Paving The Way For Bitcoin Mass Adoption: A Causal Approach For Performance Modelling

Virgile ROUMENS, Maxime LE FLOCH, Mathieu LARUELLE, Hugo Xinghe CHEN, Matteo LUSARKI and Paul-Hakim TROUVE

Supervised by:
Jean Gabriel ATTALI (Head of Finance Department - Leonard de Vinci Graduate School of Engineering (ESILV))
Mathieu VAISSIÉ (Research Director - Ginjer AM)

March 16, 2024

Contents

1 Abstract

Our project aimed to delve into the key determinants of Bitcoin's performance and develop an investment strategy based on the insights gained. By analyzing approximately 60 features sourced from Bloomberg, we meticulously cleaned and preprocessed the data, ensuring it was properly aligned with the corresponding timestamps.

Thanks to Jean Gabriel Attali's advices, we decided to shift our analysis from daily to weekly frequency datasets to address potential correlation underestimation caused by exposure to different market time zones. Through extensive research from relevant literature and valuable guidance from Mathieu Vaissié, we identified the optimal methodology for implementing a machine learning-based investment strategy.

Leveraging the power of the Random Forest algorithm, we successfully predicted Bitcoin's weekly returns from September 2014 to December 2023 shifting our timeframe analysis month after month. Each iteration provided valuable insights into the key features and sectors influencing Bitcoin's return behavior, as well as predicted future returns, enabling us to understand the evolving determinants impacting Bitcoin over time. Employing causal inference techniques, we sought to establish a causal relationship between selected factors and Bitcoin's price movements.

Our findings largely confirmed that the most relevant features identified by the Random Forest algorithm exhibited causal influence on Bitcoin's returns, as indicated by the Granger causality test. For the implementation of our long-only investment strategy, we focused on detecting signals with a monthly rebalancing frequency. Based on the predicted returns generated by our Random Forest models, we defined four distinct signals: "Hold" for negative predicted returns and three long signals corresponding to various thresholds of positive returns.

Dynamically adjusting our portfolio allocation based on these signals, we effectively captured downside movements in Bitcoin, resulting in a reduction of our strategy's annualized volatility compared to Bitcoin's annualized volatility. This approach enabled us to mitigate risk while maintaining exposure to potential upside movements. In conclusion, our project provides valuable insights into the determinants of Bitcoin's performance, empowering institutional investors to better understand the factors driving its behavior.

By leveraging state-of-the-art machine learning algorithms to validate explanatory features through a causal approach, and implementing a dynamic investment strategy, we aspire to make a modest yet meaningful contribution to the advancement of knowledge in the realm of cryptocurrency investments.

2 Explanation of features

2.1 On-chain metrics

2.1.1 MVRV

The Market Value to Realized Value (MVRV) ratio is a pivotal metric in cryptocurrency analysis, particularly for Bitcoin. It compares the market capitalization (current price multiplied by coins in circulation) with the realized capitalization (valuating each Bitcoin by the price at its last movement). This ratio helps assess if Bitcoin is overvalued or undervalued: a high MVRV indicates a possibly inflated price, while a low MVRV suggests undervaluation. Essential for market sentiment analysis, MVRV fluctuations can signal potential volatility shifts or price trends, serving as a key indicator for investment decisions.

2.1.2 Active address count

This metric tracks the daily, weekly, or monthly unique active Bitcoin addresses, offering insights into network activity and user engagement. High counts often suggest increased usage, which could correlate with market growth or price movements, while low counts might indicate decreased activity, potentially impacting Bitcoin's valuation. This indicator is vital for gauging the network's health and can significantly influence investment decisions by highlighting trends in user adoption and network utilization.

2.1.3 Address count with balance $>$ 10K and 100K BTC

These metrics denote the count of unique active addresses holding over 10,000 BTC and 100,000 BTC. It's a powerful indicator of substantial Bitcoin holdings, reflecting the activity of large-scale investors or entities, often referred to as "whales." A high count signals significant holding concentration, potentially affecting market liquidity and price volatility. Conversely, a decrease may imply distribution or selling pressures from major holders. This metric is crucial for understanding the market's power dynamics, offering insights into the influence of major stakeholders on Bitcoin's price movements and market stability.

2.1.4 Average block size

This metric measures the average amount of data processed in each block over a specific time period, reflecting network throughput in the BTC blockchain. A larger block size indicates higher transactional volume, suggesting increased network usage and activity.

2.1.5 Hashrate

The hashrate reflects Bitcoin's network computational power used to mine blocks by the miners, influencing its security and miner participation. A high hashrate, often associated with rising Bitcoin prices, suggests robust network health and bullish sentiment, while a decrease might indicate miner withdrawal and potential price volatility.

2.1.6 Mean difficulty

The mean difficulty of Bitcoin adjusts to the network's mining power (hashrate), balancing block creation rates (one every 10 minutes). It mirrors economic shifts: rising with increased mining (often due to higher Bitcoin prices) and decreasing when mining lessens. This metric indirectly hints at Bitcoin's price trends and mining engagement, influencing the network's stability and security.

2.1.7 Miner revenue in BTC and USD

This metric reflects the earnings (in BTC/USD) miners receive for each block mined, comprising block rewards and transaction fees. This revenue influences Bitcoin's ecosystem by affecting mining participation and network security, potentially impacting Bitcoin's price and market volatility, as variations in mining earnings lead to shifts in the available BTC supply and miner activity.

2.1.8 NVT

The Network Value to Transactions (NVT) ratio compares Bitcoin's market cap to its transaction volume, serving as a valuation metric. A high NVT indicates potential overvaluation and speculation, while a low NVT may suggest undervaluation or high utility. Fluctuations in NVT can offer insights into Bitcoin's price movements and investor sentiment.

2.1.9 Transaction count per second

This metric measures the number of Bitcoin transactions processed every second, reflecting the network's activity level and throughput. It's a direct indicator of the blockchain's usage, as heightened activity may indicate greater demand or adoption, potentially affecting the cryptocurrency's price stability and market perception.

2.1.10 Transaction fees in BTC and USD

Transaction fees (in BTC or USD) reflect the cost to process transactions on the blockchain, signaling network demand. High fees can indicate heavy usage or congestion, indirectly influencing Bitcoin's market sentiment and potentially its price dynamics.

2.1.11 Value in addresses with balance $>$ 10K and $>$ 100K BTC

These metrics track the total Bitcoin held in wallets with balances exceeding 10,000 BTC/100,000 BTC, indicating the distribution of wealth and potential market influence of large holders, often called 'whales.' Significant movements in these accounts can impact Bitcoin's price, as large-scale buy or sell orders can lead to increased volatility and influence market sentiment.

2.2 Equity Indices

2.2.1 DowJones

The Dow Jones Industrial Average, a key indicator of the U.S. industrial sector's health, often reflects broader economic trends. When central bank bond yields are low, offering meager returns, investors might gravitate towards Bitcoin, seeking higher yields, especially if the Dow indicates strong market confidence.

2.2.2 Nasdaq

This tech-heavy index is pivotal in representing investor sentiment towards the technology sector. In periods of low bond yields from central banks, investors might increasingly correlate Bitcoin with high-growth tech stocks listed on the Nasdaq, viewing it as an attractive alternative investment.

2.2.3 SP 500

As a broad measure of the U.S. stock market, the SP 500's trends can significantly impact Bitcoin. It's often observed that lower interest rates, leading to reduced bond yields, can drive investors towards riskier assets like Bitcoin, aiming to capitalize on its potential high returns amidst a bullish stock market.

2.2.4 CAC 40

Reflecting the French and broader European market health, the CAC40's performance can influence Bitcoin investment, particularly when European central bank bond yields are low, prompting investors to look for alternative assets that could offer better returns, such as Bitcoin.

2.2.5 FTSE 100

The FTSE 100 index provides insights into the UK's economic condition. Movements in this index, coupled with the UK's central bank bond yield trends, can affect Bitcoin's appeal. Investors might pivot to Bitcoin when traditional bonds offer unattractive yields, using it as a tool for portfolio diversification and potential gains.

2.2.6 NKY 225

Japan's Nikkei 225 index is a significant indicator of its market health. The persistent low-interest-rate environment in Japan often leads investors to seek better returns than what is offered by traditional bonds, potentially increasing interest in alternative investments like Bitcoin, especially when the Nikkei shows positive momentum.

2.2.7 VIX

Known as the fear gauge, the VIX index can indirectly affect Bitcoin's market. High volatility and uncertainty, along with low yields from central bank bonds, might make Bitcoin an appealing option for investors seeking non-correlated assets that offer a hedge against traditional market swings.

2.2.8 TSEMIL (Tel Aviv Stock Exchange MIL Index)

This index tracks the performance of the Tel Aviv Stock Exchange. While it's more localized compared to other global indices, significant movements could reflect regional economic conditions or investor sentiment, potentially affecting Bitcoin through shifts in regional investment strategies or risk appetite.

2.3 Individual Stocks metrics

2.3.1 Amazon

Amazon' stock performance can influence Bitcoin through its reflection of tech sector health and investor sentiment towards innovative, high-growth companies.

2.3.2 Apple

Apple's market movements can impact Bitcoin by indicating broader tech industry trends and investor confidence in technology investments.

2.3.3 Google

As a major player in the tech industry, Google's stock performance might correlate with Bitcoin's price, reflecting investor sentiment towards technology and speculative assets.

2.3.4 Tesla

Tesla' stock is often aligned with high-risk, high-reward investments, similar to Bitcoin. Its performance can mirror investor sentiment towards innovative and speculative market opportunities.

2.4 Commodity Indices

2.4.1 Gold

Gold is viewed as a safe-haven asset, with its value often reflecting global economic stability. Bitcoin, also called "digital gold," shares similarities, such as serving as a store of value and an inflation hedge. Both assets attract investors during economic uncertainty, with Bitcoin offering a modern, divisible, and transferable alternative, enhancing its comparison to traditional gold.

2.4.2 CL1 COMB Comdty (Brent Crude) / CO1 COMB Comdty (WTI Crude) / NG1 COMB Comdty (Natural Gas)

These commodities act as indicators of global economic health. Their price fluctuations can impact Bitcoin's valuation as shifts in economic conditions influence investor behavior and market sentiment.

2.5 Interest Rates Yields

2.5.1 USYC2Y10 (US 2-Year vs. 10-Year Treasury Yield Curve)

This metric compares short-term (2-year) to long-term (10-year) U.S. Treasury yields, providing insights into the economic outlook. A flattening or inversion of this curve (where short-term yields are higher than long-term) often signals economic uncertainty or impending recession, potentially increasing Bitcoin's appeal as a non-correlated asset or safe haven.

2.5.2 DEYC2Y10 (Germany's 2-year vs. 10-year yield spread)

This yield spread provides insights into the German bond market's shape, reflecting economic outlook and monetary policy expectations in one of Europe's largest economies. A flattening or inversion might signal economic worries, potentially increasing Bitcoin's attractiveness as a diversification tool.

2.5.3 JPYC2Y10 (Japan's 2-year vs. 10-year government bond yield spread)

This spread gives insight into the Japanese yield curve, affecting investor sentiment regarding Japan's economic future and interest rate expectations. A steepening curve might indicate economic optimism and reduced demand for Bitcoin as a safe haven, while a flattening or inversion could heighten Bitcoin's appeal amid economic or deflationary concerns.

2.5.4 FED Policy Rates

The Federal Reserve's policy rates directly influence U.S. economic activity by dictating borrowing costs. When rates are low, investors might seek higher returns in riskier assets like Bitcoin, whereas high rates can make traditional investments more attractive, possibly dampening Bitcoin's appeal.

2.5.5 US Federal Rate Target

The target federal funds rate is a benchmark for short-term interest rates. Changes to this rate can influence Bitcoin's attractiveness; lower rates generally boost interest in Bitcoin as investors look for yield, while higher rates might lead to a preference for safer, interest-bearing assets.

2.5.6 ECB Policy Rates

The European Central Bank's policy rates affect liquidity and interest rates across the eurozone. Similar to the Fed's rates, lower ECB rates can enhance Bitcoin's appeal as an alternative investment, whereas higher rates might deter investment in Bitcoin in favor of traditional assets.

2.5.7 TED SPREAD EUR, TED SPREAD US, TED SPREAD JPN

The TED Spread measures the difference between the three-month interbank interest rate and the three-month government Treasury bill rate, reflecting credit risk and liquidity in the market. A widening spread indicates higher perceived risk or potential financial stress, which could lead investors to seek alternative investments like Bitcoin for diversification or as a hedge.

2.5.8 EURR002W (Euro Overnight Index Average - EONIA)

This is the average interest rate at which eurozone banks offer to lend unsecured funds to other banks overnight. Movements in EONIA can reflect the eurozone's banking liquidity conditions; lower rates may push investors towards Bitcoin in search of higher yields.

2.5.9 GJGB3M (Japanese 3-Month Government Bond Yield), GCNY3M (Chinese 3-Month Government Bond Yield)

These yields reflect short-term government debt obligations in Japan and China, respectively. Lower yields on these bonds could indicate a search for alternative investments, potentially increasing interest in Bitcoin, especially in regions affected by negative interest rates or low yield environments.

2.6 Economic Indices Surprises

The Citi Economic Surprise Indices (CESI) for the U.S., Eurozone, Japan, and China provide insights into how actual economic data measures up against market expectations. Positive surprises in any of these indices could bolster market optimism and risk appetite, potentially uplifting Bitcoin's value as investors might see it as a lucrative high-risk investment. However, negative economic surprises might heighten market caution, driving investors towards Bitcoin as a hedge or safety asset amidst increased uncertainty. These indices collectively offer a global perspective on economic health, influencing Bitcoin's appeal either as a speculative asset during times of economic strength or as a protective haven in times of unpredictability.

2.7 Foreign Exchange Rates

Each currency pair's movements can indirectly signal shifts in Bitcoin's market, as they encapsulate broader economic sentiments, central bank actions, and investor responses to fiat currency valuations.

2.7.1 eurodollar (EUR/USD)

The EUR/USD exchange rate's variations are significant as they mirror the economic dynamics between the Eurozone and the United States. When the euro strengthens, it might boost the attractiveness of Bitcoin by presenting it as a viable alternative to traditional fiat currencies. Conversely, a strengthening U.S. dollar could lessen Bitcoin's appeal, as investors might favor the perceived stability or higher returns of holding or investing in the dollar.

2.7.2 gbpusd (GBP/USD)

Changes in the pound's value against the dollar can influence Bitcoin's attractiveness in the UK, with a rising pound potentially reducing Bitcoin's appeal, and a falling pound possibly boosting it as a value reserve.

2.7.3 renminbiusd (CNY/USD)

The CNY/USD rate affects Bitcoin, especially in the context of China's economic stance. A strengthening renminbi may decrease Bitcoin's attractiveness, whereas depreciation could heighten its status as a hedge.

2.7.4 yenusd (JPY/USD)

The yen's exchange rate against the dollar can sway Bitcoin's appeal in Japan, with yen depreciation making Bitcoin more attractive as a hedge and appreciation potentially reducing its necessity.

2.8 Miscellaneous Financial Indicators

2.8.1 Baltic Dry Index (BDIY)

This index tracks global shipping prices of various dry bulk commodities, serving as an indirect indicator of global economic health and trade flows. A rising BDIY suggests increasing demand for shipping, potentially signaling economic growth, which might influence Bitcoin as global economic confidence grows. Moreover, a declining BDIY, indicating potential economic slowdowns, could lead investors to consider Bitcoin as a hedge against economic uncertainty.

2.8.2 SOX (Philadelphia Semiconductor Index)

The SOX index represents the performance of the semiconductor industry. As semiconductors are vital for various technology sectors, and especially for Bitcoin miners, the SOX index can reflect tech industry health and investor sentiment towards technology investments, but also a higher amount of computers mining blocks on Bitcoin. That's why fluctuations in this index might correlate with Bitcoin's price and activity.

2.8.3 Federal Funds Target Rate:

This is the interest rate at which depository institutions lend balances at the Federal Reserve to other depository institutions overnight. Changes in the federal rate target can influence Bitcoin's attractiveness. Lower rates may encourage investment in riskier assets like Bitcoin for higher returns, while higher rates could drive investors towards safer, interest-bearing assets.

2.8.4 IUTPELC (Utility Sector Performance)

Utilities are typically considered stable investments; thus, this sector's performance can act as a gauge for investor risk tolerance. Strong performance might indicate a preference for safety, possibly detracting from Bitcoin's appeal, whereas underperformance might drive investors towards alternative assets, including Bitcoin.

2.8.5 BDIY (Baltic Exchange Dry Index)

Similar to the Baltic Dry Index, it tracks international shipping prices of dry bulk cargo. An increasing BDIY points to heightened global trade and economic activity, potentially limiting Bitcoin's value as market confidence grows, while a decrease may signal economic concerns, possibly boosting Bitcoin's safe-haven appeal.

2.8.6 GETB1 (Generic 1st 'GE' Future)

This tracks the futures prices of General Electric shares. While seemingly company-specific, significant movements can reflect broader industrial or economic trends, potentially impacting Bitcoin through general market sentiment or industrial sector health.

2.8.7 BOJDPBAL (Bank of Japan's Deposit Balance)

It indicates the total balance of deposits at the BOJ, reflecting Japan's monetary policy stance and liquidity in the financial system. Significant changes can affect global market liquidity and risk sentiment, potentially influencing Bitcoin's market as investors assess global monetary conditions.

3 Correlation analysis

We started by searching if there was some correlation between our variables, because the causality implied the correlation so a variable which is correlated with another variable has more can possibly cause this variable.

3.1 Daily Correlation

When we chose our variables, we took datasets representing indices or share prices from different countries in different parts of the world. At first, we wanted to process our datasets in Daily, and we obtained this correlation matrix:

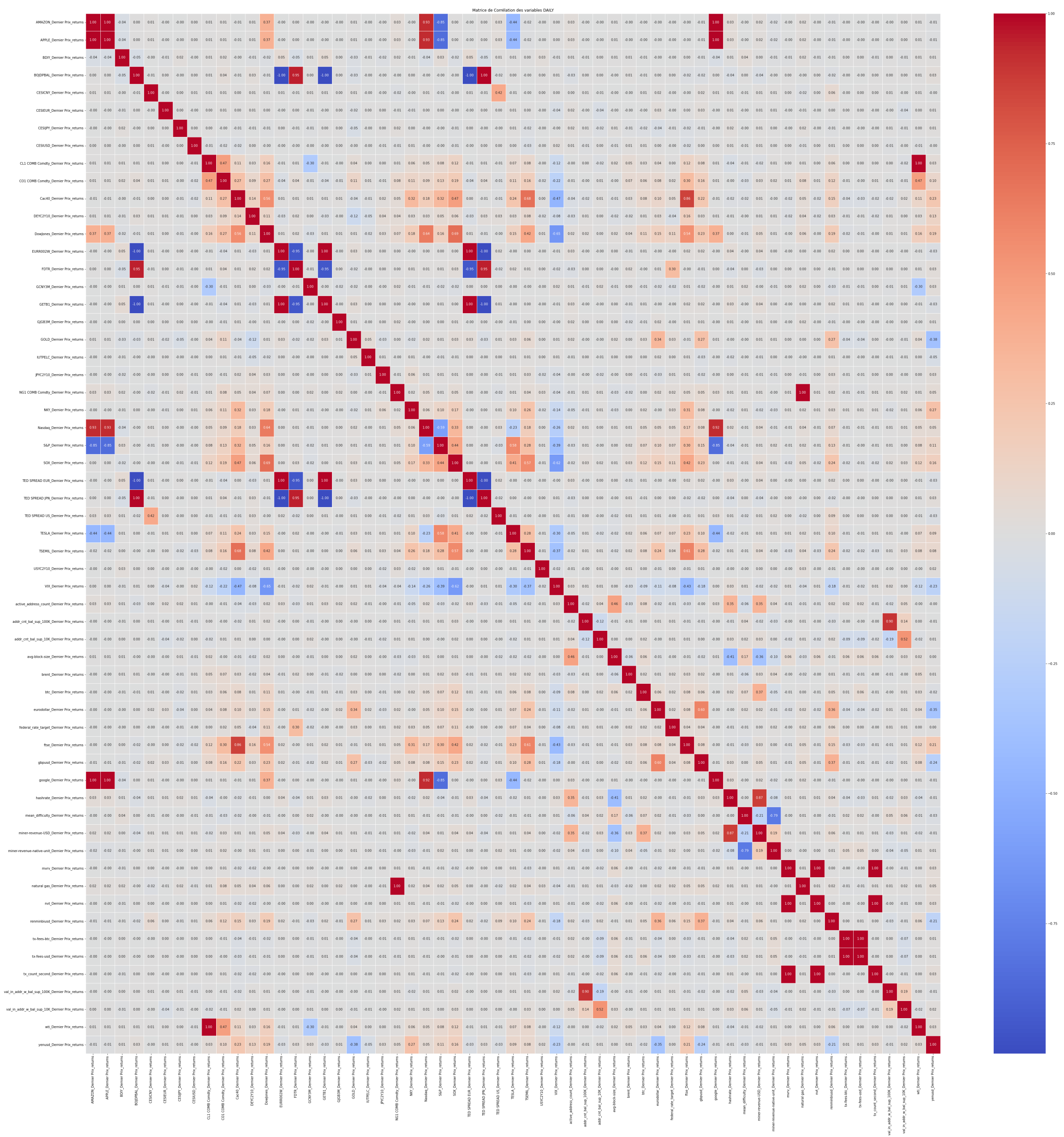

Figure 1: Correlation Matrix daily returns

There is not much correlation, particularly between stock market indices such as the CAC and S&P. This is because stock markets do not open and close at the same time.

Take, for example, the New York Stock Exchange, which opens from 3.30pm to 10pm French time, the Japan Exchange Group, which opens from 1am to 7am French time, and Euronext, which opens from 9am to 5.30pm French time.

If we take the index values or shares listed on these 3 different exchanges and calculate the correlations of their returns, we will get biased results because we will not have all the information because the exchanges do not open at the same time.

That's why the correlations in our matrix didn't come out well and why our results were inconclusive, because we were losing a lot of information.

3.2 Weekly Correlation

After discussing the matter with Jean Gabriel Attali and Mathieu Vaissié, we decided to take the data on a weekly basis rather than on a daily basis, which would greatly reduce the loss of information because we would no longer have day-to-day time differences but only at the end of the week, so the loss would be much smaller. Our new correlation matrix is therefore as follows:

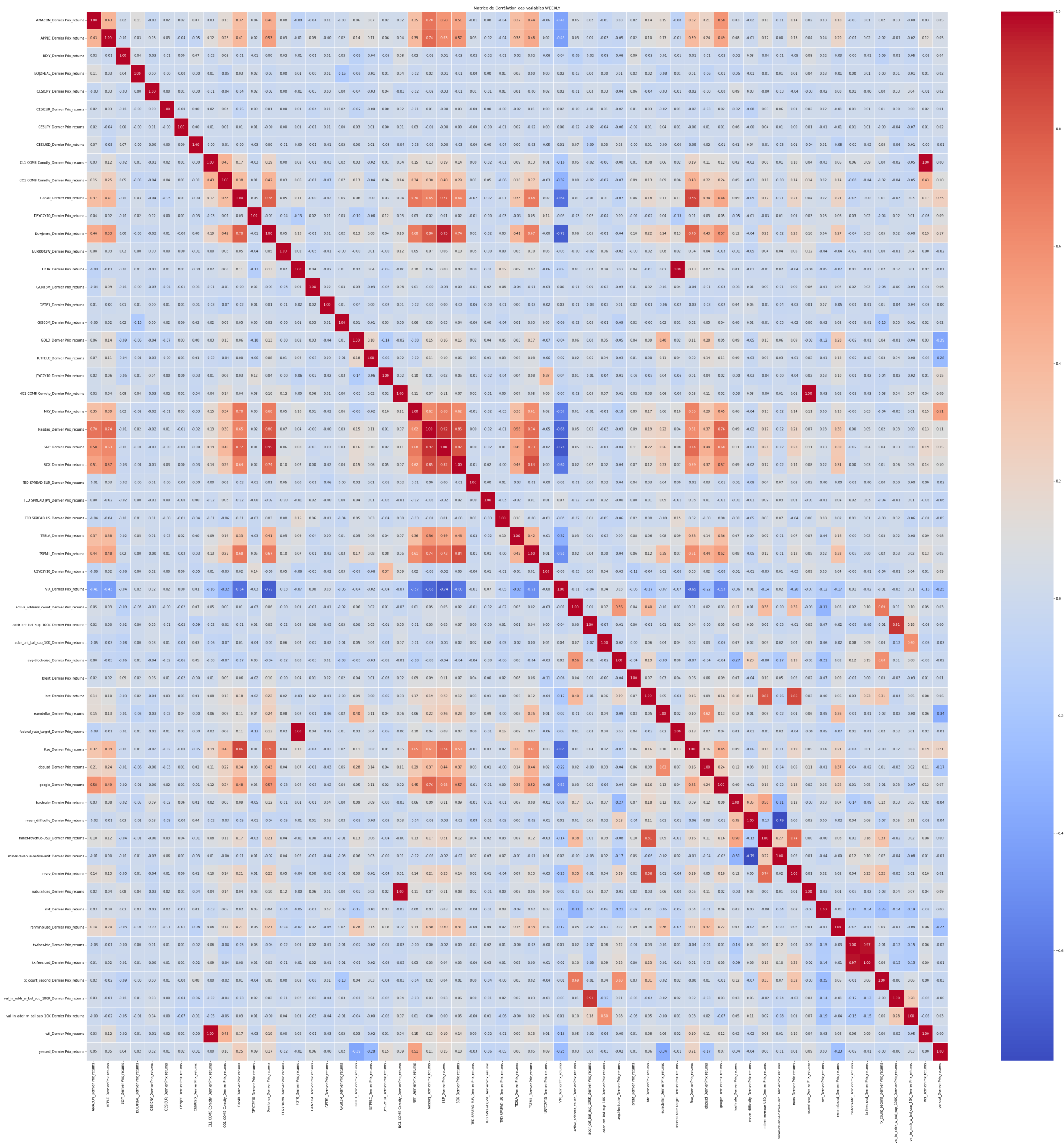

Figure 2: Correlation Matrix weekly returns

We can now see that we have stronger correlations and, above all, more interesting correlations for the rest of our project. We can see that the correlations between the indices are more pronounced, which is logical because if the US stock market opens sharply higher or lower, the other markets generally react in the same direction. This give us an overview of which variables could cause or not what we want to predicted.

4 Generalized least squares (GLS regression)

4.1 Explication of GLS regression

GLS, or generalized least squares, is an extension of classical linear regression that aims to overcome the limitations associated with heteroscedasticity and error correlation. Unlike the ordinary least squares (OLS) method, which assumes homoscedastic and independent errors, GLS allows for variable error variances and correlation between these errors.

By adjusting the weights assigned to the residuals, GLS seeks to minimize the weighted least squares, thereby promoting more efficient estimates of the model parameters. GLS regression adapts the model to the particularities of the errors, thereby improving the robustness of the estimates while offering flexibility in the face of data irregularities.

We did the GLS regression from the Bitcoin prices returns in weekly on all our data.

We obtained two noteworthy results, while the others were having R^2 close to zero and deemed uninteresting for further analysis. The two significant results are the $Mvrv\_last\_price\_return$ (weekly returns) and the $miner\_revenue\_USD\_last\_price\_return$ (weekly returns).

We will look at two results inside the GLS regression, the first one is the result of the R-squared and the second one is the result of the t-statistic and the p-value.

The first one is obtained by: $\boxed{R^2 = \frac{\text{Var}(Y)}{\text{Var}(\hat{Y})}}$

Where :
– $\hat{Y}$ is the predicted value of the dependent variable.
– Y is the observed dependent variable.
– $\text{Var}(\hat{Y})$ is the variance of the predicted values.
– $\text{Var}(Y)$ is the variance of the observed values.

To interpret the result of R^2, an R^2 of 1 indicates that the model explains all the variance of the dependent variable, while an R^2 of 0 means that the model explains no variance.

The second one is obtained by: $\boxed{t_i = \frac{b_i}{\text{standard error of } b_i}}$

Where :
– b_i is the estimated coefficient associated with the independent variable.
– The standard error of b_i is a measure of the uncertainty of the coefficient estimate.

The t-statistic follows a Student distribution, and the significance of the coefficient is assessed by comparing the t-statistic with a reference distribution.

The p-values (P>|t| in the output) indicate the probability that the t-statistic is as extreme as that observed, under the null hypothesis that the coefficient is equal to zero. A low p-value suggests the statistical significance of the coefficient.

4.2 GLS regression on MVRV

For the first variable, our results are:

```
Results for adding mvrv_Dernier Prix_returns:
                            GLS Regression Results
=============================================================================
Dep. Variable:     btc_Dernier Prix_returns   R-squared:                   0.738
Model:                                  GLS   Adj. R-squared:              0.738
Method:                       Least Squares   F-statistic:                 1807.
Date:                      Thu, 07 Mar 2024   Prob (F-statistic):      1.08e-188
Time:                              17:00:51   Log-Likelihood:             972.76
No. Observations:                       643   AIC:                        -1942.
Df Residuals:                           641   BIC:                        -1933.
Df Model:                                 1
Covariance Type:                  nonrobust
=====================================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------
const                     0.0153      0.002      7.277      0.000       0.011       0.019
mvrv_Dernier Prix_returns 1.1012      0.026     42.512      0.000       1.050       1.152
=======================================================================
Omnibus:                 719.405   Durbin-Watson:                 0.954
Prob(Omnibus):             0.000   Jarque-Bera (JB):          51233.175
Skew:                      5.296   Prob(JB):                       0.00
Kurtosis:                 45.427   Cond. No.                       12.3
=======================================================================
```

Figure 3: GLS Estimator result for MVRV

Interpretation of GLS regression result for MVRV:

R-squared (R^2) :

– R^2 measures the proportion of the variance in the dependent variable (*btc_Last_Price_returns*) explained by the model. In this case, R^2 is equal to 0.738, which means that the model explains approximately 73.8% of the variation in the dependent variable.

Significance of the coefficients (t-statistic):

– *const :* The coefficient associated with the constant (0.0153) is significant with a t-statistic of 7.277 and a very low p-value (0.000). This suggests that the constant is significantly different from zero.

– *mvrv_Last_returns_price*: The coefficient associated with this variable (1.1012) is also significant with a t-statistic of 42.512 and a very low p-value (0.000). This indicates that the independent variable "*mvrv_Last_returns_price*" has a significant effect on the dependent variable.

In summary, the model appears to be statistically significant (Prob (F-statistic) very low) and the two individual coefficients (*const* and *mvrv_Last_returns_price*) are significantly different from zero. The "*mvrv_Last_returns_price*" variable appears to have a significant impact on the dependent variable, given its high coefficient and significant t-statistic.

4.3 GLS regression on Miner Revenue

For the first variable, our results are:

```
Results for adding miner-revenue-USD_Dernier Prix_returns:
                            GLS Regression Results
==============================================================================
Dep. Variable:     btc_Dernier Prix_returns   R-squared:                       0.650
Model:                                  GLS   Adj. R-squared:                  0.650
Method:                       Least Squares   F-statistic:                     1192.
Date:                      Thu, 07 Mar 2024   Prob (F-statistic):          2.14e-148
Time:                              17:00:51   Log-Likelihood:                 879.75
No. Observations:                       643   AIC:                            -1756.
Df Residuals:                           641   BIC:                            -1747.
Df Model:                                 1
Covariance Type:                  nonrobust
==========================================================================================================
                                            coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------------------------
const                                     0.0063      0.002      2.574      0.010       0.002       0.011
miner-revenue-USD_Dernier Prix_returns    0.6176      0.018     34.528      0.000       0.582       0.653
==============================================================================
Omnibus:                      174.214   Durbin-Watson:                   1.996
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              897.788
Skew:                           1.111   Prob(JB):                    1.12e-195
Kurtosis:                       8.345   Cond. No.                         7.35
==============================================================================
```

Figure 4: GLS Estimator result for Miner Revenue

Interpretation of GLS regression result for Miner Revenue:

R-squared (R^2):

– R^2 measures the proportion of the variance in the dependent variable (*btc_Last_Price_returns*) explained by the model. In this case, R^2 is equal to 0.650, which means that the model explains approximately 65.0% of the variation in the dependent variable.

Significance of the coefficients (t-statistic):

– *const :* The coefficient associated with the constant (0.0063) is significant with a t-statistic of 2.574 and a very low p-value (0.010). This suggests that the constant is significantly different from zero.

– *miner_revenue_USD_last_price_return*: The coefficient associated with this variable (1.1012) is also significant with a t-statistic of 34.528 and a very low p-value (0.000). This indicates that the independent variable "*miner_revenue_USD_last_price_return*" has a significant effect on the dependent variable.

We got the same summary for the last variable, so we can say that these two variables are likely the most significant in explaining the movement of weekly returns on bitcoin prices.

However, it's important to note that while these variables are statistically significant, it doesn't necessarily imply causation. It suggests a statistical relationship, but not a necessary causation.

5 Random Forest Algorithm

5.1 Prices Forecasting

5.1.1 Features and Label selection

We'll employ both stationary and non-stationary datasets to forecast BTC prices. To achieve this, we'll utilize a dataset comprising raw data, returns, and volatilities.

To prepare the dataset for analysis, we will drop the labels that pertain to BTC features, namely prices, returns, and volatilities.
Let's start our Bitcoin price forecasting thanks to Random Forest algorithm.

5.1.2 Rolling Random Forest

For the analysis, we have selected a window of 200 weeks, which corresponds to a representative Bitcoin cycle period. Within this window, we allocate 160 weeks for training the model and 40 weeks for testing.

To make predictions on the price of Bitcoin, we employ the Random Forest algorithm and iterate using a rolling window of 4 weeks. In fact, we've chosen not to make an expanding Random Forest analysis (incrementing our timeframe analysis for every month shift) to really stay focus on the features that are important inside our timeframe.

This rolling approach allows us to capture the evolving dynamics that contribute to Bitcoin's narratives. By focusing on a shorter time frame, we can better understand the features that explain Bitcoin's behavior, as they change over time.

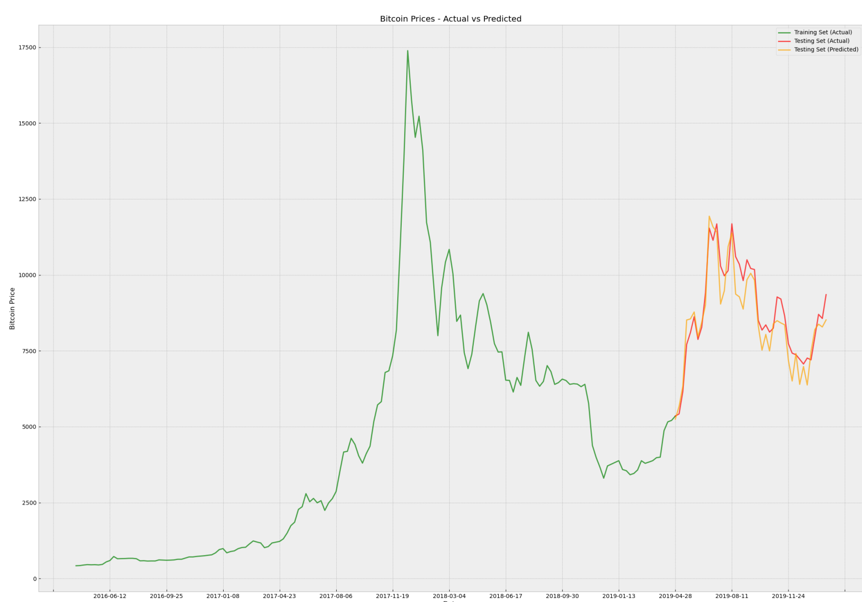

Figure 5: Forecasted price VS actual price from May 2019 to February 2020

See bellow our video predicting Bitcoin prices from September 2014 to December 2023:
https://youtu.be/zZy09jzpjxg

5.2 Returns Forecasting

5.2.1 Features and Label selection

We will be working with stationary data to predict BTC returns. For this purpose, we will utilize the dataset including returns and normalized volatilities (ie: returns divided by their volatilities, aka Z-Scores) in order to introduce volatility information in a stationary way.

To prepare the dataset for analysis, we will drop the labels that pertain to BTC features, succinctly returns and z-scores.
Let's start our Bitcoin returns forecasting thanks to Random Forest algorithm.

5.2.2 Rolling Random Forest

For the analysis, we have selected a window of 200 weeks, which corresponds to a representative Bitcoin cycle period. Within this window, we allocate 160 weeks for training the model and 40 weeks for testing.

To make predictions on the price of Bitcoin, we employ the Random Forest algorithm and iterate using a rolling window of 4 weeks. In fact, we've chosen not to make an expanding Random Forest analysis (incrementing our timeframe analysis for every month shift) to really stay focus on the features that are important inside our timeframe.

This rolling approach allows us to capture the evolving dynamics that contribute to Bitcoin's narratives. By focusing on a shorter time frame, we can better understand the features that explain Bitcoin's behavior, as they change over time.

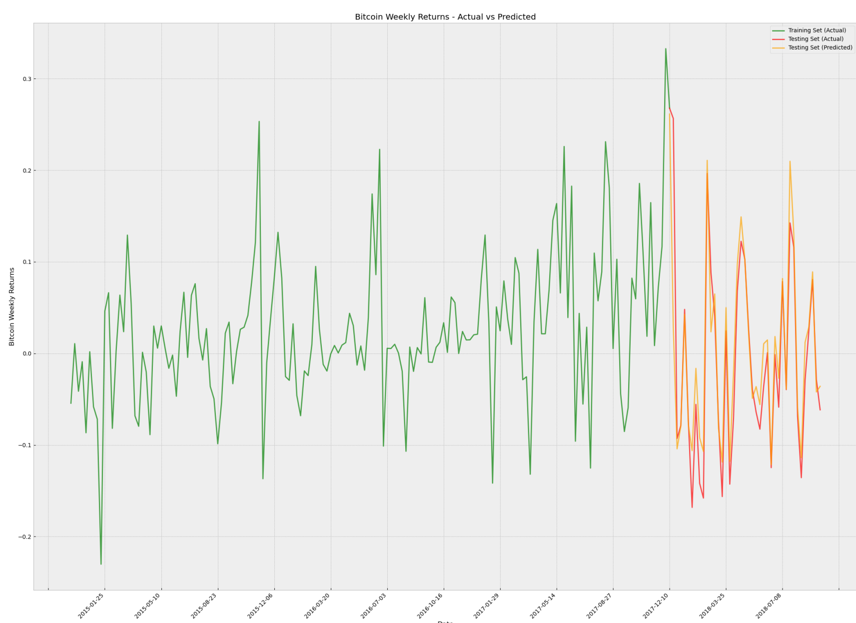

Figure 6: Forecasted returns VS actual price from December 2017 to September 2018

See bellow our video predicting Bitcoin returns from September 2014 to December 2023:

https://youtu.be/SmR2YWjO2d4

6 Key Determinants of Bitcoin

6.1 Stationary Variables

6.1.1 Features Importance

On the code of the RandomForest we did two more thing to have more information.

The first one is the importance of the variables at each period of the RandomForest, the importance of a variable is calculated based on the Mean Decrease in Impurity (or Gini impurity) across all decision trees in the ensemble.

For example, for the period :

Table 1: Period Example

Type of Period	**Date**
Start Train :	2011-11-27
End Train :	2014-12-14
Start Test :	2014-12-14
End Test :	2015-09-20

The result is :

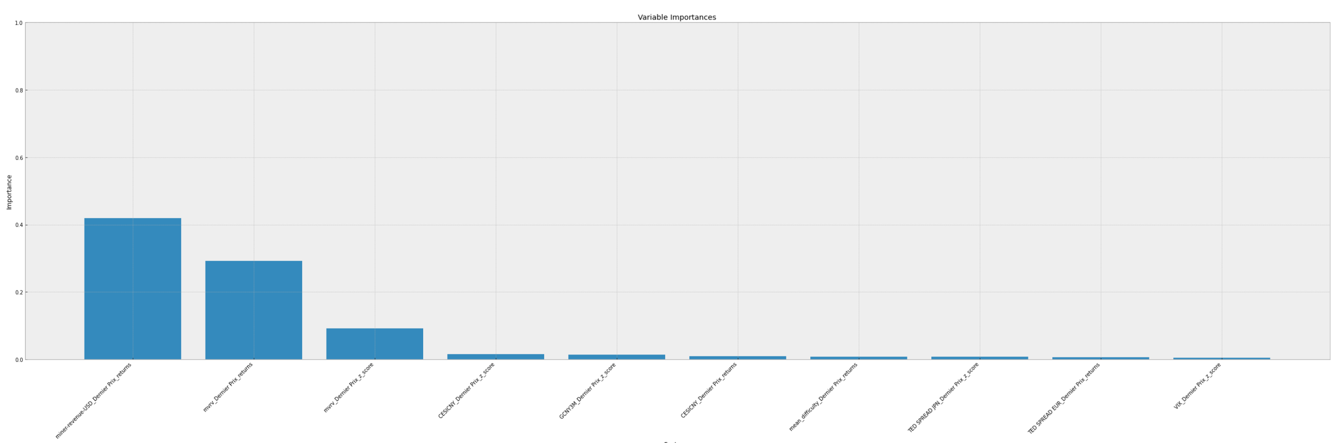

Figure 7: Importance Plot Example

Table 2: Importance Result

Feature	**Importance**
miner_revenue_USD_returns	0.4197
mvrv_returns	0.2926
mvrv_z_score_returns	0.0919
CESICNY_z_score	0.0149
GCNY3M_z_score)	0.0130
CESICNY_Prix_returns	0.0085
mean_difficulty_returns	0.0081
TED_SPREAD_JPN_returns	0.0072
TED_SPREAD_EUR_returns	0.0055
VIX_z_score	0.0048

We can see that the variables which have the most importance during this period are the *miner_revenue USD_Last_price_returns* and the *mvrv_Last_returns_price* (these are also the 2 variables that we found with the GLS regression). We computed those importances for every month roll, recalculating them with our Random Forest algorithm. The 5 variables that appear the most are:

Table 3: Counting Result

Feature	**Count**
miner_revenue_USD_returns	111
mvrv_returns	109
mvrv_z_score	109
tx_fees_usd_returns	54
active_address_count_returns	46

We got the variables that are the most important variables to predict the *btc_Dernier_Prix_returns*.

However, we need to be careful with some of these variables, as variables such as *minerrevenue* and *mvrv* are largely influenced by Bitcoin.

These variables are very much driven by upward or downward movements in Bitcoin, so we find that they can be used to predict its returns, but in the event of strong movements in this currency these variables will be very much affected and so the results of a strategy based on them may not be very good, which we will confirm or not later in this document.

6.1.2 Sectors Importance

We have created features groups, in order to obtain sectorial packages. (cf Explanation of features section)

Furthermore, we applied the importance that we saw before on these groups. If we take the same period than before, we got this result :

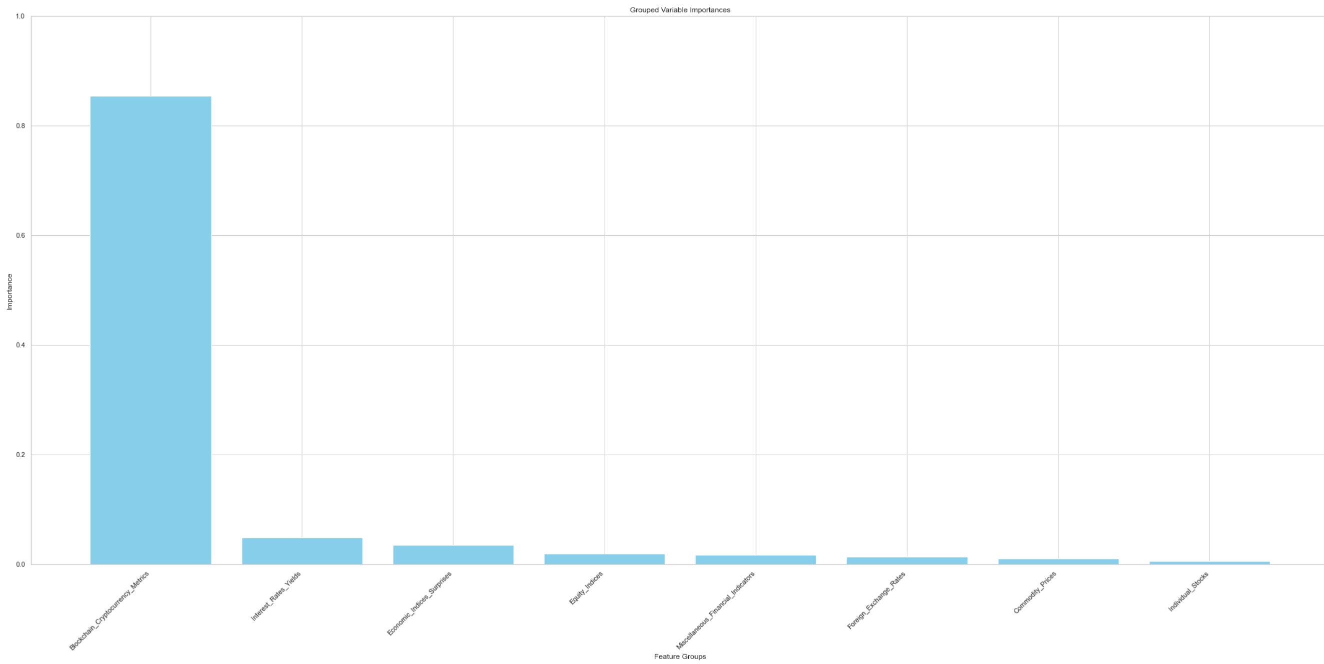

Figure 8: Sector's Importance Example

Table 4: Sector's Importance Result

Sector	**Importance**
Blockchain Cryptocurrency Metrics	0.854166
Interest Rates Yields	0.048600
Economic Indices Surprises	0.034262
Equity Indices	0.018585
Miscellaneous Financial Indicators	0.016022
Foreign Exchange Rates	0.012829
Commodity Prices	0.010192
Individual Stocks	0.005344

We can see that the most important sector to predicted the return of the Bitcoin is the "Blockchain Cryptocurrency Metrics".

We saw in the part *Features Importance* that the 5 most important variables was :

– $miner\_revenue\_USD\_returns$
– $mvrv\_returns$
– $mvrv\_z\_score$
– $tx\_fees\_usd\_returns$
– $active\_address\_count\_returns$

All of these variables are in the "Blockchain Cryptocurrency Metrics" sector so it's logical to have this sector in first and with a very high score.

6.2 Non Stationary Variables

In this part we are going to look at the result when we gave prices, returns and volatilities features to our machine learning model. Those outputs are the results that we get with the Prices Forecasting Random Forest.

We will examine the outcomes during the same time frame as the example we previously utilized.

6.2.1 Features Importance

These are the results that we get with the non-stationary variables :

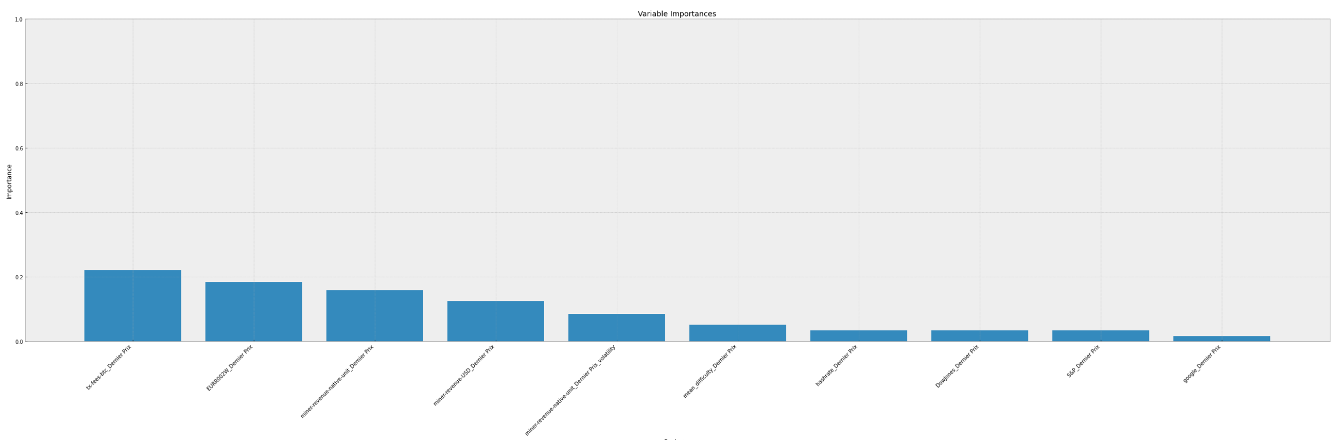

Figure 9: Importance Plot Example

Table 5: Importance Result

Feature	**Importance**
tx_fees_btc	0.2214
$EURR002W$	0.1844
$miner\_revenue\_native\_unit$	0.1594
$miner\_revenue\_USD$	0.1258
$miner\_revenue\_native\_unit\_volatility$	0.0848
$mean\_difficulty$	0.0518
$hashrate$	0.0347
$DowJones$	0.0346
$S\&$	0.0339
$Google$	0.0169

We can see that the variables which have the most importance during this period are tx_fees_btc and $EURR002W$ but there importance are less than the stationary variables.

These results are intriguing because we observed variables that were not present when we ran the Returns Random Forest.

It's difficult to explain why the $EURR002W$ has an high importance in this period so we gonna look at the final result to see if whether or not this variable is always present.

Table 6: Counting Result

Feature	**Count**
$miner\_revenue\_USD$	110
$DowJones$	71
$S\&P$	69
$mvrv$	68
$mean\_difficulty$	56

We can deduce that $EURR002W$ isn't relevant there. The number of occurrence of this variable is : 7. We may infer that this variable lacks significance in understanding the outcomes during the period we selected in the previous example.

We also see that the $miner\_revenue\_USD$ feature is the most important variable but this time with its value and not its returns or its z_score. It's intriguing to note that the second and third values were also the most significant in the other Random Forest. Therefore, we can conclude that this variable holds considerable importance for us. Its prominent position in the analysis is logical, considering its inherent structure.

However, we must be careful with these results, because, as we say before, it's structure is strongly based on the Bitcoin price and movement.

As part of our strategy, we'll see whether it's worth holding on to or whether bitcoin's movements will outweigh its impact on the strategy.

Furthermore, we observe a notable difference compared to the stationary variable analysis: the second and third most important variables are the $DowJones$ and the $S\&P500$, both of which are US stock market indices.

These findings are unsurprising, considering the growing interconnection between the cryptocurrency market, including Bitcoin, and traditional financial markets. The $DowJones$ and the $S\&P500$ are two of the most influential stock indices and are often used as barometers of overall economic health.

So, fluctuations in these indices can influence investment decisions across various assets, including Bitcoin.

6.2.2 Sectors Importance

We also applied this Random Forest on our variables groups, such as we did with the stationary variables. The results are :

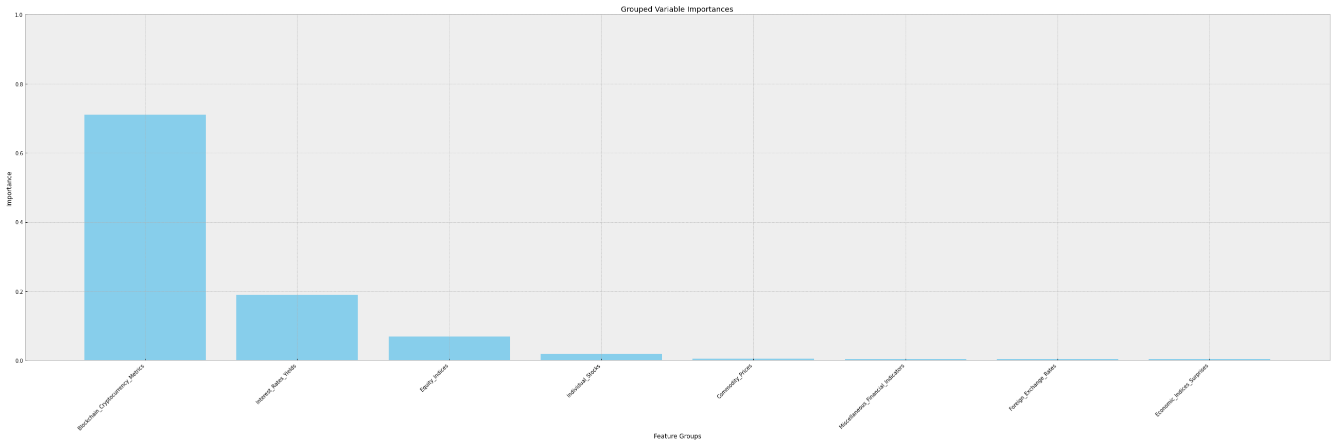

Figure 10: Sector's Importance Example

Table 7: Sector's Importance Result

Sector	Importance
Blockchain Cryptocurrency Metrics	0.709902
Interest Rates Yields	0.189637
Equity Indices	0.069399
Individual Stocks	0.018240
Commodity Prices	0.004257
Miscellaneous Financial Indicators	0.003141
Foreign Exchange Rates	0.002803
Economic Indices Surprises	0.002622

We can observe that the "Blockchain Cryptocurrency Metrics" sector is the most crucial in predicting Bitcoin prices, akin to predicting returns (even though here we have values and volatilities of variables rather than just returns).

The second most important sector is the "Interest Rates Yields" sector, which also held the second-highest importance in the Returns Random Forest. Therefore, the consistency of these top two results is noteworthy, as it suggests we can base our analysis on the variables within these two sectors.

Ultimately, we observe that the "Equity Indices" sector ranks third in importance when considering price, returns, and volatility in our Random Forest analysis. Despite its relatively low importance value (0.069399), it's intriguing because in the other Random Forest analysis, this sector was the least important (0.005344).

At the project's outset, we anticipated that this sector would feature prominently in the results, given that it includes four technological stocks. Logically, we speculated that the technology sector could be linked to the price of Bitcoin. Therefore, it's not surprising for us to see this result, even though its importance is low.

7 Granger Causality

7.1 Construction of the Granger Test

Test from Y to X. Y and X are two time series. We say that Y has a causal relationship in the sense of Granger with X if X can be statistically predicted by past data from Y.
We start by posing two hypotheses :

- H_0: causal effect and H_1: no causal effect.

We specify the number of lags, i.e., we specify that we want to predict the value X_n with the values of $Y(n - \text{lags})$. We then fit a regression model including the lags of the series X and those of Y plus the error, so we can, for example, have :

$X(t) = aX(t-1) + bX(t-2) + \ldots + cY(t-1) + dY(t-2) + \ldots +$ error with a, b, c, d coefficients in $\mathbb{R}$.

In Granger's sense, the series $Y(t)$ causes the series $X(t)$ if conditional on the past values of $Y(t)$, the mean square error of prediction of $X(t+1)$ is lower than when the information relating to the past values of $Y(t)$ was omitted:

$$E[(X(t) - E(X(t)|.))^2 | X(t-1), X(t-2), \ldots ; Y(t-1), Y(t-2), \ldots] \leq E[(X(t) - E(X(t)|.))^2 | X(t-1), X(t-2), \ldots].$$

This test often uses the Fisher test, which is a hypothesis test that tests the null hypothesis that two normal distributions have the same variance.

7.2 Application on the stationary features

We conducted the Granger causality test on the 10 most important features for each period of the RandomForest, using the same period as in the previous example (cf Figure 7). The plot displays the *p_values* of the variables as follows:

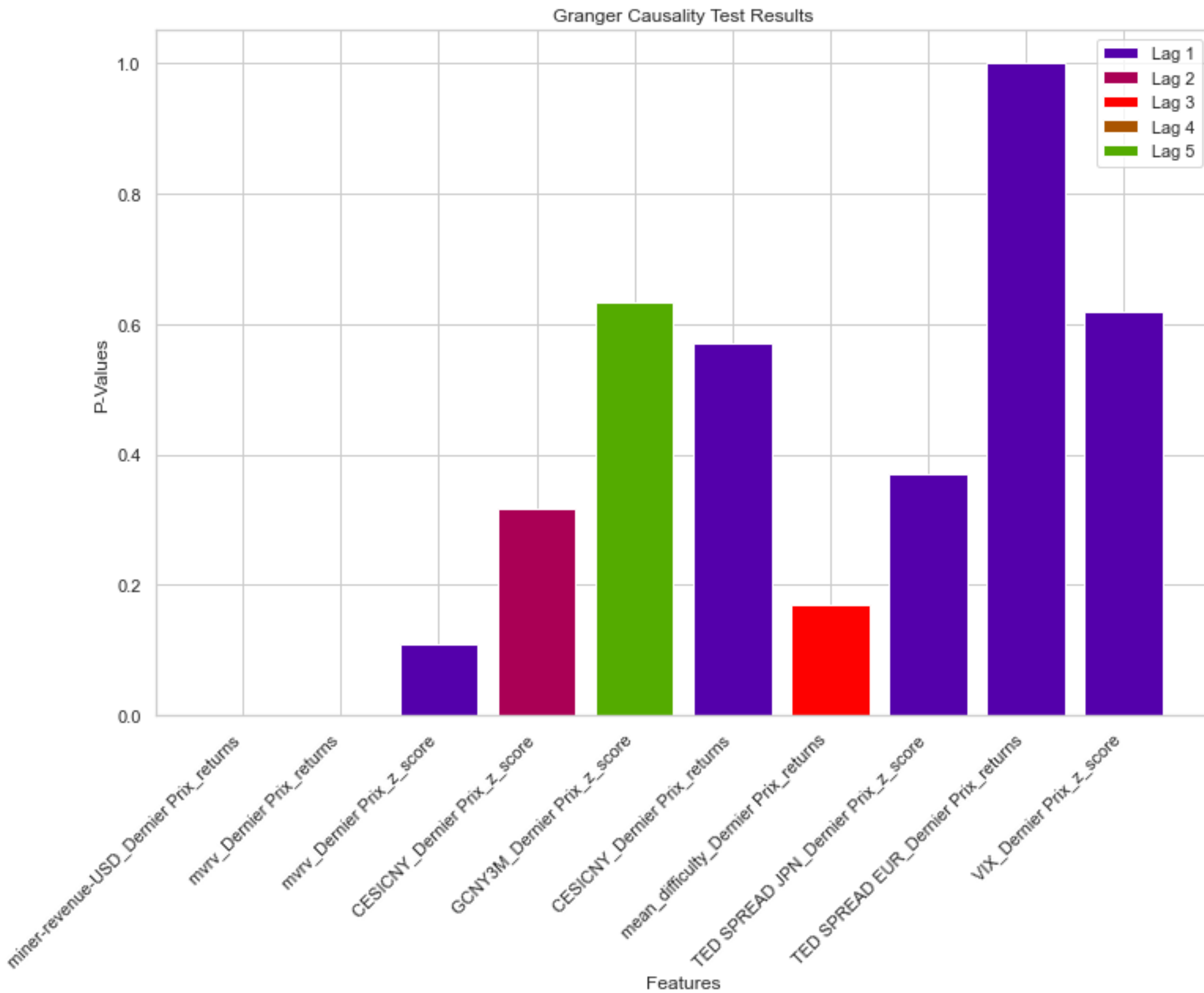

Figure 11: Granger Test p_values for one period

The p_values of the 2 first variables are not 0, this is because the p_values are in a range from 2×10^{-9} to 1.0, so we also did a table with the minimum p_values and the corresponding lag, this table is :

Table 8: P_Values Table

Feature	**p_values**	**lag**
$miner\_revenue\_USD\_returns$	$1.47{\times}10^{-3}$	1
$mvrv\_returns$	$1.99{\times}10^{-9}$	2
$mvrv\_z\_score\_returns$	$1.10{\times}10^{-1}$	1
$CESICNY\_z\_score$	$3.16{\times}10^{-1}$	2
$GCNY3M\_z\_score)$	$6.34{\times}10^{-1}$	5
$CESICNY\_Prix\_returns$	$5.71{\times}10^{-1}$	1
$mean\_difficulty\_returns$	$1.70{\times}10^{-1}$	3
$TED\_SPREAD\_JPN\_returns$	$3.69{\times}10^{-1}$	1
$TED\_SPREAD\_EUR\_returns$	1.00	1
VIX_z_score	$6.20{\times}10^{-1}$	1

A p_value close to 1 suggests that the results observed are very probably due to chance.

A p_value close to 0 indicates that the results observed are unlikely under the null hypothesis.

So, for the $mvrv\_Last\_price\_returns$, we have strong evidence to reject the null hypothesis, supporting the idea that Y has a Granger causal effect on X based on the chosen model and lag specifications.

If we need to treat these results on each period, it will take a long time, so we plot the features with the number of occurrences (+1 if it occurs in the 10 most important variables on a RandomForest period) and we add the p_values to do an average at the end :

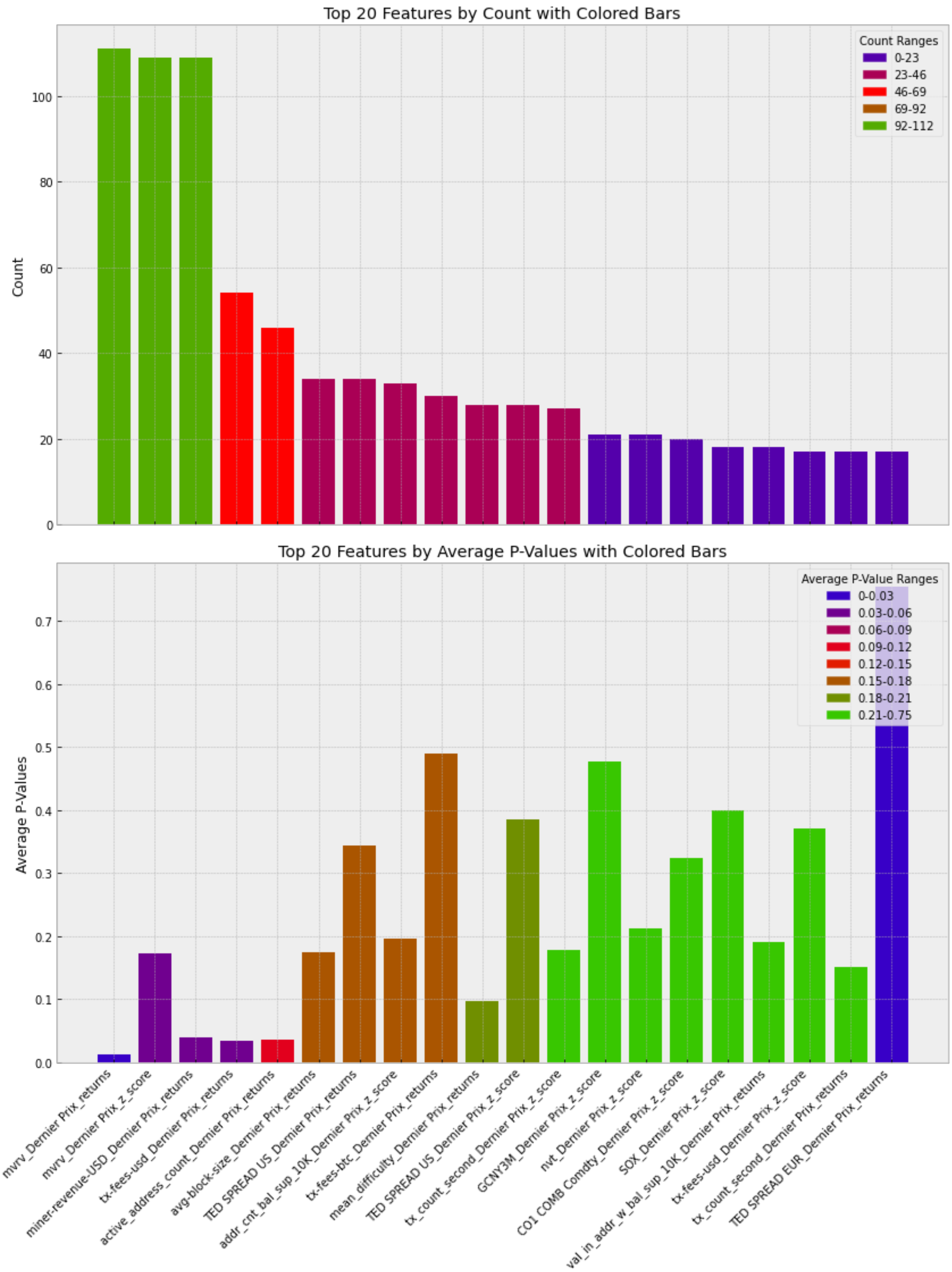

Figure 12: Granger Test Average P_Values result

We can see on these plots that the 5 variables that occurs the most are also the 5 variables with the lowest p_values after applying a Granger causality test (excepted for the *mvrv_Dernier_Prix_z_score* because the test is realized on the *btc_Last_Price_returns* and not on the *btc_Last_Price_z_score*).

The results for the 5 most important variables are :

Table 9: Mean P_Values of the 5 most important variables

Feature	**Count**	**p_values**
mvrv_returns	111	0.013183
mvrv_z_score_returns	109	0.172449
miner_revenue_USD_returns	109	0.040044
tx_fees_usd_returns	54	0.034729
active_address_count_returns	46	0.035643

If we set a significance level at 0.05, we can reject the null hypothesis at this level, supporting the presence of a Granger causal effect of these variables on *btc_Last_Price_returns*.

However, caution is advised for two reasons.

The first is that the Granger causality test is highly controversial among many researchers in machine learning and statistics, such as Marcos Lopez De Pardo, who extensively explore the causality question. According to him in his paper **CAUSAL FACTOR INVESTING (2023)** : "Granger causality may be used as a simple tool to help decide the direction of causal flow between two unconfounded variables (rather than the existence of causal flow)"

The second reason is that we employ a mean method that is not entirely accurate, and it is more precise to delve deeper and conclude causality for each period. For example, the *p_values* for *_Last_price_returns* may be close to 0 from 2015 to 2019, but afterward, they may be close to 0.7 from 2019 to 2023. The results might indicate that the average p-value is below the threshold value (e.g., 0.05), and we would then reject the null hypothesis, accepting Granger causality.

8 Investment Strategies

8.1 Strategy based on price predictions

In this section, we used our Random Forest with **stationary** and **non stationary** features, thus using our weekly raw datas, returns and volatilities to predict Bitcoin **prices**.

8.1.1 Momentum Strategy

The goal of the strategy is to capture the momentums of the features that are important according to our Random Forest. Basically, for each rolling month, we keep the 10 most important features.

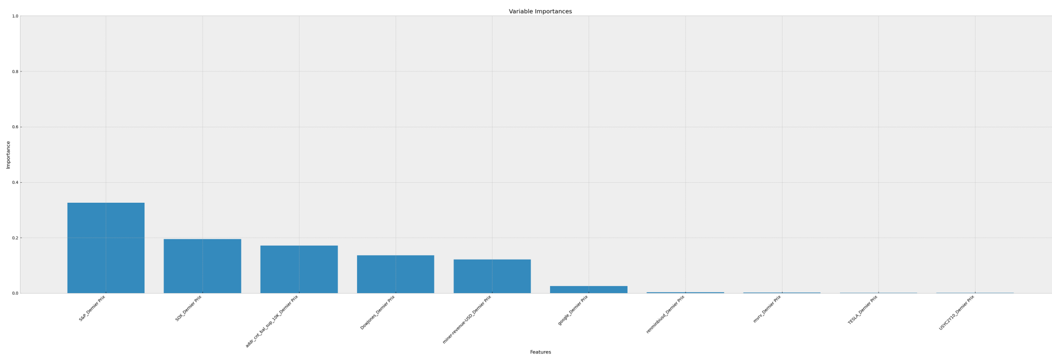

Figure 13: Features importance from June 2022 to March 2023

Once we have these feature that are apparently linked somehow to Bitcoin's price, we compute the 3 month momentums of each one of them.

We then make further calculations in order to normalize our 10 momentums using min max scaler.

Moreover, we weight each one of the 10 momentums by their importance coefficient.

Finally, we sum up these 10 normalized and weighted momentums in order to obtain the total momentum that the 10 most substancial features from a specific timeframe would have.

Let's go through these steps one by one to make things clearer.

Table 10: 10 most important features predicting Bitcoin price of the period: October 2020 - July 2021

Feature	**Importance**	**Momentum**	**Normalized**	**Weighted**
DowJones_Dernier Prix	0.1418	32100.95	0.42	0.0597
Nasdaq_Dernier Prix	0.1417	12309.28	0.39	0.0557
gbpusd_Dernier Prix	0.1417	1.23	0.44	0.0637
NKY_Dernier Prix	0.1074	26765.30	0.32	0.0345
TSEMIL_Dernier Prix	0.0889	4135.63	0.58	0.0517
miner-revenue-USD_Dernier Prix	0.0803	26051946.66	0.41	0.0334
USYC2Y10_Dernier Prix	0.0737	16.15	0.63	0.0471
S&P_Dernier Prix	0.0703	4000.28	0.41	0.0290
SOX_Dernier Prix	0.0553	2842.95	0.54	0.0299
TESLA_Dernier Prix	0.0544	254.56	0.30	0.0164

The 'Momentum' column is calculated by the three month momentum from the last predicted date.
The 'Normalized' column represents the min/max scaled momentums.
Finally, the 'Weighted' column shows the product of the 'Importance' column and the 'Normalized' one.

The sum of all weighted momentums is:

$$\text{Momentum of the timeframe analysis} = \sum_{i=1}^{10} \text{Weighted Momentum}_i = \mathbf{0.421}$$

We computed the momentum from the 10 most important features in a specific window of analysis. The goal of the strategy is now to recalculate every four weeks (one month) our period momentum. We then calculate dynamically the quantiles of our increasing momentums datasets. It allows us to determine signals that will be used for this strategy.

As we are rebalancing our portfolio every month, we base our signals likeso :

Table 11: Signals detections

Date	**Momentum**	**Threshold**	**Signal**
2023-07-16	0.378	$\in [q_{25}, q_{50}[$	Long 1
2023-08-13	0.460	$\in [q_{50}, q_{75}[$	Long 2
2023-09-10	0.450	$\in [q_{50}, q_{75}[$	Long 2
2023-10-08	0.613	$\geq q_{75}$	Long 3
2023-11-05	0.345	$<$ q25	Hold

Finally depending on the signals made of our momentums computations, we reallocate our portfolio for each month rebalancing with different amount of Bitcoin and risk free asset exposure. Here is the different cases of exposures of Bitcoin or 10 years US Government Bonds:

Table 12: Allocations of our portfolio for each signals

Signal	**BTC Exposure**	**USGG10YR**
Hold	0 %	100 %
Long 1	50 %	50 %
Long 2	75 %	25 %
Long 3	100 %	0 %

Strategy Performances Measures

Given an initial investment of **$100**, here is the evolution of our portfolio value from **July 2015** to **December 2023**

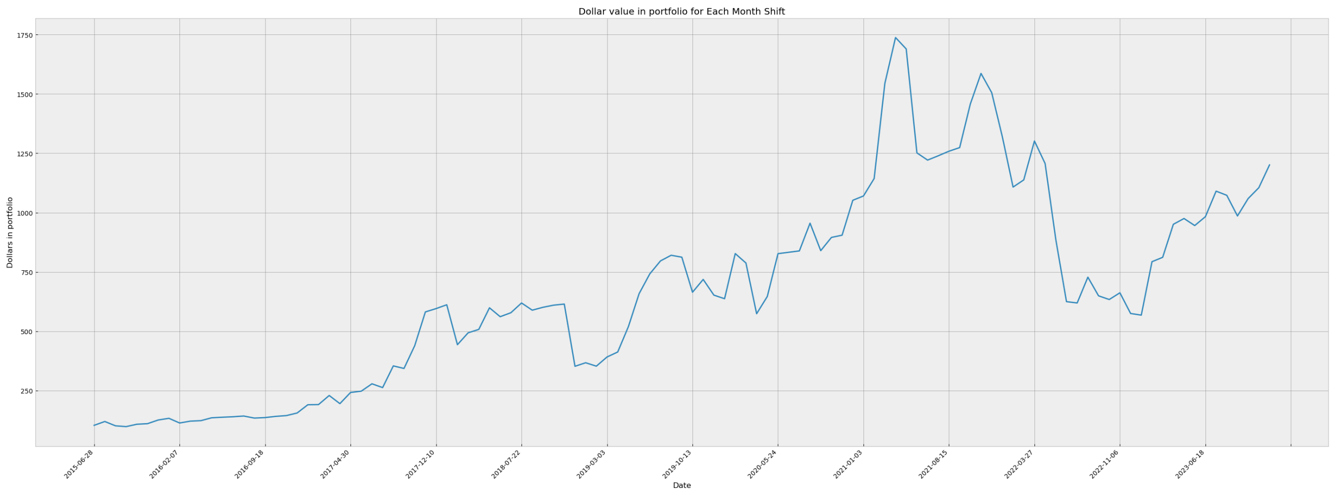

Figure 14: Portfolio value with the momentum strategy

Here are the performances measures of our strategy:

Table 13: Performances measures of our momentum strategy

Metric	**Value**
Annualized Returns	0.466347
Annualized Volatility	0.484937
Sharpe Ratio	0.934049
Maximum Drawdown	-0.673132

Here is the performances measures of a **$100** portfolio fully exposed to **BTC** within the same period:

Table 14: Performances measures of the BTC portfolio

Metric	**Value**
Annualized Returns	1.242468
Annualized Volatility	0.787901
Sharpe Ratio	1.559937
Maximum Drawdown	-0.775469

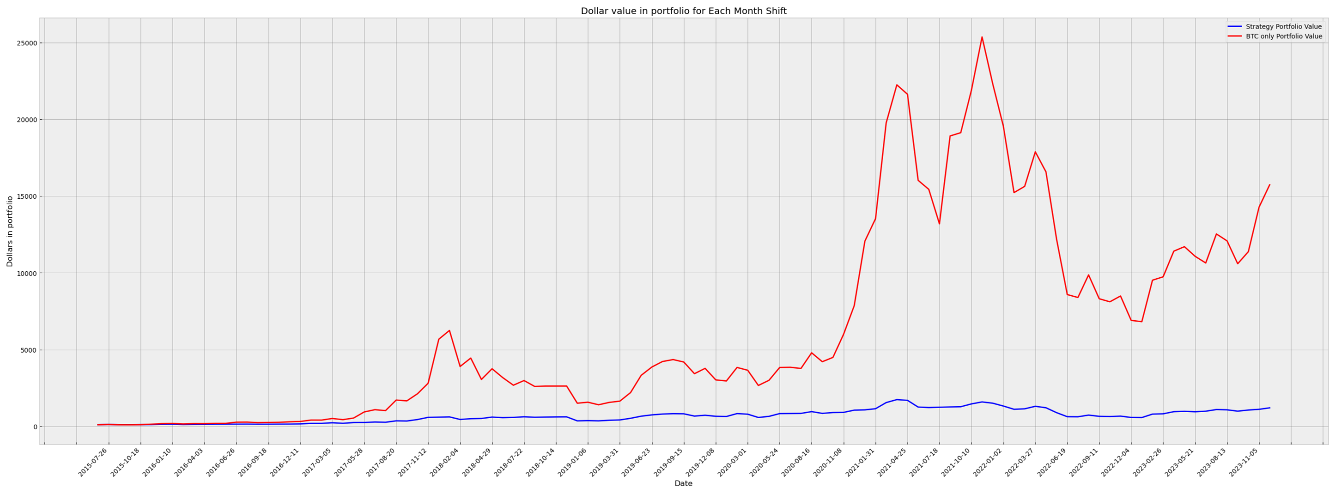

Figure 15: Momentum Strategy VS Full Bitcoin Strategy

Our momentums strategy is definitely not capturing well the upside potential of Bitcoin, reducing by a lot the annualized returns. Nevertheless, we minimized the annualized volatility and the maximum drawdown. The results of this strategy are understandable. In fact, a more effective approach to monetize a strategy on Bitcoin would be to compute its momentums and make investment decisions based on the Bitcoin momentums, rather than relying on the computed relevant features from our Random Forest.

8.1.2 Average price forecast Strategy

The goal of the strategy is to make decisions of exposing our portfolio to Bitcoin regarding the average price we forecast for the last 4 weeks (thus the average monthly price forecasted thanks to our Random Forest).

Here is the forecasted prices over our 111 month of prices predictions over the real value of Bitcoin:

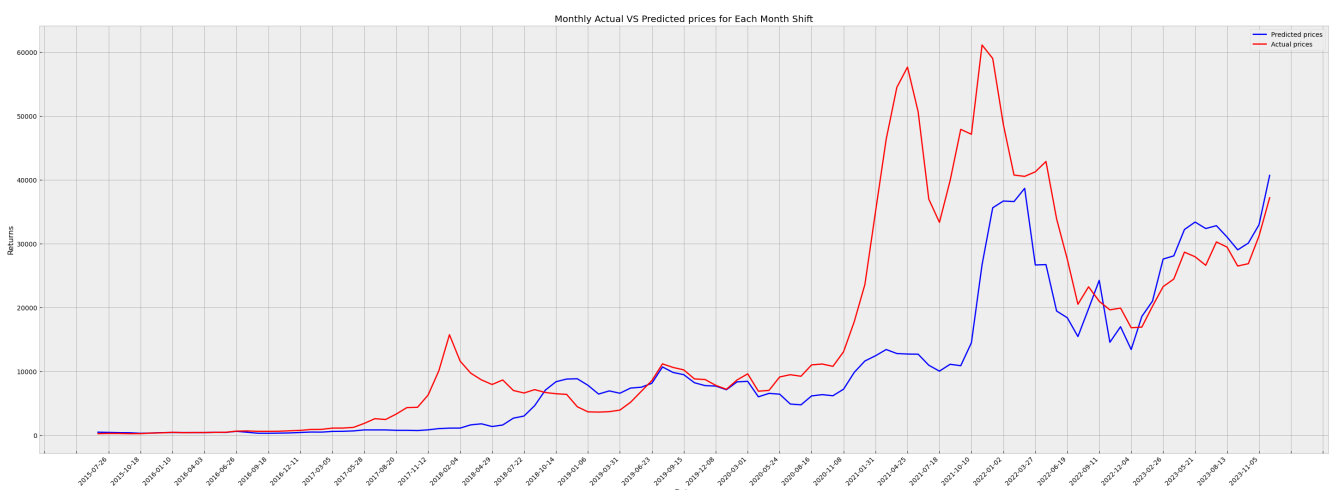

Figure 16: Predicted BTC Returns VS Actual BTC Returns

We then compute monthly returns with our averaged monthly predicted prices and define with different thresholds our signals:

Table 15: Signals detections

Predicted Return Threshold	Signal
$\in [0\%, 2\%[$	Long 1
$\in [2\%, 4\%[$	Long 2
$\geq 4\%$	Long 3
$< 0\%$	Hold

As previously, we allocate our portfolio depending of those signals with the same percentages of Bitcoin and 10 Years US Governments Bonds:

Table 16: Allocations of our portfolio for each signals

Signal	**BTC Exposure**	**USGG10YR**
Hold	0 %	100 %
Long 1	50 %	50 %
Long 2	75 %	25 %
Long 3	100 %	0 %

Strategy Performances Measures

Given an initial investment of **$100**, here is the evolution of our portfolio value from **July 2015** to **December 2023**

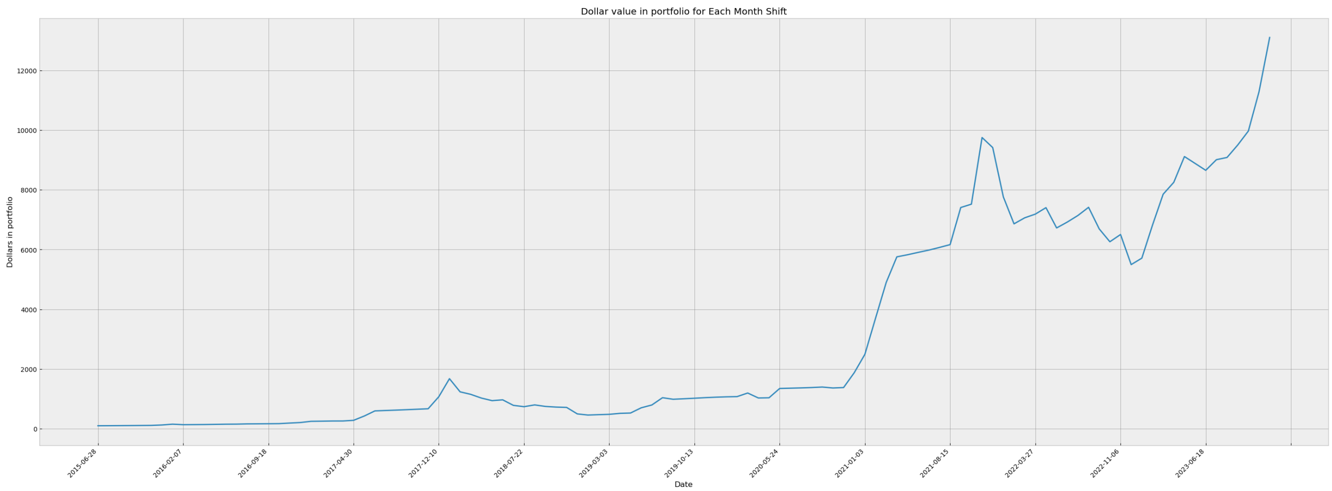

Figure 17: Portfolio value with the average prices forecasting strategy

Here are the performances measures of our strategy:

Table 17: Performances measures of our momentum strategy

Metric	Value
Annualized Returns	0.897776
Annualized Volatility	0.521177
Sharpe Ratio	1.696897
Maximum Drawdown	-0.725264

Here is the performances measures of a **$100** portfolio fully exposed to **BTC** within the same period:

Table 18: Performances measures of the BTC portfolio

Metric	Value
Annualized Returns	1.242468
Annualized Volatility	0.787901
Sharpe Ratio	1.559937
Maximum Drawdown	-0.775469

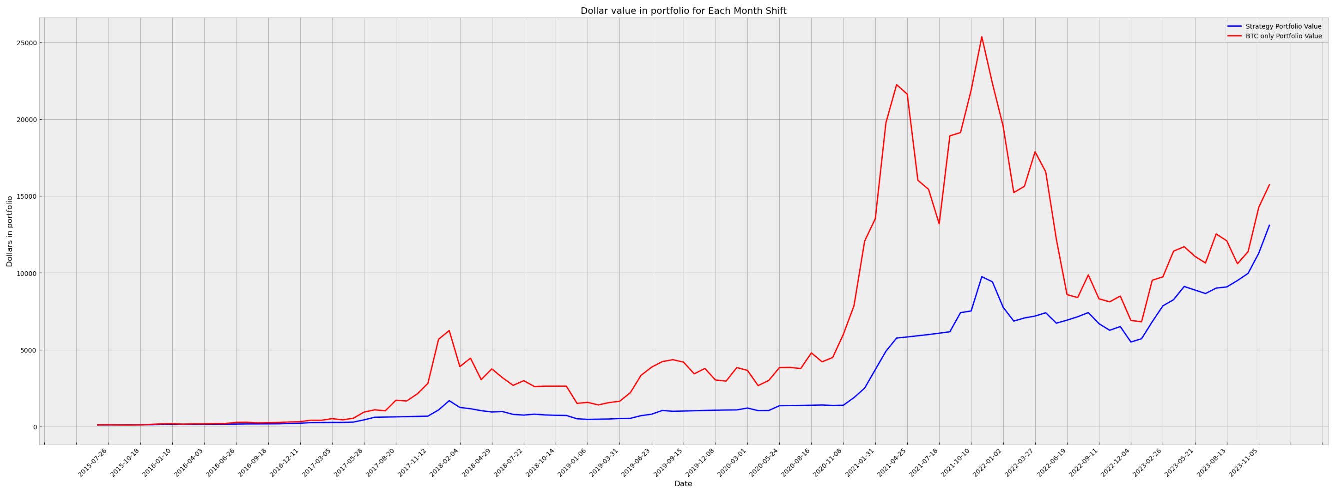

Figure 18: Average Price Forecasting Strategy VS Full Bitcoin Strategy

Our price prediction strategy did pretty well capture Bitcoin's bullish movements while mitigating annualized volatility. Still, a 52.12% annualized volatility strategy would undeniably not be very appealing to any risk averse investor. The risk adjusted performance is improved next to the full Bitcoin strategy.

8.2 Strategy based on returns predictions

In this section, we used our Random Forest with only **stationary** features, thus using our weekly returns and weekly normalized returns (divided by their volatilities) to predict Bitcoin **returns**.

The goal of this strategy is to make decisions of exposing our portfolio to Bitcoin regarding the average of the last 4 weeks of returns forecasting thanks to the Random Forest algorithm.

Here are the predicted monthly returns by our machine learning algorithm compared to the actual Bitcoin returns:

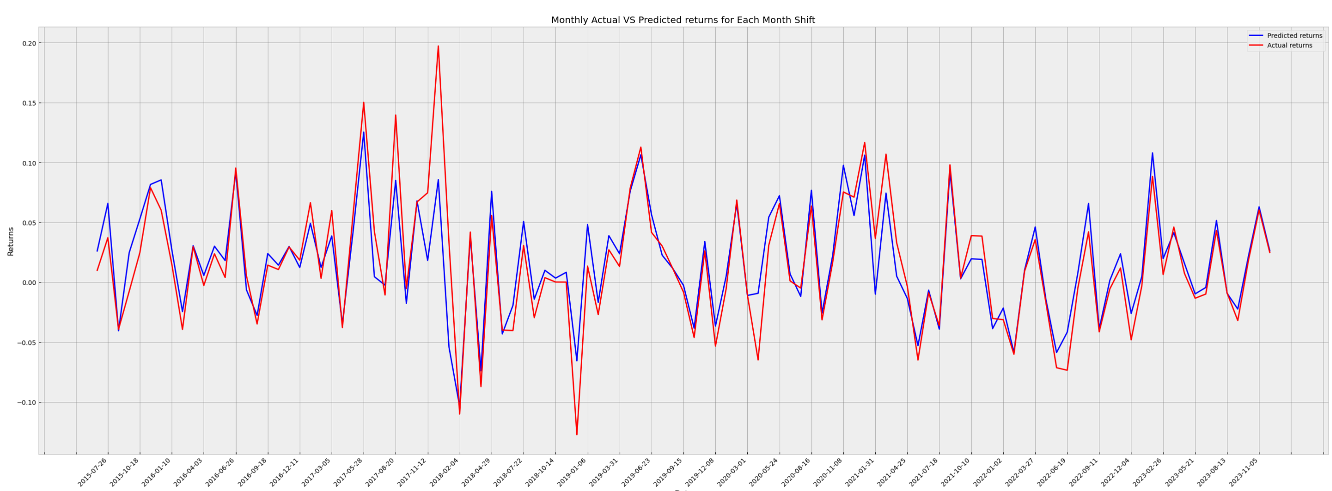

Figure 19: Predicted BTC Returns VS Actual BTC Returns

As previously, we define our investment signals thanks to our forecasted returns:

Table 19: Signals detections

Predicted Return Threshold	**Signal**
$\in [0\%, 2\%[$	Long 1
$\in [2\%, 4\%[$	Long 2
$\geq 4\%$	Long 3
$< 0\%$	Hold

Again, we allocate our portfolio depending of those signals with the same percentages of Bitcoin and 10 Years US Governments Bonds:

Table 20: Allocations of our portfolio for each signals

Signal	**BTC Exposure**	**USGG10YR**
Hold	0 %	100 %
Long 1	50 %	50 %
Long 2	75 %	25 %
Long 3	100 %	0 %

Strategy Performances Measures

Given an initial investment of **$100**, here is the evolution of our portfolio value from **July 2015** to **December 2023**

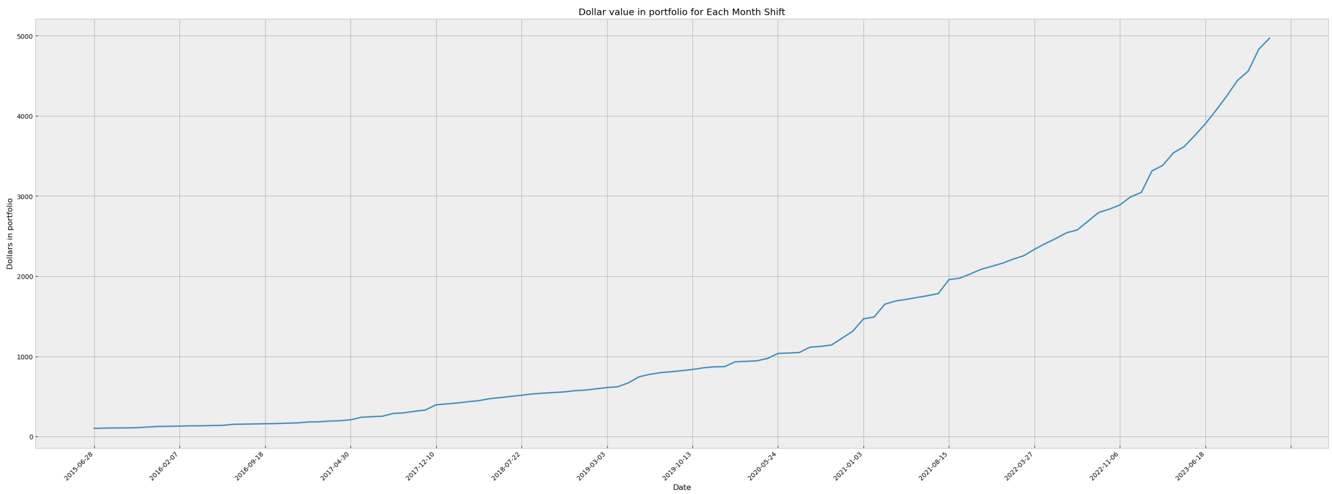

Figure 20: Portfolio value with the average returns forecasting strategy

Here are the performances measures of our strategy:

Table 21: Performances measures of our momentum strategy

Metric	Value
Annualized Returns	0.537686
Annualized Volatility	0.112839
Sharpe Ratio	4.646385
Maximum Drawdown	-0.000615

Here is the performances measures of a **$100** portfolio fully exposed to **BTC** within the same period:

Table 22: Performances measures of the BTC portfolio

Metric	Value
Annualized Returns	1.242468
Annualized Volatility	0.787901
Sharpe Ratio	1.559937
Maximum Drawdown	-0.775469

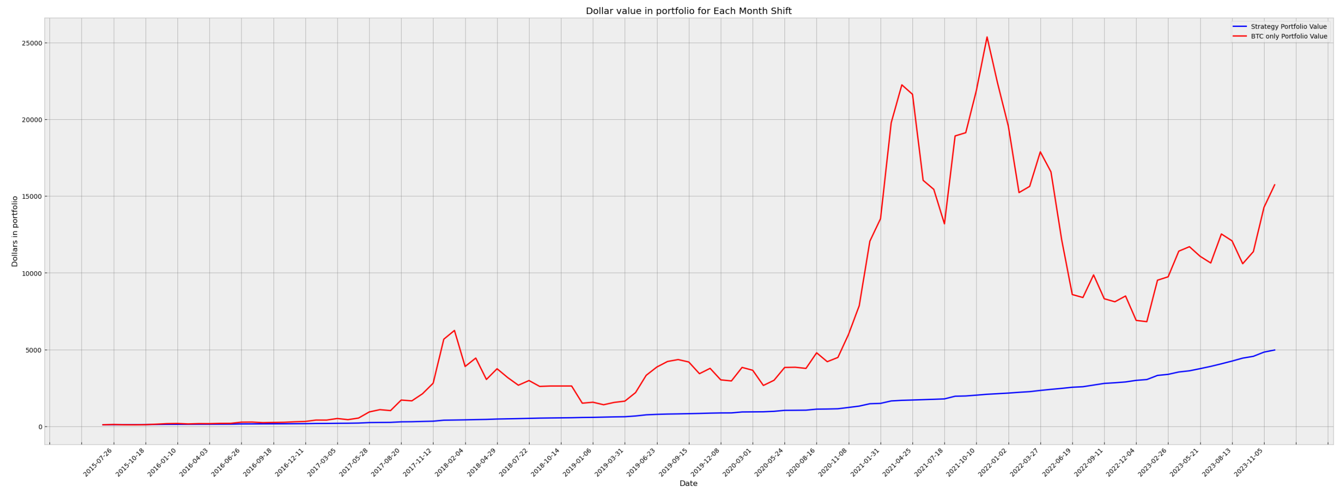

Figure 21: Average Returns Forecasting Strategy VS Full Bitcoin Strategy

The strategy using stationary features seems to predict very accurately the Bitcoin returns. To be honest, it looks non legit as we are anticipating every bullish and bearish returns precisely. Nonetheless, we decrease a lot the annualized returns (still 53.77 % though). The advantage of this strategy is really the reduction of the annualized volatility, shrinking from 78.79 % with the full BTC invested portfolio to 11.28 % with our method.

Obviously, the risk adjusted performance of our approach is undoubtedly high, with a sharpe ratio of 4.65. Ultimately, the maximum drawdown of -0.06 % deduces a constantly increasing value of our portfolio, which sounds inconceivable.

Nevertheless, from our three strategies, this one would definitely be the most appealing one for institutional investors has it completely cuts risks of loss.

8.3 Critical aspect of our investment strategies

To implement our strategies in real-life scenarios, we must also account for the time lag between gathering the latest data and running our forecasting algorithm, and the moment we execute orders to purchase Bitcoin and US Government Bonds. Our data collection process occurs weekly on Sundays at midnight, and our algorithm typically runs for around 15 minutes.
If the investors utilize Centralized Exchanges, the lag would be entirely eliminated for purchasing Bitcoin, as Centralized Exchanges enable continuous trading of crypto assets.
However, if the investors opt for Bitcoin ETFs from funds, there would be a larger lag, as they would need to wait for the markets to open on Monday to execute orders for Bitcoin.

NB: Lag is not an issue for the **10 Years US Gov Bonds** as it's our risk free asset (thus not volatile at all).

When implementing our strategy, we also need to take into account the transaction costs associated with Bitcoin shares in order to know their impact on our performance, and there are several ways of doing this. First of all, we can use Centralised Exchanges (CEX). Here, fees can be categorized into three types: deposit fees, withdrawal fees and transaction fees. These fees can vary from platform to platform, but are often below 0.3% for transactions, with deposit fees, and fees associated with the Bitcoin blockchain network for withdrawals.

This is why it makes more sense to use tokens backed by Bitcoin in Layer 2 (such as wBTC on a layer2 of the Ethereum blockchain, for example), as the associated network fees will be much lower: around a few cents on a layer 2 rather than around ten dollars on the main Bitcoin blockchain. For example, fees on the Binance platform are $0 for the deposit, 0.1% for spot taker/maker transaction fees, and withdrawal fees equal to the fees of the network used (e.g. Bitcoin). For Kraken, transaction fees are 0.26% for spot taker and 0.20% for maker.

Concerning transaction costs from Bitcoin ETF, they vary from 0.24% to 1.5% depending on the issuers. Here are the transaction costs associated with the current existing Bitcoin ETF:

Name	Ticker	Issuer	Fee (after Waiver)	Waiver Details	Exchange	Most Recent Filing
Grayscale Bitcoin Trust(Re-file)*Conversion*	GBTC	Grayscale	1.5%	None	NYSE	1/8/24
ARK 21Shares Bitcoin ETF(Re-filing)	ARKB	21Shares & ARK	0.0% *(0.25%)*	6 Months or $1 Billion	CBOE	1/8/24
iShares Bitcoin Trust	IBIT	BlackRock	0.20% *(0.30%)*	12 Months or $5 Billion	Nasdaq	1/8/24
Bitwise Bitcoin ETP Trust (Re-filing)	BITB	Bitwise	0.0% *(0.24%)*	6 Months & $1 Billion	NYSE	1/8/24
VanEck Bitcoin Trust(Re-filing)	HODL	VanEck	0.25%	None	CBOE	1/8/24
Wisdomtree Bitcoin Trust(Re-filing)	BTCW	Wisdomtree	0.50%	None	CBOE	1/8/24
Invesco Galaxy Bitcoin ETF(Re-filing)	BTCO	Invesco & Galaxy	*0.0% *(0.59%)*	6 Months & $5 Billion	CBOE	1/8/24
Fidelity Wise Origin Bitcoin Trust(Re-filing)	FBTC	Fidelity	0.39%	None	CBOE	1/8/24
Valkyrie Bitcoin Fund(Re-filing)	BRRR	Valkyrie	0.80%	None	Nasdaq	1/8/24
Hashdex Bitcoin ETF *Strategy Change*	DEFI	Hashdex	0.90%	None	NYSE	12/26/23
Franklin Bitcoin ETF	EZBC	Franklin	0.29%	None	CBOE	1/8/24

Source: Bloomberg Intelligence, SEC.gov Bloomberg

Figure 22: Transaction Costs from different BTC ETF Funds

It should be noted that it is not possible to use Decentralized Exchanges (DEX) at present, as it is not possible to exchange FIAT currencies for BTC on these Exchanges - you would have to do everything on-chain, and use Tokenised Treasury Bonds on the blockchain, as well as dollar-indexed tokens (USDC, USDT, etc.). But the users of our strategy are not necessarily comfortable enough with blockchain technology to go through all these steps associated with DEX.

9 Conclusion

The dynamics of Bitcoin are increasingly complex, especially with the influx of new participants in the market, necessitating a constant search for fresh data and features that could explain its changing behavior. Searches of new features must be carried out in macroeconomics indicators, trends among wider risk assets, microeconomics stocks, ...

It's also imperative to regularly update our datasets every time we compute our strategy, ensuring that the most recent data is incorporated. This continuous refreshment of data is vital to accurately reflect the current market conditions, adapt to the rapid changes in the Bitcoin ecosystem, and maintain the relevance and effectiveness of our strategies.

Moreover, it was important to consider in our models the cyclical nature of Bitcoin's evolution, significantly influenced by the periodic halving. These events historically trigger notable market shifts and valuation adjustments. However, the effects of these halving are becoming increasingly nuanced over time, tempered by the expanding and diversifying network of market participants. The forthcoming halving, scheduled for next April, will be a focal point of interest.

Combined with new hedge funds integrating bitcoin ETFs into their portfolios, it offers a unique vantage point to observe the emergent market dynamics and adaptations within the Bitcoin ecosystem. This period promises to unveil new features, spotlighting how Bitcoin's characteristics evolve as it further entrenches itself in the strategic frameworks of institutional finance.

The quest for causality in the relationship between these features and Bitcoin's price dynamics is notably complex. While there is no doubt that there exist features with significant causal links, the presence of confounders and colliders combined with the difficulty of computing relations between them makes the construction of a reliable causality model extremely challenging.

The current analytical tools and the time required to develop a robust model are substantial, and the fast-paced evolution of the BTC market means that any established causality links might soon become outdated. The transient nature of these causal relationships suggests that, even if we successfully identify them, the findings may only be relevant for a short duration before necessitating reevaluation.

This temporal limitation underscores the inherent complexity of the Bitcoin market, where the interplay of various factors is dynamic and can change abruptly. As such, the efficacy of any causality model is inherently constrained by the rapid evolution of market conditions, necessitating continual reassessment and adaptation of our analytical frameworks. The difficulty to pinpoint causal relationships is a testament to the sophisticated and multifaceted nature of the BTC market, reflecting the broader challenges inherent in financial market analysis where variables are numerous, interconnected, and subject to swift changes driven by external factors and internal market dynamics.

Also, an integral facet of understanding Bitcoin's landscape is scrutinizing the regulatory environment, particularly the positions of influential central banks such as the FED, the ECB, the SEC, ...
The regulatory postures and potential future actions of these institutions can profoundly influence Bitcoin's market dynamics. Any action from them might lead to increased market volatility, shifts in investor confidence, and potentially redefined legal and operational frameworks for Bitcoin and other cryptocurrencies. Therefore, any new regulatory stance by central banks must be factored into our models, as well as the new features that will come with these changes.

10 Bibliography

MARCOS LOPEZ DE PRADO : "CAUSAL FACTOR INVESTING: CAN FACTOR INVESTING BECOME SCIENTIFIC?" (2023)

CLIVE W.J. GRANGER, NAMWON HYUNG, YONGIL JEON : "SPURIOUS REGRESSIONS WITH STATIONARY SERIES" (1998)

DEVRIM UNAL, MOHAMMED MUDASSIR, SHADA BENNBAIA, MOHAMMAD HAMMOUDEH : "Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach" (2020)

JUDEA PEARL : "THE FOUNDATIONS OF CAUSAL INFERENCE" (2010)

MARCOS LOPEZ DE PRADO, VINCENT ZOONEKYND : "WHY HAS FACTOR INVESTING FAILED?: THE ROLE OF SPECIFICATION ERRORS" (2024)

TAKASHI NICHOLAS MAEDA, SHOHEI SHIMIZU : "Use of Prior Knowledge to Discover Causal Additive Models with Unobserved Variables and its Application to Time Series Data" (2024)

DONALD B. RUBIN : "ESTIMATING CAUSAL EFFECTS OF TREATMENTS IN RANDOMIZED AND NONRANDOMIZED STUDIES" (1974)

KAY H. BRODERSEN, FABIAN GALLUSSER, JIM KOEHLER, NICOLAS REMY AND STEVEN L. SCOTT : "INFERRING CAUSAL IMPACT USING BAYESIAN STRUCTURAL TIME-SERIES MODELS" (2015)