INSTITUTO
UNIVERSITÁRIO
DE LISBOA

An Assessment of Historical Simulation Techniques for Alternative Investments

Sebastião Maria Cavaco Falcato

Master in Financial Mathematics

Supervisor:
PhD Joaquim Paulo Viegas Ferreira de Carvalho, Assistant Professor, ISCTE-IUL

September 2025

Department of Finance

Department of Mathematics

An Assessment of Historical Simulation Techniques for Alternative Investments

Sebastião Maria Cavaco Falcato

Master in Financial Mathematics

Supervisor:
PhD Joaquim Paulo Viegas Ferreira de Carvalho, Assistant Professor, ISCTE-IUL

September 2025

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Professor Paulo Viegas de Carvalho, for his guidance, availability, and constant support throughout the development of this dissertation. His valuable insights and encouragement have been fundamental to the completion of this work.

I am profoundly grateful to my family for their constant encouragement and support throughout this process. Their patience, understanding, and belief in me have provided the foundation to face challenges and stay motivated until the very end.

I also wish to thank my friends, whose presence has been a source of balance and encouragement. Their companionship, words of support, and ability to bring lightness to demanding moments have made this journey not only more manageable but also more meaningful.

Resumo

Esta dissertação compara três métodos não paramétricos do VaR - Simulação Histórica (HS), o método ponderado de Boudoukh–Richardson–Whitelaw (BRW, 1998) e o modelo ponderado pela volatilidade de Hull and White (1998), usando dados diários da Bitcoin e do S&P 500 entre 2018 e 2022, um período que inclui condições normais de mercado e períodos de stress (COVID-19). A HS é simples de implementar, mas reage lentamente a mudanças abruptas no risco. O modelo de BRW, ao dar mais peso às observações recentes, adapta-se mais rapidamente e oferece uma cobertura mais estável quando as condições do mercado mudam. O modelo de Hull–White, que ajusta os retornos passados pela volatilidade condicional atual, é o mais fiável nos nossos testes e permanece conservador em todos os períodos. No geral, as evidências mostram que levar explicitamente em conta a dinâmica da volatilidade melhora o VaR não paramétrico. A HS é útil em períodos calmos, o BRW oferece um meio-termo prático e o Hull–White fornece a orientação mais confiável quando a incerteza aumenta.

Palavras-chave: *VaR*, Simulação Histórica, *BRW, HW, Backtesting, Bitcoin*

Classificação JEL: C58, G32

Abstract

This dissertation compares three non-parametric Value-at-Risk methods—Historical Simulation (HS), the Boudoukh–Richardson–Whitelaw weighted method (BRW, 1998), and the volatility-scaled procedure of Hull–White (1998)—using daily data for Bitcoin and the S&P 500 over 2018–2022, a period that includes both tranquil markets and the COVID-19 disruption. HS is straightforward to implement but reacts slowly to abrupt changes in risk. BRW, by giving more weight to recent observations, adapts faster and delivers steadier coverage when market conditions shift. The Hull–White variant, which rescales past returns by current conditional volatility, is the most reliable in our tests and remains conservative across regimes. Overall, the evidence shows that explicitly accounting for volatility dynamics improves non-parametric VaR. HS is adequate in calm periods, BRW offers a practical middle ground, and Hull–White provides the most dependable guidance when uncertainty rises.

Keywords: *VaR*, Historical Simulation, *BRW, HW, Backtesting, Bitcoin*

JEL Classification: C58, G32

Index

Table Index

Figures Index

Glossary

ARCH – Autoregressive Conditional Heteroskedasticity

BRW – Boudoukh, Richardson e Whitelaw

BTC – Bitcoin

COVID-19 – Coronavirus Disease 2019

ES – Expected Shortfall

EWMA – Exponentially Weighted Moving Average

FRTB – Fundamental Review of the Trading Book

GARCH – Generalized Autoregressive Conditional Heteroskedasticity

HS – Historical Simulation

HW – Hull and White

LRcc – Likelihood Ratio Conditional Coverage

LRind – Likelihood Ratio Independence Test

LRuc – Likelihood Ratio Unconditional Coverage

P&L – Profit and Loss

POF – Proportion of Failures

S&P 500 – Standard & Poor's 500 Index

VaR – Value-at-Risk

1. Introduction

Financial institutions face a lot of risks, but market risk is one of the most important. This risk is associated with the potential loss of money due to fluctuations in interest rates, equity prices, exchange rates, and other market variables. The Basel Committee defines Value at Risk (VaR) as the standard way to measure how much capital is needed in this situation. This made it one of the most popular tools for managing risk (BIS, 2019). Value at Risk (VaR) is a single number that indicates the maximum amount a portfolio could lose over a specified time period and confidence interval.

Non-parametric VaR estimation avoids strong distributional assumptions by deriving risk measures from the observed return distribution. Historical Simulation (HS) is the canonical implementation: the VaR is the relevant sample quantile computed from ordered past returns. Because each observation enters with equal weight, HS tends to absorb volatility breaks only gradually, a limitation noted by Pritsker (2006). An age-weighted variant due to Boudoukh, Richardson, and Whitelaw (1998) addresses this inertia by letting the weight on older data decay, so that more recent price changes have a greater influence on the reported quantile. A different approach is proposed by Hull and White (1998), who scale each historical return by the ratio of current to past conditional volatility. This rescaling aligns the empirical sample with today's risk level and, in environments with pronounced volatility clustering, yields quantiles that track regime shifts more faithfully than equal-weight HS while retaining the non-parametric spirit. In practice, these three procedures span a spectrum from simplicity to adaptability.

This study focuses on Bitcoin as a decentralized digital asset introduced in 2008 whose market design and trading environment differ from those of conventional securities. Over the sample period, Bitcoin exhibits higher day-to-day variability than broad equity benchmarks, as well as episodes of abrupt repricing. Unlike equities, its pricing is not tied to cash-flow–based valuation anchors, and market dynamics are shaped by adoption, liquidity conditions, and microstructure features. Examining Bitcoin alongside the S&P 500, therefore, provides a useful contrast: it allows the evaluation of risk-measurement frameworks across assets that exhibit distinct distributional properties and volatility regimes.

The COVID-19 pandemic in 2020 led to a unique opportunity to study the conditions of VaR approaches. In March 2020, the S&P 500 declined significantly, and Bitcoin decreased promptly and then recovered rapidly again. These events showed how harmful traditional risk assessment is and how easy it is for financial markets to be affected. Corbet (2020) has found that the pandemic caused asset classes to become more volatile and contagious, which makes standard models less reliable. Other research has also demonstrated that Bitcoin was not a safe place to put money during this time; it actually made portfolio risk higher (Conlon, 2020).

This dissertation's findings reveal significant disparities in the efficacy of the three non-parametric approaches. The volatility-adjusted method proposed by Hull and White (1998) is the most reliable and prudent, particularly in conditions of high volatility. The BRW approach is effective and reliable, owing to its weighting method that prioritizes recent market conditions. Conversely, the basic Historical Simulation becomes inadequate during market stress, since it fails to promptly adjust to abrupt fluctuations in risk. The results indicate that HW is the optimal model for extreme conditions, although BRW serves as a viable alternative, and HS remains effective in stable scenarios.

This dissertation includes five chapters. Chapter 2 offers an extensive literature review, focusing on the main concepts of market risk, the functions of VaR and ES, and the evolution of non-parametric methodologies, including Historical Simulation, BRW, and Hull and White, alongside the backtesting framework. Chapter 3 delineates the dataset, the sampled periods, and the methodological approaches employed to estimate Value at Risk (VaR) and evaluate its adequacy. Chapter 4 reports the empirical evidence, discussing the results obtained for Bitcoin and the S&P 500 across the three regimes surrounding the COVID-19 crisis, as well as the outcomes of the backtesting procedures applied to each model. Finally, Chapter 5 concludes by summarising the main findings, discussing their implications for risk management, and suggesting directions for future research.

2. Literature Review

2.1 Value at Risk and Expected Shortfall

Evaluating market risk is essential to modern financial risk management, particularly with J.P. Morgan's implementation of VaR in the early 1990s. VaR rapidly emerged as the predominant method for illustrating the potential monetary loss of a portfolio over a certain timeframe and with a specified confidence level (Danielsson & Zhou, 2016). This strategy is employed due to its simplicity; it transforms complex portfolio risk into a quantifiable potential loss with a defined probability (Manganelli & Engle, 2001).

Let X represent a loss random variable characterized by its cumulative distribution function. Nolde and Ziegel (2017) define Value at Risk at a confidence level $\alpha \in (0,1)$:

$$VaR_{\alpha}(X) = inf\,\{ x \in R : F\_X(x) \geq \alpha \} \quad (1)$$

In this context, X denotes the random variable of portfolio losses, F_X is the corresponding cumulative distribution function, $\alpha \in (0,1)$ represents the chosen confidence level, $VaR_{\alpha}(X)$ corresponds to the α - quantile of the distribution of X, and the operator "inf" indicates the greatest lower bound of the set of $x \in \mathbb{R}$ such that $F_X(x) \geq \alpha$. The Basel regulatory framework stipulates that α = 0.99 is the appropriate value for determining the capital required to mitigate market risk (Bank for International Settlements, 2019).

VaR has been criticized for its failure to appropriately represent the severity of tail events, despite its widespread application. Du and Escanciano (2015) assert that VaR overlooks significant losses by focusing exclusively on the quantile threshold. Expected Shortfall (ES) has been suggested as a remedy to mitigate this limitation. Value at Risk and Expected Shortfall are interconnected; however, ES quantifies the potential magnitude of losses that may occur if the VaR threshold is exceeded.

Nolde and Ziegel (2017) characterize Expected Shortfall in the following manner: The ES is defined for an integrable loss variable X and a level $\nu \in (0,1)$ as follows:

$$ES_{\nu}(X) = \frac{1}{1-\nu} \int_{\nu}^{1} VaR_{u}(X)\, du \quad (2)$$

In this equation, X stands for the portfolio's loss distribution, and ν stands for the chosen level of confidence, $VaR_u(X)$ is the Value at Risk at level u, and the integral finds the average of all the quantiles in the loss distribution that are higher than the ν - quantile. $ES_\nu(X)$ is the estimated loss that is greater than the VaR at level ν. People often call this number Conditional VaR, Expected Tail Loss, or Average VaR. The Basel Committee says that ν = 0.975 should be the legal limit. Data that follows a normal distribution gives risk estimates that are similar to $VaR_{99\%}$ (Bank for International Settlements, 2019).

Their mathematical characteristics can clarify the difference between VaR and ES. Artzner, Delbaen, Eber, and Heath (1999) delineated four criteria for a risk measure ρ to be deemed coherent. Chen (2018) posits that a risk measure ρ, when applied to X, $X1$, and $X2$, is considered coherent alone if it meets the following criteria:

i. Subadditivity: $\rho(X1 + X2) \leq \rho(X1) + \rho(X2)$. This ensures that diversification does not increase the measured risk.

ii. Monotonicity: $R(X1) \geq R(X2) \Rightarrow \rho(X1) \leq \rho(X2)$, where $R(X)$ represents the return associated with portfolio X.

iii. Positive Homogeneity: $\rho(\theta X) = \theta\rho(X),\ \theta > 0$, implying that scaling a portfolio by a positive factor scales its risk proportionally.

iv. Translation Invariance: $\rho(X + k) = \rho(X) - k$, where $k \in \mathbb{R}$ is a risk-free cash amount. Adding cash reduces the overall risk by that same amount.

ES fits all four of the conditions, hence it is thought to be possible (Acerbi & Tasche, 2002). VaR, on the other hand, doesn't always match the subadditivity criteria. This indicates that the advantages of diversity might be misrepresented, particularly in markets with substantial tails and unpredictable return distributions (Ballotta & Fusai, 2017).

One big difference between the two metrics is how well they acquire statistical data. You can utilize consistent scoring systems to assess and confirm Value at Risk, which shows that it can be useful for getting information. However, ES cannot be induced independently, complicating straightforward backtesting. Nolde and Ziegel (2017) demonstrate that the simultaneous utilization of Value at Risk (VaR) and Expected Shortfall (ES) facilitates a comprehensive assessment of projections.

Var is the most used measure because it's simple to grasp, and regulators agree with it. Basel III picked ES as the standard since it has a number of theoretical issues. Jorion (2007) asserts that VaR remains a viable communication tool among practitioners. However, Chu and Bhattacharyya (2020) demonstrate that ES is superior in reducing tail risk, particularly in markets marked by significant volatility. The two measurements fit well together because VaR offers a defined criterion and makes it easy to test things again, and ES, on the other hand, makes sure that the manner in which extreme losses are measured is always the same.

2.2 Historical Simulation

The Historical Simulation (HS) method is one of the easiest and most obvious ways to find Value at Risk. The main benefit is that it doesn't need any assumptions about the underlying distribution of returns; it uses the empirical distribution of the data that was seen (Hendricks, 1996).

To use this method to find the VaR, you need a big enough sample of past portfolio results. After that, the returns are put in order from least to most favorable. Finally, you can find the VaR at the stated confidence level by choosing the right quantile from the ordered distribution. A 99% one-day VaR means that just 1% of past observations show losses that are higher than that amount (Beder, 1995).

This strategy is appealing since it is clear and easy to apply. On the other hand, the HS framework presents important limitations. It implies that the distribution of returns remains constant over the whole observation period, based on the assumption that all returns in the sample are independently and identically distributed (Žiković &Aktan, 2010).

This assumption stops the model from changing to fit the market. When conditions are steady, estimates could be too cautious. When they are unstable, losses might seem less important than they really are. Basu (2009) talks about the "ghost effect," which is when big historical losses have a bigger impact on current risk evaluations than they should, which raises VaR values.

The duration of the historical window also influences how accurate the method is. A lengthy span may include data from outdated market regimes, potentially yielding insights that do not apply to present conditions. Shorter samples, on the other hand, may not have enough severe events to correctly show tail risk (Alexander, 2009).

These shortcomings are compounded during periods of financial restriction. Pritsker (2006) contends that under certain conditions, the HS methodology may produce inaccurate estimates, especially where accurate risk assessment is essential. Regulators have therefore acknowledged the shortcomings of HS. The Basel Committee has said that HS doesn't do a good job of capturing high-risk events, which is why VaR is no longer the only measure of regulatory capital (Bank for International Settlements, 2019). Even yet, HS is still a common benchmark because it's easy to understand and use, and it's a good starting point for more complicated non-parametric or hybrid methods (Pérignon & Smith, 2010).

2.3. Boudoukh, Richardson & Whitelaw

In risk management, it is widely acknowledged that recent observations yield a more precise evaluation of current portfolio risk. Boudoukh, Richardson, and Whitelaw (1998) put forward an enhancement to the basic Historical Simulation method called the BRW methodology. This technique puts more weight on recent data than on old data. This helps the model understand how volatility regimes change and gives a more accurate estimate of VaR.

The BRW approach gives each return a weight that gets smaller and smaller with time. The next statement tells you how much weight to give to a return seen i days ago:

$$w_i = \frac{\lambda^i(1-\lambda)}{(1-\lambda)^{\mathrm{T}}} \quad (3)$$

where $\lambda \in (0,1)$ is the decay factor and T is the total number of observations. The total of all the weights adds up to one, which keeps the model in line with the probabilistic framework of the historical distribution.

The decay factor λ is the most important part of the BRW model. When λ becomes close to one, it means that the observations are slowly fading away, which means that older ones still have some effect. This makes the VaR estimates more consistent, but it makes them less able to react to unexpected changes in the market. By contrast, lower values of λ make it easier to react to recent returns, which makes it easier to adjust to changes in volatility more quickly but also adds more noise to risk evaluations (Žiković & Aktan, 2010).

According to Žiković and Aktan (2011), choosing λ is very important for how well the BRW model works because it sets the balance between stability and responsiveness. Empirical evidence frequently indicates satisfactory performance with decay factors of about 0.99. However, a universally optimal selection is lacking.

Mehta, Neukirchen, Pfetsch, and Poppensieker (2012) note that financial organizations using the BRW technique usually use longer observation periods than those using traditional Historical Simulation. These windows usually last for two to three years. This is because of how the model is built. Older results aren't thrown away; instead, they are given less weight over time.

The BRW model is a big step forward from basic Historical Simulation as a whole. It gives you risk estimates that are more accurate for the current market by utilizing a system that updates volatility with weights that get smaller. This keeps the natural appeal of non-parametric methods (Boudoukh, Richardson, and Whitelaw, 1998).

2.4 Hull & White

Hull and White (1998) made a change to the basic Historical Simulation by adding changes in volatility to the risk evaluation framework. Their method changes the prior results by using a volatility factor to make them fit with the market's current level of volatility. The adjustment can be derived as follows:

$$r_t^* = \frac{\sigma^T}{\sigma^t} r^t \tag{4}$$

In this equation, r^t is the return on day t, σ^t is the volatility on day t, and σ^T is the current market volatility. This change improves returns when volatility is higher than it has been in the past. When volatility is lower than it has been in the past, returns go down. This makes VaR estimates more responsive, but it could also make things less stable, since adjusted losses could be higher than the highest loss in the historical series (Pritsker, 2006). Mehta, Neukirchen, Pfetsch, and Poppensieker (2012) observe that alterations are frequently implemented solely when the volatility factor exceeds one. Basu (2011) emphasizes that volatility clustering, characterized by extended periods of increased or decreased volatility, necessitates precise volatility forecasts for this methodology.

2.4.1. GARCH model

Bollerslev (1986) created the GARCH model (Generalized Autoregressive Conditional Heteroskedasticity) based on Engle's (1982) ARCH model. It was made to find groups of volatility. The conditional variance changes in the following way in the GARCH (p,q) model:

$$\sigma_t{}^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{(t-1)}{}^2 + \sum_{i=1}^{p} \beta_j \sigma_{(t-1)}{}^2 \tag{5}$$

In this equation, $\sigma_t{}^2$ denotes the conditional variance at time t, ω is a positive constant, α_i (where $\alpha_i \geq 0$) measures the impact of preceding squared shocks $\varepsilon_{(t-1)}{}^2$, andβ_j (where $\beta_j \geq 0$) signifies the persistence of prior conditional variances $\sigma_{(t-1)}{}^2$. The GARCH (1,1) is one of the most common forms:

$$\sigma_t{}^2 = \omega + \alpha r_{(t-1)}{}^2 + \beta \sigma_{(t-1)}{}^2 \tag{6}$$

where $r_{(t-1)}{}^2$ is the return from the previous period, α is the measure of how volatility reacts to market shocks, β is the measure of how long volatility lasts, and ω is the measure of the long-term variance component. This simple criterion is often used to figure out how risky a market is (Hull & White, 1998).

2.4.2. EWMA model

The Exponentially Weighted Moving Average (EWMA) model is a specific instance of the GARCH (1,1) model, omitting long-term mean reversion. This model calculates the conditional variance as a weighted average of the variance from the previous period and the squared return from the most recent observation:

$$\sigma_t{}^2 = \lambda \sigma_{(t-1)}{}^2 + (1-\lambda) r_{(t-1)}{}^2 \tag{7}$$

In this case, $\sigma_t{}^2$ is the conditional variance at time t, $\sigma_{(t-1)}{}^2$ is the variance from the previous period, $r_{(t-1)}{}^2$ is the squared return from the previous day, and $\lambda \in (0,1)$ is the decay factor. A smaller λ gives more weight to recent outcomes, which makes the model more responsive but less stable. On the other hand, a bigger λ makes volatility projections less accurate but makes the model less responsive (Žiković & Aktan, 2011).

Hull and White (1998) demonstrated the integration of EWMA into volatility-adjusted Historical Simulation, thereby enhancing the model's sensitivity relative to traditional HS. Basu (2011) states that EWMA makes it easier to respond to volatility, but if λ is not set correctly, it might cause instability. Mehta, Neukirchen, Pfetsch, and Poppensieker (2012) note that institutions adjust λ according to the portfolio's traits and the current market conditions.

2.5 Backtesting

Backtesting is a key aspect of risk management that ensures that a Value at Risk model's predictions are right. The model checks to see if the number and kind of exceptions match what it expected by comparing the estimated losses with the actual profit and loss (P&L) data (Christoffersen, 2008).

As part of the regulations for the internal models approach to market risk, the Basel Committee defined standards for backtesting in 1996 (BIS, 1996). The Fundamental Review of the Trading Book (FRTB) made these regulations a lot harder to follow. Even though the regulations changed from Value at Risk to Expected Shortfall (BIS, 2019), backtesting was still necessary.

The evaluation of VaR forecasts begins with the definition of a hit sequence, a binary variable that records whether a violation occurs:

$$Z_{t+1} = \begin{cases} 1, & if \;\; L_{t+1} < -VaR_t{}^{\alpha} \\ 0, & if \;\; L_{t+1} \geq -VaR_t{}^{\alpha} \end{cases} \tag{8}$$

L_{t+1} tells how much the portfolio lost on day t+1, and $VaR_t{}^{\alpha}$ tells you how much it is likely to lose on day t with a confidence level of α. The series Z_t is a binary time series that indicates the days when the actual losses are bigger than the projected losses (Christoffersen, 2008).

Pajhede (2015) emphasizes that this binary encoding enables the utilization of statistics to ascertain whether the sequence fulfills the criteria for an effective VaR model. Christoffersen (1998) asserts that a VaR model is considered legitimate if it satisfies three criteria:

i. Unconditional coverage: the unconditional likelihood of a violation is equal to the nominal coverage rate k = 1 − α.

$$H_{UC} = P\,(Z_t = 1) = k \tag{9}$$

ii. Independence: violations must happen on their own over time.

$$H_{IND} = P\,(Z_t = 1|\,\mathcal{F}_{t-1}) = P\,(Z_t = 1) \tag{10}$$

iii. Conditional coverage: unconditional coverage and independence are both true at the same time.

$$H_{CC} = P\,(Z_t = 1|\,\mathcal{F}_{t-1}) = P\,(Z_t = 1) = k \tag{11}$$

2.5.1 Unconditional Coverage Test (Kupiec, 1995)

The Proportion of Failures (POF) test, which Kupiec devised in 1995, checks to determine if the quantity of exceptions is what it should be based on the coverage level that was set. The hit sequence's likelihood function is: if N_0 is the number of times something goes wrong and N_1 is the number of times it goes right, then:

$$\mathcal{L}(\pi) = (1 - \pi)^{N_0}\,\pi^{N_1} \tag{12}$$

where π is the true probability of a violation. The maximum likelihood estimate is:

$$\hat{\pi} = \frac{N_1}{N} \tag{13}$$

The restricted likelihood under the null hypothesis that π = k is:

$$\mathcal{L}(\kappa) = (1 - \kappa)^{N_0}\,\kappa^{N_1} \tag{14}$$

The likelihood ratio statistic is:

$$LR_{UC} = -2\ln\left[\frac{(1-\kappa)^{N_0}\,\kappa^{N_1}}{(1-\hat{\pi})^{N_0}\,\hat{\pi}^{N_1}}\right] \tag{15}$$

It possesses a chi-squared distribution with a single degree of freedom. Rejecting the null hypothesis indicates that the model fails to provide accurate unconditional coverage.

2.5.2 Independence Test (Christoffersen, 1998)

The unconditional coverage test does not consider the temporal dependency of exceptions. Empirical studies demonstrate that financial returns often exhibit volatility clustering, which may lead to breaches throughout consecutive periods (Pritsker, 2005).

Christoffersen (1998) created the Independence Test, employing a first-order Markov chain. The transition probability matrix is defined in this way:

$$\pi_1 = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix} \tag{16}$$

where $\pi_{01} = \mathrm{P}(Z_{t+1} = 1 \mid Z_t = 0)$ and $\pi_{11} = \mathrm{P}(Z_{t+1} = 1 \mid Z_t = 1)$. The likelihood function of the observed sequence is:

$$\mathcal{L}(\Pi_1) = (1-\pi_{01})^{N_{00}}\,\pi_{01}^{N_{01}}\,(1-\pi_{11})^{N_{10}}\,\pi_{11}^{N_{11}} \tag{17}$$

with maximum likelihood estimates:

$$\hat{\pi}_{01} = \frac{N_{01}}{N_{00}+N_{01}} \tag{18}$$

and

$$\hat{\pi}_{11} = \frac{N_{11}}{N_{10}+N_{11}} \tag{19}$$

The likelihood ratio statistic for independence is:

$$LR_{IND} = -2\ln\left[\frac{\mathcal{L}(\hat{\pi})}{\mathcal{L}(\widehat{\Pi_1})}\right] \tag{20}$$

which follows a chi-squared distribution with one degree of freedom. Rejection implies that exceptions are clustered, violating independence.

2.5.3 Conditional Coverage Test (Christoffersen, 1998)

Christoffersen (1998) also came up with the Conditional Coverage Test, which is a combination of the unconditional coverage and independence criterion. The test statistic is:

$$LR_{CC} = LR_{UC} + LR_{IND} \tag{21}$$

It has two degrees of freedom and is chi-squared as it gets closer to zero. If rejected, the VaR model does not meet the requirements of unconditional coverage and independence at the same time, which makes it useless.

2.6. COVID-19 and Market Stress

The COVID-19 crisis provided a context for assessing non-parametric Value at Risk algorithms, underscoring their efficacy and constraints. Numerous banks and financial institutions employed the Historical Simulation (HS) method, which failed to respond promptly to the abrupt increase in volatility. In early 2020, among substantial daily market declines, HS misassessed the actual risk level due to the application of an equal weighting technique that neglected the concentration of losses in the left tail of the distribution (Shaik & Padmakumari, 2022). The problem was evident in stock indices such as the S&P 500, where backtesting revealed a significantly higher number of outliers than the theoretical confidence level would indicate.

The Hull and White (HW) volatility-adjusted historical simulation demonstrated enhanced precision. This method enables a more rapid response to sudden changes in market risk by adjusting previous data based on the relationship between current and past volatility. The HW model demonstrated particular proficiency in detecting sudden volatility changes during the pandemic, yielding VaR estimates that closely matched actual losses (Omari et al., 2020).

Additional studies in international markets support these findings. In the BRICS economies, HS consistently underestimated losses in 2020. However, HW adjustments provided more accurate backtesting results in both unconditional and conditional coverage evaluations (Ho, 2021). In March 2020, banks in major economies experienced a significant volume of backtesting exceptions under HS. Conversely, volatility-weighted methodologies such as HW exhibited superior effectiveness in illustrating market changes (Abboud, 2021).

2.7. Bitcoin and Risk Measurement

Since its establishment in 2009, Bitcoin has attracted considerable attention as both a groundbreaking technology and a financial instrument. Originally presented as a decentralized peer-to-peer payment system, it has transformed into what numerous investors consider a speculative asset and, more recently, a prospective store of value commonly termed “digital gold” (Baur, Hong, & Lee, 2018). However, its price history has exhibited considerable volatility, complicating portfolio diversification and, more importantly, the quantification of market risk (Dyhrberg, 2016).

Individuals have included Bitcoin into their financial portfolios to enhance diversification and safeguard against inflation and market volatility (Klein, Pham Thu, & Walther, 2018). Research indicates that its minimal correlation with conventional assets makes it attractive for diversified portfolios (Bouri et al., 2017). Further research indicates that Bitcoin often exhibits a correlation with global equities during systemic crises, such as the COVID-19 market collapse, thereby challenging its classification as a safe haven (Corbet, Larkin, & Lucey, 2020). Recent data from the same timeframe indicated that Bitcoin represented a riskier investment compared to a stable one for investors (Conlon & McGee, 2020).

Given that Bitcoin possesses both advantages and disadvantages, it is essential to scrutinize its deficiencies comprehensively. A prevalent method for quantifying this is VaR. Conventional Value at Risk methodologies frequently prove inadequate for Bitcoin due to its pronounced heavy tails and ongoing volatility clustering in returns. Chkili and Naoui (2021) contend that during periods of increased volatility, conventional VaR models often underestimate potential losses due to their inability to account for significant market movements. Ardia, Bluteau, and Rüede (2019) contend that parametric and semi-parametric Value at Risk models are unsuitable for Bitcoin because they overlook its skewness and excess kurtosis.

Conversely, HS has demonstrated greater reliability for cryptocurrencies. Likitratcharoen, Chudasring, Pinmanee, and Wiwattanalamphong (2023) determined that HS was the most efficacious strategy for Bitcoin and other digital assets during periods of stress, including the COVID-19 pandemic. Their findings indicate that HS yielded superior downside risk predictions compared to Delta-Normal or Monte Carlo simulations, while also resulting in fewer backtesting violations.

Multiple methodologies have been analyzed to improve the effectiveness of VaR concerning the unique characteristics of Bitcoin. Kwon (2021) employs the Conditional Autoregressive Value-at-Risk (CAViaR) framework to demonstrate that factors such as trading activity, investor sentiment, and policy uncertainty significantly influence Bitcoin's tail risk. Gkillas, Konstantatos, and Tsagkanos (2023) illustrate that specific models work adequately in steady conditions yet fail to comply with regulatory norms during crises, underscoring the challenges of risk modeling in cryptocurrency markets. Volatility-based approaches support these findings, as Katsiampa (2017) shows that GARCH-type models improve the accuracy of VaR forecasts by including the conditional heteroskedasticity and volatility clustering observed in Bitcoin returns.

The literature shows that VaR is a key tool for measuring Bitcoin's market risk. However, its effectiveness depends on the method used and the state of the market at the time. In stable settings, classical models may provide satisfactory approximations, although they often fail to accurately assess the risks associated with extreme values. Data shows that Historical Simulation can still provide accurate and useful predictions during crises, showing that it works well for controlling risk in bitcoin markets.

3. Data and Methodology

This study evaluates the effectiveness of three non-parametric methods for assessing market risk through VaR: HS, BRW, and HW. Per BIS rules, VaR is calculated at a 99% confidence level. This is universally recognized as the benchmark for assessing market risk (BIS, 2019).

The empirical study employs daily data from January 1, 2018, to December 31, 2022, covering both typical market conditions and a stress scenario associated with the COVID-19 pandemic. The sample was divided into three subperiods to encompass these distinct settings:

- **Before COVID-19:** 01/01/2018 to 31/12/2019
- **COVID-19:** 01/01/2020 to 31/12/2020
- **After COVID-19:** 01/01/2021 to 31/12/2022

The assets under evaluation include Bitcoin and the S&P 500 Index. Bitcoin was selected due to its significant volatility and speculative nature (Bouri et al., 2017), while the S&P 500 serves as a benchmark for the conventional stock market. All amounts were converted to euros (EUR) utilizing the USD/EUR exchange rate. A hypothetical investment of €1,000,000 was allocated to each asset, and the profit and loss (P&L) series was produced through daily rebalancing of the investment.

It was employed logarithmic returns to calculate the daily returns:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{22}$$

P_t denotes the asset's price at time t. Jorion (2007) argued that geometric returns offer a more accurate representation of investment performance than arithmetic returns, hence supporting their widespread use in Finance.

Table 1 presents the maximum and minimum daily fluctuations, as well as the annualized volatility for both assets for the whole sample period. This constitutes the initial phase in describing the data.

Table 1: Maximum and Minimum Variations, and Volatility

	Bitcoin	S&P500
Maximum Variation	24,3%	9,6%
Minimum Variation	-47,1%	-13,4%
Volatility	4,1%	1,3%

Table 1 indicates that Bitcoin's daily price fluctuations varied from a maximum of +24.3% to a minimum of −47.1%. The annualized volatility was 4.1%. The S&P 500 experienced a peak daily increase of +9.6% and a trough of −13.4%, accompanied by an annualized volatility of 1.3%.

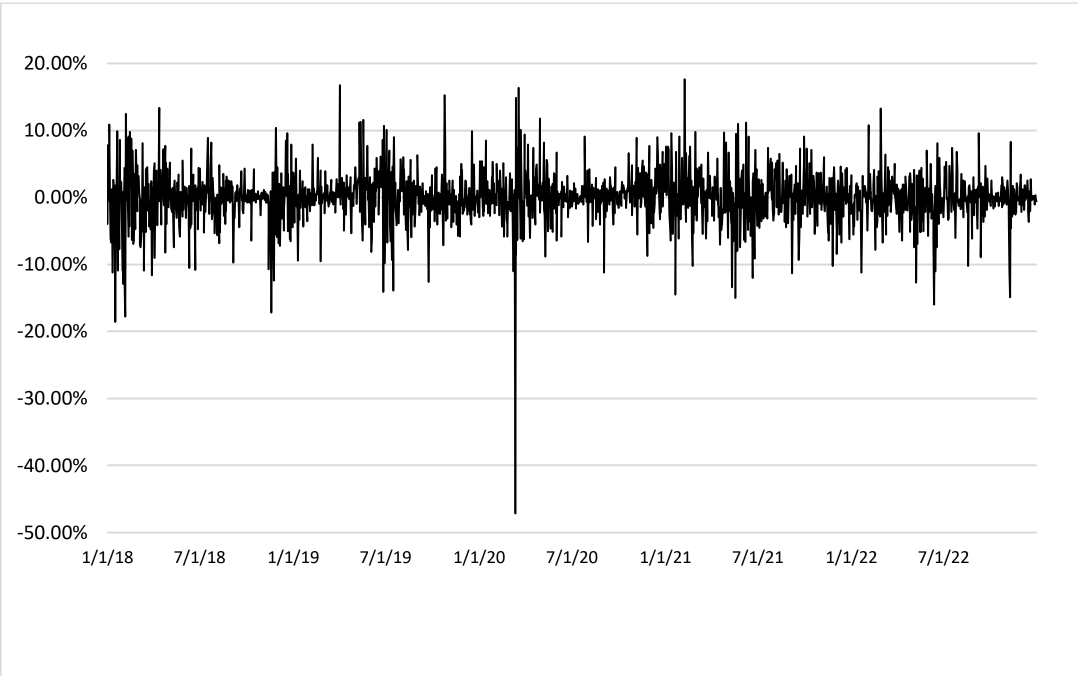

Figure 1: Bitcoin Daily Logarithmic Returns

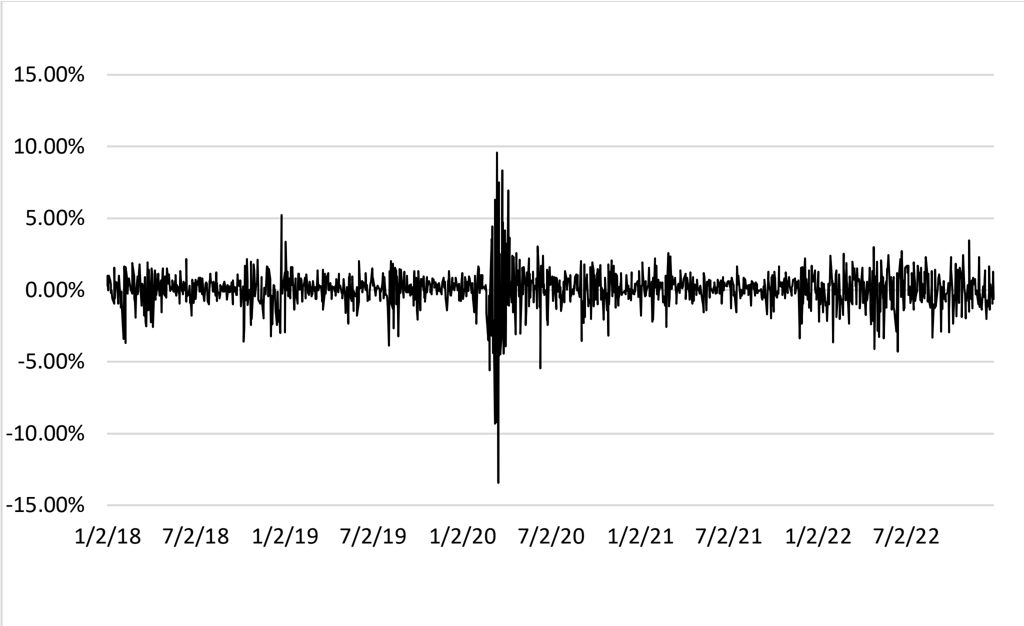

Figure 2: S&P500 Daily Logarithmic Returns

Figure 1 and Figure 2 reveal that Bitcoin and S&P500 have very different behaviors. Bitcoin has much sharper fluctuations, reflecting higher volatility. On the other hand, The S&P 500 shows more stable patterns, with changes that have a smaller magnitude. This reinforces the notion that Bitcoin is riskier and unpredictable when compared with S&P500.

4.Results

4.1 VaR under Normal Market Conditions

The following analysis outlines the VaR results for Bitcoin and the S&P 500 under normal market conditions, namely before and after the COVID-19 pandemic. The three non-parametric approaches under consideration—HS, BRW, and HW—are applied to both assets. The results are presented individually for each method and subperiod to facilitate a comprehensive comparison.

HS model computes Value at Risk as the empirical 1% quantile of the distribution of returns, offering the most direct and unmodified representation of previous losses. Table 2 presents the results from the pre-pandemic period.

Table 2: HS VaR, Before COVID-19

Before COVID-19		
Historical Simulation	**Bitcoin**	**S&P500**
VaR	**125 047,72**	**32 118,67**

Values in EUR

The HS estimates indicate that Bitcoin is significantly more volatile than the S&P 500. This is due to the thicker tails and heightened volatility of bitcoin returns. As expected, the method integrates both contemporary and past extreme events, as all observations are assigned equal weight. This explains the relative steadiness of the stock index compared to the more erratic behavior of Bitcoin.

The BRW technique, illustrated in Table 3, enhances HS by assigning exponentially diminishing weights to historical returns, hence prioritizing the most recent observations. Boudoukh, Richardson, and Whitelaw (1998) recommended setting the decay factor at λ=0.99. This ensured responsiveness to current volatility while maintaining consistency with historical data.

Table 3: BRW VaR, Before COVID-19

Before COVID-19		
BRW	**Bitcoin**	**S&P500**
VaR	**94 842,69**	**29 869,14**

Values in EUR

BRW generates lower revenue than HS at this period. This phenomenon is particularly pronounced for Bitcoin, as several of its most significant losses occurred earlier in the dataset and were thus assigned diminished weight. The disparity between the S&P 500 and HS is diminished, aligning with the index's more stable return distribution. BRW is designed to mitigate the effects of distant shocks, which is referred to as the "ghost effect" in basic Historical Simulation.

The performance of the BRW model confirms the conclusion of Boudoukh, Richardson and Whitelaw (1998) and Zikovic and Aktan (2011), who demonstrated that introducing a decay factor improves the model's responsiveness to prevailing market dynamics.

The HW approach, illustrated in Table 4, enhances the estimate by adjusting returns according to conditional volatility. We employed an Exponentially Weighted Moving Average (EWMA) method to ascertain volatility, utilizing maximum likelihood to determine the smoothing parameter λ. The maximal log-likelihood function is:

$$\max \sum_{t=1}^{T} \left[-\ln(\sigma_t^2) - \frac{r_t^2}{2\,\sigma_t^2} \right] \tag{23}$$

Thus, λ equals 0.9652 for Bitcoin and 0.9067 for the S&P 500. It was maintained that these variables were consistent throughout the analysis. The conditional variance processes are as follows:

$$\sigma_{\text{BTC},t}^2 = 0.9652\,\sigma_{\text{BTC},t-1}^2 + (1 - 0.9652)\,r_{\text{BTC},t-1}^2 \tag{24}$$

and

$$\sigma_{\text{S\&P},t}^2 = 0.9067\,\sigma_{\text{S\&P},t-1}^2 + (1 - 0.9067)\,r_{\text{S\&P},t-1}^2 \tag{25}$$

Annex A shows how the EWMA shows the conditional volatility for Bitcoin and the S&P 500. The S&P 500 has an estimated volatility that fluctuates between 0 and 3%. Most trading days exhibit a modest degree of volatility, and there were only brief spikes during the COVID-19 shock before things went back to normal. Bitcoin is less predictable since it has a higher baseline, different volatility clusters, and a lot of abrupt shifts. This was evident not just during the stress crisis in 2020, but also in more stable markets, with substantial movements in 2018 and recurring surges in 2021 and 2022. Bitcoin is more volatile and its price changes more often, while the stock index has a more stable path.

Table 4: HW VaR, Before COVID-19

Before COVID -19		
HW	**Bitcoin**	**S&P500**
VaR	**209 550,42**	**64 527,41**

Values in EUR

As expected, HW yields results that are more conservative than those of HS and BRW. This is particularly applicable to Bitcoin, as its tail estimates are significantly influenced by volatility clustering. HW also yields significantly elevated values for the S&P 500, indicating that even in relatively stable markets, accounting for volatility can substantially enhance risk assessments.

The robustness of the HW approach in calm periods is in line with Basu (2009), who emphasized the superiority of volatility-adjusted historical simulation over purely non-parametric methods.

The results for the post-pandemic period indicate similar disparities among methods. Table 5 presents the estimations for HS.

Table 5: HS VaR, After COVID-19

After COVID-19		
Historical Simulation	**Bitcoin**	**S&P500**
VaR	**112 778,74**	**33 068,35**

Values in EUR

During this sub-period, HS indicates that Bitcoin's VaR has decreased compared to pre-epidemic levels, while S&P 500 values are slightly elevated. The technique, however, provides a static representation of tail losses, disregarding conditional dynamics.

Table 6 presents the BRW results during the post-epidemic period.

Table 6: BRW VaR, After COVID-19

After COVID-19		
BRW	**Bitcoin**	**S&P500**
VaR	**111 894,07**	**33 562,84**

Values in EUR

The BRW and HS estimations for Bitcoin are presently closely aligned, contrasting with their previous lower values. The heightened focus on recent observations, specifically the significant declines in 2021, precipitated this alteration. BRW assigns the S&P 500 somewhat elevated values compared to HS, which exhibits a similar phenomenon. The model's weighting approach enhances its sensitivity to recent fluctuations in volatility.

Table 7 presents the HW outcomes after the outbreak.

Table 7: HW VaR, After COVID-19

After COVID-19		
HW	**Bitcoin**	**S&P500**
VaR	**212 650,08**	**39 846,24**

Values in EUR

HW still has the highest VaR values of the three models. The value of Bitcoin is still high, which shows that volatility clustering is still happening. The difference between the S&P 500 and HS, and BRW gets less, which is in line with a lower degree of conditional volatility in the next quarter. The model's design makes sure that its results take these modifications into account.

Distinct trends emerge when comparing the two time periods together. For Bitcoin, HS indicates a decrease in risk post-COVID-19, while BRW experiences an increase and aligns more closely with HS. HW remains the most prudent in both subperiods, with just minor fluctuations. Post-COVID-19, HS and BRW exhibit a slight increase in risk for the S&P 500, but HW demonstrates a significant decline relative to its pre-pandemic assessment. These divergent trends indicate that each strategy emphasizes certain aspects of the return distribution: HS encompasses the entire historical tail, BRW is very responsive to recent disturbances, and HW considers volatility persistence, resulting in consistently elevated risk evaluations.

4.2 VaR under Market Stress

The COVID-19 pandemic of 2020 represented a period of significant instability in modern financial history. The markets exhibited unprecedented instability. Initially, they incurred significant financial losses in March. However, they subsequently recovered robustly. Subsequently, substantial shifts and market stress emerged. This environment serves as a significant test of the effectiveness of non-parametric Value-at-Risk approaches. The findings indicate that the three methods - HS, BRW, and HW - performed effectively for Bitcoin and the S&P 500 during periods of high volatility.

Table 8 presents the HS estimates for the COVID-19 period.

Table 8: HS VaR, During COVID-19

During COVID-19		
Historical Simulation	**Bitcoin**	**S&P500**
VaR	**87 603,13**	**73 545,16**

Values in EUR

The data indicate that the VaR for the S&P 500 is significantly elevated compared to typical market conditions. Many equities experienced a decline in value in March 2020, which explains the subsequent events. HS indicates that Bitcoin's risk profile has improved compared to previous periods, suggesting that its most significant adverse events occurred prior to the crisis year. The data indicates that the stress induced by the pandemic exerted a more pronounced direct impact on stocks compared to cryptocurrencies, which experienced a smaller decline in value in 2020.

Table 9 presents the outcomes obtained from the BRW approach.

Table 9: BRW VaR, During COVID-19

During COVID-19		
BRW	**Bitcoin**	**S&P500**
VaR	**87 783,00**	**49 070,44**

Values in EUR

BRW assigns a value to Bitcoin that closely approximates HS's value. The weighted method yields comparable results when losses are uniformly distributed throughout the entire term. Conversely, BRW assigns a significantly lower value to the S&P 500 compared to HS. As the year progresses, the significant losses from March 2020 are assigned diminishing importance, thereby reducing their impact on the quantile estimate. The comparison illustrates BRW's response to the distribution of shocks across the sample. The strategy assigns reduced significance to simultaneous shocks, such as those affecting stocks. Conversely, when shocks are more dispersed, as observed with Bitcoin, the outcomes resemble those of HS more closely.

Table 10 presents the homework results for the identical time frame.

Table 10: HW VaR, During COVID-19

During COVID-19		
HW	**Bitcoin**	**S&P500**
VaR	**157 861,78**	**532 525,23**

Values in EUR

The HW approach yields the most conservative estimations, as anticipated. The values of Bitcoin remain higher than those of HS and BRW. However, the associated risk level has decreased compared to earlier subperiods, aligning with predictions from previous models. The HW prediction for the S&P 500 significantly exceeds the predictions made by HS and BRW. Significant fluctuations in volatility occurred in March 2020, which explains this phenomenon. The increased size of these spikes is attributable to the model's volatility-scaling algorithm. The HW model indicates that stocks exhibit significant conditional risk, while Bitcoin's values remain more stable, albeit at lower levels during this period. This clarifies the contrast between the two assets.

The COVID-19 pandemic had distinct implications for each of the two markets. All three models indicate that the VaR for Bitcoin was reduced during the pandemic compared to the periods before and after. Before the crisis year, cryptocurrencies experienced the greatest financial losses. The pandemic increased the risk associated with the S&P 500. HS and BRW experienced greater financial losses, whereas HW demonstrated the volatility of March 2020 by reporting figures significantly higher than those of the other methods. This disparity illustrates the differential impact of COVID-19 on various asset classes. Stocks experienced the most significant decline; however, Bitcoin did not encounter its largest drops in 2020, despite its considerable volatility.

4.3 Backtesting Results

A range of backtesting methodologies was employed to evaluate the effectiveness of non-parametric techniques in forecasting Value-at-Risk. Backtesting directly evaluates the predictive efficacy of models by comparing the frequency and pattern of violation days when actual losses exceed the VaR estimate with the theoretical expectation derived from the assumed 99% confidence level. The statistical framework is based on three tests: the unconditional coverage test by Kupiec (1995), the independence test by Christoffersen (1998), and the conditional coverage test, which combines the two. A rolling window of around one year was used, with 250 trading-day observations for the S&P 500 and 360 observations for Bitcoin, which trades all week, to make daily VaR forecasts that were used in backtesting. This distinction ensures that assets with varied trading schedules can be compared.

The first step in evaluating the models consists of counting the number of violations observed. Tables 14, 15, and 16 in Annex B report these values for HS, BRW, and HW, respectively.

The comparison of raw violation counts already reveals clear differences between the three models. Before COVID-19, all methods were relatively close in both assets, with HW consistently producing the lowest number of violations and HS the highest. During the pandemic, Bitcoin experienced fewer violations across all models, with only one under HW, confirming that the most extreme cryptocurrency losses occurred outside 2020. By contrast, the S&P 500 registers a sharp contrast between models: HS produces eight exceptions, BRW reduces the number to four, and HW only two. After the pandemic, Bitcoin exhibits an increase in exceptions, peaking at eleven under HS, while the S&P 500 displays a more moderate pattern, ranging between five and nine.

The null hypotheses of unconditional and conditional coverage are not rejected at the 1% level, indicating that the Historical Simulation (HS) model meets standard backtesting criteria for both assets across the pre- and post-pandemic periods. However, HS does not fit these requirements for the S&P 500 during the COVID-19 period. A discrepancy between the actual frequency and temporal pattern of exceptions and the nominal confidence level is indicated by the rejection of the null by both the Christoffersen conditional coverage test and the Kupiec proportion-of-failures test. On the other hand, for Bitcoin, the HS model remains statistically acceptable in all three periods, with no evidence against either unconditional or conditional coverage.

The BRW model exhibits consistent backtesting performance across assets and periods, with no rejections at the 1% significance level. It adjusts better during the COVID-19 episode than the pure HS benchmark: it meets both unconditional and conditional coverage requirements for the S&P 500, while HS does not. The exponentially decaying weighting scheme, which gives more weight to recent observations and enables quicker adjustment to changes in market volatility, is consistent with this outcome.

The HW procedure provides the most reliable performance among the competing specifications. It does not fail any backtesting criterion and generates the fewest exceptions among all asset–period combinations. With one exception for Bitcoin and two for the S&P 500, it continues to be extremely conservative during the COVID-19 pandemic while meeting the requirements for unconditional, independent, and conditional coverage. These outcomes align with the HW approach's volatility-scaling step, which reduces return clustering and adapts to changes in the market regime. The greater consistency of HW model is consistent with Likitratcharoen et al. (2023), who identified superior performance of these approaches for cryptocurrencies under heightened uncertainty.

The backtesting evidence demonstrates substantial disparities among models, assets, and market regimes. In the pre-COVID period, all requirements fulfilled the conventional coverage criteria. HW generally has the fewest exceptions. The HS approach doesn't work as well for the S&P 500 during the COVID-19 period, but the BRW method works better. The HW method is still the most reliable. After the crisis, all models pass, but HS still looks weaker than the others. These results highlight the limitations of basic historical simulation, the marginal benefits of recency weighting in BRW, and the consistent reliability of HW. In line with previous research, they demonstrate that volatility adjustment in the Hull and White (1998) model provides the most reliable non-parametric VaR during market stress.

5. Conclusion

This dissertation explores how non-parametric techniques can be applied to the measurement of market risk, with particular emphasis on the Historical Simulation method (HS) and its extensions proposed by Boudoukh, Richardson and Whitelaw (BRW), as well as Hull and White (HW). The empirical framework juxtaposes two assets with contrasting characteristics: the S&P 500, which reflects the performance of a broad and established equity market, and Bitcoin, which epitomises the extreme volatility typical of cryptocurrencies. The evaluation is carried out over three distinct intervals that capture different market environments—spanning the period preceding the COVID-19 pandemic, the crisis itself, and the subsequent recovery phase—thus allowing the models to be tested under both normal conditions and a major episode of systemic stress.

The results confirmed that the choice of technique significantly influences the accuracy of VaR forecasts. The HS approach, although simple and transparent, revealed structural weaknesses, particularly in periods of stress. This observation aligns with Pritsker (2006), who noted that HS frequently fails when precise measurement is critical. HS assessed that the S&P 500 posed less risk compared to its status during the COVID-19 crisis, resulting in numerous violations in backtesting and indicating that the model was inadequate for managing swift fluctuations in volatility.

Boudoukh, Richardson, and Whitelaw (1998) developed the BRW approach to enhance responsiveness by assigning greater weight to recent observations. Backtesting indicated that BRW outperformed HS during the stress period analyzed in this dissertation. The method avoided the distortions caused by remote shocks characteristic of HS by emphasizing recent data. However, BRW encountered issues as its performance was significantly influenced by the distribution of shocks throughout the sample period. This aligns with the findings of Žiković and Aktan (2011).

The Hull and White (1998) HW technique proved to be the most reliable strategy. HW provided more conservative and precise risk assessments by adjusting returns according to conditional volatility. The backtesting results indicated that HW successfully met all statistical criteria across all assets and time frames. This indicates robustness in both normal and stressed conditions. This aligns with Basu's (2009) assertion that incorporating volatility is crucial for effective stress testing and risk assessment.

There are clear distinctions between Bitcoin and the S&P 500. In line with Bouri et al. (2017), Bitcoin shows heavy-tailed return distributions and consistently high volatility.

Backtesting of the non-parametric VaR specifications shows the same structural pattern observed for equities: the simple HS approach deteriorates under stress, BRW improves through recency weighting but remains short of full adequacy, and the HW volatility-scaled procedure delivers the most robust coverage across regimes. In line with Corbet et al. (2020), the S&P 500 absorbed the COVID-19 shock more abruptly, whereas Bitcoin's drawdowns materialized more gradually, yielding a less pronounced left tail over 2020 within our sample.

The backtesting methodologies developed by Christoffersen (1998) were essential for validating the results' accuracy. The statistical data revealed that HW demonstrated the most effective coverage, BRW adjusted to the circumstances, and HS was insufficient under challenging conditions.

From these results, it becomes evident that non-parametric techniques for Value-at-Risk estimation can still serve as relevant instruments in contemporary risk management. Their effectiveness, however, is not uniform across specifications. Historical Simulation, despite its computational simplicity, does not provide reliable forecasts in periods of exceptional stress. The BRW variant improves upon this by attaching greater importance to recent outcomes, although its stability remains sensitive to the pattern of shocks. By contrast, the HW framework stands out as the most prudent and dependable option, delivering consistent risk assessments across different assets and under diverse market regimes.

The quantitative findings and the conclusions of this dissertation suggest an assessment of the regulatory approach to cryptocurrencies. Corbet's (2020) findings, which suggested that the pandemic increased volatility and exacerbated contagion among asset classes, contrast with the finding that Bitcoin's VaR levels decreased during the COVID-19 period. Such a divergence suggests that non-parametric VaR methods may not be sufficiently sensitive to the tail behavior of cryptoassets in periods of turbulence. Conlon (2020) reached a similar conclusion, noting that Bitcoin tended to increase the downside risk of portfolios instead of acting as a hedge. Taken together, these perspectives indicate that frameworks originally designed for conventional financial instruments, such as those set out in Basel standards, may underestimate the risks associated with highly speculative assets. For this reason, supervisors may eventually need to consider tailored approaches, including stricter capital requirements, when institutions hold significant exposures to cryptocurrencies.

Future investigations should test hybrid schemes that join non-parametric estimators with parametric volatility models, starting with GARCH, which captures volatility clustering in Bitcoin returns (Katsiampa, 2017). The analysis should also be repeated outside the COVID-19 window, re-centering it on other shocks - geopolitical events, monetary policy surprises, or liquidity squeezes - to check whether the main results carry over to different market conditions. A further refinement is to report Expected Shortfall alongside Value at Risk so that tail frequency and tail magnitude are assessed together, as proposed by Acerbi and Tasche (2002). Taken together, these extensions would give a fuller picture of risk in markets that move quickly and change regimes often.

References

Abboud, A., Anderson, C., Game, A., Iercosan, D., Inanoglu, H., & Lynch, D. (2021). Banks' Backtesting Exceptions during the COVID-19 Crash: Causes and Consequences. *FEDS Notes*, *2021.0*(2942). https://doi.org/10.17016/2380-7172.2939

Acerbi, C., & Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking & Finance*, *26*(7), 1487–1503. https://doi.org/10.1016/s0378-4266(02)00283-2

Ahelegbey, D. F., Giudici, P., & Mojtahedi, F. (2020). Tail risk measurement in crypto-asset markets. *International Review of Financial Analysis*, *73*, 101604. https://doi.org/10.1016/j.irfa.2020.101604

Alexander, C. (2009). Market Risk Analysis - Value at Risk Models (Vol. 4). *John Wiley & Sons*

Ardia, D., Bluteau, K., & Rüede, M. (2018). Regime changes in Bitcoin GARCH volatility dynamics. *Finance Research Letters*, *29*, 266–271. https://doi.org/10.1016/j.frl.2018.08.009

Artzner, P., Delbaen, F., Eber, J., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, *9*(3), 203–228. https://doi.org/10.1111/1467-9965.00068

Ballotta, L., & Fusai, G. (2017). A gentle introduction to value at risk. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2942138

Bank for International Settlements BIS (2019). Minimum capital requirements for market risk. Bank for International Settlements, 10, 89, 94 https://www.bis.org/bcbs/publ/d457.pdf

Basu, S. (2009). Stress testing for Market Risk: A comparison of VaR method, 11. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1370965

Basu, S. (2011). Comparing simulation models for market risk stress testing. European journal of operational research, 336. https://doi.org/10.1016/j.ejor.2011.02.023

Baur, D. G., Hong, K., & Lee, A. D. (2017). Bitcoin: Medium of exchange or speculative assets? *Journal of International Financial Markets Institutions and Money*, *54*, 177–189. https://doi.org/10.1016/j.intfin.2017.12.004

Beder, T. S. (1995). VAR: Seductive but dangerous. *Financial Analysts Journal*, *51*(5), 12–24. https://doi.org/10.2469/faj.v51.n5.1932

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

Chen, J. (2018). On Exactitude in Financial Regulation: Value-at-Risk, Expected Shortfall, and Expectiles. Risks. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3136278

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, *39*(4), 841. https://doi.org/10.2307/2527341

Christoffersen, P. (2008). Backtesting. *University of Toronto - Rotman School of Management; Copenhagen Business School; Aarhus University*, 1 - 8. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2044825

Chu, J., Chan, S., Nadarajah, S., & Osterrieder, J. (2017). GARCH modelling of cryptocurrencies. *Journal of Risk and Financial Management*, *10*(4), 17. https://doi.org/10.3390/jrfm10040017

Conlon, T., & McGee, R. (2020). Safe haven or risky hazard? Bitcoin during the Covid-19 bear market. *Finance Research Letters*, *35*, 101607. https://doi.org/10.1016/j.frl.2020.101607

Corbet, S., Meegan, A., Larkin, C., Lucey, B., & Yarovaya, L. (2018). Exploring the dynamic relationships between cryptocurrencies and other financial assets. *Economics Letters*, *165*, 28–34. https://doi.org/10.1016/j.econlet.2018.01.004

Corbet, S., Larkin, C., & Lucey, B. (2020). The contagion effects of the COVID-19 pandemic: Evidence from gold and cryptocurrencies. *Finance Research Letters*, *35*, 101554. https://doi.org/10f.1016/j.frl.2020.101554

Danielsson, J., Zhou, C. (2015). Why Risk Is So Hard to Measure. *De Nederlandsche Bank Working Paper No. 494*, 2. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2597563

Du, Z., & Escanciano, J. C. (2015). Backtesting expected shortfall: Accounting for tail risk. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2548544

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, *50*(4), 987. https://doi.org/10.2307/1912773

Hendricks, D. (1996). Evaluation of Value-at-Risk models using historical data. *Federal Reserve Bank of New York Economic Policy Review*, *2*(1), 39–69. https://www.newyorkfed.org/medialibrary/media/research/epr/96v02n1/9604hend.pdf

Ho, C. M., Pham, T. T., Nguyen, H. L., & Vo, D. H. (2021). Does the COVID-19 pandemic matter for market risks across sectors in Vietnam? *Heliyon*, *7*(12), e08453. https://doi.org/10.1016/j.heliyon.2021.e08453

Hull, J., & White, A. (1998). Incorporating volatility updating into the historical simulation method for value-at-risk. *The Journal of Risk*, *1*(1), 5–19. https://doi.org/10.21314/jor.1998.001

J.P. Morgan. (1996). *RiskMetrics – Technical Document* (4th ed.). New York: J.P. Morgan

Jorion, P. (2007). Value at Risk: The New Benchmark for Managing Financial Risk (3º ed.). The McGraw-Hill Companies.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, *3*(2), 73–84. https://doi.org/10.3905/jod.1995.407942

Kwon, J. H. (2021). On the factors of Bitcoin's value at risk. *Financial Innovation*, *7*(1). https://doi.org/10.1186/s40854-021-00297-3

Likitratcharoen, D., Kronprasert, N., Wiwattanalamphong, K., & Pinmanee, C. (2021). The Accuracy of Risk Measurement Models on Bitcoin Market during COVID-19 Pandemic. *Risks*, *9*(12), 222. https://doi.org/10.3390/risks9120222

Likitratcharoen, D., Chudasring, P., Pinmanee, C., & Wiwattanalamphong, K. (2023). The Efficiency of Value-at-Risk Models during Extreme Market Stress in Cryptocurrencies. *Sustainability*, *15*(5), 4395. https://doi.org/10.3390/su15054395

Mehta, A., Neukirchen, M., Pfetsch, S., & Poppensieker, T. (2012). Managing market risk: Today and tomorrow. McKinsey & Company McKinsey Working Papers on Risk. https://www.binhphuongnguyen.com/wp-content/uploads/2019/09/The-McKinsey-on-Risk-Evaluation-System.pdf

Nolde, N., & Ziegel, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation. *The Annals of Applied Statistics*, *11*(4). https://doi.org/10.1214/17-aoas1041

Omari, C., Mundia, S., & Ngina, I. (2020). Forecasting Value-at-Risk of Financial Markets under the Global Pandemic of COVID-19 Using Conditional Extreme Value Theory.

Journal of Mathematical Finance, *10*(04), 569–597. https://doi.org/10.4236/jmf.2020.104034

Pérignon, C., & Smith, D. R. (2009). Diversification and Value-at-Risk. *Journal of Banking & Finance*, *34*(1), 55–66. https://doi.org/10.1016/j.jbankfin.2009.07.003

Pritsker, M. (2005). The hidden dangers of historical simulation. *Journal of Banking & Finance*, *30*(2), 561–582. https://doi.org/10.1016/j.jbankfin.2005.04.013

Richardson, M. P., Boudoukh, J., & Whitelaw, R. (1998). The Best of both Worlds: a hybrid approach to calculating value at risk. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.51420

Shaik, M., & Padmakumari, L. (2022). Value-at-risk (VAR) estimation and backtesting during COVID-19: Empirical analysis based on BRICS and US stock markets. *Investment Management and Financial Innovations*, *19*(1), 51–63. https://doi.org/10.21511/imfi.19(1).2022.04

Žiković, S. & Aktan, B. (2010). Decay Factor Optimisation in Time Weighted Simulation - Evaluating VAR Performance, 2-3. University of Rijeka - Faculty of Economics; University of Bahrain. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1622704

Annex A: EWMA estimated volatilities for Bitcoin and S&P500

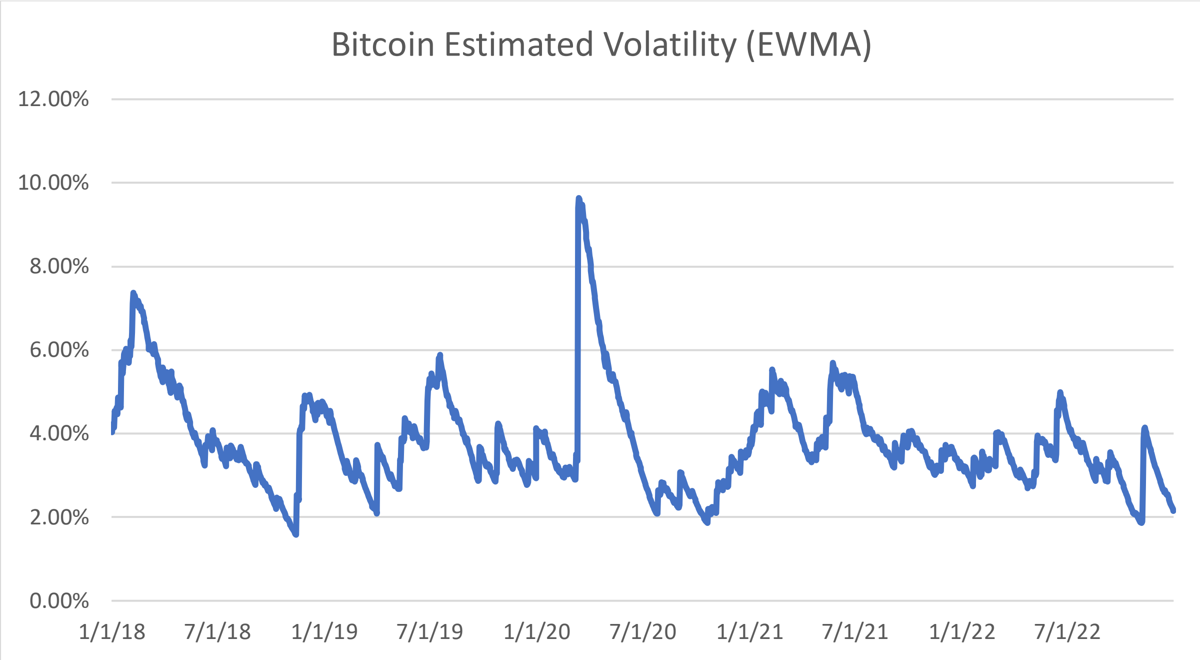

Figure 3: Graphic representation of EWMA estimated volatilitie for Bitcoin

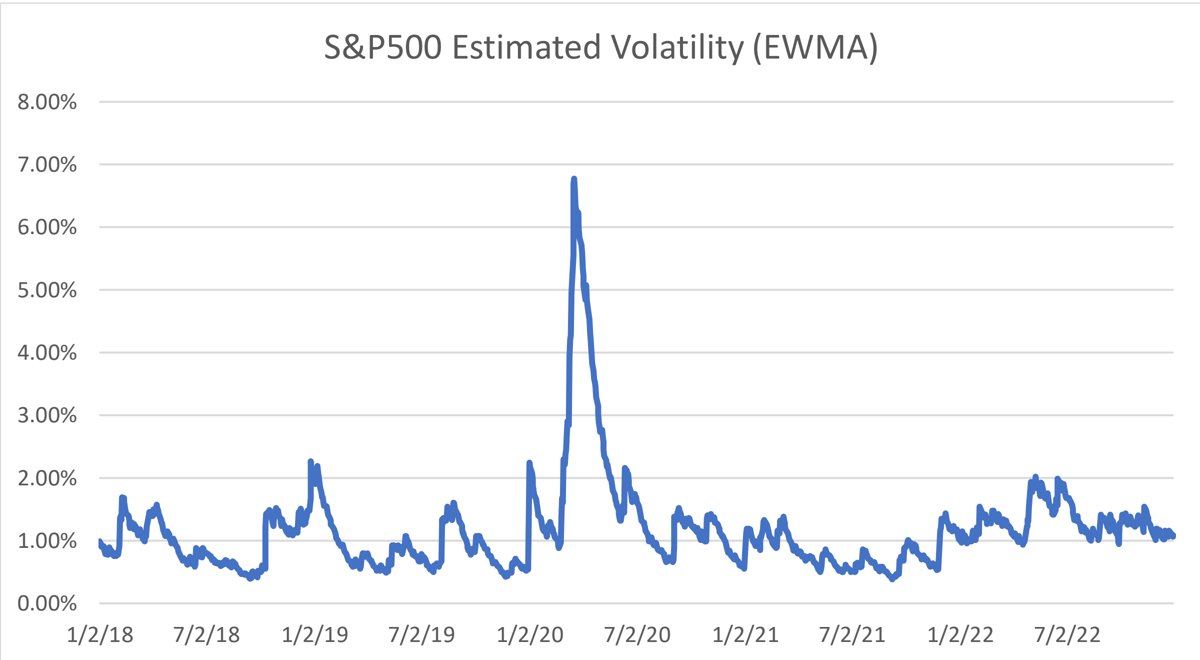

Figure 4: Graphic representation of EWMA estimated volatilitie for S&P500

Annex B: Backtesting Results

Table 11: Number of violations observed for HS approach

Historical Simulation No. Violations	Bitcoin	S&P500
Before COVID-19 Crisis	8	9
COVID-19 Crisis	4	8
After COVID-19 Crisis	11	9

Table 12: Number of violations observed for BRW approach

BRW No. Violations	Bitcoin	S&P500
Before COVID-19 Crisis	7	5
COVID-19 Crisis	4	4
After COVID-19 Crisis	9	5

Table 13: Number of violations observed for HW approach

HW No. Violations	Bitcoin	S&P500
Before COVID-19 Crisis	6	4
COVID-19 Crisis	1	2
After COVID-19 Crisis	6	6

Table 14: Backtesting results for HS model before COVID-19 Crisis

Before COVID-19 Crisis		
Historical Simulation	**Bitcoin**	**S&P 500**
No. violations	***8***	***9***
LR_{uc}	0,06575	2,56422
p-value	*0,79762*	*0,10931*
Result	***Do not reject***	***Do not reject***
LR_{ind}	3,26381	2,14036
p-value	*0,07082*	*0,14347*
Result	***Do not reject***	***Do not reject***
LR_{cc}	3,32956	4,70457
p-value	*0,06804*	*0,03008*
Result	***Do not reject***	***Do not reject***

Table 15: Backtesting results for HS model during COVID-19 Crisis

COVID-19 Crisis		
Historical Simulation	**Bitcoin**	**S&P 500**
No. violations	**4**	**8**
LRuc	0,03097	7,59989
p-value	*0,86031*	*0,00584*
Result	***Do not reject***	***Reject***
LRind	0,08840	5,64430
p-value	*0,76622*	*0,01751*
Result	***Do not reject***	***Do not reject***
LRcc	0,11937	13,24419
p-value	*0,72972*	*0,00027*
Result	***Do not reject***	***Reject***

Table 16: Backtesting results for HS model after COVID-19 Crisis

After COVID-19 Crisis		
Historical Simulation	**Bitcoin**	**S&P 500**
No. violations	***11***	***9***
LR_{uc}	1,63944	2,56422
p-value	*0,20040*	*0,10931*
Result	***Do not reject***	***Do not reject***
LR_{ind}	2,05890	0,32795
p-value	*0,15132*	*0,56687*
Result	***Do not reject***	***Do not reject***
LR_{cc}	3,69834	2,89217
p-value	*0,05447*	*0,08901*
Result	***Do not reject***	***Do not reject***

Table 17: Backtesting results for BRW model before COVID-19 Crisis

Before COVID-19 Crisis		
BRW	**Bitcoin**	**S&P 500**
No. violations	***7***	***5***
LR_{uc}	0,01263	0,00018
p-value	*0,91053*	*0,98926*
Result	***Do not reject***	***Do not reject***
LR_{ind}	0,13555	4,49540
p-value	*0,71275*	*0,03399*
Result	***Do not reject***	***Do not reject***
LR_{cc}	0,14817	4,49558
p-value	*0,70029*	*0,03398*
Result	***Do not reject***	***Do not reject***

Table 18: Backtesting results for BRW model during COVID-19 Crisis

COVID -19 Crisis		
BRW	**Bitcoin**	**S&P 500**
No. violations	***4***	***4***
LR_{uc}	0,03097	0,73324
p-value	*0,86031*	*0,39183*
Result	***Do not reject***	***Do not reject***
LR_{ind}	0,08840	0,12852
p-value	*0,76622*	*0,71997*
Result	***Do not reject***	***Do not reject***
LR_{cc}	0,11937	0,86176
p-value	*0,72972*	*0,35325*
Result	***Do not reject***	***Do not reject***

Table 19: Backtesting results for BRW model after COVID-19 Crisis

After COVID-19 Crisis		
BRW	**Bitcoin**	**S&P 500**
No. violations	***9***	***5***
LR_{uc}	0,37231	0,00018
p-value	*0,54175*	*0,98926*
Result	***Do not reject***	***Do not reject***
LR_{ind}	2,80430	0,10040
p-value	*0,09401*	*0,75135*
Result	***Do not reject***	***Do not reject***
LR_{cc}	3,17660	0,10058
p-value	*0,07470*	*0,75113*
Result	***Do not reject***	***Do not reject***

Table 20: Backtesting results for HW model before COVID-19 Crisis

Before COVID-19 Crisis		
HW	**Bitcoin**	**S&P 500**
No. violations	***6***	***4***
LRuc	0,24896	0,22912
p-value	*0,61781*	*0,63217*
Result	***Do not reject***	***Do not reject***
LRind	4,44031	0,06413
p-value	*0,03510*	*0,80009*
Result	***Do not reject***	***Do not reject***
LRcc	4,68927	0,29325
p-value	*0,03035*	*0,58814*
Result	***Do not reject***	***Do not reject***

Table 21: Backtesting results for HW model during COVID-19 Crisis

COVID-19 Crisis		
HW	**Bitcoin**	**S&P 500**
No. violations	***1***	***2***
LRuc	2,74455	0,22912
p-value	*0,09759*	*0,63217*
Result	***Do not reject***	***Do not reject***
LRind	0,00548	0,03187
p-value	*0,94099*	*0,85831*
Result	***Do not reject***	***Do not reject***
LRcc	2,75003	0,26100
p-value	*0,09725*	*0,60944*
Result	***Do not reject***	***Do not reject***

Table 22: Backtesting results for HW model after COVID-19 Crisis

After COVID-19 Crisis		
HW	**Bitcoin**	**S&P 500**
No. violations	***6***	***6***
LRuc	0,24896	0,17796
p-value	*0,61781*	*0,67313*
Result	***Do not reject***	***Do not reject***
LRind	0,09945	0,14487
p-value	*0,75249*	*0,70348*
Result	***Do not reject***	***Do not reject***
LRcc	0,34841	0,32284
p-value	*0,55502*	*0,56991*
Result	***Do not reject***	***Do not reject***