long position because of the significant deviation from the long-run equilibrium. Using a shorter window size results in significantly more transaction than the longer 90-days window. Table 18 summarizes the key facts of all out-of-sample trading strategies.

Table 18: Summary of all trading strategies (Out-of-sample)

Strategy	Trading strategy 1	Trading strategy 2	Trading strategy 3	Trading strategy 4	Trading strategy 5	Trading strategy 6
Total return realizations	4	12	5	10	3	11
Total transactions	8	17	10	17	6	16
Largest log-return	7.74%	6.44%	8.76%	8.76%	40.43%	39.42%
Lowest log-return	-21.01%	-31.45%	-20.95%	-18.09%	-39.43	-119.34%
Average log-return	-0.34%	0.24%	0.97%	0.50%	7.84%	5.13%
Total log-return	-1.37%	2.84%	4.86%	5.04%	23.51%	56.44%

The results show that all trading strategies except strategy 1 produce a positive return at the end of the period. Another interesting finding is that the strategies with a shorter window size seem to outperform the trading strategies with a 90-day window length. Just like it is the case in-sample, trading strategy 2 based on $Spread_t^{EG}$ yields the lowest positive overall profit. Likewise, trading strategy 6 on the basis of $Spread_t^{J,2}$ yields the highest total return with 56.44% despite a loss of more than 100%. This outcome could provide evidence that a good in-sample trading performance with a cointegrated Cryptocurrency portfolio is also a good indicator for a promising out-of-sample performance. This can be pinned down by the fact that regardless of the window size both trading strategies based on $Spread_t^{J,2}$ outperform all other trading strategies. In any case, the main takeaway of the out-of-sample backtesting analysis is that the strategies with the 10-days window length are superior to the strategies with a window size of three months by resulting in a considerable higher total return. When comparing the first two trading strategies one can see that using a short window generates a positive profit while using the 90-days window leads to a loss. In the next chart the cumulative returns of all six trading strategies are compared to the passive investment approach.

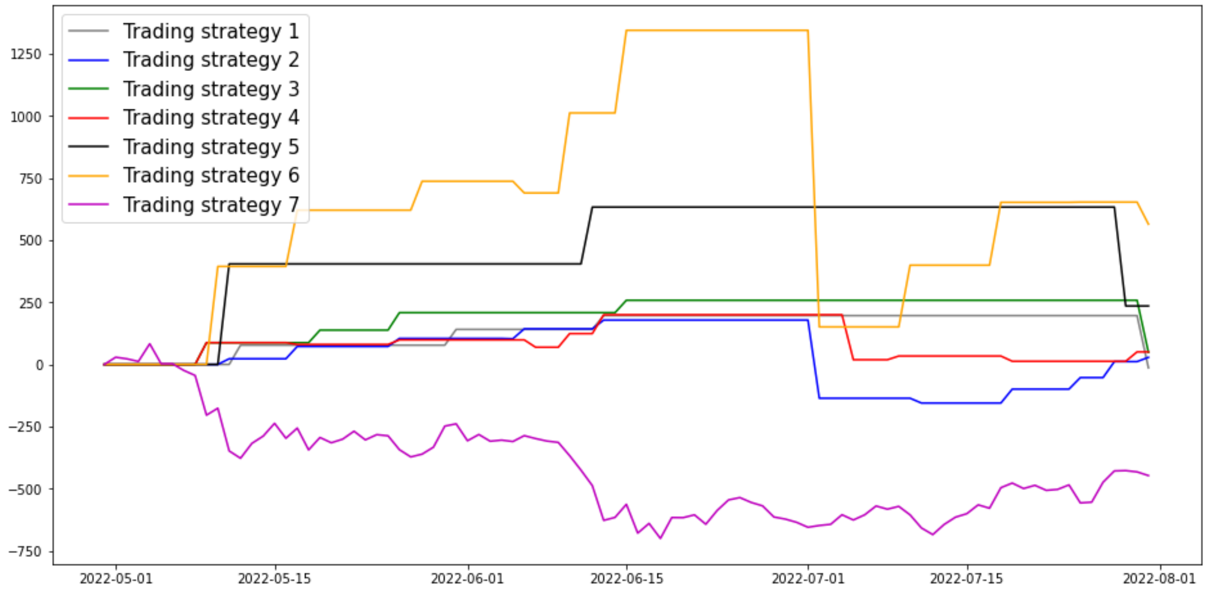


Figure 30: Cumulative returns (Out-of-sample)

The first thing to notice is that all trading strategies yield a higher total return than the passive investment approach because of the crypto bear market during the year 2022. These results suggest that cointegrated Cryptocurrency portfolios can be useful for hedging against a downtrending market. As already implied by table 18 strategy 6 is by far the best performing trading strategy. In addition, trading strategy 5 tends to outperform the first four trading strategies and even strategy 6 at the beginning of July but the costly unwinding of the last long position leads to a huge loss and hence a significant weaker performance than trading strategy 6. However, the bottom line is that the superiority of trading strategy 6 based on $Spread_t^{J,2}$ is also prevailing out-of-sample. When looking at figure 30 it is obvious that trading strategy 1 and the passive investment strategy result in a loss while the other strategies yield positive profits. To relate this outcome to the high inflation rates in Europe and the US nowadays, these findings indicate that statistical arbitrage strategies based on cointegrated Cryptocurrency portfolios are indeed a profitable shield against high inflation rates. When considering an annual inflation of 10% for simplicity, it is necessary to generate a final wealth of at least $\$1{,}000 * 1.1^{\frac{3}{12}} = \$1{,}024.11$ after the three months test period, which can be achieved with five out of six statistical arbitrage strategies.

Table 19: Final wealth of all trading strategies (Out-of-sample)

Strategy	Final wealth
Trading strategy 1	$986.29
Trading strategy 2	$1,028.46
Trading strategy 3	$1,048.62
Trading strategy 4	$1,050.44
Trading strategy 5	$1,235.12
Trading strategy 6	**$1,564.36**
Trading strategy 7	$552.77

4.4.5. Volatility and the spread

In this chapter I investigate the impact of market volatility on the spread that is used for statistical arbitrage. The time series of the three spreads show that there are huge spikes occasionally, which may be caused by high volatility in the market. This becomes especially apparent when looking at $Spread_t^{J,1}$ and $Spread_t^{J,2}$ in section 4.4.3. As a proxy for the overall volatility in the Cryptocurrency market I use the Crypto Volatility Index (CVI), which is the counterpart to the well-known CBOE Volatility Index (VIX). The next plot shows the CVI during the in-sample period. Figure 31 indicates that there are three excessive spikes caused by, among other things, the start of the COVID-19 pandemic. It is noticeable that both moments where $Spread_t^{J,1}$ and $Spread_t^{J,2}$ are in disequilibrium are periods of high market volatility implied by the CVI. This suggests that there may be a connection between the spread and market volatility.

Figure 31: Crypto Volatility Index

In order to find empirical evidence for a potential relatedness between volatility and the spread I apply a simple granger causality test introduced by Granger (1969). Usually, granger causality is tested via a restriction test in a VAR. For the following analysis I estimate three bivariate VAR models to test for granger causality between the CVI and each individual spread. To ensure a stable VAR it is necessary that each variable is I(0). $Spread_t^{EG}$, $Spread_t^{J,1}$ and $Spread_t^{J,2}$ are covariance stationary by definition, therefore it is only necessary to run a unit root test on the CVI. The results can be found in table 20.

Table 20: Unit root test results for CVI

Variable	Test statistic	Type	Critical value 1%	Critical value 5%
CVI	-3.2392	DF_μ	-3.43	-2.86

The DF_μ test rejects the null of a unit root at the 5% significance level indicating that CVI is covariance stationary, resulting in stable VAR models.

Table 21: Granger causality results

H_0	P-value	Conclusion	VAR lag order
CVI DON'T granger cause $Spread_t^{EG}$	0.4265	Not reject	1
CVI DON'T granger cause $Spread_t^{J,1}$	0.0256	Reject	5
CVI DON'T granger cause $Spread_t^{J,2}$	0.2531	Not reject	8

Notes: The VAR lag order is chosen by minimizing the AIC for lags up to ten

The results of table 21 suggest that there is only one case where the CVI granger causes the spread. This seems at first appearance a weak result for a strong connection between volatility and the spread used for establishing the trading strategies. However, the CVI is only a noisy proxy for the overall Cryptocurrency market volatility and not a reliable indicator for the volatility of the three cointegrated portfolios based on just ten specific Cryptocurrencies. Nevertheless, I found evidence that the CVI granger cause $Spread_t^{J,1}$ implying that the CVI is useful in forecasting $Spread_t^{J,1}$. Thus, it is likely that estimating the volatility of each individual Cryptocurrency in the portfolio helps to generate a more accurate forecast for the spread. Hence, the incorporation of volatility for statistical arbitrage strategies most likely leads to a better out-of-sample performance especially due to the high volatile crypto market. On the basis of this conjecture, the next chapter deals with a rigorous volatility analysis of the most popular Cryptocurrency BTC. In addition, I make out-of-sample one-step ahead forecasts and consider potential price jumps. Of course, the same analysis can be applied to the other nine Cryptocurrencies that are used for the cointegrated portfolios.

5. Volatility modelling

Financial time-series analysis is one of the most important research fields in financial econometrics. Understanding volatility in financial markets is crucial in risk management, portfolio allocation, derivative pricing and other related fields. In this section, I introduce two kinds of volatility models, which are used for estimating and forecasting the volatility of BTC in the period from 1st Apr 2015 until 31st Mar 2021. On the one hand, I use different types of GARCH models including the standard GARCH model, the EGARCH, the GJR-GARCH and the more recent realized GARCH model. On the other hand, six different HAR-RV models are also estimated.

5.1. Theoretical framework

In this chapter, the methodology of the four different GARCH models and the HAR-type models is described.

5.1.1. GARCH models

5.1.1.1. Standard GARCH model

Generalized autoregressive conditional heteroskedasticity (GARCH) models of order p and q as proposed by Bollerslev (1986) are used for modelling the conditional variance and allow for a high persistence in the process as it can be shown that $GARCH(p,q) = ARCH(\infty)$ as long as the volatility process is stable and $q > 0$. The GARCH(p,q) process is given by

$$r_t = \mu_t + \sigma_t z_t \tag{17}$$

$$\mathbb{V}[r_t|\mathcal{F}_{t-1}] = \sigma_t^2 = \omega + \sum_{j=1}^{p} \alpha_j r_{t-j}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{18}$$

where r_t denotes the observed log-returns of Bitcoin and μ_t the conditional mean. Z_t is a standardized i.i.d. error term with $\mathbb{E}[z_t] = 0 \; and \; \mathbb{V}[z_t] = 1$, formally $z_t \sim i.i.d.(0,1)$. In the following analysis z_t is considered as it follows a normal distribution and a Student's t-distribution. σ_t is a conditionally deterministic function depending on the history of the process and $\mathcal{F}_{t-1}$ denotes the information set up to and including t-1. Furthermore, p denotes the number of lags of the squared log-returns as specified in equation (17) and q the number of lags of the conditional variance. In order to ensure the positivity of the conditional variance it must also hold that $\omega > 0, \alpha_j \geq 0\ for\ j = 1, \ldots, p\ and\ \beta_j \geq 0$ for $\mathrm{j} = 1, \ldots, \mathrm{q}$. In this paper in the following only the simple case where $p = q = 1$ is considered, which reduces the volatility process to

$$\sigma_t^2 = \omega + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{18a}$$

where $\omega > 0,\ \alpha_1 \geq 0\ and\ \beta_1 \geq 0$. To ensure stability it must hold that $\alpha_1 + \beta_1 < 1$.

5.1.1.2. EGARCH model

The exponential GARCH (EGARCH) model is an asymmetric GARCH model introduced by Nelson (1991) that considers leverage effects between positive and negative shocks. The conditional variance in the EGARCH(1,1) model is given by

$$\log(\sigma_t^2) = \omega + \alpha_1 z_{t-1} + \gamma_1(|z_{t-1}| - E[|z_{t-1}|]) + \beta_1 \log(\sigma_{t-1}^2), \tag{19}$$

where α_1 describes the sign effect and γ_1 captures the magnitude of z_t. Moreover, the expectation of the absolute value of the standardized returns z_t is given by $E[|z_t|] = \int_{-\infty}^{\infty} |z_t| f(z, 0, 1, \ldots) dz$. An advantage of the EGARCH model is that because of the exponential functional form the model does not need any constraints on coefficient parameters, it is always defined. However, the stationarity conditions of the exponential GARCH model are a little bit more complicated. Nelson (1991) found that covariance stationarity has stronger requirements than strict stationarity. Strict stationarity requires $|\beta_1| < 1$ (He, Teräsvirta, and Malmsten, 2002). Here, covariance stationarity implies strict stationarity but not vice versa since Nelson (1991) shows that the distribution of z_t is crucial for the existence of finite unconditional moments. If z_t for example follows a Student-t distribution with finite degrees of freedoms and $|\beta_1| < 1$ holds, then the unconditional variance does still not exist which means that volatility process is strictly stationary but not covariance stationary.

5.1.1.3. GJR-GARCH model

The GJR-GARCH model introduced by Glosten, Jagannathan, and Runkle (1993) is another modification of the standard GARCH-model. It is also an asymmetric model which values positive and negative shocks of the conditional variance differently using the indicator function $\mathrm{I}(.)$. The conditional variance of the GJR-GARCH(1,1) is given by

$$\sigma_t^2 = \omega + \alpha_1 r_{t-1}^2 + \gamma_1 r_{t-1}^2 \mathrm{I}_{\{r_{t-1} < 0\}} + \beta_1 \sigma_{t-1}^2, \tag{20}$$

where γ_1 covers the leverage effect. To guarantee the non-negativity of the variance equation it must hold: $\omega > 0, \alpha_1 \geq 0$, $\beta_1 \geq 0$ and $\alpha_1 + \gamma_1 \geq 0$. Because of the use of the indicator function the stability condition is not straightforward and depends on the conditional distribution of z_t. In general, the stability condition for the GJR-GARCH(1,1) requires $\alpha_1 + \beta_1 + \gamma_1 \rho < 1$, where ρ denotes the expected value of the negative standardized squared returns. Formally, $\rho = E\left[\mathrm{I}_{\{r_t<0\}} z_{t-1}^2\right] = \int_{-\infty}^{0} f(z, 0,1, \ldots) dz$. In case of the symmetric normal distribution $\rho = \frac{1}{2}$ which implies $\alpha_1 + \beta_1 + \frac{1}{2}\gamma_1 < 1$ as stability condition.

5.1.1.4. Realized GARCH model

The realized GARCH model by Hansen, Huang, and Shek (2012) exploits realized measures of volatility by including a measurement equation. The measurement equation links the observed realized measure to the latent volatility of the returns (Hansen, Huang, and Shek, 2012). Another advantage of the realized GARCH model is that it also accounts for leverage effects. This asymmetric reaction to shocks is a useful property when modelling the conditional variance of stock returns because as pointed out by Black (1976) positive and negative news may affect future volatility in an asymmetric way. In the following analysis the realized GARCH(1,1) model with a log-linear specification suggested by Hansen, Huang, and Shek (2012) is used. Formally, the model is described by the following GARCH and measurement equation:

$$\log(\sigma_t^2) = \omega + \alpha_1 \log(x_{t-1}) + \beta_1 \log(\sigma_{t-1}^2) \tag{21}$$

$$\log(x_t) = \xi + \delta \log(\sigma_t^2) + \tau(z_t) + u_t, \;\; u_t \sim N(0, \lambda) \tag{22}$$

In contrast to the standard GARCH(1,1) model the logarithm of the conditional variance depends now on the first lag of the conditional variance and the first lag of a realized measure x_t. In order to avoid a measurement error in x_t, equation (22) relates the realized measure to the conditional variance of the returns. The volatility process is stable as long as $\beta_1 + \delta\alpha_1 \in (-1,1)$. The leverage function $\tau(z)$ is responsible for the asymmetric reaction to return shocks. Hansen, Huang, and Shek (2012) showed that without loss of generality it could be assumed that $E[\tau(z_t)] = 0$. This ensures that the leverage function is described by a simple quadratic function on the basis of Hermite polynomials given by

$$\tau(z_t) = \eta_1 z_t + \eta_2 (z_t^2 - 1) \tag{23}$$

The authors also highlighted the relevance of the news impact curve that was presented by Engle and Ng (1993) which is very similar in terms of content to the leverage function $\tau(z)$. The news impact curve of the realized GARCH model is defined by

$$\nu(z) = E[\log(\sigma_{t+1}^2) | z_t = z] - E[\log(\sigma_{t+1}^2)] = \alpha_1 \tau(z), \tag{24}$$

so that $100 * \nu(z)$ estimates the percentage impact on volatility as a function of z (Hansen, Huang, and Shek, 2012). In this thesis, I use two different realized measures, i.e. $x_t \in \{realized\ variance\ (RV), MedRV\}$. The realized variance in period t is defined as

$$RV_t = \sum_{i=1}^{N} r_{t,i}^2, \tag{25}$$

where $r_{t,i} = \log(P_{t,i}) - \log(P_{t,i-1}), i = 2, \ldots, N$ and $\sum_{i=1}^{N} r_{t,i}^2$ are N squared returns over the entire day.

Furthermore, $N = {}^{1}/_{\Delta}$, where Δ denotes the sampling frequency. $P_{t,i}$ is the *i*th closing price of BTC in period t. For calculating the RV as stated in equation (25) I use 5-min high-frequency data as suggested by Andersen et al. (2000).

The second realized measure is the median realized variance (MedRV) estimator introduced by Andersen, Dobrev, and Schaumburg (2012) that is robust to price jumps and to the presence of zero intraday returns in finite samples. The estimator can be calculated with the following formula:

$$\mathrm{MedRV}_t = \frac{\pi}{6 - 4\sqrt{3} + \pi} * \left(\frac{N}{N-2}\right) * \sum_{i=2}^{N-1} med(|\Delta P_{i-1}|, |\Delta P_i|, |\Delta P_{i+1}|)^2 \tag{26}$$

5.1.2. HAR-RV models

Using high-frequency data and measuring the realized variance ex-post has become a popular research field in the early 2000s. One of the most popular estimators for the ex-post variance of an asset is the realized variance estimator given in (25). Modelling the RV was very challenging in the beginning since it is difficult to account for the long memory property of the ex-post measure of the return variance with simple ARIMA models. One solution to overcome this issue is to use so-called autoregressive fractionally integrated moving average (ARFIMA) models, which can handle the exact long memory of time series. In contrast to the simple ARIMA models, the differencing parameter d of ARFIMA models is a real number instead of an integer. However, one disadvantage of such models is that they are relatively difficult to estimate especially the estimation of d is a demanding task. Corsi (2009) developed the first simple long memory model for estimating and forecasting the realized ex-post variance. The Heterogeneous Autoregressive model of Realized Volatility (HAR-RV) is based on the Heterogenous Market Hypothesis by Müller et al. (1997) which claims that agents have heterogeneous preferences in terms of their time horizon of their investment decisions. On the one hand, there are dealers and market makers who trade on a daily basis and on the other hand, there are insurance companies and pension funds that trade at a lower frequency (Corsi, 2009). Therefore, agents react to new information differently and as a result create volatility. Corsi (2009) exploits these assumptions by assuming that the daily realized volatility depends on the last daily, weekly and monthly realized volatility caused by short-term, medium-term and long-term investors, respectively. Of course, many extensions of the model have emerged and were intensively discussed in the literature. As a next step, the theoretical framework of the simple HAR-RV model is described.

Let p_t denote the logarithmic price of Bitcoin in period t. Then the diffusion process is given by the following stochastic differential equation

$$dp_t = \mu_t dt + \sigma_t dW_t, \tag{27}$$

where μ_t describes a drift with a finite and continues variation process, σ_t is a cadlág stochastic volatility process independent of W_t and W_t denotes a standard Brownian motion. The integrated variance (IV) which is equivalent to the quadratic variation (QV) of this process for one trading day is then defined by

$$QV_t = IV_t = \int_{t-1}^{t} \sigma_s^2 ds. \tag{28}$$

It can be shown that the integrated variance can be consistently estimated using the realized variance (RV) from equation (25) if the number of squared intraday returns goes to infinity, which is equivalent to $\Delta \rightarrow 0$. Formally,

$$\underset{N\rightarrow\infty}{\text{plim}}\, RV_t = IV_t = QV_t. \tag{29}$$

However, this simple framework does not incorporate any jumps in the price process. But Hung, Liu, and Yang (2020) found that Bitcoin is very prone to jumps. Therefore, as a next step I consider a jump-diffusion model which was introduced by Merton (1976). Now the jump diffusion process is a Brownian semimartingale given by

$$\mathrm{d}p_t = \mu_t dt + \sigma_t dW_t + \kappa_t dq_t, \tag{30}$$

where q_t is a Poisson process counting the number of jumps in the process and κ_t measures the magnitude of the corresponding discrete jumps. The quadratic variation of this process has now two components, namely

$$QV_t = IV_t + JV_t \tag{31}$$

or formally

$$QV_t = \int_{t-1}^{t} \sigma_s^2 ds + \sum_{t-1<s\leq t} \kappa_s^2\,, \tag{32}$$

where $\sum_{t-1<s\leq t} \kappa_s^2$ is the jump variation (JV) or discontinuous variation which is simply the sum of squared jump sizes for a given period. IV_t is the same as in (28).

Even though the quadratic variation now has two components the realized variance estimator is still consistent in a sense that if the sample points within period t approach infinity

$$\operatorname*{plim}_{N\to\infty} RV_t = QV_t = IV_t + JV_t. \tag{33}$$

As a result, RV_t contains the continuous variation and the discontinuous variation of the jumps in the price process. One may be interested in estimating just the continuous part (IV_t) of the process. Barndorff-Nielsen and Shephard (2004) introduced the realized bipower variation estimator given by

$$BV_t = \mu_1^{-2} \sum_{i=2}^{N} |r_{t,i}||r_{t,i-1}|, \tag{34}$$

where $\mu_p = E[|Z|^p] = 2^{p/2}\frac{\Gamma((1/2)(p+1))}{\Gamma(1/2)}, Z \sim N(0,1), p \geq 0$ and $\Gamma(.)$ denotes the Gamma function. Given these assumptions $\mu_1 = \sqrt{\frac{2}{\pi}}$, which reduces (34) to

$$BV_t = \frac{\pi}{2}\sum_{i=2}^{N} |r_{t,i}||r_{t,i-1}|, \tag{34a}$$

The authors show that in the presence of jumps $\operatorname*{plim}_{N\to\infty} BV_t = IV_t$, which is an important property. However, the estimator is not robust to different kinds of microstructure noise. To account for this issue the modified version of the realized bipower variation estimator suggested by Andersen, Bollerslev, and Huang (2008) is used. The modified estimator is defined by

$$BV_t = \frac{\pi}{2}\frac{N}{N-2}\sum_{i=3}^{N} |r_{t,i}||r_{t,i-2}|, \tag{35}$$

Consequently, the difference between the realized variance and the realized bipower variation consistently measures the jump variation (JV), which means in mathematical terms

$$\operatorname*{plim}_{N\to\infty}(RV_t - BV_t) = JV_t = \sum_{t-1<s\leq t} \kappa_s^2 \tag{36}$$

.

Of course, any other consistent estimator of the integrated variance in the presence of jumps can be used. The MedRV estimator given by (26) is an alternative to the realized bipower variation estimator as it was shown by Andersen, Dobrev, and Schaumburg (2012) that in the presence of jumps

$$\operatorname*{plim}_{N\to\infty} MedRV_t = IV_t. \tag{37}$$

In the following, I also use this newer jump robust estimator when estimating the discontinuous part of the realized variance. To ensure the positivity of the jump variation I apply the max-function and define the discontinuous jump variation as

$$JV_t = \max\{RV_t - MedRV_t, 0\} \tag{38}$$

Equation (38) would imply that there is a non-negative jump variation every day, which is not plausible since there should be significant and insignificant jumps in the price process. To detect significant price jumps the JO Jump test by Jiang and Oomen (2008) is performed to test for the presence of jumps in the high-frequency price series of Bitcoin. Theodosiou and Zikes (2011) compared several tests for jumps in the price series and they found that the JO jump test has the highest power among all other tests considered at a 5-min sampling frequency. In addition, the swap variance test by Jiang and Oomen (2008) also performs well in the presence of zero intraday returns because of the usage of the integrated sixticity in the denominator of the test statistic (Theodosiou and Zikes, 2011). Jiang and Oomen (2008) found that the accumulated difference of the simple arithmetic return $R_{i,n}$ and log-return $r_{i,n}$ should equal half the realized variance in the absence of jumps. If this difference is too large (in absolute terms), this would indicate the existence of jumps. In the literature, of course, there exist various different jump tests (see Aït-Sahalia and Jacod (2009) and Barndorff-Nielsen and Shephard (2006) for instance)

The null and alternative hypothesis of the JO jump test is given by:
H_0: There are no jumps in period t
H_A: There is at least one jump in period t
If there are N equispaced returns in period t and $N \to \infty$, the test statistic of the ratio test is given by

$$JOjumpTest_{t,N} = \frac{NBV_t}{\sqrt{\Omega_{S\omega V_t}}}\left(1 - \frac{RV_t}{S\omega V_t}\right) \xrightarrow{d} N(0,1), \tag{39}$$

where $S\omega V_t = 2\sum_{i=1}^{N}(R_{t,i} - r_{t,i})$, BV is the bipower variation given by (19) and RV denotes the realized variance stated in equation (9). Furthermore

$$\Omega_{S\omega V_t} = \frac{\mu_6}{9}\frac{N^3\mu_{6/p}^{-p}}{N-p+1}\sum_{i=0}^{N-p}\prod_{k=1}^{p}\left|r_{t,i+k}\right|^{6/p}$$

is an estimator of the integrated sixticity, $\int_{t-1}^{t}\sigma_s^6 ds$ and

$$\mu_p = E[|Z|^p] = \ 2^{p/2}\frac{\Gamma(1/2)(p+1)}{\Gamma(1/2)},$$

where Z are independent standard normal random variables and p is a parameter (power) and Γ(.) denotes the Gamma function. The estimator of the integrated sixticity is jump-robust for $p = 4$ and $p = 6$ (Jiang and Oomen, 2008). However, using higher powers of returns can make the estimator upward-biased, which leads to a deterioration of the power of the test (Jiang and Oomen, 2008). That is why I use $p = 4$ in this paper. It is also noteworthy that Jiang and Oomen (2008) found that using either the realized quadpower or sixthpower sixticity estimator makes little difference, which is also in line with the findings of Theodosiou and Zikes (2011).

Now it is possible to distinguish the continuous and discontinuous variation of the realized variance. Formally, the continuous part is given by

$$C_t = \mathrm{I}\left(JOjumpTest_{t,N} \leq \Phi_\alpha\right) * RV_t + \mathrm{I}\left(JOjumpTest_{t,N} > \Phi_\alpha\right) * MedRV_{t,} \tag{40}$$

where $\mathrm{I}(.)$ is the indicator function and Φ_α is the α-quantile of the standard normal distribution.

The corresponding discontinuous jump variation of the quadratic variation is defined as

$$J_t = \mathrm{I}\left(JOjumpTest_{t,N} > \Phi_\alpha\right) * \max\{RV_t - MedRV_t, 0\} \tag{41}$$

In this thesis, I use the 5% significance level for identifying statistically significant jumps.

5.1.2.1. Standard HAR-RV model

The simplest version of all HAR models was introduced by Corsi (2009) and assumes that the price process of Bitcoin is generated by equation (27) which implies a continuous price process and the absence of any jumps. Then, the HAR-RV model has the following form:

$$RV_t^{(d)} = \beta_0 + \beta_1 RV_{t-1}^{(d)} + \beta_2 RV_{t-1}^{(w)} + \beta_3 RV_{t-1}^{(m)} + \varepsilon_t, \tag{42}$$

where ε_t is a mean-zero error term and $RV_{t-1}^{(d)}, RV_{t-1}^{(w)} and\ RV_{t-1}^{(m)}$ are the corresponding first lag of daily, average weekly and average monthly realized variances which are defined as

$$RV_{t-1}^{(w)} = \frac{1}{7} * (RV_{t-7} + RV_{t-6} + \cdots + RV_{t-1})$$
$$RV_{t-1}^{(m)} = \frac{1}{30} * (RV_{t-30} + RV_{t-29} + \cdots + RV_{t-1})$$

In contrast to many other papers, I use seven days and 30 days when computing the average weekly and monthly realized variance, respectively. The reason for this is that BTC can be traded seven days a week and 24 hours a day since there do not exist any trading hours and trading days as it is the case for typical stock exchanges. Apart from that, there exist ample studies, which show that the typical assumption of a homoscedastic and normally distributed error term is violated, therefore I also use the logarithmic transformation of equation (42) as suggested by Corsi (2009). Using the logarithm of the realized variance has two advantages. On the one hand it ensures the positivity of the partial variances and on the other hand it reduces the heteroscedasticity of the error term and the assumption $\varepsilon_t \sim N(0, \sigma_\epsilon^2)$ holds approximately due to the log-normal property of the realized variance. The logarithmic version takes the form

$$\log\left(RV_t^{(d)}\right) = \beta_0 + \beta_1 \log\left(RV_{t-1}^{(d)}\right) + \beta_2 \log\left(RV_{t-1}^{(w)}\right) + \beta_3 \log\left(RV_{t-1}^{(m)}\right) + \varepsilon_t \tag{43}$$

5.1.2.2. HAR-RV-J model

The HAR-RV-J model was first introduced by Andersen, Bollerslev, and Diebold (2007) and in contrast to the HAR-RV model the HAR-RV-J model does not assume a continuous price process but a jump diffusion process given by (30). This model uses an additional explanatory variable when predicting the daily realized variance. The HAR-RV-J is defined as

$$RV_t^{(d)} = \beta_0 + \beta_1 RV_{t-1}^{(d)} + \beta_2 RV_{t-1}^{(w)} + \beta_3 RV_{t-1}^{(m)} + \beta_4 J_{t-1}^{(d)} + \varepsilon_t, \quad (44)$$

where $J_{t-1}^{(d)}$ is a jump component estimated by (41). Applying a log-transformation of (44) is a little bit more complicated because there are also days with no jump and hence, the jump component should be zero. To deal with this issue Andersen, Bollerslev, and Diebold (2007) suggest the following logarithmic HAR-RV-J model

$$\log\left(RV_t^{(d)}\right) = \beta_0 + \beta_1 \log\left(RV_{t-1}^{(d)}\right) + \beta_2 \log\left(RV_{t-1}^{(w)}\right) + \beta_3 \log\left(RV_{t-1}^{(m)}\right) + \beta_4 \log\left(1 + J_{t-1}^{(d)}\right) + \varepsilon_t \quad (45)$$

5.1.2.3. HAR-RV-CJ model

The next model is a further extension of the standard HAR-RV model again introduced by Andersen, Bollerslev, and Diebold (2007). The main idea of the HAR-RV-CJ model is to use the property of the realized variance and decompose the RV into its continuous and discontinuous component like described in (40) and (41). After doing this there are now six explanatory variables instead of the three in HAR-RV model. Formally, the HAR-RV-CJ model is defined as

$$RV_t^{(d)} = \beta_0 + \beta_1 C_{t-1}^{(d)} + \beta_2 C_{t-1}^{(w)} + \beta_3 C_{t-1}^{(m)} + \beta_4 J_{t-1}^{(d)} + \beta_5 J_{t-1}^{(w)} + \beta_6 J_{t-1}^{(m)} + \varepsilon_t, \quad (46)$$

where

$$C_{t-1}^{(w)} = \frac{1}{7} * (C_{t-7} + C_{t-6} + \cdots + C_{t-1})$$

$$J_{t-1}^{(w)} = \frac{1}{7} * (J_{t-7} + J_{t-6} + \cdots + J_{t-1})$$

$$C_{t-1}^{(m)} = \frac{1}{30} * (C_{t-30} + C_{t-29} + \cdots + C_{t-1})$$

$$J_{t-1}^{(m)} = \frac{1}{30} * (J_{t-30} + J_{t-29} + \cdots + J_{t-1})$$

The logarithmic HAR-RV-CJ model takes the form

$$\log\left(RV_t^{(d)}\right) = \beta_0 + \beta_1 \log\left(C_{t-1}^{(d)}\right) + \beta_2 \log\left(C_{t-1}^{(w)}\right) + \beta_3 \log\left(C_{t-1}^{(m)}\right) + \beta_4 \log\left(1 + J_{t-1}^{(d)}\right) + \beta_5 \log\left(1 + J_{t-1}^{(w)}\right) + \beta_6 \log\left(1 + J_{t-1}^{(m)}\right) + \varepsilon_t \quad (47)$$

5.2. Evaluation of the model performances

5.2.1. Methodology

For the evaluation, the data set is divided into a training set from 31st Mar 2015 to 31st Mar 2020 containing 1827 data points and a test set of the last year of the data set (1st Apr 2020 to 31st Mar 2021) containing 365 data points. In total sixteen different models are estimated and subject to an in-sample and out-of-sample performance evaluation. The following ten GARCH models are estimated:

- Standard GARCH with normally distributed standardized error terms (sGARCH-norm)
- Standard GARCH with Student-t distributed standardized error terms (sGARCH-std)
- EGARCH with normally distributed standardized error terms (EGARCH-norm)
- EGARCH with Student-t distributed standardized error terms (EGARCH-std)
- GJR-GARCH with normally distributed standardized error terms (GJR-GARCH-norm)
- GJR-GARCH with Student-t distributed standardized error terms (GJR-GARCH-std)
- Realized GARCH with normally distributed standardized error terms and realized variance as realized measure (RealGARCH-RV-norm)
- Realized GARCH with Student-t distributed standardized error terms and realized variance as realized measure (RealGARCH-RV-std)
- Realized GARCH with normally distributed standardized error terms and median realized variance as realized measure (RealGARCH-MedRV-norm)
- Realized GARCH with Student-t distributed standardized error terms and median realized variance as realized measure (RealGARCH-MedRV-std)

Furthermore, six different HAR-RV models are estimated, containing

- HAR-RV
- Log-HAR-RV
- HAR-RV-J
- Log-HAR-RV-J
- HAR-RV-CJ
- Log-HAR-RV-CJ

For comparing the in-sample fit of the models four different information criteria including the Akaike (AIC), Bayesian (BIC), Hannan-Quinn (HQIC) and Shibata (SIC) information criteria are used as well as the maximum value of the Log-Likelihood function. The model with the lowest information criteria and highest Log-Likelihood is considered as the best model. Formally, the information criteria are defined as follows:

$$AIC = \frac{-2LL}{N} + \frac{2m}{N} \tag{48}$$

$$BIC = \frac{-2LL}{N} + \frac{m * \log(N)}{N} \tag{49}$$

$$HQIC = \frac{-2LL}{N} + \frac{2m * \log(\log(N))}{N} \tag{50}$$

$$SIC = \frac{-2LL}{N} + \log\left(\frac{N+2m}{N}\right), \tag{51}$$

where LL denotes the maximum value of the Log-Likelihood function, N is the number of observations and m stands for the number of parameters.

However, this procedure does not make sense when comparing GARCH models with the HAR-RV models since the conditional variance of daily returns estimated from GARCH models might not, on average, correspond to the realized measures estimated from intraday data. Even though these two types of models are not directly comparable because they do not exactly predict the same volatilities, in order to perform an in-sample comparison I use the mean squared error (MSE) and QLIKE loss function.

When performing the forecasting analysis, I conduct a one-step ahead forecast where I estimate the model with the training data and re-estimate the model after each shift in start and end of the estimation interval. For evaluating the performance, the MSE and QLIKE loss functions are used as these two are robust when evaluating the out-of-sample performance by divergence between the predicted values with the volatility proxy of an observable variable like it is the case with GARCH models (Patton, 2011).The two loss functions are defined by

$$MSE = \frac{1}{N}\sum_{t=1}^{N}\left(\sigma_t^2 - \widehat{\sigma_t^2}\right)^2 \tag{52}$$

$$QLIKE = \frac{1}{N}\sum_{t=1}^{N}\left(\frac{\sigma_t^2}{\widehat{\sigma_t^2}} - \log\left(\frac{\sigma_t^2}{\widehat{\sigma_t^2}}\right) - 1\right) \tag{53}$$

Moreover, for an out-of-sample robustness check the realized variance and the MedRV serve as a proxy for the true variance σ_t^2 in order to compare GARCH and HAR models. In Panel A of table 27, the HAR-RV is considered as a benchmark model by looking whether any other model outperforms the standard HAR-RV model when applying the Diebold-Mariano test at the 5% significance level. Panel B uses the MedRV as a variance proxy and here I evaluate which model is outperformed by the RealGARCH-MedRV-std. Again, also the Diebold-Mariano test at the same significance level is applied. The Diebold-Mariano test was introduced by Diebold and Mariano (1995). Two years later the modified version suggested by Harvey, Leybourne, and Newbold (1997) was published. The newer version is also implemented in this thesis. The test works as follows: Let $d_t = L\left(\hat{x}_t^{(1)}\right) - L\left(\hat{x}_t^{(2)}\right)$, where $L\left(\hat{x}_t^{(i)}\right)$ is an out-of-sample loss function, i.e. $L\left(\hat{x}_t^{(i)}\right) \in \{MSE, QLIKE\}$ for forecast $\hat{x}_t^{(i)}$ obtained with model $\mathcal{M}_i$. Assuming d_t is covariance-stationary and has finite moments, the null hypothesis is equal predictive accuracy or formally, $H_0: E(d_t) = 0$. Moreover, the test statistic is given by

$$DM = \frac{\frac{1}{N}\sum d_t}{\sqrt{\hat{\sigma}_d^2}} \xrightarrow{d} N(0,1), \tag{54}$$

where $\hat{\sigma}_d^2$ is a long-run variance estimate of $\frac{1}{N}\sum d_t$. Further details of the long-run variance estimator are provided in Diebold and Mariano (1995) and Harvey, Leybourne, and Newbold (1997). If DM is significantly <0 (>0), one can conclude that model $\mathcal{M}_1$ is better (worse) than $\mathcal{M}_2$.

Furthermore, I use the Mincer-Zarnowitz test proposed by Mincer and Zarnowitz (1969) for out-of-sample evaluation by using the same two proxies mentioned previously. The test works as follows:

1. Estimate the simple OLS regression (Mincer-Zarnowitz regression): $\theta_t = \alpha + \beta\widehat{\theta_t} + e_t$, where θ_t denotes the true value and the regressor $\widehat{\theta_t}$ describes the value of the forecast
2. Test the joint hypothesis $H_0: \alpha = 0, \beta = 1$ by using an F-test and focus on the R^2
3. If the null hypothesis gets rejected the forecast is considered as being biased and inefficient

However, as pointed out by Andersen and Bollerslev (1998) θ_t is often subject to an estimation error which may cause a biased estimate for $\widehat{\theta_t}$. To overcome this problem the authors suggest concentrating on the R^2 of the Mincer-Zarnowitz regression instead.

5.2.2. Empirical results

5.2.2.1. Jump test results

The following table shows the numbers of rejections of the null hypothesis for different significance levels:

Table 22: JO Jump test results

Number of days: 2192			
Significance level	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 0.1\%$
Number of rejections	713	537	392

The results indicate that there are many rejections of the null, which suggests the usage of a jump-robust estimator like the MedRV estimator stated in equation (26). Even with a significance level of 0.1% the null hypothesis is rejected 392 times which equates approximately every 5th trading day.

5.2.2.2. In-sample results

The following two tables present the point estimates of all ten GARCH models and the six HAR-RV models. In addition, the corresponding R^2 of all HAR-RV models is also presented in table 24. For making a possible comparison between GARCH and HAR models, table 25 exhibits the in-sample MSE and QLIKE of all sixteen models.

Table 23: In-sample fit (GARCH Models)

Model	*Point estimates*									
	μ	ω	α_1	β_1	γ_1	δ	ξ	η_1	η_1	λ
sGARCH-norm	0.0022**	0.0001***	0.206***	0.791***						
sGARCH-std	0.0017***	0.00002***	0.131***	0.868***						
EGARCH-norm	0.0016*	-0.4501***	-0.058***	0.925***	0.331***					
EGARCH-std	0.0016***	-0.06	0.053**	0.99***	0.312***					
GJR-GARCH-norm	0.0015**	0.0001***	0.159***	0.779***	0.103***					
GJR-GARCH-std	0.0017***	0.00001***	0.146***	0.879***	-0.052**					
RealGARCH-RV-norm	0.0016**	-0.988***	0.406***	0.432***		1.22***	1.237***	-0.028*	0.065***	0.615***
RealGARCH-RV-std	0.0016***	0.2538	0.508***	0.456***		0.953***	-1.152***	-0.061**	0.108***	0.624***
RealGARCH-MedRV-norm	0.0014	-1.299***	0.378***	0.405***		1.373***	2.111***	-0.024	0.055***	0.606***
RealGARCH-MedRV-std	0.0015***	0.390	0.554***	0.409***		0.948***	-1.345***	-0.046**	0.106***	0.601***

*Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level and *** denotes significance at the 1% level, Values in parenthesis are HAC standard errors to account for the presence of autocorrelation and heteroscedasticity in the error terms*

Table 24: In-sample fit (HAR Models)

Model	*Point estimates and R^2*							
	β_0	β_1	β_2	β_3	β_4	β_5	β_6	R^2
HAR-RV	0.0005*** (0.0002)	0.3885*** (0.0668)	0.1559*** (0.0457)	0.2401*** (0.0825)				0.267
Log-HAR-RV	-0.7769*** (0.1447)	0.5045*** (0.0281)	0.2737*** (0.0339)	0.1213*** (0.0398)				0.6227
HAR-RV-J	0.0005*** (0.0002)	0.4293*** (0.0846)	0.1645*** (0.0472)	0.2344*** (0.0808)	-0.5656* (0.3317)			0.2716
Log-HAR-RV-J	-0.5905*** (0.1453)	0.5425*** (0.0294)	0.2645*** (0.0342)	0.1156*** (0.0388)	-100.58*** (27.4939)			0.6253
HAR-RV-CJ	0.0006*** (0.0002)	0.4103*** (0.0702)	0.1267** (0.0575)	0.2908*** (0.0814)	-0.0098 (0.3891)	0.5187 (0.5490)	-1.0096 (0.8421)	0.2699
Log-HAR-RV-CJ	-0.6903*** (0.1592)	0.5258*** (0.0270)	0.2358*** (0.0360)	0.1368*** (0.0437)	15.219 (38.2334)	-2.4985 (103.1854)	-49.4889 (161.6001)	0.628

*Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level and *** denotes significance at the 1% level, Values in parenthesis are HAC standard errors to account for the presence of autocorrelation and heteroscedasticity in the error terms*

Table 25: In-sample results – Loss functions

	Panel A: RV as proxy		Panel B: MedRV as proxy	
Model	*MSE*	*QLIKE*	*MSE*	*QLIKE*
sGARCH-norm	**0.1451**	0.3653	0.1446	0.3610
sGARCH-std	0.1585	0.4748	0.1604	0.4531
EGARCH-norm	0.1784	0.3783	0.1579	0.3770
EGARCH-std	0.1815	0.4396	0.1930	0.4796
GJR-GARCH-norm	0.1454	0.3773	**0.1427**	0.3712
GJR-GARCH-std	0.1639	0.4868	0.1666	0.4625
RealGARCH-RV-norm	0.1666	**0.3480**	0.1734	**0.3348**
RealGARCH-RV-std	0.3143	0.4810	0.3453	0.5280
RealGARCH-MedRV-norm	0.1661	0.3517	0.1724	0.3353
RealGARCH-MedRV-std	0.4083	0.5072	0.4426	0.5542
HAR-RV	0.1580	0.3529		
Log-HAR-RV	0.1606	0.3924		
HAR-RV-J	0.1570	0.3576		
Log-HAR-RV-J	0.1606	0.3907		
HAR-RV-CJ	0.1574	0.3571		
Log-HAR-RV-CJ	0.1586	0.3908		

Notes: The MSE is multiplied by 10^5

Table 23 shows the point estimates of all ten considered GARCH models. Almost all coefficients are significant at least at the 5% level. Furthermore, the conditional mean of all models is very similar and close to zero and all models are stable (asymptotically stationary). It is also interesting that both RealGARCH models with t-distributed innovations put more weight on the realized measure (α_1 = 0.508 and 0.554) than when considering normally distributed returns (α_1 = 0.406 and 0.378). These results suggest that the RealGARCH-RV-std and RealGARCH-MedRV-std put more weight on the contemporaneous signal and less weight on the historical signals, which should lead to a faster reaction to a sudden change in the volatility. Therefore, the realized GARCH models with t-distributed standardized returns should better estimate the conditional variance in case of a huge volatility shock like it was the case at the start of the COVID-19 crisis.

The results in table 24 indicate that the log transformation of any HAR model yields a much higher R^2 than the corresponding model with levels. In addition, the descending order of the parameters with increasing time horizon ($\beta_1 > \beta_2 > \beta_3$) only holds in the logarithmic versions. This order seems natural, as the first lag of the daily realized variance should have more impact on the current daily-realized variance than the past weekly average RV and the past weekly average RV should have more weight than the first lag of the monthly average RV. Moreover, all coefficients of the continuous component of the realized variance are highly significant. It is also striking that the parameter of the jump component in the HAR-RV-J is only significant at the 10% significance level while β_4 in the Log-HAR-RV-J model is significant at the 1% level. Another surprising result is that the parameters of the jump variation in the HAR-RV-CJ and Log-HAR-RV-CJ are all insignificant with relatively high standard errors. A possible explanation would be that this is due to the usage of Newey–West standard errors and due to the fact that Bitcoin is traded all the time and there is no overnight component like it is the case with stocks traded at a stock exchange.

Table 25 allows making a potential performance evaluation in-sample by comparing the two different types of volatility models with robust loss functions. The in-sample results are quite interesting since it seems that the GARCH family outperforms the HAR family. Panel A indicates that the sGARCH-norm has the lowest MSE and the RealGARCH-RV-norm the lowest QLIKE. Panel B of table 25 shows similar results, again, the RealGARCH-RV-norm has the lowest QLIKE but in contrast to Panel A the GJR-GARCH-norm has the smallest MSE. This is very interesting because one would expect that the HAR-RV models could better capture the long memory of the realized variance because the GARCH models only estimate the conditional variance of the daily log-returns and not a realized measure based on high-frequency data. Whether these findings also hold out-of-sample remains a topic for the next chapter.

Table 26 presents the values of the four different information criteria and the maximum value of the Log-Likelihood function (LL) for the ten estimated GARCH models. In order to compare classical GARCH models with the realized GARCH models I only use the partial log-likelihood of the realized GARCH models. All four considered information criteria suggest the EGARCH-std as the model with the best in-sample fit because it has the smallest AIC, BIC, HQIC, SIC and the highest Log-Likelihood value. It is interesting that the realized GARCH models have a worse in-sample performance than the other GARCH models without a realized measure.

Table 26: Information criteria and Log-Likelihood

Model	*AIC*	*BIC*	*HQIC*	*SIC*	*LL*
sGARCH-norm	-3.824	-3.8115	-3.8191	-3.8236	3496.9
sGARCH-std	-4.141	-4.1254	-4.1349	-4.1405	3787.3
EGARCH-norm	-3.8257	-3.8106	-3.8201	-3.8257	3499.8
EGARCH-std	**-4.1624**	**-4.1444**	**-4.1558**	**-4.1625**	**3808.4**
GJR-GARCH-norm	-3.8287	-3.8136	-3.8231	-3.8287	3502.5
GJR-GARCH-std	-4.1418	-4.1237	-4.1352	-4.1419	3789.6
RealGARCH-RV-norm	-3.7768	-3.7496	-3.7668	-3.7768	3459.1
RealGARCH-RV-std	-4.1233	-4.0961	-4.1132	-4.1233	3775.6
RealGARCH-MedRV-norm	-3.7585	-3.7314	-3.7485	-3.7586	3442.4
RealGARCH-MedRV-std	-4.1123	-4.0852	-4.1023	-4.1124	3765.6

5.2.2.3. Out-of-sample results

Table 27: Out-of-sample results – Loss functions

	Panel A: RV as proxy		*Panel B: MedRV as proxy*	
Model	*MSE*	*QLIKE*	*MSE*	*QLIKE*
sGARCH-norm	0.0527	0.3041	0.0524	0.3318**
sGARCH-std	0.0548	0.3068	0.0541	0.3189**
EGARCH-norm	0.0536	0.3358	0.0527	0.3619**
EGARCH-std	0.0697	0.4244	0.0749**	0.4975**
GJR-GARCH-norm	0.0539	0.3295	0.0532	0.3507**
GJR-GARCH-std	0.0542	0.2879	0.0537	0.3035**
RealGARCH-RV-norm	0.0711	0.4767	0.0749**	0.5502**
RealGARCH-RV-std	0.0493	0.2490*	0.0487	0.2537**
RealGARCH-MedRV-norm	0.0993	0.5719	0.1050**	0.6558**
RealGARCH-MedRV-std	0.0455	**0.2399***	**0.0434**	**0.2203**
HAR-RV	0.0475	0.3136		
Log-HAR-RV	0.0462	0.2665*		
HAR-RV-J	0.0502	0.9138		
Log-HAR-RV-J	0.0463	0.3131		
HAR-RV-CJ	0.0476	0.3171		
Log-HAR-RV-CJ	**0.0451***	0.2594*		

Notes: The MSE is multiplied by 10^5, * in Panel A denotes that the MSE/QLIKE of the corresponding model is significantly less than the MSE/QLIKE of the HAR-RV model when applying the Diebold-Mariano test at the 5% significance level, ** in Panel B denotes that the MSE/QLIKE of the corresponding model is significantly greater than the MSE/QLIKE of the RealGARCH-MedRV-std model when applying the Diebold-Mariano test at the 5% significance level

The out-of-sample results in table 27 yield interesting insights. First, the realized GARCH models with t-distributed error terms outperform all other considered GARCH models. This holds for Panel A as well as for Panel B. Second, the RealGARCH-MedRV-std has a lower MSE and QLIKE than the RealGARCH-RV-std, which means that a jump-robust realized measure like the MedRV seems to provide a more accurate forecast. Third, the EGARCH-std with the best in-sample fit in terms of information criteria has the third highest MSE and QLIKE and hence a weak out-of-sample performance. This result shows that a good in-sample fit does not necessary go hand in hand with a decent forecasting performance out-of-sample. Fourth, Panel A indicates that the Log-HAR-RV-CJ has the smallest MSE and is the only model that has a statistically significant smaller MSE than the HAR-RV. When doing the same exercise with the QLIKE loss function I found four models that outperform the HAR-RV, namely both realized GARCH models with t-distributed innovations, the Log-HAR-RV and the Log-HAR-RV-CJ. Fifth, the findings in Panel B show that the RealGARCH-MedRV-std has the smallest MSE and QLIKE but the difference of the MSE is only significant for three GARCH models. However, when considering the QLIKE loss function the RealGARCH-MedRV-std outperforms all other GARCH models at the 5% significance level. Finally, the more recent realized GARCH model with t-distributed error terms is the only GARCH model that can compete with the HAR models, especially the RealGARCH-MedRV-std that surprisingly also has the smallest QLIKE across all models. Because the QLIKE loss function is asymmetric and hence penalizes negative deviations from the true RV much more than positive deviation in contrast to the symmetric MSE loss function it is likely that the RealGARCH-MedRV-std systematically overestimates the RV, which leads to a smaller QLIKE. Figure 32 illustrates the two (and five more) loss functions graphically. Furthermore, as expected the log specification of any HAR model yields a better forecasting performance than using levels. Within the HAR models, the Log-HAR-RV-CJ has the smallest MSE and QLIKE. This finding suggests that the decomposition of the realized variance in its continuous and discontinuous components, as it is the case with the HAR-RV-CJ model results in a better forecasting performance even though the three jump components are all insignificant, which implies that the continuous part of the realized variance is the main driver in explaining the realized variance. Several plots illustrating and comparing the forecasting performance of the models can be found in the appendix.

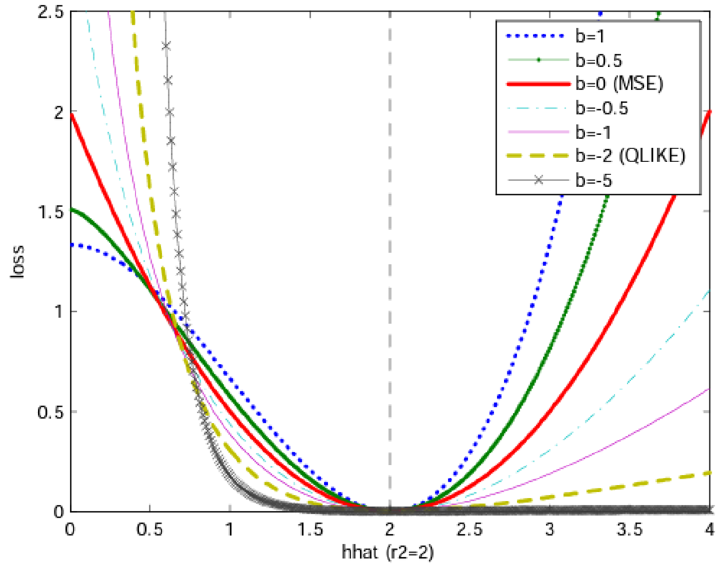


Figure 32: Robust loss functions for different choices of b (Source: (Patton, 2011, 252))

The results of the Mincer-Zarnowitz test in table 28 are little ambiguous and unexpected. On the one hand, the null hypothesis in Panel A is rejected eleven out of sixteen times and six out of ten times in Panel B at the 5% significance level, which would imply that the forecasts are predominantly biased. On the other hand, the models with the highest R^2, and hence the models with best forecasting performance in terms of explaining the variation of dependent variable, are all models where the null is rejected. In Panel A for example, the R^2 of the three GARCH models where the null cannot be rejected is relatively small in contrast to other GARCH models with a p-value of approximately zero. Apart from that, I found that the RealGARCH-MedRV-std with a R^2 of 0.3864 has the overall highest R^2 in Panel A. This is a very strong result since it suggests that RealGARCH-MedRV-std outperforms all GARCH and HAR models in terms of R^2 even though the null is rejected at a very small significance level and $\beta = 0.71$ which is far away from 1 and implies that the model yields, on average, systematically overestimated forecasts. A similar pattern can be observed within the HAR-type models. Most of the models where the null hypothesis cannot be rejected have a significantly smaller R^2 than HAR models with a relatively small p-value. The Log-HAR-RV-CJ for example has a p-value smaller than 1% but at the same time the highest R^2 across all HAR models. In conclusion, the results in Panel A indicate that the null of every single of the three models with a R^2 above 30% is rejected at the 5% level. Panel B of table 28 yields identical results because again the RealGARCH-MedRV-std outperforms all other GARCH models in terms of R^2 ($\mathrm{R}^2 = 0.4111$) but produces biased forecasts ($\beta = 0.723$). The R^2 of the four GARCH models with a p-value above 5% are about 50% smaller than the R^2 of the RealGARCH-MedRV-std. The bottom line is that the biased forecasts (forecasts with low p-values) tend to have a higher R^2, which is a very surprising result since one would expect that the models with unbiased forecasts are more likely to yield a higher R^2, but the results presented in table 28 demonstrate the opposite. This outcome is consistent with Andersen and Bollerslev (1998) as described in section 5.2.1. Therefore, it makes sense to concentrate on the R^2 and consequently the RealGARCH-MedRV-std has proven to be the best forecasting model when applying the Mincer-Zarnowitz test.

Table 28: Out-of-sample results - Mincer-Zarnowitz test

	Panel A: RV as proxy				*Panel B: MedRV as proxy*			
	α	β	p-value H_0:α=0, β=1	R^2	α	β	p-value H_0:α=0, β=1	R^2
Model								
sGARCH-norm	-0.0001 (0.0002)	1.202 (0.1276)	0.0857	0.1963	-0.0002 (0.0002)	1.103 (0.1145)	0.5778	0.2037
sGARCH-std	0.000003 (0.0002)	1.150 (0.1230)	0.0849	0.194	-0.0001 (0.0002)	1.085 (0.1229)	0.7064	0.1766
EGARCH-norm	-0.0006 (0.0003)	1.56 (0.1621)	0.0003	0.2034	-0.0007 (0.0003)	1.453 (0.1448)	0.0078	0.2172
EGARCH-std	-0.0002 (0.0002)	0.662 (0.0618)	≈0	0.2406	-0.0003 (0.0002)	0.632 (0.0617)	≈0	0.2242
GJR-GARCH-norm	-0.0001 (0.0002)	1.228 (0.1363)	0.0249	0.1828	-0.0002 (0.0002)	1.133 (0.1222)	0.5469	0.1916
GJR-GARCH-std	0.00002 (0.0002)	1.099 (0.1160)	0.2492	0.1984	-0.0001 (0.0002)	1.039 (0.1159)	0.9445	0.1812
RealGARCH-RV-norm	-0.0002 (0.0002)	0.688 (0.0601)	≈0	0.2654	-0.0004 (0.0002)	0.677 (0.0522)	≈0	0.3169
RealGARCH-RV-std	0.0003 (0.0002)	0.725 (0.0555)	≈0	0.3185	0.0001 (0.0002)	0.729 (0.0544)	≈0	0.3309
RealGARCH-MedRV-norm	-0.0003 (0.0002)	0.5764 (0.047)	≈0	0.2917	-0.0006 (0.0002)	0.5713 (0.0406)	≈0	0.3353
RealGARCH-MedRV-std	0.0004 (0.0001)	0.710 (0.0468)	≈0	**0.3864**	0.0002 (0.0001)	0.723 (0.0454)	≈0	**0.4111**
HAR-RV	-0.0001 (0.0002)	0.9738 (0.0785)	0.2401	0.2975				
Log-HAR-RV	0.0001 (0.0002)	1.1669 (0.0862)	0.0013	0.3335				
HAR-RV-J	0.00001 (0.0002)	0.8601 (0.0743)	0.0127	0.2696				
Log-HAR-RV-J	0.0002 (0.0002)	1.0460 (0.0793)	0.0230	0.3238				
HAR-RV-CJ	-0.0001 (0.0002)	0.9351 (0.0751)	0.0842	0.2995				
Log-HAR-RV-CJ	0.0002 (0.0002)	1.0965 (0.0794)	0.0092	0.3443				

Notes: Values in parenthesis are the corresponding standard errors

6. Conclusion

The cointegration analysis showed that the ten used Cryptocurrencies share common stochastic trends permitting the construction of mean-reverting portfolios. In addition, the out-of-sample results suggest that all but one trading strategies are also profitable during the test period indicating a long-run relationship. Moreover, I found that the cointegrated portfolios based on the Johansen procedure yield, on average, a higher overall profit than the portfolio suggested by the Engle-Granger two-step procedure. Another important finding is that $Spread_t^{J,2}$ with the best in-sample performance also generates the highest final wealth out-of-sample. In contrast to the volatility modelling part where a good in-sample fit does not lead to a promising forecasting performance out-of-sample, the evaluation of the trading strategies in the test period demonstrates exactly the opposite. Here, both trading strategies on the basis of $Spread_t^{J,2}$ yield the highest total returns in-sample and out-of-sample as well. The main take-away from the presented cointegration analysis is that because of the unregulated crypto market there exist arbitrage opportunities that can be exploited using cointegrated portfolios. Such mean-reverting portfolios are able to generate positive returns during a bear market and also outperform a “buy-and-hold” strategy. However, there are some limitations that are not considered in this master thesis. The first issue are possible short-selling constraints that can limit the introduced statistical arbitrage strategies. Nevertheless, there are Cryptocurrency exchanges (*Binance* for example) that enable taking a short position, which speaks for the practical feasibility of the trading strategies used in this thesis. Another shortcoming is the neglect of transaction costs, which also affect the profitability. A future research direction could be the implementation of short-selling costs and transaction fees when constructing statistical arbitrage strategies based on cointegrated portfolios. Moreover, the granger causality results show that there is a link between the presented trading strategies and volatility, therefore chapter 5 aims to extend the state of the art of modelling the volatility of Bitcoin by using the more recent realized GARCH model with different realized measures and distributions and compare it to the HAR-RV model. So far, there hardly exists literature that makes a similar comparison. In line with Zahid et al. (2022) and Qian et al. (2022) and many other researches I also found that considering jumps when forecasting the conditional variance with GARCH models can substantially improve the forecasting accuracy. Furthermore, my findings indicate that the realized GARCH model with a jump-robust realized measure and t-distributed innovations is the only GARCH model that is a serious competitor to the HAR-RV model. Another important finding is that the returns of BTC are far away from a normal distribution, which become apparent when looking at the out-of-sample forecasting performance of both realized GARCH models with a normally distributed error term. Here, the

MSE and QLIKE are almost twice as high as when using the Student's t-distribution for the standardized returns regardless which volatility proxy is used. The most important finding is that the RealGARCH-MedRV-std outperforms all other considered GARCH models including five HAR-RV type models when using the realized variance as proxy. However, the difference of the MSE is not statistically significant while the QLIKE is significantly smaller when comparing with the forecasting performance of the HAR-RV model. I show that the Log-HAR-RV-CJ is the only model with a statistically significant smaller MSE than the HAR-RV. Moreover, the results of the Mincer-Zarnowitz regression are also interesting and rather surprising since they suggest that models with biased and inefficient forecasts seem to yield a higher R^2 than models where the F-test does not reject. One of the main conclusions to be drawn from the Mincer-Zarnowitz test is that the RealGARCH-MedRV-std is the superior model in terms of R^2. In conclusion, it can be said that jumps matter and are important to take into account when modelling and forecasting the volatility of Bitcoin. In contrast to other papers like Bergsli et al. (2022) for instance, this thesis provides evidence that GARCH models are able to provide more accurate forecasts than the standard HAR-RV model. Even though it is necessary to include additional information based on high-frequency data as it is the case with the realized GARCH model. In addition, the realized measure requires being jump-robust and it is crucial to employ a heavy-tailed distribution like the Student's t-distribution for the innovations.

To investigate the robustness of the provided results when also considering HAR-RV models with a leverage effect and using longer forecasting horizons (e.g. 1 week and 1 month) remains a topic for further research. Building on that a future research topic could be the incorporation of a volatility component of each Cryptocurrency when constructing statistical arbitrage strategies based on cointegrated portfolios. The inclusion of volatility as a second indicator apart from the optimal threshold for taking a long or short position most likely leads to a more profitable out-of-sample performance. I suggest building on the volatility analysis, which was done for BTC and extend it to the other nine Cryptocurrencies. This allows making a one-day ahead volatility forecast for each individual Cryptocurrency, which may prevent traders from huge losses because of costly long or short positions.

7. References

Adebola, Solarin Sakiru, Luis A. Gil-Alana, and Godfrey Madigu. 2019. “Gold Prices and the Cryptocurrencies: Evidence of Convergence and Cointegration.” *Physica A: Statistical Mechanics and Its Applications,* 523 (June): 1227–1236. https://doi.org/10.1016/j.physa.2019.04.123.

Aït-Sahalia, Yacine, and Jean Jacod. 2009. “Testing for Jumps in a Discretely Observed Process.” *The Annals of Statistics,* 37(1): 184–222. https://doi.org/10.1214/07-AOS568.

Alexander, C. 1999. “Optimal Hedging Using Cointegration.” *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 357(1758): 2039-2058. https://doi.org/10.1098/rsta.1999.0416.

Andersen, Torben G., and Tim Bollerslev. 1998. “Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts.” *International Economic Review,* 39(4): 885–905. https://doi.org/10.2307/2527343.

Andersen, Torben G., Tim Bollerslev, and Francis X. Diebold. 2007. “Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility.” *The Review of Economics and Statistics,* 89(4): 701–720. https://doi.org/10.1162/rest.89.4.701.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys. 2000. “The Distribution of Realized Exchange Rate Volatility.” SSRN Scholarly Paper. Rochester, NY. https://doi.org/10.2139/ssrn.267791.

Andersen, Torben G., Tim Bollerslev, and Xin Huang. 2008. “A Reduced Form Framework for Modeling Volatility of Speculative Prices Based on Realized Variation Measures.” SSRN Scholarly Paper. Rochester, NY. https://doi.org/10.2139/ssrn.1150050.

Andersen, Torben G., Dobrislav Dobrev, and Ernst Schaumburg. 2012. “Jump-Robust Volatility Estimation Using Nearest Neighbor Truncation.” *Journal of Econometrics*, 169(1): 75–93. https://doi.org/10.1016/j.jeconom.2012.01.011.

Aras, Serkan. 2021. “On Improving GARCH Volatility Forecasts for Bitcoin via a Meta-Learning Approach.” *Knowledge-Based Systems,* 230 (October): 107393. https://doi.org/10.1016/j.knosys.2021.107393.

Barndorff-Nielsen, Ole E., and Neil Shephard. 2004. “Power and Bipower Variation with Stochastic Volatility and Jumps.” *Journal of Financial Econometrics,* 2(1): 1–37. https://doi.org/10.1093/jjfinec/nbh001.

———. 2006. “Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation.” *Journal of Financial Econometrics,* 4(1): 1–30. https://doi.org/10.1093/jjfinec/nbi022.

Bergsli, Lykke Øverland, Andrea Falk Lind, Peter Molnár, and Michał Polasik. 2022. “Forecasting Volatility of Bitcoin.” *Research in International Business and Finance,* 59 (January): 101540. https://doi.org/10.1016/j.ribaf.2021.101540.

Black, F. 1976. "Studies of stock price volatility changes." *Proceedings of the 1976 Meeting of the Business and Economic Statistics Section, American Statistical Association*, 177-181.

Bollerslev, Tim. 1986. “Generalized Autoregressive Conditional Heteroskedasticity.” *Journal of Econometrics,* 31(3): 307–327. https://doi.org/10.1016/0304-4076(86)90063-1.

Bollerslev, Tim, Andrew J. Patton, and Rogier Quaedvlieg. 2016. "Exploiting the Errors: A Simple Approach for Improved Volatility Forecasting." *Journal of Econometrics,* 192(1): 1–18. https://doi.org/10.1016/j.jeconom.2015.10.007.

Bouri, Elie, Konstantinos Gkillas, Rangan Gupta, and Christian Pierdzioch. 2021. "Forecasting Realized Volatility of Bitcoin: The Role of the Trade War." *Computational Economics,* 57(1): 29–53. https://doi.org/10.1007/s10614-020-10022-4.

Catania, Leopoldo, and Stefano Grassi. 2017. "Modelling Crypto-Currencies Financial Time-Series." SSRN Scholarly Paper. Rochester, NY. https://doi.org/10.2139/ssrn.3028486.

Cermak, Vavrinec. 2017. "Can Bitcoin Become a Viable Alternative to Fiat Currencies? An Empirical Analysis of Bitcoin's Volatility Based on a GARCH Model." SSRN Scholarly Paper. Rochester, NY. https://doi.org/10.2139/ssrn.2961405.

Chiu, Mei Choi, and Hoi Ying Wong. 2015. "Dynamic Cointegrated Pairs Trading: Mean–Variance Time-Consistent Strategies." *Journal of Computational and Applied Mathematics,* 290 (December): 516–534. https://doi.org/10.1016/j.cam.2015.06.004.

"CME Group Announces Launch of Micro-Sized Bitcoin and Ether Options - CME Group." Accessed on April 28, 2022. https://www.cmegroup.com/media-room/press-releases/2022/3/28/cme_group_announceslaunchofmicro-sizedbitcoinandetheroptions.html.

Corsi, Fulvio. 2009. "A Simple Approximate Long-Memory Model of Realized Volatility." *Journal of Financial Econometrics,* 7(2): 174–196. https://doi.org/10.1093/jjfinec/nbp001.

Dickey, David A., and Wayne A. Fuller. 1979. "Distribution of the Estimators for Autoregressive Time Series With a Unit Root." *Journal of the American Statistical Association,* 74(366): 427–431. https://doi.org/10.2307/2286348.

Diebold, Francis, and Roberto Mariano. 1995. "Comparing Predictive Accuracy." *Journal of Business & Economic Statistics,* 13(3): 253–263.

Engle, Robert F., and C. W. J. Granger. 1987. "Co-Integration and Error Correction: Representation, Estimation, and Testing." *Econometrica,* 55(2): 251–276. https://doi.org/10.2307/1913236.

Engle, Robert F., David F. Hendry, and Jean-Francois Richard. 1983. "Exogeneity." *Econometrica,* 51(2): 277–304. https://doi.org/10.2307/1911990.

Engle, Robert F., and Victor K. Ng. 1993. "Measuring and Testing the Impact of News on Volatility." *The Journal of Finance,* 48(5): 1749–1778. https://doi.org/10.1111/j.1540-6261.1993.tb05127.x.

Gatev, Evan, William N. Goetzmann, and K. Geert Rouwenhorst. 2006. "Pairs Trading: Performance of a Relative Value Arbitrage Rule." SSRN Scholarly Paper. Rochester, NY. https://doi.org/10.2139/ssrn.141615.

"GDP (Current US$) - Spain | Data." 2022. Accessed on September 11, 2022. The World Bank. 2022. https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=ES.

George J. Jiang, and Roel C. A. Oomen. 2008. "Testing for Jumps When Asset Prices Are Observed with Noise–a 'Swap Variance' Approach." *Journal of Econometrics,* 144(2): 352–370. https://doi.org/10.1016/j.jeconom.2008.04.009.

"Global Cryptocurrency Market Charts | CoinMarketCap." Accessed April 27, 2022. https://coinmarketcap.com/charts/.

Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle. 1993. "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on

Stocks.” *The Journal of Finance,* 48(5): 1779–1801. https://doi.org/10.1111/j.1540-6261.1993.tb05128.x.

Granger, C. W. J. 1969. “Investigating Causal Relations by Econometric Models and Cross-Spectral Methods.” *Econometrica,* 37(3): 424–438. https://doi.org/10.2307/1912791.

———. 1981. “Some Properties of Time Series Data and Their Use in Econometric Model Specification.” *Journal of Econometrics,* 16 (1): 121–130. https://doi.org/10.1016/0304-4076(81)90079-8.

Granger, C. W. J., and P. Newbold. 1974. “Spurious Regressions in Econometrics.” *Journal of Econometrics,* 2(2): 111–120. https://doi.org/10.1016/0304-4076(74)90034-7.

Hansen, Peter R., Asger Lunde, and James M. Nason. 2011. “The Model Confidence Set.” *Econometrica,* 79(2): 453–497. https://doi.org/10.3982/ECTA5771.

Hansen, Peter Reinhard, Zhuo Huang, and Howard Howan Shek. 2012. “Realized GARCH: A Joint Model for Returns and Realized Measures of Volatility.” *Journal of Applied Econometrics,* 27(6): 877–906. https://doi.org/10.1002/jae.1234.

Harvey, David, Stephen Leybourne, and Paul Newbold. 1997. “Testing the Equality of Prediction Mean Squared Errors.” *International Journal of Forecasting,* 13(2): 281–291. https://doi.org/10.1016/S0169-2070(96)00719-4.

He, Changli, Timo Teräsvirta, and Hans Malmsten. 2002. “Moment Structure of a Family of First-Order Exponential GARCH Models.” *Econometric Theory,* 18(4): 868–885.

Hou, Ai Jun, Weining Wang, Cathy Yi-Hsuan Chen, and Wolfgang K. Härdle. 2019. “Pricing Cryptocurrency Options: The Case of Bitcoin and CRIX.” SSRN Scholarly Paper. Rochester, NY. https://doi.org/10.2139/ssrn.3159130.

Hung, Jui-Cheng, Hung-Chun Liu, and J. Jimmy Yang. 2020. “Improving the Realized GARCH’s Volatility Forecast for Bitcoin with Jump-Robust Estimators.” *The North American Journal of Economics and Finance,* 52 (April): 101165. https://doi.org/10.1016/j.najef.2020.101165.

Jiang, George J., and Roel C. A. Oomen. 2007. “Testing for Jumps When Asset Prices Are Observed with Noise - A Swap Variance Approach.” SSRN Scholarly Paper. Rochester, NY. https://papers.ssrn.com/abstract=850324.

Johansen, Søren. 1988. “Statistical Analysis of Cointegration Vectors.” *Journal of Economic Dynamics and Control,* 12(2): 231–254. https://doi.org/10.1016/0165-1889(88)90041-3.

———. 1991. “Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models.” *Econometrica,* 59(6): 1551–1580. https://doi.org/10.2307/2938278.

———. 1995. “Likelihood-Based Inference in Cointegrated Vector Autoregressive Models.” OUP Catalogue. Oxford University Press. https://econpapers.repec.org/bookchap/oxpobooks/9780198774501.htm.

Kubát, Max. 2015. “Virtual Currency Bitcoin in the Scope of Money Definition and Store of Value.” *Procedia Economics and Finance*, 30: 409–416. https://doi.org/10.1016/S2212-5671(15)01308-8.

Leung, Tim, and Hung Nguyen. 2018. “Constructing Cointegrated Cryptocurrency Portfolios for Statistical Arbitrage.” SSRN Scholarly Paper. Rochester, NY. https://doi.org/10.2139/ssrn.3235890.

Liu, Ruyi, Zhen Wu, and Qing Zhang. 2020. “Pairs-Trading under Geometric Brownian Motions: An Optimal Strategy with Cutting Losses.” *Automatica,* 115 (May): 108912. https://doi.org/10.1016/j.automatica.2020.108912.

MacKinnon, James G. 2010. “Critical Values for Cointegration Tests.” *QED Working Paper,* No. 1227.

MacKinnon, James, Alfred Haug, and Leo Michelis. 1999. “Numerical Distribution Functions of Likelihood Ratio Tests for Cointegration.” *Journal of Applied Econometrics,* 14(5): 563–577.

Merton, Robert C. 1976. “Option Pricing When Underlying Stock Returns Are Discontinuous.” *Journal of Financial Economics,* 3(1): 125–144. https://doi.org/10.1016/0304-405X(76)90022-2.

Mincer, Jacob, and Victor Zarnowitz. 1969. “The Evaluation of Economic Forecasts.” *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, 3-46. https://econpapers.repec.org/bookchap/nbrnberch/1214.htm.

Mittal, Sonal. 2012. “Is Bitcoin Money? Bitcoin and Alternate Theories of Money.” SSRN Scholarly Paper. Rochester, NY. https://doi.org/10.2139/ssrn.2434194.

Müller, Ulrich A., Michel M. Dacorogna, Rakhal D. Davé, Richard B. Olsen, Olivier V. Pictet, and Jacob E. von Weizsäcker. 1997. “Volatilities of Different Time Resolutions — Analyzing the Dynamics of Market Components.” *Journal of Empirical Finance*, 4(2-3): 213–239. https://doi.org/10.1016/S0927-5398(97)00007-8.

Nelson, Daniel B. 1991. “Conditional Heteroskedasticity in Asset Returns: A New Approach.” *Econometrica,* 59(2): 347–370. https://doi.org/10.2307/2938260.

Ngo, Minh-Man, and Huyên Pham. 2016. “Optimal Switching for the Pairs Trading Rule: A Viscosity Solutions Approach.” *Journal of Mathematical Analysis and Applications,* 441(1): 403–425. https://doi.org/10.1016/j.jmaa.2016.03.060.

Ni, Sophie X., Jun Pan, and Allen M. Poteshman. 2008. “Volatility Information Trading in the Option Market.” *The Journal of Finance,* 63(3): 1059–1091. https://doi.org/10.1111/j.1540-6261.2008.01352.x.

Omane-Adjepong, Maurice, Paul Alagidede, and Nana Kwame Akosah. 2019. “Wavelet Time-Scale Persistence Analysis of Cryptocurrency Market Returns and Volatility.” *Physica A: Statistical Mechanics and Its Applications,* 514 (January): 105–120. https://doi.org/10.1016/j.physa.2018.09.013.

Patton, Andrew. 2011. “Volatility Forecast Comparison Using Imperfect Volatility Proxies.” *Journal of Econometrics,* 160(1): 246–256.

Pesaran, M. Hashem, Yongcheol Shin, and Richard J. Smith. 2001. “Bounds Testing Approaches to the Analysis of Level Relationships.” *Journal of Applied Econometrics,* 16(3): 289–326.

Phillips, Peter, and Sam Ouliaris. 1990. “Asymptotic Properties of Residual Based Tests for Cointegration.” *Econometrica,* 58(1): 165–193.

Pollet, Joshua M., and Mungo Wilson. 2010. “Average Correlation and Stock Market Returns.” *Journal of Financial Economics,* 96(3): 364–380. https://doi.org/10.1016/j.jfineco.2010.02.011.

Qian, Lihua, Jiqian Wang, Feng Ma, and Ziyang Li. 2022. “Bitcoin Volatility Predictability–The Role of Jumps and Regimes.” *Finance Research Letters,* 47 (June): 102687. https://doi.org/10.1016/j.frl.2022.102687.

Shen, Dehua, Andrew Urquhart, and Pengfei Wang. 2020. “Forecasting the Volatility of Bitcoin: The Importance of Jumps and Structural Breaks.” *European Financial Management,* 26(5): 1294–1323. https://doi.org/10.1111/eufm.12254.

Song, Qingshuo, and Qing Zhang. 2013. “An Optimal Pairs-Trading Rule.” *Automatica,* 49(10): 3007–3014. https://doi.org/10.1016/j.automatica.2013.07.012.

Sovbetov, Yhlas. 2018. “Factors Influencing Cryptocurrency Prices: Evidence from Bitcoin, Ethereum, Dash, Litcoin, and Monero.” SSRN Scholarly Paper. Rochester, NY. https://papers.ssrn.com/abstract=3125347.

Tan, Zhengxun, Yilong Huang, and Binuo Xiao. 2021. “Value at Risk and Returns of Cryptocurrencies before and after the Crash: Long-Run Relations and Fractional Cointegration.” *Research in International Business and Finance,* 56 (April): 101347. https://doi.org/10.1016/j.ribaf.2020.101347.

Tapia, Sebastian, and Werner Kristjanpoller. 2022. “Framework Based on Multiplicative Error and Residual Analysis to Forecast Bitcoin Intraday-Volatility.” *Physica A: Statistical Mechanics and Its Applications,* 589 (March): 126613. https://doi.org/10.1016/j.physa.2021.126613.

Theodosiou, Marina, and Filip Zikes. 2011. “A Comprehensive Comparison of Alternative Tests for Jumps in Asset Prices.” *Working Papers* 2011–2. Central Bank of Cyprus. https://ideas.repec.org/p/cyb/wpaper/2011-2.html.

Tokat, Ekin, and Ahmet Cevdet Hayrullahoğlu. 2022. “Pairs Trading: Is It Applicable to Exchange-Traded Funds?” *Borsa Istanbul Review,* 22(4): 743–751. https://doi.org/10.1016/j.bir.2021.08.001.

Urquhart, Andrew. 2018. “What Causes the Attention of Bitcoin?” *Economics Letters,* 166 (May): 40–44. https://doi.org/10.1016/j.econlet.2018.02.017.

Yan, Tingjin, and Hoi Ying Wong. 2022. “Equilibrium Pairs Trading under Delayed Cointegration.” *Automatica,* 144 (October): 110498. https://doi.org/10.1016/j.automatica.2022.110498.

Zahid, Mamoona, Farhat Iqbal, Abdul Raziq Abdul Raziq, and Naveed Sheikh. 2022. “Modeling and Forecasting the Realized Volatility of Bitcoin Using Realized HAR-GARCH-Type Models with Jumps and Inverse Leverage Effect.” *Sains Malaysiana,* 51(3): 929–942. https://doi.org/10.17576/jsm-2022-5103-25.

A. Appendix

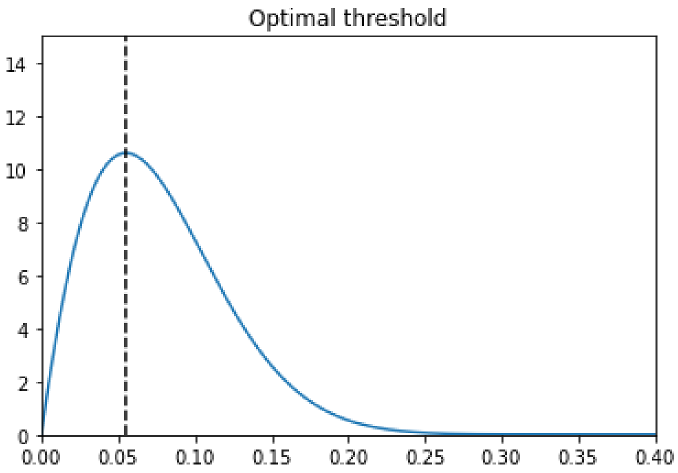


Figure 33: Optimal threshold of $Spread_t^{EG}$

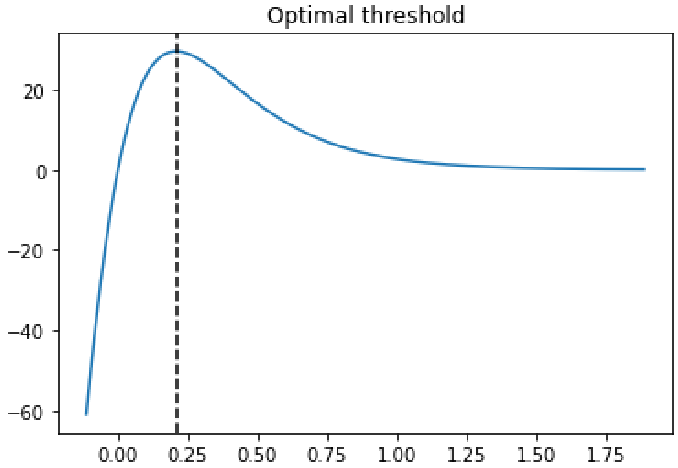


Figure 34: Optimal threshold of $Spread_t^{J,1}$

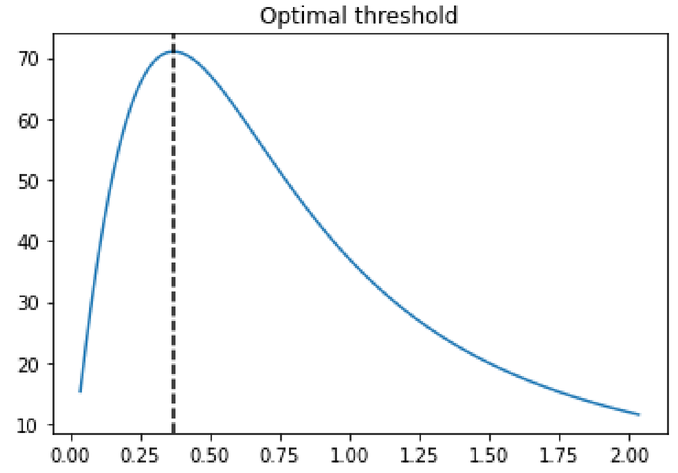


Figure 35: Optimal threshold of $Spread_t^{J,2}$

Figure 36: ECT of BTC-BNB pair

Figure 37: ECT of BTC-XRP pair

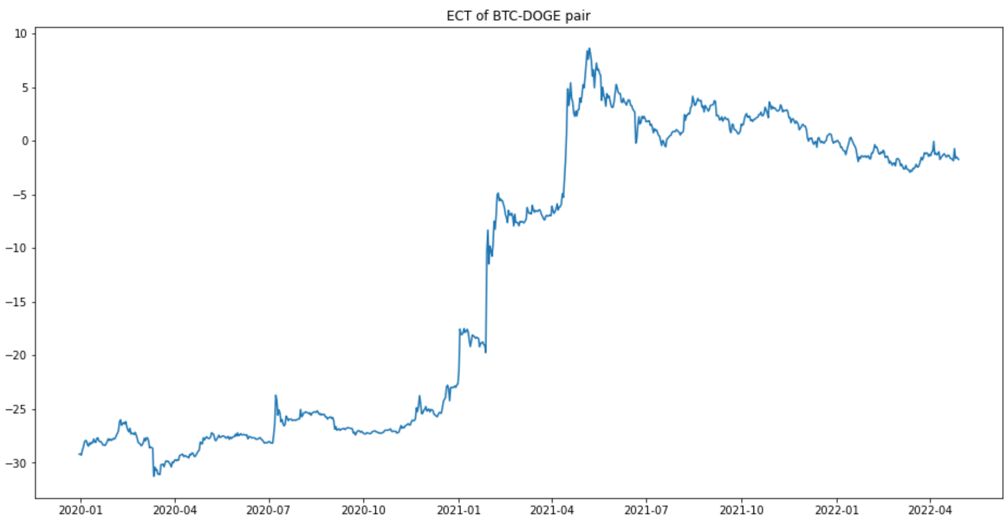


Figure 38: ECT of BTC-DOGE pair

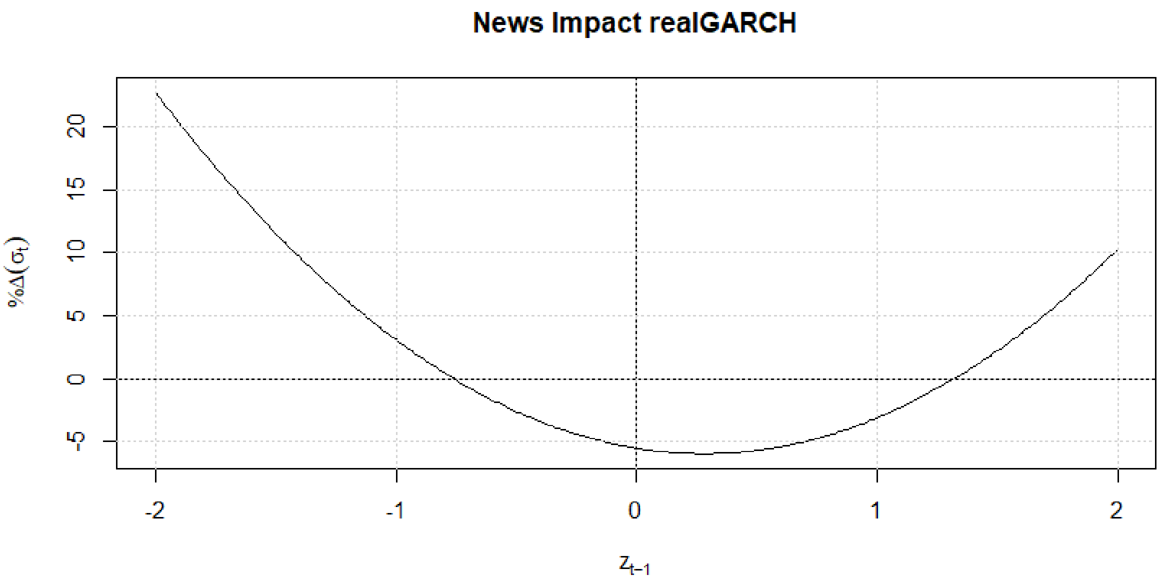


Figure 39: News impact Curve (RealGARCH-RV-std)

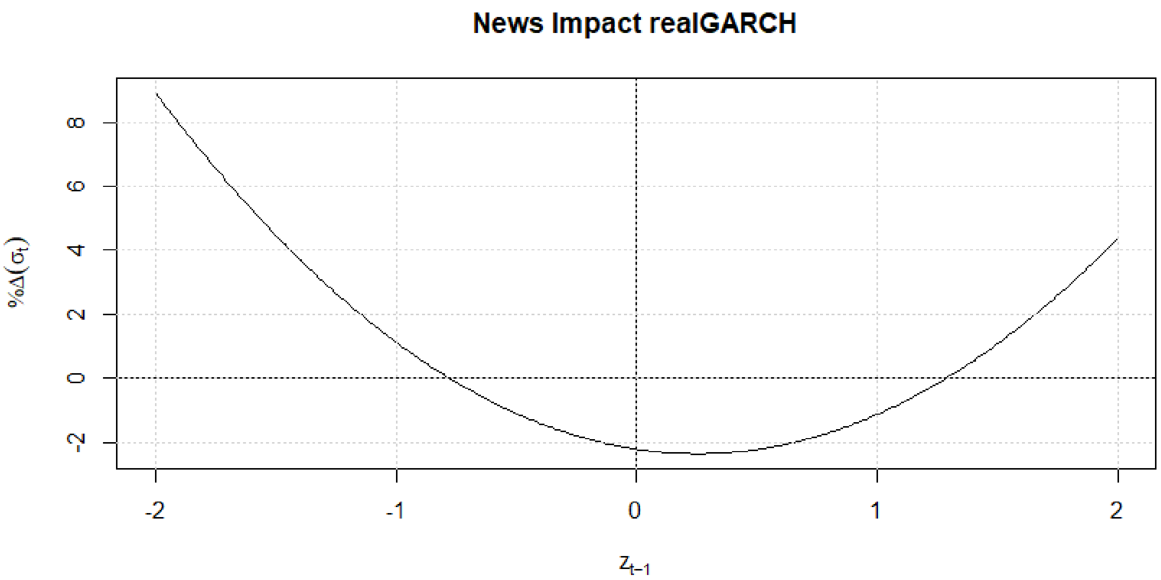


Figure 40: News impact Curve (RealGARCH-RV-norm)

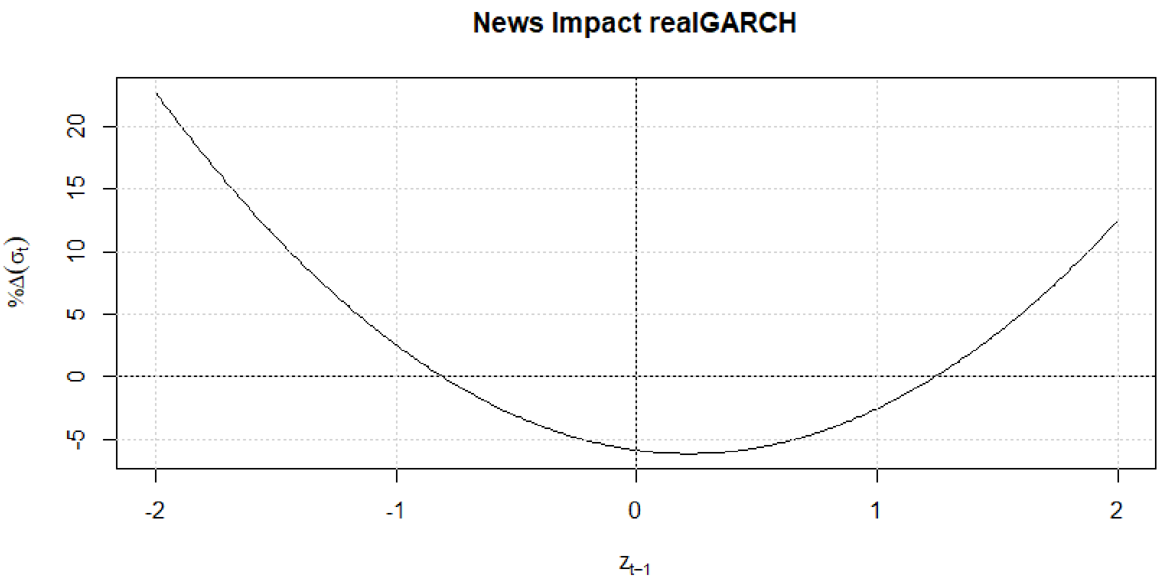


Figure 41: News impact Curve (RealGARCH-MedRV-std)

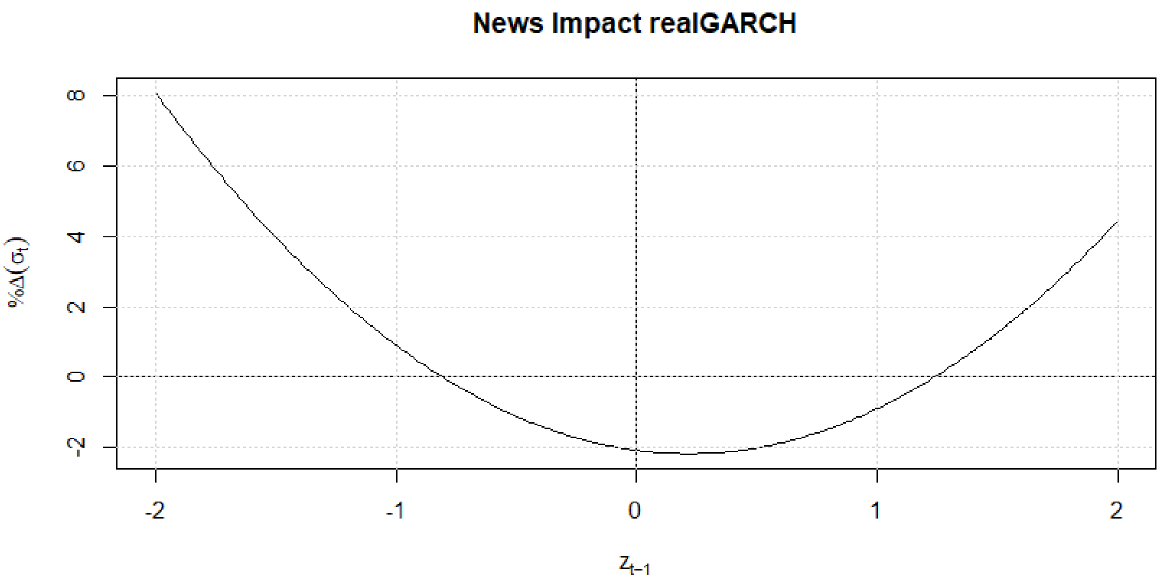


Figure 42: News impact Curve (RealGARCH-MedRV-norm)

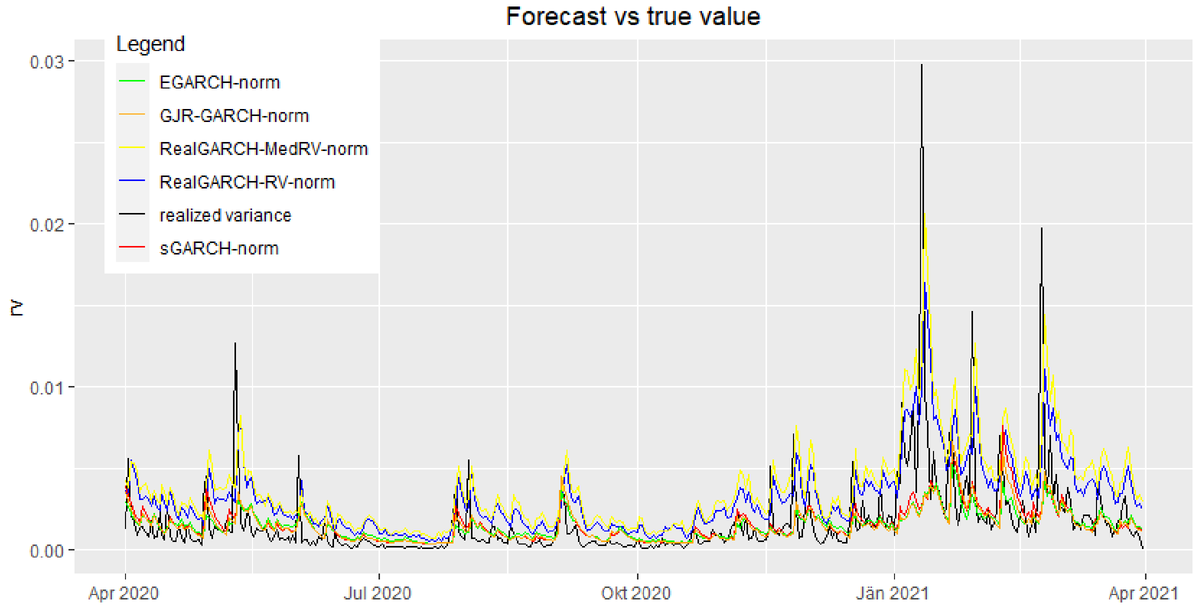


Figure 43: Forecasting performance (GARCH-norm models)

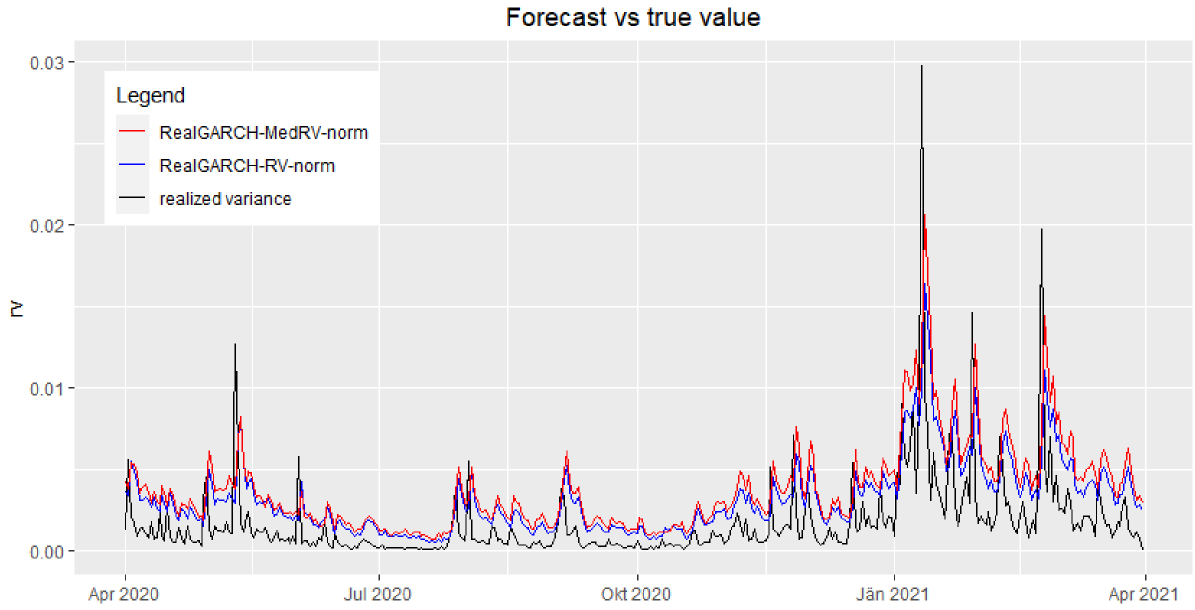


Figure 44: RealGARCH-RV-norm vs RealGARCH-MedRV-norm

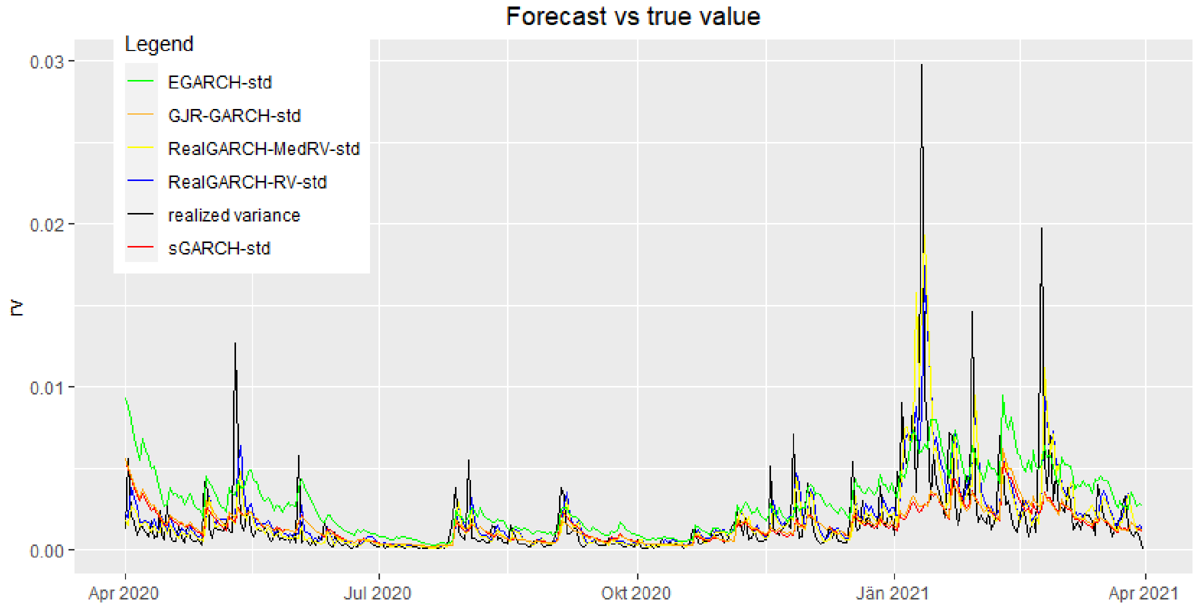


Figure 45: Forecasting performance (GARCH-std models)

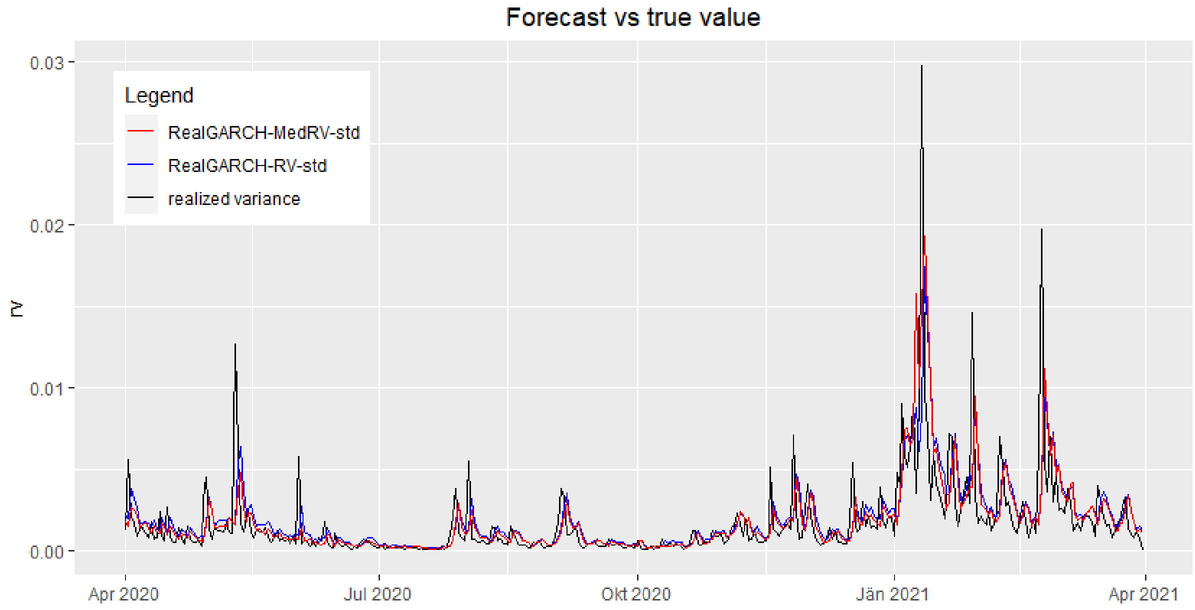


Figure 46: RealGARCH-RV-std vs RealGARCH-MedRV-std

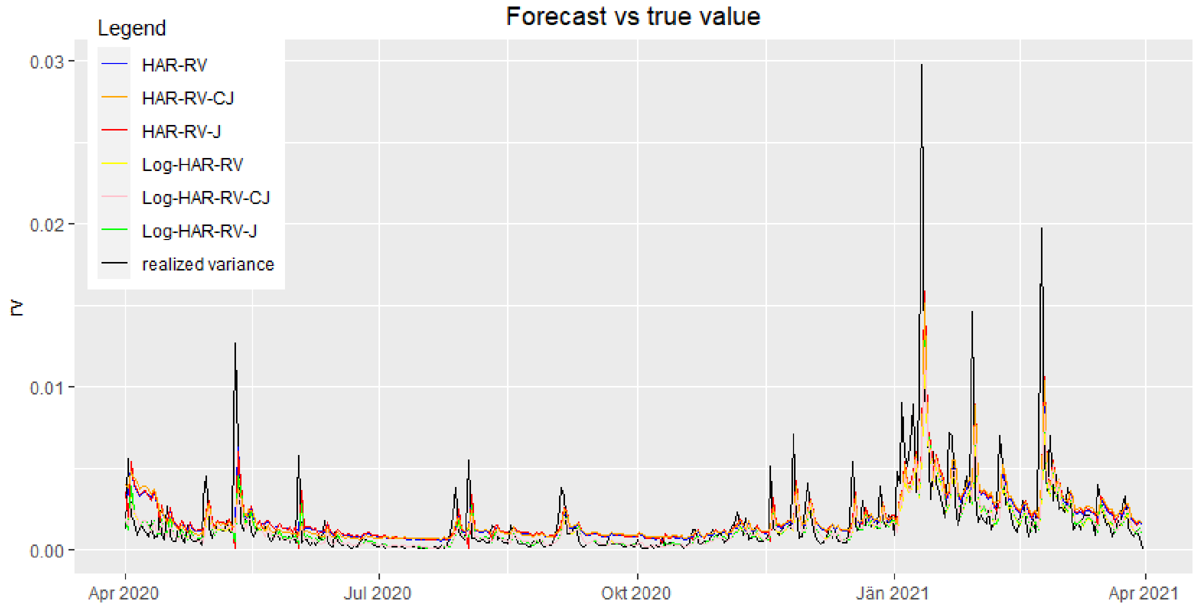


Figure 47: Forecasting performance (HAR-RV models)

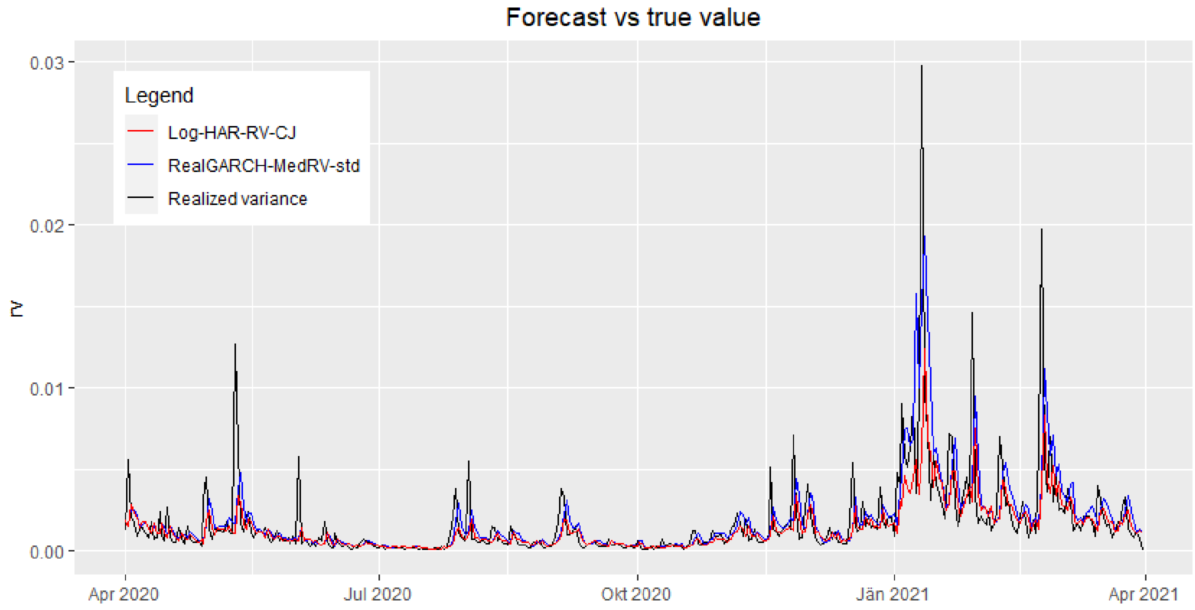


Figure 48: RealGARCH-MedRV-std vs Log-HAR-RV-CJ