POST-ETF CRYPTO MARKET EFFICIENCY

by

Tymur Tleuberlin

A thesis submitted in partial fulfillment of
the requirements for the degree of

MA in Economic Analysis

Kyiv School of Economics

2026

Thesis Supervisor: Professor Olesia Verchenko

Approved by _____
Head of the KSE Defense Committee, Professor

_____

_____

_____

Date _____

Kyiv School of Economics

Abstract

POST-ETF CRYPTO MARKET EFFICIENCY

by Tymur Tleuberlin

Thesis Supervisor: Professor Olesia Verchenko

In this thesis I investigate whether the approval of spot Bitcoin ETFs and spot Ethereum ETFs by the US Securities and Exchange Commission changed cryptocurrency market efficiency. For the analysis I used a one-minute trading data from Binance of five major cryptocurrencies (BTC, ETH, XRP, SOL, ADA) from January 2021 and through end-2025.

Informational efficiency is tested with Random Forest classifiers trained on microstructure features, realized volatility of asset itself and BTC with ETH to capture spillover effects. Results are that AUC converged towards 0.5 in the period after BTC ETF approval, but after that reverted to pre BTC ETF approval levels with the similar feature importance indicating that ETF's effect on crypto markets informational efficiency is not sustainable. To complement ML driven approach I also tested weak-form efficiency under Effective and Adaptive market Hypotheses by conducting VR and AVR tests over static and rolling window. All tests fail to reject random walk null hypothesis indicating weak-form efficiency across all assets and regimes. To investigate effect of ETF on crypto

markets I utilized Bai Perron Multiple Structural breaks test for realized volatility, results confirm volatility regime change around BlackRock's spot ETF filling in June 2023 and then after October 2023 when ETF approval started to get huge attention in media, but no actual structural break in the time of event. That suggests that investors behavior followed the mental model of "buy the rumor sell the news". And pattern holds for BTC, ETH, ADA and SOL, but not XRP. And a final piece of testing is a economic significance test, using predicted sign of change in autocorrelation and available market data, to test whether short term predictability with microstructure features practically exploitable. Test reveals that detected predictable short-term inefficiencies are expensive to exploit making market practically efficient.

TABLE OF CONTENTS

LIST OF FIGURES iii
LIST OF TABLES iv
LIST OF ABBREVATIONS v
AI USE DECLARATION vi
Chapter 1. INTRODUCTION 1
Chapter 2. LITERATURE REVIEW 9
Chapter 3. METHODOLOGY 19
Chapter 4. DATA 39
Chapter 5. ESTIMATION RESULTS 44
Chapter 6. CONCLUSIONS AND POLICY RECOMMENDATIONS 59
WORKS CITED 62
APPENDIX A 66

LIST OF FIGURES

Number *Page*

Figure 1. BTC Daily Log-transformed Realized Volatility with Bai-Perron Structural Breaks 45

Figure 2. Lo-MacKinlay Random Walk Rejection Heatmap 46

Figure 3. Chow-Denning Robust Test Statistic results 47

Figure 4. Rolling window Variance Ratio test result at k=1 48

Figure 5. Rolling window AVR test results at k=3 49

Figure 6. Mean AUC for Sign of Change in Returns Autocorrelation 51

Figure 7. CPCV Annualized Sharpe Distribution with Config B at 5 bps 54

Figure 8. Probability of Backtest Overfitting Heatmap. Model Config A, Cost 10 pbs. 55

Figure 9. Probability of Backtest Overfitting Heatmap. Model Config B, Cost 10 pbs. 56

Figure A1. Feature importance chart for regularized Random Forest model 68

LIST OF TABLES

Number *Page*

Table 1. Random Forest Model Configuration 25

Table 2. BTC Descriptive statistics, 1 minute data 40

Table 3. ETH Descriptive statistics 41

Table 4. ADA Descriptive statistics 41

Table 5. SOL Descriptive statistics 42

Table 6. XRP Descriptive statistics 42

Table 7. Bai-Perron structural break results for Log (Realized Volatility) 45

Table 8. Mean Daily Returns (bps) and Statistical Test Results 53

Table 9. Cumulative Sharpe Ratios over Transaction Cost Levels 53

Table A1. Lo-MacKinlay Variance Ratio Statistics (Daily Realized Volatility) 66

Table A2. Chow-Denning Multiple Variance Ratio Test Results 67

LIST OF ABBREVATIONS

US. United States

ETF. Exchange-traded fund

SEC. Securities and Exchange Commission

AUC. Area Under the Curve

BTC. Bitcoin

ETH. Ethereum

SOL. Solana

XRP. Ripple

ILLIQ. Amihud Illiquidity Ratio

VPIN. Volume Synchronized Probability of Informed Trading

AI USE DECLARATION

During the preparation of this thesis, Generative AI (Claude) was used as a supplementary tool to:

1. Support grammar correction and improvement of sentence clarity in Introduction, Literature Review and Methodology chapters.
2. Support with documentation on LaTeX syntax to write formulas in Methodology chapter.
3. Support with documentation and ideas generation during code development for data engineering pipeline.
4. Code structuring and formatting to make it easier to maintain it and understand.

Any suggestions produced by the AI tool were critically evaluated and substantially modified by the author before being incorporated into the thesis.

The tool was not used to generate research findings, produce literature summaries used directly in the thesis, create citations, or conduct any form of data analysis. All research questions, theoretical framing, methodology, analysis, interpretation of results, and conclusions were developed independently by the author with advises from advisor.

The author takes full responsibility for the originality, validity, and academic integrity of the work presented in this thesis.

Chapter 1

INTRODUCTION

When the U.S. SEC approved spot Bitcoin ETFs in January 2024 it opened a regulated investment channel that may have changed crypto market quality and efficiency. Previously institutions that wanted to invest in crypto had to find work arounds, typically through closed end trusts or public companies like MicroStrategy carrying large Bitcoin treasuries, and of course those substitutes were traded at significant premiums or discounts to their underlying value. So spot ETFs offered something more straightforward - a direct and regulated access for institutional capital into crypto markets.

But it is not obvious if this shift actually improved market quality. My initial assumption was that it should because institutional entry and ETF-enabled arbitrage should have been reduced crypto's short-horizon predictability and push prices towards efficiency. But the evidence for pre-ETF market showed informational inefficiency and rejected weak form efficiency. Easley, O'Hara, Yang and Zhang (2024) documented that standard microstructure features predicted short-horizon trend of returns with significant out-of-sample AUC for five major cryptocurrencies they tested - a direct violation of the Efficient Market Hypothesis.

There are other findings on efficiency but all of them studied Bitcoin only, we have Urquhart's (2016) evidence that Bitcoin is not efficient but "moving towards efficiency" that was studied within Adaptive Market Hypothesis Framework. Then Nadarajah and Chu (2017) that worked over Urquhart's findings to show that actually Bitcoin is efficient if you apply odd integer transformation to returns. And lastly Yi et al. (2023) showed that Bitcoin is "near efficient", but they argued

using Variance Ratio Tests for efficiency hypothesis testing and used Quantum Harmonic Oscillator model instead.

In this thesis I try to answer those questions: first, did the ETF launch actually make the crypto spot market more efficient which would mean that institutional inflows reduced the return predictability that Easley et al. (2024) documented? Second, whether efficiency measures for all five major cryptos will differ when using different approaches that previously showed controversial result?

So in this thesis I test four hypotheses:

- **H1 (Efficiency & Predictability):** The introduction of ETFs has improved informational efficiency, evidenced by a decline in the out-of-sample predictive power of microstructure features for short-horizon dynamics for all five cryptocurrencies compared to the pre-ETF period.

- **H2 (Convergence to Random Walk):** Hypothesis is that the absolute deviation of the Variance Ratio from 1 significantly decreased after ETF approval. Samuelson (1965) showed that informationally efficient prices must follow a random walk; Fama (1970) gave turned that into the weak-form efficiency framework. We already have analysis of Urquhart (2016) that showed that while BTC generally inefficient but some tests showed that in latter part of the sample BTC returns moved towards efficiency. That draws to hypothesis that such events as BTC institutionalization make asset more mature and drive towards weak form efficiency. And then Nadarajah and Chu (2017) built their analysis over Urquhart's (2016) findings where they came to conclusion that if they apply odd integer power transformation to bitcoin's returns this doesn't lead to information loss and their wild bootstrapped automatic variance ratio test showed that BTC returns are weakly efficient.

- **H3 (Existence of a Structural Break):** Hypothesis is that there is a statistically significant structural break of realized volatility detected by the Bai-Perron procedure at the time of the ETF approval for all five cryptocurrencies. Since the ETF approval alters market participant composition, liquidity provision, and arbitrage activity, these changes are expected to take affect primarily in volatility dynamics rather than mean returns, we have evidence for that in ETF literature for US stock market (Ben-David et al. 2018). This is strongly supported by Charfeddine and Maouchi (2019) who investigated whether shocks to cryptocurrency returns and volatility are truly persistent. They applied the Bai-Perron procedure to both the mean returns and the volatility series of major cryptocurrencies. Their empirical results provided strong evidence against the presence of structural breaks in the raw returns series. However, when applied to the volatility series, the tests successfully identified multiple significant structural breaks. And it drives me to hypothesis that assessing volatility dynamics is robust method for capturing structural market shifts such as ETF approvals compared to assessing breaks in returns.

- **H4 (Time-varying efficiency):** Hypothesis is that rolling window variance ratio of all five major cryptocurrencies that are used in research (Bitcoin, Etherium, Ardano, Solana, Ripple) converging towards 1 after ETF approval of Bitcoin and Etherium. Urquhart (2016) in his work assess efficiency of bitcoin according to Efficient Market Hypothesis by using Variance Ratio test over whole dataset he had and also tested efficiency under Adaptive Market Hypothesis which assumes that efficiency varies with conditions. And findings were that Bitcoin is moving towards efficiency, when using time factor as condition the more recent data showed variance ratios converge to 1. Also we have later

works that assessed efficiency under Adaptive Market Hypothesis such as Chu et al. (2019) where they concluded that efficiency of Bitcoin and Etherium varies over time. And Khursheed (2020) with the same findings, specifically that Bitcoin showed long periods of efficiency and that efficiency varies over time.

- **H5 (Economic (non-)exploitability):** Hypothesis is that if Hypothesis 1 shows inefficiency and opportunity to predict short horizon dynamics post ETF approval - then using those predictions we could earn abnormal returns after considering transactions cost. We have Jensen (1978) who reframed market efficiency as no possibility of earning economic profits net of all costs and not statistical patterns in returns. Then Grossman and Stiglitz (1980) worked over Jensen's work and proved that in perfectly informationally efficient markets are impossible therefore some degree of statistical predictability must persist as compensation for informed trading. In terms of crypto context Hudson and Urquhart (2021) tested 15,000 trading rules on Bitcoin and three altcoins to see if they can earn profits, considering findings in that period that Bitcoin is not efficient. And they built a strategy that showed profits in a back tests but when applied to fresh data - profits vanished right after launch of Bitcoin Futures in December 2017 by CME Group. And my theory here is that ETF event is similar to launch of Bitcoin Futures and even if predictability didn't change, due to more institutional money, more trading volume, and more arbitrageurs hunting for opportunities - crypto currency will be practically efficient as inefficiencies are not exploitable after transactions costs.

To test market efficiency, I replicate and extend the machine learning framework of Easley, Yang, and Zhang (2024). Using one-minute trading data from Binance,

I construct a vector of microstructure features - including VPIN (Easley, López de Prado, and O'Hara 2012), Amihud (2002) illiquidity, Kyle's (1985) lambda, Roll's (1984) spread and impact measures, and realized returns and volatility at multiple horizons - to train Random Forest classifiers following the purged cross-validation methodology of López de Prado (2018). The target variable is the sign of next-period return autocorrelation, following Easley et al. (2024): an out-of-sample AUC significantly exceeding 0.5 indicates that publicly observable microstructure information retains predictive power that prices have failed to incorporate, and act as evidence against weak-form informational efficiency. The feature set for each asset also includes cross-market microstructure variables from the other dominant asset(s): BTC models include ETH-derived features, ETH models include BTC-derived features, and altcoin models (XRP, SOL, ADA) include features from both BTC and ETH. The idea is that these cross-market variables of realized volatility, returns, liquidity computed at matching roll back windows (50 and 100 bars) could capture informational linkages between assets. The logic is that if the ETF approval altered information transmission across the market, then price discovery became more and cross-market features lost predictive power.

Next to test economic significance I test whether it is possible to use predicted sign of change in returns autocorrelation and other available market data to earn abnormal returns net of transaction costs. Idea was to have prediction for a next day sign of change of returns autocorrelation and take average of daily bets. And test over three strategies: 1) conditional momentum strategy that goes long, short or neutral based on recent price momentum and predicted sign, 2) no short positions strategy that does same as first but never goes short, only long or neutral, 3) and only prediction model - that uses only model's prediction to decide if go long, short or neutral. To assess performance of strategies I used Sharpe Ratio, Probabilistic Sharpe Ration and to not fall into same back testing

problem as Urquhart (2021) had in their research I do Combinatorial Purged Cross-Validation which employs simulations to see if my back testing results are biased.

Next to test Efficient Market Hypothesis with Variance Ratio test I conduct Lo and MacKinlay's (1988) heteroskedasticity-robust Variance Ratio test, with the Chow-Denning (1993) joint multiple-horizon specification as a robustness check. With that I test whether post-ETF returns more closely approximate a random walk relative to the pre-ETF period.

And additionally, I utilize wild-bootstrapped automatic variance ratio test that Nadarajah and Chu (2017) used but for rolling window of returns following Adaptive Market Hypothesis approach that Urquhart (2016) used to see if ETF approval affected crypto market efficiency and compare findings together with others mentioned above.

Finally, to validate the regime specification imposed by the ETF approval dates, I apply the Bai-Perron (1998, 2003) multiple structural break procedure. Following the theoretical expectation outlined in H3 and the empirical evidence of Charfeddine and Maouchi (2019), this procedure is applied to realized volatility series rather than daily log returns, as the evidence shows that there is a lot of noise and not enough power to detect reliable shift in mean returns at appropriate significance levels.

As to assets selection, I will follow Easley et al. (2024), who selected Bitcoin (BTC), Ethereum (ETH), Ripple (XRP), Solana (SOL), and Cardano (ADA) as the top five cryptocurrencies by market capitalization in January 2021. Retaining the same assets will allow me maintaining comparability with their study, while leveraging the ETF natural experiment to introduce some kind of a treatment-control dimension. BTC and ETH serve as treatment assets as only they received

SEC spot ETF approval, indicating their formal integration into regulated institutional investment channels. Their microstructure variables will be used as cross-market features in altcoin models, enabling direct measurement of informational transmission from ETF-affected assets to the broader crypto market. Other assumption here is that ADA and SOL are major-capitalization assets without direct ETF exposure. And I use them to see if ETF approvals affected them and therefore majority of crypto market. I test for cross-market spillover effects - whether the microstructure changes created by BTC and ETH ETFs spilled over to correlated assets. If the cross-market BTC and ETH features, that were predictive for altcoin dynamics in the pre-ETF regime, lose their predictive power post-ETF - then this would be an evidence that institutional integration improved price discovery not only for the directly affected assets but across broader crypto market.

It is also important to highlight that XRP here plays a distinct role as a natural placebo: its regulatory isolation from the ETF ecosystem and unique legal history that is different from the clear regulatory path of BTC ETF. In 2020 SEC filled a lawsuit against XRP which was resolved in 2023 and conclusion was that XRP is not a security that created pretty long regulatory isolation to this coin. And in other side - Brazil started process of XRP ETF approval, which was announced at the end of 2024 and the resulting approval happened in Feb 19, 2025. Including this asset allows to test whether market reacted simply to the world "ETF" or specifically to the US SEC's institutional approval. The logic is that if XRP shows a regime change before its Brazil ETF due to BTC spillover effects or only after BTC and/or ETH ETFs approval - then it is a cross market effect. If it reacts only to its own Brazil approval - then market is responding to local regulatory events and not a global trend.

In my opinion, findings could be useful to several audiences. For researchers and policymakers, this thesis extends Easley et al. (2024) by using ETF approvals as natural pre-post experiment. It gives a clear, outside cause on how policy approvals affect crypto market microstructure, and whether it helps to integrate crypto into existing financial system. For traders and market-makers, evaluating whether detected predictability survives transaction costs gives real world pragmatic insights: whether the inefficiency was ever actually exploitable.

Chapter 2

LITERATURE REVIEW

Institutional Context and Market Evolution

Bitcoin is not a conventional financial asset as it is just a piece of code that runs in a decentralized system: it is a cryptographically secured store of value that has no sovereign backing. Over the time it has grown into a benchmark for the broader cryptocurrency ecosystem due to its considerable market capitalization and trading volume (Lee Kuo Chuen, 2024). The market around it has shifted as well - from a small, retail-dominated environment with extreme volatility and low liquidity to a fast-growing sector where institutions are getting increasingly involved.

The biggest marker of that shift was the U.S. SEC approval of spot Bitcoin ETFs on January 10, 2024, and then spot Ethereum ETFs on July 23, 2024. These approvals show bigger confidence in those cryptos as security for public and created direct channels for institutional capital. That pulled the two largest cryptocurrencies into closer contact with traditional financial infrastructure.

Prior to the ETF approval, the literature characterized the crypto market as highly segmented. Makarov and Schoar (2020) documented large persistent price gaps between exchanges and attributed them to capital controls and practical limits to arbitrage. That led them to conclusion that liquidity was fractured, preventing the efficient flow of capital. Regulated investment channels act as bridges between traditional finance and digital assets (Lee Kuo Chuen, 2024) and spot ETFs are the clearest example of this in cryptocurrency markets. There is an early post-ETF evidence that suggests some stabilization, for example, Babalos, Bouri, and Gupta (2024) find altcoin volatility decreased

after BTC ETF approval. Borri et al. (2025) have similar investigation of how the introduction of the U.S. spot Bitcoin ETFs affects spot returns and volatility. This thesis extends that investigation from aggregate volatility to microstructure-level informational efficiency across ETF regimes, distinctiveness of BTC as asset and efficient market hypothesis tests.

Market Efficiency: Theoretical Foundations

The Efficient Market Hypothesis (EMH) was first formalized by Fama (1965, 1970) and it states that asset prices should fully reflect all available information making systematic return prediction impossible. The theoretical foundation for this is based on Samuelson's (1965) proof that properly anticipated prices must fluctuate randomly. Overall, Fama defined three forms of market efficiency: weak-form, which states that prices reflect all past trading data; semi-strong form, which states that prices additionally reflect all publicly available information; and strong form, which states that prices reflect all information including private signals. In this thesis I test weak-form market efficiency to test that past price and volume data should contain no exploitable information about future returns.

In the context of market microstructure Foucault, Pagano, and Röell (2024) argue that "frictions such as transaction costs, asymmetric information and inventory risk prevent immediate price adjustment". These frictions are especially relevant for cryptocurrency markets and include the following features mentioned above: fragmented liquidity across exchanges, no designated market makers, limited regulatory oversight and informational asymmetries.

The question here is whether ETF-led institutional integration reduced these frictions enough to affect Bitcoin's weak-form efficiency.

Microstructure Measures

In thesis I used standard metrics to quantify market frictions, but also those metrics utilized by Easley et al. (2024) so I kept exactly the same set to maintain comparability of results to their results. The first important feature is the implicit spread measure, introduced by Roll (1984), who demonstrated that effective bid-ask spread can be inferred from the serial covariance of price changes. Also Easley et al. (2024) proposed a Roll Impact measure which is basically a Roll (1984) measure but divided by the dollar volume over a specified time period. It is used alongside with Roll (1984) and idea is that by scaling Roll measure by training activity it will let us to capture price impact of trading volume and increase accuracy for inefficiency predictions and return distributions.

Kyle (1985) formalized the concept of price impact as the sensitivity of prices to order flow: Kyle's lambda quantifies by how much price moves when an order to buy or sell is placed. It also serves as a direct proxy for market depth and the presence of informed trading: it shows market makers' response to their perceived probability of informed trading. A higher lambda shows bigger perceived risk that traders, who place orders, have some private information on real price of an asset. It works like this: when informed traders trade, they increase uncertainty about the asset's true value. And meanwhile uninformed traders create noise which is random price movement based on liquidity needs and not real information. So basically Kyle's lambda is a regression slope and it shows how sensitive prices are to the ratio of informed trading to uninformed trading.

Amihud (2002) developed the Amihud Illiquidity Ratio (ILLIQ), which captures the absolute price response per unit of trading volume. Rather than measuring effect of placing market orders on prices, it gives an actual number of how

much price changes per one dollar traded. It calculated as absolute daily return divided by daily dollar trading volume.

Easley et al. (2012) suggested a Volume-Synchronized Probability of Informed Trading (VPIN) measure. It estimates the probability of informed trading by measuring the absolute imbalance between buy- and sell-initiated volume over fixed volume buckets. According to the authors, higher VPIN indicates "toxic" order flow, when informed traders are disproportionately active and force market makers to act upon that. VPIN is particularly relevant to the ETF hypothesis tested here: if institutional integration through ETFs improves information efficiency and attracts informed arbitrageurs, then predictive power of VPIN for future dynamics of returns should decrease as informed trading signals are more rapidly incorporated into prices.

Machine Learning and Microstructure Predictability

López de Prado (2018) states that traditional econometric models often fail to capture the complex, nonlinear interactions between microstructure features, for instance, how order flow toxicity interacts with depth imbalance under changing volatility regimes. Easley et al. (2021) addressed this by applying machine learning methods to microstructure research in equity futures markets. They showed that some microstructure features have high in-sample explanatory power but have low out-of-sample predictive power, while others with medium explanatory power prove being highly predictive for out-of-sample prediction. The authors chose Random Forest classifier, because its design excels at identifying complex trends and nonlinear relationships, but still this classifier lets researchers analyze which predictors were the most important,

compared to more complicated algorithms such as neural networks that lack explainability of predicted results.

Easley, Yang, and Zhang (2024) extended this framework to cryptocurrency markets, constructing microstructure feature vectors from high-frequency data for BTC, ETH, XRP, SOL, and ADA over the 2021-2023 period. As a result, they got a statistically significant out-of-sample predictability of short-horizon return dynamics, and concluded that crypto markets exhibit information inefficiencies that algorithmic strategies could exploit.

In this thesis I directly extend their work by introducing the ETF natural experiment: if the post-ETF regime eliminates or reduces the predictability they had then this shows evidence that institutional integration improved informational efficiency.

It is important to note that Random Forest was not designed specifically for time series data and by naively training it with default options will introduce a lookahead bias in the model, because the standard k-fold cross-validation assumes independent and identically distributed observations. López de Prado (2018) proposed purged cross-validation with embargo periods to control this: purging removes from the training set any observations whose labels overlap temporally with the test set and then the embargo excludes additional observations following each purged block to account for serial correlation in features. I employ these procedures as well to ensure that any observed changes in AUC across ETF regimes reflect actual shifts in market efficiency rather than artifacts of information leakage between training and test sets.

Testing Weak-Form Efficiency: Variance Ratio Tests

The Efficient Market Hypothesis was first proposed by Samuelson (1965) and then formalized by Fama (1970). The idea is that in a weak-form efficient

market, prices must fully reflect all historical information making them unpredictable and therefore follow a random walk.

The existing literature reports mixed findings Bitcoin's and crypto efficiency. For example, Urquhart (2016) showed early evidence that BTC returns were inefficient for his full sample (daily closing prices from August 1st 2010 to 31st July 2016). But when analyzed subsample with later period (August 1st 2013 to 31st July 2016) it showed a movement toward efficiency. So in his work he established a comparison of Effective Market Hypothesis and Adaptive Market Hypothesis, where under EMH there were no efficiency, whilst under AMH there were signs of efficiency over time. And Nadarajah and Chu (2017) showed that by applying odd-integer transformation of returns and wild bootstrap Bitcoin is weak form efficient, same conclusion had Yi et al (2023) that used Quantum Harmonic Oscillator to assess efficiency and their conclusion was that Bitcoin is near efficient.

And there were more researches of Bitcoin efficiency that adopted Adaptive Market Hypothesis such as Chu et al. (2019) where they concluded that efficiency of Bitcoin and Etherium varies over time. And Khursheed (2020) with the same findings, specifically that Bitcoin showed long periods of efficiency and that efficiency varies over time.

And a last piece of literature examines efficiency through economical significance. Jensen (1978) iterated over Fama (1970) and redefined market efficiency as not an absence of statistical pattern but inability to make economic profit net of all costs. But Grossman and Stiglitz (1980) extend it and proved that informationally efficient market is impossible so some predictability must persist as a cost for informed trading. And as I see implication of their findings is that practically statistical signal of inefficiency alone doesn't prove market is actually inefficient.

Structural Break Identification

There is a theoretical expectation that ETF introduction affects market participant structure, liquidity structure, and arbitrage activity and changes expected to be indicated primarily in the second moment of returns rather than the first (Ben-David, Franzoni, and Moussawi, 2018) because ETFs reshape trading condition but don't change fundamentals. So their impact might be seen in volatility, co-movement, and liquidity and not in mean returns. This leads to idea to apply structural break tests to realized volatility series rather than returns. This choice is further supported by Charfeddine and Maouchi (2019), who analyzed whether shocks to cryptocurrency returns and volatility are actually persistent or are consequences of regime shifts. Their conclusion was that structural shifts in cryptocurrency markets are detectable in volatility but not in raw returns and attributed it to extreme skewness and noise in return series of crypto. However, since Nadarajah and Chu (2017) showed that applying odd integer transformation to BTC returns normalizes their distributional properties without informational loss, it is worth verifying whether the transformed returns themselves exhibit structural breaks, and not just the volatility of raw returns.

So to investigate this, I apply Bai Perron multiple structural break procedure to realized volatility, realized returns and to the odd-integer-transformed return series. If breaks are detected in the transformed returns and are aligned with those found in volatility - it would suggest that the transformation recovers information lost in raw returns. If this is not the case, then this would confirm Charfeddine and Maouchi's conclusion that volatility dynamics is uniquely informative about structural change in crypto markets and return-based analysis cannot substitute direct volatility examination.

Economic Significance test

Jensen (1978) iterated over Fama (1965) work and suggested other definition of market efficiency - rather than absence of statistical pattern he offered that market efficiency will result in absence of economic profit net of all costs. Then Grossman and Stiglitz (1980) worked over it that resulted in "Grossman-Stiglitz Paradox", so in their work they argued that perfectly informationally efficient markets are impossible because if prices perfectly reflect all available information then there would be no economical profits and therefore no motivation to trade and markets would eventually collapse because of it. Therefore in efficient market could be statistical predictability but it calibrated to cost so that informed trader earn just enough to cover information acquirement costs. So in Grossman Stiglitz equilibrium we expect to find statistical predictability, but if predictability doesn't survive realistic cost - then it would mean market is efficient. And to support my point of view on their theory we have Hudson and Urquhart (2021) with empirical test - they used 15,000 trading rules on Bitcoin and three altcoins and they found significant predictability and profitability for these trading rules compared to buy-and-hold strategy in earlier Bitcoin periods before 2017. But when they tested those strategies for recent data - they discovered that profits vanished right after CME Group offered Bitcoin Futures in Chicago Mercantile Exchange in December 2017. And it complements Grossman and Stiglitz findings because it shows that before futures it was cheaper to get information and earn profits. But futures launch attracted more money and more market participants that resulted in cheaper arbitrage and equilibrium predictability shrank towards break even with cost. So statistical signal may still exist but economic profit gone. And in my thesis I frame ETF approval as a similar event for spot crypto, but considerably bigger scale, because ETF let to trade in

NASDAQ and NYSE which have much higher trading volume compared to Chicago Mercantile Exchange. And hypothesis is that even if short horizon predictability persists - there would be little to no opportunity to make profit using that predictability, and it can be detected only with economic test as statistical tests blind to it by design.

And to correctly conduct economic test I used Lopez de Prado (2018) methodologies: from chapter 10 I used his warning that fixed-horizon infuses cost to signal ratio and suggested averaging active bets for continuous position. Then in ch. 12 he proposed combinatorial purged cross-validation to simulate 252 paths for strategy, combined with his Probability of Backtest Overfitting - should give unbiased view on strategy performance and close to real world results. Overall his book Advances in Financial Machine Learning (2018) is a foundation for machine learning algorithms design and its testing in my thesis.

Contribution to Literature

This thesis contributes to the existing literature by combining several efficiency measures and efficiency hypotheses that typically was examined in isolation. Those are predictability test using market microstructure theory and its metrics as predictors that previously showed inefficiency, variance ratio tests for testing EMH's weak form efficiency with mixed historical evidence, variance ratio tests over rolling window for AMH weak form efficiency that showed some periods of efficiency of BTC, and economic significance testing. Bringing those together around big institutional event of US SEC's spot ETF approval allows me to evaluate together effects on market efficiency, sensitivity of different approaches, and turn findings into one consistent framework rather than disconnected studies with different samples and assumptions.

In my opinion findings can be useful to three audiences:

1. For policy makers - they provide evidence whether regulated investment channels actually improve market efficiency and price discovery, and reduces opportunities for informed trading.

2. For traders and quantitative researchers - comparison of statistical and economic predictability across different regimes, with evidence whether microstructure features can predict price dynamics and whether this predictability can be practically exploitable.

3. For academics applying EMH or AMH frameworks - the joint study of statistical and economical tests, comparison of results for each framework for altcoins (not only BTC as majority of studies have) and all around clearly identified institutional event. That can help to contribute to development Efficient Market and Adaptive Market Theories.

Chapter 3

METHODOLOGY

In thesis I use publicly available high-frequency trade data from Binance for five cryptocurrencies: Bitcoin (BTC), Ethereum (ETH), Ripple (XRP), Solana (SOL), and Cardano (ADA). The raw data consists of one-minute resolution records containing open, high, low, close prices, trading volume, and number of trades. The sample period spans from January 1, 2023 to the December 31, 2025, denominated in USDT trading pairs.

The sample is partitioned into three regimes defined by exogenous regulatory events:

- Pre-ETF regime: January 1, 2023 - January 9, 2024 (the last trading day before BTC spot ETF approval).
- BTC-ETF regime: January 10, 2024 - July 22, 2024 (from BTC ETF approval to the last day before ETH spot ETF approval).
- Full-ETF regime: July 23, 2024 - end of sample (from ETH ETF approval onward).

Log returns are computed as:

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

where P_t denotes the close price at time t. All microstructure features are computed from this one-minute data and then aggregated to the target prediction horizon as described below.

Microstructure Feature Construction

Following Easley, Yang, and Zhang (2024), the feature set consists of six categories of microstructure measures, each computed over rolling windows of $w \in \{50,100\}$ one-minute bars.

Realized Volatility

Realized volatility is estimated as the square root of the sum of squared intraday log returns over a rolling window of w observations:

$$\mathrm{RV_t(w)} = \sqrt{\sum_{\mathrm{i=t-w+1}}^{\mathrm{t}} \mathrm{r_i^2}} \tag{2}$$

where r_i is the one-minute log return at time i. This provides a model-free, non-parametric estimate of integrated variance that does not require distributional assumptions (Andersen, Bollerslev, Diebold, and Labys 2003).

Roll Spread

Roll (1984) demonstrates that under the assumption of an efficient market with a fixed bid-ask spread, the effective spread can be inferred from the serial covariance of price changes. The Roll spread estimator over window w is:

$$\mathrm{Roll}_t(w) = \begin{cases} 2\sqrt{-\mathrm{Cov}(\Delta P_i, \Delta P_{i-1})} & \text{if } \mathrm{Cov}(\Delta P_i, \Delta P_{i-1}) < 0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where $\Delta P_i = P_i - P_{i-1}$ and the covariance is computed over the rolling window $[t - w + 1, t]$. A positive serial covariance, which would yield an undefined

square root, is set to zero following standard practice. Higher values indicate wider effective spreads and greater trading costs.

Roll Impact

To capture the price impact per unit of spread, I compute the Roll impact as:

$$\text{RollImpact}_t(w) = \frac{\text{Roll}_t(w)}{\sum_{i=t-w+1}^{t} V_i} \tag{4}$$

where V_i is the dollar volume at time i and the denominator is the cumulative dollar volume over the window, following Easley et al. (2024). This normalizes the spread estimate by trading activity and providing a measure of how much the effective spread costs relative to the liquidity available in the market.

Kyle's Lambda

Kyle's lambda (1985) is estimated via OLS regression over a rolling window:

$$\Delta P_i = \alpha + \lambda \cdot S_i + \varepsilon_i, \quad i \in [t - w + 1, t] \tag{5}$$

where S_i is the signed volume at time i (positive for buyer-initiated, negative for seller-initiated trades, classified using the tick rule or the BVC algorithm as applicable). The coefficient λ measures the permanent price impact per unit of signed order flow. Higher λ indicates lower market depth and greater information asymmetry, as larger price movements are required to absorb informed order flow.

Amihud Illiquidity

Amihud (2002) illiquidity ratio is the average absolute return per unit of dollar volume:

$$\text{ILLIQ}_t(w) = \frac{1}{w} \sum_{i=t-w+1}^{t} \frac{|r_i|}{V_i} \quad (6)$$

where r_i is the log return and V_i is the trading volume (in USDT) at time i. This ratio captures the price response to trading activity: higher values indicate that a given volume of trading moves prices more, reflecting lower liquidity.

Volume-Synchronized Probability of Informed Trading (VPIN)

VPIN, introduced by Easley, López de Prado, and O'Hara (2012), measures the imbalance between buy and sell volume.

I implement a time-bar approximation to VPIN following the Bulk Volume Classification (BVC) approach. This approximation operates directly on the one-minute bar data rather than partitioning volume into fixed-size buckets. This choice motivated by computational compromise considering the 2.6M+ rows per asset. Within each one-minute bar i, buy volume V_i^B and sell volume V_i^S are classified based on the standardized price change:

$$V_i^B = V_i \cdot \Phi\left(\frac{\Delta P_i}{\sigma_{\Delta P}}\right), \quad V_i^S = V_i - V_i^B \quad (7)$$

where $\Phi(\cdot)$ is the standard normal CDF and $\sigma_{\Delta P}$ is the standard deviation of price changes computed over the rolling window. But the normal CDF can be approximated as the logistic sigmoid scaled by 1.7, which is accurate to within ≈ 0.01 of the true CDF, so I used this approximation to simplify computation.

Then the informed volume fraction within each bar is then $|V_i^B - V_i^S|/V_i$. VPIN is computed as the rolling sum of informed volume divided by total volume over window w:

$$\text{VPIN}_t(w) = \frac{\sum_{i=t-w+1}^{t} |V_i^B - V_i^S|}{\sum_{i=t-w+1}^{t} V_i} \tag{8}$$

Increase in VPIN would happen in periods when there is disproportion of informed traders activity relative to uninformed order flow. In the context of my thesis if VPIN retains predictive power for future returns as it was seen in Easley et al. (2024) then it would imply that the market hasn't incorporated the information signaled by probability of informed training yet.

Cross-Market Features

For each asset, the feature set is augmented with cross-market microstructure variables from the other dominant asset(s). Specifically:

- BTC models include ETH-derived features (realized volatility, Amihud, Roll, Kyle's lambda, Roll impact, and VPIN) at both window lengths.
- ETH models include the corresponding BTC-derived features.
- Altcoin models (XRP, SOL, ADA) include cross-market features from both BTC and ETH.

This design captures informational spillovers between assets. If BTC and ETH microstructure variables predict altcoin return dynamics pre-ETF but lose predictive power post-ETF, this constitutes evidence that institutional integration improved cross-market price discovery.

Target Variable

Following Easley et al. (2024), the target variable is the sign of next-period return autocorrelation. Define the forward-looking return autocorrelation over a horizon of $h = 1500$ one-minute bars (approximately one trading day) as:

$$\rho_t(h) = \text{Corr}(r_i, r_{i-1}), \quad i \in [t+1, t+h] \tag{9}$$

The binary target variable is then:

$$y_t = \begin{cases} 1 & \text{if } \rho_t(h) > 0 \\ 0 & \text{if } \rho_t(h) \leq 0 \end{cases} \tag{10}$$

This target captures the sign of short-horizon serial dependence in returns. If microstructure features can predict whether returns will exhibit positive or negative autocorrelation in the next period, the market has failed to incorporate publicly available information about its own dependency structure. An out-of-sample AUC significantly exceeding 0.5 therefore constitutes evidence against weak-form informational efficiency.

Random Forest Classification

Model Specification

Following Easley et al. (2021, 2024) and the methodological recommendations of López de Prado (2018), I employ Random Forest (RF) classifiers. The RF is an ensemble of B decision trees, each trained on a bootstrapped sample of the data with a random subset of m features considered at each split. For a feature vector x_t, the predicted probability of class k is:

$$\hat{P}(y_t = k \mid x_t) = \frac{1}{B} \sum_{b=1}^{B} \hat{P}_b(y_t = k \mid x_t) \tag{11}$$

where $\hat{P}_b$ is the class probability estimate from tree b. The RF is chosen for three reasons: it captures nonlinear interactions between features without imposing parametric functional form assumptions; it is robust to overfitting relative to individual decision trees due to bagging and feature subsampling; and it maintains comparability with the anchor studies (Easley et al. 2021, 2024).

Model Configurations

Two configurations are estimated for each asset-regime combination to assess sensitivity to regularization. Both configurations share $B = 100$ trees, max_features $= \lfloor \sqrt{p} \rfloor$ where p is the feature count, balanced class weights, and a fixed random seed.

Table 1. Random Forest Model Configuration

Parameter	Config A (Non-regularized)	Config B (Regularized)
N Estimators (B)	100	100
Max Features	36	36
Max Depth	None (unlimited)	5
Min Weight Fraction Leaf	0.0	0.05
Class Weight	balanced	balanced

Config A (Non-regularized) mirrors the Easley et al. (2024) specification with no depth or leaf constraints. It allows the model maximum flexibility to fit the training data.

Config B (Regularized) limits tree depth to 5 and requires each leaf to represent at least 5% of the sample weight. Idea was to have model with this configuration as

an additional measure against overfitting in cost of providing a more conservative estimate of out-of-sample predictive power.

Both configs are included for comparability with Easley et al. 2024.

Purged Cross-Validation with Embargo

Following López de Prado (2018), I implement purged K-fold cross-validation with embargo to prevent information leakage. The procedure uses $K = 10$ folds, a purge size equal to the target horizon $h = 1{,}500$ bars, and an embargo of $e = 10{,}080$ bars (approximately one week of one-minute observations).

Purging logic is that for each fold any training observation whose label window $[t + 1, t + h]$ overlaps temporally with any test observation's feature window is removed from the training set. Formally, training observation j is purged if:

$$\left[t_j + 1, t_j + h\right] \cap [t_i - w, t_i] \neq \emptyset \qquad (12)$$

for any test observation i, where w is the maximum feature lookback window and h is the forward-looking target horizon.

Embargoing logic is that for nn additional buffer of e observations following each purged block is excluded from the training set to account for residual serial correlation in features that may persist beyond the explicit overlap window:

$$\text{Training set} = \text{Training set} j : t_j \in \left[t_{\text{test,end}} + 1, t_{\text{test,end}} + e\right] \qquad (13)$$

This procedure ensures that the RF classifier cannot benefit from look-ahead bias, and that any observed out-of-sample predictive accuracy reflects real informational content of the features rather than data leakage.

Evaluation Metrics

Model performance is assessed using two metrics:

Area Under the ROC Curve (AUC): The AUC measures the probability that the classifier assigns a higher predicted probability to a randomly chosen positive example than to a randomly chosen negative example:

$$\text{AUC} = P\left(\hat{P}(y = 1 \mid x^{+}) > \hat{P}(y = 1 \mid x^{-})\right) \tag{14}$$

where x^{+} and x^{-} are drawn from the positive and negative classes respectively.

An AUC of 0.5 corresponds to random guessing, while AUC > 0.5 indicates better-than-chance predictive power. In the my work, AUC > 0.5 constitutes evidence that publicly observable microstructure information contains signal about future return autocorrelation that the market has failed to incorporate.

Economic Significance Test

Objective of economic significance test is to answer whether predictive signal from Random Forest classified can be used to earn profits from trading after realistic transaction cost. Theoretically classifier may achieve statistically significant AUC scores but fail to generate economic value and this test is used to distinguish detectable inefficiencies from profitable ones.

Test consumes out-of-sample (OOS) predictions generated by model that was described in section above to predict sign of change in returns autocorrelation with horizon of 1500 bars (which ~1 trading day). Then we use this prediction as a signal for decision making, but using naïve approach to simply trade on each bar's prediction directly and flipping position whenever signal changes blows up

cost to signal ratio and makes this strategy deeply unprofitable regardless of predictive accuracy. So I used Averaging Active Bets approach (Lopez de Prado 2018, ch. 10).

Lopez de Prado explicitly warns against fixed time horizon rebalancing which is in a nutshell naïve approach I describe. And he suggest instead for each minute's prediction create a "bet" with K=1500 lifespan. Then at any time t, the portfolio position equals the rolling mean of all active bets target signals over [t-K+1, t]:

$$\text{position}(t) = \frac{1}{K} \sum_{i=t-K+1}^{t} \text{target_signal}\,(i) \qquad (15)$$

And this produces a continuous position in [-1, +1] rather than separate flips and minimizes turnover.

And since data is at 1-minute frequence there are ~1440 prediction per day (24-hour crypto trading). So the averaging mechanism resolves multiple predictions with this logic:

- Each minute's prediction contributes one target signal to the rolling window
- The position at any moment is the average of the most recent K=1500 predictions
- New predictions shift the rolling window by one bar and changing position by max at |new_signal - dropped_signal|/K
- So the typical per bar position change: ~1/K=0.0007

It is important to mention that there is no daily rebalance point. The position updates continuously with each bar's prediction blending into aggregated, and that also eliminates the need for arbitrary timing decisions (e.g. rebalance at 00:00 UTC)

Trading strategies based on signal types

Test implements three signal construction methods for comparison:

Condition Momentum Strategy (primary strategy):

Combines autocorrelation prediction with recent price momentum:

$$\text{recent_momentum}[t] = \text{sgn}\left(\sum_{i=t-K+1}^{t} r_i\right) \tag{16}$$

$$\text{autocorr_view}[t] = \begin{matrix} +1 \; if \; P(t) > 0.55 \\ -1 \; if \; P(t) < 0.45 \\ 0 \; otherwise (neutral \; zone = 0.05) \end{matrix} \tag{17}$$

$$\text{target_signal}[t] = \text{recent_momentum}[t] \times \text{autocorr_view}[t] \tag{18}$$

$$\text{position}[t] = \frac{1}{K}\sum_{j=0}^{K-1} \text{target_signal}[t-j] \tag{19}$$

Recent_momentum uses trailing K-bar cumulative return and first K bars are set to 0.

Autocorr_view translates non directional prediction into directional signal via a neutral zone.

And the product yealds a signal in {-1, 0 +1} before averaging.

Binary Averaged Strategy

Same as Conditional Momentum strategy but long-only signal. It goes long when P(y=1) > 0.5, flat otherwise and then average over K bars

$$\text{target_signal}[t] = 1[P(t) > 0.5] \tag{20}$$

$$\text{position}[t] = \frac{1}{K}\sum_{j=0}^{K-1} \text{target_signal}[t-j] \tag{21}$$

Position is always in [0,1]. This strategy serves as a baseline capturing market beta during trending periods.

Neutral Zone Averaged

Directional signal with neutral zone, but without momentum multiplier, basically conditional momentum strategy that only uses prediction.

$$\text{target_signal}[t] = \begin{matrix} +1 \; if \; P(t) > 0.55 \\ -1 \; if \; P(t) < 0.45 \\ 0 \; otherwise (neutralzone = 0.05) \end{matrix} \tag{22}$$

$$\text{position}[t] = \frac{1}{K}\sum_{j=0}^{K-1} \text{target_signal}[t-j] \tag{23}$$

Idea is to use this one to isolate effect of the autocorrelation prediction alone.

Transaction Costs

Strategy returns are computed as:

$$\text{gross_return}[t] = \text{position}[t] \times \text{realized_return}[t] \tag{24}$$

$$\text{cost}[t] = \frac{\text{cost_bps}}{10000} \times |\text{position}[t] - \text{position}[t-1]| \tag{25}$$

$$\text{net_return}[t] = \text{gross_return}[t] - \text{cost}[t] \tag{26}$$

Costs are proportional to the absolute change in position not to discrete trades. Four cost scenarios are tested: 5, 10, 15, 25 bps one-way, where 10 bps is the baseline.

Performance Metrics:

Annualized Sharpe Ratio

1-minute returns are aggregated to daily by truncating to the nearest multiple of 1440 bars and summing within each day. The daily Sharpe ratio is then annualized:

$$\text{SR}_{\text{annual}} = \frac{\bar{r}_{\text{daily}}}{\sigma_{\text{daily}}} \times \sqrt{365} \tag{27}$$

Daily aggregation avoids the autocorrelation-induced inflation identified by Lo (2002) that afflicts high-frequency Sharpe ratios computed directly on 1-minute data.

Probabilistic Sharpe Ratio (PSR)

Bailey and Lopez de Prado (2012) show that the standard Sharpe ratio is biased for non-normal return distributions. PSR adjusts for skewness and kurtosis:

$$\text{PSR} = \Phi\left(\frac{\left(\widehat{SR} - SR^*\right)\sqrt{T-1}}{\sqrt{1 - \hat{\gamma}_3\widehat{SR} + \frac{\hat{\gamma}_4 - 1}{4}\widehat{SR}^2}}\right) \tag{28}$$

where $\hat{\gamma}_3$ is skewness, $\hat{\gamma}_4$ is excess kurtosis, T is sample size, and $SR^* = 0$ is the benchmark Sharpe. PSR > 0.95 indicates that the observed Sharpe is statistically significantly positive after accounting for higher moments.

Combinatorial Purged Cross-Validation (CPCV)

A single backtest equity curve is fragile because its Sharpe ratio depends on the specific temporal ordering of Out of Sample (OOS) folds. CPCV (Lopez de Prado, 2018, Ch. 12) generates C(N, k) = C(10, 5) = 252 distinct backtest paths by recombining OOS fold predictions, providing an empirical distribution of performance.

Procedure:

1. Partition the net strategy return series into N=10 equal temporal segments (aligned with the 10 CV folds).

2. For each of the 252 combinations of k=5 segments, concatenate the selected segments into one backtest path.

3. Compute the annualized daily Sharpe ratio for each path.

4. The result is a distribution of 252 Sharpe ratios per (asset, regime, config, cost, signal_type).

And then I use simulated distribution to derive those metrics:
Probability of Backtest Overfitting - which is a fraction of 252 with Sharpe <= 0.

Median Path Sharpe - median of the 252-path Sharpe distribution.
Sharpe Confidence Interval - 90% confidence interval from the simulated distribution.

Variance Ratio Tests

Lo-MacKinlay Variance Ratio Test

Under the random walk hypothesis, the variance of q-period returns should equal q times the variance of one-period returns. The variance ratio statistic is:

$$VR(q) = \frac{\text{Var}\big(r_t(q)\big)}{q \cdot \text{Var}(r_t)} = 1 + 2\sum_{k=1}^{q-1}\left(1 - \frac{k}{q}\right)\hat{\rho}(k) \tag{29}$$

where $\hat{\rho}(k)$ is the k-th order autocorrelation of returns and $r_t(q) = \sum_{j=0}^{q-1} r_{t-j}$ is the q-period return. Under the null hypothesis of a random walk, $VR(q) = 1$.

Lo and MacKinlay (1988) derive a heteroskedasticity-robust test statistic:

$$M_1^*(q) = \frac{VR(q) - 1}{\sqrt{\hat{\phi}^*(q)}} \xrightarrow{d} N(0,1) \tag{30}$$

where $\hat{\phi}^*(q)$ is a heteroskedasticity-consistent variance estimator:

$$\hat{\phi}^*(q) = \sum_{k=1}^{q-1}\left(\frac{2(q-k)}{q}\right)^2 \hat{\delta}(k) \tag{31}$$

and

$$\hat{\delta}(k) = \frac{\sum_{t=k+1}^{T}(r_t - \hat{\mu})^2(r_{t-k} - \hat{\mu})^2}{[\sum_{t=1}^{T}(r_t - \hat{\mu})^2]^2} \quad (32)$$

$VR(q) > 1$ indicates positive serial correlation (momentum), while $VR(q) < 1$ indicates negative serial correlation (mean reversion). The heteroskedasticity-robust formulation is essential for cryptocurrency data, which exhibits pronounced volatility clustering and fat tails.

The test is applied to daily realized returns at holding periods $q \in \{2,4,8,16,32\}$ for each asset within each of the three ETF regimes. The primary comparison is whether $|VR(\mathrm{q})| \rightarrow 1$ post-ETF, indicating convergence toward random walk behavior.

Chow-Denning Joint Variance Ratio Test

Chow and Denning (1993) test considers the maximum absolute value of the individual VR statistics across a set of holding periods $\{q_1, q_2, \dots, q_m\}$:

$$MV^*(q) = \max_{1 \le i \le m} |M_1^*(q_i)| \quad (33)$$

Under the null hypothesis that VR(q) = 1 for all q simultaneously (joint random walk). Rejection of the Chow-Denning test indicates that at least one holding period has significant serial dependence.

Nadarajah and Chu (2017) VR Test with odd integer transformations

To test whether efficiency varies with market conditions, I combine AMH rolling window approach that Urquhart(2016) used with odd-integer power transformed AVR test of Nadarajah and Chu (2017). This combination is my own

methodological contribution as Urquhart used fixed VR lags over rolling windows to assess BTC's time-varying efficiency over time, when Nadarajah and Chu used transformation and wild bootstrap over full sample.

For each asset, I resample one minute close prices to daily log returns and construct rolling windows of 730 calendar days which is approximately two years, following Urquhart 2016, with a step forward monthly. This produces 36 overlapping windows per asset from January 2021 to December 2025. And within each window I apply Kim (2009) wild-bootstrapped automatic variance ratio test to two transformed return series r_t^k for $k \in \{1,3\}$. The k = 1 is basically untransformed returns that let's me to replicate Urquhart (2016) and k=3 case adapts Nadarajah and Chu (2017). Odd powers are $\text{sign}(r^k) = \text{sign}(r)$ for all odd k, so each transformation preserves the directional information of the original k. In their work Nadarajah and Chu (2017) used power of 17 for transformation but mentioned that they got same results for smaller and bigger powers. However they analyzed pre 2017 data which had different price regime and where daily price swings of 50% were pretty usual. However, in a current state of market with prices 70 times higher than average in 2016 - transformation to power of 17 structurally ruins test where one extreme observation affects whole test results. Therefore I test versus k=1 (i.e. no transformation) which is direct replication of Urquhart (2016) test and k=3 as a non linear extension that can capture skewness-related serial structure without extreme sensitivity to outliers.

The test statistic is $VR(q^*)$ where q^* is selected automatically by the Andrews (1991) data-driven bandwidth rule: an AR(1) model is fitted to the demeaned series to estimate the optimal lag, clipped to $[2, n/4]$. Inference uses the Kim (2009) wild bootstrap with $B = 499$ normal-distributed replications: pseudo-

samples are generated as $r_t^* = (r_t - \bar{r})\eta_t$ where $\eta_t \sim \mathcal{N}(0,1)$, and the two-sided bootstrap p-value is the fraction of replicates satisfying $| VR^*(q^*) - 1 | \geq | VR(q^*) - 1 |$. The null hypothesis of a random walk is rejected at the 5% level when this p-value falls below 0.05.

Each rolling window is assigned a regime label based on its temporal position: "pre-ETF" if the window ends before January 10, 2024; and "transitional" if the window extends into or beyond the post-ETF period. Because the two-year window length exceeds the six-month gap between BTC ETF approval (January 2024) and ETH ETF approval (July 2024) - no window falls entirely within the BTC-ETF-only regime. Therefore transitional windows capture the post-ETF environment and do not allow us to distinguish if observed efficiency patterns are attributable to the BTC ETF, or to the ETH ETF, or their combined effect. This is a trade-off I had to take because the 730-day window provides sufficient statistical power for the AVR test (each window contains approximately 730 daily observations), at the cost of temporal granularity in regime attribution. However I think that the ML-based predictability analysis and the static VR tests, which operate on regime-specific subsamples, complement this limitation by providing regime-separated efficiency assessments.

Bai-Perron Structural Break Test

Specification

The Bai-Perron (1998, 2003) procedure detects multiple structural breaks at unknown dates in a linear regression framework. Consider the model:

$$z_t = \mu_j + \varepsilon_t, \quad t = T_{j-1} + 1, \dots, T_j, \quad j = 1, \dots, m+1 \qquad (34)$$

where z_t is the dependent variable (realized volatility), m is the number of breaks, and $T_1, \dots, T_m$ are the break dates. The procedure estimates break dates by minimizing the global sum of squared residuals:

$$(\hat{T}_1, \dots, \hat{T}_m) = \arg\min_{T_1,\dots,T_m} \sum_{j=1}^{m+1} \sum_{t=T_{j-1}+1}^{T_j} (z_t - \bar{z}_j)^2 \qquad (35)$$

where $\bar{z}_j$ is the regime-specific mean. The sequential $\sup F_T(l+1 \mid l)$ test is used to determine the number of breaks, testing the null of l breaks against the alternative of $l+1$ breaks.

Application

The Bai-Perron procedure is applied to daily realized volatility rather than daily log returns. This choice is motivated by two considerations. First, the theoretical expectation that ETF introduction alters market participant composition, liquidity provision, and arbitrage activity - changes that manifest primarily in the second moment of returns (Ben-David, Franzoni, and Moussawi 2018). Second, the well-documented low statistical power of mean-shift tests applied to daily return series, where the variance-to-mean ratio renders break detection infeasible at conventional significance levels (Pettenuzzo and Timmermann 2011). This approach is further supported by Charfeddine and Maouchi (2019), who applied the Bai-Perron procedure to both return and volatility series of major cryptocurrencies and found significant breaks only in volatility.

Daily realized volatility for the Bai-Perron test is computed as:

$$RV_t^{\text{daily}} = \sqrt{\sum_{i=1}^{M} r_{t,i}^2} \qquad (36)$$

where $r_{t,i}$ is the i-th intraday return on day t and M is the number of intraday observations.

If the endogenously estimated break dates aligns with the ETF approval dates (January 10, 2024 and July 23, 2024), this validates the exogenous regime specification used in the RF and VR analyses. If breaks are detected for BTC, ETH, ADA, and SOL but not for XRP, this supports the interpretation that structural changes are attributable to ETF introduction specifically rather than to whole crypto market effect.

Chapter 4

DATA

Tables 2 - 6 present descriptive statistics for high-frequency, 1-minute data on four major cryptocurrencies: Bitcoin (BTC), Ethereum (ETH), Cardano (ADA), and Solana (SOL). The data is split into three periods reflecting successive regulatory events:

Pre-ETF: Jan 1, 2021 - Jan 9, 2024.

BTC-ETF: spot Bitcoin ETFs, but prior to the approval of Ethereum ETFs. (Jan 10, 2024 - Jul 22, 2024)

Full-ETF: The period after spot Ethereum ETFs (Jul 23, 2024 - Dec 31, 2025).

Data shows clear split in trading volume patterns between assets with ETF approval and altcoins. Overall volatility dropped across the board indicating market maturing.

In Table 2 we can see BTC's average minute by minute trading volume fell from $2.07 million before ETF to $1.52 million after, reason for that could be that volume shifting from Binance exchange to exchange traded funds. For ETH (Table 3) we can see different effect with volume raising after ETF approval which can indicate higher trust in ETH as an asset that ETF approval caused. Also Solana (Table 5) have trading volume tripled, reason for that can be that speculative money that left Bitcoin rotated into Solana.

To highlight again - realized volatility dropped for all assets, indicating market maturing. But despite that - liquidity indicators show that liquidity got worse. Roll

Spread became wider for all pairs, that could indicate that market makers on Binance widened spread because exchanges brought back trading fees and big liquidity providers moved to direct investment in US. Kyle's lambda spiked as well so the same trading volume that barely moved prices before ETFs now moves them a lot more.

Returns show the usual financial data pattern - heavy tails and non normal distributions across all assets and regimes. Massive excess kurtosis and Jarque-Bera stats of 0 confirm Nadarajah and Chu (2017) observations of distributional characteristics of crypto. But it is worth mentioning that excess kurtosis fell for Bitcoin (Table 2) during BTC-ETF and even though it grew in Full ETF period - still it is lower than in pre ETF. Solana (Table 5) also shows decrease in excessive kurtosis in full ETF compared to pre ETF, but other coins otherwise show that excessive kurtosis increased dramatically - almost 3 times higher for ETH and 6 times to 10 times higher for XRP (Table 6) and ADA (Table 4).

Table 2. BTC Descriptive statistics, 1 minute data

Metric	Pre-ETF	BTC-ETF	Full-ETF
Return (1-min log)	2.83×10^{-7} (9.85×10^{-4})	1.36×10^{-6} (7.42×10^{-4})	3.44×10^{-7} (6.65×10^{-4})
Volume (USD/min)	2.07×10^{6} (3.00×10^{6})	1.58×10^{6} (2.46×10^{6})	1.52×10^{6} (2.72×10^{6})
Realized Vol (100 window)	0.0077840 (0.0060350)	0.0063317 (0.0038765)	0.0055604 (0.0036488)
Amihud (100 window)	4.52×10^{-10} (2.58×10^{-10})	4.44×10^{-10} (1.62×10^{-10})	4.77×10^{-10} (2.16×10^{-10})
Roll Spread (100 window)	8.5620 (13.8382)	12.379 (18.271)	15.177 (23.106)
Kyle λ (100 window)	0.93350 (0.77407)	2.0928 (0.9376)	3.1081 (1.6772)
VPIN-BVC (w100)	0.54293 (0.07235)	0.55470 (0.07266)	0.55526 (0.07304)
Skewness	-0.06204	-0.5951	-0.5069

Metric	Pre-ETF	BTC-ETF	Full-ETF
Exc. Kurtosis	94.51	32.84	78.65
JB stat	5.91×10^{8}	1.26×10^{7}	1.96×10^{8}
N obs	1,588,587	280,800	758,880

Table 3. ETH Descriptive statistics

Metric	Pre-ETF	BTC-ETF	Full-ETF
Return (1-min log)	7.20×10^{-7}	1.37×10^{-6}	-1.93×10^{-7}
	(1.23×10^{-3})	(8.64×10^{-4})	(1.01×10^{-3})
Volume (USD/min)	9.37×10^{5}	8.89×10^{5}	1.10×10^{6}
	(1.56×10^{6})	(1.53×10^{6})	(2.07×10^{6})
Realized Vol (100 window)	0.0096427	0.0073473	0.0085025
	(0.0075966)	(0.0045472)	(0.0053850)
Amihud (100 window)	1.25×10^{-9}	1.08×10^{-9}	1.20×10^{-9}
	(5.00×10^{-10})	(4.49×10^{-10})	(6.41×10^{-10})
Roll Spread (100 window)	0.64117	0.86672	0.82896
	(1.01587)	(1.31450)	(1.26261)
Kyle λ (100 window)	0.0059914	0.011119	0.0078967
	(0.0049937)	(0.005268)	(0.0042839)
VPIN-BVC (100 window)	0.55642	0.55987	0.56123 (0.07265)
	(0.07418)	(0.07336)	
Skewness	-0.1752	-1.407	-2.294
Exc. Kurtosis	116.7	117.8	304
JB stat	9.02×10^{8}	1.62×10^{8}	2.92×10^{9}
N obs	1,588,587	280,800	758,880

Table 4. ADA Descriptive statistics

Metric	Pre-ETF	BTC-ETF	Full-ETF
Return (1-min log)	6.47×10^{-7}	-6.56×10^{-7}	-3.23×10^{-7}
	(1.60×10^{-3})	(1.17×10^{-3})	(1.53×10^{-3})
Volume (USD/min)	1.73×10^{5}	51558.53	96259.37
	(4.35×10^{5})	(108315.97)	(253216.32)
Realized Vol (100 window)	0.012715	0.0099749	0.011894
	(0.009747)	(0.0060890)	(0.009585)
Amihud (100 window)	5.18×10^{-8}	4.96×10^{-8}	4.94×10^{-8}
	(1.08×10^{-7})	(4.90×10^{-8})	(5.94×10^{-8})
Roll Spread	4.53×10^{-4}	1.66×10^{-4}	2.56×10^{-4}

Metric	Pre-ETF	BTC-ETF	Full-ETF
(100 window)	(7.75×10^{-4})	(2.23×10^{-4})	(4.38×10^{-4})
Kyle λ	7.39×10^{-9}	4.16×10^{-9}	4.64×10^{-9}
(100 window)	(7.36×10^{-9})	(1.93×10^{-9})	(2.75×10^{-9})
VPIN-BVC (100 window)	0.55814 (0.07471)	0.56508 (0.07439)	0.56255 (0.07146)
Skewness	-0.3685	-4.16	-10
Exc. Kurtosis	209.3	308.1	1924
JB stat	2.90×10^{9}	1.11×10^{9}	1.17×10^{11}
N obs	1,588,587	280,800	758,880

Table 5. SOL Descriptive statistics

Metric	Pre-ETF	BTC-ETF	Full-ETF
Return (1-min log)	2.60×10^{-6}	2.09×10^{-6}	-4.76×10^{-7}
	(1.96×10^{-3})	(1.38×10^{-3})	(1.21×10^{-3})
Volume (USD/min)	1.69×10^{5}	4.89×10^{5}	4.95×10^{5}
	(3.83×10^{5})	(7.28×10^{5})	(9.08×10^{5})
Realized Vol (100 window)	0.015927 (0.011440)	0.012078 (0.006751)	0.010488 (0.006053)
Amihud	5.13×10^{-8}	3.55×10^{-9}	3.19×10^{-9}
(100 window)	(1.25×10^{-7})	(1.62×10^{-9})	(1.54×10^{-9})
Roll Spread (100 window)	0.027620 (0.054749)	0.058479 (0.080415)	0.056342 (0.084381)
Kyle λ	3.90×10^{-5}	6.27×10^{-5}	6.24×10^{-5}
(100 window)	(6.44×10^{-5})	(3.15×10^{-5})	(3.66×10^{-5})
VPIN-BVC (100 window)	0.56340 (0.07384)	0.55742 (0.06388)	0.55450 (0.06692)
Skewness	1.01	-0.6888	0.1889
Exc. Kurtosis	132.1	74.91	32.37
JB stat	1.16×10^{9}	6.57×10^{7}	3.31×10^{7}
N obs	1,588,586	280,800	758,880

Table 6. XRP Descriptive statistics

Metric	Pre-ETF	BTC-ETF	Full-ETF
Return (1-min log)	5.84×10^{-7}	2.49×10^{-7}	1.46×10^{-6}
	(1.64×10^{-3})	(1.02×10^{-3})	(1.39×10^{-3})

Metric	Pre-ETF	BTC-ETF	Full-ETF
Volume (USD/min)	2.47×10^5 (5.60×10^5)	1.64×10^5 (3.06×10^5)	3.64×10^5 (8.54×10^5)
Realized Vol (100 window)	0.012113 (0.011045)	0.0082017 (0.0060035)	0.010489 (0.009074)
Amihud (100 window)	8.72×10^{-9} (7.82×10^{-9})	7.83×10^{-9} (5.55×10^{-9})	6.00×10^{-9} (4.67×10^{-9})
Roll Spread (100 window)	2.62×10^{-4} (5.02×10^{-4})	1.61×10^{-4} (2.21×10^{-4})	6.55×10^{-4} (1.08×10^{-3})
Kyle λ (100 window)	2.85×10^{-9} (2.13×10^{-9})	1.78×10^{-9} (7.78×10^{-10})	1.28×10^{-8} (9.31×10^{-9})
VPIN-BVC (100 window)	0.54622 (0.07202)	0.55106 (0.07305)	0.55337 (0.06887)
Skewness	-1.636	-4.044	-6.324
Exc. Kurtosis	264.8	296.7	1815
JB stat	4.64×10^9	1.03×10^9	1.04×10^{11}
N obs	1,588,588	280,800	758,880

Chapter 5

ESTIMATION RESULTS

Structural Breaks

The Bai-Perron multiple structural break procedure for realized volatility identifies a break in mid-July 2023 for BTC which is aligned with BlackRock's filing for a spot ETF. A second break appears at the end of 2023, it is consistent with the period of intensive speculation on ETF approval and the "buy-the-rumor sell-the-news" dynamics that followed the actual January 2024 announcement. Both breaks appear in all assets analyzed except XRP, where structural breaks identified looks more aligned with Brazil's ETF announcement and approval. Figure 2 plots the structural breaks for BTC; Table 5 details the estimated break dates across assets. No structural breaks identified when used for realized returns. So Hypothesis 3 is partially supported. Structural breaks exist in realized volatility, but they occurred 3-6 months earlier than ETF approval date. Volatility regime changed when the market became convinced approval was likely to happen not when formally occurred. And pattern consistent across all four coins BTC, ETH, SOL and ADA but absent for XRP which is likely happened due to unique regulatory situation of XRP.

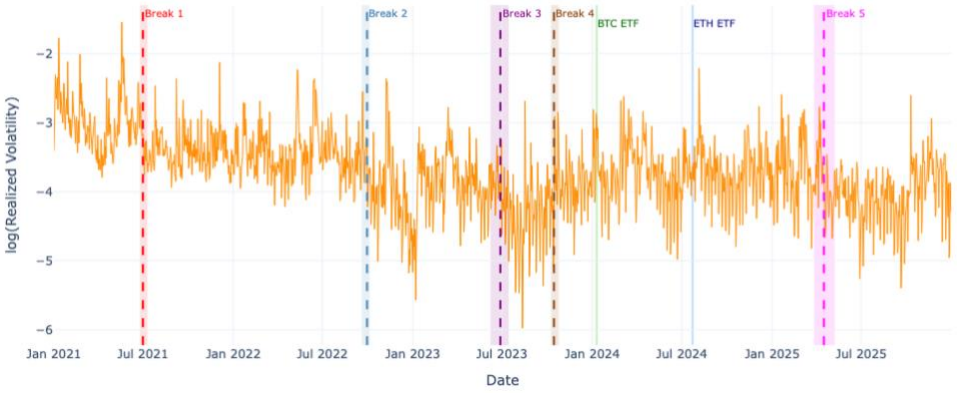


Figure 1. BTC Daily Log-transformed Realized Volatility with Bai-Perron Structural Breaks

Table 7. Bai-Perron structural break results for Log (Realized Volatility)

Asset	Break 3 [CI]	Break 4 [CI]	Break 5 [CI]
BTC	2023-06-28	2023-10-15	2025-04-16
	[06-08, 07-15]	*[10-09, 10-25]*	[03-27, 05-08]
ETH	2023-05-12	2023-10-22	2024-11-05
	[05-03, 05-20]	*[10-16, 10-29]*	[09-27, 12-14]
ADA	2023-10-30	2024-11-05	2025-03-13
	[09-26, 12-05]	[10-29, 11-10]	*[03-07, 03-22]*
SOL	2023-03-24	2023-10-18	2024-05-02
	[03-06, 04-25]	*[10-09, 10-27]*	*[04-20, 05-15]*
XRP	2022-11-25	2024-11-09	2025-03-13
	[11-04, 12-10]	*[11-04, 11-14]*	*[03-08, 03-20]*

Variance Ratio Tests for Weak-Form Efficiency under Efficient Market Hypothesis

Figure 2 presents Lo and MacKinlay (1988) variance ratio statistics on daily realized volatility across holding periods $q \in \{2, 4, 8, 16, 32\}$ days. Full results in Table A1 in Appendix A.

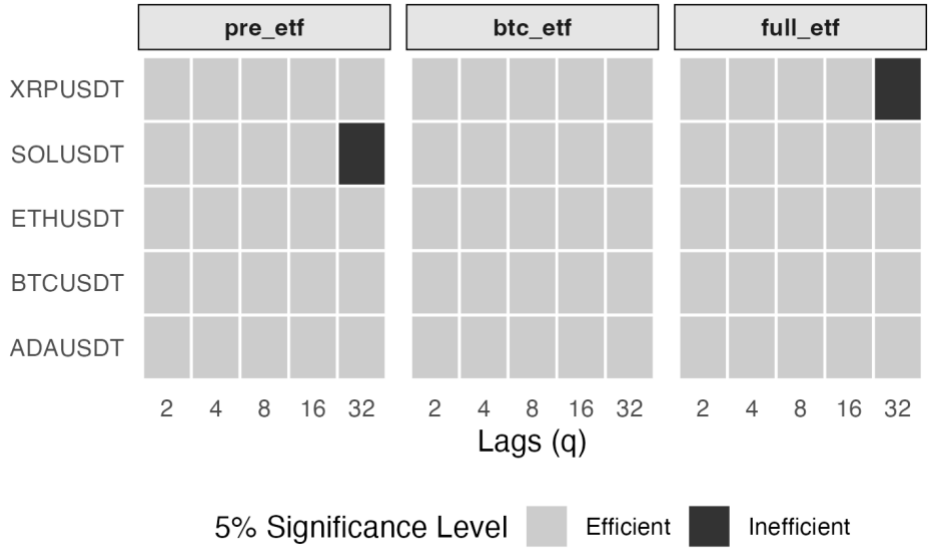


Figure 2. Lo-MacKinlay Random Walk Rejection Heatmap

Overall dominant pattern is that VR are statistically indistinguishable from 1 across all almost coins and regimes, only SOL and XRP showing significant efficiency gains after ETF approval. Out of 75 individual VR tests only two reject random walk null and 95% confidence level: SOL pre-ETF and XRP in the full-ETF both at q=32. This means daily realized volatility generally follow a random walk across all five assets, consistent with weak-form efficiency.

Chow-Denning (1993) join test evaluates random walk null simultaneously across all five holding periods and corrects for multiple comparison problem. Figure 3 presents results with basically no statistically significant results at 95% confidence

after Bonferroni correction. And this test reinforces Lo-MacKinlay test findings when the multiple comparison problem is properly addressed, so there is a strong probability that two statistically significant results I observed in LoMac test is just an error, which observed by the joint test framework. Full results in Appendix A, Table A2.

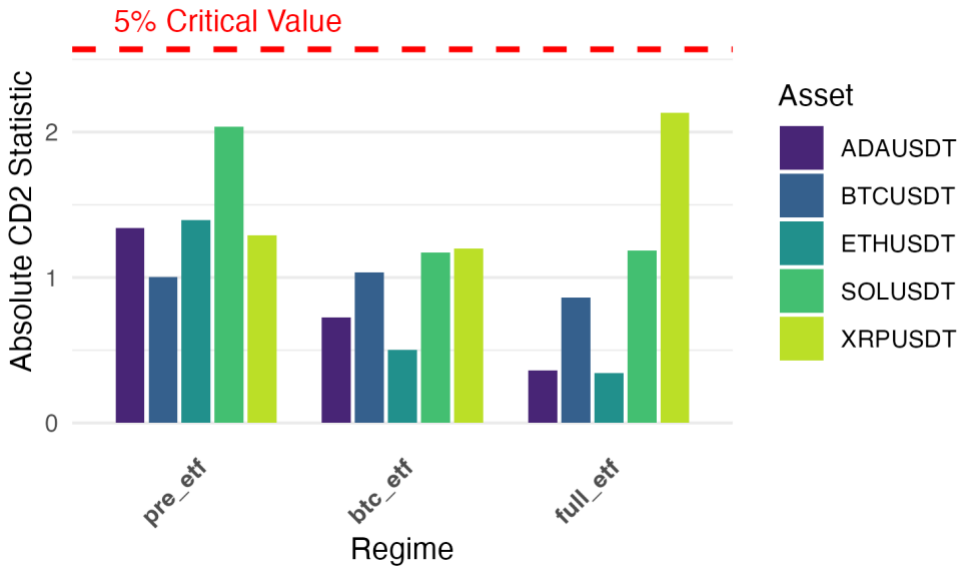


Figure 3. Chow-Denning Robust Test Statistic results

So Hypothesis 2 is not supported. Data doesn't show that VR deviations from 1 decreased after ETF approval and VR tests fail to reject random walk null in almost all cases which implies weak-form efficiency.

Time-Varying Efficiency Under the Adaptive Market Hypothesis

In Figure 4 and 5 presented results from rolling two-year window with wild bootstrapped automatic variance test for k=1 and k=3 respectively. Results are different to static tests as expected and shows that results BTC, ETH, ADA and

SOL shows 100% efficient windows across both k-values and rolling periods from 2021 through 2025 and no statistically significant rejection of the random walk hypothesis at the 95% confidence level. XRP is an exception here, at k=1 six out of 36 windows were classified as inefficient, but at the k=3 transformation (following Nadarajah and Chu 2017) all windows tested as efficient. And actually this finding aligns with their similar finding on early BTC era suggesting that odd integer power transformation eliminated detectable autocorrelation structure for XRP without information loss, so I can conclude that it holds not only for Bitcoin as they showed but for all five cryptocurrencies that I used.

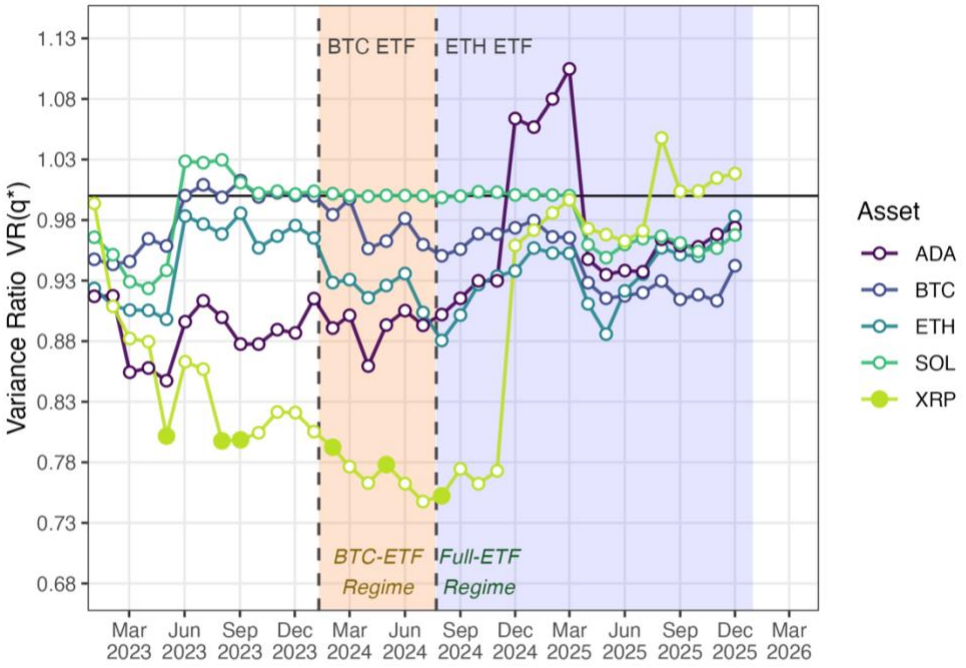


Figure 4. Rolling window Variance Ratio test result at k=1. Solid point = inefficient, hollow points = efficient.

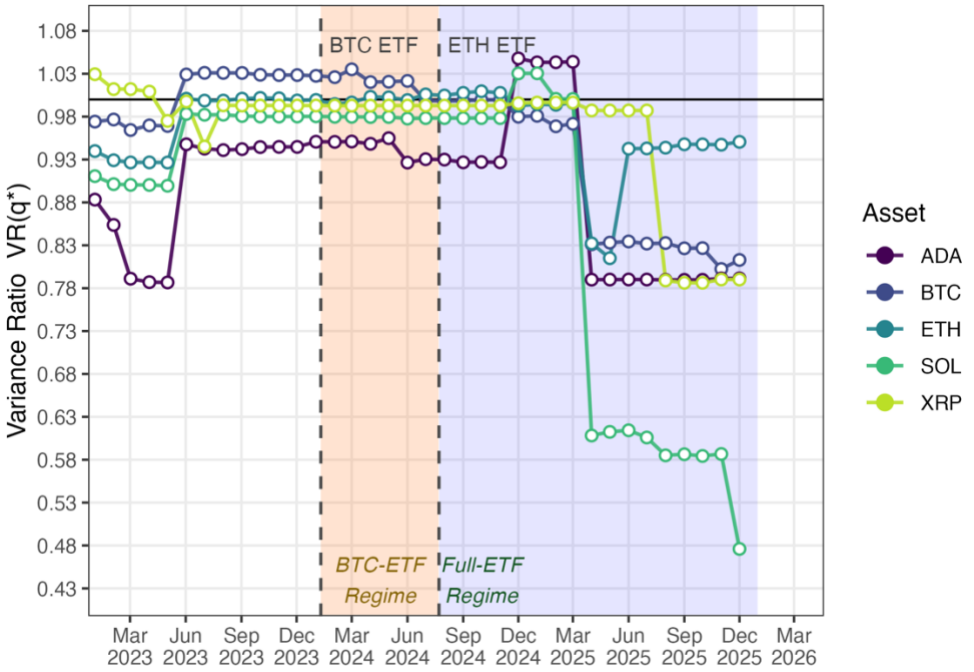


Figure 5. Rolling window AVR test results at k=3
Solid point = inefficient, hollow points = efficient.

Idea of Adaptive Market Hypothesis Framework (Lo 2004) is that efficiency should vary with market conditions. What we see here is that for the four major assets efficiency doesn't vary and uniformly maintained, and after transformation for all assets. This could mean one of two things: either market conditions didn't change enough to push these assets away from efficiency, or the two-year rolling window is too long to capture some episodes of inefficiency. So hypothesis 4 is not supported, results show that assets were efficient prior ETF approval and event did not change it. Urquhart (2016) finding that Bitcoin was "moving towards efficiency" appears to be completed before my sample begins (2021). Results also consistent with Nadarajah and Chu (2017) findings.

Out of Sample Predictability of Microstructure Features

The Random Forest results, shown in Figure 6, tell a more detailed story: for BTC, mean AUC converged toward 0.5 only during the BTC-ETF regime. But in the Full-ETF period it returned to its pre-ETF level - and this pattern holds across four cryptocurrencies where XRP is a partial exception. In BTC-ETF regime AUC is 0.536 barely drops from pre-ETF period and stays the same in Full-ETF period. So XRP's weaker response is consistent with its retail driven trading base, own regulatory events history and lower institutional overall with BTC/ETH ecosystem. Overall findings does not support the hypothesis that ETF approval permanently improved informational efficiency in crypto spot markets.

As per feature importance analysis for model with unlimited tree depth configuration (Figure A1): in pre-ETF the top predictor for BTC were Roll spread (L=100) and it aligns with Easley et al. (2024), other were Realized Volatility, Kyle lambda and Etherium's VPIN, which were not mentioned as top features in original paper. All other cryptocurrencies also had Roll Spread as main predictor in pre-ETF and full-ETF periods, which is also consistent with Easley et al. (2024). In BTC-ETF period for BTC almost all features has near 0 feature importance which is expected considering AUC in this period around 0.5. And in Full ETF I observe that Roll Spread gained the most predictive power compared to previous periods and remained dominant predictor, while other top predictors are Realized Volatility. My interpretation is that feature importance structure didn't change indicating that ETF approval didn't affect market microstructure as well and made previous findings of Easley et al. (2024) even more pronounced.

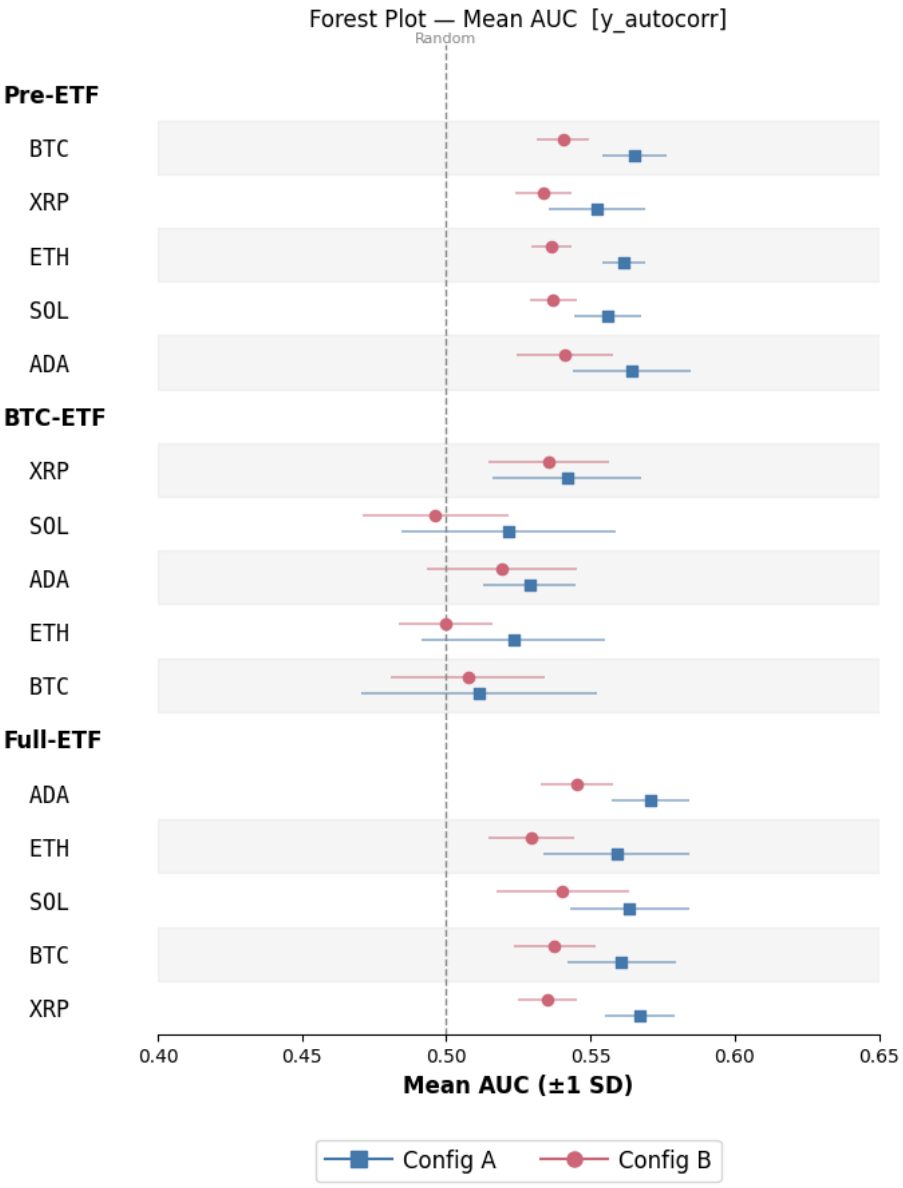


Figure 6. Mean AUC for Sign of Change in Returns Autocorrelation

Economic Significance Test

In this test we have an answer if inefficiencies captured right after BTC ETF approval can be exploitable. Table 8 reports mean daily strategy returns in basis points (bps) and their statistical significance for two model configurations across pre-ETF and full-ETF regime.

Results are not straightforward - strategy returns are negative. Config A with un-regularized model and lower AUC scores produces significant losses for BTC in both regimes, ETH full ETF, SOL and XRP full ETF, as expected, predictions are almost random and transaction cost simply produces losses.

Config B regularized predictions slightly better - returns are closer to zero and mostly statistically insignificant. ETH pre ETF shows tine positive mean and SOL full ETF as well, but neither results approaches statistical significance. Overall the only interesting results from economical standpoint of view is absence of significant losses for regularized config across most combinations and it suggests regularized strategy at least avoids destruction of capital and maybe more sophisticated strategy would show better performance using model's prediction.

In Table 9 reported cumulative Sharpe ratios at different transaction cost level showing cost sensitivity and its implications on strategy performance evaluation. (Note: selected rows shown in Table 9, full results in appendix with all 30 asset-regime-config combinations).

Table 8. Mean Daily Returns (bps) and Statistical Test Results

Asset	Config	Pre-ETF Mean (bps)	t-stat	p-value	Full-ETF Mean (bps)	t-stat	p-value
BTC	A	−0.462	−2.94	0.003**	−0.511	−4.41	<0.001**
BTC	B	−0.077	−0.33	0.742	−0.317	−2.11	0.035**
ETH	A	−0.376	−1.91	0.057*	−0.566	−3.39	<0.001**
ETH	B	+0.059	+0.21	0.833	−0.023	−0.10	0.923
ADA	A	−0.380	−1.36	0.173	−0.071	−0.22	0.825
ADA	B	+0.014	+0.04	0.965	−0.033	−0.10	0.924
SOL	A	−0.833	−2.01	0.044**	−0.360	−1.73	0.084*
SOL	B	−0.156	−0.37	0.708	+0.054	+0.20	0.839
XRP	A	−0.524	−1.58	0.114	−0.580	−2.20	0.028**
XRP	B	+0.004	+0.01	0.990	−0.607	−1.60	0.109

Table 9. Cumulative Sharpe Ratios over Transaction Cost Levels

Asset	Regime	Config	TC=5	TC=10	TC=15	TC=25
BTC	Pre-ETF	A	−0.70	−2.05	−3.40	−6.11
BTC	Pre-ETF	B	+0.19	−0.22	−0.62	−1.43
BTC	BTC-ETF	A	−3.30	−4.91	−6.53	−9.74
BTC	BTC-ETF	B	−0.82	−1.51	−2.19	−3.55
BTC	Full-ETF	A	−1.63	−4.13	−6.63	−11.60
BTC	Full-ETF	B	−1.05	−1.94	−2.84	−4.60

Asset	Regime	Config	TC=5	TC=10	TC=15	TC=25
ETH	Pre-ETF	B	+0.44	+0.13	−0.18	−0.79
ETH	Full-ETF	B	+0.46	−0.07	−0.61	−1.68
SOL	BTC-ETF	B	+2.11	+1.76	+1.40	+0.68
XRP	BTC-ETF	B	+2.48	+2.20	+1.93	+1.36

In Table 9 we can see important structure that even with the most favorable assumption of 5 bps transaction cost, Config A with unregularized trees produce negative Sharpe Ratios with BTC-ETF regime being particularly costly, I think it is because strategy trades in a regime where predictions are least accurate.
Also in Config B we see really high Sharpe Ratios for SOL (+2.11 at 5 bps) and XRP (+2.48 at 5 bps) with positive SR even at 25 bps. The CPCV Sharpe distribution (Figure 7) confirms that these are not point estimates with median across 252 CPCV paths is positive for both assets with IQ for XRP lies entirely above zero.

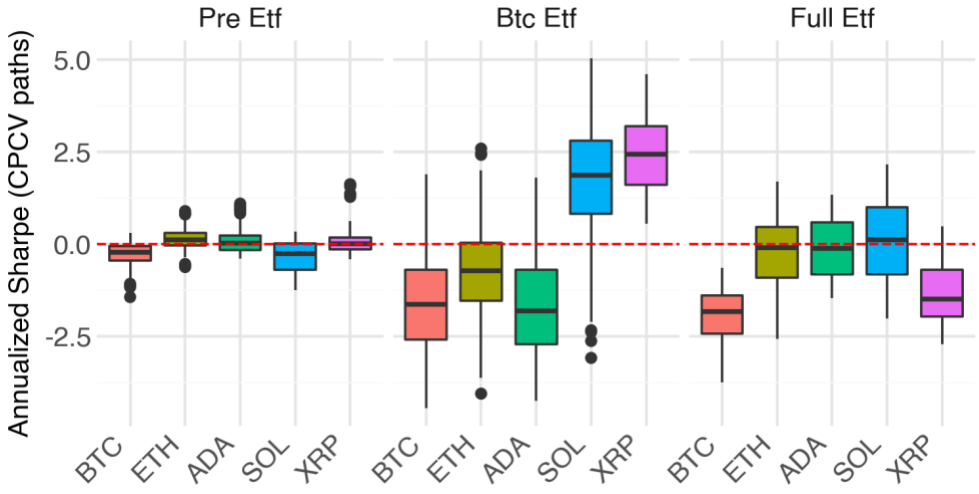


Figure 7. CPCV Annualized Sharpe Distribution
with Config B at 5 bps

Also results of Probability of Backtest Overfitting (PBO) confirms those results and showing more insights. In Chart 8 visualized PBO for model with config A and cost pbs = 10, where probability below 0.5 suggest results are robust. And here conditional and neutral zone strategies show PBO of 1 or close for almost all coins. The Binary strategy which goes only long or does nothing shows smaller PBO values and I think it says that model predictions are valuable if they go together with bullish market trend.

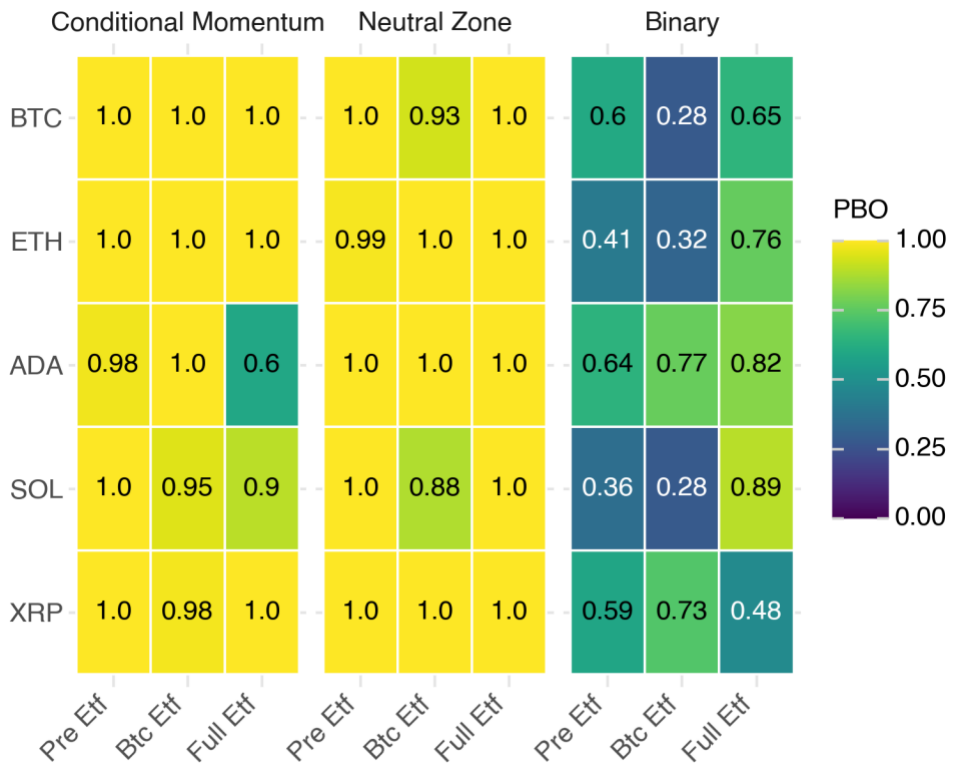


Figure 8. Probability of Backtest Overfitting Heatmap. Model Config A, Cost 10 pbs.

In Figure 9 presented heatmap for PBO for regularized model (config B) and cost bps 10. Here what I see it that conditional momentum strategy shows robust performance for some asset-regime combinations. XRP in the BTC-ETF regime achieves PBO = 0 meaning none of the 252 paths produced overfitting. SOL in

BTC ETF regime has PBO = 0.05 which is also well below threshold. For ETH and ADA pre-ETF similarly low PBO indicate that strategy captured real signal. However all those results don't persist - XRP PBO jumps to 0.9 in full-ETF regime, BTC to 0.95 so it leads to conclusion that signal was real but later was consumed by market.

The neutral zone averaged strategy shows high PBO across nearly all combinations showing that strategy that relies only on prediction doesn't bear profitable signals.

The binary averaged strategy is somewhere in a middle with some combinations below 0.5 that suggests some signal but still inconsistent across regimes.

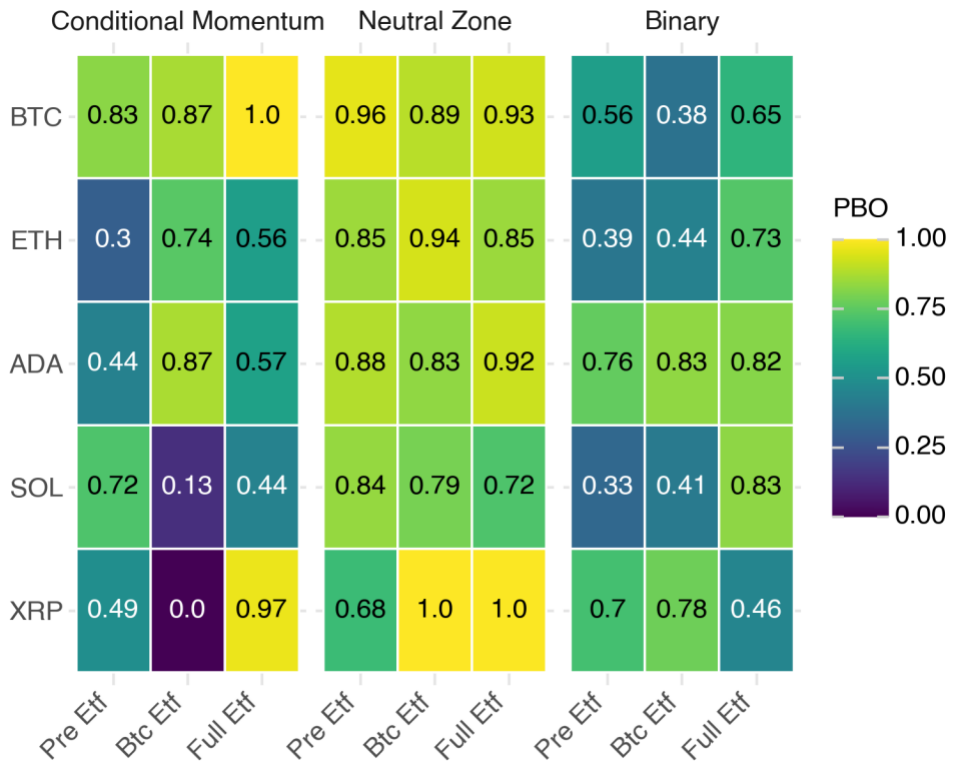


Figure 9. Probability of Backtest Overfitting Heatmap. Model Config B, Cost 10 pbs.

So CPCV analysis adds important insight to economic significance test that positive SR for SOL and XRP during BTC-ETF regime are real and exploitable during a specific six-month window when BTC ETF introduction happened. And I think this finding actually strengthens practical efficiency conclusion that signal was real and exploitable in narrow window for one specific strategy type and two specific assets. It didn't generalize to other regimes and trader would need to know in advance which strategy to deploy and when to exit so practically opportunity was not possible to identify.

Synthesis of findings

Statistical predictability exists but rarely translates into exploitable inefficiency. ML models detected above random AUC consistently but economic significance tests who that this signal generally produced losses. CPCV analysis showed that conditional momentum strategy for SOL and XRP during BTC-ETF was real arbitrage opportunity that appeared during the market shock caused by ETF launch and closed as market adapted. And it is actually consistent with Grossman-Stiglitz equilibrium that some brief windows of exploitability emerge and then market corrects it, with long-run expected profit is zero.

High frequency predictability coexists with low-frequency VR efficiency. So difference in results between ML models and VR tests lays in data frequency as ML models captures high frequency inefficiency but when tested over daily realized volatility there is no inefficiencies left. So the market can simultaneously fail to be informationally efficient while still satisfying weak form efficiency conditions.

XRP behaves differently from the rest as it is the only that showed VR rejection at long horizon in Full-ETF Regime (however not significant after Bonferroni correction), also shows rolling AVR inefficiency episodes at k=1 and no ETF

related Bai-Perron breaks, and highest economic gains during BTC-ETF period. My thought is that it is due unique regulatory events that occurred around XRP.

Another finding is that Nadarajah and Chu transformation generalizes to other currencies. With k=3 transformation eliminates any rolling AVR rejections which extends Nadarajah and Chu (2017) finding on BTC to all five assets.

And finally, that Efficiency should be understood as spectrum and not a binary. As combined evidence supports this view that crypto markets are weak-form efficient in the traditional Variance Ratio sense, practically efficient in the Jensen (1978) sense of non-exploitability, but not informationally efficient at the microstructure level where short-horizon patterns are possible to detect with Machine Learning models, and I guess it just aligns again with Jensen that degree of informational inefficiency is calibrated to the cost of extracting it.

Chapter 6

CONCLUSIONS AND POLICY RECOMMENDATIONS

Conclusions

The ETF approval had real but temporary effect on crypto market efficiency. In a window between BTC ETF approval and prior to ETF approval predictability of next day sign of change in returns autocorrelation converged to 0.5 but returned to pre ETF level (closer to 0.6), also Bai Perron breaks cluster around ETF approval event. So market became harder to predict, then adapted and returned to previous state. And the microstructure didn't change - consistent with Easley et al. (2024) findings Roll spread is a dominant predictor for all coins and regimes that I analyzed. But you can't make money off those short term predictions under realistic transaction costs, tests showed negative returns or statistically indistinguishable from zero for all coins. The only exception in results was SOL and XRP during BTC-ETF window, which had high Sharpe Ratio which held under Combinatorial Purged Cross-Validation, and showed low probability of back test overfitting. But caveat is that trader would need to know in advance which coins to trade, which strategy and when to exit, so in my opinion practically it wasn't identifiably. So overall results shows real example of Grossman-Stiglitz paradox where some predictability exists but it is too expensive to exploit which make market practically efficient.

And it is also consistent under EMH and AMH, robust variance tests showed that all coins are weak-form efficient pre-ETF and post-ETF. Also all of them are efficient under AMH at every horizons, however XRP showed efficiency after k=3 transformation.

Also I want to flag findings on structural breaks - volatility regime changes in the middle of 2023 and end of 2023 which consistent with BlackRock's filling and media speculation peak. So it was 3-6 months before actual approval and no break were detected at the approval event itself. Market prices reflected ETF once investors became convinced that approval was coming and formal announcement was a not shock event - which is clear example of "buy the rumor, sell the news" market behavior. Interesting that it hold for all coins but XRP, which had its own regulatory timeline (lawsuit and Brazil ETF), so the market responded to that coin's own regulatory situation and not "ETF was approved somewhere signal, let's buy all".

I would say that my main takeaway is that efficiency isn't binary, crypto market is simultaneously weak-form efficient on daily data, not informationally efficient at the microstructure level and practically efficient in a way that you can't exploit microstructure inefficiencies. Major cryptocurrencies looks like normal and mature financial securities from this perspective.

Policy Recommendations

The first takeaway for policy makers is that approving ETF for the underlying asset is not enough to improve informational efficiency of this asset. And if the goal is a better functioning market then ETF doesn't move a needle a lot and you need you look at the spot market microstructure directly.

The second is about fee structures in spot market. It was not something I tested directly and rather extension of Grossman-Stiglitz, but my findings are that detectable predictability stays not exploitable only because transactions costs eat all profits. So when crypto exchanges start some marketing campaigns such as zero-fee trading - it can break equilibrium and make short term inefficiencies

exploitable. And my recommendation for regulators to watch for promotional or unstable fee structures as it could create negative effect on market efficiency.

And a last thing I want to flag is that realized volatility fell across all coins after ETF approval, but liquidity metrics got worse, Roll spreads widened and Kyle's lambda spiked. My hypothesis is that with ETF approval big liquidity providers moved from Binance into regulated US ETF channel. And if it is right and it was a goal of US SEC - then takeaway that ETF improved access for institutions but harmed trading channel that retail investors still used.

Limitations

Several constraints on interpretation should be acknowledged.

The analysis relies exclusively on Binance data, one venue in a fragmented global market. Results may not generalize to other spot exchanges, regulated futures markets (CME), or the ETF market itself. A direct extension would replicate the RF analysis on Coinbase - the exchange most closely linked to the U.S. institutional ecosystem - to test whether efficiency improvements were venue-specific.

The Full-ETF regime sample (July 2024 to end-2025) encompasses significant macroeconomic and crypto-specific events, including the November 2024 U.S. election and its associated crypto policy expectations. Disentangling ETF effects from concurrent macro regime shifts is inherently difficult without a longer post-ETF window.

WORKS CITED

Acharya, Viral V., and Lasse Heje Pedersen. (2005). *Asset Pricing with Liquidity Risk*. **Journal of Financial Economics** 77, no. 2: 375-410.

Amihud, Yakov (2002). *Illiquidity and stock returns*. **Journal of Financial Markets,** 5(1), 31-56.

Babalos, Vassilios & Bouri, Elie & Gupta, Rangan. (2025). *"Does the introduction of US spot Bitcoin ETFs affect spot returns and volatility of major cryptocurrencies?,"* **The Quarterly Review of Economics and Finance, Elsevier,** vol. 102(C).

Bai, Jushan & Perron, Pierre. (1998, 2003). *Estimating and testing linear models with multiple structural changes*; *Critical values*…. **Econometrica**, Vol. 66, No. 1, 47-48.

Borri, Nicola & Liu, Yukun & Tsyvinski, Aleh & Wu, Xi. (2025). *Cryptocurrency as an Investable Asset Class: Coming of Age.* 10.48550/arXiv.2510.14435.

Brière, Marie & Oosterlinck, Kim & Szafarz, Ariane. (2013). *Virtual Currency, Tangible Return: Portfolio Diversification with Bitcoin*. SSRN Electronic Journal. 16. 10.2139/ssrn.2324780.

Easley, David and O'Hara, Maureen and Yang, Songshan and Zhang, Zhibai. (2024). **"***Microstructure and Market Dynamics in Crypto Markets".* SSRN: https://ssrn.com/abstract=4814346

David Easley, Marcos M. López de Prado, Maureen O'Hara. (2012). *"Flow Toxicity and Liquidity in a High-frequency World"*. The Review of Financial Studies, Volume 25, Issue 5, May 2012, Pages 1457–1493.

David Easley & Marcos López de Prado & Maureen O'Hara & Zhibai Zhang & Wei Jiang. (2021). *"Microstructure in the Machine Age [The risk of machine learning],*" The Review of Financial Studies, Society for Financial Studies, vol. 34(7), pages 3316-3363.

Foucault, Thierry, Marco Pagano, and Ailsa Röell (2024). *Market Liquidity: Theory, Evidence, and Policy.* 2nd ed. Oxford: Oxford University Press.

Kyle, Albert S. (1985). "*Continuous auctions and insider trading".* **Econometrica,** 53(6), 1315-1335.

Lee Kuo Chuen, David, ed. 2024. "*Handbook of Digital Currency: Bitcoin, Innovation, Financial Instruments, and Big Data."* 2nd ed. Amsterdam: Elsevier/Academic Press.

Lo, Andrew W & MacKinlay, A. Craig. (1988). "*Stock market prices do not follow random walks.*" **Review of Financial Studies,** 1(1), 41-66.

López de Prado, Macros. (2018). *Advances in Financial Machine Learning.* Hoboken, NJ: John Wiley & Sons.

Makarov, Igor and Schoar, Antoinette. (2020). *Trading and arbitrage in cryptocurrency markets***. Journal of Financial Economics**, 135(2), 293-319.

O'Hara, Maureen (1995). *Market Microstructure Theory*. Hoboken, NJ: Wiley & Sons

Ben-David, Itzhak, Franzoni, Francesco and Moussawi, Rabih, (2018). "*Do ETFs Increase Volatility?*"**. Journal of Finance**, 73, issue 6, p. 2471-2535.

Lanouar Charfeddine, Youcef Maouchi, (2019). *"Are shocks on the returns and volatility of cryptocurrencies really persistent?".* **Finance Research Letters,** Volume 28, Pages 423-430, ISSN 1544-6123, https://doi.org/10.1016/j.frl.2018.06.017.

Fama, Eugene, F., and Kenneth R. French. 2004. *"The Capital Asset Pricing Model: Theory and Evidence."* **Journal of Economic Perspectives,** 18 (3): 25-46.

Bollerslev, Tim & Diebold, Francis & Andersen, Torben & Labys, Paul. (2003). "*Modeling and Forecasting Realized Volatility.*" Econometrica. 71. 579-625. 10.2139/ssrn.267792.

Andrew Urquhart. (2016). "*The inefficiency of Bitcoin,"* **Economics Letters,** Volume 148, Pages 80-82, ISSN 0165-1765.

Nadarajah, S., Chu, J. (2017). "*On the Inefficiency of Bitcoin."* **Economics Letters,** 150, 6-9. *https://doi.org/10.1016/j.econlet.2016.10.033*

Fama, Eugene F. *(1965) "Random Walks in Stock Market Prices."* Financial Analysts Journal 21, no. 5: 55-59.

Fama, Eugene F. (1970) *"Efficient Capital Markets: A Review of Theory and Empirical Work."* The Journal of Finance 25, no. 2: 383-417.

Samuelson, Paul A. (1965*). "Proof That Properly Anticipated Prices Fluctuate Randomly".* Industrial Management Review. 6 (2): 41-49.

Chow, K. Victor and Denning, Karen C. (1993) "*A Simple Multiple Variance Ratio Test.*" Journal of Econometrics, 58, 385-401.

Michael C. Jensen. (1978). "*Some anomalous evidence regarding market efficiency.*" Journal of Financial Economics,Volume 6, Issues 2-3, Pages 95-101.

Andrews, Donald W. K. (1991). "*Heteroskedasticity and autocorrelation consistent covariance matrix estimation.*" Econometrica: Journal of the Econometric Society, 59(3), 817-858. doi.org

Yi, Eojin & Yang, Biao & Jeong, Minhyuk & Sohn, Sungbin & Ahn, Kwangwon. (2023). "*Market efficiency of cryptocurrency: evidence from the Bitcoin market.*" Scientific Reports. 13. 10.1038/s41598-023-31618-4.

Chu, Jeffrey, Yuanyuan Zhang, and Stephen Chan. (2019). *"The adaptive market hypothesis in the high frequency cryptocurrency market."* International Review of Financial Analysis 64 (2019). 10.1016/J.IRFA.2019.05.008

Khursheed, Ambreen, Muhammad Naeem, Sheraz Ahmed, and Faisal Mustafa. (2020). *"Adaptive market hypothesis: An empirical analysis of time -varying market efficiency of cryptocurrencies."* Cogent economics & finance 8.

Grossman, Sanford J., and Joseph E. Stiglitz. (1980). *"On the Impossibility of Informationally Efficient Markets."* The American Economic Review 70, no. 3: 393-408. http://www.jstor.org/stable/1805228.

Robert Hudson & Andrew Urquhart. (2021). *"Technical trading and cryptocurrencies".* Annals of Operations Research, Springer, vol. 297(1), pages 191-220, February.

APPENDIX A

Table A1. Lo-MacKinlay Variance Ratio Statistics
(Daily Realized Volatility)

Asset	Regime	VR(2)	VR(4)	VR(8)	VR(16)	VR(32)
BTC	Pre-ETF	0.962	0.967	0.996	0.996	0.986
BTC	BTC-ETF	0.904	0.906	0.969	0.906	1.052
BTC	Full-ETF	0.969	0.946	0.870	0.843	0.896
ETH	Pre-ETF	0.940	0.930	0.935	0.904	0.883
ETH	BTC-ETF	0.954	0.946	1.057	1.054	1.008
ETH	Full-ETF	0.989	0.988	0.949	0.945	1.071
ADA	Pre-ETF	0.938	0.936	0.940	1.029	1.151
ADA	BTC-ETF	0.942	0.960	0.917	0.749	0.851
ADA	Full-ETF	0.980	0.988	0.954	1.034	1.132
SOL	Pre-ETF	0.978	0.982	1.082	1.208	1.503**
SOL	BTC-ETF	0.894	0.876	0.911	0.855	0.912
SOL	Full-ETF	0.968	0.910	0.842	0.733	0.806
XRP	Pre-ETF	0.944	0.877	0.892	0.891	0.920
XRP	BTC-ETF	0.911	0.840	0.828	0.546	0.439
XRP	Full-ETF	1.047	1.067	1.194	1.457*	1.770**

*Note: ** denotes $p < 0.05$, * denotes $p < 0.10$.*

Table A2. Chow-Denning Multiple Variance Ratio Test Results

Asset	Regime	CD_1	CD_2	Critical Value at 90%	Critical Value at 95%
BTC	Pre-ETF	1.255	1.001	2.311	2.569
BTC	BTC-ETF	1.343	1.037	2.311	2.569
BTC	Full-ETF	1.012	0.862	2.311	2.569
ETH	Pre-ETF	1.999	1.393	2.311	2.569
ETH	BTC-ETF	0.642	0.504	2.311	2.569
ETH	Full-ETF	0.395	0.344	2.311	2.569
ADA	Pre-ETF	2.043	1.340	2.311	2.569
ADA	BTC-ETF	0.805	0.726	2.311	2.569
ADA	Full-ETF	0.476	0.360	2.311	2.569
SOL	Pre-ETF	2.617	2.036	2.311	2.569
SOL	BTC-ETF	1.479	1.170	2.311	2.569
SOL	Full-ETF	1.391	1.184	2.311	2.569
XRP	Pre-ETF	2.189	1.291	2.311	2.569
XRP	BTC-ETF	1.439	1.201	2.311	2.569
XRP	Full-ETF	2.770	2.131	2.311	2.569

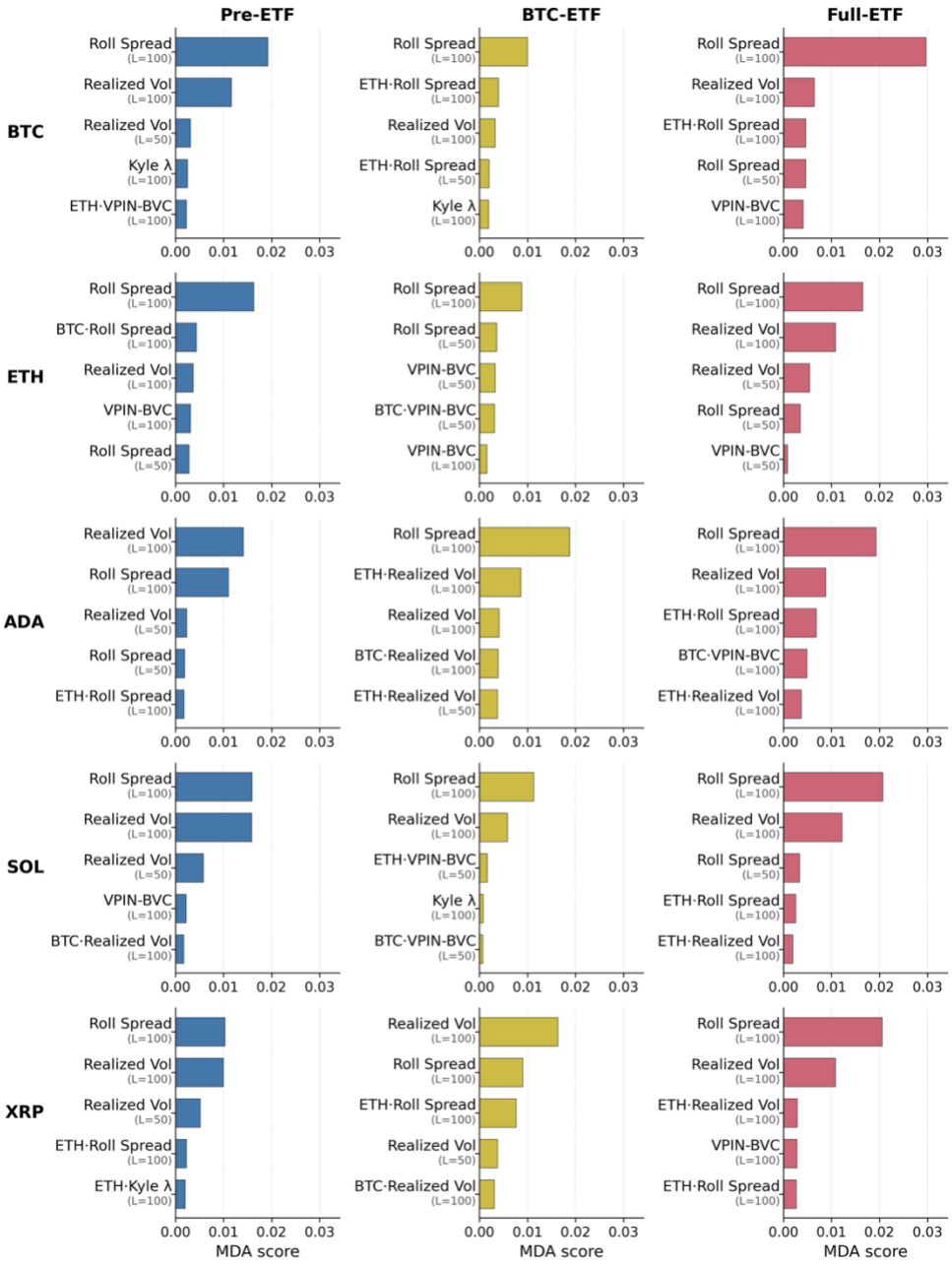


Figure A1. Feature importance chart for regularized Random Forest model