

Forecasting Digital Asset return: an Application of Machine Learning Model

Abstract:

In this study, we aim to identify the machine learning model that can overcome the limitations of traditional statistical modelling techniques in forecasting Bitcoin prices. Also, we outline the necessary conditions that make the model suitable. We draw on a multivariate large dataset of Bitcoin prices and its market microstructure variables and apply three machine learning models, namely Double Deep Q-learning, XGBoost and ARFIMA-GARCH. The findings show that the Double Deep Q-learning model outperforms the others in terms of returns and Sortino ratio and is capable of one-step-ahead sign forecast of the returns even on synthetic data. This critical insights in forecasting literature will support practitioners and regulators to identify an economically viable cryptocurrency forecasting return model.

Keywords: Digital asset; Forecasting price; Bitcoin; Time-series; Machine Learning; Reinforcement Learning , Double Deep Q-learning

1. Introduction

In recent years, there has been growing interest in Bitcoin investment as the cryptocurrency gains global popularity and acceptance in some countries (Xie et al., 2020; Rehman et al., 2020). There are more than 81 million crypto wallets user across the world as of November 2022 (Statista, 2023). The rapid evolution of Bitcoin trading over the past years has often raised concerns among investors in terms of overvaluation, overreaction, and irrational behavior of the cryptocurrency prices (Amini, 2013; Borgards and Czudaj, 2020; Corbet and Katsiampa, 2020; Mattke et. al 2021). Investors, market practitioners, and regulators have shown vigorous interest in understanding and explaining the movements of cryptocurrency prices in detail (Raimundo et al., 2020; Signature Bank failure, March 2023). Nevertheless, understanding the drivers of changes in cryptocurrency prices remains an open question as the application of econometric and statistical modelling has largely failed to adequately provide actionable insights in forecasting Bitcoin prices (Chen et al., 2021; Wang et al., 2023).

Given that Bitcoin transactions generate large datasets that can provide critical insights, it is therefore important to explore if big data analytical tools such as machine learning could be useful solution to overcome the limitation in forecasting Bitcoin prices (Tofangchiet al., 2021). In addition, cryptocurrencies like Bitcoin are less efficient when compared to the traditional financial assets (Al-Yahyaee et al., 2018), in the context of volatility. Even though, we observe a decrease in this volatility over the time, but, the historical volatilities of Bitcoin remains almost ten times higher than gold and several conventional currencies (Bianchetti et al., 2018). Moreover, Bitcoin, possess a combination of properties of other traditional financial and speculative asset and has a low correlation with other financial instruments traded in the financial market (Klein et al., 2018). Thus, following the literature, we use the highly liquid cryptocurrency, Bitcoin in this study (Amiram et al., 2022). In Appendix D, we summarise the main socio-economic impact of Bitcoin. Over the years, we mainly observe that researchers either focus on one type of value creation mechanism (Kitchens et al., 2018) or adopt the additive approach (Grover et al., 2018). Such approaches are difficult to apply in a complex set-up, such as explaining complex financial relationships (Gradojevic et al., 2021; Newell and Marabelli, 2015). However, in recent years, there have been several recommendations on the

importance of applying big data analytics to generate valuable insights about business operation (Grover et al., 2018). As suggested by Hendershott et al. (2021), the adoption of machine learning models can be a game changer in the context of investment in cryptocurrency trading (Müller et al., 2016). However, application of Machine Learning model invites the challenge of identifying the algorithm that possess the capability to forecast the cryptocurrency return with real time data . Thus, in this research we ask the following question: Does Double Deep Q-learning model outperforms the other popular model (XGBoost and ARFIMA-GARCH) in forecasting cryptocurrency return? Because of wider discussion about prediction efficiency of XGBoost and ARFIMA-GARCH model in forecasting, we decided to compare their performance with the Double Deep Q-learning model.

Extant literature documents the challenges in smoothing and forecasting Bitcoin prices (Milleret et al., 2019; Jana et al., 2021; Kraaijeveld and De Smedt, 2020), especially due to its high volatility (Li et al., 2021; Yaya et al., 2021; Gradojevic and Tsiakas, 2021). Moreover, while advanced machine learning algorithms are capable to deliver exceptional in-sample performances, the ability to generalize out-of-sample remains inherent to a more limited reach (Keilbar and Zhang,2021; Anyfantaki et al., 2021). Out-of-sample performance is the most important performance indicator to find whether a financial model will deliver the expected performance in the real world (Catania and Grassi, 2022; Liang et al., 2020). The lack of logical understanding of outputs generated by complex algorithms often regarded as black boxes and because of such complication the application of machine learning models remains limited. So far, prior literature has largely focused on complex models such as reinforcement learning (Tofangchi et al., 2021). Nevertheless, practitioners such as regulators, and some individual investors are unlikely to favor the implementation of complex, black box deemed algorithms over simpler ones where the relation between cryptocurrency returns and explanatory variables can be easily explained and interpreted. Thus, to contribute to the ongoing discussion on forecasting cryptocurrency returns, we examine the research question by applying machine learning algorithms and displaying various levels of their complexity. In addition, to provide critical insights about the model interpretability, we conduct out-of-sample performance test.

The sample time-series dataset consists of daily, open, spot prices of Bitcoin for the period February 2012–December 2023 sourced from Quandl (3290 daily observations). We randomly select the above period to set up a synthetic dataset for the simulation purpose. In the simulation process, we aim to invest according to the algorithms and calculate the investment performance. We selected Bitcoin in this study because of its popularity, maturity, market position as the leading cryptocurrency (Gradojevic et al., 2021) and for its long-term social impact¹. Our unique finding shows that the Double Deep Q-learning model outperforms the others in terms of returns and Sortino ratio and is capable of one-step-ahead sign forecast of the returns even on synthetic data. According to these results, the success of machine learning models in prediction of cryptocurrency returns is re-established and Double Deep Q-learning model adds an extra layer of confidence about its predictability in the forecasting literature.

Our contributions are threefold - firstly, previous research has conflicting views on the suitability of these models. To our best knowledge this is the first study to resolve debates with empirical evidence on the effectiveness of machine learning models in predicting cryptocurrency returns. Secondly, the study uses the Sortino ratio instead of the Sharpe ratio, focusing on downside volatility to provide a more accurate risk assessment. Thus, it highlights the difference in risk considerations between proprietary traders and investment funds/banks, emphasizing the latter's focus on controlled risk scenarios. Finally, to address the unreliability of historical data (Pintelas et al., 2020) we propose using a Variational Autoencoder to create synthetic datasets for out-of-sample performance evaluation.

The rest of the paper is organized as follows. Section 2 further offers a general overview of the relevant literature and an in-depth discussion of reinforcement learning models. Section 3 outlines the research methodology by discussing the data and the training architectures. Section 4 introduces the out-of-sample backtesting methodology based on synthetic data extracted from the Variational Autoencoder model. Section 5 compares the out-of-sample performances

¹ Socio-economic impact of Bitcoin is summarized in Appendix D

in terms of investment strategies and classification statistics. Section 6 concludes and addresses potential future work.

2. Literature Review

Introduced by Sutton and Barto (1998), the literature of reinforcement learning relishes several extensions that enrich its original scope and application opportunities to various industries (Van Moffaert and Nowé, 2014). The application of reinforcement learning includes self-driving cars, mastering board games such as the AlphaZero chess engine (Silver et al., 2017), etc. Among the extensions of reinforcement learning, the basic Q-learning algorithm (Watkins, 1989) was revised as Double Q-learning and can address an overestimation bias of the basic Q-learning model (Van Hasselt et al., 2016). Moreover, recent reinforcement learning literature focuses on prioritized experience replay to improve data efficiency (Schaul et al., 2015), the dueling network architecture (Wang et al., 2016) and Noisy Double Q-learning (Fortunato et al., 2017) for stochastic network layers to improve exploration. These several contributions are blended into a rainbow model showing that most of the extensions are complementary and capable to produce outperforming performances (Hessel et al., 2018).

On the other hand, financial literature has adopted reinforcement learning models in the recent years. In our survey of literature, we find an excellent application of reinforcement learning in financial markets (Fischer, 2018). Lee et al. (2007) applies multiple Q-learning agents to a stock-trading framework focused on Korean stock market. Jiang et al. (2017) use a thirty-minute cryptocurrency trading strategy and apply an ensemble of identical independent reinforcement learning evaluators based on a convolutional neural network, a recurrent neural network, and a long-short term memory model. Sadighian (2020) applies deep reinforcement learning to create an intelligent market-making strategy testing seven reward functions, extending the previous reinforcement learning market-making models based on time-based event environments. Xiong et al. (2018) show how a Deep Deterministic Policy Gradient can build an optimal portfolio that outperforms the traditional mean-variance asset allocation and a buy and hold strategy on the Dow Jones Industrial Average. Wu et al. (2020) apply the Gated Recurrent Unit model to extract informative financial features that are eventually used to extract intrinsic characteristics of the US

stock market. Besides reinforcement learning, several neural network applications have been deployed to the problems of financial forecasting, portfolio optimization, investment strategies and risk management. In the work of Chen et al. (2003), we observe one of the first applications of neural networks in finance, where they predicted the return direction of the Taiwanese Stock Exchange index by means of a probabilistic neural network and showing its capability to outperform non-neural network-based strategies. In the field of time series forecasting, recurrent neural networks have proven to be particularly useful, thanks to their stateful architecture which allows modelling of serial autocorrelation. Edet (2017), predicts the movements of the S&P 500 index using a recurrent neural network and its variations, namely the long-short-term memory and the gated recurrent unit. They applied the networks to fourteen economic variables and four levels of hidden layers.

Following Baillie et al. (1996) and Gianfreda, & Grossi (2012) we use the ARFIMA-GARCH regression model and Chen & Guestrin (2016) for the XGBoost model. After critically examining the relevant literature, we cannot find any evidence of studies focusing on the out-of-sample performance via synthetic datasets produced with a Variational Autoencoder of a Bitcoin investment strategy based on reinforcement learning, XGBoost and the ARFIMA-GARCH regression model.

In the extant literature, some reinforcement learning approaches, frameworks and models have been proposed, see table in Appendix A. Despite the contributions of these studies, some limitations exist. First, we observe that models in existing studies (see Wu et al., 2020) have largely relied on an out-of-sample performance evaluation on a single set of historical data, making it difficult to generalize the results. In literature, we mainly observe out-of-sample performance evaluation on historical data, with limited focus on synthetic data (Catania and Grassi, 2022). When the training data are highly imbalanced (especially relevant for cryptocurrencies given the highly volatile and leptokurtic distributions), then models using synthetic data could generate more accurate results when applied on real data. One of the most efficient ways to generate a synthetic dataset is by means of a Variational Autoencoder (VAE). In this technique, the encoder compresses the original dataset into a more compact structure, which is, in turn, transmitted to the decoder to

generate an output which represents the original dataset with some noise. The lack of attention by scholars on synthetic datasets motivates us to focus on the out-of-sample performance. Second, we observe that models in existing studies (see Li et al., 2019) target the maximization of total returns or cumulated profits and do not use explanatory variables, rather only focus on time-series dependencies. Algorithms targeting total returns or cumulated profits result in extreme portfolios, with large exposures in a single asset that widely vary over time. On the other hand, targeting a risk-adjusted measure - such as the Sortino ratio - results in more stable, less extreme investment strategies. However, cryptocurrencies are characterized by complex distributions that cannot be explained by their respective univariate time series, rather the usage of explanatory variables is deemed necessary.

3. Methodology

3.1 Reinforcement Learning – model specification

Reinforcement learning is a reward-driven process where an agent learns to interact with a complex environment via trial-and-error to achieve rewarding outcomes (Sutton and Barto, 1998). The agent learns to maximize the reward by choosing the best action in each state of the environment. At the heart of reinforcement learning lies the explore-exploit dilemma. In practice, the agent faces the dilemma of either exploiting what has been learnt thus far or exploring to gain additional knowledge at the risk of recording lower payoffs.

Consider an agent within the environment Ω in discrete time with single step $t = 1, 2, \dots, n$ coupled with the triplet action, state, and reward (a_t, s_t, r_t) . At each time t , the agent is in state s_t and selects an action a_t . The interaction with the environment Ω returns the next reward r_{t+1} and the next state s_{t+1} . The entire set of states and environment rules for transitioning from one state to another may be represented as a Markov decision process. In fact, the current state s_t encompasses all the information needed by the environment for processing state transitions and assigning rewards. Therefore, an agent tries to choose an action $r_t \in A$ that maximizes the expected conditional future reward. This approach is named Q-learning (Watkins, 1989), a form of temporal difference learning (Sutton and Barto, 1998).

Deep reinforcement learning involves the usage of deep neural network architectures to serve as function approximators. A deep-Q-network is a multi-layered neural network $f(x): \mathbb{R}^n \rightarrow \mathbb{R}^m$ that outputs $Q(a_t, s_t)$, where $a_t \in A$, $s_t \in S$ and $r_t \in R$. As a result, the objective of the reinforcement learning becomes learning the optimal set of neural network weights $w_t \in W$ that minimizes a loss function. The latter, however, is an unobservable process which depends on the future combinations of (a_t, s_t) . As such, one needs to solve a dynamic programming algorithm in the form of a Bellman equation.

This optimization mechanism, however, would lead to quickly forgetting rare outcomes as well as it is prone to strongly correlated updates that violate the *i.i.d.* assumption of stochastic gradient descent algorithms. Experience replay (Lin, 1992) addresses these issues as experience is stored in a replay memory from where the network can draw input values, thus potentially including long-term learning and rare outcomes. At the same time, this allows mixing more with less recent experiences for the updates, leading to an update distribution closer to being i.i.d. (Mnih et al., 2015) introduce experience replay to the deep-Q-network architecture. Moreover, it would be more efficient to sample more frequently replay batches where there is more to learn. To do so, (Schaul et al., 2015) introduced prioritized experience replay.

We define the following reinforcement learning environment composed of:

- State $S = [p, h]$: a set including the univariate time-series of prices $p \in \mathbb{R}_+$ and the number of contracts held $h \in \mathbb{R}_+$;
- Action $S = [1, -1]$: a set of actions where 1 represents a buying order and -1 a selling one. The action leads to changes in the holding balance $h \in \mathbb{R}_+$;
- Reward $r(s_t, a_t, s_{t+1})$: the change of the cumulated return of the investment strategy when action a_t is taken in state s_t and eventually leading to the new state s_{t+1} and where $r_t = \ln\left(\frac{p_t}{p_{t-1}}\right)$;
- No contract accumulation is possible; hence a single contract can be traded each time.

The goal is to design an investment strategy that maximizes the Sortino ratio SR_t for the investors in the Bitcoin exchange:

$$SR_t = \frac{r_t - rf_t}{\sigma_{t,semi}} \quad (1)$$

where $\sigma_{t,semi}$ is the semi-standard deviation of the returns generated by the investment strategy, and rf_t is the risk-free rate which we set equal to the three months LIBOR/ SOFR rate. We choose to target the Sortino ratio to limit the downside volatility on the strategy since the Bitcoin market is characterized by frequent and pronounced volatility spikes.

4.2 XGBoost – model specification

XGBoost is an implementation of gradient boosting machines belonging to the broader collection of tools under the umbrella of the Distributed Machine Learning Community. Its widespread adoption followed winning the Higgs Machine Learning Challenge. The XGBoost library provides two wrapper classes that allow the random forest implementation provided by the library to be used with the scikit-learn machine learning library. One of the most important differences between XGBoost and Random Forest is that the XGBoost always gives more importance to functional space when reducing the cost of a model while Random Forest tries to give more preferences to hyperparameters to optimize the model. As such, while the XGBoost model often achieves higher accuracy than decision trees, it sacrifices the interpretability of the explanatory variables. Unlike gradient boosting that works as gradient descent in function space, a second order Taylor approximation is used in the loss function to make the connection to the Newton Raphson method. For an overview of XGBoost models, see Chen and Guestrin (2016).

4.3 GARCH model – model specification

Generalized Autoregressive Conditional Heteroskedasticity (GARCH) is a statistical model used for analyzing time-series data where the variance error is serially autocorrelated. GARCH models assume that the variance of the error term follows an autoregressive moving average process. GARCH models are commonly employed in modelling financial time series that exhibit time-varying volatility and volatility clustering. For an overview of GARCH models, see Bollerslev (1987).

4. Data

The dataset used is a time-series of daily, open, spot prices of Bitcoin futures (BTC) for the period February 2012 – December 2023 sourced from Quandl (3290 daily observations). The choice of the dataset has been driven by data completeness and availability of many explanatory variables. Due to the persistent correlation and Granger causality between Bitcoin prices and other cryptocurrencies (Ghorbe land Jeribi, 2021), investigation of the former allows establishing a gauge over the entire market. In Table 1, we define every explanatory variable used in this study. [Table 1 here]

We fractionally differentiate each time series to achieve stationarity. To estimate the fractional parameter, we used the algorithm of Geweke and Porter-Hudak (2008), whose estimator is based on the regression equation using the periodogram function as an estimate of the spectral density. We trained each model on a training sample composed of 60% of the observations and cross-validated the in-sample estimations by means of k -fold cross-validation, with $k = 5$. Since a Portmanteau test rejected the null hypothesis of identically and independently distributed data at any confidence interval, the cross-validation was purged to take into consideration the serial correlation of the data. As such, in Figure 1 we formed the validation folds with adjacent observations, rather than with randomly picked ones.

[Figure 1 here]

The model specifications are fine-tuned via grid search. For the multilinear perceptron within the reinforcement learning model, we use two hidden layers with leaky RELU activation function, the Glorot kernel initializer, a Ridge regularizer of 0.01 and a Lasso regularizer of 0.01. The training is done via Stochastic Gradient Descent with a Nesterov momentum of 0.6. The loss function is binary cross-entropy. We chose the multilinear perceptron model as more complex models, such as convolutional neural networks or recurrent neural networks, improved the in-sample fitting but worsened the model capability to generalize out-of-sample. In other words, the reduced bias due to enhanced model complexity is more than offset by the increased variance. For the XGBoost, we use a Ridge regularizer of 0.05 and a Lasso regularizer of 0.02, a maximum tree depth of 5 and accuracy as the evaluation metric. In both cases, a validation sample of 20% of the training

observation was used to apply the purged five-folds cross-validation algorithm. For the univariate ARFIMA-GARCH model, we use an eGARCH(1,1) specification with skewed t-student distribution and an ARIMA(2,0,2) for the mean model.

5. Out-of-sample validation methodology

To address the research question, we follow the literature and apply the Double Deep Q-learning, as it avoids the overestimation problem associated with Q learning. Application of machine learning is challenging in case of stock market forecasting because of noisy nature of the historical data. Competitive machine learning approaches mostly act in a supervised manner, ignoring several macro factors affecting the financial market, which leads to over-fitting. As reinforcement learning approaches can learn the process to maximize a return function during the training stage, we can minimize the overfitting problem. Thus, we use a Q-learning agent, which can be trained several times using the same training data and can be important in the real-world stock markets (Carta et al., 2021). However, Double Deep Q-learning might underestimate the action values at times. Since neural networks and machine learning models are prone to overfitting in the training sample, the focus should be on the capability to generalize out-of-sample. For this reason, we evaluate the models on a strict out-of-sample framework based on ten synthetic datasets generated by means of a Variational Autoencoder (VAE), which was introduced by Kingma (2013). The VAE reduces the reconstruction error between the input and output of the network when applied on real data. Thus, VAE improves the generated data quality by minimizing the distribution distances between the real posterior and the estimated one.

The next of this paragraph introduces the VAE model used to produce the synthetic datasets.

Consider a dataset $X = \{X^{(i)}\}_{i=1}^N$ composed of N *i.i.d.* samples coming from a random variable x . Let's assume that the data is generated by a random process involving an unobserved continuous random variable z . The process consists of first generating a value z from some prior distribution $p_{\theta}(z)$ to then generating a value x^i from the conditional distribution $p_{\theta}(x|z)$. Let's assume that the PDFs of $p_{\theta}(z)$ and $p_{\theta}(x|z)$ are differentiable almost everywhere with respect to z, θ . However, the true parameters θ and the values of the latent variable z are unknown. The objective

is to find an efficient neural network approximation for the latent variable z as this would allow to mimic the hidden random process and generate a synthetic dataset that resembles the real data. To do so, we employ a Variational Auto-Encoder. Assume the prior over the latent variables to be a centred isotropic multivariate Gaussian $p_\theta(z) = N(z, 0, I)$. Let $p_\theta(x|z)$ be a multivariate Gaussian with θ estimated via a fully connected neural network with a single hidden layer. The true posterior is intractable but assuming that is approximated by a Gaussian distribution with an approximately diagonal covariance, then the variational approximate posterior is a multivariate Gaussian with a diagonal covariance structure:

$$\log q_\theta(z|x^i) \approx \log N(z, \mu^i, \sigma^i I) \quad (2)$$

where $q_\theta(z|x^i)$ is based on an alternative technique for sampling z such as Monte Carlo and (μ^i, σ^i) are the mean and standard deviation of the approximate posterior which are outputted by the neural network as nonlinear functions of x^i and the variational parameters ϕ .

Afterwards, one simply need to sample from the posterior $z^{i,l} \sim q_\theta(z|x^i)$ with $z^{i,l} = g_\theta(x^i, \epsilon^l) = \mu^i + \sigma^i \epsilon^l$, where $\epsilon^l \sim N(0, I)$. It can be proven that the Kullback-Leibler divergence can be computed without estimation and the resulting estimator for the datapoint x^i is given by:

$$\mathcal{L}(\theta, \phi, x^i) \cong \frac{1}{2} \sum_{j=1}^J \left(1 + \log \left(\left(\sigma_j^{(i)} \right)^2 \right) - \left(\mu_j^{(i)} \right)^2 - \left(\sigma_j^{(i)} \right)^2 \right) + \frac{1}{L} \sum_{l=1}^L \log p_\theta(x^i | z^{i,l}) \quad (3)$$

Where $\log p_\theta(x^i | z^{i,l})$ is a Gaussian fully connected neural network decoding term.

The robustness of the VAE used in our study is coherent with other studies (Camuto et al. 2021). Given this framework, we produce ten multivariate synthetic datasets composed of 1316 observations (40% of out-of-sample observations). Once again, focusing on out-of-sample performance is essential in financial applications to avoid in-sample overfitting. On the other hand, the models have been trained and validated on the 60% of in-sample observations. The model specification is fine-tuned via grid search. For the VAE, we use two hidden layers with five hidden units activated by means of the leaky RELU function, initialised with the Glorot kernel initialiser, with a Ridge regularizer of 0.02 and a Lasso regularizer of 0.01. The training is done via Stochastic Gradient Descent with a Nesterov momentum of 0.6. The loss function targets the representation error via the mean squared error.

Figure 4,5,6,7,8,9, 10 in Appendix B compare each variable’s density plot in the original data with those in the ten synthetic datasets. The grey shaded area is the distribution of the synthetic variable in each dataset, while the red shaded one is the distribution of the same variable in the original dataset. As is visible, the synthetic distributions closely match the original dataset with few discrepancies, which are mostly limited to higher standard deviation around the mean and rare differences in the tails.

6. Results

In this section, we discuss our findings. We follow the extant literature (e.g., Ding et al. 2022; Wang et al., 2023; Kumar et al 2024) to evaluate the performance of our Double Deep Q-learning model and compare its performance with two existing models, namely XGBoost and ARFIMA-GARCH. In line with our aim of offering an approach to forecast Bitcoin prices, we focused our evaluation on ten synthetic datasets in terms of an investment strategy based on their one-step-ahead sign forecasts and in terms of their performances as classifiers. According to Huang (2021) the significance of model-predicted signs is crucial for investment strategies and indicates while ordinary least squares (OLS) estimators generally yield better Sharpe ratios, sign regression can outperform for certain assets. Similarly using sign prediction, Sebastião and Godinho (2021) examine the predictability of digital currency using linear models, random forests, and support vector machines. They show how the combination of multiple models can achieve annualized Sharpe ratios of 80.17% for Ethereum and 91.35% for Litecoin (despite changes in trading costs and market volatility). So, we first illustrate how the sign predictions are translated into investment strategies. Starting from the Double Deep Q-learning model, the output of the learning process is the triplet state, action, and reward (a_t, s_t, r_t) . As such, given a state s_t , the chosen actions $a_t \in [1, -1]$ are directly translated into rewards r_t . Hence, the profit and loss of the investment strategy are the direct output of the learning algorithm.

The XGBoost model, instead, forecasts the one step-ahead probability p_t of the next fractional return r_{t+1} being positive (Chen et al., 2021). Hence, we convert the probability into a trading action by using a static threshold of 0.5 with a buying mechanism triggered when $p_t > 0.5$, and

vice versa. This is a direct yield of having a binary classification as the evaluation objective. Finally, the ARFIMA-GARCH is a univariate regression model that forecasts the one-step-ahead fractional return. Hence, the strategy is to ‘buy’ when the one-step-ahead predicted return, r_{t+1} , is positive and vice versa. The fractional differentiation algorithm does not alter the sign of the price change. As such, increase in price reflects in positive fractional returns, and vice versa. This property is important since both the XGBoost and the ARFIMA-GARCH models use fractional returns of Bitcoin prices as these are not stationary (Almaafi et al, 2023).

Following Dos Santos and Aguilar (2024), at the next stage, we compare the investment performance achieved by the three models on the ten synthetic datasets (also see Arian et al 2024). To showcase the applicability of these models, we add to the comparison a naive buy-and-hold strategy on the BTC. Tables (Table 4,5 and 6) in Appendix C report the descriptive statistics of the returns as well as the Sharpe and Sortino ratios achieved by each strategy in each of the ten out-of-sample synthetic datasets. To improve comparability, we also include the average across the ten datasets. Double Deep Q-learning achieves the highest average annualized mean return of 15.8%. This suggests that the model is capable of generating substantial returns over time. However, high returns often come with higher risk. Investors should evaluate their risk tolerance and ensure they are comfortable with the potential volatility associated with this strategy. While XGBoost outperforms in terms of average annual median return (10.5%) which indicates a more consistent and stable return profile. This could be appealing to investors who prioritize stability and predictability in their investments. In addition, XGBoost may be less prone to extreme outcomes, making it a potentially safer choice for risk-averse investors. Both Double Deep Q-learning and XGBoost outperform the buy-and-hold strategy (7.9% annual mean and 4.8% annual median), ARFIMA-GARCH is the worst performer in terms of both metrics. At the same time, ARFIMA-GARCH strategies achieve unstable performances across the ten datasets compared to the Double Deep Q-learning and XGBoost whose statistics are stable across the synthetic datasets. [Table 2A, B and C here]

Nevertheless, the Double Deep Q-learning records the highest annualized standard deviation across the ten datasets, averaging 33.2% compared to 26.9% of XGBoost and 24% of ARFIMA-

GARCH. Investors might use Double Deep Q-learning as part of a diversified portfolio to balance risk and reward, leveraging its potential for high returns while mitigating overall portfolio risk. All the models have a larger standard deviation compared to the buy-and-hold strategy whose annualized standard deviation is 20.7% (also see Grot et al. 2022). At the same time, Double Deep Q-learning records a large average kurtosis of 9.7x, while XGBoost and ARFIMA-GARCH achieve a kurtosis below three, improving on the 6.7x of the buy-and-hold strategy. However, the larger standard deviation and kurtosis of Double Deep Q-learning are offset by the largest, positive average skewness of 2.4 and the lowest downside volatility of 3.1%. In fact, the objective function of Double Deep Q-learning is to maximize the Sortino ratio, which embeds minimizing the downside volatility while maximizing the returns. This implies that its higher average standard deviation is the result of higher upside volatility. Figure 2 plots the average standard deviations while Figure 3 plots the average downside volatilities both calculated on a rolling window of one hundred days. ARFIMA-GARCH returns, instead, are mostly symmetrical (positive and negative returns are roughly equal in magnitude and frequency) while the downside volatility is the highest across all the synthetic datasets. This means that while the returns are symmetrical, the negative returns (losses) can be quite large and frequent, leading to higher risk during downturns. So, the model's performance could be more erratic during the financial crisis, requiring investors to be vigilant and possibly adjust their strategies accordingly. Investors may need to implement robust risk management strategies to mitigate this downside risk. From these figures, it is evident that both Double Deep Q-learning and XGBoost can effectively curtail the standard deviations and the downside volatility of the investment strategies. From this follows an expectation of superior risk-adjusted performances, such as Sharpe and Sortino ratios, wherein an improved performance is linked to lower volatility, all else equal.

[Figure 2 and 3 here]

In terms of Sharpe ratio, Double Deep Q-learning and XGBoost achieve a similar performance of 0.63x, while ARFIMA-GARCH underperforms due to lower returns not sufficiently offset by lower standard deviation. The findings are consistent with the existing studies (Wang et al., 2020). For the ARFIMA-GARCH strategy the same applies in terms of Sortino ratio. Double Deep Q-

learning, instead, is the best performing model in terms of Sortino ratio (best risk-adjusted returns by focusing on downside risk) on the back of lower downside volatility (meaning it is less likely to experience significant losses, which is crucial for risk-averse investors) and higher returns, followed by the XGBoost strategy which also outperforms the buy-and-hold strategy. This makes the machine learning model an attractive option for investors seeking high returns with controlled risk.

Moving to the performance of the classification models, Table 3 reports statistics of the confusion matrices generated by each classifier. All the confusion matrices are based on the model's ability to correctly classify the sign of the one-step-ahead realized return on the average of the ten out-of-sample synthetic datasets. The sign is extracted in the same way as presented in the previous section. The realized classes in the average out-of-sample datasets are well balanced, with 50.3% of the observations belonging to the buying class and the rest to the selling one. Double Deep Q-learning and XGBoost outperform ARFIMA-GARCH in terms of all the metrics proposed. The latter, in fact, can barely improve on the performance of a random classifier as it records a 51.9% accuracy and a 95% confidence interval lower bound below the 50% threshold. Double Deep Q-learning and XGBoost, instead, achieve fairly similar results. The latter records the highest out-of-sample classification accuracy of 76.35%. Moreover, both Double Deep Q-learning and XGBoost have larger specificity compared to sensitivity. In other words, both models are better suited in identifying days when selling is the best strategy compared to buying (Filos 2019). ARFIMA-GARCH, on the other hand, records higher sensitivity compared to specificity, yet not far enough from the performance of a random classifier.

[Table 3 here]

7. Conclusions

In this study, we investigate whether a machine learning inspired model can successfully forecast cryptocurrency returns. We need machine learning models to find easy explanation for investors and policy makers and also to address the limitations of statistical models explained in the existing recent studies (Chen et al., 2021; Wang et al., 2023). To address the above question, we evaluated the performance of three models, namely Double Deep Q-learning, XGBoost and ARFIMA-GARCH in forecasting Bitcoin prices as well as to a buy-and-hold strategy. Our results show that

the Double Deep Q-learning model outperforms the other models in terms of returns and Sortino ratio while the ARFIMA-GARCH model represented the worst performing model across all tests. In layman terms, by using above models we identified one of the best models that can consider the risk factor appropriately in forecasting and can be suitable for any investors with any level of risk-taking behavior. The findings suggest that in practice, an investment strategy will only be penalized for volatility in a down-moving market, which is a great assurance for investors. It is also important for the policy makers to know that our findings suggest a less damage of economic value of the market during the time of extreme volatility. Based on these results, the study offers three main contributions.

First, our study contributes to the literature on cryptocurrency returns by settling the debates on the suitability of machine learning models in forecasting cryptocurrency returns. In prior finance literature, there are debates about the suitability of machine learning models in forecasting (Sun Yin et al., 2019). While some studies (Chen et al., 2021; Gradojevic et al., 2021) argue that machine learning models can forecast cryptocurrencies returns, others (Christodoulou et al., 2019) disagree, leading to inconclusive debates. Moreover, existing research (Xie et al., 2020) tends to use regression approaches or primarily focus on out-of-sample performance evaluation on a single subset of historical data opening the results to more criticisms as to the suitability of machine learning models. However, in this study, we move beyond the existing research by evaluating the suitability of three machine learning models as well as used the entire historical data of Bitcoin to overcome the single historical snapshot criticism. By doing so, the study offers critical insights that address the criticism of existing research as well as attempts to settle the ongoing debate in the literature.

Second, whereas existing research mainly uses the total cumulated profit as target function, only focusing on the time series of returns and out-of-sample backtesting on historical data, our study contributes new insights by targeting a risk adjusted measure such as the Sortino ratio, which penalizes an investment strategy only for volatility in a down-moving market, as opposed to Sharpe ratio which penalizes for volatility in any market movement. For instance, prior studies (Li et al., 2019) use cumulated profit, which is criticized for not accurately reflecting risk considerations.

However, our approach of risk adjusting using the Sortino ratio offers a more holistic representation of risk targeting cumulated profit results in an investment strategy with extreme allocations, without considering the riskiness of the position. The difference becomes particularly relevant when different types of investors are considered. While, on one hand, proprietary traders are focused on maximizing the profit and loss function of their investment strategies, investment funds and banks are focused on delivering higher returns amid controlled risk scenarios. For these reasons, these institutions are mostly evaluated against risk-adjusted measures, such as the Sortino ratio. A widespread adoption of cryptocurrencies to foster broad societal consequences also passes through the inclusion of these instruments among the traded instruments of such large institutions. Lastly, our study enriches the literature on financial asset forecasting by offering an alternative perspective in forecasting Bitcoin returns. By conducting this study on Bitcoin price forecasting, we enrich the investment literature (Mattke et al., 2021; Mai and Hranac, 2013; Gefen, 2002). The extant literature (Ibrahim et al., 2021) predominantly uses time-series data of cryptocurrency prices, which often do not take other critical peculiarities - such as market microstructure - into consideration. By relying solely on time series data of cryptocurrency price, the results of these studies are sometimes criticized for robustness. Our study overcame this challenge by using a set of explanatory variables (average block size, has rate, transaction volume, transaction per block, transactions excluding popular addresses and number of transactions), in addition to time-series data of cryptocurrency prices. Thus, this approach allows us to take into consideration the peculiar market microstructure of cryptocurrencies. By using explanatory variables to augment limitations in solely relying on time-series data, this study contributes a novel process that advances cryptocurrency returns forecasting research. In addition, Bitcoin time-series exhibit high volatility and leptokurtosis, which, coupled with the short trading history, makes out-of-sample evaluations based on historical data highly unreliable (Pintelas et al., 2020). For this reason, we contribute to the finance literature by proposing the usage of a Variational Autoencoder to simulate the original distribution of the underlying data in ten synthetic datasets and evaluate the out-of-sample performances on these.

Practically, the study also offers some critical insights. First, the results demonstrate that it is possible to use machine learning models to successfully predict cryptocurrency returns. This

means practitioners using and those thinking of using machine learning models can be more confident in applying machine models. It is difficult for the practitioners to use the findings of the prior studies because most of them cannot accurately reflect risk consideration by using cumulated profit (Li et al., 2019). However, we use a more holistic approach where the Sortino ratio is relevant for different types of investors. Second, the results of this study offer practitioners a benchmark and reference point for their application of machine learning models since existing research has only provided anecdotal evidence. Lastly, the study offers some strategies to cope with the cryptocurrency volatility. For instance, we define an environment where an agent learns to choose the best suited between two actions, buy or sell a BTC future contract, during each trading day to maximize the Sortino ratio of the investment strategy. We choose to target the Sortino ratio to limit the downside volatility on the strategy since the Bitcoin market is characterized by frequent and pronounced volatility spikes. Creating a consistently profit-making investment algorithm based on online learning would attract more long-term investors and potentially win the regulatory consensus (Sun Yin et al., 2019) for creating regulated spot trading venues. Therefore, this would result in improved long-term market liquidity and lower market volatility. The virtuous cycle would complete with the more widespread adoption of digital coins, fueling positive societal impact hidden beneath the merely speculative aims. As such, the scope of this manuscript is to propose an online machine learning model, the Double Deep Q-learning, and analyze its performance both in terms of investment strategy and as a classifier. Like all research, this study has some limitations, which presents an avenue for future studies. First, this research only focused on Bitcoin, therefore future studies can use other cryptocurrencies such as Ethereum, Litecoin, Doge coin, etc., validate our findings towards wide generalization. Second, our study used three popular machine learning models, namely Double Deep Q-learning, XGBoost and ARFIMA-GARCH in forecasting Bitcoin prices. Second, future studies can explore further development and refinement of cross-validation methods tailored to financial data, particularly focusing on mitigating overfitting and improving model robustness. It is also important to examine the application of the Combinatorial Purged Cross-Validation (CPCV) method to real-world financial markets to validate its effectiveness and practicality. Finally, financial institutions can explore the validation techniques on regulatory compliance.

References

- Adhami, S. and Guegan, D. [2020], 'Crypto assets: the role of ico tokens within a well-diversified portfolio', *Journal of Industrial and Business Economics* 47(2), 219–241.
- Al-Yahyaee, K. H., Mensi, W., & Yoon, S. M. (2018). Efficiency, multifractality, and the long-memory property of the Bitcoin market: A comparative analysis with stock, currency, and gold markets. *Finance Research Letters*, 27, 228-234.
- Al-Yahyaee, K. H., Mensi, W., Ko, H.-U., Yoon, S.-M. and Kang, S. H. [2020], 'Why cryptocurrency markets are inefficient: The impact of liquidity and volatility', *The North American Journal of Economics and Finance* 52, 101168.
- Almaafi, A., Bajaba, S., & Alnor, F. [2023]. Stock price prediction using ARIMA versus XGBoost models: the case of the largest telecommunication company in the Middle East. *International Journal of Information Technology*, 15(4), 1813-1818.
- Alonso-Monsalve, S., Suárez-Cetrulo, A. L., Cervantes, A. and Quintana, D. [2020], 'Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators', *Expert Systems with Applications* 149, 113250.
- Amini, S., Gebka, B., Hudson, R. and Keasey, K. [2013], 'A review of the international literature on the short-term predictability of stock prices conditional on large prior price changes: Microstructure, behavioral and risk related explanations', *International Review of Financial Analysis* 26, 1–17.
- Amiram, D., Jørgensen, B. N., & Rabetti, D. (2022). Coins for Bombs: The Predictive Ability of On-Chain Transfers for Terrorist Attacks. *Journal of Accounting Research*, 38(1), 483-517
- Ana, R., Ghosh, I. and Das, D. [2021], 'A differential evolution-based regression framework for forecasting bitcoin price', *Annals of Operations Research* pp. 1–26.
- Anyfantaki, S., Arvanitis, S. and Topaloglou, N. [2021], 'Diversification benefits in the cryptocurrency market under mild explosivity', *European Journal of Operational Research*.
- Baillie, R. T., Chung, C.-F. and Tieslau, M. A. [1996], 'Analysing inflation by the fractionally integrated ARFIMA–GARCH model', *Journal of applied econometrics* 11(1), 23–40.
- Arian, H. R., Norouzi M, D., & Seco, L. A. [2024]. Backtest Overfitting in the Machine Learning Era: A Comparison of Out-of-Sample Testing Methods in a Synthetic Controlled Environment. *Available at SSRN*.

- Bianchetti, M., Ricci, C., & Scaringi, M. (2018). Are cryptocurrencies real financial bubbles? Evidence from quantitative analyses. *Evidence from Quantitative Analyses* (February 24, 2018). A version of this paper was published in *Risk*, 26.
- BitGreen [2021], ‘Sustainability on the blockchain’.URL: <https://bitg.org/>.
- Bollerslev, T. [1987], ‘A conditionally heteroskedastic time series model for speculative prices and rates of return’, *The review of economics and statistics* pp. 542–547.
- Borgards, O. and Czudaj, R. L. [2020], ‘The prevalence of price overreactions in the cryptocurrency market’, *Journal of international financial markets, institutions, and money* 65, 101194.
- Borrageiro, G., Firoozye, N. and Barucca, P. [2022], ‘The recurrent reinforcement learning crypto agent’, *CoRR abs/2201.04699*. URL: <https://arxiv.org/abs/2201.04699>.
- Caitlin Ostroff, T. W. S. J. [2021], ‘Bitcoin price slips on Elon Musk’s breakup meme tweet’. URL:<https://www.wsj.com/articles/bitcoin-price-slips-on-elon-musks-breakup-meme-tweet-11622805958>.
- Cardano [2021], ‘Making the world work better for all’.URL: <https://cardano.org/>.
- Catania, L., and Grassi, S. [2022]. ‘Forecasting cryptocurrency volatility’, *International Journal of Forecasting*, 38(3), 878-894.
- Carta, S., Ferreira, A., Podda, A. S., Recupero, D. R., & Sanna, A. (2021). Multi-DQN: An ensemble of Deep Q-learning agents for stock market forecasting. *Expert systems with applications*, 164, 113820.
- Chen, A.-S., Leung, M. T. and Daouk, H. [2003], ‘Application of neural networks to an emerging financial market: forecasting and trading the Taiwan stock index’, *Computers & Operations Research* 30(6), 901–923.
- Chen, T. and Guestrin, C. [2016], XGBoost: A scalable tree boosting system, in ‘Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining’, pp. 785–794.
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H. et al. [2015], ‘XGBoost: extreme gradient boosting’, *R package version 0.4-2* 1(4), 1–4.
- Chen, W., Xu, H., Jia, L. and Gao, Y. [2021], ‘Machine learning model for bitcoin exchange rate prediction using economic and technology determinants’, *International Journal of Forecasting* 37(1), 28– 43.

- Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y. and Van Calster, B. [2019], 'A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models', *Journal of clinical epidemiology* 110, 12–22.
- Corbet, S. and Katsiampa, P. [2020], 'Asymmetric mean reversion of bitcoin price returns', *International Review of Financial Analysis* 71, 101267.
- Dempster, M. A. and Leemans, V. [2006], 'An automated FX trading system using adaptive reinforcement learning', *Expert Systems with Applications* 30(3), 543–552.
- Deng, Y., Bao, F., Kong, Y., Ren, Z. and Dai, Q. [2016], 'Deep direct reinforcement learning for financial signal representation and trading', *IEEE transactions on neural networks and learning systems* 28(3), 653–664.
- Ding, S., Cui, T., & Zhang, Y. [2022]. Futures volatility forecasting based on big data analytics with incorporating an order imbalance effect. *International Review of Financial Analysis*, 83, 102255.
- Dos Santos, R., and Aguilar, J. [2024], 'A synthetic data generation system based on the variational-autoencoder technique and the linked data paradigm', *Progress in Artificial Intelligence*, 1-15.
- Edet, S. [2017], 'Recurrent neural networks in forecasting S&P 500 index', Available at SSRN 3001046 .
- Filos, A. [2019]. Reinforcement learning for portfolio management. arXiv preprint arXiv:1909.09571.
- Fischer, T. G. [2018], Reinforcement learning in financial markets-a survey, technical report, FAU Discussion Papers in Economics.
- Fortunato, M., Azar, M. G., Piot, B., Menick, J., Osband, I., Graves, A., Mnih, V., Munos, R., Hassabis, D., Pietquin, O. et al. [2017], 'Noisy networks for exploration', arXiv preprint arXiv:1706.10295.
- Gartner [2021], 'Survey: Most financial executives not planning to use bitcoin as corporate asset'. URL: <https://www.cfodive.com/news/cryptocurrency-corporate-asset-gartner/595226>.
- Gefen, D. [2002], 'Customer loyalty in e-commerce', *Journal of the association for information systems* 3(1), 2.
- Geweke, J. and Porter-Hudak, S. [2008], 'The estimation and application of long memory time series model', *Journal of Time Series Analysis* 4, 221–238.
- Ghorbel, A. and Jeribi, A. [2021], 'Investigating the relationship between volatilities of cryptocurrencies and other financial assets', *Decisions in Economics and Finance* pp. 1–27.

- Gianfreda, A., & Grossi, L. (2012). Forecasting Italian electricity zonal prices with exogenous variables. *Energy Economics*, 34(6), 2228-2239.
- Gort, B. J. D., Liu, X. Y., Sun, X., Gao, J., Chen, S., & Wang, C. D. [2022]. Deep reinforcement learning for cryptocurrency trading: Practical approach to address backtest overfitting. arXiv preprint arXiv:2209.05559)
- Gradojevic, N. and Tsiakas, I. [2021], 'Volatility cascades in cryptocurrency trading', *Journal of Empirical Finance* 62, 252–265.
- Gradojevic, N., Kukolj, D., Adcock, R., and Djakovic, V. [2021]. 'Forecasting Bitcoin with technical analysis: A not-so-random forest?'. *International Journal of Forecasting* (forthcoming)
- Grover, V., Chiang, R. H., Liang, T.-P. and Zhang, D. [2018], 'Creating strategic business value from big data analytics: A research framework', *Journal of Management Information Systems* 35(2), 388–423.
- Hendershott, T., Zhang, X., Zhao, J. L. and Zheng, Z. [2021], 'Fintech as a game changer: Overview of research frontiers', *Information Systems Research* 32(1), 1–17.
- Hessel, M., Modayil, J., Van Hasselt, H., Schaul, T., Ostrovski, G., Dabney, W., Horgan, D., Piot, B., Azar, M. and Silver, D. [2018], Rainbow: Combining improvements in deep reinforcement learning, in 'Proceedings of the AAAI Conference on Artificial Intelligence', Vol. 32.
- Huang, W. [2021]. Sign prediction and sign regression. *Journal of Investment Strategies*, 9(3).
- Ibrahim, A., Kashef, R. and Corrigan, L. [2021], 'Predicting market movement direction for bitcoin: A comparison of time series modeling methods', *Computers & Electrical Engineering* 89, 106905.
- Jiang, Z., Xu, D. and Liang, J. [2017], 'A deep reinforcement learning framework for the financial portfolio management problem', arXiv preprint arXiv:1706.10059.
- Karisma, K. [2022], 'Comparative review of the regulatory framework of cryptocurrency in selected jurisdictions', *Regulatory Aspects of Artificial Intelligence on Blockchain* pp. 82–111.
- Keilbar, G. and Zhang, Y. [2021], 'On cointegration and cryptocurrency dynamics', *Digital Finance* pp. 1–23.
- Kingma, D. P. and Welling, M. [2013], 'Auto-encoding variational bayes', arXiv preprint arXiv:1312.6114 .

- Kitchens, B., Dobolyi, D., Li, J. and Abbasi, A. [2018], ‘Advanced customer analytics: Strategic value through integration of relationship-oriented big data’, *Journal of Management Information Systems* 35(2), 540–574.
- Klein, T., Thu, H. P., & Walther, T. (2018). Bitcoin is not the New Gold—A comparison of volatility, correlation, and portfolio performance. *International Review of Financial Analysis*, 59, 105-116.
- Kraaijeveld, O. and De Smedt, J. [2020], ‘The predictive power of public twitter sentiment for forecasting cryptocurrency prices’, *Journal of International Financial Markets, Institutions and Money* 65, 101188.
- Kumar, S., Alsamhi, M. H., Kumar, S., Shvetsov, A. V., & Alsamhi, S. H. [2024]. Early MTS Forecasting for Dynamic Stock Prediction: A Double Q-Learning Ensemble Approach. *IEEE Access*.
- Lee, J. W., Park, J., Jangmin, O., Lee, J. and Hong, E. [2007], ‘A multiagent approach to q-learning for daily stock trading’, *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 37(6), 864–877.
- Lee, K., Ulkuatam, S., Beling, P. and Scherer, W. [2018], ‘Generating synthetic bitcoin transactions and predicting market price movement via inverse reinforcement learning and agent-based modeling’, *Journal of Artificial Societies and Social Simulation* 21(3).
- Li, Y., Zheng, W. and Zheng, Z. [2019], ‘Deep robust reinforcement learning for practical algorithmic trading’, *IEEE Access* 7, 108014–108022.
- Li, Z., Dong, H., Floros, C., Charemis, A. and Failler, P. [2021], ‘Re-examining bitcoin volatility: A caviar-based approach’, *Emerging Markets Finance and Trade* pp. 1–19.
- Liang, C., Zhang, Y., Li, X. and Ma, F. [2020], ‘Which predictor is more predictive for bitcoin volatility? and why?’, *International Journal of Finance & Economics*.
- Lin, L.-J. [1992], ‘Self-improving reactive agents based on reinforcement learning, planning and teaching’, *Machine learning* 8(3-4), 293–321.
- Mai, E. and Hranac, R. [2013], *Twitter interactions as a data source for transportation incidents*, technical report.
- Mattke, J., Maier, C., Reis, L. and Weitzel, T. [2021], ‘Bitcoin investment: a mixed methods study of investment motivations’, *European Journal of Information Systems* 30(3), 261–285.
- Michael Wursthorn, T. W. S. J. [2021], ‘Wall Street watchdog keeps bitcoin ETFs on ice’. URL: <https://www.wsj.com/articles/wall-street-watchdog-keeps-bitcoin-etfs-on-ice-11626616801>.

- Miller, N., Yang, Y., Sun, B. and Zhang, G. [2019], ‘Identification of technical analysis patterns with smoothing splines for bitcoin prices’, *Journal of Applied Statistics*.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G. et al. [2015], ‘Human-level control through deep reinforcement learning’, *nature* 518(7540), 529–533.
- Moody, J., Wu, L., Liao, Y. and Saffell, M. [1998], ‘Performance functions and reinforcement learning for trading systems and portfolios’, *Journal of Forecasting* 17(5-6), 441–470.
- Müller, O., Junglas, I., Brocke, J. v. and Debortoli, S. [2016], ‘Utilizing big data analytics for information systems research: challenges, promises and guidelines’, *European Journal of Information Systems* 25(4), 289–302.
- Newell, S. and Marabelli, M. [2015], ‘Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of ‘datification’’, *The Journal of Strategic Information Systems* 24(1), 3–14.
- Paul Vigna, T. W. S. J. [2021], ‘Bitcoin on the balance sheet is an accounting headache for tesla, others’. URL: <https://www.wsj.com/articles/bitcoin-price-slips-on-elon-musk-breakup-meme-tweet-11622805958>.
- Pintelas, E., Livieris, I. E., Stavroyiannis, S., Kotsilieris, T. and Pintelas, P. [2020], Investigating the problem of cryptocurrency price prediction: a deep learning approach, in ‘IFIP International conference on artificial intelligence applications and innovations’, Springer, pp. 99–110.
- Raimundo Junior, G. d. S., Palazzi, R. B., Tavares, R. d. S. and Kotze, M. C. [2020], ‘Market stress and herding: a new approach to the cryptocurrency market’, *Journal of Behavioral Finance* pp. 1–15.
- Ripple [2021], “XRP: The best digital asset for global payments”. URL: <https://ripple.com/xrp/>.
- Sadighian, J. [2020], ‘Extending deep reinforcement learning frameworks in cryptocurrency market making’, arXiv preprint arXiv:2004.06985.
- Samek, W., Wiegand, T. and Müller, K.-R. [2017], ‘Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models’, arXiv preprint arXiv:1708.08296 .
- Schaul, T., Quan, J., Antonoglou, I. and Silver, D. [2015], ‘Prioritized experience replay’, arXiv preprint arXiv:1511.05952.
- Schnaubelt, M. [2022], ‘Deep reinforcement learning for the optimal placement of cryptocurrency limit orders’, *European Journal of Operational Research* 296(3), 993–1006.

- Sebastião, H., & Godinho, P. [2021]. Forecasting and trading cryptocurrencies with machine learning under changing market conditions. *Financial Innovation*, 7, 1-30.
- Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Ku- maran, D., Graepel, T. et al. [2017], ‘Mastering chess and shogi by self-play with a general reinforcement learning algorithm’, arXiv preprint arXiv:1712.01815.
- SolarCoin [2021], ‘Solarcoin is a cryptocurrency that incentivizes a solar-powered planet’.URL: <https://solarcoin.org/>
- Statista [2021], ‘Number of blockchain wallet users worldwide from November 2011 to September 13, 2021’. URL: <https://www.statista.com/statistics/647374/worldwide-blockchain-wallet-users/>
- Stellar [2021], ‘Stellar is an open network for storing and moving money.’.URL: <https://www.stellar.org/?locale=en>.
- Sun Yin, H. H., Langenheldt, K., Harlev, M., Mukkamala, R. R. and Vatrappu, R. [2019], ‘Regulating cryptocurrencies: a supervised machine learning approach to de-anonymizing the bitcoin blockchain’, *Journal of Management Information Systems* 36(1), 37–73.
- Sutton, R. S. and Barto, A. G. [1998], ‘Reinforcement learning: an introduction MIT press’, Cambridge, MA 22447.
- Tofangchi, S., Hanelt, A., Marz, D. and Kolbe, L. M. [2021], ‘Handling the efficiency–personalization trade-off in service robotics: a machine-learning approach’, *Journal of Management Information Systems* 38(1), 246–276.
- Van Hasselt, H., Guez, A. and Silver, D. [2016], Deep reinforcement learning with double q- learning, in ‘Proceedings of the AAAI Conference on Artificial Intelligence’, Vol. 30.
- Van Moffaert, K. and Nowé, A. [2014], ‘Multi-objective reinforcement learning using sets of pareto dominating policies’, *The Journal of Machine Learning Research* 15(1), 3483–3512.
- Wang, Z., Schaul, T., Hessel, M., Hasselt, H., Lanctot, M. and Freitas, N. [2016], Dueling network architectures for deep reinforcement learning, in ‘International conference on machine learning’, PMLR, pp. 1995–2003.
- Wang, L., Ma, F., Liu, J., and Yang, L. [2020], ‘Forecasting stock price volatility: New evidence from the GARCH-MIDAS model’, *International Journal of Forecasting*, 36, 684–694.

- Wang, Y., Andreeva, G., and Martin-Barragan, B. [2023], ‘Machine learning approaches to forecasting cryptocurrency volatility: Considering internal and external determinants’, *International Review of Financial Analysis*, 90, 102914.
- Watkins, C. J. C. H. [1989], ‘Learning from delayed rewards’.
- Wei, W. C. [2018], ‘Liquidity and market efficiency in cryptocurrencies’, *Economics Letters* 168, 21–24.
- Wu, X., Chen, H., Wang, J., Troiano, L., Loia, V. and Fujita, H. [2020], ‘Adaptive stock trading strategies with deep reinforcement learning methods’, *Information Sciences* 538, 142–158.
- Xie, P., Chen, H. and Hu, Y. J. [2020], ‘Signal or noise in social media discussions: The role of network cohesion in predicting the bitcoin market’, *Journal of Management Information Systems* 37(4), 933–956.
- Xiong, Z., Liu, X.-Y., Zhong, S., Yang, H. and Walid, A. [2018], ‘Practical deep reinforcement learning approach for stock trading’, arXiv preprint arXiv:1811.07522.
- Yang, H., Liu, X.-Y., Zhong, S. and Walid, A. [2020], Deep reinforcement learning for automated stock trading: An ensemble strategy, in ‘Proceedings of the First ACM International Conference on AI in Finance’, pp. 1–8.
- Yaya, O. S., Ogbonna, A. E., Mudida, R. and Abu, N. [2021], ‘Market efficiency and volatility persistence of cryptocurrency during pre-and post-crash periods of bitcoin: Evidence based on fractional integration’, *International Journal of Finance & Economics* 26(1), 1318–1335.
- Zhang, J. and Maringer, D. [2016], ‘Using a genetic algorithm to improve recurrent reinforcement learning for equity trading’, *Computational Economics* 47(4), 551–567.
- Zhang, Z., Zohren, S. and Roberts, S. [2020], ‘Deep reinforcement learning for trading’, *The Journal of Financial Data Science* 2(2), 25–40.

Table 1 The variables composing the dataset used in this study

Name	Description
BTC	Bitcoin price in USD
AVBLS	Average block size
HRATE	Hash rate (diff.)
ETRAV	Estimated transaction volume (diff.)
NTRBL	Transactions per block (diff.)
NADDU	Transactions excluding popular addresses (diff.)
NTREP	Number of transactions (diff.) -1.6

Note: In Table 1 we define the variables used in this study. The abbreviations are defined in the second column and in other tables and in the text we used the abbreviations

Table 2A Annualized descriptive statistics for Double Deep Q-learning investment strategy

Sample	Mean	Median	Std. deviation	Skewness	Kurtosis	Volatility Skew	Downside volatility	Sharpe ratio	Sortino ratio
V1	15.629%	8.214%	33.005%	1.990	6.023	60.054	3.011%	0.592	4.190
V2	15.385%	7.950%	33.718%	2.194	6.150	61.466	3.258%	0.589	4.333
V3	15.811%	7.746%	31.953%	1.844	5.455	64.304	2.735%	0.649	4.258
V4	14.885%	7.646%	32.691%	2.493	6.517	68.208	3.085%	0.655	4.412
V5	15.762%	6.941%	34.191%	2.310	5.753	71.336	3.213%	0.636	4.422
V6	15.666%	8.247%	33.215%	1.931	5.866	66.233	2.698%	0.627	4.097
V7	15.915%	8.146%	33.869%	2.583	6.273	66.259	3.104%	0.603	4.386
V8	15.865%	7.123%	33.040%	2.321	6.483	67.120	2.948%	0.611	4.380
V9	15.756%	7.016%	33.817%	1.762	5.636	71.153	2.739%	0.618	4.180
V10	15.396%	7.246%	33.684%	2.358	6.912	69.235	3.232%	0.598	4.137
Average	15.607%	7.627%	33.318%	2.179	6.107	66.537	3.002%	0.618	4.279

Notes: In Table 2A, we report the annualized descriptive statistics for Double Deep Q-learning investment strategy in each of the ten out-of-sample synthetic datasets. V1-V10 denote each of the out of sample synthetic datasets. In the last row of the table, we included the average of the ten datasets.

Table 2B: Annualized descriptive statistics for **XGBOOST** investment strategy

Sample	Mean	Median	Std. deviation	Skewness	Kurtosis	Volatility Skew	Downside volatility	Sharpe ratio	Sortino ratio
V1	16.169%	7.398%	35.145%	2.167	6.239	72.821	2.938%	0.598	4.657
V2	16.086%	7.523%	32.517%	1.761	5.691	64.059	3.240%	0.591	4.475
V3	16.419%	6.651%	33.157%	2.129	6.938	67.732	2.740%	0.675	4.630
V4	16.439%	7.066%	32.237%	2.275	7.721	69.766	3.341%	0.621	4.436
V5	15.751%	7.879%	34.236%	2.055	5.461	64.796	3.121%	0.604	4.301
V6	15.750%	7.197%	34.442%	2.738	6.406	71.401	3.196%	0.620	4.606
V7	15.946%	6.335%	33.251%	2.448	5.630	65.679	2.771%	0.618	4.355
V8	15.624%	7.871%	31.230%	2.180	5.977	69.754	3.199%	0.645	5.277
V9	15.551%	6.578%	31.988%	2.286	6.716	56.727	3.325%	0.660	4.693
V10	16.288%	7.903%	31.758%	2.391	7.046	68.861	2.790%	0.651	4.507
Average	16.002%	7.240%	32.996%	2.243	6.382	67.160	3.066%	0.628	4.594

Notes: In Table 2B, we report the annualized descriptive statistics for **XGBOOST** investment strategy in each of the ten out-of-sample synthetic datasets. V1-V10 denote each of the out of sample synthetic datasets. In the last row of the table, we included the average of the ten datasets.

Table 2C: Annualized descriptive statistics for **ARFIMA-GARCH** investment strategy

Sample	Mean	Median	Std. deviation	Skewness	Kurtosis	Volatility Skew	Downside volatility	Sharpe ratio	Sortino ratio
V1	15.142%	7.900%	38.452%	2.544	6.850	68.406	4.136%	0.573	4.182
V2	13.772%	11.948%	36.895%	2.329	9.079	78.723	3.233%	0.598	4.090
V3	14.834%	11.427%	34.989%	2.248	9.985	74.034	4.219%	0.542	3.956
V4	13.769%	9.690%	35.835%	3.125	9.272	86.801	4.164%	0.533	3.264
V5	13.921%	9.357%	34.081%	3.700	6.567	68.946	3.316%	0.519	3.329
V6	15.352%	11.693%	34.977%	4.191	7.457	69.753	4.222%	0.611	3.602
V7	14.423%	9.612%	43.990%	2.840	6.558	81.582	3.699%	0.582	3.826
V8	14.697%	8.286%	33.505%	3.701	7.763	75.759	3.305%	0.488	3.470
V9	11.811%	11.263%	34.425%	2.438	7.026	71.307	4.049%	0.493	3.321
V10	15.682%	10.647%	33.529%	3.779	6.954	66.600	4.214%	0.601	2.581
Average	14.340%	10.182%	36.068%	3.089	7.751	74.191	3.856%	0.554	3.562

Notes: In Table 2 C, we reported the annualized descriptive statistics for **ARFIMA-GARCH** investment strategy in each of the ten out-of-sample synthetic datasets. V1-V10 denote each of the out of sample synthetic datasets. In the last row of the table, we included the average of the ten datasets.

Table 3 Statistics of confusion matrices

	Double Deep Q-Learning	XGBoost	ARFIMA-GARCH
Accuracy (%)	75.27	76.35	51.93
95% CI lower (%)	77.6	78.34	54.68
95% CI upper (%)	72.8	73.94	49.17
Sensitivity (%)	74.11	74.57	52.4
Specificity (%)	76.42	78.1	51.45
Balanced Accuracy (%)	75.26	76.34	51.93

Note: Table 3 reports statistics of the confusion matrices generated by each classifier. The classifiers are Double Deep Q-Learning; XGBoost and ARFIMA- GARCH. The statistics are reported in percentage.

Appendix

A. Comparative Table for existing literature

Articles	Reinforcement learning model	Application domain	Limitations
Wu et al., (2020)	Gated Deep Q-learning	Equity single stock	Out of sample performance evaluated on a single set of historical data.
Borrageiro et al., (2022)	Recurrent Reinforcement Learning	Bitcoin versus US Dollars; trading perpetual swap derivatives contract	Monte Carlo simulation of 250 trials, obtained reasonable variability of returns. No out-of-sample data analysis.
Zhang and Maringer, (2016)	Genetic Algorithm- Re-current Reinforcement Learning (GA-RRL)	Daily prices, trading volume, price-earning, price-cash flow, debt-market value of S&P 500US firms	GA-RRL trading system did not out-perform the buy-and-hold strategy by producing a greater number of positive Sharpe ratio.
Li et al., (2019)	Deep Reinforcement Learning	Equity single stock	Out of sample performance evaluated on a single set of historical data. Low volatility time series removed. Maximizes cumulated profit. Cumulates large positions (n contracts). No explanatory variables (Time-dependency)

Zhang et al., (2020)	Deep Q Reinforcement Learning	Multi assets future contracts	Out of sample performance evaluated on a single set of historical data. Maximizes cumulated profit. Cumulates large positions (n contracts). No explanatory variables (Time-dependency)
Deng et al., (2016)	Fuzzy Deep Direct Reinforcement	Equity indexes and commodity futures	Out of sample performance evaluated on a single set of historical data. Maximizes cumulated profit. Cumulates large positions (n contracts). No explanatory variables (Time-dependency)
Moody et al., (1998)	Recurrent Reinforcement Learning	Equity index	Out-of-sample performance evaluated on a single set of historical data.
Yang et al., (2020)	Deep Q-Reinforcement Learning	Equity Index	Out-of-sample performance evaluated on a single set of historical data. Maximizes cumulated profit.
Lee et al., (2018)	Inverse Reinforcement Learning	Bitcoin	Out-of-sample performance evaluated on multiple sets of historical data. Maximizes cumulated profit.
Adhami and Guegan (2020)	DCC, ADCC	Multi assets, cryptocurrencies (bitcoin, tokens)	Evolution of the economic impact of ICOs on the real economy and financial stability is still to be tested. No use of machine learning.

Schnaubelt, (2022)	Backward-induction Q-learning, Deep double Q-networks	Multiple assets	Used out-of-sample performance of reinforcement learning algorithms and benchmark strategies. No implementation of feature representations from the data using CNN and routing orders to multiple exchanges
Alonso-Monsalve et al., (2020)	Compares the performance of four different network architectures	Multiple cryptocurrency forecasting	For Dash and Ripple because of noise and temporal behavior. The data generation parameters are not sufficient. Short-term trend prediction has its own limitation with network architecture.
Dempster and Leemans, (2006)	Adaptive Reinforcement learning. The parameters are dynamically optimized to maximize a trader's utility. Adaptive reinforcement utility	Historical data on foreign exchange markets	Risk management layer can be extended to control several automated FX trading systems that trade different currencies. Out-of-sample cumulative profit measured.

Note: Table in Appendix A summarises the most relevant literature related to our study

Appendix B Synthetic distributions (obtained via VAE) vs original ones [Figure 4-10]

Appendix C Annualized descriptive statistics for investment strategies

Table 4 Annualized descriptive statistics for XGBoost investment strategy

	Mean	Median	Std. deviation	Skewness	Kurtosis	Volatility Skew	Downside volatility	Sharpe ratio	Sortino ratio
V1	12.96%	11%	27.87%	84	4.2	9.25	6.44%	0.62	1.68
V2	13.71%	11.34%	26.03%	78	3.01	12.75	5.39%	0.7	2.12
V3	11.09%	9.56%	27.94%	60	2.16	6.68	7.17%	0.53	1.29
V4	12.81%	10.62%	25.84%	66	2.29	10.82	5.66%	0.66	1.89
V5	13.14%	10.10%	27.43%	71	2.53	10.39	6.06%	0.63	1.81
V6	13.44%	10.91%	26.79%	64	1.71	11.65	5.69%	0.66	1.97
V7	12.79%	10.64%	26.98%	71	2.36	9.75	6.12%	0.63	1.74
V8	12.98%	11.28%	26.72%	79	2.59	10.91	5.80%	0.64	1.87
V9	13.07%	10.39%	25.85%	47	1.48	10.21	5.85%	0.67	1.86
V10	12.32%	9.65%	27.92%	104	5.02	9.3	6.34%	0.58	1.62
Average	12.83%	10.55%	26.94%	72	2.73	10.17	6.05%	0.63	1.78

Note: Tables 4 reports the descriptive statistics of the returns as well as the Sharpe and Sortino ratios achieved by XGBoost investment strategy in each of the ten out-of-sample synthetic datasets. To improve comparability, we also include the average across the ten datasets. V1-V10 denotes the out of sample synthetic datasets.

Table 5 Annualized descriptive statistics for ARFIMA-GARCH investment strategy

	Mean	Median	Std. deviation	Skewness	Kurtosis	Volatility Skew	Downside volatility	Sharpe ratio	Sortino ratio
V1	3.49%	1.85%	17.38%	0.27	3.08	1.16	7.44%	0.03	0.04
V2	5.59%	4.02%	26.34%	-0.04	2.13	1.07	11.54%	0.03	0.04
V3	4.92%	3.07%	23.21%	-0.08	1.93	1.09	10.12%	0.03	0.04
V4	6.27%	11.28%	24.31%	-0.18	1.79	1.03	10.74%	0.03	0.05
V5	4.83%	0.21%	21.49%	0.22	1.81	1.21	9.11%	0.03	0.04
V6	15.06%	14.01%	30.66%	-0.02	1.32	1.22	13%	0.06	0.1
V7	2.82%	6.75%	25.13%	-0.13	1.89	0.98	11.23%	0.01	0.02
V8	1.79%	3.59%	19.89%	-0.14	2.08	0.98	8.89%	0.01	0.02
V9	5.56%	1.86%	29.51%	0.05	1.11	1.12	12.77%	0.02	0.04
V10	4.83%	9.29%	22.60%	-0.22	3.92	0.98	10.11%	0.03	0.04
Average	5.52%	5.59%	24.05%	-0.03	2.11	1.08	10.50%	0.03	0.04

Note: Tables 5 reports the descriptive statistics of the returns as well as the Sharpe and Sortino ratios achieved by ARFIMA-GARCH investment strategy in each of the ten out-of-sample synthetic datasets. To improve comparability, we also include the average across the ten datasets. V1-V10 denotes the out of sample synthetic datasets.

Table 6 Annualized descriptive statistics for Bitcoin buy-and-hold investment strategy

	Mean	Median	Std. deviation	Skewness	Kurtosis	Volatility Skew	Downside volatility	Sharpe ratio	Sortino ratio
BTC	7.96%	4.79%	20.70%	1.21	6.73	8.71	4.69%	20.2	1.41

Note: Tables 6 reports the descriptive statistics of the returns as well as the Sharpe and Sortino ratios achieved by Bitcoin buy -and hold investment strategy.

Appendix D

Socio-Economic Impact of Bitcoin

In 2021, the traditional Bitcoin mining process came under severe scrutiny due to its intense energy consumption and heavy environmental impact. In May 2021, Tesla's chief executive officer Elon Musk tweeted that he had suspended accepting bitcoin as a means of payment amid concerns about the rapidly increasing use of fossil fuels for Bitcoin mining and transactions (Caitlin Ostroff, 2021). However, the aim of most cryptocurrencies is to create a long-lasting, positive societal impact (Islam et al., 2021). Following is a list of few prominent examples, Cardano (2021) is a decentralized blockchain platform using an environmentally sustainable blockchain network by relying on less energy-consuming cryptocurrency miners. SolarCoin (2021) mines a new coin for each megawatt hour generated from solar technology. BitGreen (2021) uses a low-energy Proof of Stake algorithm. Stellar (2021) involves no transaction costs for using its network enabling more cost-effective cross-country transactions. Ripple (2021) uses the Ripple Protocol Consensus Algorithm wherein at least 80% of the network's validators must approve a transaction before being added to the Ripple ledger, resulting in a secure network, and allowing users to transfer money with little expense and greater speed.

Enhancing the liquidity on cryptocurrency markets would result in lower volatility and wider long-term cryptocurrency adoption, hence creating a durable social imprint beyond the merely speculating scopes. Al-Yahyaee et al. (2020) find that higher market liquidity diminishes market volatility, in turn resulting in enhanced market efficiency via a more fluid non-arbitrage pricing activity. Market efficiency is more pronounced in highly liquid cryptocurrency markets as active traders are most likely to arbitrage away any sign of return predictability (Wei et al., 2018).

The scarce liquidity also halts a wider cryptocurrency adoption by long-term investors such as large corporates. In fact, as per the current accounting standards, digital coins are considered intangible assets with an indefinite life. As such, if the mandatory periodic impairment tests evidence any descent of market value below the carrying balance sheet value, a write-down pressuring the profitability of the business would be necessary. Conversely, corporates can only record a revaluation gain when the assets are disposed

(Paul Vigna, 2021). A survey from research firm Gartner (2021) in February 2021 found that only 5% of chief financial officers considered holding cryptocurrencies as a corporate asset. Moreover, the regulators have so far adopted a cautious approach in regulating the cryptocurrency markets due to their volatile nature (Karisma, 2022). As of August 2021, the SEC is close to accepting a regulated futures trading venue for cryptocurrencies since future trading comprises stricter rules than spot trading like which those mutual funds are subject to (Michael Wursthorn, 2021).

Therefore, creating a consistently profit-making investment algorithm based on online learning that requires little maintenance and supervision would attract more long-term investors and potentially win the regulatory consensus for creating regulated spot trading venues. Consequently, this would result in improved long-term market liquidity and lower market volatility. The virtuous cycle would complete with the more widespread adoption of digital coins, fueling the positive societal impact hidden beneath the merely speculative aims.

Forecasting Digital Asset return: an Application of Machine Learning Model

Figure 1 shows the scatter plots and densities for each pair of variables in the dataset as well as their univariate distributions along the main diagonal.

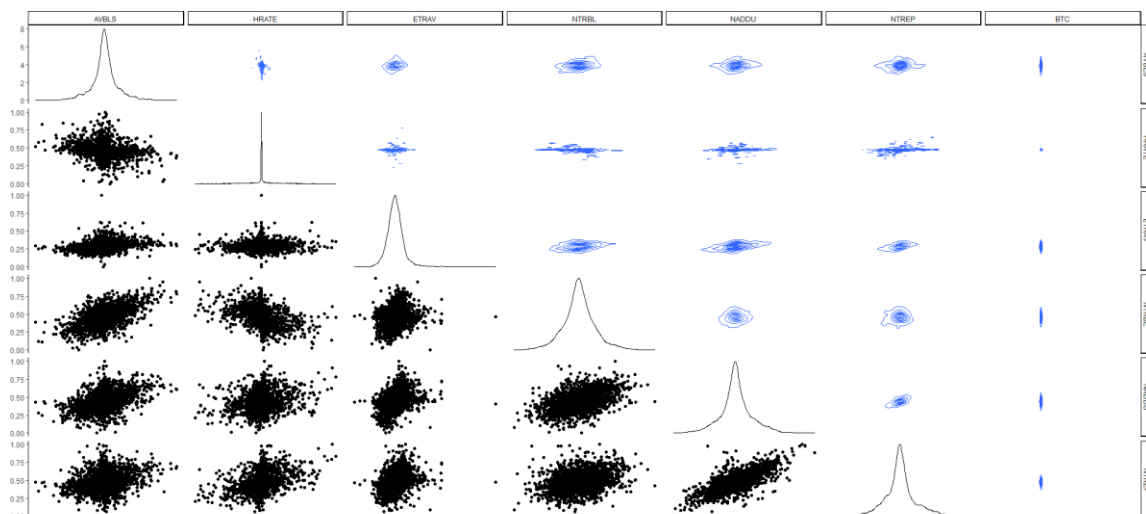


Figure 1 Scatter plots and densities for each pair of variables as well as univariate distributions

Figure 3 Rolling downside volatilities obtained by each strategy

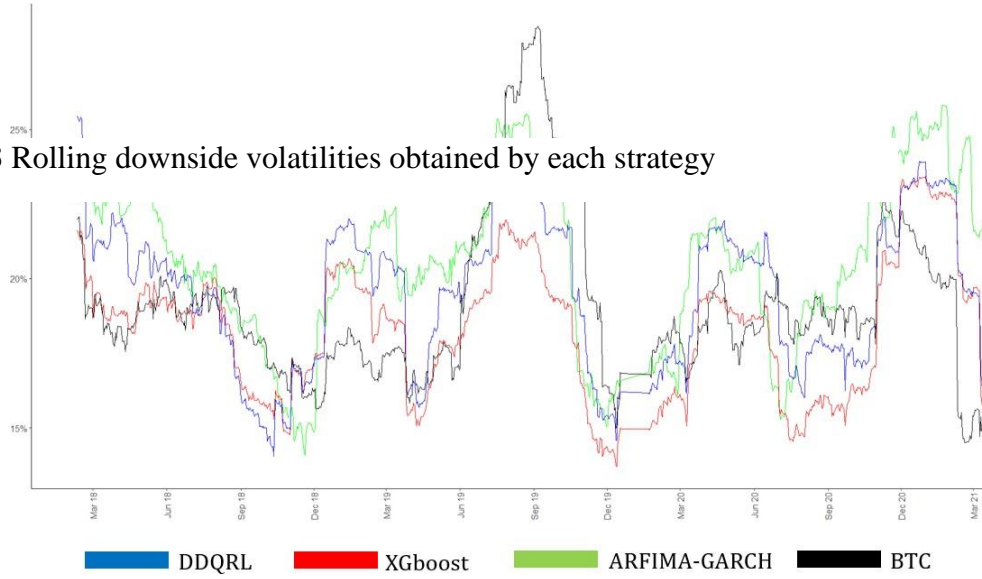
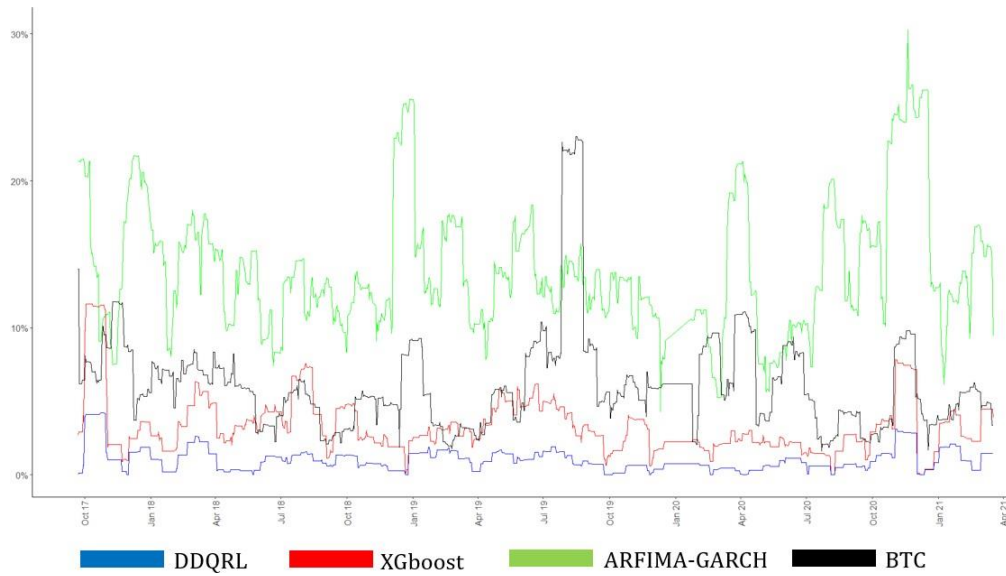


Figure 2 Rolling standard deviations obtained by each strategy



Appendix B Synthetic distributions (obtained via VAE) vs original ones [Figure 4-10]

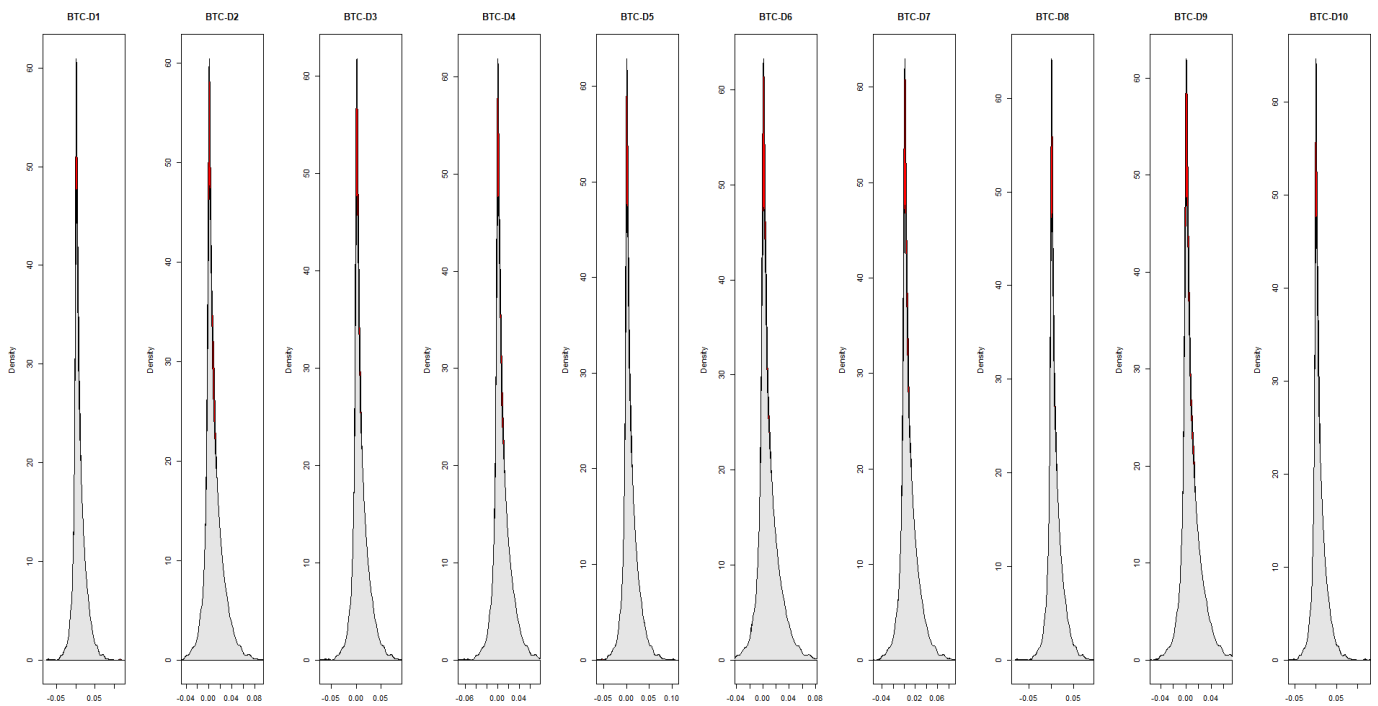


Figure 4 BTC - Bitcoin Price

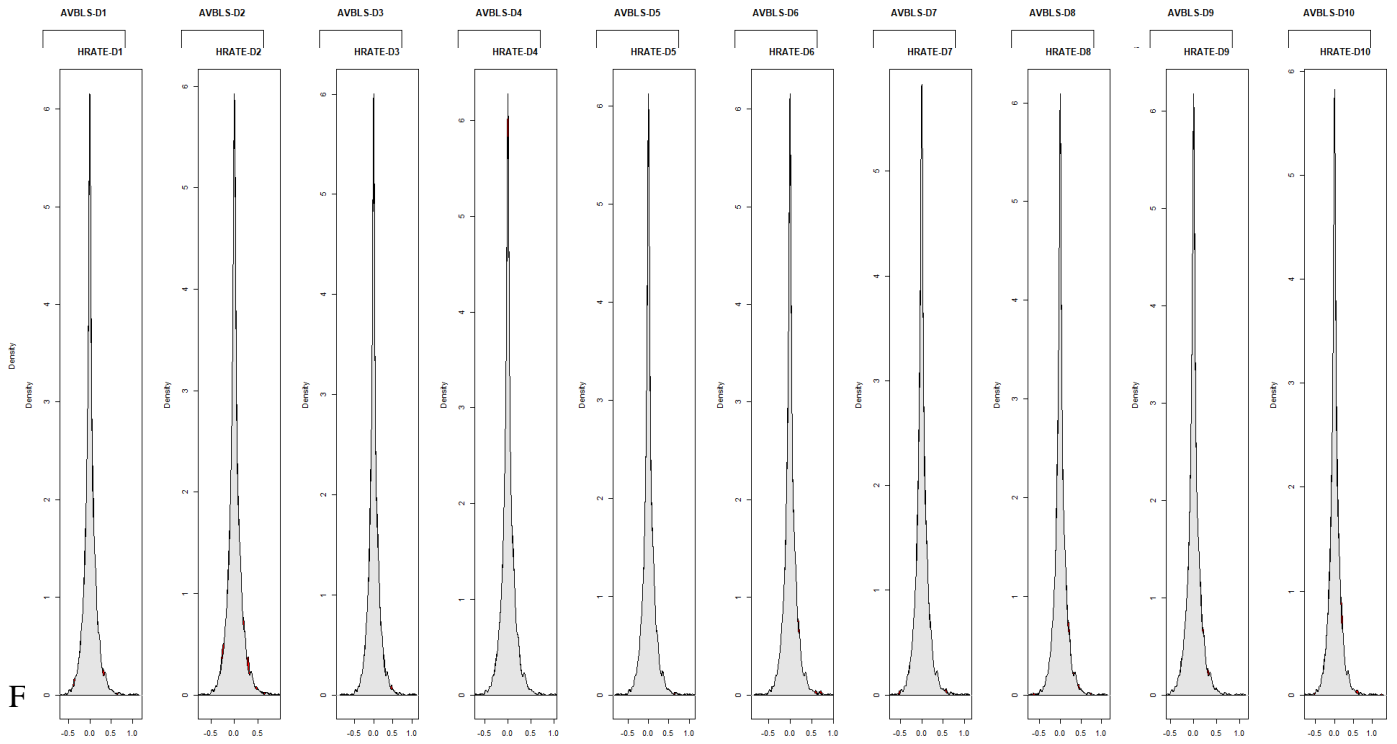


Figure 6 HRATE - Hash Rate

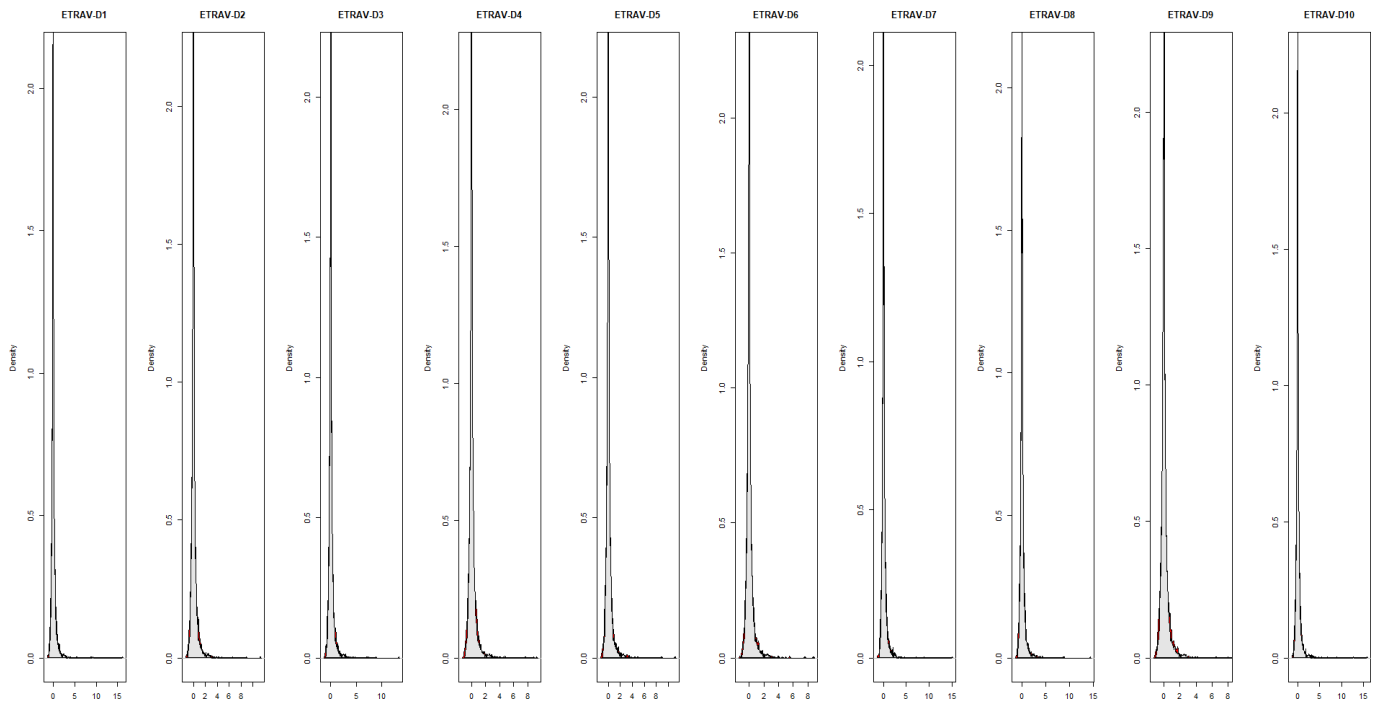


Figure 7 ETRAV - Estimated Transaction Volume

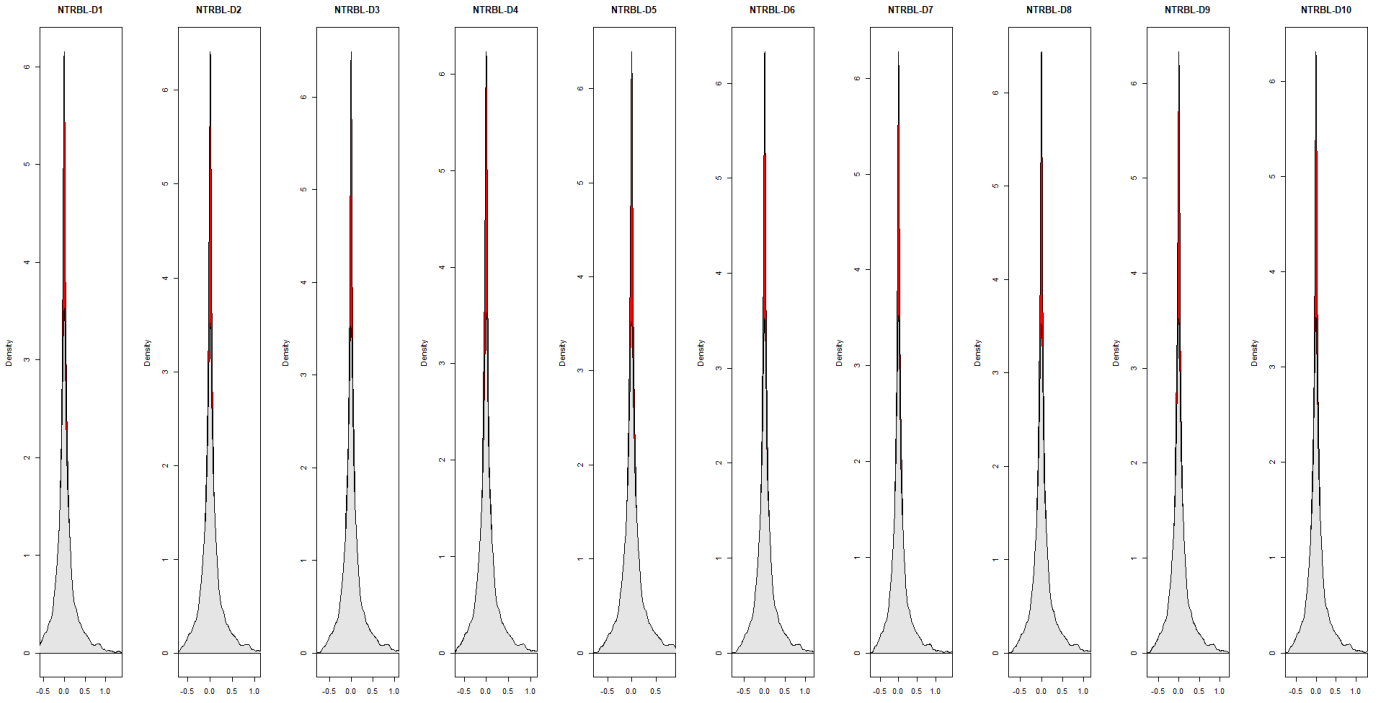


Figure 8 NTRBL - Transaction per block

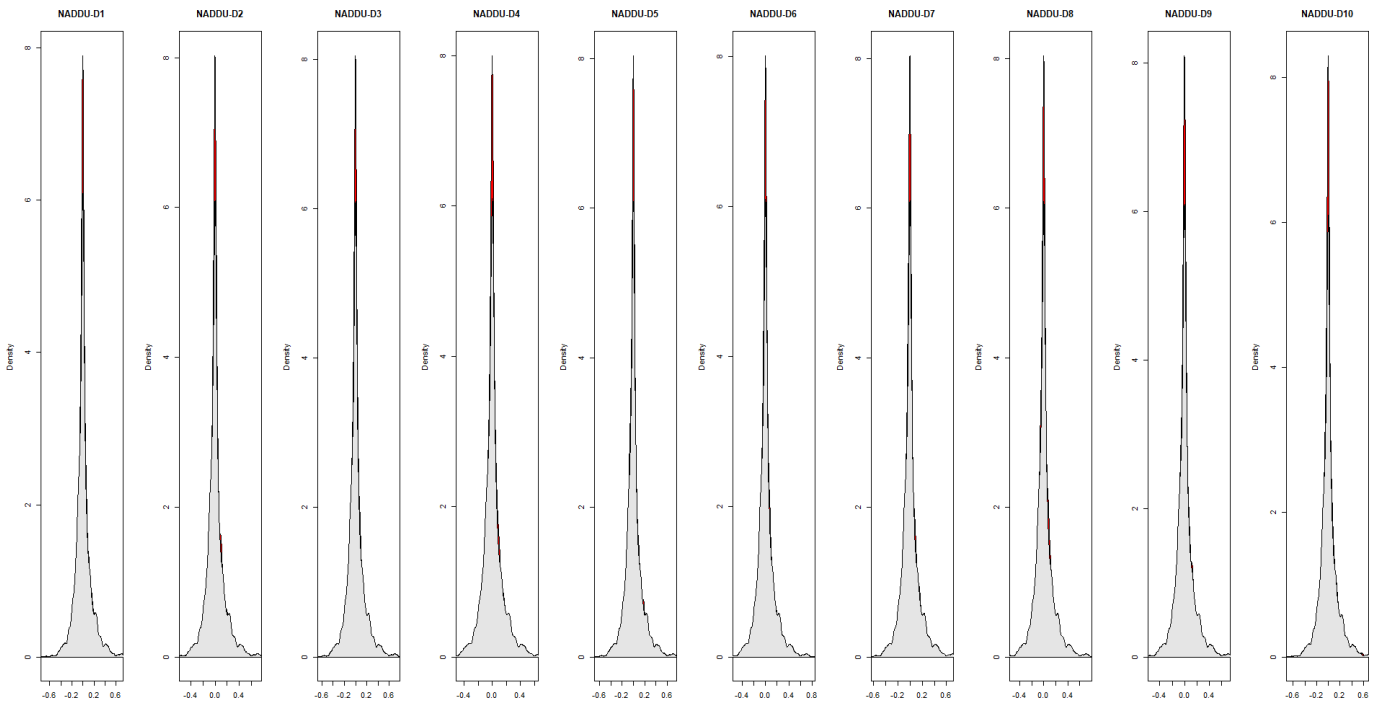


Figure 9 NADDU - Transactions excluding popular addresses

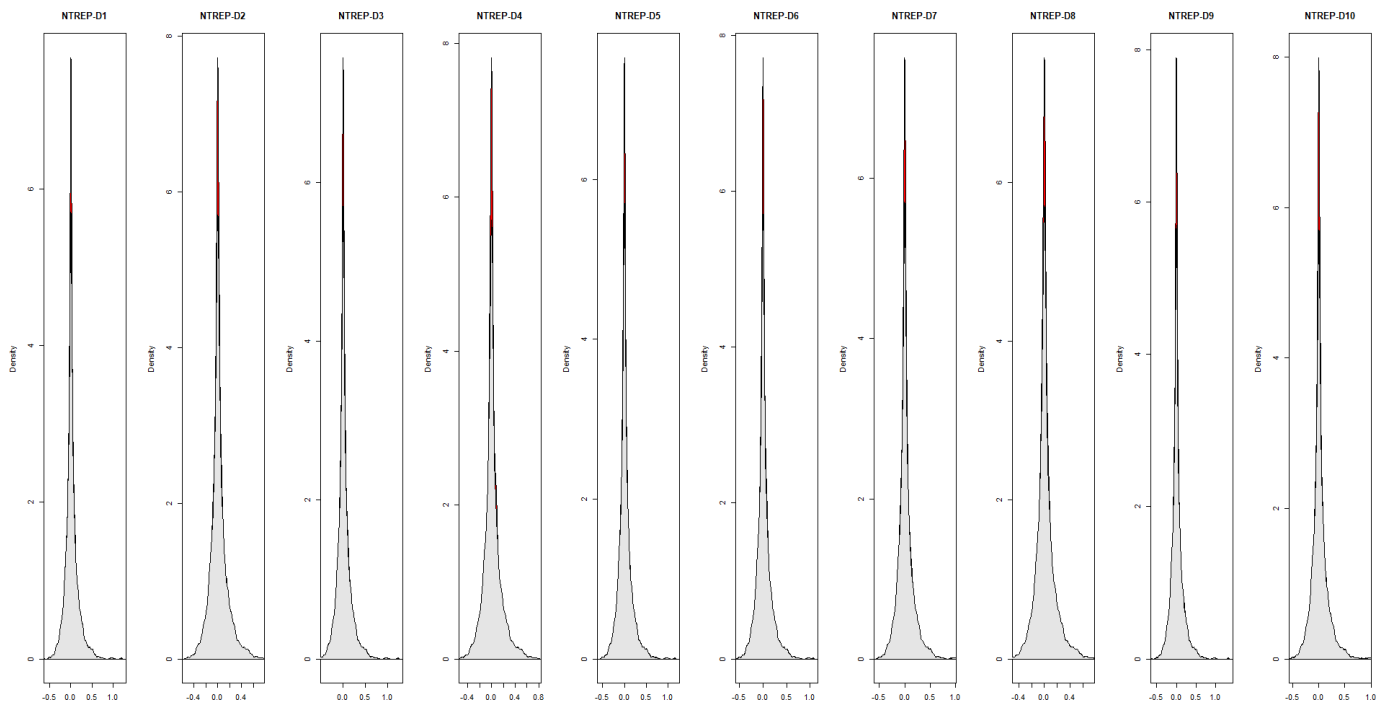


Figure 10 NTREP Number of Transactions