MDSAA

Master Degree Program in

Data Science and Advanced Analytics

Cross-Market Influences on Cryptocurrencies: A Multivariate Markov Chain Perspective

Afonso Reyna Lopes Ferreira

Master Thesis

presented as partial requirement for obtaining a Master's Degree in Data Science and Advanced Analytics

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

Cross-Market Influences on Cryptocurrencies: A Multivariate Markov Chain Perspective

by

Afonso Reyna Lopes Ferreira

Master Thesis presented as partial requirement for obtaining the Master’s degree in Data Science and Advanced Analytics, with a specialization in Data Science

Supervised by

Bruno Damásio, PhD, NOVA IMS

July, 2024

Statement of Integrity

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Lisbon, July 2024

Abstract

The predictability of the cryptocurrency market is a current and highly researched topic. However, the cross-market influences on the returns of the cryptocurrencies would benefit from different perspectives. This study examines the dynamic interplay between Bitcoin, Ethereum, and traditional financial indices spanning from 2017 and 2023 using the Mixture Transition Distribution (MTD) and Generalized Multivariate Markov Chain (GMMC) models. The findings indicate the presence of autocorrelation within each cryptocurrency, as well as an influence of a diverse set of financial indices. Evidence suggest different sector indices affect the cryptocurrencies with distinct degrees of influence. Additionally, the results demonstrate cross-currency influence, notably Ethereum's influence on Bitcoin, suggesting a more interconnected market.

Keywords

Cryptocurrency; Markov Chains; Mixed Transition Distribution; Ethereum, Bitcoin

Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

BTC: Bitcoin

ETH: Ethereum

VIX: CBOE Volatility Index

PCR: Total Put/Call Ratio

SP500: S&P 500 Index

FTSE: Financial Times Stock Exchange 100 Index

OIL: Crude Oil Prices

EPU: Economic Policy Uncertainty

VGT: Vanguard Information Technology ETF

VFH: Vanguard Financials ETF

VHT: Vanguard Health Care ETF

XLP: Consumer Staples Select Sector SPDR Fund

VDE: Vanguard Energy ETF

XLI: Industrial Select Sector SPDR Fund

VPU: Vanguard Utilities ETF

MTD: Mixture Transition Distribution Model

GMMC: Generalized Multivariate Markov Chain

1 Introduction

Cryptocurrencies have experienced an extraordinary surge in interest and investment over the past decade. These digital assets have transformed from niche technological curiosities into key components of the global financial system. This rapid rise in popularity, coupled with the peculiarities of these assets, calls for a similarly rapid expansion in research to understand their unique characteristics and implications from different scientific perspectives.

One of the persistent challenges in the field of cryptocurrency is the unpredictability of the markets, which remains a hurdle for any time-series analysis with the intent on forecasting. The volatility of cryptocurrencies was already considerable until 2020, but it was particularly evident after the COVID-19 outbreak. On the 13th of March 2020, the price of Bitcoin was recorded at $4970, and by 13th of March 2021, it had risen to $61.243, representing a 1232% increase. However, just 2 months later, this price subsequently declined to $34.616 in 29th of May 2021 (CoinMarketCap, 2024).

Despite the perceived volatility and speculative nature of cryptocurrencies, research has identified patterns and factors that can assist in forecasting return movements. The application of advanced data analytics and machine learning has contributed to the development of a growing body of literature with the objective of predicting the movements of cryptocurrencies with greater accuracy. These developments suggest that, although the market may display erratic return patterns on the surface, there are underlying regularities that can be leveraged to many ends. Findings link Bitcoin's returns to a multitude of factors, including market sentiment, social media, regulatory developments, technological advancements, and macroeconomic trends (Benhamed et al., 2023).

This paper seeks to contribute to the understanding of cryptocurrency dynamics and predictability by analysing the impact of various financial indices on the daily returns of Bitcoin and Ethereum with the perspectives offered by Markov chains. Markov chains are powerful tools for modelling the probabilistic nature of financial markets, with an extensive repertoire in the literature. By capturing the stochastic behaviour of the transitions between states of the daily returns, this methodology allows for the measurement of the strength of autocorrelation, cross-currency influence and, in particular, the influence that parallel market indicators exert over cryptocurrencies. Despite the capacities of this framework, there still remain few applications in the literature in this context. Therefore, the approach with the adopted models offers an innovative perspective on this topic.

The research questions adressed in this study can be summarized as follows:

- What are the strongest cross-market influences on cryptocurrency returns?
- Is there evidence of cross-currency influence and autocorrelation in cryptocurrency returns?
- Are Markov chains frameworks adequate for this modeling context?

The following section will present a comprehensive review of the existing literature on the subject, starting with an analysis of the historical development, characteristics and growth of cryptocurrencies. This will be followed by an overview of previous research on the application of Markov chains to financial data, as well as an examination of the various modelling approaches employed in the analysis of cryptocurrency returns. The methodology adopted in this study will then be described in detail, accompanied by an account of the data and its sources. Finally, the results section will present the estimation results and a factor analysis, which aims to distil the information into a more concise representation. The findings have the potential to inform investment strategies, improve forecasting models and guide regulatory frameworks.

2 Literature Review

2.1 Inception of the Cryptocurrency

The first cryptocurrency was introduced by the pseudonymous Satoshi Nakamoto (2008) in the paper “Bitcoin: A Peer-to-Peer Electronic Cash System”. Bitcoin is a decentralised peer-to-peer (P2P) electronic cash system designed to facilitate direct online payments between parties without relying on a trusted intermediary. The system employs cryptographic proof to secure transactions, thereby eliminating the necessity for third-party trust.

The peer-to-peer network ensures all participants have simultaneous access to a continuously updated and immutable ledger of transactions. Each transaction, or “block,” is linked to the previous one, forming a continuous “chain” of records. This structure, from which the term “Blockchain” originates, guarantees data integrity and transparency, as each block is cryptographically secured and validated by network consensus mechanisms, such as proof of work or proof of stake.

The fast appreciation of Bitcoin prompted the cryptocurrency ecosystem to evolve, giving rise to a diverse range of alternatives. One such alternative is Ethereum, created by programmer Vitalik Buterin in 2013 and launched in 2015. Ethereum innovated by introducing smart contracts, which enabled a plethora of decentralized applications and expanded the functionality and potential of blockchain technology. Ethereum is currently the second leading cryptocurrency in terms of market capitalization after Bitcoin. (CoinMarketCap, 2024)

The primary mode of cryptocurrency trading is conducted on centralized exchanges (CEX), such as Binance, Coinbase, and Kraken. These platforms facilitate the matching of buyers and sellers within their respective platforms. In a CEX, users deposit funds into the exchange’s wallets, where the exchange handles order matching and transaction processing. This provides high liquidity and a user-friendly interface. While this offers convenience, it also introduces security risks and requires users to trust the exchange with their funds.

In contrast, decentralised exchanges (DEX), such as Uniswap and Sushiswap, facilitate peer-to-peer trading directly between users without a central authority. This enhances security and privacy through the use of smart contracts, which automatically execute trades when predefined conditions are met, eliminating intermediaries. However, DEX often results in lower liquidity and greater complexity, as users must manage their wallets and understand blockchain interactions directly. Transactions on DEX may be slower and may result in higher fees due to network congestion and gas prices (small amounts of cryptocurrency used to compensate for the computational energy required to process and validate transactions) (Hägele, 2024).

Despite the current surge in popularity, the complexity of the underlying technology and the fundamental differences from traditional currencies continue to present challenges for the general population and policymakers alike in fully comprehending the nuances and risks involved with these digital assets.

Furthermore, the decentralized and uncontrolled nature of blockchain technology poses significant governance challenges, as the lack of central authority complicates the process of making policy changes and settling disputes. This structure exacerbates issues related to tax enforcement and money laundering, leading to increased uncertainty and unpredictability in the cryptocurrency market (Makarov & Schoar, 2022).

2.2 Markov Chains in the Literature

Markov chains are employed in a multitude of domains within the literature. Their utility stems from their capacity to model systems that undergo transitions from one state to another in a probabilistic manner, rendering them optimal in capturing and forecasting random processes. Their ”memorylessness”, wherein the subsequent state hinges solely on the current state and disregards the sequence of events preceding it, simplifies the analysis of complex systems. (Ching et al., 2013)

In physics, Albert and Barabási (2002) use Markov chains to model and analyze the behavior of networks in statistical mechanics. In Biology, Huelsenbeck and Ronquist (2001) perform Bayesian inference of phylogeny using a variant of Markov chain Monte Carlo algorithm. In economic history, Damásio and Mendonça (2019) highlight the advantages of Markov chains when compared to Vector Autoregressive (VAR) approaches to model insurgent/incumbent dynamics in periods of technological evolution. The authors explore the application of Markov framework on these dynamics further in Damásio and Mendonça (2023). In Engineering, Au and Beck (2001) model the probability of transitions between different states in the context of reliability analysis of repairable systems, which allows for the evaluation of system performance and the prediction of failure events over time. Ching et al. (2013) detail the advances in Markov chain modelling and their applications in queueing systems, the Internet, remanufacturing systems, inventory management, DNA sequence analysis, genetic networks, and various further practical systems. Damásio and Nicolau (2014) propose the usage of Multivariate Markov Chains (MMC) as covariates in a regression model.

Markov chains have also been extensively used in financial modeling to capture the stochastic nature of various financial phenomena. Their ability to model state transitions over time makes them particularly useful for understanding asset returns, risk management and market dynamics. Studies such as Hamilton (1989) and Jarrow et al. (1997) have demonstrated the effectiveness of Markov switching models and credit risk models, respectively. Additionally, Ang and Bekaert (2002) applied regime-switching models to international equity returns. Nicolau and Riedlinger (2015) enhance the existing Multivariate Markov chains methodologies by offering a more efficient and less restrictive estimation approach, with demonstrated practical applications in prediction of stocks or investment recommendations. Damásio et al. (2018) utilize Markov chains to model the transitions between different states of GDP growth rates to estimate the expected time for the economy to recover from recessions to higher growth levels. Damásio and Nicolau (2024) introduce a novel approach for identifying and examining multiple structural breaks in a Markov chain, where the timing of these breaks is not predetermined.

2.3 Modeling the Cryptocurrency Market

The field of cryptocurrency prediction has attracted considerable interest, with a substantial number of initiatives being undertaken at the vanguard of this area. The variables selected for this thesis were significantly informed by Panagiotidis et al. (2024), who employ the Bayesian LASSO model to investigate the effects of various economic, financial, and technological factors, as well as indicators of uncertainty and public attention on Bitcoin returns. The results indicate that sentiment and technological variables significantly impact Bitcoin returns, whereas among economic and financial factors, stock market returns and volatility indices are the most influential.

Khan et al. (2023) utilize several machine learning models to predict the return volatility of four cryptocurrencies. The models include neural network autoregressive (NNETAR), cubic smoothing spline (CSS), and group method of data handling neural network (GMDH-NN) algorithm. The authors demonstrate the effectiveness of the frameworks, concluding that no single model performs uniformly across all cryptocurrencies.

Papadimitriou et al. (2022) employ a combination of GARCH and SVM methodologies to forecast steep fluctuations in Bitcoin returns, achieving a high level of predictive accuracy. The authors provide evidence that the returns of alternative cryptocurrencies offer significant insights into Bitcoin return spikes. This finding suggests that cryptocurrency markets are not isolated; rather, they are increasingly interconnected, with information in one cryptocurrency market spreading to others. Despite testing a wide range of variables, the authors highlight that none play a significant role in forecasting Bitcoin returns and their spikes.

Berger and Koubová (2024) compare the effectiveness of econometric time series models versus machine learning techniques in forecasting Bitcoin returns. The study demonstrates the superiority of machine learning approaches over traditional econometric models for Bitcoin price forecasting, highlighting that both ARMA-GARCH and machine learning models outperform naive benchmarks, with machine learning models showing the highest precision.

Although the application of Markov chains is still relatively limited in the literature, several studies have already demonstrated their utility in modelling and forecasting cryptocurrency returns. Shaikh (2020) investigate the relationship between economic policy uncertainty (EPU) and Bitcoin returns using various estimation techniques, including quantile regression and Markov regime-switching models. The primary findings suggest that EPU impacts Bitcoin returns, albeit differently depending on the volatility regime. High volatility regimes show a negative impact of EPU on Bitcoin returns, while in low volatility regimes, the impact varies by geographical regions and specific economic indicators.

Koki et al. (2022) apply Bayesian Hidden Markov Models (HMM), particularly the Non-Homogeneous Hidden Markov Model (NHHM) with four states, to forecast returns of Bitcoin, Ether, and Ripple. The study demonstrates that HMMs significantly outperform traditional models by capturing regime-switching dynamics, identifying distinct market behaviors (bull, bear, calm states), and integrating various financial and cryptocurrency-specific predictors. This approach enhances predictive accuracy and provides valuable insights for portfolio management, risk diversification, and policy-making in volatile cryptocurrency markets, highlighting the utility of Markov chains in modeling time-varying volatility and structural breaks in financial time series.

Araújo and Barbosa (2023) endeavor to reconstruct cryptocurrency processes using Markov chains, in order to forecast the dynamics of intra-day returns. The researchers manage to obtain predictions that surpass the accuracy of a random choice process.

Pennoni et al. (2022) utilize a multivariate Hidden Markov Model to analyze the returns of five major cryptocurrencies from 2017 to 2021, identifying six regimes that capture dynamic market changes, including phases of negative returns and high volatility, as well as periods of significant price increases. The model reveals an increasing market correlation with Bitcoin, reflecting the market's maturation and stronger USD association due to stablecoins, and suggests further improvements using different conditional distributions and regime-switching copula models.

Nascimento et al. (2023) explore the use of high-order Markov chain models to predict future returns of several cryptocurrencies using data between 2016-2019. The research emphasizes the computational challenges related to high-order Markov chains and suggests further exploration into different categorization methods and forecasting horizons. The results highlight that cryptocurrencies exhibit long-range memory.

In summary, while the existing literature has extensively explored various aspects of cryptocurrency dynamics and return prediction, significant gaps remain. Notably, there is a limited application of Multivariate Markov Chains (MMC) to model the interactions between cryptocurrency returns and financial indices. Existing research has primarily employed econometric models and machine learning techniques, which, although effective, lack the ability of Markov chains to capture dependencies and influences with a probabilistic perspective. Financial markets frequently exhibit interdependencies among different assets. Multivariate Markov chains are well-suited to capture these correlations and model the joint dynamics of multiple time series simultaneously. However, existing Markov chain applications lack a more extended time frame consideration and the inclusion of financial variables as covariates. In the context of this study, the resulting estimates of probability transition matrices are highly interpretable, offering a clear map of state transition of returns when under different market conditions. The findings may prove useful in informing investment strategies and policy frameworks by providing insight into how cryptocurrencies react to different values of financial indices.

3 Data & Methods

This section addresses the data and the methods that are a basis for the analysis. The variables and the procedures employed to make them suitable will be delineated, along with the theoretical framework of Markov chains and how they can be leveraged as a powerful tool to model the relationships sought to be understood.

3.1 Data

3.1.1 Variables

The information is extracted from datasets obtained from various well regarded financial data portals: Barchart.com, the Chicago Board Options Exchange (CBOE.com), Investing.com, Yahoo Finance (finance.yahoo.com) and the Federal Reserve Bank of St.Louis (fred.stlouisfed.org). The variables are of daily granularity and vary across different time spans, sharing a common interval relevant to the study that extends from the start of 2017 to the end of 2023, a window that captures the rise and boom of cryptocurrencies. The main variables of study are the returns of Bitcoin and Ethereum, the two most popular cryptocurrencies, which represent approximately 70% of the market share. In their raw form, these two variables consist of the last recorded price in the day, so the daily log-returns were computed for each cryptocurrency, as follows:

$$r_{i,t} = \ln\left(\frac{P_{i,t}}{P_{i,t-1}}\right), \text{ where } P_{i,t} \text{ is the closing price.} \tag{1}$$

To track the shifts in market sentiment and the broader economic conditions that happen in parallel to the cryptocurrency markets, several financial indices were also taken into account. The choice of the variables is substantiated in the literature (see Panagiotidis et al., 2024 and Koki et al., 2022), with aim to capture the most impactful documented drivers of the cryptocurrency returns. The evidence points to a significant importance of volatility and market sentiment indices, as well as stock indices. Furthermore, there is also basis to employ exchange rates, interest rates and commodity prices (such as gold and oil). Many of these variables were also transformed to log-returns. This decision was to ensure the time series were stationary. Although it is not absolutely necessary when handling Markov chain models, having variables with a linear trend or that do not return to past values can indeed be problematic. This was also to capture the effect of different degrees of change and not the effect of the value of the variable itself.

The cryptocurrency market operates continuously, unlike the stock market, which closes on weekends. The consequence is that all the data that derives from indexes/stocks will be missing those days. Initially, the idea of imputing data was enticing, but the high volume of missing data quickly led to the decision against pursuing this approach further. Thus, the decision was taken to exclude weekend data from the analysis. The following table details the variables and their sources.

Table 1: Summary of Variables and Sources

Variable Name	Detail
BTC (returns)	Bitcoin-USD price https://www.barchart.com/crypto/quotes/%5EBTCUSD
ETH (returns)	Ethereum-USD price https://www.barchart.com/crypto/quotes/^ETHUSD
VIX	CBOE Volatility Index https://www.cboe.com/tradable_products/vix/vix_historical_data/
PCR	Total Put/Call Ratio https://www.barchart.com/stocks/quotes/$CPC
SPREAD	Interest Rate Spread https://fred.stlouisfed.org/series/T10Y3M
SP500 (returns)	S&P500 index price https://fred.stlouisfed.org/series/SP500
FTSE (returns)	FTSE 100 Index price https://www.investing.com/indices/uk-100-historical-data
OIL (returns)	Crude Oil WTI May '24 Futures https://www.barchart.com/futures/quotes/CLK24
EPU	Economic Policy Uncertainty Index for United States https://fred.stlouisfed.org/series/USEPUINDXD
VGT (returns)	Vanguard Information Technology Index Fund https://finance.yahoo.com/quote/VGT
VFH (returns)	Vanguard Financials Index Fund https://finance.yahoo.com/quote/VFH
VHT (returns)	Vanguard Health Care Index Fund https://finance.yahoo.com/quote/VHT
XLP (returns)	Consumer Staples Select Sector SPDR Fund https://finance.yahoo.com/quote/XLP
VDE (returns)	Vanguard Energy Index Fund https://finance.yahoo.com/quote/VDE
XLI (returns)	Industrial Select Sector SPDR Fund https://finance.yahoo.com/quote/XLI
VPU (returns)	Vanguard Utilities Index Fund https://finance.yahoo.com/quote/VPU

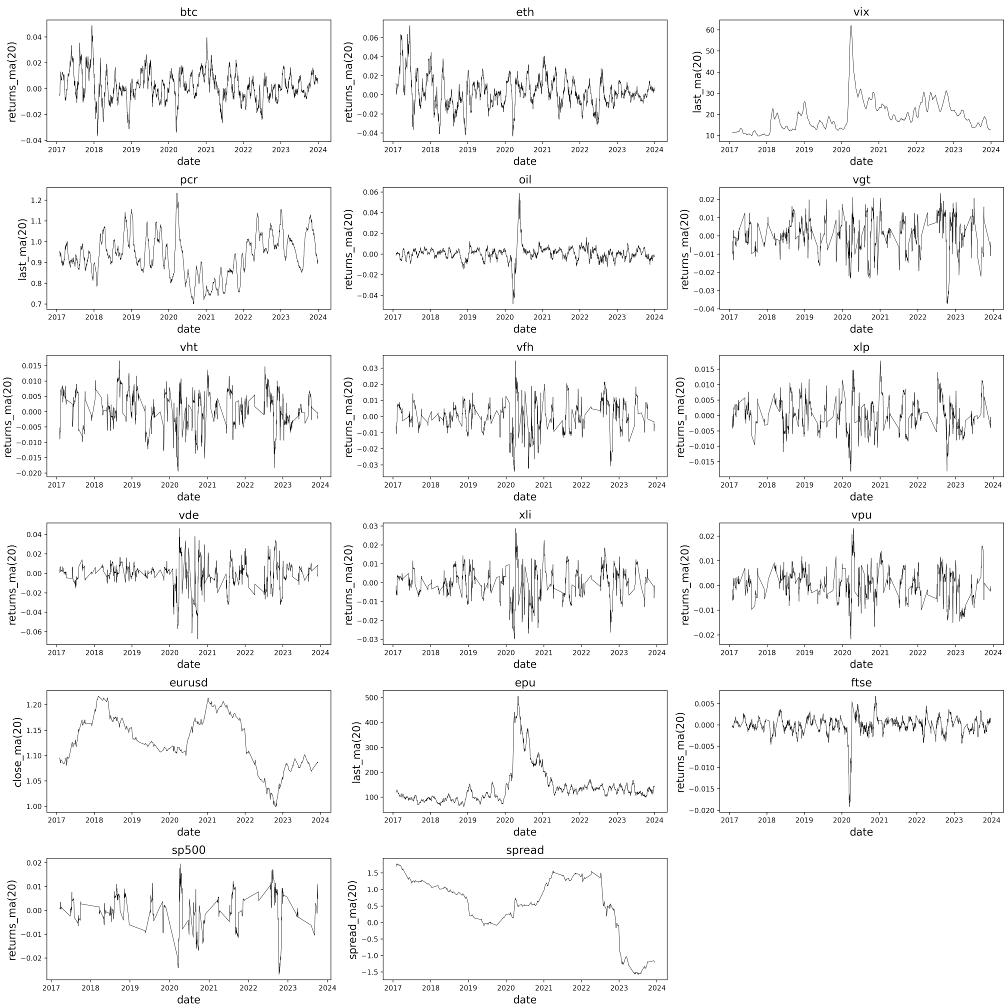

Figure 1: 20-day rolling averages of collected variables (2017-2024)

Table 2: Summary Statistics

Variable	Count	Mean	Std	Min	Q1	Median	Q3	Max
BTC returns	1824	0.002	0.043	-0.311	-0.016	0.002	0.021	0.218
ETH returns	1824	0.003	0.058	-0.401	-0.021	0.002	0.029	0.402
VIX	1824	19.210	8.083	9.140	13.393	17.585	22.790	82.690
PCR	1824	0.932	0.149	0.600	0.830	0.920	1.011	1.910
OIL returns	1822	0.000	0.030	-0.282	-0.011	0.002	0.014	0.320
VGT returns	1744	0.000	0.065	-0.436	-0.010	0.002	0.019	0.284
VHT returns	1744	0.000	0.036	-0.265	-0.008	0.001	0.011	0.187
VFH returns	1744	0.000	0.054	-0.431	-0.011	0.001	0.015	0.395
XLP returns	1744	0.000	0.032	-0.208	-0.007	0.001	0.009	0.149
VDE returns	1744	-0.001	0.083	-0.644	-0.017	0.000	0.016	0.656
XLI returns	1744	0.000	0.049	-0.382	-0.009	0.001	0.013	0.329
VPU returns	1744	0.000	0.039	-0.245	-0.009	0.002	0.011	0.206
EUR/USD	1761	1.127	0.058	0.960	1.087	1.127	1.176	1.251
EPU	1824	140.676	100.677	4.700	78.543	111.775	165.420	807.660
FTSE returns	1794	0.000	0.010	-0.115	-0.004	0.001	0.005	0.087
SP500 returns	1641	0.000	0.043	-0.309	-0.007	0.001	0.014	0.244
SPREAD	1732	0.543	0.971	-1.890	0.057	0.760	1.260	2.200

3.1.2 Parsing and Transformation

In order to parse the data and handle any missing values, all variables' time series were aligned to all days within the specified period. This guarantees that each day is represented in the time series, incorporating missing values in time series that omitted days. Subsequently, the weekends were excluded, and the remaining missing values were substituted with the mean of the preceding and subsequent day.

The selection of the mean of the previous and next day as an imputation method is underpinned by several methodological and practical considerations, which make it particularly suitable for maintaining data integrity and ensuring a robust analysis in this context. The primary justification is that this method leverages the local context, which is crucial in time series data where values tend to exhibit strong temporal dependencies. The reliance on adjacent observations for imputation inherently respects the local data structure, thereby reducing the risk of introducing systemic bias that could distort underlying patterns. Furthermore, for time series data with short intervals of missing data such as this case, this method works well because the values on either side of the gap are likely to be similar. In such cases, the value in the middle of ten represents a transition between these two values, making the mean a reasonable estimate.

In very specific cases where there are consecutive missing values, these are imputed with the mean of the distribution of the variable.

Markov chain models can only be applied when the variables are categorical. To make the transformation from continuous to categorical, a function was created to apply binning to the variables, splitting the observations into m groups (states) according to their position in the distribution.

For an arbitrary m, for each segment $s = \{1, ..., m\}$, the corresponding percentile (p) interval is given by $\{\frac{100(s-1)}{m}, \frac{100s}{m}\}$.

As an example, let j be a variable to be transformed. If we apply $m = 5$, the process is described as follows:

$$S_t^{(j)} = \begin{cases} 1 & \text{if } obs_{jt} \leq q_{j,20} \\ 2 & \text{if } q_{j,20} < obs_{jt} \leq q_{j,40} \\ 3 & \text{if } q_{j,40} < obs_{jt} \leq q_{j,60} \\ 4 & \text{if } q_{j,60} < obs_{jt} \leq q_{j,80} \\ 5 & \text{if } obs_{jt} > q_{j,80} \end{cases} \tag{2}$$

where $q_{j,p}$ represents the p-th percentile of the marginal distribution of the variable j.

The end decision was to apply m = 5 for all variables, a common number of states in the literature dedicated to applying Markov chains to financial assets. This number of states provides a median value while including two extreme values and two mid-range values, offering a reasonable balance of complexity and interpretability. This approach is not overly demanding on the computation of the applied models and helps avoid estimation errors that could arise from using a higher number of states with a small sample size such as the one used in this study.

To illustrate the impact of this approach, the resulting variables are displayed in the following table of graphs. These graphs provide a visual representation of how each variable is distributed across the five states throughout the years of analysis.

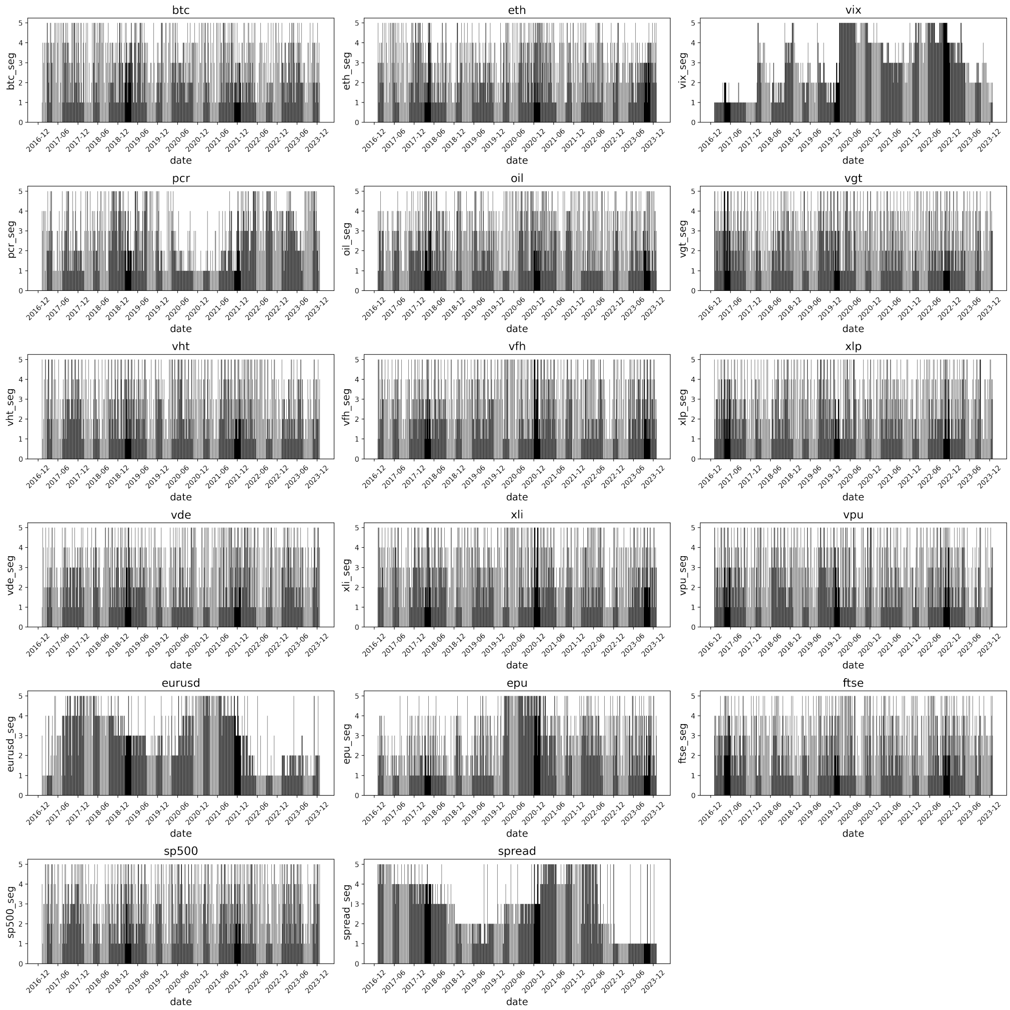

Figure 2: Plots of Segmented Variables

3.2 Markov Chains

Markov chains are a mathematical concept used in statistics to model the transition probabilities between states within a countable space, denoted as $M = \{1, 2, \dots, m\}$. This framework is utilized to describe the behavior of a stochastic process, represented as $S_t, S_{t-1}, \dots, S_0$, which transitions between these states over time.

The key assumption of a Markov chain is that the current state depends only on the state that precedes it, such that:

$$P(S_t = k_0|\mathcal{F}_{t-1}) = P(S_t = k_0|S_{t-1} = k_1) \tag{3}$$

where $\mathcal{F}_{t-1}$ is the σ-algebra generated by the available information until time $t-1$.

A multivariate stochastic process $\{S_{1t}, ..., S_{st}; t = 0, 1, 2, ...\}$ is said to be a Multivariate Markov Chain (MMC) process if it can be defined as:

$$P(S_{jt} = k|\mathcal{F}_{t-1}) = P(S_{jt} = k|S_{1t-1} = i_1, ..., S_{st-1} = i_s) \tag{4}$$

In multivariate Markov chains, the transition probabilities are the main focus of analysis and can be estimated through maximum likelihood estimation. The MLE is given by:

$$\hat{P}^{(jk)}_{i_j i_k} = \frac{n^{(jk)}_{i_j i_k}}{\sum\limits_{i_k=1}^{m} n^{(jk)}_{i_j i_k}} \tag{5}$$

The probability of observing a transition from from state i_j of process j and state i_k of process k is given by the quotient of the total number of transitions observed between these states ($n^{(jk)}_{i_j i_k}$) and the total number of transitions from i_j and every state of the process k.

3.2.1 The Mixture Transition Distribution Model

The Mixture Transition Distribution (MTD) model, proposed by Raftery (1985), is designed for modeling lth-order Markov chains, where the present state is dependent on the l previous states. The key innovation of the MTD model is that it expresses the transition probabilities as a linear combination of the transition probabilities of lower-order Markov chains. This approach significantly reduces the number of parameters from $(m-1)m^l$, typical in conventional lth-order Markov chains, to a more manageable number, making it feasible to estimate higher-order dependencies while maintaining a balance between realism and parsimony.

As an adaptation of Raftery's specification, Ching (2002) suggested a way of simplifying the estimation with an alternative of order of 1 and fewer parameters. The proposed model intends to represent the behaviour of multiple categorical sequences generated by similar sources or the same source, an application very relevant to this study. It is defined as follows:

Let $X_t^{(j)}$ be the state vector of the jth sequence at time t. If the jth sequence is in state j at time t, then it is written:

$$X_t^{(j)} = (0, ..., 0, \underbrace{1}_{j\text{th entry}}, 0, ..., 0)^t \tag{6}$$

the model is given by:

$$X_{t+1}^{(j)} = \sum_{k=1}^{s} \lambda_{jk} P^{(jk)} X_t^{(k)} \text{ for } j = 1, 2, ..., s \tag{7}$$

where $\lambda_{jk} \geq 0$ for $1 \leq j, k \leq s$ and $\sum_{k=1}^{s} \lambda_{jk} = 1$ for $j = 1, 2, ..., s$.

The state probability distribution of the kth sequence at time $(t+1)$ depends on the weighted average of $P^{(jk)} X_t^{(k)}$. Here, $P^{(jk)}$ is a transition probability matrix from the states in the kth sequence to the states in the jth sequence, and $X_t^{(k)}$ is the state probability distribution of the kth sequences at time t.
In matrix form it is written:

$$\boldsymbol{X}_{t+1} \equiv \begin{pmatrix} X_{t+1}^{(1)} \\ X_{t+1}^{(2)} \\ \vdots \\ X_{t+1}^{(s)} \end{pmatrix} = \begin{pmatrix} \lambda_{11} P^{(11)} & \lambda_{12} P^{(12)} & \cdots & \lambda_{1s} P^{(1s)} \\ \lambda_{21} P^{(21)} & \lambda_{22} P^{(22)} & \cdots & \lambda_{2s} P^{(2s)} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{s1} P^{(s1)} & \lambda_{s2} P^{(s2)} & \cdots & \lambda_{ss} P^{(ss)} \end{pmatrix} \begin{pmatrix} X_t^{(1)} \\ X_t^{(2)} \\ \vdots \\ X_t^{(s)} \end{pmatrix} \equiv Q \boldsymbol{X}_t \tag{8}$$

The parameters of interest are the coefficients λ_{jk} and the transition probability matrices $P^{(jk)}$. The lambda estimates provide a measure of impact associated with the transitions between states of j and k processes - a higher λ hinting at a higher impact. While the probability matrices are not subject to optimization and solely represent observed data, they provide valuable insights into the dynamics of state transitions.

The MTD model is used as the primary reference to understand the impact of various variables on the returns of Bitcoin and Ethereum. The methodology applied was to estimate the model using three variables. The first two variables, Bitcoin and Ethereum, remain constant throughout the analysis, providing a measure of auto-correlation and cross-currency influence. The third variable varies with each estimation and is the primary focus of the study, as the lambdas provide a measure of impact associated with the transitions between the states of the the financial indices and the cryptocurrencies.

The estimations were performed using the GenMarkov R package authored by Vasconcelos and Damásio (2022). The probability matrices $P^{(jk)}$ are computed as proposed in Ching (2002) and the MTD model parameters are estimated through Berchtold (2001) optimization algorithm.

3.2.2 The Generalized Multivariate Markov Chain (GMMC)

A different model was employed to understand the directional impact of the various financial indices on the probability transition matrices of Bitcoin and Ethereum returns. This model is the Generalized Multivariate Markov Chain (GMMC) model with covariates, proposed by Vasconcelos and Damásio (2022) as a generalization of the Multivariate Markov Chains (MMC) model by Ching (2002). This specification considers exogenous or pre-determined covariates in the σ-algebra generated by the available information until $\mathcal{F}_{t-1}$. The model is expressed as:

$$P(S_{jt} = k \mid \mathcal{F}_{t-1}) = P(S_{jt} = k \mid S_{1t-1} = i_1, S_{2t-1} = i_2, \ldots, S_{st-1} = i_s, \mathbf{x}_t) \tag{9}$$

It can also be specified as proposed by Ching (2002) using Raftery's notation:

$$\begin{aligned} &P(S_{jt} = i_0 \mid S_{1t-1} = i_1, \ldots, S_{st-1} = i_s, \mathbf{x}_t) = \\ &\quad \lambda_{j1} P(S_{jt} = i_0 \mid S_{1t-1} = i_1, \mathbf{x}_t) + \cdots + \\ &\quad \lambda_{js} P(S_{jt} = i_0 \mid S_{st-1} = i_s, \mathbf{x}_t) \end{aligned} \tag{10}$$

The adopted process with this model involves estimating it with three variables, similar to the MTD approach: Bitcoin and Ethereum, which remain constant across estimations, and different financial variables as the covariate. The primary focus is on the probability transition matrices, which are computed for three distinct values of the covariate: the maximum, the minimum, and the median. This allows to have a directional perspective on the probabilities across different values of the covariate.

An important note is that the median value matrices are merely computed to assess stability, and showed to be identically distributed across the different covariates. Due to this, in the proceeding analysis, the focus is pivoted to the maximum and minimum results. The package GenMarkov in R was also used to compute these estimations.

3.3 Factor Analysis

The GMMC model generated 60 5x5 probability transition matrices with numerous insights and complex interrelationships. To enhance the presentation of the results and manage the impracticality of individually detailing the 60 matrices, an exploratory factor analysis is employed. This statistical method aims to identify underlying relationships between observed variables by reducing them to a smaller set of latent factors that explain their correlations. The goal was to capture the variation of the 25 individual matrix data points of the 60 matrices into representative factors with expressive loading profiles.

The initial step involved flattening the 5x5 matrices into vectors of length 25. Following this, a master dataset was created, incorporating all 60 individual vectors as rows. The resulting dataset has a shape of 60 rows and 25 columns.

Let M_i be the i-th 5x5 matrix, where $i = 1, 2, \ldots, 60$. Each matrix M_i can be represented as:

$$M_i = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{15} \\ p_{21} & p_{22} & \cdots & p_{25} \\ \vdots & \vdots & \ddots & \vdots \\ p_{51} & p_{52} & \cdots & p_{55} \end{bmatrix} \tag{11}$$

Where p_{ij} represents the probability of transitioning from the state i to state j.

Each matrix M_i is flattened into a vector v_i of length 25:

$$v_i = \text{vec}(M_i) = [p_{11}, p_{12}, \cdots, p_{15}, p_{21}, \cdots, p_{25}, \cdots, p_{55}]^\top \tag{12}$$

Let D be the master dataset. This dataset is formed by stacking all the vectors v_i as rows in a matrix. Thus, D will have 60 rows and 25 columns.

$$D = \begin{bmatrix} v_1^\top \\ v_2^\top \\ \vdots \\ v_{60}^\top \end{bmatrix} \tag{13}$$

The results of Bartlett's Test of Sphericity (Chi-Square = 7894.33, p-value = 0.0) indicate that the correlation matrix is significantly different from an identity matrix, meaning that there are significant relationships among the variables. This is a positive indication for the suitability of the data for factor analysis. Nevertheless, the Kaiser-Meyer-Olkin (KMO) measure is comparably low (0.48) to the literature minimum standard of 0.6, suggesting that the data may not be adequately suited for factor analysis.

Regardless, the factor analysis was conducted to summarize, simplify, and illustrate this complex amount of data, rather than aiming for maximum precision. Despite the adequacy measures indicating many inadequacies in performing factor analysis on this data, the results were nonetheless interesting and informative. This justifies its continued consideration within the context of this study.

The rationale behind the decision to retain three factors is informed by a combination of statistical measures and interpretability considerations. Firstly, Figure 3 illustrates that there is a distinct inflection point after the third factor, which suggests that additional factors would contribute only marginally to explaining the variance. Furthermore, the subsequent factors have eigenvalues lower than one, which indicates that they should be disregarded in accordance with Kaiser's criterion. The possibility of retaining two or four factors was also considered for testing purposes. However, the loading profiles of the option with three factors were ultimately found to be the most expressive.

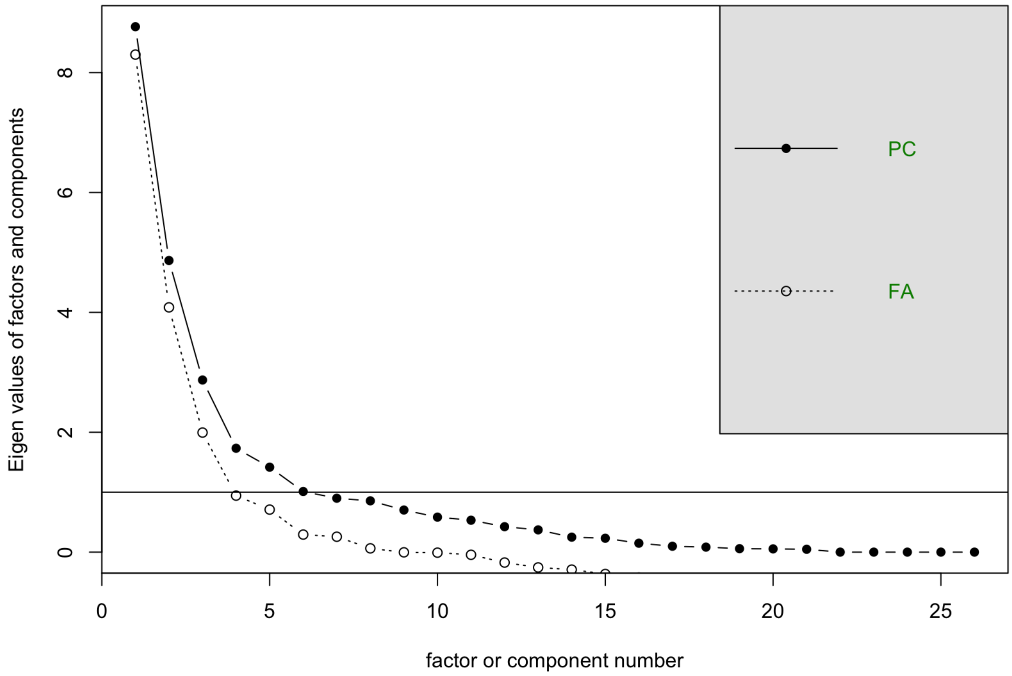

Figure 3: Scree Plot

One key aspect of factor analysis is that the factor axes can be rotated within the multidimensional variable space. The rotations make the pattern of loadings (coefficients that represent the relationship between the observed variables and the underlying latent factors) more evident, helping understand which variables are strongly associated with which factors.

Different rotations were performed in the fitting process, with the goal in mind to maximize a clear segmentation between the distinguishing aspects of the matrices - such as the cryptocurrency, the value of the covariate and the covariates themselves. While rotation techniques such as Varimax and Quartimax are commonly employed to achieve a simpler and more interpretable factor solution, the performed analysis revealed that the unrotated solution provided more meaningful results. Specifically, the factor loadings obtained without rotation provided profiles that better segmented the aspects of the matrices. The loadings will be detailed in the results section.

4 Results and Discussion

This section depicts the estimation results of the models described in the last chapter for the period of 2017-2023.

4.1 MTD Results

Each MTD model estimation with this methodology yields three equations, reflecting the effects of all variables respectively on Bitcoin, Ethereum, and a financial index (which varies with each estimation). Each equation has three lambda estimates, indicating the influence of each variable on the equation's subject. The third equation comprises the effects of the variables on a financial index, which is outside the scope of this study. Therefore, this equation is ignored and the focus is directed solely to the cryptocurrency equations.

4.1.1 Bitcoin equation

The following table contains the estimates for the lambdas of the first equation (pertaining to bitcoin) of the MTD model. These variables are divided into the group of ”Market Indices”, including indices that describe the overall health, direction, and trends of financial markets, and the group of ”Sectorial Indices”, encompassing the various indices that track different sectors of the economy.

Table 3: Estimates for Bitcoin MTD Equation

(a) Market Indices

	λ BTC	λ ETH	λ Index	LL
VIX	0.4537 *** (0.1227)	0.2261 (0.1705)	0.3202 ** (0.142)	-8791.16
OIL	0.4471 *** (0.121)	0.2221 (0.1397)	0.3309 *** (0.0824)	-8791.16
EUR/USD	0.3430 *** (0.1256)	0.3087 ** (0.1338)	0.3483 *** (0.0898)	-8779.06
EPU	0.4457 *** (0.121)	0.2212 (0.1522)	0.333 *** (0.1049)	-8781.99
PCR	0.3364 *** (0.1256)	0.2656 * (0.1437)	0.398 *** (0.1141)	-8784.97
SP500	0.2991 ** (0.1216)	0.1211 (0.1533)	0.5798 *** (0.1116)	-8789.65
FTSE	0.3576 *** (0.136)	0.2784 ** (0.1356)	0.364 *** (0.1277)	-8777.22
SPREAD	0.2508 ** (0.1274)	0.2526 ** (0.1231)	0.4965 *** (0.0693)	-8785.10

(b) Sectorial Indices

	λ BTC	λ ETH	λ Index	LL
VPU	0.3636 *** (0.1289)	0.2517 * (0.1448)	0.3847 *** (0.1254)	-8784.61
VDE	0.3877 *** (0.1213)	0.2359 (0.1651)	0.3764 *** (0.1148)	-8785.66
XLI	0.3803 *** (0.1305)	0.2466 * (0.1364)	0.3731 *** (0.1169)	-8784.92
VFH	0.4248 *** (0.1246)	0.2584 * (0.1556)	0.3168 ** (0.1287)	-8786.11
VHT	0.3754 *** (0.1234)	0.2284 (0.1421)	0.3962 *** (0.1011)	-8786.14
VGT	0.3699 *** (0.1227)	0.2250 (0.1454)	0.4051 *** (0.1006)	-8785.82
XLP	0.3165 ** (0.1328)	0.3535 ** (0.153)	0.3301 ** (0.1484)	-8781.59

estimates are displayed with se's below between parenthesis, LL is the LogLikelihood score

***, ** and * represent the significance levels, respectively 1%, 5% and 10%

The layout of the table is designed to represent in each row an entirely different estimation. Each row corresponds to a model where the stated index is used as the third variable. The columns represent the lambda estimates for Bitcoin (λ_{BTC}), Ethereum (λ_{ETH}), and the index in question (λ_{Index}).

The first insight pertains to the significance of the lambda estimates. All specifications with the different third variable show that the dynamics that govern the bitcoin returns are characterized by a persistent dependence on its own past states, with significance of λBTC at the 5% level and in most of the cases at the 1% level. On the other hand it can be seen that the effect of Ethereum's returns is not as impactful on the current state of bitcoin returns, with many lambda estimates failing to show significance at any of the conventional significance levels. However, it is important to note that also in a lot specifications, λETH does show significance at the 10% and even 5% level, an intriguing result, considering the drastically lower market share of Ethereum compared to Bitcoin. In regards to the lambdas of the financial variables, every single one displays significant lambdas with p-values lower than 5% and in most cases lower than 1%, solidifying the fact that bitcoin's returns are indeed deeply tethered to the parallel economic movements.

It can be seen that the weight of Bitcoin's influence on itself outclasses the remaining influences in the majority of specifications, with notable exception of the returns of S&P500 and the spread of interest rates, which take a significantly higher slice of the total influence ($\lambda SP500 = 0.5798$ vs $\lambda BTC = 0.2991$ and $\lambda SPREAD = 0.4965$ vs $\lambda BTC = 0.2508$). The high influence of the S&P500 returns and interest rate spreads on Bitcoin's returns can be attributed to their role in being some of the more fundamental measures of the economic situation and expectations for the future, which directly impact investment decisions. Some other notably high influences include the Put/Call ratio ($\lambda PCR = 0.398$), the returns of the Technological Sector Index ($\lambda VGT = 0.4051$) and the Health Sector Index ($\lambda VHT = 0.3962$).

The Put/Call ratio (PCR) is another metric that captures market sentiment at a given moment through the ratio of opposite directional bets on the state of the market. High PCR reflects a market with negative expectations and conversely a low PCR reflects positive expectations. This relationship is noteworthy, as the expectation for the close future of the market has an impact on the returns of Bitcoin. The significant influence of past Health Sector Index returns on Bitcoin returns is likely due to the analysis period, which largely coincides with the Covid-19 outbreak. During this time, the health sector experienced substantial growth and volatility due to the increased demand for medical supplies, vaccines, and healthcare services. Given the high stakes of the healthcare sector's performance in society and the economy during this period, it is unsurprising that cryptocurrency returns closely mirrored the sector's fluctuations.

4.1.2 Ethereum equation

Ethereum's equations display very analogous behaviour to Bitcoin's, exhibiting very significant influence of its own past values on itself, as well as relatively uniformly significant effect of the financial variables. The same financial indices that highly influence Bitcoin also show large influence on Ethereum ($\lambda SP500 = 0.5567$, $\lambda SPREAD = 0.4827$), with notable increases in the influence of the exchange rate of EUR/USD, the Financial Sector Index and the Industrial Sector Index ($\lambda EUR/USD = 0.4619$, $\lambda VFH = 0.4827$ and $\lambda XLP = 0.4573$, respectively). However, where they distinguish drastically is in the fact that Bitcoin's past values do not influence Ethereum's current values in any specification. This is an intriguing finding, considering the fact that in Bitcoin's equation, Ethereum's past values manage to stay relevant in some cases.

The market share of Bitcoin represents as of 2024 more than 50%, whereas Ethereum retains closer to 15%. The first intuition is that the steps the giant takes are the ones that influence the surroundings. However, these results

Table 4: Estimates for Ethereum MTD Equation

(a) Market Indices

	λ BTC	λ ETH	λ Index	LL
VIX	0.2332 (0.2446)	0.4469 *** (0.1348)	0.3198 * (0.1921)	-8769.49
OIL	0.1365 (0.1764)	0.5547 *** (0.1363)	0.3088 *** (0.088)	-8791.18
EUR/USD	0.1608 (0.1512)	0.3773 *** (0.1281)	0.4619 *** (0.085)	-8775.76
EPU	0.2480 (0.1673)	0.4816 *** (0.1336)	0.2703 ** (0.1103)	-8780.18
PCR	0.1785 (0.1763)	0.4771 *** (0.1356)	0.3444 *** (0.107)	-8780.02
SP500	0.1026 (0.1863)	0.3407 ** (0.1355)	0.5567 *** (0.1358)	-8783.84
FTSE	0.2346 (0.1745)	0.4501 *** (0.1366)	0.3153 *** (0.1007)	-8788.93
SPREAD	0.1248 (0.1655)	0.3926 *** (0.133)	0.4827 *** (0.0686)	-8781.09

(b) Sectorial Indices

	λ BTC	λ ETH	λ Index	LL
VPU	0.1765 (0.1861)	0.4725 *** (0.1343)	0.351 *** (0.1141)	-8784.60
VDE	0.1936 (0.1777)	0.4542 *** (0.1342)	0.3522 *** (0.1084)	-8783.27
XLI	0.1956 (0.1724)	0.4087 *** (0.1357)	0.3956 *** (0.1233)	-8778.59
VFH	0.1936 (0.1839)	0.3541 ** (0.1391)	0.4523 *** (0.1631)	-8772.77
VHT	0.1966 (0.179)	0.3611 *** (0.1356)	0.4423 *** (0.1449)	-8775.67
VGT	0.1943 (0.168)	0.4057 *** (0.1369)	0.3999 *** (0.1169)	-8779.42
XLP	0.1921 (0.1825)	0.3505 *** (0.136)	0.4573 *** (0.1459)	-8774.42

estimates are displayed with se's below between parenthesis, LL is the LogLikelihood score

***, ** and * represent the significance levels, respectively 1%, 5% and 10%

seem to suggest that Ethereums's market movements exert influence on Bitcoin, but not vice versa. This seems like a wildly far-fetched proposition. The correlation of the price of Bitcoin and Ethereum is of 0.713 in the period considered. But when looking at the distribution of the returns, Ethereum has wider tails with more extreme values. One the one hand, this could be the result of an idiosyncratic interaction between the variables involved, the nature of this model of Markov Chains itself or the binning algorithm chosen to apply to the variables. On the other hand, this may suggest a more complex interaction between these cryptocurrencies and more sophisticated modeling would be necessary to untangle these dynamics and verify the robustness of this finding. However, even if significant, the lambda estimates of Ethereum's effect on Bitcoin retain a considerably low influence in the broad spectrum of the other influential factors.

4.2 GMMC Results

As stated prior, the results of this model are quite extensive and for this reason a factor analysis was performed to summarize the results and illustrate the analysis.

4.2.1 Factor loadings

Linearity- 51.9% var explained

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	-0.21	-0.66	-0.25	0.75	0.67
[2,]	-0.90	0.09	0.09	0.78	0.50
[3,]	-0.91	-0.15	0.21	0.72	0.64
[4,]	-0.48	-0.41	0.06	0.44	0.42
[5,]	-0.91	-0.69	-0.03	0.49	0.82

U-shape - 31.44% var explained

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	-0.36	0.48	0.57	-0.31	-0.55
[2,]	-0.26	0.22	0.81	0.29	-0.73
[3,]	-0.27	0.05	0.62	0.11	-0.33
[4,]	-0.74	0.50	0.60	-0.13	-0.49
[5,]	0.05	0.20	0.48	-0.16	-0.31

Stability- 16.66% var explained

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	0.23	0.24	-0.62	0.06	0.04
[2,]	0.14	-0.43	0.02	-0.11	0.13
[3,]	-0.07	-0.06	0.54	-0.18	-0.18
[4,]	0.19	-0.37	0.70	-0.52	-0.13
[5,]	-0.03	0.01	-0.60	0.40	0.10

Figure 4: Factor Loadings

The resulting factor loadings for each of the 25 variables are displayed in the original matrix format, a perspective that allows to identify the patterns with more clarity.

The first loading profile displays a clear contrast in the correlation of the factor with the transitions to states of low returns and the transitions to states of higher returns. If the transition matrices score high in this factor, the cryptocurrency returns display a higher tendency to transition to a state of higher returns - state [,4] and [,5] - from whichever state they are in the previous day. Conversely, if they score low in this factor, there is a tendency to transition from whichever state to more negative states of returns - [,1] and [,2].

The chosen name for this factor was *Linearity*, as it is the factor that captures the direction of the returns with a clear dichotomy between transitions to negative and positive returns. This factor explains 51.9% of the variance in the transition patterns of cryptocurrency returns, indicating that more than half of the variability in these transitions can be attributed to the linear direction of returns. This significant percentage underscores the importance of the *Linearity* factor in understanding how returns move from one state to another, either towards higher or lower values.

The second loading profile captures the essence of another very usual pattern in PTMs, the U-shape (reference). U-shape transition matrices express volatility, characterized by a higher likelihood of transitioning from any previous state to one of the two extreme states the following day. In contrast, the reversed U-shape expresses a higher likelihood to transition from any previous state to median states in the following day - a centre heaviness and balance. High scores in this factor point to an inverse U-shape, while low scores suggest a U-shape pattern.

This factor explains 31.44% of the variance in the transition patterns of cryptocurrency returns, highlighting its significant role in capturing the volatility and central tendency dynamics within the data. One important note is that this factor displays a high degree of orthogonality with the factor *Linearity*, therefore, when values away from zero of *Linearity* are seen, it can be expected that the values of *U-shape* are closer to zero and vice-versa.

The third factor shows a more asymmetric and peculiar loading profile, with high positive correlations in the transition from the third state into itself, from the fourth state to the third and from the fifth to the fourth. On the other hand, high negative correlations can be seen from the first state to the third, second to itself, forth to second and itself and fifth to the third state.

One important note is that the matrices that score high in this factor are associated with adverse and deteriorated market/economic conditions - as these matrices are influenced by the adverse counterpart of the values assigned to the covariate variables. When the matrices are influenced by the value that expresses a good economic climate, the values of this factor tend to be closer to zero. Moreover, for matrices with values away from zero, this factor creates a clear segregation of the data pertaining to Bitcoin and Ethereum. Therefore, it seems to capture a certain idiosyncratic behaviour relating to the stability of the centre transitions of the two currencies. Due to this, this factor was named *Stability*.

While *Stability* only represents 16.66% of the explained variance, as the following section will address, this factor manages to be quite interesting due to how well it segments the two cryptocurrencies.

4.3 Analysis of Factor Scores of Matrices

The following plot illustrates the distribution of factor scores of the probability transition matrices along the axes of *Linearity* (x axis) and *Stability* (y axis). The data is segmented along 4 data series:

- Ethmin: the matrices of Ethereum's transitions affected by the minimum value of the respective covariate. Represented as the points in a darker shade of grey.
- Ethmax: the matrices of Ethereum's transitions affected by the maximum

value of the respective covariate. Represented as the points in a lighter shade of grey.

- Btcmin: the matrices of Bitcoin's transitions affected by the minimum value of the respective covariate. Represented as the points in a darker shade of orange.

- Btcmax: the matrices of Bitcoin's transitions affected by the maximum value of the respective covariate. Represented as the points in a lighter shade of orange.

The respective minimum and maximum matrices of the same cryptocurrency are linked by a dotted grey line, pointing at the minimum value matrix. The covariates relating to each matrix can be seen written above its respective minimum value point.

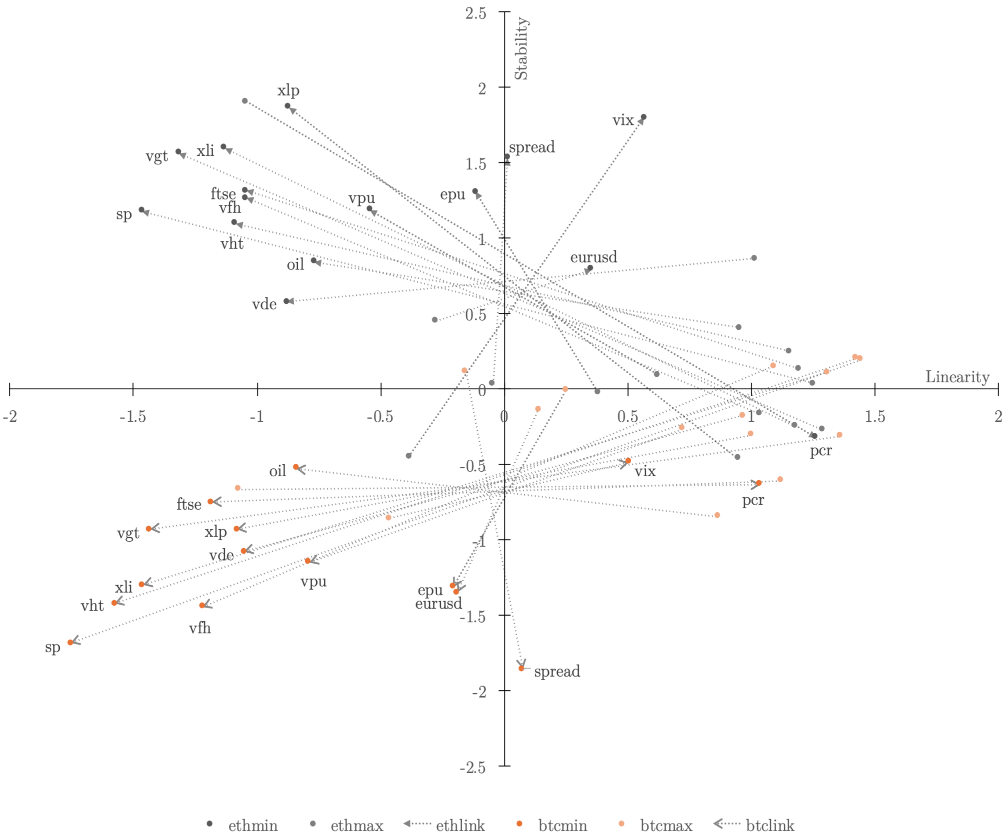

Figure 5: Factor Analysis of Bitcoin and Ethereum Return Transitions Influenced by Financial Indices

Firstly, we can observe in the limits of the first and fourth quadrant a big cluster of matrices affected by the maximum value of the covariates. This cluster does not distinguish between Bitcoin and Ethereum and it is characterized by a positive score in $Linearity$ and a score of $Stability$ close to zero. In the context of market behavior, when the economy demonstrates strength through positive economic indices, the tendency is for both cryptocurrencies to transition to states of higher returns. In other words, when the economic indices show positive values or returns, the cryptocurrencies tend to follow suit.

On the opposite end of the spectrum, in the second and third quadrant, we can see that as the covariates shift towards adverse values, the $Linearity$ scores pivot towards negative values for the majority of the variables. Thus, when the economic and market conditions deteriorate, both cryptocurrencies' returns tend to steer towards the negative states. It's in this context that the factor of $Stability$ shows values further away from zero, with Bitcoin exhibiting negative values and Ethereum exhibiting positive values. Therefore, during economic distress, both cryptocurrencies have their idiosyncratic transition nuances, with Bitcoin manifesting a less persistent and more volatile nature and Ethereum displaying more likelihood to persist in the median transitions.

Some outliers in these plots are the matrices influenced by EPU, SPREAD, VIX, PCR and EUR/USD, as they display different behaviour from the matrices in these two large clusters. First and foremost, in regards to VIX and PCR, the nature of these variables is fundamentaly different from the remaining variables, as they express volatility, uncertainty and instability with high values and economic robustness with low values. Due to this innate inverse relationship, its natural for them to resemble an opposite behaviour relating to the matrices in the clusters in the chart. If we were to inverse the coordinates of $Linearity$, they clearly fit the two main clusters and display the same behaviour.

SPREAD and EUR/USD are the only high scorers in the *U-shape* factor, depicting high market volatility at high values and stability at low values. Due to the aforementioned orthogonality of $Linearity$ and *U-shape*, its only natural for these matrices to be the 2 lowest scoring in $Linearity$.

The following figure depicts the absolute difference of the factor $Linearity$ between matrices affected by the same Financial Index, with segmentation by cryptocurrency. It allows to understand which Indices have a more accentuated impact in the linearity of the matrices between moments of economic prosperity and economic difficulties. The indices that score high in this graph affect the matrices in a more radical way between economic states, with very high probabilities to transition to lower states of returns in a bad economic climate and a very high probability to tranistion to the highest states of returns in an optimistic economic climate.

As we can see, the effects of S&P500, VHT (Health), VGT (Technology) and XLI (Industry) are the most pronounced, indicating that the cryptocurrency market shifts polarity aggressively between the most extreme states of these indices. The health sector seems to have an impact of a different calibre when it comes to Bitcoin and Ethereum, affecting Bitcoin in a more emphatic way.

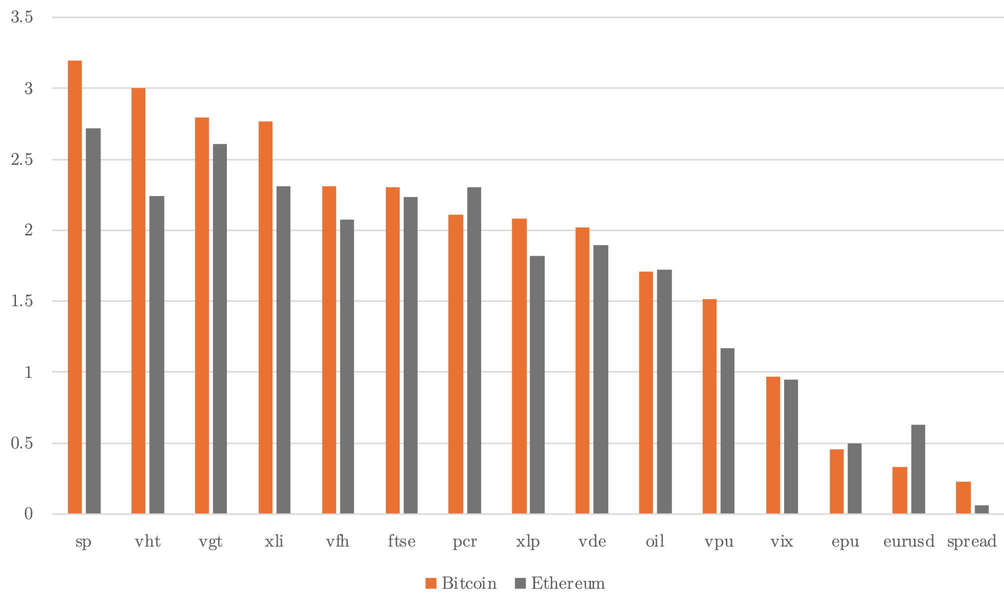

Figure 6: Absolute Difference Between Maximum and Minimum Values of $Linearity$ Factor of Probability Transition Matrices

Naturally, SPREAD, EUR/USD have the least amount of impact in the $Linearity$ transition, as these indices score high in *U-shape* and therefore low in $Linearity$. This, however, does not mean that the transitions between high and low values of these indices do not have a prominent effect in the cryptocurrency market. If we look at the absolute difference of the scores in the *U-shape* factor, these indices are completely isolated with scores roughly 5x higher than the rest of the variables, as these matrices are the only ones that exhibit a pronounced U-shape.

The SPREAD variable measures the difference in interest rates between long-term and short-term US government bonds. If this variable is high, investors expect the economy to grow in the future, so they demand higher interest rates for longer-term investments. Conversely, in the opposite case, investors are worried about the economy's future, possibly expecting slower growth or a recession. This expectation seems to drive a wedge in cryptocurrency returns.

The VIX and EPU indices prove to be the variables that trigger the least movement in the matrices, as they score low in both $Linearity$ and *U-shape*. This translates to PTMs with a balanced and even distribution of probabilities across the transition states for both the high and low values of the covariates.

As these measures can be expressed as the ability to cause movement and change, which can be likened to influence and impact, these results bear close resemblance to the results of the MTD model. The same indices that take a big slice of the total influence from the lambda estimates also score high in the absolute differences of the factors, solidifying both results.

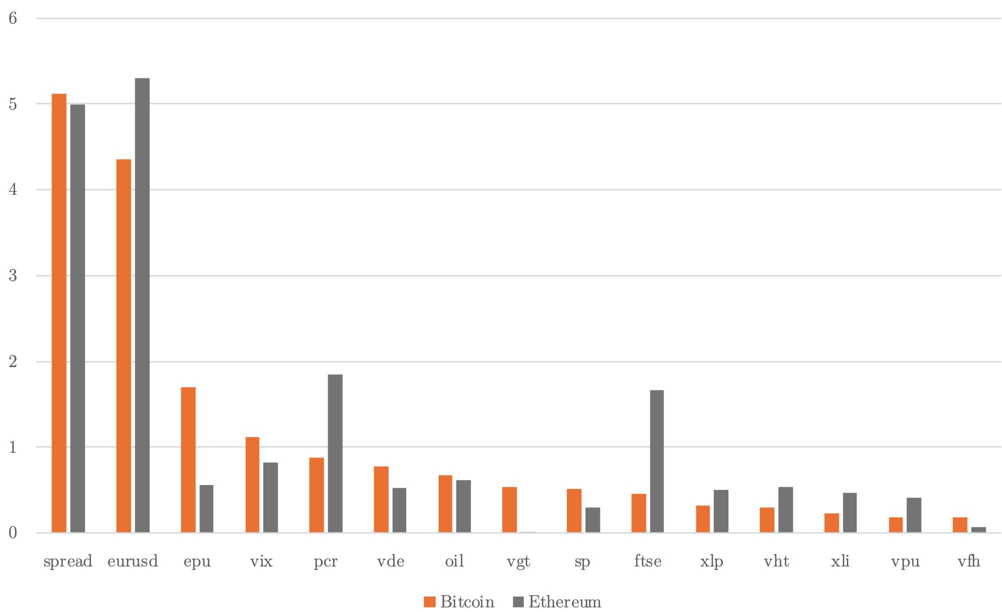

Figure 7: Absolute Difference Between Maximum and Minimum Values of *U-shape* Factor of Probability Transition Matrices

4.4 Discussion

The results are in accordance with the findings in the literature that stock indices, market sentiment measures, volatility indices, interest rates, policy uncertainty and exchange rates affect the returns of cryptocurrencies. Specifically, Bitcoin and Ethereum are found to be highly sensitive to variations in financial indices, with particular salience of the S&P 500 Index, spread of interest rates, the exchange rate of dollars and euros, and the Technological/Financial/Health Sector Indices. These indices also emerge as strong predictors in the aforementioned studies.

In regards to auto-correlation, Bitcoin and Ethereum's daily returns display a high dependence between consecutive days. The findings on Ethereum's partially unreciprocated influence on Bitcoin appear to be consistent with the conclusions of Papadimitriou et al. (2022) that "the returns of alternative cryptocurrencies provide important information on Bitcoin return spikes" and that there is "evidence that the cryptocurrencies markets are not segmented between them and are becoming more integrated with information spillovers from one cryptocurrency to the other". However, this study differs from Papadimitriou in its conclusion that there is a link between the financial markets and the cryptocurrency markets. It is important to note that the span of the analysed data in Papadimitriou is from 10 May 2013 to 29 April 2019. It could be argued that as cryptocurrencies became more established in the economy, their relationship to parallel markets became more apparent.

The difference in conclusions could also be rooted in the difference between the GARCH/SVM and Markov chain methodologies. Damásio and Mendonça (2019) demonstrate that the Markov chain methodology (specifically the MTD-Probit specification) is effective in capturing multivariate relationships and dependencies between two technologies, surpassing traditional parametric econometric techniques like vector autoregressions that only capture linear relationships. Unlike methods such as multivariate GARCH models that focus on conditional mean and variance, non-linear methodologies such as Markov chains can capture complex relationships beyond these moments. The MTD-Probit model, free from parametric assumptions and constraints, allows for the identification of a broader range of associations between variables, demonstrating the advantage of non-parametric approaches in economic modeling.

5 Conclusion

This study aimed to explore the cross-market influences on cryptocurrencies, specifically Bitcoin and Ethereum, through the lens of multivariate Markov chains. By analysing the transitions in returns influenced by various financial indices, the goal was to provide a clearer understanding of how these digital assets interact with traditional financial markets. The Mixture Transition Distribution (MTD) and Generalized Multivariate Markov Chain (GMMC) models were utilised as the framework to capture the intricate dependencies and state transitions.

The findings reveal significant auto-correlation in the daily returns of both Bitcoin and Ethereum, consistent with existing literature. Notably, Ethereum exhibited a partial influence on Bitcoin, challenging the conventional notion that Bitcoin solely leads the market. The study also highlighted the strong predictive power of certain financial indices, such as the S&P500 Index, interest rate spreads, and exchange rates, on cryptocurrency returns. These results align with those of previous studies that underscore the growing integration between cryptocurrency and traditional financial markets. The primary objective was to decipher the dynamic interactions between cryptocurrencies and financial market indicators. The results demonstrate that cryptocurrency return states are significantly affected by the direction of traditional financial indices. This finding answers the research question by confirming that financial market conditions play a crucial role in shaping cryptocurrency returns.

It should be noted that the study is not without limitations. Firstly, the Markov property assumes that the future state depends only on the current state and not on the sequence of events that preceded it. This may result in an oversimplification of the actual dependencies observed in financial time series. This issue could be addressed by adopting a high-order Markov chain methodology, which was not within the scope of this study. Secondly, the study period may not encompass the full spectrum of economic cycles, including prolonged recession or boom periods, which could limit the generalisability of the findings. Similarly, this study does not account for the potential existence of structural breaks. For instance, the study period encompasses specific market conditions, such as the global response to the pandemic caused by the COVID-19 virus, which may not be representative of typical market behaviour.

From a policy perspective, these findings suggest that regulators and investors should consider the intertwined nature of cryptocurrency and traditional financial markets. While the inherent structure of Blockchain fosters the volatility and lack of regulation, there is evidence to suggest that the cryptocurrencies do in fact react to the market similarly to traditional currencies. Nevertheless, the likelihood of a more widespread formal adoption remains remote, and any such developments are unlikely to occur in the near future. However, there is a possibility of taking steps to improve the integration and stability of cryptocurrencies within the broader financial ecosystem, and studies such as this may provide assistance in that matter.

It would be beneficial for future research to extend the granularity of analysis and include a broader range of financial indices in order to explore additional possible influences and enhance the robustness of the findings. Additionally, it would be advantageous to explore other non-linear modelling approaches as well as taking into consideration the existence of structural breaks and memory in the processes. Expanding the scope to include other major cryptocurrencies and assessing their interactions with Bitcoin and Ethereum could further enrich the understanding of cross-market influences in the cryptocurrency ecosystem.

References

Albert, R., & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, *74*(1), 47–97. https://doi.org/10.1103/RevModPhys.74.47

Ang, A., & Bekaert, G. (2002). International Asset Allocation With Regime Shifts. *Review of Financial Studies*, *15*(4), 1137–1187. https://doi.org/10.1093/rfs/15.4.1137

Araújo, T., & Barbosa, P. (2023). Reconstructing Cryptocurrency Processes via Markov Chains. *Computational Economics*. https://doi.org/10.1007/s10614-023-10512-1

Au, S.-K., & Beck, J. L. (2001). Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, *16*(4), 263–277. https://doi.org/10.1016/S0266-8920(01)00019-4

Benhamed, A., Messai, A. S., & El Montasser, G. (2023). On the Determinants of Bitcoin Returns and Volatility: What We Get from Gets? *Sustainability*, *15*(3), 1761. https://doi.org/10.3390/su15031761

Berchtold, A. (2001). Estimation in the Mixture Transition Distribution Model. *Journal of Time Series Analysis*, *22*(4), 379–397. https://doi.org/10.1111/1467-9892.00231

Berger, T., & Koubová, J. (2024). Forecasting Bitcoin returns: Econometric time series analysis vs. machine learning. *Journal of Forecasting*, for.3165. https://doi.org/10.1002/for.3165

Ching, W.-K., Huang, X., Ng, M. K., & Siu, T.-K. (2013). *Markov Chains: Models, Algorithms and Applications* (Vol. 189). Springer US. https://doi.org/10.1007/978-1-4614-6312-2

Ching, W.-K. (2002). A multivariate Markov chain model for categorical data sequences and its applications in demand predictions. *IMA Journal of Management Mathematics*, *13*(3), 187–199. https://doi.org/10.1093/imaman/13.3.187

CoinMarketCap. (2024). Bitcoin Historical Data.

Damásio, B., Louçã, F., & Nicolau, J. (2018). The changing economic regimes and expected time to recover of the peripheral countries under the euro: A nonparametric approach. *Physica A: Statistical Mechanics and its Applications*, *507*, 524–533. https://doi.org/10.1016/j.physa.2018.05.089

Damásio, B., & Mendonça, S. (2019). Modelling insurgent-incumbent dynamics: Vector autoregressions, multivariate Markov chains, and the nature of technological competition. *Applied Economics Letters*, *26*(10), 843–849. https://doi.org/10.1080/13504851.2018.1502863

Damásio, B., & Mendonça, S. (2023). Leader-follower dynamics in real historical time: A Markovian test of non-linear causality between sail and steam (co-)development. *Applied Economics*, *55*(17), 1908–1918. https://doi.org/10.1080/00036846.2022.2100868

Damásio, B., & Nicolau, J. (2014). Combining a regression model with a multivariate Markov chain in a forecasting problem. *Statistics & Probability Letters*, *90*, 108–113. https://doi.org/10.1016/j.spl.2014.03.026

Damásio, B., & Nicolau, J. (2024). Time inhomogeneous multivariate Markov chains: Detecting and testing multiple structural breaks occurring at unknown dates. *Chaos, Solitons & Fractals*, *180*, 114478. https://doi.org/10.1016/j.chaos.2024.114478

Hägele, S. (2024). Centralized exchanges vs. decentralized exchanges in cryptocurrency markets: A systematic literature review. *Electronic Markets*, *34*(1), 33. https://doi.org/10.1007/s12525-024-00714-2

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, *57*(2), 357. https://doi.org/10.2307/1912559

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, *17*(8), 754–755. https://doi.org/10.1093/bioinformatics/17.8.754

Jarrow, R. A., Lando, D., & Turnbull, S. M. (1997). A Markov Model for the Term Structure of Credit Risk Spreads. *Review of Financial Studies*, *10*(2), 481–523. https://doi.org/10.1093/rfs/10.2.481

Khan, F. U., Khan, F., & Shaikh, P. A. (2023). Forecasting returns volatility of cryptocurrency by applying various deep learning algorithms. *Future Business Journal*, *9*(1), 25. https://doi.org/10.1186/s43093-023-00200-9

Koki, C., Leonardos, S., & Piliouras, G. (2022). Exploring the predictability of cryptocurrencies via Bayesian hidden Markov models. *Research in International Business and Finance*, *59*, 101554. https://doi.org/10.1016/j.ribaf.2021.101554

Makarov, I., & Schoar, A. (2022). *Cryptocurrencies and Decentralized Finance (DeFi)* (tech. rep. No. w30006). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/w30006

Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System.

Nascimento, K. K. F. D., Santos, F. S. D., Jale, J. S., Júnior, S. F. A. X., & Ferreira, T. A. E. (2023). Extracting Rules via Markov Chains for Cryptocurrencies Returns Forecasting. *Computational Economics*, *61*(3), 1095–1114. https://doi.org/10.1007/s10614-022-10237-7

Nicolau, J., & Riedlinger, F. I. (2015). Estimation and inference in multivariate Markov chains. *Statistical Papers*, *56*(4), 1163–1173. https://doi.org/10.1007/s00362-014-0630-6

Panagiotidis, T., Papapanagiotou, G., & Stengos, T. (2024). A Bayesian approach for the determinants of bitcoin returns. *International Review of Financial Analysis*, *91*, 103038. https://doi.org/10.1016/j.irfa.2023.103038

Papadimitriou, T., Gogas, P., & Athanasiou, A. F. (2022). Forecasting Bitcoin Spikes: A GARCH-SVM Approach. *Forecasting*, *4*(4), 752–766. https://doi.org/10.3390/forecast4040041

Pennoni, F., Bartolucci, F., Forte, G., & Ametrano, F. (2022). Exploring the dependencies among main cryptocurrency log-returns: A hidden Markov model. *Economic Notes*, *51*(1), e12193. https://doi.org/10.1111/ecno.12193

Raftery, A. E. (1985). A Model for High-Order Markov Chains. *Journal of the Royal Statistical Society: Series B (Methodological)*, *47*(3), 528–539. https://doi.org/10.1111/j.2517-6161.1985.tb01383.x

Shaikh, I. (2020). Policy uncertainty and Bitcoin returns. *Borsa Istanbul Review*, *20*(3), 257–268. https://doi.org/10.1016/j.bir.2020.02.003

Vasconcelos, C., & Damásio, B. (2022). GenMarkov: Modeling Generalized Multivariate Markov Chains in R [Version Number: 1]. https://doi.org/10.48550/ARXIV.2202.00333

NOVA
IMS
Information
Management
School