Department of Statistics
Savitribai Phule Pune University
MSc-II ST-402 Project

Project Report

Exploring the Metaverse: A Statistical Study on Cryptocurrencies and Virtual Reality

Julio Hasiniaina Andrianasolo Randrianarimanana (2226)
Ayan Bashir Sheikh (2244)

Guided by:- Dr. Madhuri G. Kulkarni

Certificate of the Guide

This is to certify that, the following students of M.Sc. Statistics,

(1) Julio Hasiniaina Andrianasolo Randrianarimanana (2226)

(2) Ayan Bashir Sheikh (2244)

have successfully completed their project titled ***Exploring the Metaverse: A Statistical Study on Cryptocurrencies and Virtual Reality*** under the guidance of Dr. Madhuri G Kulkarni and have submitted this project report on 22 May 2024 as a part of the course ST-402, towards partial fulfillment of requirements for the degree of M.Sc. Statistics in Savitribai Phule Pune University in the academic year 2022-2024.

Dr. Madhuri G. Kulkarni
(Project Guide)

Prof. T. V. Ramanathan
(Head Of The Department)

Acknowledgments

We would like to express our deepest gratitude to our project guide, **Dr. Madhuri Kulkarni**, for her invaluable guidance and insightful feedback throughout this research project. Her expertise and encouragement were instrumental in shaping the direction of this research. We would like to thank Head of Department, **Prof. T. V. Ramanathan**, for providing the essential resources and facilities that allowed us to successfully complete this project. Special thanks to our fellow classmates for their technical assistance and moral support throughout the project.

Contents

1 Introduction

> *What is real? How do you define 'real'? If you're talking about what you can feel, what you can smell, what you can taste and see, then 'real' is simply electrical signals interpreted by your brain.*
>
> —Morpheus to Neo.(1999 **The Matrix** science fiction film)

1.1 Metaverse

The Metaverse can be defined as an internet-supported three-dimensional virtual environment that can be accessed through computers and augmented reality devices. The term Metaverse was originally coined by science fiction writer Neal Stephenson in his 1992 Cyberpunk novel Snow Crash, in which he presented a 3D virtual world where people represented as avatars could interact with each other and be represented as artificially intelligent agents.The Metaverse universe can be viewed as a virtual universe where individuals have a customizable digital avatar and can navigate different virtual worlds, play games, watch concerts, socialize, or collaborate,and travel in a virtual environment. This can be accomplished through a fully virtual environment using virtual reality (VR) technologies or with a virtual content layer using augmented reality (AR) or mixed reality (MR) technologies.Metaverse is also suitable for remote working conditions, providing an environment where managers can communicate with employees, read their body language, and maintain face-to-face interaction. Additionally, Metaverse is an open-source platform where anyone can produce worthwhile projects and trade NFT's for financial gain. Metaverse's potential extends beyond entertainment and business, with implications for the healthcare sector. The increased Internet speed and decreased delay times can open the way for surgery via Metaverse. It is also a valuable tool for healthcare professionals who face geographic restrictions, enabling doctors to interact with patients, examine them, and understand their health status.

Metaverse has its handicaps. There are problems with security and privacy issues The technologies that power Metaverse platforms come with their own set of risks.Cyber-attacks are the first thing that comes to mind when discussing the digital world.With Metaverse, the digital space will be closely intertwined with the real world, and the risk of cyber-attacks will increase. Numerous concerns, such as network credential theft, identity theft, social engineering attacks, and ransomware attacks, can arise due to the AR and VR technologies powering the Metaverse. People need a real social life, family, and friends because of their thousands of years of routine. Virtual friends and families will not provide the same happiness and satisfaction as real ones. Although the exact measurement of the Metaverse has not been made, there is no risk of it affecting our perception of real time and space.

2 Motivation

From the dawn of the civilization to the emergence of the Metaverse, advances in communication technology and human imagination have led to transformative changes in our perception of time and space. These technological developments have made vast distances seem closer and reduced the time and effort required to complete tasks, from months to days, minutes, and even milliseconds. In particular, the rise of web technologies has transformed the Internet into a nexus for communication, with unprecedented levels of multimedia content in forms such as writing, audio, and images. On September 13, 2023, Veyond a Metaverse successfully completes historic digital surgery, connecting surgeon in France and patient in Myanmar. The successful remote surgery shows how metaverse technology can change healthcare globally, encouraging more study into its many uses and potential impacts. This real-world example highlights the importance of analyzing and using metaverse capabilities for the future advancements and innovations.In Summary the goal of the study is to provide a comprehensive understanding of the evolution of web and the impacts of new technologies in our daily life.

3 Elements of Metaverse

3.1 Web 3.0

Web 3.0 is the evolution of the internet into a decentralized ecosystem built on blockchain technology. It leverages artificial intelligence to create a 'semantic web' that processes information with human-like intelligence. The semantic web enables machines to handle complex tasks and supports pervasive data sharing across various boundaries, enhancing user privacy through encrypted digital identities.

3.2 Blockchain

Blockchain is a decentralized ledger that records transactions across a distributed network of computers. It structures data in blocks linked in a chain, ensuring immutability and eliminating intermediaries. Originating from Bitcoin's implementation, blockchain now underpins various applications, including tokenization and decentralized finance (DeFi), offering enhanced security and efficiency.

3.3 Cryptocurrencies and Metaverse Tokens

Cryptocurrencies are digital assets classified as 'coins' or 'tokens'. Coins operate on their own blockchain, while tokens use existing blockchains. In the metaverse, cryptocurrencies facilitate transactions and support the sustainability and security of blockchain platforms.

Platforms like Ethereum have native cryptocurrencies, such as Ether, integral to their network's functionality.

3.4 Extended Reality (VR, AR and MR)

This technology is an umbrella term that refers to virtual reality, augmented reality, and mixed reality technologies which empower and support reality.In Metaverse, which is based on the convergence of technologies that enable multi-sensory interactions between extended reality technology, digital objects, and people.individuals can get together with their friends, meet with their friends, attend meetings, concerts, fashion shows, and events, buy a digital product and try it on their avatars without changing their position in the real world, using XR devices. Therefore, these technologies provide users with realistic, interactive environments and offer unforgettable experiences in different contexts.Considering this study, it is aimed to discuss the technology of extended reality, which is one of the significant building blocks of Metaverse, and its use in Metaverse.

3.5 NFTs

Non-Fungible Tokens (NFTs) are unique digital assets on a blockchain, each with a distinct identifier that traces ownership and history. They can encapsulate various forms of digital media and are immutable once minted. NFTs are not interchangeable like cryptocurrencies, making them akin to unique art pieces in the digital realm.

3.6 Gamification in the Metaverse

The Metaverse, a virtual universe accessed through virtual reality, offers a space parallel to real life where users interact via avatars. It supports a variety of activities, from social interactions to educational experiences. Gamification within the Metaverse applies game-design elements to enhance engagement, improve problem-solving abilities, and encourage creativity. It transforms the Metaverse into an active, user-focused environment, transcending physical, temporal, and linguistic barriers.

4 Recent Contributions

4.1 Denisa Elena BĂLĂ1, Stelian STANCU2. Cryptocurrencies in the Digital Era. Composite Index-Based Analysis for the Top Ten Virtual Currencies Traded -June 2022

- **Introduction:-** Digital currencies are recent form of currency relying on Blockchain technology. Both cryptocurrencies and Blockchain are the interest of investors and researchers alike. Bitcoin, which is the pioneer digital currency, relies on Blockchain and holds the top position in the market. Despite of their popularity, many people remain unfamiliar with Blockchain and cryptocurrencies. Nevertheless, some experts believe that Blockchain has the potential to revolutionize various industries, similar to the impact of the Internet. Blockchain ensures secure and efficient transactions across different sectors and brings together experts from mathematics, computer science, and cryptography. Interest in the cryptocurrency market has surged, with its market capitalization reaching approximately 1.8 trillion USD in April 2022. Bitcoin facilitates transactions without the need of intermediaries due to its decentralized system. Blockchain records transactions securely without the involvement of third parties. The perception of Bitcoin varies among experts, with some view it as a secure investment and others as a speculative bubble. Cryptocurrencies are highly volatile and influenced by global economic activities and trading volumes. Some studies investigate the relationship between public sentiment and cryptocurrency prices, with influential figures like Elon Musk impacting market trends through social media. The paper aims to analyse the cryptocurrency market by looking at the top ten digital currencies based on their market value. Firstly, they created a combined index using two important indicators: the returns of the cryptocurrencies and their trading volumes. They want to see how the cryptocurrencies' performance would vary by adjusting the weights of these indicators in the index. Additionally, they want to check if there are any causal relationship between the indices of the different cryptocurrencies.

- **Methodology**

  To achieve the research goals, the research paper gathers data on the top ten digital currencies . To ensure a comprehensive dataset, only cryptocurrencies launched after 2017 are included. The cryptocurrencies considered are: Bitcoin (BTC), Ripple (XRP), Binance Coin (BNB), Ethereum (ETH), Cardano (ADA), Dogecoin (DOGE), Tether (USDT), Litecoin (LTC), Dash (DASH), and Monero (XMR). For each currency, the daily returns is calculated as follow:

$$r_i = \frac{P_{t+1} - P_t}{P_t} \times 100 \tag{1}$$

1) r_i represents the return of the cryptocurrency.
2) P_{t+1} represents the price of cryptocurrency i at t+1 moment.
3) P_t represents the price of cryptocurrency i at t+1 moment.

The traded volume in term of growth rates is considered within the analysis. After preparing these two indicators, an index is created to summarize the performance of the top ten digital currencies.
A combined index is created, which is explained below:

$$\text{Cryptoindex}_{(i,t)} = \alpha_i r_{i,t} + \beta_i V_{i,t}$$

where

- $Cryptoindex_{(i,t)}$: represents the index associated with the cryptocurrency i at the time period t.
- α_i represents the weight associated with r in constructing the composite index
- β_i represents the weight associated with v in constructing the composite index
- $r_{i,t}$ represents the return of the cryptocurrency i at the time t
- $V_{i,t}$ represents the traded volume, expressed as growth rate, associated with the cryptocurrency i,at the time t. The weights α_i and β_i can be determined using the actual data series and simulation techniques.

  Additionally, the assess the correlation between the ten cryptocurrencies using the Pearson correlation coefficient. This coefficient helps to understand the strength and direction of the linear relationship between pair of currency. The Granger test (1969) is employed to evaluate causality between the composite indices linked to the chosen digital currencies.

- **Discussion:-** Over the range period, there were notable fluctuations in the returns and trading volumes of these cryptocurrencies. Binance (BNB Index), Dash (DASH Index), Dogecoin (DOGE Index), and Monero (XMR Index) experienced significant increases

in the fourth quarter of 2020, followed by declines in both returns and trading volumes. Bitcoin and Ethereum, being the most popular, exhibited fluctuations, yet the index value remained relatively stable, as depicted in the charts. Comparing the two graphs, we observed minimal differences in the index movements, except for Ethereum (ETH Index), Litecoin (LTC Index), and Tether (USDT Index). These cryptocurrencies displayed strong positive correlations between trading volumes and returns. The simulations indicated that adjusting the weights between the index components had little impact on the index value. Correlation analysis revealed strong connections between the indices of various cryptocurrencies, regardless of the weighting scheme. Bitcoin's index demonstrated strong correlations with most others, except for Monero (XMR Index), whereas Monero and Ripple (XRP Index) showed weak correlations. The Augmented Dickey-Fuller stationarity test revealed nonstationarity in the level time series. However, after applying first-order differentiation, all indicators became stationary. The Johansen cointegration test confirmed a long-term relationship between the variables. Granger causality tests detected significant bidirectional connections between Bitcoin and Binance, Bitcoin and Dogecoin, Bitcoin and Ethereum, Dash and Monero, and Ripple and Litecoin, indicating mutual influence. For expample, BTC Index can predict BNB Index.

- **Conclusion:-** This study focuses on understanding the cryptocurrency market by looking at how major virtual currencies are connected and how they affect each other. Before starting the analysis, an index was created to summarize how key cryptocurrencies are doing. This index includes calculated returns and trading volumes of selected virtual currencies, with weights assigned based on observed or simulated data. Modifying the weights did not lead to significant changes in the performance of major digital assets, except for cryptocurrencies where returns and trading volumes were correlated. Using Pearson correlation coefficient, strong positive connections were identified among the ten currencies studied. Furthermore, Granger causality analysis revealed bidirectional connections between Bitcoin and Binance, Bitcoin and Dogecoin, Bitcoin and Ethereum, Dash and Monero, and Ripple and Litecoin. This suggests that past prices of these cryptocurrencies may serve as predictors for the prices of others.
- **Limitation:-**
  - The cryptocurrency market is known for its high volatility and rapid fluctuations, which can make it challenging to detect stable relationships over time. Market conditions can change quickly, impacting the validity of historical correlations and causal relationships.
  - They did not explicitly specify the method for calculating the values representing the weights α_i and β_i in constructing the composite index. This step is very important as the composite index is be used for the rest of the analysis.

4.2 Shivangi Sahay, Nishtha Mahajan, Shivangi Malik, Jasdeep Kaur. Metaverse: Research Based Analysis and Impact on Economy and Business-(August 2022)

- **Data used :-** The research paper investigates the Metaverse cryptocurrency tokens, specifically AXS, MANA, SAND, and ILV. The dataset comprises actual closing prices of these cryptocurrencies, spanning from March 2021 to March 2022. The analysis is based on approximately 1460 observations.

- **Emerging Features and analysis:-** The research paper explores the about the feature related actual closing price of the Metaverse token and the details about the VR(Virtual reality), AR(Augmented Reality), and Metaverse can be described as the network of the three-dimensional virtual world focusing. Also, the metaverse can be called the amalgamation of a virtual intensified physical world and a physically tenacious simulated place. The paper also explores the business game construction for training. (It is a high-level plan for operating a business profitably in the marketplace). They used $ARIMA$ and $SARIMAX$ models for the analysis of the actual closing price of respective Metaverse tokens, They also used the Augmented Dickey fuller test for checking the stationarity of the data. test for checking the stationarity of the data.

  H_0**: It is recognized when the input series contains a unit root.**
  H_1**: It is recognized when the input series does not contain a unit root**
  this research paper also forecasts the closing price of the respective metaverse tokens.

- **Result and conclusions;-** In this study, they used time series analysis for forecasting and model fitting for considered cryptocurrency/ Metaverse Tokens. The $ARIMA$ and $SARIMAX$ models are built and trained with the actual closing price on the training data set. the paper considered four different metaverse tokens (cryptocurrency coins) and they find out that the ARIMA model is appropriate for the following tokens

  - (Axs coin) Axie infinity
  - Illusion (ILV coin)
  - Decentraland(MANA coin)

  They have also find out that the for the forecasting for the SAND the $ARIMA$ model is not appropriate. for that they use $SARIMA$ model. for the forecasting the data was too non stationary For checking the performance of the model they used four types of accuracy criteria that

  MSE(Mean quare error), MAE(Mean Absolute Error), RMSE(Root mean square error), MAPE(Mean Absolute Percentage Error).

4.3 Xiao-Ting Huang, Jiahui Wang, Zhihui Wang, Linqiang Wang 1 and Chenfei Cheng 1. Experimental study on the influence of virtual tourism spatial situation on the tourists temperature comfort in the context of metaverse

- **Data used:-** In this study, data was gathered from 180 volunteers who took part in a lab experiment. They were subjected to various virtual tourism scenarios with different environmental temperatures, sounds, and visuals. The researchers recorded the participants' thermal feelings, physiological markers, and comfort levels related to temperature. A seven-point scale was used to gauge thermal sensation, ranging from "cold" to "hot". Electrocardiogram devices were used to measure physiological indicators by recording heart rate variability. A five-point scale was used to assess temperature comfort, ranging from "very uncomfortable" to "very comfortable". The data collection process was thorough and all necessary ethical considerations were adhered to.

- **Methods used for analysis:-** The research team employed a multivariate analysis of variance (MANOVA) to examine the primary effects of the independent variables (environmental temperature, audio, and visual conditions) on the dependent variables (thermal sensation, physiological indicators, and temperature comfort). This statistical method enabled them to comprehend the influence of each independent variable on the dependent variables while accounting for the effects of the other independent variables. In addition to MANOVA, the researchers utilized the PROCESS macro, a computational tool for path analysis, mediation, moderation, and conditional process modeling, to examine the mediation effects of thermal sensation and physiological indicators on the relationship between tourism spatial situations and temperature comfort. This method enabled them to comprehend the indirect effects of the independent variables on the dependent variable through the mediators.

- **Result and conclusion:-** The study's findings revealed that environmental temperature, audio, and visual conditions significantly influenced the participants' thermal sensation, physiological indicators, and temperature comfort. Specifically, higher environmental temperatures, louder audio, and brighter visual conditions were linked with higher thermal sensation, higher physiological indicators, and lower temperature comfort. The mediation analysis revealed that thermal sensation and physiological indicators mediated the effects of tourism spatial situations on temperature comfort. This implies that the influence of tourism spatial situations on temperature comfort was partially explained by their effects on thermal sensation and physiological indicators. The researchers concluded that the creation of virtual tourism spatial situations should take into account the physiological and psychological data of the experiencers. They also proposed that virtual tourism can serve as a sustainable alternative to real tourism in the metaverse

era, reducing the environmental impact of tourism while providing a comparable level of experience to tourists

5 Objectives and proposed Methodology

1 Explore the long-term relationships among Metaverse cryptocurrencies (Decentraland, Axie Infinity, Sandbox, and Enjin) .

2 Predict future states (price movements) of Metaverse Tokens.

3 Assess the risk associated with different cryptocurrencies.

4 Examine the dynamics of these cryptocurrencies within the Metaverse by investigating the error correction term and lagged values in the VECM model.

5 **User Experience Analysis:-** Analyze the overall user experience when using the virtual reality (VR) application.

   * **Immersion:-** Measure the level of immersion users feel when using the VR application.
   * **Presence:-** Evaluate how present users feel in virtual environment.
   * **Interactivity:-** Assess the extent to which users feel they can interact with and engage in the virtual environment.

6 **Emotional State Analysis:-** Investigate the correlation between the use of VR technology and the user's emotional state.

   * **Flow of States:-** Analyze the level of flow states users experience during the use of VR.
   * **Time Perception:-** Evaluate how users perceive the passage of time during the VR experience.
   * **Distraction from External Concerns:-** Assess the extent to which users forget about external concerns during the VR experience.

7 **Aesthetic Appeal Analysis:-** Analyze users' perceptions of the aesthetic aspects of the website interface design.

8 **Travel Intent Analysis:-** Examine the influence of VR experiences on users' intentions to travel to Da Nang.

   * **Intention to travel:-** Analyze users' intentions to travel to Da Nang after experiencing the destination through VR.

9 **Demographic Analysis:-** Analyze the demographic data (gender, age group, marital status, income level) to identify any patterns or correlations with the responses to the VR experience.

10 Determine the primary factors influencing motion sickness levels in virtual reality (VR) users.

11 Evaluate the influence of variables such as gender, VR headset type, duration of VR experience, and age on motion sickness experiences.

We explore long-term relationships among Metaverse cryptocurrencies using Johansen's cointegration analysis. If cointegration is found, we employ Vector Error Correction Models (VECM) to understand dynamics. For price movement prediction, we use time series techniques (ARIMA, GARCH). Risk assessment involves calculating Value at Risk (VaR) and Conditional Value at Risk (CVaR). In VR analysis, we evaluate immersion, presence, and interactivity using non-parametric tests (Wilcoxon, Kruskal-Wallis). Aesthetic appeal and travel intent are also explored. For the travel intention, we fit the logistic regression, Demographic correlations, and motion sickness factors to complete our approach. Additionally, Markov chains predict future token states.

6 Analysis

6.1 Exploratory Data Analysis

A time series is a collection of observations made sequentially through time.Much statistical theory is concerned with random samples of independent observations. The special feature of time series analysis is the fact that successive observations are usually not independent and that the analysis must take into account the time order of the observations. When successive observations are dependent, future values may be predicted from past observations. If a time series can be predicted exactly, it is said to be deterministic. However, most time series are stochastic in that the future is only partly determined by past values, so that exact predictions are impossible and must be replaced by the idea that future values have a probability distribution, which is conditioned on a knowledge of past values.Many time series are routinely recorded in economics and finance. Examples include share prices on successive days, export totals in successive months, average incomes in successive months, company profits in successive years and so on.

financial time series shows the daily returns (or percentage change) of the adjusted closing prices. The mean of the return series seems to be stable with an average return of approximately zero, but the volatility of data changes over time. the first step of the time series analysis is to plot the time plot which can help to identify is there is any trend in the data it also get the idea about the stationarity , seasonality and the outliers in time series data

ACF (Autocorrelation Function)
ACF plot used to identify the presence of the autocorrelation in the data. Autocorrelation is the correlation of a time series with its own past and future values if the autocorrelation plot shows the significant autocorrelations at certain lags, it suggests that the time series data is not random and past values have an effect on current values.

PACF plot(partialautocorrelation Function)

PACF plotshows the correlation between the time series and its lags after removing the effect of any correlations due to the terms at shorter lags. It also helps in getting the idea about the order of an autoregressive model.

ACF and PACF plots are used to understand the correlation structure of a time series data, which is crucial in identifying the appropriate model to predict future values. They help us determine if the data is random (white noise), if it's an autoregressive or moving average process, and the order of the process if it's autoregressive or moving average.

Stationarity

A univariate time series z_t is said to be weakly stationary if

(a) $E(z_t) = \mu$, a constant, and

(b) $\text{Cov}(z_t, z_{t-k}) = \gamma_k$, a constant for each lag k.

Here, $E(z)$ and $\text{Cov}(z_t, z_{t-k})$ denote the expectation and autocovariance function of the random variable z, respectively. Thus, the mean and autocovariance of a weakly stationary time series z_t do not depend on time, that is, the first two moments of z_t are time invariant. Implicit in the definition, we require that the mean and autocovariance of a weakly stationary time series exist.

A univariate time series z_t is strictly stationary if the joint distribution of the m collection, $(z_{t_1}, \ldots, z_{t_m})$, is the same as that of $(z_{t_1+j}, \ldots, z_{t_m+j})$, where m, j, and $(t_1, \ldots, t_m)$ are arbitrary positive integers. In statistical terms, strict stationarity means that the statistical properties of the time series remain constant over time.

Strict stationarity requires that the probability distribution of an arbitrary collection of z_t is time invariant. An example of strictly stationary time series is the sequence of independent and identically distributed random variables of standard normal distribution. From the definitions, a strictly stationary time series z_t is weakly stationary provided that its first two moments exist.

There are test for checking the Stationarity and ARCH effect of the data re as follows

- **Augmented Dickey-Fuller (ADF) Test**:

  - Null Hypothesis (H_0): The time series has a unit root (i.e., it is non-stationary).
  - Alternative Hypothesis (H_1): The time series has no unit root (i.e., it is stationary).

- **Phillips-Perron (P.P) Test**:
  - Null Hypothesis (H_0): The time series has a unit root (i.e., it is non-stationary).
  - Alternative Hypothesis (H_1): The time series has no unit root (i.e., it is stationary).

- **Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test**:
  - Null Hypothesis (H_0): The time series is stationary.
  - Alternative Hypothesis (H_1): The time series is non-stationary.

- **Autoregressive Conditional Heteroskedasticity (ARCH) Effect Test**:
  - Null Hypothesis (H_0): There is no ARCH effect in the time series.
  - Alternative Hypothesis (H_1): There is an ARCH effect in the time series.

- If the p-value ≤ 0.05, then the null hypothesis is **rejected**. This suggests that the observed data is inconsistent with the null hypothesis, and the alternative hypothesis is considered to be supported.
- If the p-value > 0.05, then the null hypothesis is **fail to rejected**. This suggests that the observed data is consistent with the null hypothesis.

6.1.1 Axie Infinity

The first, and most important, step in any time-series analysis is to plot the observations against time.This plot called time plot will show up important features of the series such as trend, seasonality, outliers and discontinuities.so we first start our analysis from plotting the time plot and along with we consider the ACF(Auto Correlation plot) and PACF(Partialuto Correlation plot)

- Axie Infinity Return Plot: The plot depicts Axie Infinity's returns from January 2021 to July 2022. The returns exhibit high volatility, indicating substantial risk and unpredictability. No discernible trend or pattern is observed, suggesting that the returns are influenced by various factors and are not easily predictable.
- ACF Plot: The ACF plot measures autocorrelation, which is the correlation of a series with its own lags. In this case, all autocorrelations are within the confidence intervals, indicating no significant autocorrelation. This suggests that Axie Infinity's returns are not significantly influenced by their past values.

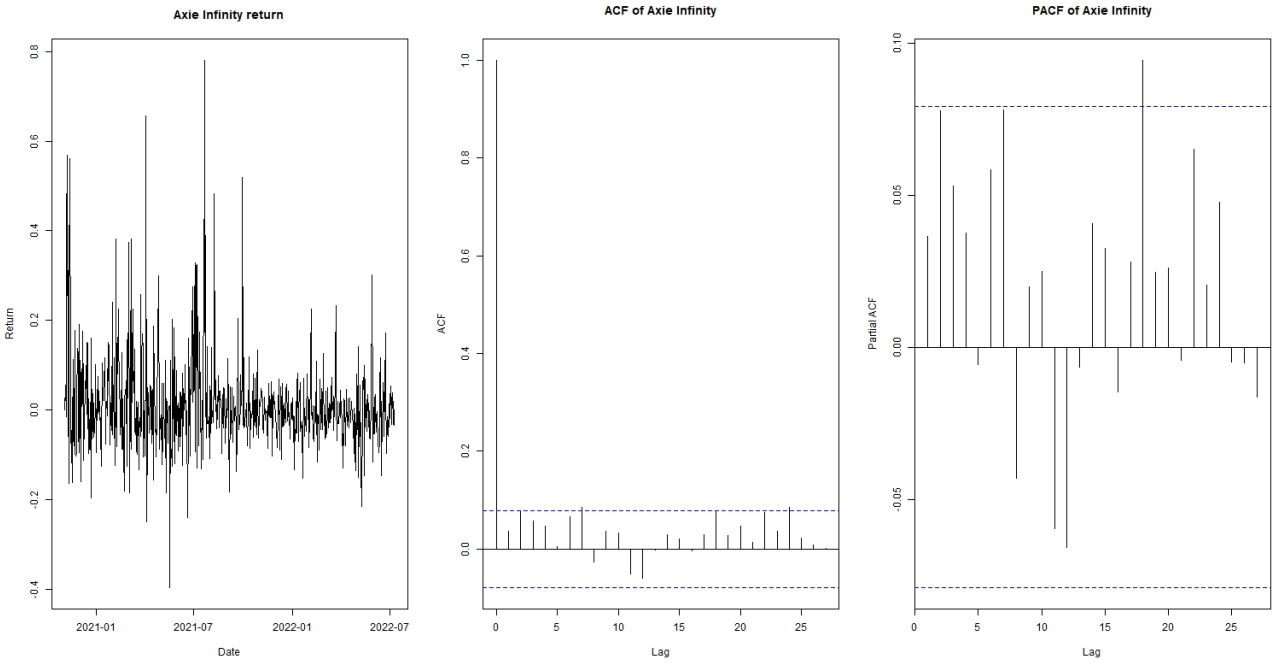

Figure 1: Axie Infinity

- PACF Plot: The PACF plot measures the correlation of a series with its own lag after removing the effect of previous lags. Similar to the ACF plot, all partial autocorrelations are within the confidence intervals, indicating no significant partial autocorrelation. This suggests that any observed correlation between the series and its lags is not due to their mutual correlations with intervening lags.

In conclusion, the analysis suggests that Axie Infinity's returns behave like a random walk, where future values are not predictable based on past values. This characteristic aligns with the efficient market hypothesis, which posits that asset prices fully reflect all available information. This implies that the market for Axie Infinity is efficient, and its price movements are largely unpredictable.

Stationarity Tests

- ADF (Augmented Dickey-Fuller) Test: The test statistic is -7.8097, and the p-value is 0.01. Since the p-value is less than the significance level (α) of 0.05, we reject the null hypothesis that the time series has a unit root, i.e., it is non-stationary. So, the time series is stationary according to the ADF test.
- Phillips-Perron (P.P) Test: The test statistic is -652.07, and the p-value is 0.01. Similar

Test	Test Statistic	Lag Order/Par	p-value	Hypothesis
Augmented Dickey-Fuller	-7.8097	8	0.01	Stationary
Phillips-Perron Unit Root	-652.07	6	0.01	Stationary
KPSS for Level Stationarity	1.0085	6	0.01	Non-stationary
ARCH LM-test	15.732	12	0.2038	No ARCH effects

Table 1: Summary of Stationarity and ARCH Effect Tests

to the ADF test, the p-value is less than the alpha of 0.05, so we reject the null hypothesis of the unit root. Thus, the time series is also stationary according to the P.P test.

- KPSS (Kwiatkowski-Phillips-Schmidt-Shin) Test: The test statistic is 1.0085, and the p-value is 0.01. The null hypothesis for the KPSS test is the opposite of the ADF and PP tests. It is the hypothesis that the time series is stationary. If the p-value is less than the significance level of 0.05, we reject the null hypothesis and infer the series is non-stationary. Here, since the p-value is less than 0.05, we conclude the series is non-stationary according to the KPSS test.
- ARCH (Autoregressive Conditional Heteroskedasticity) Effect: The test statistic is 15.732, and the p-value is 0.2038. If the p-value is less than the significance level of 0.05, we reject the null hypothesis that there is no ARCH effect. Here, since the p-value is greater than 0.05, we do not reject the null hypothesis, suggesting there is no ARCH effect in the time series.

Model fitting

```
Series: d\$Return
ARIMA(1,1,1)

Coefficients:
ar1      ma1
0.0126  -0.9835
s.e.  0.0422   0.0131

sigma^2 = 0.01218:  log likelihood = 480.56
AIC=-955.12   AICc=-955.08   BIC=-941.86
```

- **Model**: The ARIMA(1,1,1) model implies that the current value of the series is influenced by its immediate past value and the error term.
- **AR(1) Coefficient (0.0126)**: Indicates a weak positive correlation with the immediate previous value. The standard error (0.0422) suggests uncertainty in this estimate.

- **MA(1) Coefficient (-0.9835)**: Suggests that the current error term has a strong negative correlation with the previous error term. The standard error (0.0131) is relatively small, indicating a precise estimate.
- **Sigma-squared (0.01218)**: Represents the variance of the error term, indicating the model doesn't explain all variability in the data.
- **Log Likelihood (480.56)**: A measure of goodness-of-fit. Higher values indicate better fit.
- **AIC (-955.12), AICc (-955.08), BIC (-941.86)**: Lower values suggest a better balance between model fit and complexity.

In conclusion, the model appears to fit the data reasonably well, The model suggests that the series is influenced by its immediate past, but with high volatility (as indicated by the MA(1) coefficient and sigma-squared).

Residual Analysis

- **Standardized Residuals**: The first plot shows standardized residuals fluctuating around zero, indicating that the model's predictions are generally close to the actual values, with no apparent bias.
- **ACF of Residuals**: The second plot displays the ACF of residuals with values close to zero across different lags, suggesting minimal autocorrelation and indicating that the model has adequately captured the time series patterns.
- **Ljung-Box Test**: The third plot shows p-values for the Ljung-Box statistic well above the 0.05 threshold, implying that there is no significant autocorrelation in the residuals at various lags.

Overall, the graph suggests that the model fits the data well, with residuals appearing to be random noise, which is a desirable property in time series modeling.

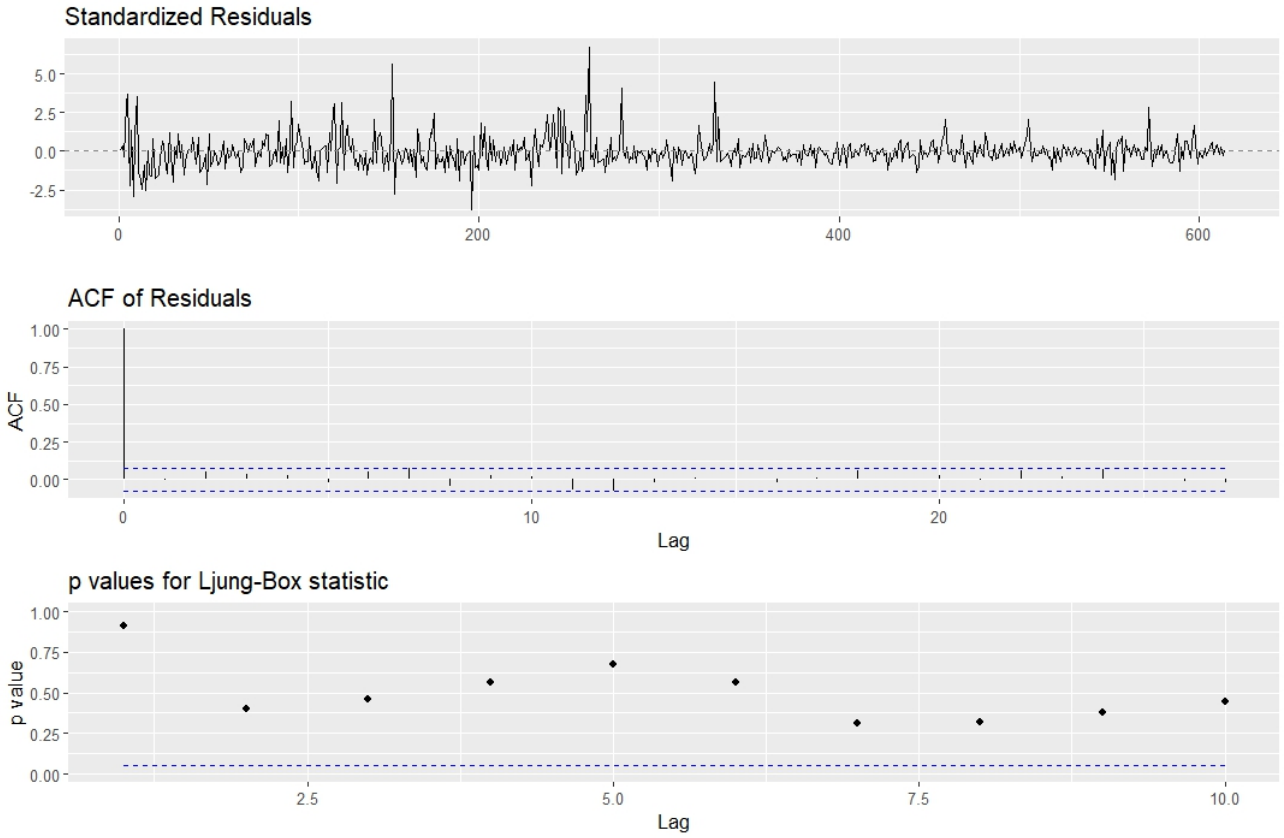

Figure 2: Axie Infinity

6.1.2 Decentraland

- **Return Over Time**: The first graph depicts the return of Decentraland from January 2021 to July 2022. The plot exhibits significant volatility, with numerous spikes indicating periods of rapid gains and losses.
- **Autocorrelation (ACF)**: The second graph, the ACF plot, shows autocorrelation factors that are mostly close to zero across different lags. This suggests minimal autocorrelation, implying that past values do not significantly influence future values.
- **Partial Autocorrelation (PACF)**: The third graph, the PACF plot, displays more variability in partial autocorrelations at different lags. However, most values are within the confidence intervals, indicating no significant partial autocorrelation.

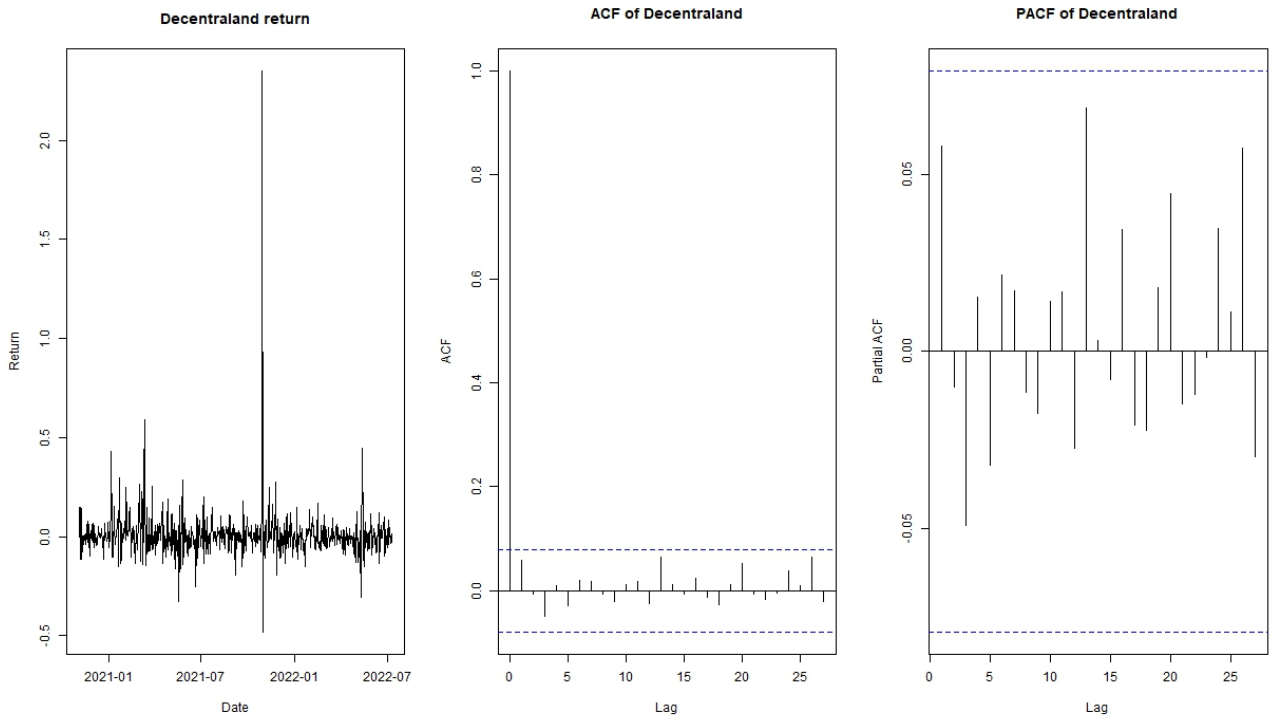

Figure 3: Time plot for Decentraland

Stationarity tests

Test	Test Statistic	Lag Order/Par	p-value	Hypothesis
Augmented Dickey-Fuller	-8.4171	8	0.01	Stationary
Phillips-Perron Unit Root	-555.62	6	0.01	Stationary
KPSS for Level Stationarity	0.22277	6	0.1	stationary
ARCH LM-test	7.5452	12	0.8196	No ARCH effects

Table 2: Summary of Stationarity and ARCH Effect Tests

Model fitting

```
Series: d\$Return
ARIMA(0,0,1) with non-zero mean

Coefficients:
ma1    mean
0.0584  0.0096
s.e.  0.0402  0.0056
```

```
sigma^2 = 0.01734:  log likelihood = 374.57
AIC=-743.15   AICc=-743.11   BIC=-729.89
```

- **Model**: The time series is modeled as an ARIMA(0,0,1) with non-zero mean. This indicates a Moving Average model of order 1 without any autoregressive or differencing terms.
- The estimated coefficient for the MA(1) model is 0.0584, with an associated standard error of 0.0402. This indicates that the current error term is influenced by the previous time point's error term to the extent of 5.84%. Initially, the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots did not reveal any significant dependence on past observations, suggesting a lack of autocorrelation. However, upon fitting the model, the MA(1) was identified as the most suitable, implying that the time series does indeed rely on the lag-1 observation. This apparent contradiction could be due to the subtlety of the autocorrelation structure which is not easily discernible in the ACF and PACF plots but is captured by the MA(1) model fitting process.
- **Mean**: The estimate for the mean of the series is 0.0096, with a standard error of 0.0056. This indicates that the average return of the series is approximately 0.0096.
- **Sigma-squared**: The estimate for sigma-squared (the variance of the error term) is 0.01734. This provides a measure of the variability in the time series that is not explained by the model.
- **Model Fit**: The log likelihood of the model is 374.57, and the AIC, AICc, and BIC values are -743.15, -743.11, and -729.89 respectively. These are measures of the relative quality of the model, taking into account the number of parameters used. Lower values are better, suggesting that the model provides a reasonable fit to the data.

In conclusion, the ARIMA(0,0,1) model seems to provide a reasonable fit to the data, but its predictive accuracy would need to be assessed on new observations.

Residual Analysis

- **Standardized Residuals**: Fluctuations around zero suggest an adequate model fit, with a potential outlier indicated by a spike.
- **ACF of Residuals**: Values within confidence bounds imply no significant autocorrelation, indicating the model's effectiveness.
- **Ljung-Box p-values**: Uniformly above the 0.05 threshold, confirming the absence of autocorrelation in the residuals.

Overall, the model appears to be well-fitted to the data, with the exception of a possible outlier.

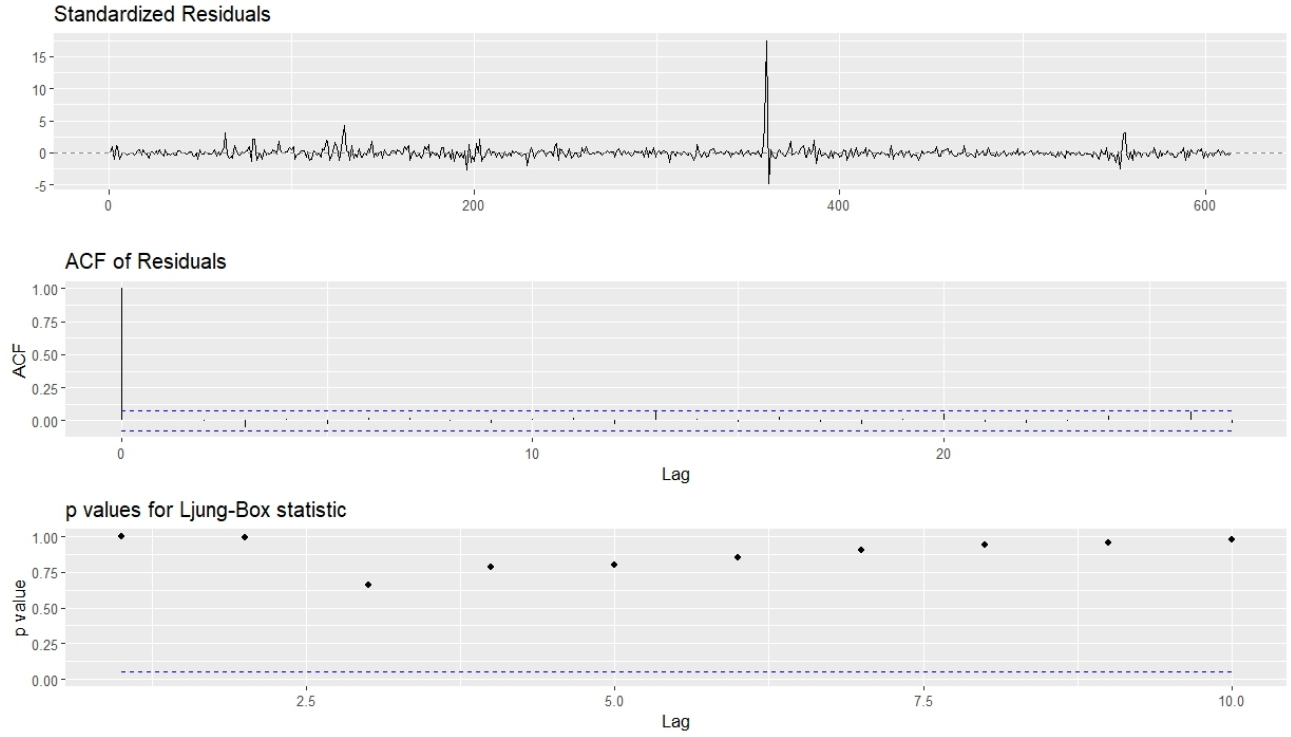

Figure 4: Residual Analysis Decentraland

6.1.3 Enjin

Time Plot

- The **Enjin Return** plot exhibits significant volatility, indicative of the fluctuating nature of Enjin's returns.
- The **ACF of Enjin** plot shows low autocorrelation factors across various lags, suggesting minimal influence of past values on future values.
- The **PACF of Enjin** plot reveals certain lags with notable partial autocorrelation, pointing to their potential predictive power in the time series.

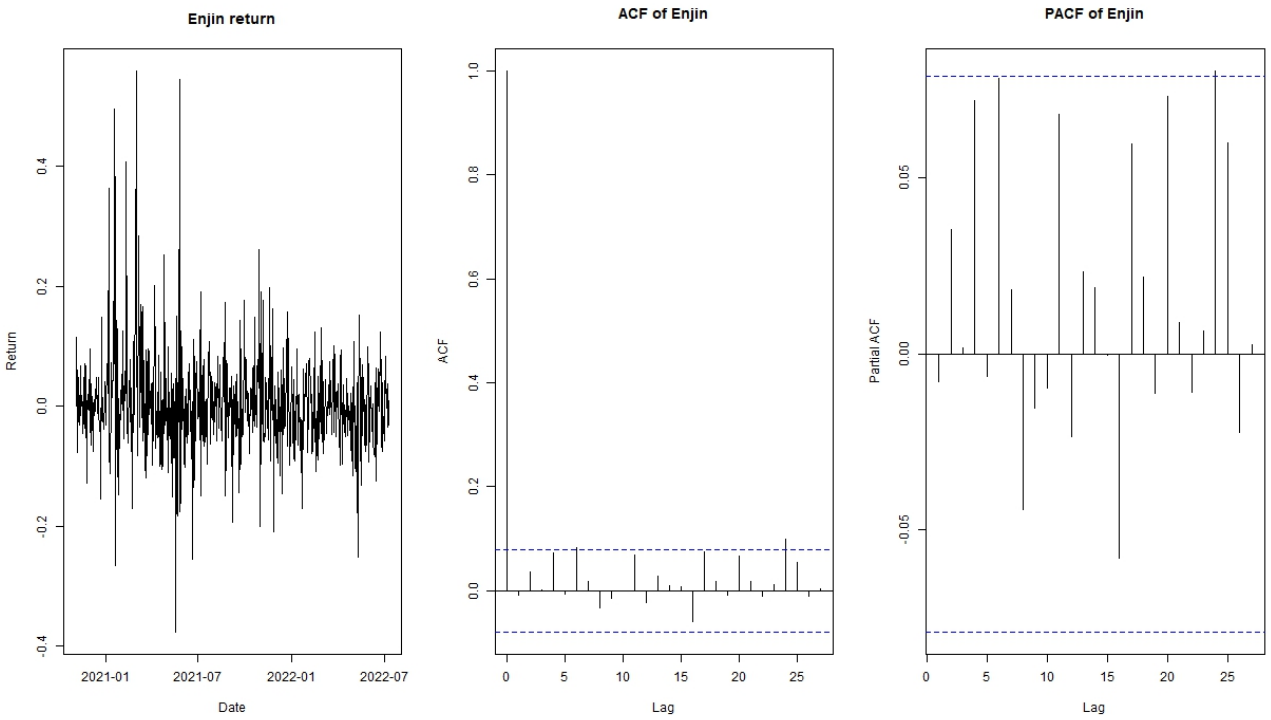

Figure 5: Residual Analysis Decentraland

Stationarity tests

Test	Test Statistic	Lag Order/Par	p-value	Hypothesis
Augmented Dickey-Fuller	-8.1084	8	0.01	Stationary
Phillips-Perron Unit Root	-652.57	6	0.01	Stationary
KPSS for Level Stationarity	0.71423	6	0.01225	Non-stationary
ARCH LM-test	43.295	12	2.012e-05	there is ARCH effects

Table 3: Summary of Stationarity and ARCH Effect Tests

Model fitting

```
Series: d\$Return
ARIMA(0,1,1)

Coefficients:
ma1
-0.9877
s.e.   0.0093
```

```
sigma^2 = 0.007637:  log likelihood = 622.94
AIC=-1241.88   AICc=-1241.86   BIC=-1233.05
```

- **Model**: The time series is modeled as an ARIMA(0,1,1), indicating a Moving Average model of order 1 with one order of differencing.
- **MA(1) Coefficient**: The estimate for the MA(1) coefficient is -0.9877, with a standard error of 0.0093. This suggests that the error at a given time point is approximately -0.9877 times the error at the previous time point.
- **Sigma-squared**: The estimate for sigma-squared (the variance of the error term) is 0.007637, providing a measure of the variability in the time series that is not explained by the model.
- **Model Fit**: The log likelihood of the model is 622.94, and the AIC, AICc, and BIC values are -1241.88, -1241.86, and -1233.05 respectively. These suggest that the model provides a reasonable fit to the data.

Residual Analysis

- Standardized Residuals: Shows residuals fluctuating around zero, suggesting no obvious pattern or bias in the residuals.
- ACF of Residuals: Indicates little to no autocorrelation, as values are close to zero across various lags.
- Ljung-Box Statistic: P-values are above the 0.05 threshold, implying no significant autocorrelation at different lags.

Overall, the graph suggests that the model residuals are random, which is a good indication that the model fits the data well. There's no evidence of autocorrelation, meaning past values do not significantly influence future values in the series. This is often a desirable property in a well-fitting time series model.

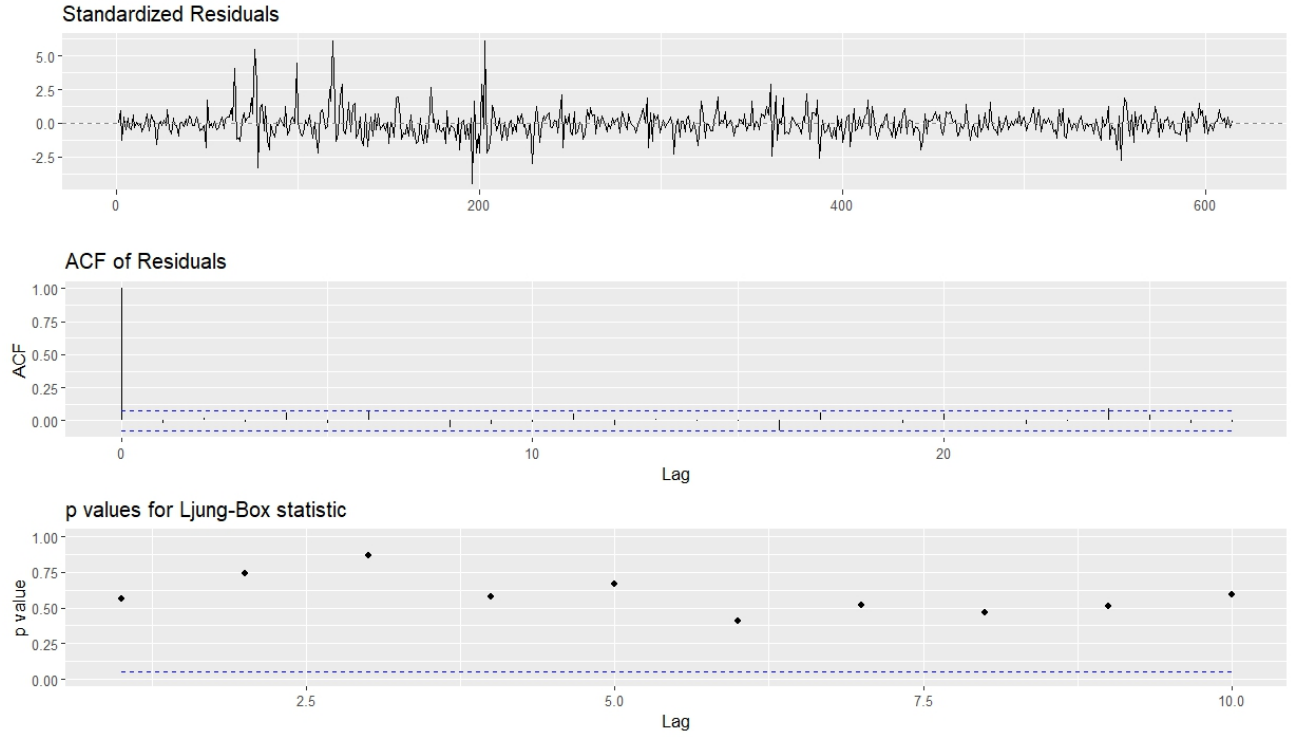

Figure 6: Residual Analysis Decentraland

6.1.4 Sandbox

Time Plot

- **Sandbox Return Fluctuations:**
  - The time series plot shows Sandbox returns fluctuating around zero from 2021-01 to 2022-07, indicating periods of both gains and losses.
- **Significant ACF at Initial Lags:**
  - The ACF plot reveals significant autocorrelation at initial lags, suggesting that past values have a strong influence on immediate future values.
- **Indicative PACF of AR Process:**
  - The PACF plot shows a few bars exceeding the significance threshold, hinting at a potential autoregressive (AR) process in the data.

Overall, the analysis suggests that the Sandbox returns exhibit time-dependent structures, which could be modeled using AR terms to predict future values based on past returns.

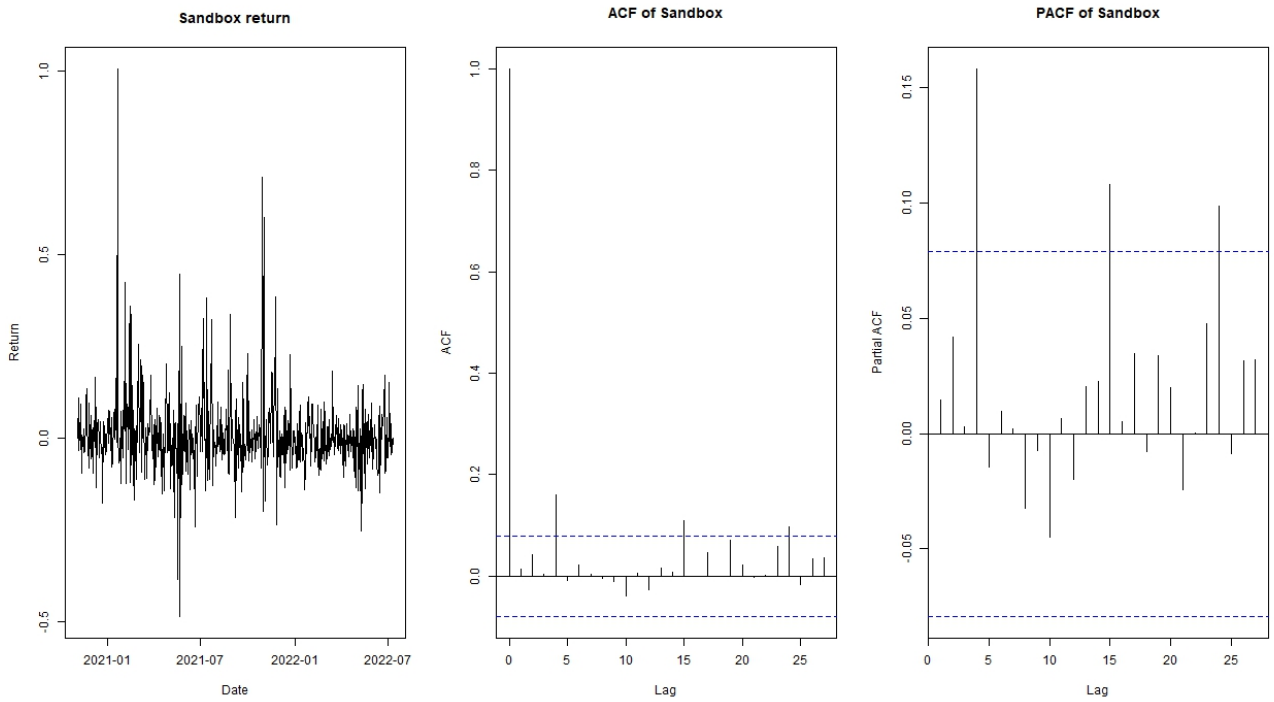

Figure 7: Residual Analysis Decentraland

stationarity tests

Test	Test Statistic	Lag Order/Par	p-value
Augmented Dickey-Fuller	-7.8655, 8	0.01	Stationary
Phillips-Perron Unit Root	-661.87, 6	0.01	Stationary
KPSS Test for Level Stationarity	0.35968, 6	0.09454	Stationary
ARCH LM-test	14.642, 12	0.2616	No ARCH effect

Table 4: Summary of Statistical Test Results

Model fitting

```
Series: d\$Return
ARIMA(0,0,0) with non-zero mean

Coefficients:
mean
0.0106
s.e.  0.0043
```

```
sigma^2 = 0.01158:  log likelihood = 498.07
AIC=-992.14   AICc=-992.12   BIC=-983.3
```

- The model is an **ARIMA(0,0,0)** with a non-zero mean, which is a simple constant model. It does not incorporate the effects of past values (p=0), differencing (d=0), or past forecast errors (q=0). This model is akin to a random walk model.
- The **mean** of the series is **0.0106**.
- The **variance** of the residuals (**sigma^2**) is **0.01158**, indicating the spread of residuals around the fitted values.
- The **log likelihood** is **498.07**, a measure of goodness-of-fit.
- The **AIC**, **AICc**, and **BIC** values are **-992.14**, **-992.12**, and **-983.3** respectively, used for comparing models.

Residual Analysis

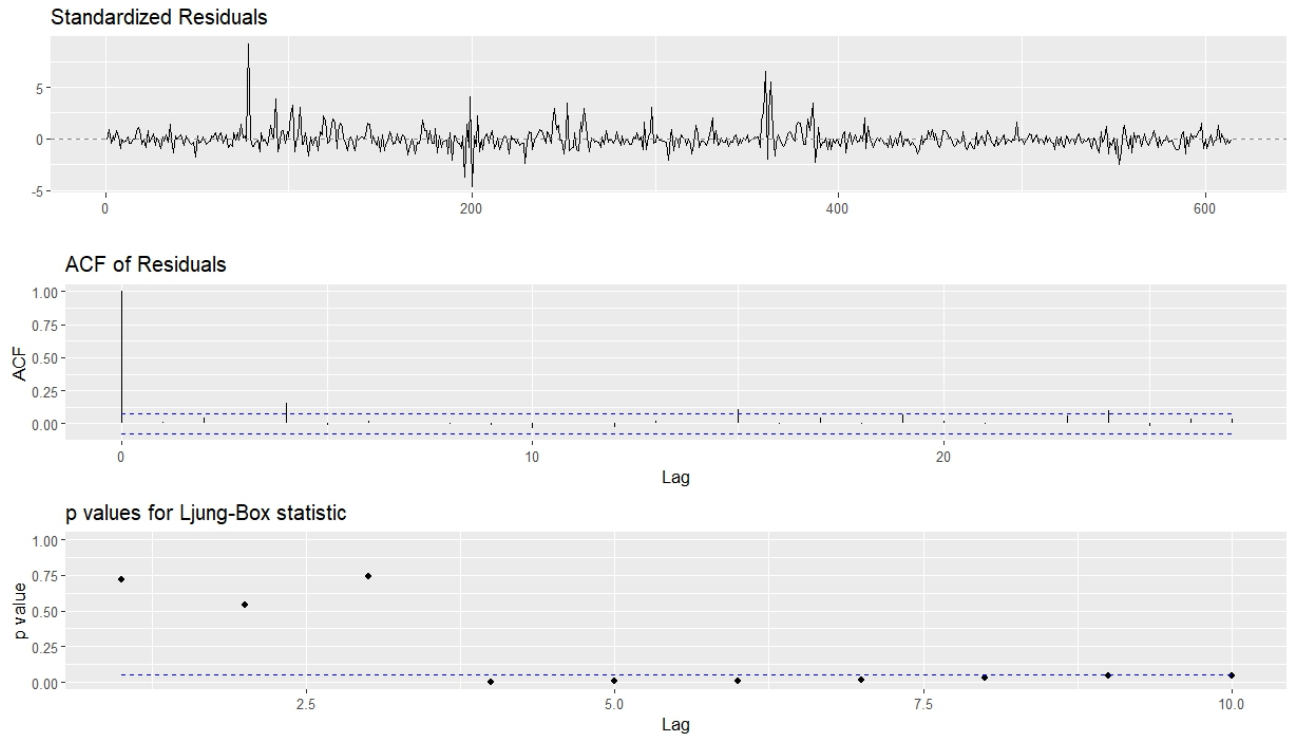

Figure 8: Residual Analysis Decentraland

- **Standardized Residuals:**
  - The top subplot shows the standardized residuals over time, which helps identify periods where the model's predictions deviate from the actual observations.
- **ACF of Residuals:**
  - The middle subplot displays the autocorrelation function (ACF) of the residuals, indicating the degree of linear predictability in the data at different lags.
- **Ljung-Box Statistic:**
  - The bottom subplot presents the p-values for the Ljung-Box statistic at various lags, with low p-values suggesting that the residuals are not independent.

Overall, the plot is used to assess the adequacy of a time series model by checking for randomness and lack of correlation in the residuals. If the standardized residuals are randomly scattered without pattern, the ACF values are within the confidence interval, and the p-values are above the significance level (usually 0.05), it suggests that the model is appropriate for the data.

6.2 Johansen's Cointegration

This study aims to explore the potential long-term relationships among the Metaverse cryptocurrencies: Decentraland, Axie Infinity, Sandbox, and Enjin. We use the Johansen cointegration test to determine if these cryptocurrencies are cointegrated, and a Vector Error Correction Model (VECM) is used to estimate any potential long-term relationships. Cointegration in the realm of cryptocurrencies is a relatively uncharted concept, making it an intriguing area for analysis. Leung and Nguyen (2018) state that the high correlation among cryptocurrencies is a reason to investigate cointegration. We intend to extend this investigation within the context of the Metaverse. The significance of this study lies in its potential implications for investors. If cointegration exists among these cryptocurrencies, it indicates a long-term relationship between them. This information can be important for the investors to make strategic trading decisions. The investment strategy focuses on identifying assets with similar price movements, which can be beneficial for traders if the assets are cointegrated. When a deviation from the long-term equilibrium occurs, investors can act on this deviation with the expectation that it will return to the long-term equilibrium. This study, therefore, provides valuable insights that could shape investment strategies within the Metaverse cryptocurrency market.

Cointegration

Cointegration, as defined by Stock and Watson (2015), occurs when two or more time series variables share a common stochastic trend. The concept of cointegration was first

introduced by Granger in 1981 and then further examined by Engle and Granger in 1987. They developed a way to analyze such data using regression. They found that just because two sets of data show a correlation does not mean they are meaningful related, especially if the data is not stationary. Applying methods meant for stationary data to non-stationary data can lead to meaningless connections, known as spurious relationships. Murray (1994) used a funny example to explain cointegration and error correction. Imagine a drunk person and her dog leaving a bar and wandering around randomly at night. The drunk's walk and the dog's path are both like random walks. Occasionally, the drunk calls for her dog, and when she does, the dog catches up. This keeps the distance between them fairly consistent. Even though both the drunk and the dog are moving randomly, the overall relationship between them stays stable in the long term. A time series is non-stationary and has a unit root if it is integrated of order 1, called an I(1) process. Financial variables commonly fall into this category. When multiple non-stationary time series move together over time, sharing a common trend, they are considered cointegrated. These cointegrated variables have a long-term relationship and might stray from it in the short term, but they tend to return to their long-term equilibrium (Brooks 2008).

Unit root testing

Firstly, we need to test the data to confirm its non-stationary, which can be done using a unit root test. In this study, we will conduct both the Dickey-Fuller test and the augmented Dickey-Fuller test to check if the variables are non-stationary and exhibit a unit-root process (Dickey and Fuller 1979). Each time series will be tested individually for a unit root. The Dickey-Fuller test is the easiest way to test for a unit root. The equation of the Dickey- Fuller test looks as follows:

$$\Delta y_t = \alpha y_{t-1} + u_t$$

Where μ_t t is a white noise variable. The Dickey-Fuller test tests the null-hypothesis of a unit root in the data, $\alpha = 0$. The alternative hypothesis is that α ¡ 0 and that there is no unit root, which means that the data is stationary. If the test statistic is more negative than the critical value, the null-hypothesis will be rejected in favor of the alternative.

The augmented Dickey-Fuller test incorporates lags into the model to check for a unit root in the time series data. Including lags helps to remove autocorrelation between the random term μ and the dependent variable. The appropriate number of lags can be determined using either Schwarz's Bayesian information criterion (SBIC) or the Akaike information criterion (AIC). Brooks (2008) notes that both criteria have their pros and cons, and neither is universally preferred. According to Brooks (2008), AIC is typically efficient but not always consistent, whereas SBIC is consistent but can be inefficient. In this study, the number of lags will be chosen based on the AIC, which is commonly used in practice (Stock and Watson 2015).

The regression that is tested in the augmented Dickey-Fuller test follows from the equation:

$$\Delta y_t = \alpha y_{t-1} + \sum_{i=1}^{p} \gamma_i \Delta y_{t-i} + u_t$$

where ρ is the number of lags of the dependent variable. The ADF-test examines the same hypotheses as the standard Dickey-Fuller test. Under the null hypothesis, y has a stochastic trend and is non-stationary, while under the alternative hypothesis, y is stationary (Brooks 2008).

Johansen's Cointegration Method

There are various methods for testing cointegration. Johansen (1988, 1991) introduced a test using maximum likelihood to determine if multiple time series establish cointegrating relationships. In this study, we use the Johansen test. There are two types of Johansen test: the maximum eigenvalue test and the trace test.

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \Gamma_2 \Delta y_{t-2} + \cdots + \Gamma_{k-1} \Delta y_{t-(k-1)} + u_t$$

where k is the number of lags, $\Pi = (\sum_{i=1}^{k} \beta_i) - I_g$ and $\Gamma = (\sum_{j=1}^{k} \beta_j) - I_g$, which are two matrices. The matrix Γ catches the short-run dynamics, while the matrix Π contains the long-run effects. In the model, there exists a set of g variables, where g is equal to or greater than two. The Johansen test focuses on the matrix Π. Each rank, r, of the matrix Π, is tested where the rank of the matrix equals the number of its eigenvalues that are different from zero.

Firstly, the trace statistic is a joint test that evaluates the null hypothesis, suggesting that the number of cointegrating vectors is zero, indicating no cointegration. Conversely, the alternative hypothesis proposes the presence of cointegration. Secondly, the max statistic tests each eigenvalue individually. The null hypothesis posits that there are r cointegrating vectors, while the alternative hypothesis suggests there are $r + 1$ cointegrating vectors.

The trace statistic is formulated as:

$$\lambda_{\text{trace}}(r) = -T \sum_{i=r+1}^{g} \ln(1 - \hat{\lambda}_i)$$

and the max statistic follows by:

$$\lambda_{\max}(r, r+1) = -T \ln(1 - \hat{\lambda}_{r+1})$$

where r is the number of cointegrating vectors under the null-hypothesis and $\hat{\lambda}_i$ is the estimated i-th ordered eigenvalue from the matrix (Brooks 2008).

If the trace statistic exceeds the critical value, the null hypothesis is rejected in favour of the alternative hypothesis. If the null hypothesis is not rejected initially, it suggests no cointegration among the variables. However, if the null hypothesis is rejected, the test continues by testing the new null hypothesis that $r = 1$ against the alternative that the cointegrating vectors are $r + 1$. The value of r increases until the null hypothesis cannot

be rejected anymore. The matrix can have a maximum of $g - 1$ ranks, meaning that if two variables are tested, it can have a maximum rank of 1 if there is cointegration. For three variables, it can have a maximum rank of 2. If the matrix is of full rank, it implies that the data is stationary (Brooks 2008).

VECM

If the cointegration tests confirm the presence of a cointegrating vector among the variables, we can employ the Vector Error Correction Model (VECM) to estimate these vectors. The equation below provides a formal representation of the VECM:

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \Gamma_2 \Delta y_{t-2} + \cdots + \Gamma_{k-1} \Delta y_{t-(k-1)} + u_t$$

The Johansen cointegration test, detailed in the previous section, investigates the matrix Π, which serves as a long-run coefficient matrix. It is formed by the multiplication of two matrices:

$$\Pi = \alpha \beta^{'}$$

where the matrix β provides the cointegrating vectors and the matrix α gives the adjustment parameters (Brooks 2008). The adjustment parameter indicates the rate at which deviations from equilibrium converge towards the long-run equilibrium, while β represents the long-run relationship among the variables.

Dynamic of the cryptocurrencies

The list of currencies included in our analysis can be found in the following table:

Name	Code	Price (USD)	Market Cap (USD)	Volume
Axie Infinity	AXI	14.64154	25054.91	176032400
Sandbox	SAN	1.177096	967.3139	211466048
Enjin	ENJ	0.5409371	903.5239	58601632
Decentraland	DEC	0.8820965	827.7193	115719336

Table 5: List of cryptocurrencies some statistics. Market capitalization as of July 10, 2022.

In total we have 614 daily price observation from November 4, 2020 until July 10, 2022. The aggregated market capitalization of our sample is around 27753 USD. By looking at the table, it appears that the crypto market within Metaverse is dominated by Axie Infinity.

The line graph provided above illustrates the daily closing price evolution of four cryptocurrencies. From the first graph, it is noticeable that the AXI exhibits a significantly different trend compared to the others, with its price soaring to much higher levels. Given this distinct behaviour, it would be beneficial to set aside the analysis of this particular cryptocurrency for

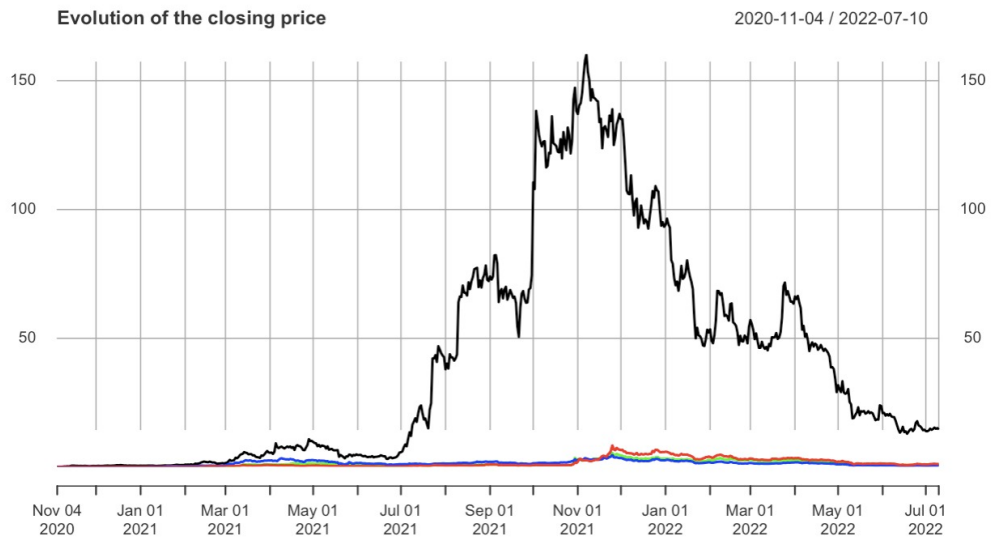

Figure 9: Evolution of Closing Price

a separate, more detailed investigation. This approach will allow us to focus on the remaining three cryptocurrencies (DEC, ENJ, SAN) which display more similar price trends, as shown it the second graph. By concentrating on these three, we can delve deeper into their respective behaviours and potential interrelationships. This focused analysis could provide valuable insights into the dynamics of these cryptocurrencies within the Metaverse, helping in the development of informed investment strategies. In the subsequent sections, we will explore these trends in more detail, aiming to uncover any underlying patterns or correlations that could inform future investment decisions. This approach ensures a comprehensive understanding of the market dynamics of these Metaverse cryptocurrencies. The multivariate time series shows a clear co-movement among the three cryptocurrencies. This suggests that they are dependent on each other not just in terms of changes over time but also in their actual levels. Therefore, it is crucial to consider cointegration when analyzing how these cryptocurrencies move together. Ignoring cointegration would lead to an incomplete understanding of their dynamics. Before conducting any cointegration analysis, it is important to ensure that all currency series are non-stationary and integrated to the same order.

Figure 10: Black: AXI, Green: DEC, Blue: ENJ, Red: SAN

Stationary test

Performing the Augmented Dickey-Fuller (ADF) test with a constant and a time trend, the null hypothesis of a unit root cannot be rejected for the individual logged prices at 90% level. The lag length k for the ADF test is determined using the Ng and Perron (1995) downtesting procedure, starting with a maximum lag of 12. However, the results show that the choice of k did not affect the ADF test outcomes, as the null hypothesis could not be rejected for any number of lagged terms in each series. In the next step, we apply differences to the time series and conducted the ADF test on the differenced data. This time, we found that the null hypothesis of non-stationarity is rejected for all indices at the 99% level, indicating that daily returns follow a stationary process. The KPSS test is used further to confirm the result from ADF. The KPSS test is an important tool that complements the ADF test, providing additional insights into the stationarity of time series data. Unlike the ADF test, which looks for unit roots to determine stationarity, the KPSS test directly examines whether the data is stationary. This different approach allows the KPSS test to detect stationarity around stochastic trends, giving a more detailed view of the underlying patterns in the data. One advantage of the KPSS test is its ability to identify stationarity in data with nonlinear trends or structural breaks, which the ADF test might miss. Since the original series need to be differenced once to achieve stationarity, we conclude that cryptocurrency prices are integrated of order one, making the vector Xt is I(1). The results of the tests are summarized in the following table.

Now that we have confirmed that all the series are integrated to the same order, we can proceed to test for cointegration.

	x_t		$\Delta\, x_t$	
	ADF	KPSS	ADF	KPSS
SAN	0.6387	<0.01	<0.01	>0.1
ENJ	0.6443	<0.01	<0.01	>0.1
DEC	0.7762	<0.01	<0.01	>0.1

Estimation results for linear VECM

The Johansen test for cointegration provides trace statistics and their corresponding critical values at different significance levels (10%, 5%, and 1%). To determine cointegration, we compare the trace statistics with the critical values. If the trace statistic is greater than the critical value at a given significance level, we reject the null hypothesis of no cointegration at that level.

r	Test	10%	5%	1%
$r \leq 2$	133.02	6.50	8.18	11.65
$r \leq 1$	333.41	15.66	17.95	23.52
$r = 0$	585.87	28.71	31.52	37.22

For $r \leq 2$, the trace statistic (133.02) is much larger than the 1% critical value (11.65), suggesting rejection of the null hypothesis of at most 2 cointegrating relations.

For $r \leq 1$, the trace statistic (333.41) is much larger than the 1% critical value (23.52), suggesting rejection of the null hypothesis of at most 1 cointegrating relation.

For $r = 0$, the trace statistic (585.87) is much larger than the 1% critical value (37.22), suggesting rejection of the null hypothesis of no cointegration.

Therefore, based on these results, there is evidence to suggest that there are at least two cointegrating relationships among your variables.

Now that we have determined the cointegration rank, we can move forward with estimating the cointegration vectors. The estimated coefficients can be found in the following table:

	SAN	ENJ	DEC
β_1	1	0	-1.6989551
β_2	0	1	-0.6875794

Table 6: Estimated cointegration vectors

To ensure a unique estimator, we normalize the j-th entry of the j-th cointegration vector to 1. This normalization means that each of the two largest currencies will have one associated vector. For instance, we can observe for β_1 that the entry for SAN is one whereas the entry for ENJ is close to zero. With these estimation results, we can plot the time series of our two stochastic trends, it is shown in the following figure:

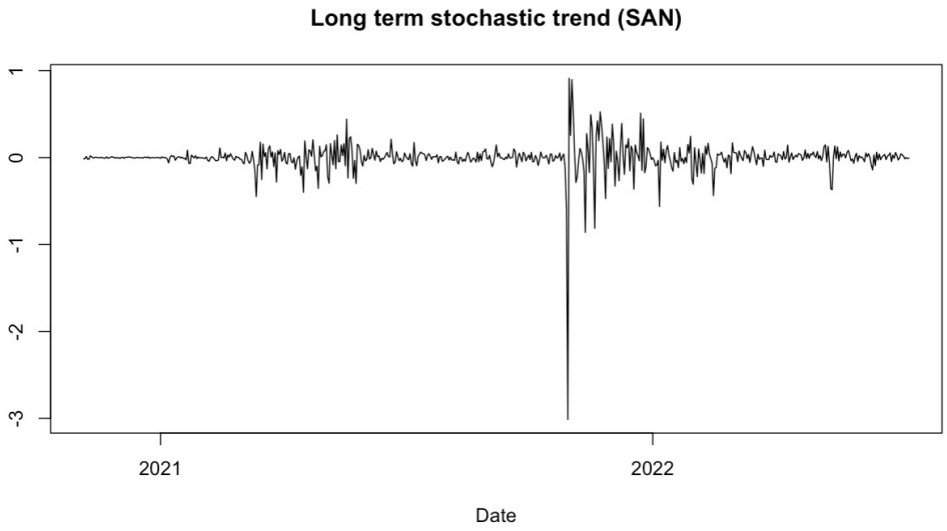

Figure 11: Long term stochastic trend (SAN)

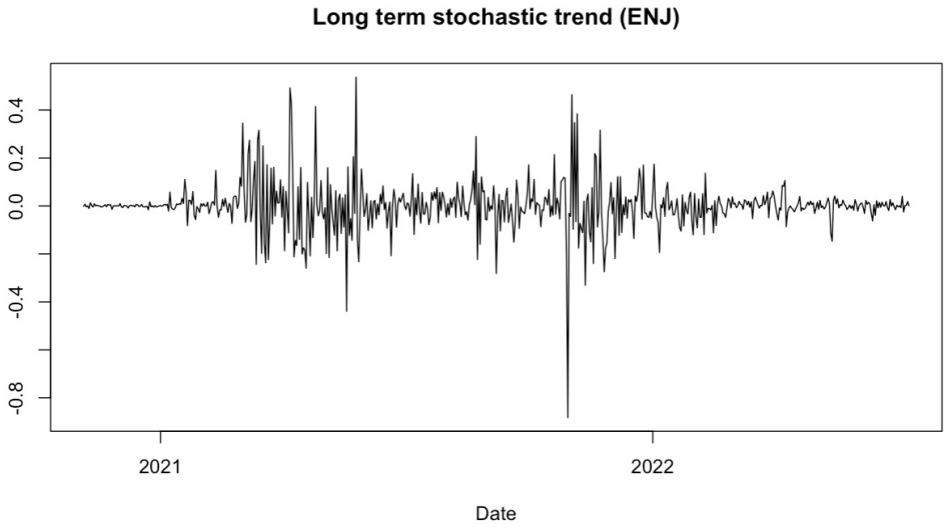

Figure 12: Long term stochastic trend (ENJ)

We can observe steady and mean-reverting stochastic trends. It suggests that the series fluctuate around a long-term average, which is 0 here. When the variable deviates significantly from this mean, there is a tendency for it to return to its average level over time. Since the results from the ADF test reject the hypothesis that these trends possess a unit root, we can

proceed to estimate the short-run parameters α and Γ. For this purpose, we select the lag order, k = 3, by Hannan-Quinn criterion (HQ). We opt for the Hannan-Quinn criterion (HQ) because it is the most parsimonious criterion among all criterion. The estimation results of our baseline VECM indicate that cointegration plays an important role for cryptocurrencies. The estimation of the loading matrix is presented in following table:

	ECT1	ECT2
SAN	-0.1788256	-0.0790017
ENJ	*0.5251891*	**-1.0311934**
DEC	*0.6793674*	0.1214925

Table 7: Estimated of the loading matrix α

Bold indicates significance of negative coefficients, *italic* indicates significance of positive coefficients, with significance at 5%, 1% and 0.1% level.

The (j, i)-th entry of the table shows how currency j is affected by error correction term i, where

$$ECT_{j,t-1} = \beta_j X_{t-1}. \tag{2}$$

ENJ and *DEC* are significantly affected by at least one stochastic trend, **SAN** is the only exceptions.

In the first error correction term, *ENJ* and *DEC* do not tend to return to the long-run equilibrium because the coefficient on the error term is positive. In the second error correction term, *ENJ* exhibits a negative sign, suggesting that the disequilibrium represented by the error correction term will by reduced over time.

The estimation results for the lagged differences can be found in the following table:

	SAN(-1)	ENJ(-1)	DEC(-01)	SAN(-2)	ENJ(-02)	DEC(-02)	SAN(-03)	ENJ(-3)	DEC(-3)
SAN	**-0.5857**	0.1399	**-0.4758**	**-0.3477**	0.1502	**-0.5331**	**-0.1414**	-0.0157	-0.0843
ENJ	**-0.3941**	-0.0199	0.0967	**-0.2448**	0.0469	-0.0663	**-0.1236**	-0.0749	0.0973
DEC	**-0.5355**	-0.121	0.1473	**-0.3767**	-0.041	0.0404	**-0.1629**	-0.0905	0.1175

Table 8: Estimated of the coefficient matrix gamma

Bold indicates significance of negative coefficients, italic indicates significance of positive coefficients, with significance at 5SAN has highly significant coefficients associated with their own lagged value and lagged value of DEC. ENJ and DEC have highly significant coefficients associated with lagged value of SAN.

Cointegration, therefore, has universal effects. The long-term connections between the SAN, ENJ and DEC imply that cryptocurrency prices are interdependent, and one can predict the behaviour of one based on information from others. Moreover, investors who want to diversify their portfolios should note that the three cryptocurrency prices in the system share

a common stochastic trend. Consequently, these markets yield similar returns in the long run. As a result, diversifying across these markets has limitations, and investors should consider incorporating other markets with lower correlation to mitigate their risks.

Conclusion

This study delved into the long-term relationship among the four Metaverse cryptocurrencies such as Decentraland, Axie Infinity, Sandbox, and Enjin. By using cointegration analysis and Vector Error Correction Models (VECM), we have seen how these digital currencies interact within the Metaverse framework. Our findings uncover substantial evidence of cointegration among these cryptocurrency prices, that is long-term relationships amonh Decentraland, Sandbox and Enjin. This suggests that investors could strategically capitalize on these connections to guide their trading decisions and craft investment strategies.

6.3 Markov Analysis

Birth-Death process: Suppose $\{X(t), t \geq 0\}$ is a continuous-time Markov chain with state space $\{0, 1, \ldots\}$, with $X(0) = a$ and with the following infinitesimal transition probabilities.

$$\begin{aligned} P_i(i+1)(h) &= P[X(t+h) = i+1 | X(t) = i] = \lambda_i h + o(h) \\ P_i(i-1)(h) &= P[X(t+h) = i-1 | X(t) = i] = \mu_i h + o(h) \\ P_{ii}(h) &= P[X(t+h) = i | X(t) = i] = 1 - (\lambda_i + \mu_i)h + o(h) \\ P_{ij}(h) &= P[X(t+h) = j | X(t) = i] = o(h) \quad \forall j \neq i+1, i-1. \end{aligned}$$

Then $\{X(t), t \geq 0\}$ is known as a birth-death process. It is to be noted that $\mu_0 = 0$, but λ_0 may or may not be 0 depending on the particular application area. For example $\lambda_0 = 0$ for Yule Furry process, but $\lambda_0 = \lambda$ in M/M/1 queuing model.

Thus, a birth and death process is a continuous-time Markov chain with states $\{0, 1, \ldots\}$ for which transitions from state i are either to state $i-1$ or to state $i+1$. The intensity rates for this process for $i \geq 0$ are given by,

$$\begin{aligned} q_i(i+1) &= \lambda_i \\ q_i(i-1) &= \mu_i \\ q_{ii} &= -(\lambda_i + \mu_i) \\ q_{ij} &= 0 \quad \forall j \neq i+1, i-1, \quad \mu_0 = 0. \end{aligned}$$

A queuing system with one service counter can be modeled as a birth death process. The state $X(t)$ of the queuing system at any time is represented by the number of people in the system at that time, that is, number of people in the queue and the one getting the service. Suppose that whenever there are i people in the system, then (i) new arrivals enter the system at an exponential rate λ_i and (ii) people leave the system at an exponential rate μ_i. Then infinitesimal transition probabilities $P_{ij}(h)$ are same as given above. The corresponding queuing model is known as M/M/1 queuing model.

From the definition of the birth-death process, it is a continuous time Markov chain with state space as a set of whole numbers and intensity rates q_{ij} as specified above. Thus, the sojourn time random variable in state $i > 0$ has an exponential distribution with rate $\lambda_i + \mu_i$, for $i = 0$, the rate is λ_0. Further if $P = [p_{ij}]$ denotes the transition probability matrix of the embedded chain, then from the transition rates we have $p_{ij} = 0$ if $j \neq i+1, i-1$, $p_{01} = 1$ and p_{ii+1} and p_{ii-1} are positive. We find the expression of these in the following.

$$\begin{aligned} p_{ii+1} &= P[\text{Process enters state } i+1 \text{ starting from } i] \\ &= P[\text{A birth occurs before a death}] \\ &= P[X < Y] \text{ where } X \sim \exp(\lambda_i) \text{ and } Y \sim \exp(\mu_i) \\ &= \frac{\lambda_i}{\lambda_i + \mu_i} \quad i \geq 0. \end{aligned}$$

Similarly,

$$\begin{aligned} p_{ii-1} &= P[\text{Process enters state } i-1 \text{ starting from } i] \\ &= P[\text{A death occurs before a birth}] \\ &= P[X > Y] \text{ where } X \sim \exp(\lambda_i) \text{ and } Y \sim \exp(\mu_i) \\ &= \frac{\mu_i}{\lambda_i + \mu_i}, \quad i \geq 1. \end{aligned}$$

It thus follows that the Markov chain embedded in a birth-death process is a random walk on a set of whole numbers, where probability of transition from i to $i+1$ is $\frac{\lambda_i}{\lambda_i+\mu_i}$, $i \geq 0$ and probability of transition from i to $i-1$ is $\frac{\mu_i}{\lambda_i+\mu_i}$, $i \geq 1$. The state 0 may or may not be an absorbing state, it depends on the value of λ_0.

Suppose $\lambda_k = \lambda$ and $\mu_k = k\mu$, then a birth-death process is known as an immigration-death process. Thus, μ denotes the death rate per individual and λ is the immigration rate, not depending on the population size.

In our study, we employ a Birth-Death process, a specific type of Markov chain, to model the price returns. This methodology is predicated on the assumption that the price returns are stationary and adhere to the Markov property. This implies that the future state of the system depends solely on the current state and is independent of the sequence of preceding events. This assumption is in alignment with the fundamental tenets of stochastic processes and the Birth-Death process. We bifurcate the price returns into two distinct states: 'birth' and 'death'. **A return greater than 0 is interpreted as a 'birth', symbolizing a positive shift or growth in the price. Conversely, a return less than 0 is interpreted as a 'death', denoting a negative shift or decline in the price.** By associating the price returns with these two states, we transform the continuous price returns into a discrete-time, discrete-state Markov chain. This transformation enables us to apply the Birth-Death process model, which is particularly suited for systems that oscillate between two states. This methodology offers a robust framework for modeling price returns and evaluating investment risks.

Methodology

- **Parameter Estimation:** The parameters of the model (e.g., λ and μ) are estimated by maximizing the likelihood function. This method finds the parameter values that make the observed data most probable.
- **Cross-Validation:** The data is divided into subsets (for example, 100 subsets). For each subset, a portion of the data is held out as a test set, and the remaining data is used as a training set. The model parameters are estimated using the training set and then used to make predictions on the test set.
- **Simulation of Returns:** For each test set, a sequence of states is simulated based on the estimated parameters (e.g., λ and μ). Random variates are generated from a binomial distribution, which models the number of "successes" in a fixed number of "trials".
- **Accuracy Calculation:** The simulated states are converted into predicted returns, which are then compared with the actual returns to calculate the Mean Squared Error (MSE). The mean of these MSE values across all subsets is taken as the final accuracy measure.

Axie Infinity

```
        death     birth
death 0.4872611 0.5127389
birth 0.5384615 0.4615385
```

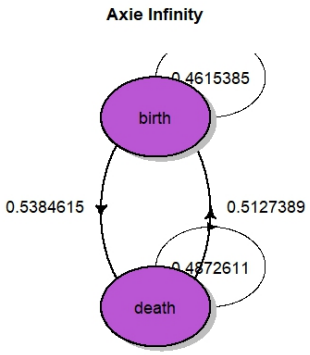

Figure 13: Axie Infinity

- **Death to Death (0.4872611):** This suggests a moderate persistence in the death state, with a 48.73% chance of remaining in this state in the next time step.
- **Death to Birth (0.5127389):** This indicates a slight inclination towards transitioning from death to birth, with a 51.27% probability.
- **Birth to Death (0.5384615):** This reveals a higher volatility when in the birth state, with a 53.85% chance of transitioning to the death state in the next time step.
- **Birth to Birth (0.4615385):** This suggests a moderate stability in the birth state, with a 46.15% chance of remaining in this state in the next time step.

Stationary Distribution

```
0.5122349 0.4877651
```

The Stationary distribution of a Markov chain represents the long-term behavior of the chain. In other words, it's the probability distribution to which the process converges over time, regardless of the initial state.

Stationary distribution of 0.5122349 and 0.4877651, this implies that in the long run, the Axie Infinity Metaverse token is expected to be in the birth state about 51.22% of the time and in the death state about 48.78% of the time.

Decentraland

```
          death     birth
death 0.4548495 0.5451505
birth 0.5159236 0.4840764
```

- **Death to Death (0.4548495):** This suggests a moderate persistence in the death state, with a 45.48% chance of remaining in this state in the next time step.
- **Death to Birth (0.5451505):** This indicates a slight inclination towards transitioning from death to birth, with a 54.52% probability.
- **Birth to Death (0.5159236):** This reveals a higher volatility when in the birth state, with a 51.59% chance of transitioning to the death state in the next time step.
- **Birth to Birth (0.4840764):** This suggests a moderate stability in the birth state, with a 48.41% chance of remaining in this state in the next time step.

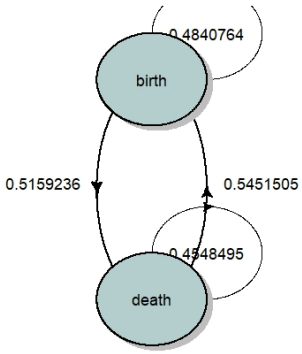

Figure 14: Decentraland

Stationary Distribution

```
0.4862277 0.5137723
```

stationary distribution of 0.4862277 and 0.5137723, this implies that in the long run, the Decentraland Metaverse token is expected to be in the birth state about 48.62% of the time and in the death state about 51.37% of the time.

Enjin

```
        death     birth
death 0.4584718 0.5415282
birth 0.5192308 0.4807692
```

- **Death to Death (0.4584718):** This suggests a moderate persistence in the death state, with a 45.85% chance of remaining in this state in the next time step.
- **Death to Birth (0.5415282):** This indicates a slight inclination towards transitioning from death to birth, with a 54.15% probability.
- **Birth to Death (0.5192308):** This reveals a higher volatility when in the birth state, with a 51.92% chance of transitioning to the death state in the next time step.

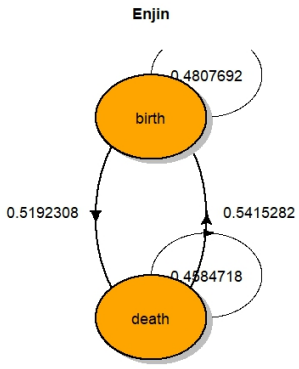

Figure 15: Enjin

- **Birth to Birth (0.4807692):** This suggests a moderate stability in the birth state, with a 48.08% chance of remaining in this state in the next time step.

Stationary Distribution

```
0.4894899 0.5105101
```

Stationary distribution of 0.4894899 and 0.5105101, this implies that in the long run, the Enjin Metaverse token is expected to be in the birth state about 48.94% of the time and in the death state about 51.05% of the time.

Sandbox

```
          death    birth
death 0.452459 0.547541
birth 0.538961 0.461039
```

- **Death to Death (0.452459):** This suggests a moderate persistence in the death state, with a 45.25% chance of remaining in this state in the next time step.
- **Death to Birth (0.547541):** This indicates a slight inclination towards transitioning from death to birth, with a 54.75% probability.
- **Birth to Death (0.538961):** This reveals a higher volatility when in the birth state, with a 53.90% chance of transitioning to the death state in the next time step.

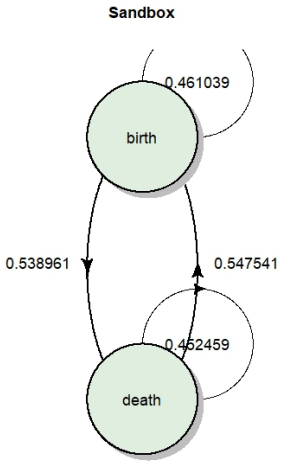

Figure 16: Enjin

- **Birth to Birth (0.461039):** This suggests a moderate stability in the birth state, with a 46.10% chance of remaining in this state in the next time step.

Stationary Distribution

```
0.4960516 0.5039484
```

- **Birth (0.4960516):** This suggests that in the long run, the system is expected to be in the death state approximately 49.61% of the time.
- **Death (0.5039484):** This indicates that in the long run, the system is expected to be in the birth state approximately 50.39% of the time.

Accuracy

Metaverse Token	Error
Axie Infinity	0.025729
Decentraland	0.053011
Enjin	0.015025
Sandbox	0.027077

Table 9: Error measure for Birth Death process

In our study, we employed a birth-death process model to predict the states of various Metaverse Tokens. The accuracy of these predictions was evaluated using the Mean Squared Error (MSE), a common statistical measure that quantifies the average squared difference between the actual and predicted values.

Our analysis revealed the following:

- **Axie Infinity:** The model demonstrated a reasonable level of accuracy for Axie Infinity, with an MSE of 0.025729. This implies that the model's predictions were generally close to the actual states for this token.
- **Decentraland:** The predictions for Decentraland exhibited a higher MSE of 0.053011, indicating a larger average discrepancy between the predicted and actual states compared to Axie Infinity.
- **Enjin:** Among the four tokens, the model's predictions for Enjin were the most precise, reflected by the smallest MSE of 0.015025. This suggests a high degree of accuracy in predicting the states for Enjin.
- **Sandbox:** The model's predictions for Sandbox yielded a moderate MSE of 0.027077, indicating a low level of prediction error.

6.3.1 Conclusion

This study evaluated a Birth - Death process model using Maximum Likelihood Estimation (MLE) and cross-validation applied to a dataset of returns for four Metaverse cryptocurrencies: Axie Infinity, Decentraland, Enjin, and Sandbox. The analysis aimed to estimate the parameters of the birth rate λ and death rate μ and assess the accuracy of the model in predicting the states of these cryptocurrencies. The evaluation process involved estimating the model parameters using MLE, followed by a 100-fold cross-validation to test the model's predictive performance. Simulations of the return states were conducted for each test set, and the accuracy of the model was measured using Mean Squared Error (MSE). The findings reveal that the Birth - Death process model, with MLE-estimated parameters, fits the data well, as indicated by low MSE values across all cryptocurrencies. Specifically, the stationary distributions indicate a balanced probability between birth and death states, reflecting their inherent volatility and stability.

6.4 Risk Analysis

Financial institutions and corporations can suffer financial losses in their portfolios or treasury department due to unpredictable and sometimes extreme movements in the financial markets. The recent increase in volatility in financial markets and the surge in corporate failures are driving investors, management, and regulators to search for ways to quantify and measure risk exposure. One answer came in the form of Value-at-Risk (VaR) being the minimum loss a portfolio will not exceed with a given probability over a specific time horizon (Jorion2007, Christoffersen2001). For a critical review of the VaR approach see (Acerbi2002). They also discuss the merits of an important and closely related risk measure, the expected shortfall. It is defined as the expected tail return conditional on a specific VaR level and provides further sensitive insights into the loss distribution, i.e., the expected portfolio loss when the portfolio value exceeds the VaR.

The VaR of some portfolio (.) may be defined as a one-sided confidence interval of expected h-periods ahead losses:

$$VaR^{(.)}_{t+h,\zeta} = \xi^{(.)}_t(1 + \bar{\xi}_{t+h,\zeta}) \tag{3}$$

where $\xi^{(.)}_t$ is the value of a portfolio in time t and $\bar{\xi}_{t+h,\zeta}$ is a time-dependent quantile of the conditional distribution of portfolio returns $\xi^{(.)}_{t+h}$ such that

$$P[\xi^{(.)}_{t+h} < \bar{\xi}_{t+h,\zeta}] = \zeta, \tag{4}$$

$$\bar{\xi}_{t+h,\zeta} = \sigma_{t+h} z_\zeta \tag{5}$$

and z_ζ is a quantile from an unconditional distribution with unit variance. In the light of the assumption of conditional normality in , we will take the quantiles z_ζ from the Gaussian distribution. the quantities $\bar{\xi}_{t+h,\zeta}$ and σ_{t+h} generally depend on the portfolio composition. For convenience, however, Our notation does not indicate this relationship. Depending on the risk averseness of the agent, the parameter ζ is typically chosen as some small probability, for instance, $\zeta = 0.005, 0.01, 0.05$.

In order to assess the performance of distinct VaR models in-sample and out-of-sample, one can employ VaR backtesting methods. Several contributions in the recent literature exploit the statistical properties of the empirical hit series. A literature review and a comparative simulation study can be found in Campbell (2006). Given ζ, a so-called hit in time $t + h$ is defined by

$$hit_{t+h}(\zeta) = 1_{\xi^{(.)}_{t+h} < VaR^{(.)}_{t+h,\zeta}}. \tag{6}$$

The indicator function 1 becomes unity if the portfolio value falls below its computed VaR and is zero otherwise. If the model is correctly specified the empirical hit rate, $\hat{\zeta} = \frac{1}{T}\sum_{t=1}^{T} hit_{t+h}(\zeta)$, for $T \to \infty$ periods converges to ζ. In the empirical part, we will

exploit this fact and compare the unconditional coverage of the estimated VaR series for the discussed volatility models.

Secondly, if the model is correctly specified, the observed hits do not provide any serial information and they are assumed to be independent. To validate the unconditional and conditional VaR coverage, Christoffersen (1998) suggests two likelihood ratio tests. These tests have been widely employed in the literature on multivariate volatility (Chib et al. 2006). A similar idea on testing the conditional coverage, Engle and Manganelli (2004) propose a dynamic quantile test assessing an autoregressive model on the series of centered hits by a Wald test for joint significance of the coefficients. A linear dependency of the hits in time contradicts the VaR model specification.

The following table provides a comparative analysis of the risk associated with two cryptocurrencies and Metaverse tokens, quantified using two key financial risk measures - Value at Risk (VaR) and Conditional Value at Risk (CVaR)

Token	**VaR**	**CVaR**
ETH	-0.07987586	-0.10390724
BTC	-0.06250000	-0.08938667
LTC	-0.07514402	-0.11226267
Axie Infinity	-0.12117114	-0.16534219
Decentraland	-0.10918033	-0.16748954
Enjin	-0.10918033	-0.16748954
Sandbox	-0.12068350	-0.17679551

Table 10: Risk Measure

- **ETH:** Ethereum exhibits a moderate level of risk, as indicated by its VaR of -0.07987586 and CVaR of -0.10390724. This suggests that potential losses on Ethereum investments could be substantial, but not extreme.
- **BTC:** Bitcoin appears to be the least risky among the listed tokens. It has the lowest VaR (-0.06250000) and a relatively low CVaR (-0.08938667), indicating that potential losses on Bitcoin investments are likely to be lower than the other tokens.
- **LTC:** Litecoin presents a moderate to high level of risk, with a VaR of -0.07514402 and a CVaR of -0.11226267. This suggests a significant potential for loss, particularly in adverse market conditions.
- **Axie Infinity:** Axie Infinity carries a high level of risk, as reflected in its VaR of -0.12117114 and CVaR of -0.16534219. This indicates a high potential for loss, especially in unfavorable market scenarios.

- **Decentraland and Enjin:** Both Decentraland and Enjin exhibit high risk levels, with identical VaR and CVaR values of -0.10918033 and -0.16748954, respectively. This suggests a high likelihood of substantial losses for these tokens.
- **Sandbox:** Sandbox is the riskiest token among those listed, with the highest VaR (-0.12068350) and CVaR (-0.17679551). This indicates a significant potential for extreme losses.

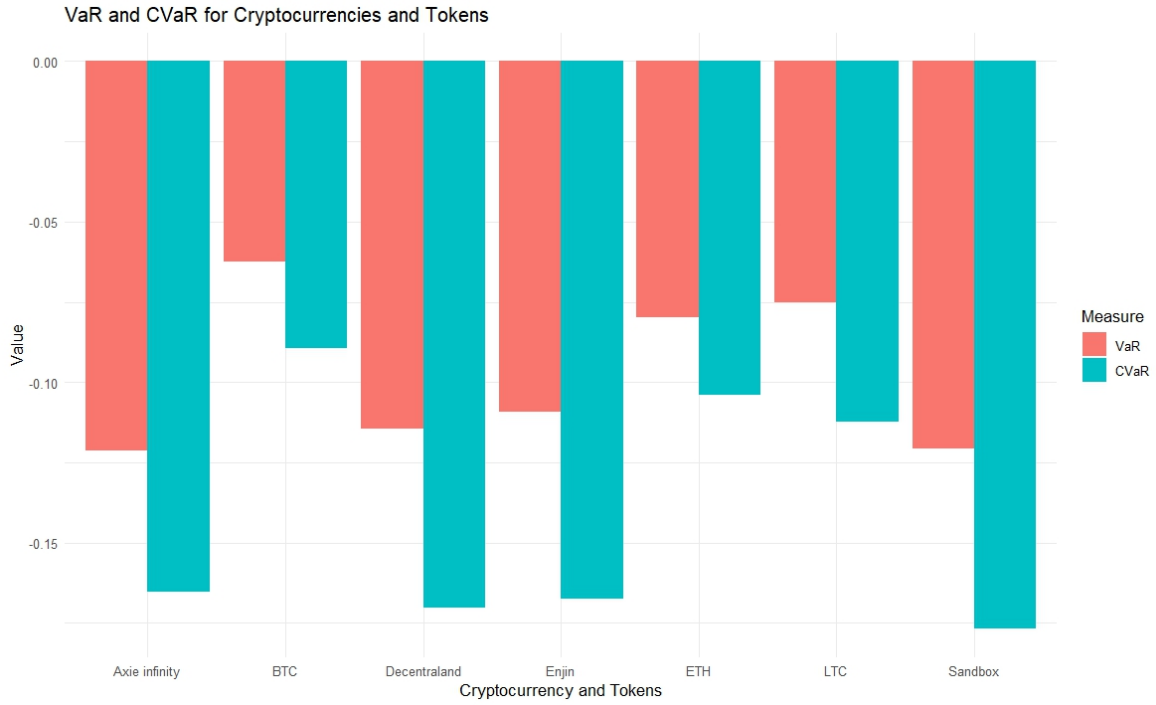

Figure 17: VaR and cVaR comparison

Cryptocurrencies vs Metaverse Tokens: From a statistical perspective, we can compare the VaR and CVaR values of cryptocurrencies and Metaverse tokens. If the bars (representing VaR and CVaR) for Metaverse tokens are longer (more negative) than those for cryptocurrencies, it indicates that investing in Metaverse tokens might be riskier than investing in the given cryptocurrencies. Conversely, if the bars for Metaverse tokens are shorter, it suggests that Metaverse tokens might be less risky.

In conclusion, based on these risk measures, Bitcoin appears to be the safest investment option among these tokens, while Sandbox carries the highest risk. However, Other factors such as market trends, investment duration,all investments carry some level of risk and past performance is not indicative of future results.

6.5 VR Tourism

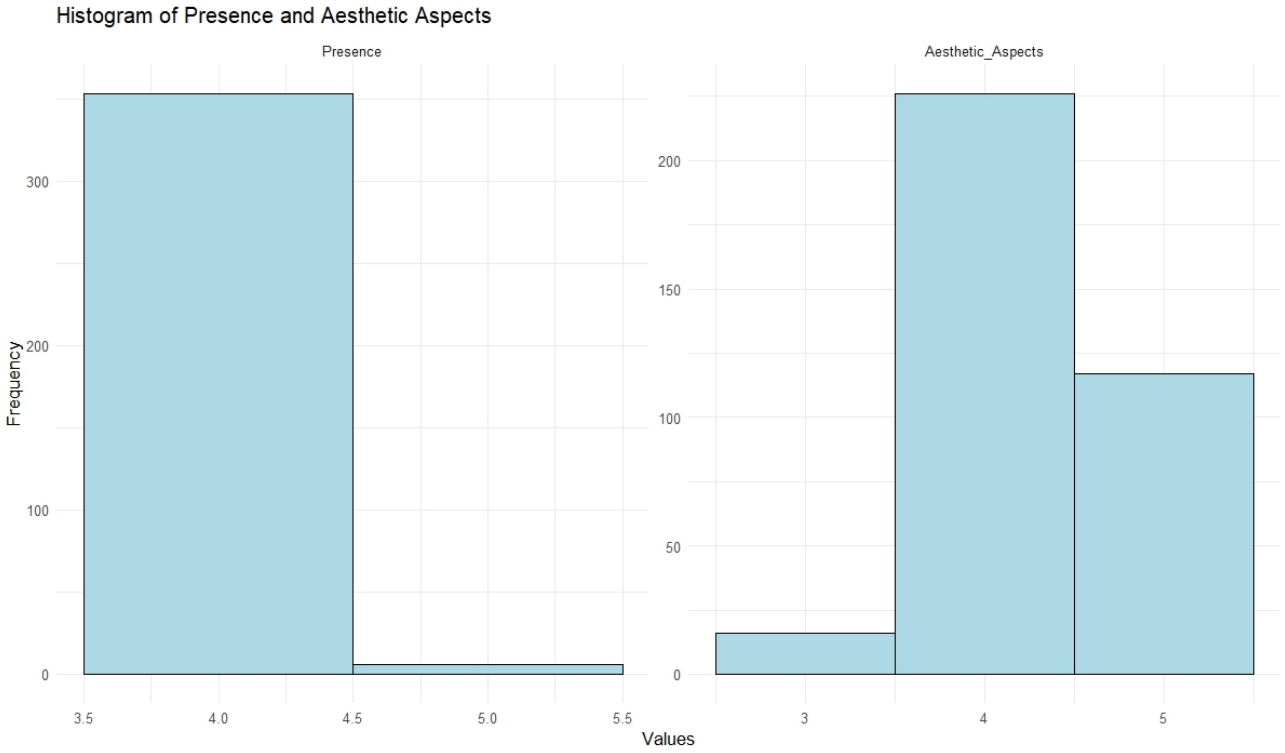

Figure 18: Presence and Aesthetic Aspects

- **Presence:** The histogram for Presence shows a significant peak at a rating of 4. This indicates that a majority of the participants felt a strong sense of presence during their virtual tourism experience. This could be interpreted as the virtual reality technology is effective in making users feel as if they were truly present in the Da Nang destination.
- **Aesthetic Aspects:** The histogram for Aesthetic Aspects shows a more varied distribution, with the majority of ratings concentrated around 4. This suggests that while most users found the aesthetic aspects of the virtual experience to be pleasing, there was more variation in these ratings compared to Presence. This could be due to individual differences in aesthetic preferences among users.

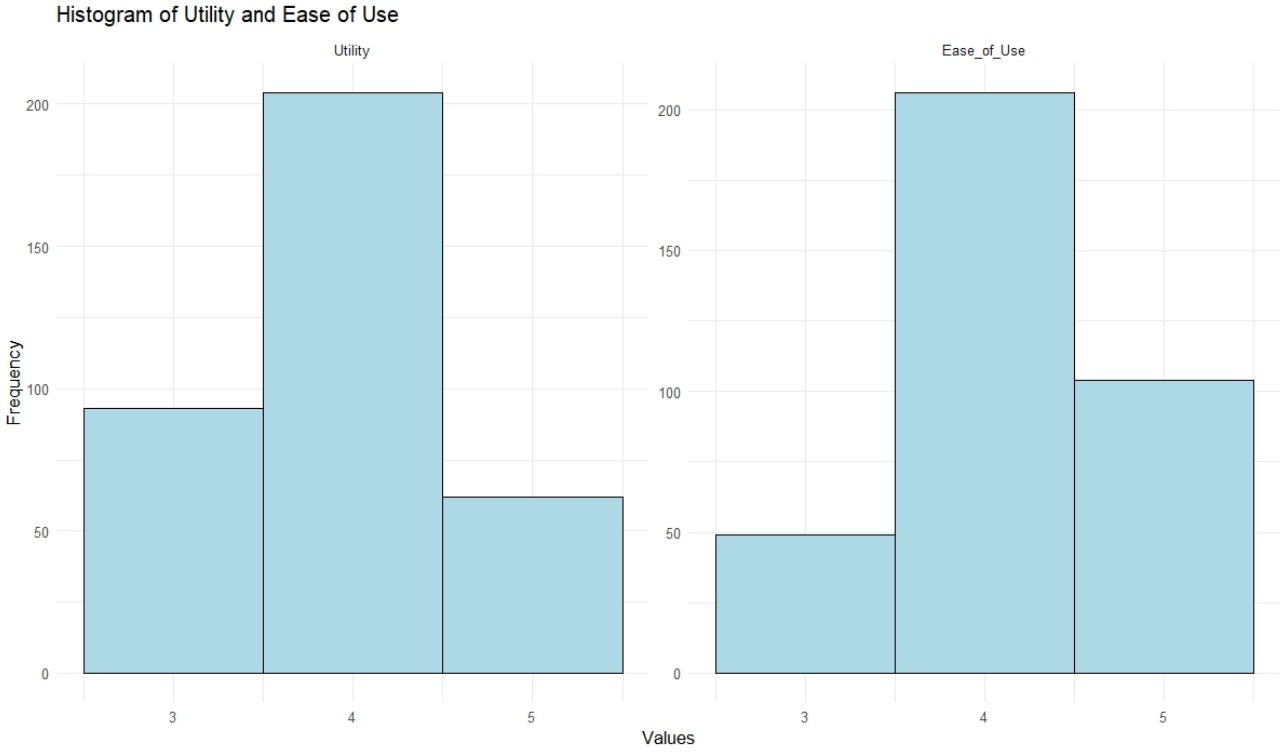

Figure 19: Utility and Ease of Use

- **Utility:** The histogram for 'Utility' shows a peak at a rating of 4, indicating that most users found the utility of the VR experience to be good. A smaller number of users rated it as moderate (3), and a few users rated it as excellent (5). This suggests that while users generally found the VR experience useful, there is room for improvement to achieve more excellent utility ratings.
- **Ease of Use:** The histogram for 'Ease of Use' shows a high frequency of ratings at value 4, indicating that a significant number of users found the VR application to be of ease of use. There were no ratings at value 4, suggesting that no users rated it as good. However, a substantial number of users rated it

6.5.1 The Wilcoxon Rank Sum Test

Suppose that we wish to derive the LMP rank test for the logistic distribution. Then we have

$$f(x) = (1 + e^{-x})^{-2}. \tag{7}$$

This can also be expressed as

$$f(x) = F'(x)(1 - F(x)). \tag{8}$$

This enables us to explicitly calculate

$$-\frac{f'(x)}{f(x)} = (2F(x) - 1). \tag{9}$$

Then $T(R)$ is a linear function of

$$W = \sum_{j=1}^{n_2} R_{n1+j}, \tag{10}$$

the rank sum of the second random sample.Besides being the LMP rank test for the logistic distribution, the Wilcoxon test has a very high asymptotic relative efficiency (ARE) of 0:95 against the two sample t-test for the Normal distribution. It is obvious that it is very simple to apply. Its mean and variance under the null hypothesis can be obtained from theorem as

$$E_{H_0}(W) = \frac{n_2(n+1)}{2} \tag{11}$$

$$Var_{H_0}(W) = \frac{n_1 n_2(n+1)}{12} \tag{12}$$

If there are ties in the observations, say in k groups with $t_1, t_2, ..., t_k$ observations in each group, then assigning the midrank to each of the observations in the tied group reduces the variance to

$$Var_{H_0,T}(W) = \frac{n_1 n_2(n+1)}{12} - \frac{n_1 n_2}{12n(n-1)} \sum_{i=1}^{k} t_i(t_i^2 - 1). \tag{13}$$

Under H_0,

$$W^* = \frac{W - E_{H_0}(W)}{\sqrt{Var_{H_0}(W)}} \underset{d}{\rightarrow} N(0,1) \quad \text{as} \quad n \rightarrow \infty. \tag{14}$$

This allows us to use standard normal tables for the choice of critical points when the sample size is large.

```
Wilcoxon rank sum test with continuity correction

Data:  Presence by Gender
W = 14615, p-value = 0.00023
Alternative hypothesis: true location shift is not equal to 0
```

P-value=0.0023 less than 0.05 is typically taken to indicate strong evidence against the null hypothesis. In this case, the p-value is much less than 0.05, which suggests that there is a significant difference in Presence ratings between the two Gender groups.

6.5.2 The Kruskal-Wallis Test

If we arrange all the $n = \sum_{i=1}^{k} n_i$ observations from the smallest to the largest and assign rank R_{ij} to X_{ij}, the j^{th}; $j = 1; 2; \ldots; n_i$ observation from the i^{th} sample, then the Kruskal-Wallis statistic is defined as

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} n_i \left(\bar{R}_{i.} - \frac{n+1}{2} \right)^2 \tag{15}$$

Here $\bar{R}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} R_{ij}$, the average rank of the observations from the i^{th}; $i = 1; 2; \ldots; k$ sample. This can be seen to be the rank version of the usual ANOVA defined as (Sum of squares due to the treatment)/(Total sum of squares) except that the total sum of squares here is a non-stochastic quantity, being the sum of squares of difference of the numbers $1; 2; \ldots; n$ from their average $\frac{n+1}{2}$. A slightly simpler equivalent expression for H is

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} n_i \bar{R}_{i.}^2 - 3(n+1) \tag{16}$$

Under H_0 we expect that the average ranks of k samples will be close to $\frac{n+1}{2}$, the overall average of ranks. On the other hand, if the locations of the populations from which these random samples are obtained are different then that should show up in the complete ranking. Then the average ranks will be (at least in some cases) away from the overall mean rank. This would yield higher values of the statistic H under the alternative H_1 than under H_0. Therefore, the suggested test is

$$\text{Reject } H_0 \text{ if } H > H_\alpha \tag{17}$$

where H_α is the upper α point of the null distribution of H. The null distribution, being based on the null rank statistics distribution is known and is tabulated for small sample

sizes. Many leading software packages include it. The asymptotic null distribution of H as $\min(n_1; n_2; \ldots; n_k)$ tends to infinity, is chi-squared with $k-1$ degrees of freedom. This is so because of the linear constraint

$$\sum_{i=1}^{k} n_i \bar{R}_{i.} = \frac{n(n+1)}{2} \tag{18}$$

on the basic random quantities $\bar{R}_{1.}; \bar{R}_{2.}; \ldots; \bar{R}_{k.}$ which are involved in the statistic. Hence the critical point $\chi^2_{k-1,\alpha}$ should be used for moderately large sample sizes. As in the two sample case, we adjust the statistic H in case there are tied observations in the data.We instead consider

$$H^* = \frac{H}{1 - \sum_{j=1}^{m} \frac{t_j^3 - t_j}{n^3 - n}} \tag{19}$$

To take care of the reduced variation in the possible values of the statistic H, we consider the following. Here t_i is the number of observations tied in the i^{th} group, there being m such groups. Of course, the distribution of H^* is to be interpreted as its conditional distribution given the number of tied groups and the number of observations in them, that is, $t_1; t_2; \ldots; t_m$.

```
Kruskal-Wallis rank sum test

Data:  Utility by Income
Kruskal-Wallis chi-squared = 3.713, df = 3, p-value = 0.2942
```

- P-value is greater than 0.05, which suggests that there is not enough evidence to reject the null hypothesis. This means that the differences in 'Utility' across the 'Income' groups are not statistically significant.
- This could imply that regardless of income level, potential tourists find similar utility in the virtual reality experience of Da Nang destination.

```
Kruskal-Wallis rank sum test

Data:  Presence by Age
Kruskal-Wallis chi-squared = 3.0092, df = 2, p-value = 0.2221
```

The p-value is 0.2221, which is greater than 0.05. This means that if the null hypothesis is true (i.e., there is no difference in the distributions of 'Presence' ratings across different 'Age' groups)

6.5.3 Spearman's Rank Correlation

We use Spearman's Rank Correlation to identify is there any relationship between the flow of states in the VR experience and the enjoyment reported by users. If a significant correlation is found this could inform the design decision of the VR experience. If a higher flow of states correlates with greater enjoyment, efforts could be made to enhance the flow of states in the VR experience.

```
Spearman's rank correlation rho

Data:  df\$Flow\_of\_States and df\$Enjoyment
S = 4680594, p-value = 1.041e-14
Alternative hypothesis: true rho is not equal to 0
Sample estimates:
rho
0.393023
```

- A rho value of 0.393023 suggests a moderate positive correlation between 'Flow of States' and 'Enjoyment'. This means that as the flow of states increases, the enjoyment also tends to increase.
- The p-value is 1.041e-14, which is much less than the typical significance level of 0.05. This suggests that the observed correlation is statistically significant.
- This could mean that improving the flow of states in the VR experience might enhance user enjoyment.

In our case, we use this test to understand if there is a correlation between the age of the user and their rating of how easy to use virtual reality. The value of Spearman's rank correlation ranges between -1 to 1 indicating the perfect correlation and -1 indicates the perfect negative correlation.

```
Spearman's rank correlation rho

Data:  df\$Age and df\$Ease\_of\_Use
S = 7583325, p-value = 0.754
Alternative hypothesis: true rho is not equal to 0
Sample estimates:
rho
0.01659829
```

- A rho value of 0.01659829 suggests a very weak positive correlation between age and ease of use. This means that as age increases, the ease of use rating slightly increases, but the correlation is so weak it might not be meaningful.
- The p-value is 0.754, which is much greater than the typical significance level of 0.05. This suggests that the observed correlation could very likely be due to chance, and we fail to reject the null hypothesis that there is no correlation between age and ease of use.

6.5.4 Logistic Regression

Since our predictor $G(x)$ takes values in a dis- crete set G, we can always divide the input space into a collection of regions labeled according to the classification. Methods that model the posterior probabilities That model the posterior probabilities $Pr(G = k|X = x)$ are also in this class. Clearly, if either the $\delta_k(x)$ or $Pr(G = k|X = x)$ are linear in x, then the decision boundaries will be linear.

Actually, all we require is that some monotone transformation of δ_k or $Pr(G = k|X = x)$ be linear for the decision boundaries to be linear. For example, if there are two classes, a popular model for the posterior probabilities is

$$Pr(G = 1|X = x) = \frac{\exp(\beta_0 + \beta^T x)}{1 + \exp(\beta_0 + \beta^T x)}, \tag{20}$$

$$Pr(G = 2|X = x) = \frac{1}{1 + \exp(\beta_0 + \beta^T x)}. \tag{21}$$

Here the monotone transformation is the logit transformation: $\log[p/(1-p)]$, and in fact we see that

$$\log\left(\frac{Pr(G = 1|X = x)}{Pr(G = 2|X = x)}\right) = \beta_0 + \beta^T x. \tag{22}$$

The decision boundary is the set of points for which the log-odds are zero, and this is a hyperplane defined by $\{x|\beta_0 + \beta^T x = 0\}$. We discuss two very popular but different methods that result in linear log-odds or logits: linear discriminant analysis and linear logistic regression. Although they differ in their derivation, the essential difference between them is in the way the linear function is fit to the training data.

```
Call:
NULL

Coefficients:
                   Estimate  Std.Error z value Pr(>|z|)
(Intercept)       -13.1181     4.7559  -2.758  0.00581 **
Presence           -0.8909     1.0895  -0.818  0.41352
Aesthetic_Aspects   1.7230     0.3101   5.557 2.75e-08 ***
Utility             1.4102     0.2761   5.107 3.27e-07 ***
Ease_of_Use         0.7425     0.2660   2.791  0.00525 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 473.31  on 358  degrees of freedom
Residual deviance: 362.77  on 354  degrees of freedom
AIC: 372.77

Number of Fisher Scoring iterations: 5
```

- **Intercept:** The intercept term is -13.1181, which represents the log-odds of ‘Intention to Travel’ when all predictors are zero.
- **Presence:** The coefficient for ‘Presence’ is -0.8909, but it’s not statistically significant (p=0.41352), suggesting ‘Presence’ may not be a crucial factor in predicting ‘Intention to Travel’.
- **Aesthetic Aspects:** The coefficient for ‘Aesthetic Aspects’ is 1.7230 and is statistically significant (p¡0.001), indicating that higher aesthetic aspects are associated with higher odds of intending to travel.
- **Utility:** The coefficient for ‘Utility’ is 1.4102 and is statistically significant (p¡0.001), suggesting that higher utility ratings are associated with higher odds of intending to travel.
- **Ease of Use:** The coefficient for ‘Ease of Use’ is 0.7425 and is statistically significant (p=0.00525), indicating that easier-to-use VR experiences are associated with higher odds of intending to travel.

6.5.5 Chi-Square Test of Independence

The Chi-Square Test of Independence is a statistical test used to determine if there is a significant association between two categorical variables. In our case, we are looking at the relationship between Intention to Travel and Gender. The test can help identify if there's a relationship between the gender of the user and their intention to travel after experiencing the VR technology. Understanding the relationship between gender and intention to travel can also inform marketing strategies. If one gender has a stronger intention to travel, marketing efforts could be targeted towards this group.

```
Pearson's Chi-squared test with Yates' continuity correction

Data:  contingency_table
X-squared = 1.9289, df = 1, p-value = 0.1649
```

The p-value is greater than the typical significance level of 0.05. This means that we fail to reject the null hypothesis, which states that there's no association between 'Gender' and 'Intention to Travel'. In other words, the test did not find a statistically significant difference in the 'Intention to Travel' between different 'Gender' groups. The result suggests that 'Gender' does not have a statistically significant effect on 'Intention to Travel' based on the data we have.

6.5.6 Conclusion

The evaluation of the VR tourism experience in Da Nang reveals several key insights about user perceptions and factors influencing their intention to travel. Overall, the VR technology effectively immerses users in the Da Nang destination, as evidenced by the strong sense of presence felt by the majority of participants. The aesthetic aspects of the VR experience were generally pleasing, although individual preferences varied more widely. Users found the VR experience to be quite useful and easy to use, though there is room for improvement to achieve higher utility ratings. Statistical analyses further support these findings, with significant differences in Presence ratings between genders, while no significant differences were found for utility across income levels or for presence ratings across age groups. A moderate positive correlation between the flow of states and user enjoyment suggests that enhancing the flow of the VR experience could improve overall enjoyment. Logistic regression analysis indicates that aesthetic aspects, utility, and ease of use are significant predictors of the intention to travel, whereas presence is not a crucial factor. Additionally, gender does not significantly influence the intention to travel. In conclusion, the VR tourism experience in Da Nang is generally well-received, providing a strong sense of presence, pleasing aesthetics, good utility, and ease of use.

7 Virtual Reality User Experience Dataset: Analyzing Immersion and Physiology

7.1 Introduction

As Virtual Reality (VR) technology is rapidly growing, experts think it could be worth $26.89 billion by 2022. However, this impressive growth comes with a downside. Users have reported feeling nauseous, experiencing dizziness, vomiting, and even breaking out in cold sweats after using VR. McCauley called this phenomenon as VR sickness (McCauley & Sharkey, 1992). Usually these unpleasant symptoms are referred to as motion sickness, and they consequently motion sickness limits the VR community in the full adaptation of this immersive technology. Many researchers have used VR in experiments to understand why it causes discomfort (Duh et al., 2004). Based on their findings, some studies have provided guidelines for reducing VR sickness (L. Rebenitsch, 2015). Previous studies have revealed various factors linked to motion sickness caused by VR systems (G. Gonçalves et al., 2018). These factors include, but are not limited to, differences between genders, the type of virtual environment, the VR experiences themselves, the visual properties, the brightness of the virtual environment, and previous experiences with motion sickness. Given the prevalence of VR sickness and its impact on user experience, the objective of this report is to identify the key factors associated with motion sickness in VR users, examining variables such as gender, virtual environment characteristics, VR experiences, visual properties, and individual susceptibility to motion sickness.

7.1.1 Description of the data

The dataset contains information about 1000 users experiences in VR environments, including physiological responses like heart rate and skin conductance, emotional states, and user preferences. Its purpose is to improve VR technology by analyzing user experiences, aiming to enhance VR design, user comfort, and customization. Here are the variables in the dataset:

- **User ID**: A unique identifier for each user in the VR experience.
- **Age**: The age of the user during the VR experience, recorded as an integer.
- **Gender**: The gender of the user, categorized as “Male”, “Female”, or “Other”.
- **VR Headset Type**: The type of VR headset used by the user, such as “Oculus Rift”, “HTC Vive”, or “PlayStation VR”.

- **Duration**: The duration of the VR experience in minutes.
- **Motion Sickness Rating**: The user's self-reported rating of motion sickness experienced during the VR experience, measured on a scale of 1 to 10.
- **Immersion Level**: A measure of how immersed the user felt during the VR experience, recorded on a scale of 1 to 5, with 5 indicating the highest level of immersion.

7.2 Data analysis

First, we need to check if motion sickness is distributed normally. It will guide us if we should use a parametric test, ANOVA, or non-parametric test, like Kruskal-Wallis test. After that, we will see if motion sickness levels are different for different genders and VR headset types. This helps us figure out if being male or female or other, or using different VR headsets, affects the level of sickness people felt while experiencing VR. We perform Shapiro-Wilk test for normality. Given below the result of normality test of "Motion sickness" within each Gender group and VR headsets type:

Gender group	p-value
Male	<0.01
Female	<0.01
Other	<0.01

Table 11: Normality test of "Motion sickness" within each Gender group

VR headset	p-value
HTC Vive	<0.01
Oculus Rift	<0.01
PlayStation VR	<0.01

Table 12: Normality test of "Motion sickness" within each VR headset type

The first table shows the p-values of the normality test for "Motion sickness" within each gender group. All p-values are less than 0.01, which is usually considered statistically significant. This suggests that the distribution of "Motion sickness" is significantly different from a normal distribution within each gender group. The second table shows the p-values of the normality test for "Motion sickness" within each VR headset type. Again, all p-values are less than 0.01, suggesting that the distribution of "Motion sickness" is significantly different from a normal distribution within each VR headset type. In other words, these results suggest that the experiences of motion sickness are not normally distributed across different genders and among users of various VR headsets. Since there are three groups for both Gender and

VR headset, the Kruskal-Wallis test is appropriate to check the variation of motion sickness across different group.

7.2.1 Kruskal-Wallis test

In our analysis, we use the Kruskal-Wallis test to investigate whether there were significant differences in motion sickness levels across different genders and VR headset types. In other words, this test allowed us to determine if gender or VR headset type had a notable effect on motion sickness experiences among users. The table below shows the result after performing Kruskal-Wallis test:

Table 13: Kruskal-Wallis rank sum test

Data	Degree of freedom	Chi-squared	p-value
Motion sickness by gender	2	0.29582	0.8625
Motion sickness by VR headset	2	2.4357	0.2959

Interpretation

Both p-value are greater than 0.05, which means we fail to reject the null hypothesis. Again the null hypothesis of the Kruskal-Wallis test is that the population medians of all groups are equal. So, the result suggests that there is no statistically significant difference in motion sickness levels across different VR headsets and Gender.

7.2.2 Random Forest

Random forests provide an improvement over bagged trees by way of a random small tweak that decorrelates the trees. As in bagging, we build a number forest of decision trees on bootstrapped training samples. But when building these decision trees, each time a split in a tree is considered, a random sample of m predictors is chosen as split candidates from the full set of p predictors. The split is allowed to use only one of those m predictors. A fresh sample of m predictors is taken at each split, and typically we choose m $\approx\sqrt{p}$ that is, the number of predictors considered at each split is approximately equal to the square root of the total number of predictors.

In other words, in building a random forest, at each split in the tree, the algorithm is not even allowed to consider a majority of the available predictors. This may sound crazy, but it has a clever rationale. Suppose that there is one very strong predictor in the data set, along with a number of other moderately strong predictors. Then in the collection of bagged trees, most or all of the trees will use this strong predictor in the top split. Consequently, all of the bagged trees will look quite similar to each other. Hence the predictions from the bagged trees will be highly correlated. Unfortunately, averaging many highly correlated quantities

does not lead to as large of a reduction in variance as averaging many uncorrelated quantities. In particular, this means that bagging will not lead to a substantial reduction in variance over a single tree in this setting. Random forests overcome this problem by forcing each split to consider only a subset of the predictors. Therefore, on average $\frac{(p-m)}{p}$ of the splits will not even consider the strong predictor, and so other predictors will have more of a chance. We can think of this process as decorrelating the trees, thereby making the average of the resulting trees less variable and hence more reliable.

After examining the scatter plot of the data, we observed that there is neither a clear linear relationship nor a discernible pattern between motion sickness and the other variables. This suggests that traditional linear regression models may not be suitable for capturing the complex relationship between motion sickness and the predictors. Therefore, we will perform Random Forest as our analysis method. Random Forest is capable of capturing non-linear relationships and interactions in the data, making it well-suited for exploring the potential influences of various factors on motion sickness levels in VR users. The figure below shows variable importance plot from Random Forest

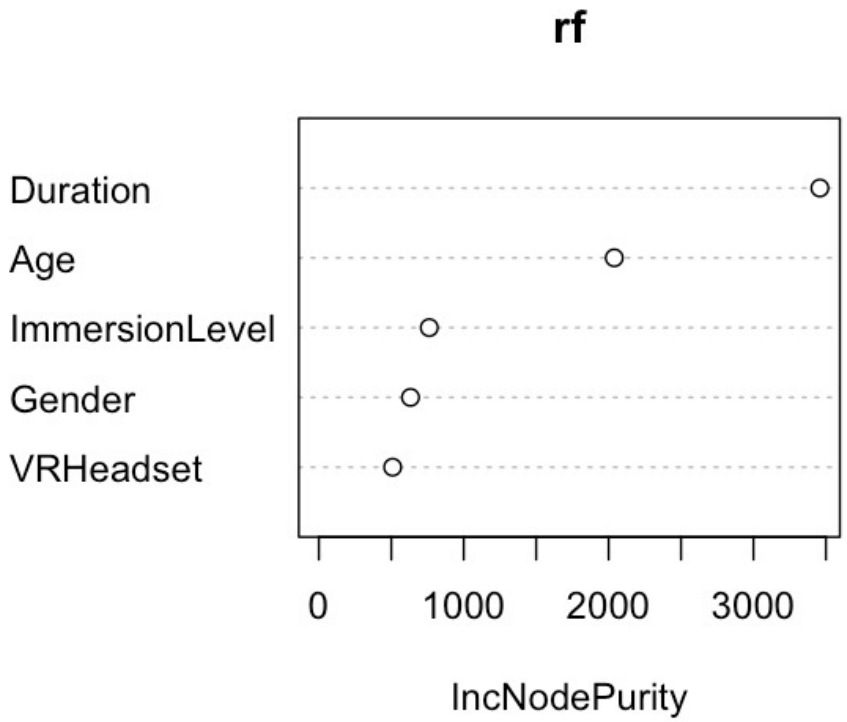

Figure 20: variable importance plot

Based on the Figure above, it appears that duration of the VR experience is the most important predictor of motion sickness. This is indicated by its highest Increase in Node Purity (IncNodePurity), a measure of the total reduction in the criterion brought about by that variable.

- **Duration:** This is the most important variable in predicting motion sickness. It suggests that the length of the VR experience has the most significant impact on motion sickness.
- **Age:** This is the second most important variable. It indicates that the age of the user also plays a significant role in predicting motion sickness during VR experiences.
- **ImmersionLevel:** This variable has a moderate level of importance. It suggests that the level of immersion can also influence motion sickness, but not as much as 'Duration' or 'Age'.
- **VRHeadset:** This variable has relatively low importance, indicating that the type of VR headset used has a lesser impact on motion sickness.
- **Gender:** This is the least important variable in the model. It suggests that the user's gender has the least influence on motion sickness during VR experiences.

Note that based on the results of the Kruskal-Wallis test and the variable importance analysis conducted using Random Forest, it appears that gender, VR headset type, have limited impact on motion sickness levels. Therefore, we have decided to exclude these variables and Immersion Level from further analysis. Instead, we will focus on exploring the potential influence of "Duration" and "Age" on motion sickness experiences in VR users.

7.3 Partial dependence of motion sickness on duration

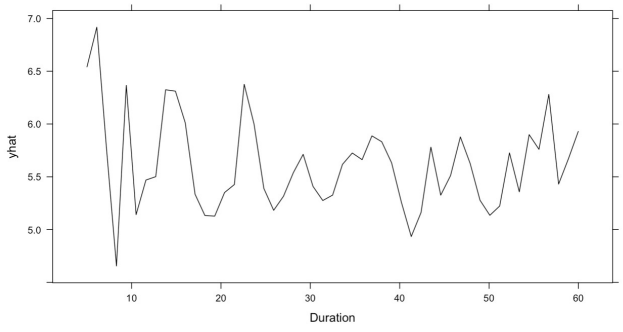

Figure 21: Partial dependence of motion sickness on duration

Interpretation:

- For very short durations (around 10 minutes), there is a sharp increase in the predicted motion sickness level. This suggests that even a small amount of time spent in VR can lead to a significant increase in motion sickness for some users.
- However, as the duration increases beyond this point, the predicted motion sickness level decreases sharply. This could indicate that users might initially experience discomfort (resulting in motion sickness) when they start using VR, but as they spend more time and get accustomed to the VR environment, their motion sickness level decreases.
- Beyond a certain point (around 20 minute), the motion sickness level starts fluctuating. This could be due to individual differences among users - some users might continue to experience decreased motion sickness with longer durations, while others might experience an increase. Note: These interpretations are based on the assumption that all other variables are held constant.

7.4 Partial dependence of motion sickness on age

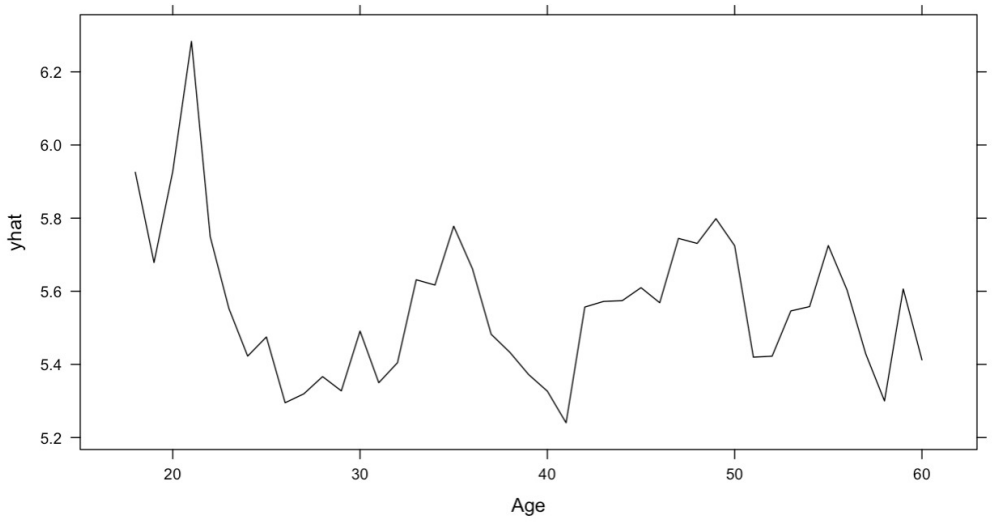

Figure 22: Partial dependence of motion sickness on duration

Interpretation:

- Young Age (around 20): The predicted motion sickness is highest around the age of 20. This suggests that younger users might experience more motion sickness during VR experiences.
- Middle Age (around 30): As age increases from 20 to about 30, the predicted motion sickness decreases. This could indicate that users in this age range might adapt better to VR and experience less motion sickness.
- Older Age (30 to 50): From age 30 to about 50, the predicted motion sickness fluctuates with a general increase. This suggests that the relationship between age and motion sickness might be influenced by other factors in this age range.
- Age 50 and above: After age 50, the predicted motion sickness increases a bit and then declines again. This could suggest that older users might experience less motion sickness, possibly due to different usage patterns or better adaptation to VR.

7.5 Interaction between "Duration" and "Age" on motion sickness

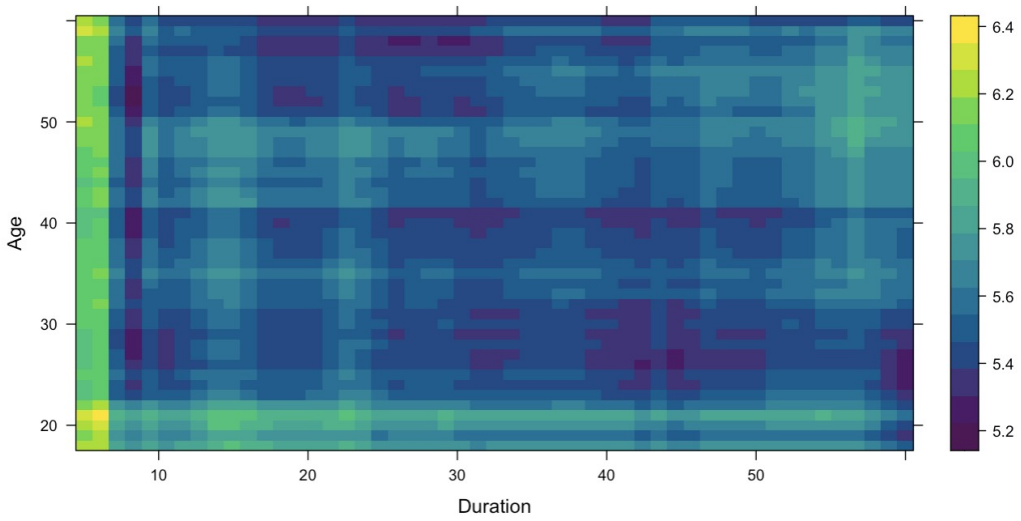

Figure 23: Interaction between "Duration" and "Age" on motion sickness

Interpretation:

- Duration and younger and older users: The lighter colours in from the bottom left to the topright left of the plot suggest that users experiencing shorter VR durations tend to have higher predicted motion sickness levels. This could indicate that users are more susceptible to motion sickness during initial exposure to VR.
- Longer Duration: As we move rightwards along the 'Duration' axis, the colours fluctuate between light and dark. This suggests that the effect of duration on motion sickness is not straightforward and might depend on other factors. For instance, longer durations might lead to more motion sickness in some users, while others might adapt better to VR over time.

Conclusion

In conclusion, our analysis provides valuable factors influencing motion sickness levels in virtual reality users. Despite the rapid growth of VR technology, motion sickness remains a significant challenge, limiting the full adaptation of this immersive technology. Our findings indicate that gender and VR headset type have limited impact on motion sickness levels, as confirmed by both the Kruskal-Wallis test and variable importance analysis using Random Forest. The most influential predictors of motion sickness are the duration of the VR experience and the age of the user. Specifically, shorter durations and younger age (partial dependence) are associated with higher predicted motion sickness levels. However, as the duration of VR experience increases, motion sickness levels tend to decrease, suggesting a potential adaptation effect over time. Additionally, users in the middle age range (around 30) tend to experience less motion sickness compared to younger and older age groups.

8 Overall Conclusion

The comprehensive evaluation of various aspects of Metaverse cryptocurrencies and VR technologies provides valuable insights into their dynamics and user experiences. The analysis of Metaverse cryptocurrencies, specifically Decentraland, Axie Infinity, Sandbox, and Enjin, revealed significant long-term relationships among their prices. The cointegration and Vector Error Correction Models (VECM) demonstrated that these cryptocurrencies are interdependent over the long term. This suggests that investors can use these relationships to make more informed trading decisions and develop strategic investment plans. The Birth-Death process model, applied to the returns of these cryptocurrencies, showed that it fits the data well, as evidenced by low Mean Squared Error (MSE) values. The stationary distributions indicated a balanced probability between birth and death states, reflecting the inherent volatility and stability of these digital assets. In the realm of VR tourism, the evaluation of the VR experience in Da Nang highlighted that users felt a strong sense of presence and generally found the aesthetic aspects pleasing, despite some variation in individual preferences. The utility and ease of use of the VR experience were positively rated, with room for improvement. Statistical analysis indicated that gender differences influenced presence ratings, but no significant differences were found for utility across income levels or presence ratings across age groups. A moderate positive correlation between the flow of states and user enjoyment suggests that enhancing the flow can improve overall enjoyment. Furthermore, higher aesthetic aspects, utility, and ease of use were significant predictors of the intention to travel, whereas presence was not crucial. Lastly, the analysis of motion sickness in VR users identified duration of VR experience and age as the most significant predictors. While gender and VR headset type had limited impact, shorter durations and younger users were more prone to motion sickness. However, increased exposure led to adaptation, reducing motion sickness over time. Overall, these studies collectively enhance our understanding of the financial dynamics within the Metaverse ecosystem and the user experiences in VR environments. They provide strategic insights for investors, highlight the importance of aesthetic and functional aspects in VR tourism, and emphasize the need for optimizing VR experiences to minimize discomfort, particularly for younger and less experienced users. Together, these analyses contribute valuable knowledge to both the financial and technological aspects of the evolving digital and virtual landscapes.

9 Limitations

- The evaluations of user experiences in VR tourism are inherently subjective. Factors such as presence, aesthetics, utility, and ease of use are influenced by individual preferences and biases, which can compromise the objectivity and reliability of the conclusions derived from these subjective assessments.
- The analysis of Metaverse cryptocurrencies is based on data from 2020 to 2022. This relatively short time frame may not capture long-term trends or account for significant market changes, limiting the robustness and future applicability of the findings.
- User ratings for aspects such as presence, aesthetics, utility, and ease of use in VR tourism are subjective and influenced by individual preferences and biases. This subjectivity limits the objectivity and reliability of the conclusions drawn from these evaluations.

10 References

- Metaverse Technologies, Opportunities and Threats (Fatih Sinan Esen Hasan Tinmaz Madhusudan Singh Editors)
- Denisa Elena BĂLĂ, Stelian STANCU. Cryptocurrencies in the Digital Era. Composite Index-Based Analysis for the Top Ten Virtual Currencies Traded -June 2022
- Shivangi Sahay,Nishtha Mahajan,Shivangi Malik,Jasdeep Kaur. Metaverse: Research Based Analysis and Impact on Economy and Business-(August 2022)
- Xiao-Ting Huang , Jiahui Wang 2 , Zhihui Wang 1, Linqiang Wang 1 and Chenfei Cheng 1 Experimental study on the influence of virtual tourism spatial situation on the tourists' temperature comfort in the context of metaverse
- Brooks, C. 2018. Introductory Econmterics for finance. 2nd ed. Cambridge: Cambridge University Press.
- Georg, K. Yanfen, Z. 2021. On cointegration and cryptocurrency dynamics. SSRN Electronic Journal
- Joline, G. 2019. A cointegration analysis of Bitcoin, Bitcoin Cash, EOS, Ethereum, Litecoin and Ripple. Spring
- Leung, T. and Nguyen, H. 2018. Constructing Cointegrated Cryptocurrency Portfolios for Statistical Arbitrage. Studies in Economics and Finance, 2019.
- Murray, M.P. 1994. A Drunk and her Dog: An Illustration of Cointegration and Error Correction. The American Statistician, 48(1): 37–39.
- Stock, J.H and Watson, M.W. 2015. Introduction to econometrics. Updated 3rd ed. Harlow : Pearson Education.
- U. A. Chattha, U. I. Janjua, F. Anwar, T. M. Madni, M. F. Cheema and S. I. Janjua. (2020) "Motion Sickness in Virtual Reality: An Empirical Evaluation," in IEEE Access, vol. 8.
- Chang, E., Kim, H. T., & Yoo, B. (2020). Virtual Reality Sickness: A Review of Causes and Measurements. International Journal of Human–Computer Interaction, 36.
- Michael E. McCauley and Thomas J. Sharkey. 1992. Cybersickness: Perception of self-motion in virtual environments. Presence: Teleoper. Virtual Environ. 1, 3 (Summer 1992), 311–318.

- Henry Been-Lirn Duh, Donald E. Parker, Thomas A. Furness. (2004) An Independent Visual Background Reduced Simulator Sickness in a Driving Simulator. Presence: Teleoperators and Virtual Environments.
- Lisa Rebenitsch. (2015). Managing cybersickness in virtual reality. XRDS 22, 1 (Fall 2015), 46–51.
- Gonçalves V, Andrade K, Carvalho K, Silva M, Pereira M, Galvao TF. (2018) Accuracy of self-reported hypertension: a systematic review and meta-analysis. J Hypertens
- Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, An Introduction to Statistical Learning with Applications in R.
- Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning Data Mining, Inference, and Prediction, Second Edition.
- Jayant V. Deshpande, Uttara Naik-Nimbalkar, Isha Dewan Non Parametric Statistics, Theory and Method.
- Bernhard Pfaff, Analysis of integrated and cointegrated Time series with R ,Second Edition.
- RUEY S. TSAY Multivariate Time series Analysis with R and financial Application, WILEY.
- Wolfgang Karl Härdle, Cathy Yi-Hsuan Chen, Ludger Overbeck Editors Applied Quantitative Finance, Third Edition.
- Chris Chatfield, haipeng Xing, The Analysis of Time Series An Introduction with R (Seventh Edition)