RESEARCH   **Open Access**

Online risk-based portfolio allocation on subsets of crypto assets applying a prototype-based clustering algorithm

Luis Lorenzo*† and Javier Arroyo†

†Luis Lorenzo and Javier Arroyo contributed equally to this work

*Correspondence: luislore@ucm.es

Faculty of Statistic Studies, Complutense University, Avda. Puerta de Hierro, s/n, 28040 Madrid, Spain

Abstract

Mean-variance portfolio optimization models are sensitive to uncertainty in risk-return estimates, which may result in poor out-of-sample performance. In particular, the estimates may suffer when the number of assets considered is high and the length of the return time series is not sufficiently long. This is precisely the case in the cryptocurrency market, where there are hundreds of crypto assets that have been traded for a few years. We propose enhancing the mean-variance (MV) model with a pre-selection stage that uses a prototype-based clustering algorithm to reduce the number of crypto assets considered at each investment period. In the pre-selection stage, we run a prototype-based clustering algorithm where the assets are described by variables representing the profit-risk duality. The prototypes of the clustering partition are automatically examined and the one that best suits our risk-aversion preference is selected. We then run the MV portfolio optimization with the crypto assets of the selected cluster. The proposed approach is tested for a period of 17 months in the whole cryptocurrency market and two selections of the cryptocurrencies with the higher market capitalization (175 and 250 cryptos). We compare the results against three methods applied to the whole market: classic MV, risk parity, and hierarchical risk parity methods. We also compare our results with those from investing in the market index CCI30. The simulation results generally favor our proposal in terms of profit and risk-profit financial indicators. This result reaffirms the convenience of using machine learning methods to guide financial investments in complex and highly-volatile environments such as the cryptocurrency market.

Keywords: Fintech, Mean-variance, Cryptocurrency, Electronic market, Portfolio allocation model, Clustering

Introduction

Blockchain technology is one of the most disruptive technologies in the last 30 years, with applications to many different domains where process transactions take place. As a result, blockchain has changed our view of contracts, logistics, and shipping, and has sparked academic research (Zhou et al. 2021). Finance (Zhao et al. 2016; Xu et al. 2019) is one of the fields where the impact of blockchain has been more important; in

particular, the use of cryptocurrencies as trading assets, which in turn has raised considerable interest from academia (Fang et al. 2022).

Since the appearance of Bitcoin in 2008, the cryptocurrency market capitalization has grown to $1676bn in February 2022, where Bitcoin and Ethereum represent 41.8% and 18.1% of the market, respectively.[1]

While the market is only in its infancy, some studies highlight a Compound Annual Growth Rate (CAGR) higher than 21% for the next 5 years,[2] which makes it extremely attractive for investors. At the same time, its volatility is also extremely high, and the number of crypto assets traded is also very high at an estimated total of around 10,000 as of February 2022.[3] Most of those crypto assets have been traded for just a few years or even less.

While the market is attractive to investors, the aforementioned characteristics render well-known portfolio optimization models, such as the Mean-Variance (MV) methods (Markowitz 1952a, 1959) unreliable. In particular, the high number of potential crypto assets and their short time in the market may hinder the estimation of the covariance matrix.

Some studies use a predefined criterion for reducing the number of considered crypto assets, such as focusing on those with higher capitalization. However, such a strategy may leave aside potentially interesting assets for the investor. Thus, we propose the use of a clustering method to partition the cryptocurrency space and the automatic selection of the partition that best suits the risk-aversion preference of the investor. In particular, we propose the use of a prototype-based clustering algorithm, such as K-Means or K-Medois, as the prototypes will be used as representative elements of the partitions. The usefulness of such methods for characterizing the crypto-asset market in a meaningful way has previously been demonstrated in Lorenzo and Arroyo (2022). Following this work, we use the bivariate representation of the average and standard deviation of the returns together with a partitional clustering algorithm as a preliminary step before the portfolio optimization. After clustering, we select the cluster that best fits one of the predefined risk-aversion profiles using the prototype as an adequate summary of the cluster. By doing so, our method reduces the number of crypto assets that will be considered by the portfolio optimization model, focusing only on those that best suit the investor strategy. Furthermore, our approach acknowledges the changing nature of the cryptocurrency market, as we repeat the process (clustering analysis, selection of the cluster, and portfolio optimization). The resulting number of clusters and their composition may change completely. In this way, our approach can be considered as an online portfolio selection methodology, where decisions are made sequentially incorporating the new information and it goes beyond the static or cross-sectional use of clustering methods in other portfolio selection approaches.

We compare the results of our clustered MV model with those of the standard MV model applied to the whole market. In addition, we also compare it with those from

---

[1] https://coinmarketcap.com/all/views/all/.

[2] https://www.reportlinker.com/p05619614/Crypto-Asset-Management-Market-by-Platform-And-Region-Global-Forecast-to.html?utm_source=GNW.

[3] https://www.statista.com/statistics/863917/number-crypto-coins-tokens/.

more sophisticated methods such as the Risk Parity (Qian 2016), and the Hierarchical Risk Parity (HRP) methods (de Prado 2016). Furthermore, we also compare it against a buy-and-hold strategy of the CCI30 cryptocurrency index that represents the overall market behavior.

The simulation entails a test period of 17 months. We perform three experiments with a different set of cryptocurrencies: the whole cryptocurrency market with data available (over 500 cryptocurrencies), and two selections of the cryptocurrencies with the higher market capitalization (175 and 250 cryptocurrencies, respectively). For each experiment and each method, we repeat the simulation 1500 times using different *investment paths*, each one represents the simulation days that an investor considers for entering the market. If an investment is made on a given day, the position is held for the next 30 days. The different methods are compared using standard profit and risk-profit financial indicators.

The rest of this paper is organized as follows. A literature review is presented in "Literature review" section, in which we support our investigation on well-stated portfolio allocation models and different approaches for clustering of financial markets with more details for the cryptocurrency domain. The "Data and methods" section includes data processing, the methods followed, and the simulation carried out. Finally, we discuss our results and present concluding remarks in the "Results" and "Conclusions" sections, respectively.

Literature review
Portfolio selection
An investment portfolio is a basket of tradable assets and portfolio optimization models are concerned with finding the best combination of assets according to due objectives. There are two main schools of principles and theories for portfolio selection: (i) Markowitz's Mean Variance models and (ii) Capital Growth Theory (Kelly jr 1956; Breiman 1960; Thorp 1975; Finkelstein and Whitley 1981). This research focuses on the first type of model. Modern Portfolio Theory (MPT), also referred to as Mean-Variance Optimization models (MVO), was first posited in the 1950s by (Markowitz 1952a; Sharpe 1964; Lintner 1965) and considers the diversification of assets as the most effective way to obtain low risk-reward ratios maximizing the *expected utility* of the returns. Diversification of capital helps to neutralize idiosyncratic risks. In this way, MVO links with the theory of rational behavior under uncertainty (Markowitz 1952b) and they are included in what are known as risk-based models. The portfolio allocation theory framework has been exponentially developed in very different works (Steinbach 2001; Kolm et al. 2014) to propose solutions to existing constraints on the different models mostly considering practical applicability to the markets. The models now include transaction costs, tax effects, estimation errors on the risk and return forecast, and inter-temporal effects as edging conditions, and the inclusion of specific features required by financial planners. We can find an exhaustive taxonomy of MVO methods in Kalayci et al. (2019), all of which are focused on reducing risk while increasing diversification. One of the weaknesses of MVO models is that it is necessary to provide an estimation of the expected returns and covariances of all the securities in the investment universe; more details on

criticism of MVO can be found in Michaud (1989) and Leland (1999). We use the acronyms MPT, MV, and MVO to refer to the same portfolio allocation model.

Risk Parity Portfolio (RPP) (Qian 2016; Roncalli 2013), also known as Equally weighted Risk Contribution (ERC) portfolio, together with MVO, belong to risk-based models. It is an approach to portfolio management that focuses on risk allocation rather than capital allocation. While the MVO methods minimize the variance, RPP models try to constrain each asset to contribute equally to the portfolio's overall volatility (Maillard et al. 2010) and equalize risk contribution. In other words, it balances the risk so that the risk contribution of every asset is equal and in this sense, it is also considered a risk-based model.

Merging classical portfolio optimization models and hierarchical methods from unsupervised learning techniques detailed in the next subsection, the Hierarchical Risk Parity (HRP) model introduced by de Prado (2016) addresses the problems of traditional risk-based portfolios to compute a portfolio on an ill-generated or even a singular covariance matrix by conducting the optimization process by a top-down recursive bisection using graph theory and machine learning techniques. Based on the same idea, Raffinot (2017) proposed Hierarchical Clustering Based Asset Allocation (HCAA) that allocates capital within and across clusters of assets in multiple hierarchical models. The Hierarchical Equal Risk Contribution Portfolio (HERC) (Raffinot 2018) merges HRP and HCAA. Several variations to this approach have also been proposed (Lohre et al. 2020; Molyboga 2020). In particular, we use the HRP model as a benchmarking method to compare our proposal, as explained below. Moreover, our approach also tackles the problem of the covariance matrix. However, it focuses only on one cluster and in doing so, reduces the number of crypto assets and makes the computation and the inversion process of the covariance matrix easier.

An exhaustive study of the latest risk-based portfolio optimization strategies applied to the 30 highest market capitalization cryptos of the cryptocurrency market can be found in Burggraf (2019). An important strand of research is focused on the effects of adding cryptocurrencies to traditional asset portfolios (Eisl et al. 2015; Chuen et al. 2017; Petukhina et al. 2021; Culjak et al. 2022). Others, as in our proposal, are devoted exclusively to crypto markets. For example, Liu (2019) analyze the invertibility of selected top-10 major cryptocurrencies demonstrating the benefits of the diversification for different portfolio optimization models as ERC, MV, RPP, maximum Sharpe ratio, and maximum utility.

We test the performance of our proposal in cryptocurrency markets using MV, RPP, and HRP models as benchmarks, which are among the most relevant in the portfolio allocation literature.

Clustering techniques in portfolio selection

The application of unsupervised models and in particular clustering techniques to find groups of assets characterized by their financial behavior arises with the seminal paper of Mantegna (Mantegna 1999) applying a hierarchical Minimum Spanning Tree (MST) that takes the linkage between stocks from the New York Stock Exchange market into account. While Mantegna's methodology has been extensively applied, it has some drawbacks. A criticism of the initial Mantegna approach was related to the employed

distance, based on a simple static correlation among the returns' time series. Alternative approaches have been developed considering auto-correlation structure (Piccolo 1990), distances based on GARCH parameters (Otranto 2008), frequency domain features, higher moments of time series, and so on. Different alternatives have been proposed. Onnela et al. (2003) investigated the distribution and dynamics of correlation coefficients. For example, Bonanno et al. (2004) compared the return and volatility networks considering different time horizons. Tumminello et al. (2006) proposed Planar Maximally Filtered Graph (PMFG) graphs instead of MST. Brida and Risso (2009) combined symbolic time series analysis with MST. Musmeci et al. (2014) applied a new linkage method known as a Directed Bubble Hierarchical Tree (DBHT) to financial markets. From our viewpoint, Mantegna's approach is a powerful methodology to determine the structures of the market in the context of a cross-sectional analysis. However, we consider it difficult to adapt it to streaming data and online portfolios due to different constraints for instance with the sampling frequency (Bonanno et al. 2001) (e.g., intra-day, daily, weekly), the length of the rolling window T (Onnela et al. 2003), and the number of assets N under study (Borysov et al. 2014). We later tackle the same challenge when N approaches the T value in portfolio optimization risk-based models. Marti et al. (2017) presents an exhaustive revision of hierarchical clustering in financial markets.

Another important strand of clustering in finance applies partitional prototype-based clustering to financial markets. D'Urso et al. (2013) and D'Urso et al. (2016) used a model-based approach with different variations of *fuzzy* clusters and different distance metrics (autoregressive, Caiado). Iorio et al. (2018) proposed a clustering based on the computation of the spline coefficients of the time series and directly measured the performance within MV, Equally Weighted (EW), and ERC portfolio allocation models. Similarly, D'Urso et al. (2020) proposed a fuzzy clustering method based on cepstral representation, using the daily Sharpe ratio as a variable of clustering. Soleymani and Vasighi (2020) adapted a K-means to cluster NYSE stocks based on Value-at-Risk (VaR) and Conditioned Value-at-Risk (CVar) measures. Gubu et al. (2020) presents a robust portfolio selection using the KAMILA algorithm on a combination of continuous and categorical variables with a robust covariance estimation. Cerqueti et al. (2021) propose a clustering time series according to their estimated conditional moments via Autocorrelation-based fuzzy C-means (A-FCM algorithm); the proposal is enhanced in Cerqueti et al. (2022), in which they computed an optimal weight for each moment. Both proposals are tested directly on different time series as empirical experiments.

Regarding the partitional clustering family, Nanda et al. (2010) applied K-means, fuzzy C-means, and Self-Organizing Maps (SOM) to returns and financial ratios from Indian stocks to classify them into different clusters and subsequently develop portfolios. The analysis considers a set of stocks with fixed-weight allocation along the investing period. However, the approach does not use out-of-sample data and the study is not replicated over time to investigate how the clusters and the results evolve. Nguyen Cong et al. (2014) proposed another precedent, which combines a stage of clustering using return and standard deviation variables and a multi-objective portfolio optimization allocating stocks from the different clusters using a genetic algorithm. The simulation is carried out using 570 stocks from the Stock Exchange of Thailand (SET) and identifies four clusters. Again, the number of clusters does not change over time. Datta and Ghosh (2015)

propose an approach that groups the daily Indian market volatility by comparing Kernel K-means, SOM, and Gaussian clustering models to achieve the right volatility prediction using the clusters as predictors. Luca and Zuccolotto (2017) propose a dynamic clustering procedure for time-series returns using the time-varying tail dependency as a dissimilarity measure. The aim is to provide a criterion for portfolio selection focusing on the lower tails of the returns distributions that are sensitive to the contagion phenomena between stocks for the FTSE MIB index. Duarte and De Castro (2020) segment the assets of the Brazilian Stock Exchange (B3) into partitional clusters of correlated assets taken as initial medoids of those assets with the lowest standard deviation of the past series of prices that feed an MV model and compared the performance with the RPP model. Instead, our approach is not based on any correlation measure but on a Euclidean distance defined on the volatility-return space.

Clustering techniques applied to the cryptocurrency market

Regarding the application of clustering methods to the cryptocurrency market, Song et al. (2012) applied Mantegna's initial ideas based on hierarchical clustering but renewed for the emerging crypto market. Similarly, Stosic et al. (2018) use clustering to characterize the cryptocurrency market using the correlations of 110 cryptocurrencies and detect hierarchical structures using the MST. Song et al. (2019) also applied MST but removed the influence of Bitcoin-Ethereum to avoid a highly-correlated matrix. Lorenzo and Arroyo (2022) applied three different prototype-based clustering techniques to conduct a cross-sectional analysis of the cryptocurrency market and identify associations between the clusters and several financial and technological descriptors. Each clustering method deals with the cryptocurrencies represented in a different way. Namely, a representation as two variables of the average and standard deviation of the daily returns, the distribution of daily returns, and the daily return time series.

Our approach uses clustering, but contrary to other works, it uses a sliding window, allowing for the number of clusters and their composition to change over time and automatically deciding the number of clusters by combining several validity indexes using a voting mechanism.

Online portfolio selection

We adopt an approach that fits into the Online Portfolio Selection (OLPS) models. The main characteristic of such portfolios is that it sequentially select a portfolio over a set of assets to achieve certain targets. In OLPS, market information arrives sequentially and the allocation decision must be made immediately. An exhaustive survey can be found in Li and Hoi (2014), which is complemented in Li et al. (2016) with an open-source MAT-LAB library to apply different online algorithms.

There are two types of methodologies in the OLPS literature: (i) *Batch learning* where the model is trained from a batch of training instances and (ii) *Online learning* where the model is successively trained from a single instance taking the price change ($x_{t,i} = \frac{p_{t,i}}{p_{t-1,i}}$) as an input vector. Our research is focused on the continuous-time MV model developed for multiple period (batch) portfolio selection for the control part and it is analytically resolved in Li and Ng (2000) and Dai et al. (2010). Our proposal suits the batch approach because it is based on the deterministic management of historical data for the

portfolio selection where there is not any dependency on the allocation decisions between different time frames. In addition, the target into each iteration for every investing window is a mean reversion formula inspired by the Online Mean Average Reversion (OLMAR) methods (Li and Hoi 2012; Li et al. 2015; Umino et al. 2022).

Jiang and Liang (2017) proposes an online portfolio approach in cryptocurrency markets. In particular, they propose a deterministic deep reinforcement learning based on a Convolutional Neural Network (CNN) applied to a training window of the historic price changes. The weight on the allocation is based on a reward function that maximizes the portfolio value although only for 12 cryptocurrencies. There are some similarities between this work and our approach since both use parameter tuning based on back-testing trading, and both combine machine learning and portfolio allocation. However, we first apply a clustering technique instead of the more complex CNN. Second, we use the classical portfolio allocation model MV instead of a reward function that does not consider any aspect of risk-aversion on the investor. Third, we apply our method to 534 cryptos instead of just 12.

Another relevant reference is Khedmati and Azin (2020), who presents an online selection algorithm based on the pattern matching principle where it uses K-means, k-medoids, spectral, and hierarchical clustering for the selection of the best investing time window.

Market efficiency

Market efficiency is a key financial subject that the researchers try to transpose from the classical to the cryptocurrency domain. Starting with the seminal works of Fama (1965) and Samuelson (1965) on traditional financial markets, there have been different attempts to understand the applicability of efficiency to the new markets. Kyriazis (2019) conducts a systematic survey on the predictability of cryptocurrency prices and concludes that the Efficient Market Hypothesis (EMH) is rejected, opening a door to speculation, a conclusion that we partially confirm in our investigation. Makarov and Schoar (2020) comes to the same conclusion but analyzes arbitrage opportunities for the price deviation across the different exchanges. One of the major implications of the inefficiencies of the cryptocurrency markets is that they offer investment opportunities to portfolio management of making excess returns based on out-performing the market (Palamalai et al. 2021). We find different examples of how we can take advantage of such inefficiencies by applying different machine learning techniques. For instance, Alessandretti et al. (2018) applied different forecasting models achieving profits over the investing period and performing better than a baseline strategy. The parameter optimization based on the Sharpe ratio achieves the best results, which is one of the strategies that we analyze herein. Livieris et al. (2020) ensemble different learning strategies that exhibit high efficiency and reliability mainly for low-frequency applications. Fang et al. (2021) analyze a data-driven approach with a retraining method to predict successful mid-price movements in cryptocurrency markets. The disadvantage of the learning algorithms that take advantage of the market inefficiency is that the methods are data-hungry (Marcus 2018) and the forecasting benefit decay in non-stationary time series. Finally, Sebastião and Godinho (2021) analyze the predictability of three important cryptocurrencies, Bitcoin, Ethereum, and Litecoin, using several machine learning methods and compare

their profitability incorporating trading costs. The results indicate that it is possible to propose profitable trading strategies in the cryptocurrency market, even under adverse market conditions, providing an example of the market inefficiencies.

Data and methods

Dataset and preprocessing

From Cryptocompare exchange,[4] we retrieve the daily closing price for all the cryptocurrencies traded from January 1, 2018, to May 31, 2021, for a total of 1,999,953 market observations along 1247 trading days.

For each cryptocurrency, we transform the price time series into the arithmetic return time series, whose use is extended and consolidated due to its more suitable statistical properties and better comparability (Gilli et al. 2019). The arithmetic returns for the cryptocurrency i at time t are:

$$R_{t,i} = \frac{P_{t,i} - P_{t-1,i}}{P_{t-1,i}} = \frac{P_t}{P_{t-1}} - 1, \; t = 1, ...T, \; i = 1, ..., m \tag{1}$$

where $P_i(t)$ is the daily cryptocurrency price for crypto-asset i at day t and T is the time series sampling.

Regarding data cleaning, we remove the observations with duplicated rows and NaN or Inf values for $R_{t,i}$.

Furthermore, we filter out the cryptocurrencies with heavier tails in the return distribution because it implies relatively frequent extreme price fluctuations and affects the consistency of the results, particularly the estimation of the returns and covariance matrix for the portfolio optimization.

Heavy-tail behavior in a return distribution is related to the finite-size effects in the number of active agents linked to the liquidity and volume of the market (Watorek et al. 2020). According to Newman (2005), a distribution has a heavy-tail behavior if the tail index is lower than 2. In our case, we discard cryptocurrencies with heavier tails, that is, those with a tail index higher than 2.3. In this way, we ensure the existence of the two-moment expectation and covariance matrix required for risk-based models. We apply this filter only to the first two-years of data, that is, before the simulation starts. In this way, we avoid the look-ahead bias. The results are reported in Table 1.

Stationarity is another important property of the underlying process in a return time series, especially if we are interested in forecasting or pricing. Random walk theory allows us to test the weak form of the EMH (Samuelson 1965; Fama 1965) within a series of asset returns. The key principle of EMH is that asset prices reflect all information, making it impossible for investors to derive benefits through trying to predict their behavior. The weak form of efficiency can be expressed as an autoregressive random walk model of stock returns:

$$R_t = \rho R_{t-1} + e_t, \; t = 1, 2, ..., T \; and \; e_t \sim N(0, \sigma^2) \tag{2}$$

---

[4] https://www.cryptocompare.com/.

Table 1 Heavy Tail and KSS inference tests of the time-series from 2018–19 of the selected 534 cryptoassets ordered by market cap; AlphaN and AlphaP are the scaling parameters (α) for left and right tails of the distribution; Sd.N and Sd.P are the standard deviation for α; KssStat and KssPvalue are the Kss non-stationarity statistic and the corresponding P-value

MkCapRk	SYM	AlphaP	Sd.P	AlphaN	Sd.N	kssStat	kssPvalue	MkCapRk	SYM	AlphaP	Sd.P	AlphaN	Sd.N	kssStat	kssPvalue	MkCapRk	SYM	AlphaP	Sd.P	AlphaN	Sd.N	kssStat	kssPvalue
1	BTC	2.944	0.101	3.128	0.112	−3.240	0.001	180	IDEX	3.643	0.262	4.206	0.300	−2.928	0.004	359	USDS	4.903	0.329	3.266	0.184	1.626	0.105
2	ETH	3.247	0.118	3.547	0.133	−2.474	0.014	181	UPP	2.326	0.087	4.647	0.207	−0.459	0.646	360	BCPT	3.030	0.110	3.799	0.142	−2.342	0.019
3	USDT	3.695	0.142	3.587	0.134	−1.941	0.053	182	RDNN	4.958	0.206	4.588	0.189	−2.969	0.003	361	CSM	3.080	0.182	3.088	0.181	−1.546	0.124
4	BNB	3.694	0.145	4.073	0.157	−3.020	0.003	183	NXT	5.617	0.243	3.377	0.124	−3.754	0.000	362	ZLA	2.458	0.095	4.481	0.201	−3.964	0.000
5	ADA	4.099	0.165	4.011	0.155	−3.560	0.000	184	COCOS	3.945	0.417	7.351	0.776	−0.757	0.451	363	HYDRO	3.003	0.131	3.625	0.141	−1.229	0.219
6	DOGE	2.431	0.084	2.316	0.063	−4.074	0.000	185	NXS	2.488	0.080	3.912	0.149	−0.775	0.439	364	DNA	2.341	0.085	2.796	0.082	−2.649	0.008
7	LINK	3.469	0.135	5.125	0.207	−3.654	0.000	186	BEAM	3.050	0.179	4.608	0.283	−3.170	0.002	365	QQQ	5.194	0.571	2.369	0.156	−3.636	0.000
8	USDC	3.804	0.185	3.433	0.165	−3.265	0.001	187	TOPN	2.589	0.116	2.973	0.085	0.497	0.619	366	XUC	3.153	0.114	3.964	0.153	−1.168	0.243
9	XLM	2.899	0.102	3.331	0.120	−3.980	0.000	188	MRPH	3.756	0.179	3.680	0.164	−3.046	0.002	367	SUR	3.345	0.129	4.312	0.165	−2.088	0.037
10	BCH	3.213	0.121	4.495	0.176	−2.957	0.003	189	BHP	3.185	0.165	2.491	0.113	−0.846	0.398	368	HYN	3.430	0.245	3.741	0.280	−0.955	0.336
11	LTC	4.072	0.167	4.491	0.177	−0.578	0.563	190	ZAP	2.886	0.110	4.699	0.177	−4.172	0.000	369	DVP	2.405	0.189	4.257	0.417	−1.278	0.205
12	THETA	3.082	0.114	3.683	0.140	−3.181	0.002	191	SNTVT	3.027	0.271	3.566	0.323	−0.314	0.754	370	PLAC	2.442	0.268	3.059	0.322	−1.973	0.054
13	TRX	3.194	0.120	3.356	0.118	−2.806	0.005	192	VTC	3.459	0.128	6.244	0.277	−5.190	0.000	371	BCDN	2.981	0.252	3.809	0.110	−4.521	0.000
14	ETC	5.883	0.257	5.916	0.256	−3.108	0.002	193	TONE	3.601	0.160	3.802	0.165	−1.388	0.166	372	SEND	3.391	0.136	5.005	0.195	−2.629	0.009
15	ALGO	3.192	0.217	3.442	0.256	−2.398	0.018	194	QSP	3.618	0.135	4.771	0.201	−1.397	0.163	373	SERO	2.515	0.196	2.507	0.195	−2.345	0.021
16	EOS	2.734	0.093	3.163	0.111	−2.952	0.003	195	GRIN	2.917	0.156	2.703	0.121	0.772	0.441	374	ART	2.921	0.114	3.560	0.121	−0.861	0.390
17	CEL	4.444	0.217	4.886	0.233	−0.389	0.697	196	NCASH	3.629	0.144	3.152	0.117	−2.554	0.011	375	SHIP	3.633	0.157	3.422	0.135	−1.538	0.124
18	NEO	3.239	0.118	3.428	0.126	−3.306	0.001	197	MTV	3.793	0.273	3.644	0.262	−1.176	0.241	376	ZCL	2.797	0.098	3.510	0.127	2.519	0.012
19	XMR	3.074	0.108	5.574	0.240	−1.781	0.075	198	BAX	3.074	0.130	3.974	0.167	2.974	0.003	377	ASAFE2	3.245	0.130	3.592	0.125	−2.399	0.017
20	OKB	3.993	0.177	3.262	0.128	−2.766	0.006	199	BURST	4.317	0.175	3.567	0.133	−6.237	0.000	378	ATH	2.308	0.112	3.108	0.121	0.022	0.982
21	BTT	3.054	0.170	2.610	0.118	−3.180	0.002	200	PPC	3.311	0.119	5.483	0.238	−0.762	0.447	379	RYO	3.063	0.127	3.907	0.178	0.259	0.796
22	RSR	2.743	0.177	3.547	0.239	−1.434	0.153	201	FOR	2.909	0.276	5.318	0.524	−1.842	0.069	380	ATL	2.692	0.101	3.827	0.133	−2.706	0.007
23	MKR	2.944	0.106	3.084	0.105	−7.634	0.000	202	YOYOW	3.024	0.104	4.338	0.179	−4.541	0.000	381	ZER	2.607	0.085	3.500	0.130	−6.451	0.000
24	MIOTA	4.290	0.172	3.354	0.124	−4.783	0.000	203	SKY	3.919	0.154	4.659	0.190	−1.800	0.072	382	ROX	3.602	0.168	3.198	0.134	−1.972	0.049
25	XTZ	3.496	0.135	3.833	0.144	−3.418	0.001	204	CND	3.223	0.120	3.988	0.152	3.935	0.000	383	SCASH	3.093	0.109	3.640	0.139	−2.400	0.017
26	TEL	2.818	0.100	3.155	0.112	−2.238	0.026	205	GO	2.939	0.121	4.171	0.185	−1.595	0.111	384	SAFE	3.449	0.188	4.855	0.266	−2.442	0.015

Table 1 (continued)

MkCapRk	SYM	AlphaP	Sd.P	AlphaN	Sd.N	kssStat	kssPvalue
27	LEO	4.983	1.626	38.853	2.606	0.049	0.961
28	SNX	3.610	0.153	3.978	0.164	−3.764	0.000
29	CTC	2.926	0.100	3.126	0.113	−3.234	0.001
30	NEXO	3.803	0.166	4.192	0.183	1.011	0.312
31	HBAR	2.606	0.213	3.347	0.350	2.349	0.021
32	GNO	4.614	0.188	3.156	0.114	−2.926	0.004
33	DASH	3.521	0.133	3.642	0.138	−0.978	0.329
34	DCR	3.034	0.106	3.804	0.147	−1.529	0.127
35	ZIL	3.703	0.146	3.786	0.147	−4.250	0.000
36	MANA	3.019	0.107	4.247	0.168	−6.796	0.000
37	XEM	3.240	0.118	3.324	0.121	−2.964	0.003
38	WAVES	3.367	0.127	3.234	0.114	−4.916	0.000
39	QNT	2.925	0.160	2.785	0.132	−1.575	0.116
40	KCS	3.178	0.119	4.111	0.157	−3.164	0.002
41	ENJ	2.488	0.080	2.525	0.078	0.191	0.849
42	PAX	3.141	0.145	3.920	0.193	−3.924	0.000
43	XWC	2.479	0.089	3.337	0.110	1.938	0.053
44	BAT	4.309	0.175	4.302	0.171	−4.007	0.000
45	BTG	3.294	0.120	3.667	0.140	−3.282	0.001
46	ONT	3.003	0.115	3.968	0.158	−4.501	0.000
47	QTUM	3.192	0.116	5.551	0.235	−1.915	0.056
48	REV	3.071	0.113	3.694	0.136	−2.242	0.025
49	ANKR	3.088	0.172	4.032	0.248	−2.786	0.006
50	ZEN	6.503	0.293	3.975	0.153	−3.610	0.000
51	NANO	3.027	0.109	3.875	0.147	−4.219	0.000
52	ZRX	6.241	0.288	5.819	0.241	−4.024	0.000
53	ONGAS	2.467	0.100	2.743	0.113	2.232	0.026
54	ICX	3.718	0.141	4.508	0.185	−3.090	0.002

MkCapRk	SYM	AlphaP	Sd.P	AlphaN	Sd.N	kssStat	kssPvalue
206	KAN	2.540	0.090	2.310	0.092	−1.965	0.050
207	PHB	3.682	0.264	3.794	0.265	−2.477	0.014
208	GAME	2.814	0.101	3.660	0.132	1.223	0.222
209	SALT	3.113	0.112	2.512	0.078	1.052	0.293
210	CSP	3.340	0.196	3.800	0.222	−0.394	0.694
211	TCT	2.306	0.073	2.637	0.084	−3.736	0.000
212	GBYTE	6.394	0.285	3.795	0.145	−0.280	0.779
213	SOC	2.770	0.104	3.841	0.146	−4.578	0.000
214	ZEL	3.511	0.162	3.867	0.148	0.040	0.968
215	NEBL	2.602	0.087	5.773	0.240	−3.520	0.000
216	ABYSS	3.138	0.133	3.763	0.156	−2.330	0.020
217	WABI	3.706	0.147	5.468	0.225	−1.408	0.160
218	AMB	3.305	0.121	4.096	0.161	−4.336	0.000
219	NAV	3.166	0.115	4.095	0.160	−4.681	0.000
220	SWTH	3.302	0.138	3.564	0.147	0.666	0.506
221	VIDT	2.656	0.156	3.025	0.151	−2.678	0.008
222	SENT	3.029	0.115	4.147	0.157	−4.030	0.000
223	MITX	3.280	0.150	5.212	0.244	−2.912	0.004
224	ABT	7.265	0.359	5.882	0.256	0.659	0.510
225	PMA	2.536	0.097	2.560	0.093	3.743	0.000
226	SWFTC	2.766	0.096	2.665	0.084	−0.347	0.729
227	QLC	3.504	0.134	2.443	0.074	−4.026	0.000
228	WTC	3.465	0.130	3.417	0.126	−0.020	0.984
229	MDS	2.773	0.095	6.630	0.294	−4.824	0.000
230	AEON	2.627	0.087	2.379	0.071	−6.430	0.000
231	PRE	2.560	0.102	2.404	0.063	−0.393	0.694
232	GOLD	3.728	0.147	4.177	0.161	−0.232	0.816
233	APPC	3.582	0.137	4.047	0.160	−3.072	0.002

MkCapRk	SYM	AlphaP	Sd.P	AlphaN	Sd.N	kssStat	kssPvalue
385	404	2.382	0.078	2.797	0.088	−4.828	0.000
386	PBT	2.815	0.099	5.837	0.243	−2.904	0.004
387	OXEN	2.973	0.126	5.927	0.282	1.533	0.126
388	PINK	3.373	0.127	3.211	0.113	−0.719	0.472
389	ZRC	3.025	0.113	4.242	0.161	0.328	0.743
390	PEPE	2.976	0.104	4.637	0.189	−4.390	0.000
391	NUT	3.346	0.287	2.462	0.172	0.402	0.689
392	BCNA	2.688	0.234	4.075	0.373	−1.022	0.309
393	ORE	2.505	0.098	3.889	0.143	−5.495	0.000
394	ORLY	3.095	0.109	3.639	0.139	−2.389	0.017
395	ORME	2.332	0.080	3.097	0.098	−2.109	0.035
396	BCA	2.985	0.101	3.416	0.133	−0.295	0.768
397	OK	3.226	0.121	4.061	0.154	−4.386	0.000
398	OMNI	3.374	0.125	4.298	0.171	−0.827	0.408
399	QBAO	2.988	0.109	2.487	0.078	−0.745	0.457
400	PTOY	2.326	0.074	3.328	0.116	−6.127	0.000
401	AXPR	3.951	0.188	4.466	0.220	2.270	0.024
402	RBY	3.205	0.115	3.239	0.118	−2.464	0.014
403	RAISE	2.306	0.099	3.598	0.136	−4.984	0.000
404	888	2.930	0.113	3.062	0.099	2.323	0.020
405	BANCA	3.030	0.118	5.618	0.244	−3.343	0.001
406	POE	3.121	0.110	3.710	0.143	−4.509	0.000
407	PLA	4.101	0.244	3.528	0.210	−2.339	0.020
408	PITCH	2.705	0.130	2.949	0.120	−0.326	0.745
409	PLU	3.934	0.165	4.883	0.191	−3.175	0.002
410	PLR	3.055	0.113	5.644	0.232	−2.328	0.020
411	VRC	3.003	0.108	4.565	0.182	−6.616	0.000
412	VITAE	2.758	0.119	4.026	0.172	−1.324	0.186

Table 1 (continued)

MkCapRk	SYM	AlphaP	Sd.P	AlphaN	Sd.N	kssStat	kssPvalue	MkCapRk	SYM	AlphaP	Sd.P	AlphaN	Sd.N	kssStat	kssPvalue	MkCapRk	SYM	AlphaP	Sd.P	AlphaN	Sd.N	kssStat	kssPvalue
55	DGB	2.920	0.104	2.555	0.079	− 3.790	0.000	234	YEE	3.350	0.128	2.550	0.080	− 3.980	0.000	413	ACTN	3.568	0.239	5.324	0.351	− 3.154	0.002
56	BNT	5.577	0.236	3.659	0.141	− 1.566	0.118	235	TRUE	2.345	0.082	5.493	0.247	− 2.775	0.006	414	VBK	2.488	0.145	3.699	0.266	− 2.624	0.009
57	SAND	2.954	0.107	4.492	0.175	− 4.392	0.000	236	LTO	3.398	0.223	6.334	0.457	− 0.255	0.799	415	VEE	3.727	0.147	4.944	0.203	1.320	0.187
58	SC	3.045	0.107	3.635	0.137	− 5.994	0.000	237	SLS	3.552	0.136	3.975	0.153	− 0.087	0.930	416	VAPOR	3.091	0.109	3.632	0.138	− 2.392	0.017
59	OMG	3.751	0.146	3.373	0.123	− 3.772	0.000	238	CDT	5.148	0.217	3.969	0.155	− 3.260	0.001	417	USE	2.939	0.145	2.708	0.121	− 1.475	0.141
60	FTM	3.201	0.164	4.849	0.274	− 1.596	0.111	239	FCT	3.118	0.111	3.124	0.111	− 4.328	0.000	418	XMY	3.636	0.141	3.533	0.130	− 4.195	0.000
61	CHSB	4.747	0.208	4.958	0.208	− 1.564	0.118	240	GVT	3.261	0.120	3.542	0.131	− 4.286	0.000	419	XST	3.109	0.115	3.981	0.150	− 3.299	0.001
62	IOST	3.756	0.149	2.494	0.077	− 5.796	0.000	241	OAX	3.200	0.116	3.767	0.143	− 4.076	0.000	420	XSH	4.481	0.191	3.802	0.151	− 3.622	0.000
63	RVN	3.341	0.143	3.554	0.136	− 4.515	0.000	242	MDA	2.545	0.087	2.523	0.075	− 2.141	0.033	421	XRL	3.599	0.145	2.935	0.096	− 2.127	0.034
64	LPT	2.672	0.126	4.810	0.280	2.556	0.011	243	HPB	3.086	0.112	6.191	0.282	− 4.358	0.000	422	32BIT	14.838	0.713	3.414	0.128	− 2.803	0.005
65	SNT	3.462	0.131	4.005	0.154	− 2.010	0.045	244	ETP	3.286	0.128	3.357	0.117	− 1.912	0.056	423	WMC	3.103	0.110	3.614	0.137	− 2.401	0.017
66	THEX	2.476	0.250	2.489	0.233	− 1.841	0.070	245	QRL	2.726	0.096	2.994	0.099	− 4.878	0.000	424	XAUR	4.195	0.169	3.734	0.141	− 0.358	0.721
67	KICK	3.150	0.118	4.536	0.177	− 2.539	0.011	246	VIA	3.895	0.152	3.853	0.149	− 6.023	0.000	425	XBC	3.003	0.105	3.895	0.151	− 1.237	0.216
68	NMR	3.209	0.120	4.427	0.173	− 6.004	0.000	247	GNX	3.351	0.127	2.464	0.075	− 1.501	0.134	426	XBI	2.553	0.114	3.705	0.141	− 6.295	0.000
69	REN	3.483	0.143	3.562	0.131	− 6.226	0.000	248	POLYS	2.877	0.111	3.270	0.127	− 3.625	0.000	427	SUB	3.339	0.126	3.939	0.150	− 3.318	0.001
70	XVG	3.063	0.109	3.849	0.147	− 4.798	0.000	249	OST	3.375	0.125	2.308	0.069	− 3.866	0.000	428	SWT	2.460	0.083	5.375	0.213	− 1.407	0.160
71	LSK	3.507	0.132	3.540	0.132	− 2.415	0.016	250	UGAS	2.457	0.142	3.158	0.213	− 1.599	0.111	429	ANAL	2.817	0.098	2.459	0.074	− 2.225	0.026
72	MCO	3.476	0.134	5.707	0.239	− 3.457	0.001	251	RUFF	3.325	0.133	3.399	0.122	− 3.757	0.000	430	SOP	3.321	0.236	3.766	0.298	− 1.113	0.267
73	LRC	2.969	0.103	4.853	0.201	− 4.955	0.000	252	ENG	3.288	0.121	6.358	0.277	− 4.531	0.000	431	SPOTS	6.398	0.279	3.133	0.113	− 3.237	0.001
74	STORJ	2.501	0.078	3.942	0.156	− 5.216	0.000	253	KRL	3.069	0.128	3.212	0.135	− 3.766	0.000	432	SSP	2.567	0.125	3.573	0.184	− 2.043	0.042
75	MVL	2.493	0.108	3.947	0.205	1.895	0.059	254	SMT	3.136	0.116	3.562	0.132	− 1.925	0.055	433	ZAIF	10.881	1.397	5.742	0.602	0.276	0.783
76	CELR	2.755	0.152	3.054	0.170	− 3.088	0.002	255	XDN	2.947	0.105	4.116	0.159	− 5.162	0.000	434	SPHR	2.421	0.079	4.563	0.177	− 0.783	0.434
77	OCEAN	2.310	0.077	4.536	0.168	− 0.241	0.809	256	CRDTS	2.943	0.109	4.418	0.184	− 2.013	0.045	435	SPK	2.975	0.128	4.433	0.164	1.749	0.081
78	CET	3.108	0.133	4.900	0.231	− 0.421	0.674	257	SAN	3.143	0.112	3.233	0.117	− 3.236	0.001	436	AMN	3.041	0.139	4.455	0.180	− 2.715	0.007
79	RLC	3.626	0.140	4.318	0.171	− 1.698	0.090	258	WPR	3.055	0.111	4.320	0.179	− 0.342	0.732	437	UCT	3.095	0.124	4.573	0.201	− 3.069	0.002
80	FET	2.385	0.115	4.410	0.268	− 2.246	0.025	259	PAY	2.925	0.103	4.277	0.168	− 5.785	0.000	438	AID	2.945	0.114	4.808	0.188	− 0.750	0.454
81	BTMX	3.307	0.220	3.802	0.243	− 1.092	0.276	260	DGTX	3.108	0.150	4.141	0.210	0.048	0.962	439	TUBE	3.688	0.146	5.383	0.221	− 4.341	0.000

Table 1 (continued)

MkCapRk	SYM	AlphaP	Sd.P	AlphaN	Sd.N	kssStat	kssPvalue	MkCapRk	SYM	AlphaP	Sd.P	AlphaN	Sd.N	kssStat	kssPvalue	MkCapRk	SYM	AlphaP	Sd.P	AlphaN	Sd.N	kssStat	kssPvalue
82	SUSD	2.874	0.145	3.524	0.177	−0.643	0.520	261	BRD	3.649	0.137	3.886	0.154	−6.409	0.000	440	TZC	2.559	0.106	2.468	0.065	0.669	0.504
83	DENT	2.716	0.095	5.428	0.221	−1.998	0.046	262	UBQ	3.321	0.123	3.631	0.136	−2.395	0.017	441	TDX	2.629	0.103	2.602	0.073	−0.949	0.343
84	VTHO	3.856	0.192	3.700	0.165	−2.310	0.021	263	MTX	2.434	0.081	2.868	0.092	−7.979	0.000	442	TDP	3.528	0.211	4.643	0.285	1.347	0.179
85	MAID	3.420	0.125	4.802	0.202	−0.899	0.369	264	HIT	2.788	0.115	3.578	0.164	−2.575	0.010	443	TAU	2.864	0.106	3.749	0.142	−2.110	0.035
86	DCN	2.489	0.087	3.548	0.122	−7.703	0.000	265	TNB	4.389	0.177	3.618	0.138	−3.251	0.001	444	AMM	3.012	0.115	5.196	0.204	−3.183	0.002
87	CVC	3.251	0.117	3.399	0.126	−2.754	0.006	266	PART	3.509	0.135	5.192	0.213	−2.874	0.004	445	TCHTRX	2.860	0.319	4.422	0.528	−0.433	0.667
88	MLN	4.373	0.180	3.824	0.145	−0.816	0.415	267	EVX	2.428	0.078	6.285	0.266	4.576	0.000	446	TKY	3.616	0.150	4.901	0.199	−2.980	0.003
89	POWR	2.388	0.072	3.926	0.154	−3.232	0.001	268	BLOCK	3.246	0.120	3.794	0.144	−9.385	0.000	447	THEMIS	3.205	0.235	4.424	0.321	−0.966	0.335
90	LAMB	3.193	0.215	2.665	0.158	−3.989	0.000	269	KCASH	2.904	0.102	6.426	0.281	−1.321	0.187	448	TIME	2.633	0.090	4.400	0.170	−5.574	0.000
91	FUN	3.761	0.143	5.989	0.264	−2.264	0.024	270	MWAT	2.419	0.079	5.704	0.250	−1.957	0.051	449	BTO	2.350	0.075	3.065	0.107	−2.311	0.021
92	PROM	2.715	0.313	2.579	0.174	−0.569	0.570	271	PVT	2.695	0.219	4.078	0.379	−2.757	0.007	450	ESBC	2.384	0.140	2.349	0.128	−2.607	0.010
93	IOTX	3.251	0.133	3.834	0.166	−3.090	0.002	272	NEU	2.881	0.104	4.282	0.163	−4.235	0.000	451	ENE	3.183	0.114	3.624	0.139	−2.305	0.021
94	NOIA	3.432	0.277	3.505	0.273	−0.265	0.791	273	ITC	2.864	0.102	6.422	0.276	−3.199	0.001	452	ENQ	2.404	0.184	3.741	0.337	−0.701	0.485
95	STMX	3.146	0.122	3.885	0.156	−1.619	0.106	274	DLT	4.002	0.164	5.076	0.205	−3.523	0.000	453	ENTER	2.758	0.111	3.762	0.126	−2.857	0.004
96	NKN	3.232	0.139	3.624	0.153	−5.471	0.000	275	AUC	2.492	0.106	4.091	0.169	0.091	0.928	454	EXCL	2.900	0.104	5.757	0.239	−4.794	0.000
97	ELF	3.580	0.134	3.169	0.115	−2.402	0.017	276	PENTA	3.324	0.137	4.146	0.178	−2.140	0.033	455	ECA	2.472	0.091	3.253	0.104	3.974	0.000
98	ARDR	3.078	0.107	5.116	0.218	−3.448	0.001	277	XMX	2.497	0.121	2.354	0.076	−1.759	0.079	456	ELLA	3.080	0.108	3.227	0.118	−1.329	0.184
99	STRAX	3.227	0.116	6.003	0.263	−3.229	0.001	278	OLT	2.860	0.116	5.101	0.246	−2.639	0.009	457	BTU	5.430	0.226	4.013	0.162	−4.838	0.000
100	STEEM	3.349	0.125	3.510	0.129	−3.359	0.001	279	ADK	4.004	0.181	5.725	0.253	−3.829	0.000	458	EMPC	3.111	0.110	3.622	0.138	−2.399	0.017
101	META	2.374	0.102	3.767	0.191	−2.563	0.011	280	SNGLS	3.092	0.108	4.138	0.166	−3.855	0.000	459	EDG	3.199	0.118	3.280	0.116	−6.296	0.000
102	BMX	3.328	0.142	2.616	0.101	−1.428	0.154	281	VIB	3.593	0.135	5.942	0.260	−2.940	0.003	460	EDN	3.411	0.176	11.059	0.717	−2.905	0.004
103	ANT	3.046	0.109	3.298	0.118	−5.214	0.000	282	MTH	2.828	0.097	4.629	0.187	−5.652	0.000	461	BTX	3.436	0.133	3.470	0.124	−3.122	0.002
104	RIF	3.296	0.186	5.097	0.302	−5.353	0.000	283	POA	5.065	0.220	4.122	0.172	0.129	0.897	462	FXT	2.483	0.101	3.289	0.119	−2.946	0.003
105	KMD	3.660	0.139	4.070	0.161	−3.333	0.001	284	COV	3.560	0.138	4.050	0.155	−2.889	0.004	463	GALI	4.373	0.272	2.387	0.103	−1.302	0.194
106	ARK	4.123	0.165	3.218	0.115	−4.321	0.000	285	QUN	2.990	0.108	2.508	0.078	−6.236	0.000	464	FREE	2.669	0.364	2.875	0.154	−1.602	0.112
107	MTL	3.516	0.132	4.414	0.178	−3.042	0.002	286	LOL	3.769	0.361	3.888	0.353	−0.694	0.489	465	FTX	2.492	0.096	4.154	0.170	1.912	0.056
108	AVA	4.043	0.244	4.026	0.222	−1.100	0.272	287	OPEN	2.553	0.123	3.148	0.105	−0.577	0.564	466	GARD	3.447	0.192	3.770	0.203	−0.378	0.706
109	XHV	4.589	0.254	3.926	0.186	−1.750	0.081	288	ABL	3.117	0.218	2.456	0.145	−1.102	0.272	467	GBX	3.451	0.137	4.148	0.155	2.335	0.020

Table 1 (continued)

MkCapRk	SYM	AlphaP	Sd.P	AlphaN	Sd.N	kssStat	kssPvalue	MkCapRk	SYM	AlphaP	Sd.P	AlphaN	Sd.N	kssStat	kssPvalue	MkCapRk	SYM	AlphaP	Sd.P	AlphaN	Sd.N	kssStat	kssPvalue
110	DNT	2.793	0.094	3.235	0.116	−1.840	0.066	289	MVP	2.759	0.111	2.387	0.082	−4.510	0.000	468	GCC	4.545	0.190	3.208	0.113	0.021	0.983
111	HDAC	2.740	0.121	2.581	0.092	−3.421	0.001	290	SRN	3.051	0.109	4.005	0.156	−1.686	0.092	469	FACE	4.150	0.191	3.923	0.167	−3.307	0.001
112	ETN	2.816	0.102	3.386	0.118	−5.644	0.000	291	PLBT	3.539	0.139	4.007	0.151	−5.799	0.000	470	FOTA	3.094	0.117	4.194	0.165	−3.761	0.000
113	MONA	2.369	0.077	4.030	0.149	0.626	0.532	292	CV	3.473	0.136	4.762	0.196	−2.092	0.037	471	FLIXX	2.709	0.095	7.003	0.297	–	0.000
114	GET	5.550	1.175	24.823	0.891	−3.800	0.000	293	CHP	3.342	0.132	3.476	0.135	−2.004	0.046	472	FLO	3.696	0.148	3.914	0.146	−2.087	0.037
115	BTS	2.950	0.100	2.323	0.071	−3.558	0.000	294	FTC	2.986	0.111	3.514	0.124	−3.715	0.000	473	CPAY	3.090	0.115	5.539	0.227	0.852	0.394
116	TOMO	12.048	0.636	3.949	0.160	−3.226	0.001	295	DDD	2.972	0.114	4.702	0.192	−3.587	0.000	474	CATT	3.182	0.163	4.049	0.216	−2.216	0.027
117	WAN	3.779	0.157	4.606	0.199	−0.888	0.375	296	LET	4.282	0.207	3.386	0.141	−3.653	0.000	475	COSM	3.570	0.166	3.986	0.182	−2.759	0.006
118	YCC	2.386	0.120	4.425	0.258	−4.110	0.000	297	INT	2.740	0.096	2.365	0.070	−6.379	0.000	476	CURE	3.014	0.108	3.863	0.146	−2.765	0.006
119	DAG	5.482	0.295	4.296	0.205	−2.792	0.005	298	BCDT	3.826	0.214	3.209	0.140	−1.432	0.153	477	CREED	3.072	0.108	3.607	0.137	−2.393	0.017
120	UTK	3.326	0.123	3.847	0.149	−3.726	0.000	299	EMC	2.841	0.099	3.370	0.120	−3.505	0.000	478	CAS	2.451	0.082	3.411	0.121	0.132	0.895
121	VRA	2.723	0.155	4.674	0.285	−4.842	0.000	300	LBA	3.105	0.127	3.580	0.147	−3.594	0.000	479	CRW	3.598	0.140	4.122	0.159	−4.129	0.000
122	C20	3.381	0.129	4.020	0.158	−4.626	0.000	301	BOX	3.066	0.134	2.630	0.099	−0.264	0.792	480	CMCT	2.852	0.122	4.303	0.189	2.616	0.009
123	GLM	2.423	0.210	3.790	0.130	−2.171	0.030	302	PLAY	2.428	0.084	2.684	0.085	−4.415	0.000	481	DIT	2.423	0.092	2.446	0.087	−4.129	0.000
124	TRAC	3.119	0.115	2.577	0.085	−6.069	0.000	303	DAT	3.182	0.121	2.762	0.087	−5.611	0.000	482	DMS	3.251	0.491	2.324	0.354	−6.884	0.000
125	REP	2.354	0.072	4.513	0.182	−1.328	0.185	304	SNET	2.774	0.190	4.151	0.305	−1.594	0.113	483	DICE	3.083	0.112	7.765	0.346	−4.916	0.000
126	BTM	2.957	0.106	2.420	0.072	−5.986	0.000	305	FAIRG	3.147	0.122	3.438	0.125	−4.049	0.000	484	DGX	2.464	0.096	2.414	0.089	−2.239	0.026
127	ADX	2.387	0.074	2.301	0.067	−4.534	0.000	306	MDCL	2.955	0.110	3.324	0.121	−5.383	0.000	485	DRA	4.310	0.174	3.498	0.130	−2.203	0.028
128	RDD	3.443	0.133	4.434	0.173	−4.132	0.000	307	WINGS	2.619	0.087	4.136	0.160	−6.532	0.000	486	DRG	3.743	0.159	3.911	0.164	1.051	0.293
129	DBC	2.333	0.071	6.683	0.295	−2.960	0.003	308	UUU	3.366	0.145	4.392	0.193	−3.032	0.003	487	DROP	2.800	0.097	3.919	0.149	−0.630	0.529
130	GAS	3.218	0.118	3.606	0.134	−3.889	0.000	309	DRT	3.342	0.138	4.067	0.146	−2.636	0.009	488	DOV	2.925	0.108	3.683	0.132	−2.993	0.003
131	PNK	3.144	0.110	3.616	0.139	−0.809	0.419	310	ACAT	2.887	0.109	5.515	0.243	−2.720	0.007	489	CANN	2.349	0.080	3.278	0.108	−6.475	0.000
132	QKC	4.026	0.179	3.412	0.144	−2.214	0.027	311	POT	3.166	0.115	3.691	0.139	−5.325	0.000	490	CVCOIN	2.639	0.104	3.509	0.149	−0.993	0.321
133	EURS	2.768	0.129	3.281	0.130	−1.730	0.084	312	SKM	2.370	0.084	6.048	0.286	−4.926	0.000	491	CXO	5.551	0.244	3.823	0.153	−2.992	0.003
134	PPT	8.131	0.380	3.423	0.125	1.719	0.086	313	BLK	3.364	0.124	4.668	0.192	−4.002	0.000	492	DERO	3.373	0.181	4.924	0.273	−1.176	0.240
135	CENNZ	2.842	0.110	4.603	0.195	−0.120	0.905	314	TOPC	2.697	0.096	3.436	0.122	−5.205	0.000	493	DEV	3.510	0.141	4.840	0.222	−2.745	0.006
136	APL	3.503	0.168	3.575	0.146	−1.570	0.117	315	RNT	2.424	0.082	3.865	0.139	−3.983	0.000	494	DAV	2.338	0.092	4.556	0.226	−4.848	0.000
137	MFT	3.653	0.160	3.668	0.164	−1.201	0.230	316	SNC	2.482	0.082	3.858	0.142	−4.724	0.000	495	DAY	2.584	0.091	7.990	0.339	−3.453	0.001

Table 1 (continued)

MkCapRk	SYM	AlphaP	Sd.P	AlphaN	Sd.N	kssStat	kssPvalue	MkCapRk	SYM	AlphaP	Sd.P	AlphaN	Sd.N	kssStat	kssPvalue	MkCapRk	SYM	AlphaP	Sd.P	AlphaN	Sd.N	kssStat	kssPvalue
138	TT	3.465	0.249	4.316	0.305	−2.665	0.008	317	TRIO	2.480	0.089	3.767	0.154	−6.960	0.000	496	DCT	2.337	0.077	2.979	0.096	−5.197	0.000
139	TROY	3.468	0.744	2.668	0.446	0.927	0.374	318	EOSDAC	3.212	0.136	4.259	0.185	−1.639	0.102	497	MCAR	3.115	0.110	3.622	0.138	−2.396	0.017
140	IGNIS	2.974	0.110	3.678	0.132	−1.004	0.316	319	SILKT	3.457	0.258	5.919	0.499	−0.626	0.532	498	BLAST	4.355	0.362	3.313	0.218	−3.644	0.000
141	CARRY	2.832	0.235	2.918	0.211	−3.298	0.001	320	JAR	2.931	0.213	4.260	0.318	−2.460	0.015	499	LCC	3.709	0.152	4.286	0.175	3.028	0.003
142	AION	3.205	0.117	2.359	0.070	−3.445	0.001	321	BU	2.756	0.143	3.158	0.179	−2.431	0.016	500	LCP	3.152	0.113	4.933	0.206	−4.161	0.000
143	ELA	3.288	0.124	4.858	0.205	−5.323	0.000	322	PST	2.836	0.101	4.082	0.154	−3.512	0.000	501	LND	2.853	0.114	3.399	0.133	−4.292	0.000
144	SNM	3.979	0.156	3.804	0.147	−2.514	0.012	323	CAPP	2.891	0.104	4.016	0.151	−2.312	0.021	502	LNKC	2.773	0.114	2.977	0.108	−4.084	0.000
145	BCN	3.067	0.110	3.869	0.148	1.940	0.053	324	LIKE	2.559	0.089	2.655	0.081	−1.642	0.101	503	LOCK	4.103	0.284	5.442	0.425	−0.035	0.972
146	STPT	2.424	0.146	5.171	0.405	−1.841	0.067	325	EXP	2.823	0.101	2.696	0.084	0.928	0.354	504	NGC	5.142	0.221	4.178	0.163	−1.516	0.130
147	NRG	3.855	0.222	4.503	0.228	−4.424	0.000	326	GEM	3.068	0.122	4.559	0.200	1.321	0.187	505	BERRY	2.975	0.114	3.396	0.132	−2.402	0.017
148	AKRO	7.443	1.563	2.959	0.196	−0.840	0.403	327	AVT	2.377	0.081	3.179	0.104	−2.615	0.009	506	NTK	3.212	0.124	3.662	0.140	−1.888	0.060
149	AERGO	3.217	0.175	4.773	0.275	−3.031	0.003	328	IDT	2.580	0.106	4.053	0.168	−4.718	0.000	507	NLC2	3.058	0.110	5.681	0.240	−4.380	0.000
150	AST	2.459	0.078	5.027	0.207	−4.783	0.000	329	IOC	2.975	0.106	3.893	0.147	−4.243	0.000	508	MOBI	2.690	0.098	4.305	0.169	−0.952	0.341
151	REQ	3.375	0.124	3.474	0.129	−3.199	0.001	330	CHX	2.425	0.091	3.081	0.123	−3.719	0.000	509	BIHU	4.139	0.324	2.440	0.140	−0.776	0.439
152	XYO	2.553	0.097	2.595	0.092	−7.376	0.000	331	BIX	3.433	0.149	3.198	0.125	−3.849	0.000	510	MUE	3.782	0.152	3.399	0.121	−3.317	0.001
153	DOCK	3.935	0.174	4.477	0.195	−3.597	0.000	332	DACC	4.788	0.248	3.136	0.134	−0.840	0.402	511	MYST	2.439	0.083	2.388	0.067	−1.665	0.096
154	QASH	3.044	0.113	3.362	0.118	−5.498	0.000	333	TPAY	2.574	0.102	4.085	0.174	−1.611	0.108	512	MT	3.878	0.148	3.900	0.154	−26.103	0.000
155	DATA	3.380	0.127	7.378	0.329	−3.557	0.000	334	VIBE	2.384	0.078	3.719	0.133	−7.134	0.000	513	BOXT	4.308	0.349	2.366	0.129	−2.854	0.005
156	LBC	3.601	0.139	3.495	0.128	−3.102	0.002	335	MEET	2.995	0.128	9.073	0.474	−0.851	0.395	514	HMQ	2.313	0.077	2.589	0.076	−2.755	0.006
157	DGD	2.960	0.105	4.363	0.172	−3.656	0.000	336	ADB	3.212	0.120	3.888	0.152	−1.191	0.234	515	HER	2.550	0.097	2.663	0.098	−5.994	0.000
158	GTO	2.357	0.072	4.376	0.176	−2.086	0.037	337	BUT	2.361	0.171	3.847	0.384	3.142	0.002	516	HGT	2.435	0.085	4.132	0.148	−2.712	0.007
159	PRO	3.057	0.114	5.285	0.212	−4.484	0.000	338	HSC	2.858	0.116	3.948	0.177	−4.946	0.000	517	HTML	2.447	0.083	4.212	0.155	1.860	0.063
160	NEW	3.923	0.277	3.313	0.234	−2.507	0.013	339	ZSC	2.375	0.075	3.939	0.148	−4.450	0.000	518	BOS	3.588	0.143	4.443	0.171	−5.444	0.000
161	BFT	3.456	0.142	4.324	0.189	−2.790	0.005	340	UBEX	2.897	0.122	3.687	0.171	1.504	0.133	519	IDH	2.937	0.108	5.456	0.227	−3.203	0.001

Table 1 (continued)

MkCapRk	SYM	AlphaP	Sd.P	AlphaN	Sd.N	kssStat	kssPvalue
162	AE	3.850	0.150	5.385	0.228	− 2.170	0.030
163	LOOM	5.056	0.231	4.911	0.214	− 1.510	0.132
164	GRS	2.446	0.079	6.339	0.268	− 1.744	0.082
165	RCN	3.054	0.109	6.863	0.302	− 3.432	0.001
166	CONT	3.263	0.279	2.507	0.174	− 1.966	0.051
167	NULS	4.594	0.187	5.995	0.263	− 4.009	0.000
168	CVT	2.897	0.130	2.851	0.104	− 6.169	0.000
169	PERL	3.390	0.294	4.038	0.392	− 1.450	0.150
170	ARPA	3.019	0.263	2.328	0.162	− 1.632	0.106
171	UTT	3.652	0.148	4.068	0.152	− 0.114	0.909
172	RFR	3.078	0.115	3.634	0.143	− 3.710	0.000
173	DRGN	2.332	0.071	4.641	0.187	− 2.908	0.004
174	DUSK	3.056	0.231	4.418	0.362	− 3.113	0.002
175	EMC2	3.127	0.114	3.053	0.105	− 1.665	0.096
176	WICC	2.748	0.096	3.965	0.152	− 2.988	0.003
177	MOC	2.315	0.131	3.467	0.237	0.026	0.979
178	PIVX	3.397	0.126	3.341	0.122	− 3.331	0.001
179	MET	2.980	0.125	3.612	0.154	− 0.442	0.658
341	COFI	2.568	0.088	4.581	0.183	− 0.909	0.364
342	TOS	2.996	0.130	3.341	0.156	− 2.986	0.003
343	STK	3.538	0.140	2.527	0.080	− 2.928	0.004
344	NCC	2.979	0.178	5.023	0.356	− 2.486	0.014
345	AIDOC	3.178	0.117	3.971	0.155	− 3.411	0.001
346	MEX	3.038	0.291	8.570	0.977	− 3.466	0.001
347	AAC	2.973	0.121	4.136	0.172	0.088	0.930
348	ELEC	3.317	0.133	6.375	0.295	− 1.312	0.190
349	FYP	2.633	0.100	5.524	0.210	− 3.199	0.001
350	LATX	2.515	0.096	3.620	0.144	− 4.221	0.000
351	DTA	2.731	0.094	3.468	0.129	− 4.945	0.000
352	FUEL	2.818	0.097	3.342	0.120	− 3.928	0.000
353	DPY	3.510	0.136	3.700	0.137	− 0.542	0.588
354	ETZ	2.853	0.107	3.153	0.106	− 0.960	0.337
355	BRYLL	2.720	0.417	3.009	0.158	− 0.209	0.835
356	SIB	3.199	0.120	4.506	0.176	− 0.131	0.896
357	SENC	2.300	0.076	4.312	0.186	2.635	0.009
358	LUN	3.138	0.113	3.780	0.144	− 2.236	0.026
520	HVN	2.977	0.107	3.869	0.145	− 0.143	0.886
521	GLT	2.831	0.102	2.432	0.071	1.026	0.305
522	BRO	2.577	0.083	4.118	0.161	− 5.000	0.000
523	BNK	2.866	0.112	3.548	0.150	− 0.654	0.513
524	KEK	2.623	0.088	4.969	0.201	− 1.095	0.274
525	KIND	2.710	0.154	3.330	0.173	− 1.671	0.096
526	KOBO	5.715	0.511	2.657	0.065	− 0.681	0.496
527	BMH	3.078	0.137	6.413	0.286	1.286	0.199
528	BMC	3.044	0.108	3.390	0.124	− 4.988	0.000
529	INK	2.834	0.101	3.794	0.140	− 4.775	0.000
530	INSTAR	5.066	0.236	4.319	0.177	− 3.891	0.000
531	ILK	3.183	0.369	3.490	0.389	0.060	0.953
532	JMC	4.292	0.264	3.843	0.208	− 0.210	0.834
533	BOLT	4.254	0.368	3.689	0.288	− 2.137	0.034
534	BOLI	2.764	0.101	3.219	0.108	− 7.651	0.000

From Eq. 2, the stock return series, R_t, is considered a random walk only if $\rho = 1$, whereas if $|\rho| < 1$, then the series is a stationary and predictable process, which violates the weak-form EMH. We apply the Kapetanios, Shin, and Snell (KSS) nonlinear unit root test (Kapetanios et al. 2003), which is more robust when there are market frictions (i.e., transaction costs) as it is proposed in Apopo and Phiri (2021) for cryptocurrency markets. We use the test implementation in the R package by Guris (2021). The null hypothesis is that the raw time series of log returns is a random-walk type against the alternative of a stationary process. Results applied to the 2018-19 window are reported in Table 1 and we demonstrate that we cannot discard the null hypotheses with a p-value higher than 0.01 in approximately 50% of the cryptos. These results are aligned with others in the cryptocurrency market (Kyriazis 2019).

The resulting number of eligible cryptocurrencies for portfolios is 534. Additionally, we also consider two smaller sets with 250 and 175 cryptocurrencies with the highest market capitalization among the eligible ones.

Simulation

In this section, we describe our methodology. We perform a Monte Carlo experiment repeating each simulation 1500 times to better assess the outcome of the different investment methods considered. The simulation period is 17 months, from January 2020 to May 2021. In each simulation, we run a sequence of investments known as a *Random Investment Path* (RIP) that consists of a sequence of t_w days where it will be considered entering the market.

The RIPs are the same for all the investment methods under analysis: the ones proposed and those used for benchmarking. For each investment at time t_w, we consider a 2-year estimation window (730 days) from t_{w-729} to t_w, which is used to estimate the portfolio. If an investment is made, the investing window will always be held for 30 days, that is, from t_{w+1} to t_{w+30}.

The investment path is created as follows. For each of the 17 months of the simulation period, there is a 50% probability of investing in that month. If the month is selected, we randomly select the day t_w when the investment will be made. Since the holding period of an investment is 30 days, we ensure that there is no overlap between the holding period of t_w and the next investment day $t_w + 1$.

The flowchart in Fig. 1 summarizes the investment process, which consists of the following steps:

1 *Data selection* As explained, for each investment time t_w, we use a two-year sliding window for the estimation of the optimal portfolio, and 30-days as the holding period for the investment.
2 *Market segmentation* At this stage, we use a prototype-based clustering algorithm to segment the market. In particular, we use a k-medoids algorithm and a quality index to automatically set the *k* value. We repeat the process 50 times to remove the uncertainty of the initialization of the algorithm. Each partition is denoted as P_n in the chart.
3 *Prototype selection strategy* Given the prototypes of the 50 segmentations of the previous step, we apply a heuristic to select the most suitable cluster for later portfo-
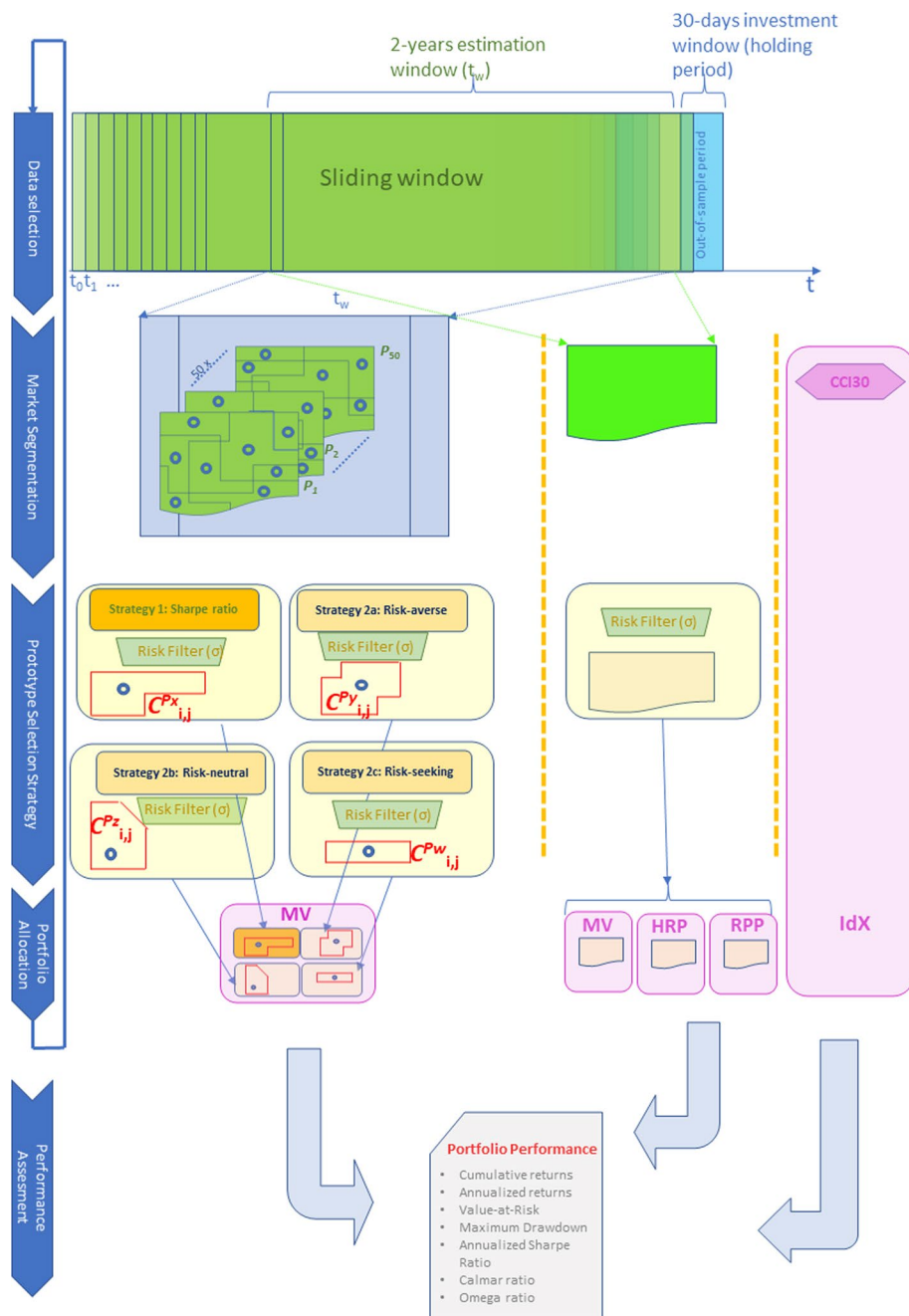
Fig. 1 Flowchart of the algorithm for portfolio allocation

lio optimization. In particular, we consider four different strategies, three of which are related to risk measures and another based on the well-known Sharpe ratio. If no prototype matches the strategy requirements, then no cluster is selected and no investment will be made at t_w. For a given strategy and an investment time t_w, we select a cluster i of the partition P_n denoted as $C_{w,i}^{P_n}$.

4 *Portfolio allocation* The MVO method is applied to the cryptocurrencies that belong to the cluster $C_{w,i}^{P_n}$ selected for each strategy once the risk filter is applied to remove

extremely volatile cryptocurrencies. It may be the case that the portfolio optimization produces no result, in such case, no investment is made. The reason behind this fact can be due to the covariance matrix being non-symmetric positive definite and hence not invertible or because the optimization problem is ill-conditioned and it does not find a solution. It should be noted that the quadratic function to minimize in our model as defined in Eq. 8 is convex and reaches a global optimum solution if and only if the covariance matrix is semi-definite positive.

Before applying the portfolio optimization method, we apply a risk filter to remove the cryptocurrencies with extreme volatility in the last month.

5 *Performance assessment* This last step is carried out once the investment path is executed. We then measure the performance of the MV model with four of the proposed strategies and of the benchmarks. The performance is measured using profit, risk, and profit-risk indicators. As benchmarks, we use well-known portfolio optimization models over all the cryptocurrencies, that is, with no market segmentation and prototype selection. In particular, the mean-variance (MV), the risk parity (RPP), and the hierarchical risk parity (HRP). In addition, we also apply the random investment paths on the market index CCI30 to compare the strategies against the main market trend.

Sampling strategy and data selection

Our dataset ranges from 1st January, 2018 to 30th May, 2021. For each investment at time t_w, we consider a 2-year estimation window (730 days) from t_{w-729} to t_w; this data is used to estimate the portfolios. The investing period is 30 days, that is, from t_{w+1} to t_{w+30}. At this stage, we apply the so-called *filter risk* to remove cryptocurrencies with extreme volatility in the last month of the estimation window from those considered. In particular, we remove those with $\sigma > 1$ in the last 30 daily returns. We also remove cryptocurrencies with missing data or that were no longer traded during the estimation period.

Market segmentation

First, we describe each cryptocurrency in our dataset using the average and standard deviation (σ, μ) of the daily returns (Eq. 1) computed along the estimation window. This representation is used later for the automatic selection of the cluster, according to the investment strategy and is also consistent with the MV portfolio optimization. In addition, it succinctly summarizes the profitability and volatility of each asset and has been successfully used for clustering cryptocurrencies (Lorenzo and Arroyo 2022). While more sophisticated representations are previous reference, the (σ, μ) variables make a faster computation possible, which is crucial due to the intensive calculations of the simulations.

For market segmentation, we use a partitional prototype-based clustering algorithm. We need the algorithm to produce a partition of disjoint subsets of assets (cryptocurrencies in our case) and we need a prototype representing each subset for some of the investment strategies explained below.

In clustering, the prototype represents the cluster elements optimally, that is, it typically minimizes the total distance between all the cluster objects and itself. The process is usually (Henning et al. 2016) formalized as the minimization of $S(\mathcal{D}, m_1, ..., m_k)$ as follows by choice of the prototype $m_1, ..., m_k$,

$$
\begin{aligned}
S(\mathcal{D}, m_1, ..., m_k) &= \sum_{i=1}^{n} d(x_i, m_{c(i)}), \ where \\
c(i) &= \operatorname*{arg\,min}_{j \in (1,...,K)} \ d(x_i, m_j), j = 1, ..., n
\end{aligned}
\tag{3}
$$

where n in Eq. 3 is the total number of objects in the space $\mathcal{D}$, d is the dissimilarity measure function, and K is the number of clusters. The prototypes $m_1, ..., m_k$ may be required to be objects in $\mathcal{D}$. d may be the given distance between the observation x_i and the cluster centroids or prototypes m_c.

We use Euclidean Distance (ED) $d^2(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|^2$ where $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^n$ because it is both simple and meaningful where the resulting space meets the appropriate mathematical properties. Furthermore, it has also been used in a similar application for prototype-based clustering of cryptocurrencies represented as the average return and volatility (Lorenzo and Arroyo 2022; Mattera et al. 2021; Nguyen Cong et al. 2014), as in our case.

In some clustering algorithms, for example, in the K-means, the prototype is the mean of the objects. However, we want the prototype to be an observed object, so we use a K-medoids or Partition Around Medoids (PAM) algorithm.

Owing to the intensive use of the clustering algorithm in our simulations, we use a computationally-efficient version of the PAM algorithm called CLARA(Clustering LARge Applications). The difference is that these algorithms use only random samples of the dataset (instead of the entire dataset) to compute the medoids. However, it is important to note that the resulting partition includes all the elements of the dataset.

The CLARA algorithm belongs to the family of prototype-based clustering algorithms (Kaufman 1986; Kaufman and Rousseeuw 1990). It is a PAM algorithm adapted to large datasets. Readers interested in an in-depth analysis of the PAM-CLARA/CLARANS algorithm can refer to the work by Schubert and Rousseeuw (2019).

In the CLARA algorithm, we have to determine three parameters: the size of the random sample sampsize, the minimum cluster cardinality, and the number of clusters K.

In our case, sampsize vary in each execution from 50 to 100, so 50 partitions are generated. These are the aforementioned executions of the clustering algorithm. In each execution, the resampling changes and therefore slightly changes the outcome of the clustering.

For each sampsize iteration, we save all the clusters with a cardinality higher than 10 ensuring that the cluster is sufficiently large to run the portfolio allocation algorithms efficiently.

There has been extensive research on cluster detection and evaluation (Kou et al. 2014; Li et al. 2021). In this case, we rely on an automatic process that consists of computing different Cluster Validity Indices (CVIs) for crisp partitions (Arbelaitz et al. 2013), including Silhouette, Dunn, COP Davies-Bouldin, Calinski-Harabasz, or

Table 2 Strategies for prototype selection

Name	Strategy ID	Type	Description
Sharpe ratio	Strategy 1	Highest sharpe ratio	Cluster with the higher training portfolio sharpe ratio
Risk-aversion criteria	Strategy 2a	Risk-averse	Cluster with the lower volatility-centroid values
	Strategy 2b	Risk-neutral	Cluster with average volatility-centroid values
	Strategy 2c	Risk-seeking	Cluster with the higher volatility-centroid values

the score function, and then apply the *majority rule* to select the number of clusters K that is best according to more CVIs.

The outcomes of clustering algorithms for different runs are groups or clusters of cryptocurrencies, each represented by a prototype. In the next subsection, we explain how we select the cluster that is used to optimize our portfolio.

Prototype selection strategies

Our method proposes clustering as a way to partition the cryptocurrency market according to the financial behavior of the cryptocurrencies. Once we have the partitions, we need a criterion to select the portion of the market that most interests us. At this point, the medoids or prototypes of the clusters will be the inputs for a simple heuristic algorithm of cluster selection according to the different investing strategies. Only clusters with cardinality equal to or higher than 10 are considered interesting for the investor. These cardinality criteria are applied to ensure that the portfolio allocation algorithms work more efficiently. The strategies that are described next are summarized in Table 2.

Strategy 1: Sharpe ratio. The Sharpe ratio is the average excess risk-free return by volatility unit or total risk. The ratio determines the risk of the investment concerning the return of an investment with zero risk:

$$SR_c = \frac{r_P - r_f}{\sigma_P},\tag{4}$$

where r_P in Eq. 4 is the portfolio return, r_f is the *risk-free* rate and σ_P is the portfolio risk (standard deviation or the volatility of the portfolio). For r_f reference, we consider the daily of the annualized T-Bill over 90 days obtained from the Federal Reserve Economic Database (FRED) hosted by the Federal Reserve Bank of St. Louis. The greater the value of the Sharpe ratio, the more attractive the risk-adjusted return of the portfolio.

In this strategy, we compute the annualized Sharpe ratio of the MV portfolios considering the cryptocurrencies during the estimation window. This strategy is represented with the *Sharpe ratio* (SR) label in the following tables and charts.

Strategy 2: Prototype selection by a risk-aversion criteria strategy. First, we compute the volatility (σ) for all cryptocurrencies traded in the estimation window and compute the quartiles of the distribution that will serve as a reference for the volatility of the period and allow us to classify the cluster prototypes. Accordingly, we consider three different risk-aversion profiles for investors following (Goetzmann et al. 2014). In particular, the profiles are:

1. Strategy 2a represents the utility function of a *risk-averse investor* as it chooses the prototypes whose volatility is within the 1st quartile. Represented by the *Low-Risk* (LR) label in the following tables and charts.
2. Strategy 2b represents the utility function of a *risk-neutral investor* as it chooses the prototypes that are between the 2nd and the 3rd quartile of the volatility distribution. Represented by the *Mean-Risk* (MR) label.
3. Strategy 2c represents the utility function of a *risk-seeking investor* as it chooses the prototype over the 3rd quartile of the volatility distribution. Represented by the *High-Risk* (HR) label.

Since the prototype is described by two variables of average return and volatility, if more than one prototype is selected, we choose the one with the highest average return.

Portfolio allocation

In our proposal, once we select the most suitable partition according to our strategy, we run portfolio allocation using the well-known MV optimization. One of its drawbacks is its tendency to maximize the effects of errors in the input assumptions on return and volatility estimations. This means that small changes in the expected returns or the computed covariance matrix can produce very different results.

Mean-variance model assumptions

We summarize some of its main assumptions below for a single-period MV model (Steinbach 2001):

1. The existence of the two-moments expectation ($\bar{\mathbf{r}}$) and covariance matrix (Σ) being the apostrophe ($'$) transpose vector:

$$\bar{\mathbf{r}} := \mathbf{E}(\mathbf{r}); \quad \Sigma := \mathbf{E}[(\mathbf{r} - \bar{\mathbf{r}})(\mathbf{r} - \bar{\mathbf{r}})'] = \mathbf{E}[\mathbf{rr}'] - \bar{\mathbf{r}}\bar{\mathbf{r}}' \tag{5}$$

2. The returns ($\mathbf{r}$) are assumed to follow a normal distribution
3. The investors have MV preferences and thus ignore skewness

where **r** is a return vector. We also consider the following definitions:

- Definition 1.1 (reward). The *reward* (γ) of a portfolio is the mean if its returns (**r**) by the corresponding weights (**w**)

$$\gamma(\mathbf{w}) := \mathbf{E}(\mathbf{r}'\mathbf{w}) = \bar{\mathbf{r}}'\mathbf{w} \tag{6}$$

- Definition 1.2 (risk). The *risk* (R) of a portfolio is the variance of the returns

$$\begin{aligned} R(\mathbf{w}) &:= \sigma^2(\mathbf{r}'\mathbf{w}) \\ &= \mathbf{E}[(\mathbf{r}'\mathbf{w} - \mathbf{E}(\mathbf{r}'\mathbf{w}))^2] \\ &= \mathbf{E}[\mathbf{w}'(\mathbf{r} - \bar{\mathbf{r}})(\mathbf{r} - \bar{\mathbf{r}})'\mathbf{w}] \\ &= \mathbf{w}'\Sigma\mathbf{w} \end{aligned} \tag{7}$$

MV models require a risk measure as we see in Eq. 7 computed as a covariance matrix that at some point in the optimization process must be inverted, for which certain

properties in the matrix are necessary; otherwise the matrix may not be invertible and the solution obtained may have too much error. The presence of noise in the series (Pafka and Kondor 2003) and the requirements of the covariance estimator itself force us to take care when optimizing the portfolio (Ledoit and Wolf 2003). In general, the sample covariance matrix is considered suitable for applications where its inverse is not required (Gatheral 2008), so there is a problem with risk-based portfolio selection because when the matrix is inverted, the noise is amplified (Ledoit and Wolf 2004). The sample covariance matrix contains substantial statistical noise that is amplified when it is inverted and in the same way, since the return matrix contains noise, the former may not estimate the true covariance matrix. An understanding of the estimation error of the covariance is important if we want to ensure better out-of-sample performance of the optimization model. The problem arises from the fact that the covariance matrix is calculated over a finite window length T, with T being the sampling of the time series and this inevitably leads to the appearance of noise (measurement error) in the estimator itself; this effect is greater as T approaches the value of N, the number of time-series. This is something that must be taken into consideration by applying appropriate estimators (we use `cov.trob` function in R MASS package (Venables and Ripley 2002)) given the narrowness of the time window (T) and considering the high number of cryptocurrencies (N). From matrix theory, the condition number of a matrix A provides a measure of the sensitivity of the solution x of the system $Ax = b$ to perturbations in b. In many situations with time series, the ill-conditioned matrix is caused because $N > T$. Even when $T > N$, the eigenstructure tends to be systematically distorted unless $T \gg N$, resulting in a numerically ill-conditioned estimator for Σ. From a different perspective, a strong correlation between some time series corresponds with a rank deficiency as well as with non-unique solutions in MV optimization. Hence, for the stability of the solution of a risk-based model, we definitely need invertible and well-conditioned covariance matrices.

We expect that we can enhance the performance of the portfolios thanks to a better estimation of the covariance matrices by reducing the cryptocurrency space by selecting the partitions according to investor goals.

Mean-variance (MV) model

The optimization goal for the MV model is to determine the best trade-off between return and risk, subject to a set of constraints assuming that the investor knows the value of the expected return vector μ and covariance matrix Σ. Rational investors always pursue the lowest risk under a specific expected return or the highest return under a particular risk. The risk measure developed by Markowitz is an asset-weighted covariance matrix, $\mathbf{w}'\Sigma\mathbf{w}$, where Σ in Eq. 8 is the covariance matrix and $\mathbf{w}$ is the portfolio weights vector. The optimization solution is obtained by setting a target portfolio return $\bar{r}$ discounting transaction costs aligned with the model proposed by Wang et al. (2014) but considering, in our case, a fixed amount per portfolio, allowing only long positions and full invested conditions with maximum and minimum holding sizes $\omega \in [0.001, 0.5]$, such that:

$$\begin{aligned}
\underset{w}{\text{minimize}} \quad & \mathbf{w}'\Sigma\mathbf{w}, \textit{Covariance Risk}, \\
\text{subject to} \quad & \mathbf{w}'\hat{\mu} \geq \bar{\mu}_{min} + \gamma, \textit{ Target Return} \\
& \mathbf{w}'\mathbf{1} + x_f = 1, \textit{ Full investment} \\
& \mathbf{w}_{min} \geq \mathbf{w} \geq \mathbf{w}_{max}, \textit{ Holding sizes for long} - \textit{only positions}
\end{aligned}$$
(8)

where $\hat{\mu}$ is the estimated mean return vector of the cryptocurrencies computed on historical return values and $\bar{\mu}$ is the target required return. After portfolio optimization, we invest the remaining budget not allocated in cryptos due to portfolio weight constraints in risk-free security (x_f).

The vectors $\mathbf{w}_{min}$ and $\mathbf{w}_{max}$ in Eq. 8 are the hold lower and upper position bounds where $\mathbf{w} = [\omega_0, \omega_1, ..., \omega_N]'$. Under a basic constraint, the weights for allocated assets in the portfolio model Eq. (8) lie between 0 and 1 (long positions), and they sum up to 1 (fully invested portfolio). TC in Eq. 9 is the Transaction Cost as defined in Eq. (9).

In the presence of market friction, there are transaction costs paid by the investor to trade on the market.[5] For the MV model, transaction costs are computed as follows,

$$TC(\mathbf{w}) = \sum_{i=0}^{N} C_i(w_i) \equiv \gamma,$$
(9)

where N is the number of cryptocurrencies allocated into the portfolio and C_i is the Costs (C) expressed in basis points ($1\,bps = 1/100\% = 1/10000\,\$$). In our case, for computation of Eq. 9, we simply consider 5 bps per portfolio, which transforms Eq. 9 in a constant γ.

The following is an explanation of the key terms in Eq. 8:

- The *targeted portfolio return* ($\bar{\mu}_{min}$) is computed based on the MAR of the last 30-days of historical returns for the $H \times n$ matrix $\mathbf{R}$ where n is the variable number of cryptoassets allocated into the portfolio during the holding period H. The assumption is that the market will behave during the holding period at least as well as the last 30 days taken from the estimation window

$$\bar{\mu}_{min} = \mathbb{E}(\mathbf{R}_w)_{30d} = \mathbf{w}'\mathbb{E}(\mathbf{R})_{30d} = (\mathbf{w}'\mu)_{30d}$$
(10)

- The *portfolio variance* (σ) for the objective function

$$\sigma_w = Var(\mathbf{R}_w) = \sum_{i,j} Cov(\mathbf{r}_i, \mathbf{r}_j)w_i w_j = \mathbf{w}'\Sigma\mathbf{w}$$
(11)

A *Quadratic Program* (QP) is an optimization problem whose objective is to minimize or maximize a quadratic function subject to a finite set of linear equality and inequality constraints. QP models are applied for solving many problems including most of *mean-variance models* Markowitz (Cornuejols and Tütüncü 2006) where Σ is part of the objective function as in Eq. 8. The R package selected to solve the quadratic programming problem is quadprog that implements the dual method of Goldfarb and Idnani (1983).

---

[5] https://www.blueskycapitalmanagement.com/how-to-trade-hedge-crypto-and-related-transaction-cost-analysis-tca/.

The optimal solution to 8 is a weight vector **w** that will produce the optimal portfolio financial return (r_p) at time t when applied to the allocated cryptocurrencies:

$$r_{p_t} = \sum_{i=1}^{n} w_i \cdot R_{t_i} = \mathbf{w}' \mathbf{r}_t \tag{12}$$

where **r** is a $n \times 1$ vector crypto-asset returns for n crypto assets allocated for the portfolio, and w_i is the weights of the crypto-asset i with a return r_i. In our case, considering the aforementioned maximum and minimum holding sizes, we only consider those cryptocurrencies with weights strictly higher than 0.001 and lower than 0.5, and consider a free-risk asset until 100% of the investment is complete. In this way, we aim to obtain portfolios with a reasonable cardinality.

The cumulative portfolio return (Rp) at the end of the holding period H days applying Eq. 12 is

$$Rp = \prod_{t}^{H}(1 + r_{p_t}) - 1 \tag{13}$$

We evaluate the performance of the investment methods using different return indicators based on Eq. 13 throughout the investing period. Regarding profit indicators, we use the arithmetic cumulative return ($r_A = \sum_{t=1}^{T} Rp$) and the geometric compounding return ($r_G = (\prod_{t=1}^{T}(1 + Rp)) - 1$); for the arithmetic average or average return per period (e.g., one year), we have ($\bar{r}_A = \frac{1}{T} \sum_{t=1}^{T} Rp$) and the geometric average return ($\bar{r}_G = (\prod_{t=1}^{T}(1 + Rp))^{\frac{1}{T}} - 1$), the compound annual return or annualized return for the annualized returns ($r_G^{ann} = (\prod_{t=1}^{T}(1 + Rp))^{\frac{n}{T}} - 1$), and the annualized arithmetic average return ($r_A^{ann} = \frac{n}{T} \sum_{t=1}^{T} Rp$). In all cases, Rp is computed based on Eq. 13, T is the number of periods under analysis and n is the number of periods within the year (monthly $n = 12$). Arithmetic ratios reflect the additive relationship and geometric ones reflect compounding relationships. Compounding rates apply to investors that reallocate the funds obtained after one investment period into the next one.

Benchmarking of different portfolio allocation models

We benchmark the proposed investing strategies with other portfolio allocation models well stated in the financial literature applied to the whole market.

- Mean-Variance (MV) (Markowitz 1952a, 1959).
- Hierarchical Risk Parity (HRP) (De Prado 2016).
- Risk Parity Portfolio (RPP) (Roncalli 2013).

In addition, we also compare the performance of the methods against investing in the market index CCI30, a weighted market cap index launched on January 1, 2017. This price index is a weighted average of the 30 largest cryptocurrencies by market capitalization, and it is a good representative of the market's overall growth and daily and long-term movement. We briefly present the HRP and the RPP models that are risk-based portfolio methods below.

Risk Parity Portfolio (RPP). A key concept in RPP model is the Marginal Risk Contribution (MRC) defined as follows:

$$MRC_i = \frac{\partial \sigma_P}{\partial \omega_i} = \frac{\omega_i \sigma_i^2 + \sum_{i \neq j} \omega_j \sigma_{i,j}}{\sigma_P} \tag{14}$$

where ω_i in Eq. 14 is the weight of the cryptocurrency i, σ_P is the portfolio volatility, and $\sigma_{i,j}$ is the covariance between crypto i and j. The Total Risk Contribution (TRC) of the ith cryptocurrency to portfolio risk is

$$TRC_i = \sigma_{P,i} = \omega_i \frac{\partial \sigma_P}{\partial \omega_i} \tag{15}$$

Hence, the portfolio risk is computed as follows:

$$\sigma_P = \sum_{i=1}^{N} \sigma_{P,i} = \sum_{i=1}^{N} \omega_i \frac{\partial \sigma_P}{\partial \omega_i} = \sum_{i=1}^{N} TRC_i \tag{16}$$

For RPP model implementation, we apply the R package riskParityPortfolio (Cardoso and Palomar 2021; Feng and Palomar 2015).

Hierarchical Risk Parity (HRP). This model merges hierarchical clustering and portfolio allocation procedures and is based on three main steps:

- Step 1: It determines the hierarchical relationships between the assets using the recursive cluster formation scheme Hierarchical Tree Clustering algorithm. Specifically, the algorithm calculates tree clusters based on the $T \times N$ matrix of asset returns, where T represents the number of samples for a due time frame and N is the number of assets. The correlation-distance matrix D, where $\rho_{i,j}$ is the correlation between time series i and j, is as follows

$$D(i,j) = \sqrt{0.5 \times (1 - \rho(i,j))} \tag{17}$$

  and Eq. 17 is transformed in $\hat{D}$ by taking the ED between all the columns in a pairwise manner as follows

$$\hat{D}(i,j) = \sqrt{\sum_{k=1}^{N} (D(k,i) - D(k,j))^2} \tag{18}$$

- Step 2: Quasi-Diagonalization, which is a seriation algorithm that rearranges the data to show the inherent clusters more clearly. The algorithm rearranges the rows and columns of the covariance matrix of assets so that similar investments are placed together and dissimilar investments are placed apart.
- Step 3: Recursive bisection is a top-down approach to split portfolio weights between sub-groups obtained by recursively bisecting the rearranged covariance matrix from the second step based on inverse proportion to their aggregated variances.

For HRP model implementation, we apply the R functions available in https://rdrr.io/github/jackylauu/hierarchicalPortfolios/src/R/HRP.R.

Performance assesment

In addition to the different return indicators, we apply some evaluation measurement for portfolios to assess different aspects of the financial performance of the investment, namely:

- *Value-at-risk* (VaR) measures the worst expected loss over a given interval under normal market conditions at a given confidence level (the lower the better).
- *Conditional VaR* (CVaR) or Expected Short Fall (ETL) is the expected loss tail VaR and tail loss that takes the shape of the tail (the lower the better) into account.
- *Maximum drawdown* ($D_{max} or MDD$): Percent the greatest fall from peak to valley on the return series (the lower the better). Drawdowns are measured as a percentage of that maximum cumulative return.
- *Annualized sharpe ratio* (SR_{ann}) is a *reward-to-variability ratio* already defined by Eq. 4 but for benchmark between strategies taken σ_P as the standard deviation of the annualized series (the higher the better).
- *Calmar ratio* (CAL) is the annualized return over the absolute value of the maximum drawdown of an investment. It is a Sharpe-type measure that uses maximum drawdown rather than standard deviation to reflect the investor's risk:

$$CR = \frac{r_p - r_T}{D_{max}} \qquad (19)$$

    where r_T is the minimum target return that we consider equal to zero.
- *Omega ratio* (OME) is a weighted risk-return ratio for a given level of expected return set to zero in our case, which helps us to identify the chances of winning in comparison to losing (the higher the better):

$$\Omega = \frac{\frac{1}{n} \sum_{i=1}^{i=n} max(r_i - r_T, 0)}{\frac{1}{n} \sum_{i=1}^{i=n} max(r_T - r_i, 0)}. \qquad (20)$$

Results

We analyze the performance of the different strategies based on some descriptive statistics and from a financial perspective applied to different cryptocurrency spaces, the full market with 534 cryptocurrencies, and the top 250 and 175 ones according to market capitalization. Additionally, we apply different visual representations to highlight some of our findings. We complement our analysis with an exhaustive study of the outcomes of MC simulations.

Performance of monthly portfolios

In this section, we compare the investment strategies and benchmark methods by aggregating the everyday results of all the investments at day t_w during the simulation window for each method or strategy. The descriptive analysis is a standard way to illustrate the performance of asset allocation models where each method is compared with the other. In Table 3, we compute the basic descriptive statistics of the portfolio returns (Rp_t) obtained every day for each strategy and method, broken-down by market size. Figures 2, 3, and 4 represent the portfolio cumulative return (r_A). In the same

(a) Time series of daily returns (Rp) for each method

(b) Time series of cumulative return (r_A) for each method
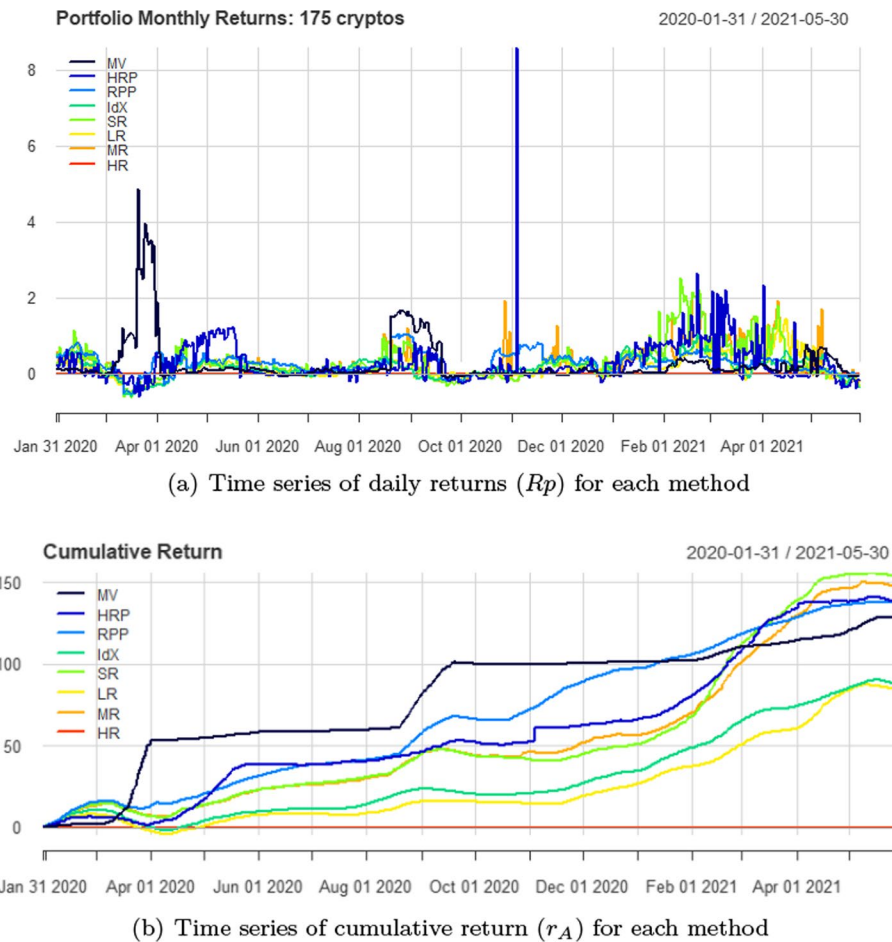
Fig. 2 Time series of the daily and cumulative returns for each method for market size 175. Mean-variance (MV), Hierarchical Risk Parity (HRP), Risk Parity Portfolio (RPP), Index CCI30 (IdX), Sharpe Ratio strategy (SR), Low-Risk strategy (LR), Mean-Risk strategy (MR) and High-Risk strategy (HR)

way, Table 4 presents the more important ratios for each model, also broken-down by market sizes. From these, we can draw the main conclusions:

- Considering the *Mean Return* values in Table 3, the SR (0.315, 0.300, and 0.931 values) and MR (0.304, 0.339, and 1.296) strategies outperform the others in terms of average values independently of the considered market size. However, as stated in the financial literature, *median return* value is considered a more representative indicator of the performance of the model when there is a heavy-tail effect on the return distribution. For *median* values, SR and MR perform slightly worse than RPP for the smaller market size (0.295).
- There are differences in the result of the benchmark approaches (MV, HRP, and RPP) depending on the market size. The impact on the performance of the models indicates that the advantage of the clustered MV model increases as the market size becomes larger. This supports our hypothesis that covariance misspecification is more likely as market size increases.

(a) Time series of daily returns (Rp) for each method

(b) Time series of cumulative return (r_A) for each method
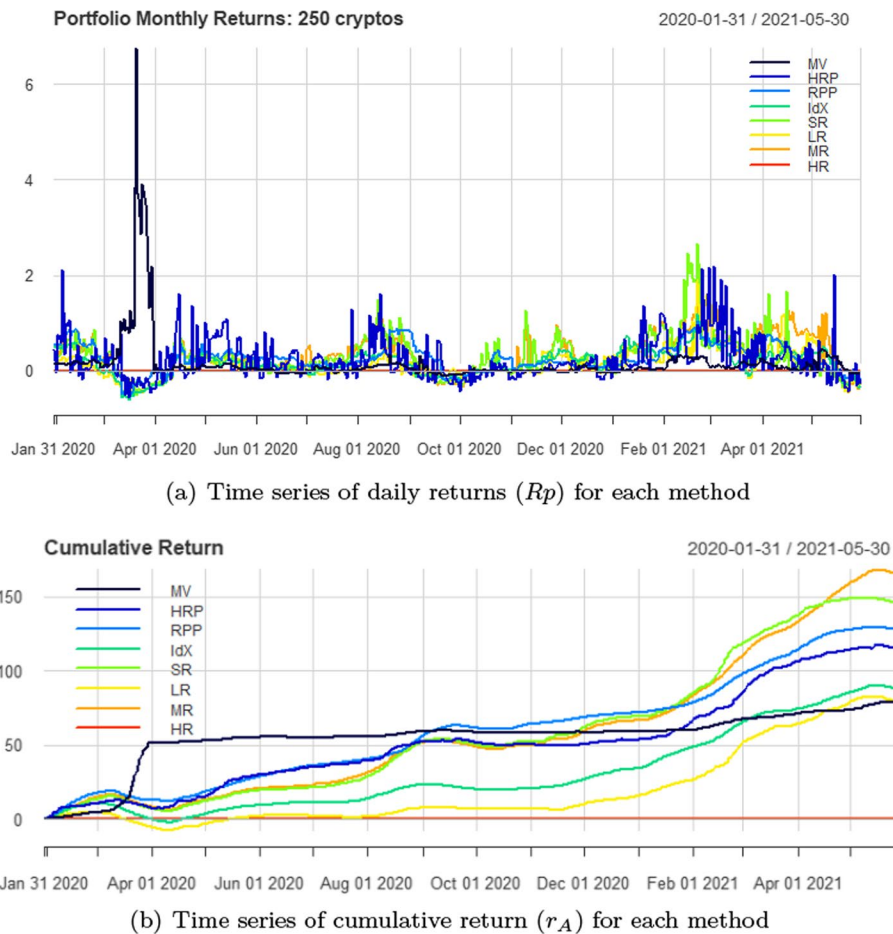
Fig. 3 Time series of the daily and cumulative returns for each method for market size 250

- In terms of the financial ratios in Table 4, we appreciate a similar result. As we increase the market size, SR and MR strategies more clearly outperform any of the benchmarks in terms of *CumRet*, *AnnRet*, and *AnnSR*, where SR strategy outperforms the others in terms of Sharpe Ratio measure (4.089), for instance.

- The HR strategy exhibits no investments (zero values on the HR columns on Tables 3 for the lower market sizes (175 and 250) and a flat red line in Figs. 2 and 3). It means that the HR strategy for the cluster prototype allocation has not found a suitable cluster in the highest risk quartiles of the market or the cardinality of the cluster is lower than 10, meaning that any cluster with fewer than 10 cryptos is not considered by any strategy). HR centroids are only selected when we consider the whole market size.

- Examining the figures of the Cumulative Returns, the MV strategy outperforms the others in the first half of the investing period for market size 175 as seen in Fig. 2 and only at the very beginning for market size 250 as shown in Fig. 3. However, in these cases, the superiority of the MV is due to one or two very profitable periods that occur on consecutive investment days. However, the MR and SR strategies have more sustained slopes, which suggests better financial behavior. In addition, the MR and SR strategies are better when considering the whole market in Fig. 4.
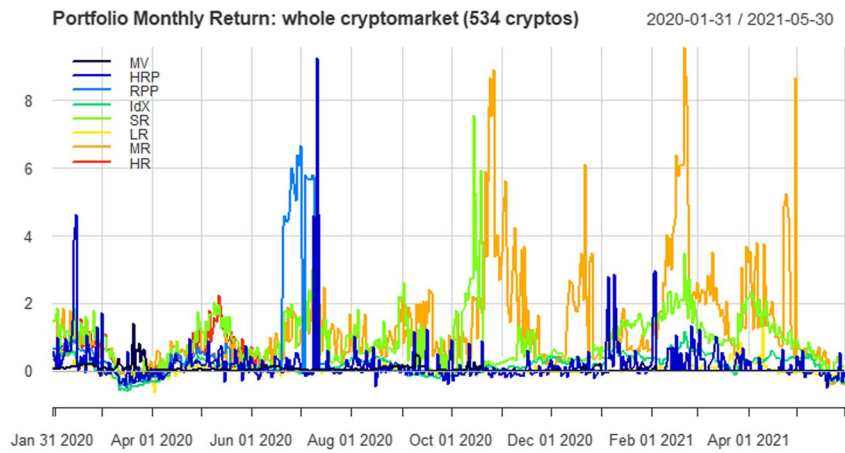
(a) Time series of daily returns (Rp) for each method
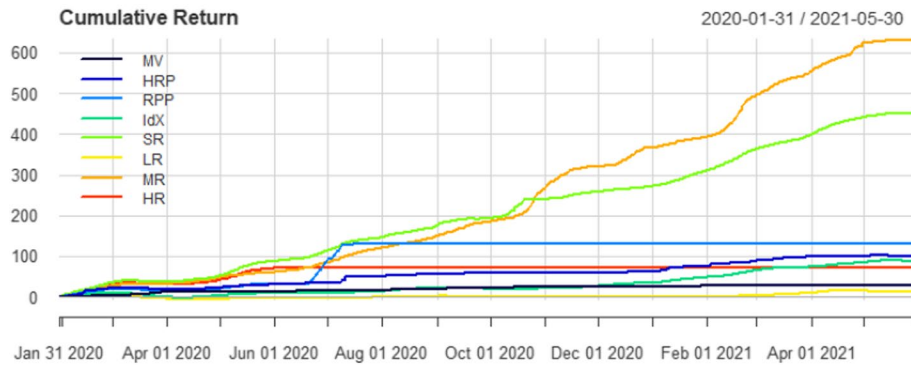
(b) Time series of cumulative return (r_A) for each method

Fig. 4 Time series of the daily and cumulative returns for each method for the whole market (534 cryptos)

Table 3 Descriptive statistics of the cumulative portfolio returns (Rp) for each strategy

MkSize	Stat	MV	HRP	RPP	IdX	SR	LR	MR	HR
175	Mean	0.265	0.283	0.283	0.180	0.315	0.172	0.304	
	Stdev	0.603	0.596	0.267	0.268	0.493	0.307	0.454	
	Minimum	− 0.101	− 0.606	− 0.550	− 0.590	− 0.619	− 0.592	− 0.564	
	Median	0.076	0.120	0.295	0.194	0.205	0.148	0.212	
	Maximum	4.878	8.589	1.065	1.185	2.521	1.470	2.147	
250	Mean	0.162	0.236	0.263	0.180	0.300	0.162	0.339	
	Stdev	0.542	0.432	0.281	0.268	0.440	0.328	0.415	
	Minimum	− 0.094	− 0.556	− 0.545	− 0.590	− 0.520	− 0.559	− 0.508	
	Median	0.055	0.123	0.223	0.194	0.233	0.084	0.303	
	Maximum	6.756	2.192	1.037	1.185	2.674	2.085	1.593	
Whole market	Mean	0.057	0.205	0.268	0.180	0.931	0.026	1.296	0.150
	Stdev	0.113	0.660	1.049	0.268	0.789	0.148	1.468	0.387
	Minimum	− 0.035	− 0.480	− 0.477	− 0.590	− 0.403	− 0.609	− 0.456	− 0.403
	Median	0.024	0.036	0.000	0.194	0.815	0.011	0.816	0.000
	Maximum	1.409	9.262	6.676	1.185	7.548	1.259	9.574	2.252

MV, mean-variance model; HRP, hierarchical risk parity; RPP, risk parity portfolio; IdX, Index CCI30; SR, sharpe ratio strategy;
LR, low-risk strategy, MR, mean-risk strategy; HR, high-risk strategy

Table 4 Financial ratios of daily returns (Rp) for each method

Mksize	FinPerf	MV	HRP	RPP	IdX	SR	LR	MR	HR
175	CumRet	128.613	137.745	137.470	87.414	153.087	83.576	147.806	
	AnnRet	96.592	103.450	103.244	65.650	114.973	62.768	111.006	
	MDD	0.035	0.682	0.269	1.078	0.555	1.346	0.527	
	ETL	− 0.070	− 0.326	− 0.264	− 0.417	− 0.367	− 0.393	− 0.357	
	VaR	− 0.053	− 0.221	− 0.128	− 0.302	− 0.246	− 0.308	− 0.242	
	AnnSR	1.521	1.648	3.673	2.327	2.214	1.941	2.321	
	CAL	0.268	0.204	0.254	0.146	0.238	0.133	0.235	
	OME	8.351	5.650	13.546	7.617	7.827	5.394	7.102	
250	CumRet	78.607	114.900	128.052	87.414	145.823	78.774	164.843	
	AnnRet	59.036	86.293	96.171	65.650	109.517	59.162	123.802	
	MDD	0.026	0.417	0.358	1.078	0.607	2.073	0.569	
	ETL	− 0.070	− 0.347	− 0.295	− 0.417	− 0.375	− 0.393	− 0.386	
	VaR	− 0.052	− 0.259	− 0.188	− 0.302	− 0.268	− 0.310	− 0.286	
	AnnSR	1.033	1.896	3.248	2.327	2.363	1.714	2.829	
	CAL	0.141	0.174	0.231	0.146	0.234	0.121	0.273	
	OME	7.186	6.195	11.466	7.617	8.244	4.867	8.739	
Whole market	CumRet	27.593	99.847	130.337	87.414	452.611	12.655	629.661	72.812
	AnnRet	20.723	74.988	97.887	65.650	339.924	9.504	472.894	54.684
	MDD	0.004	0.183	0.349	1.078	0.058	2.050	0.039	0.114
	ETL	− 0.014	− 0.305	− 0.285	− 0.417	− 0.163	− 0.308	− 0.204	− 0.188
	VaR	− 0.004	− 0.217	− 0.144	− 0.302	− 0.008	− 0.170	− 0.017	0.000
	AnnSR	1.746	1.079	0.886	2.327	4.089	0.608	3.057	1.341
	CAL	0.569	0.130	0.122	0.146	0.832	0.015	1.011	0.106
	OME	21.729	4.116	17.000	7.617	35.699	3.628	36.576	17.298

`CumRet` is cumulative return (r_A), `AnnRet` is annualized return (r_A^{ann}), `MDD` is the maximum drawdown (D_{max}), `ETL` is the Conditional VaR, `AnnSR` is the annualized Sharpe ratio (SR_{ann}), `CAL` is the Calmar ratio and `OME` is the Omega ratio

- In general, the MV strategy outperforms the others in the risk indicators (MDD, ETL, and VaR) for all market sizes. However, SR and MV present a better trade-off between risk and returns as we can see in the combined indicators (AnnSR, CAL, and OME) for the market size 250 and the whole market.
- The flat lines on the Cumulative Returns curves for MV, RPP, and HRP models for the whole market highlight the case that the allocation models are not working properly when the market size is large. The optimization algorithms probably do not converge to a feasible solution due to covariance misspecification, which causes a zero portfolio return for that holding period.

Financial comparison of the simulations

We now compare the aggregated results of the random investment paths for all the methods considered and the benchmarks in the three market sizes (175, 250, and the whole market).

In Figs. 5, 6, and 7, we present the aggregation graphically by means of the Cumulative Distribution Function of the Annualized Sharpe Ratio. In these figures, we present the distributions of the 1500 values that correspond to the realization of the simulation.
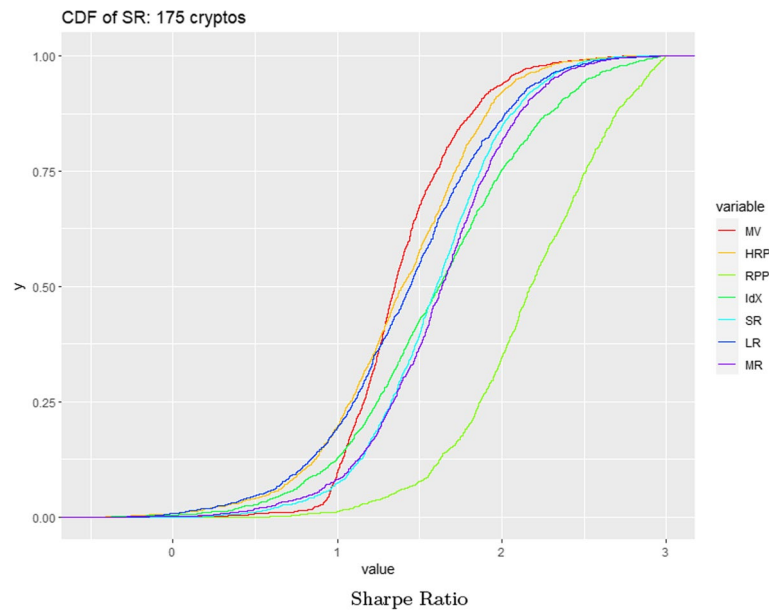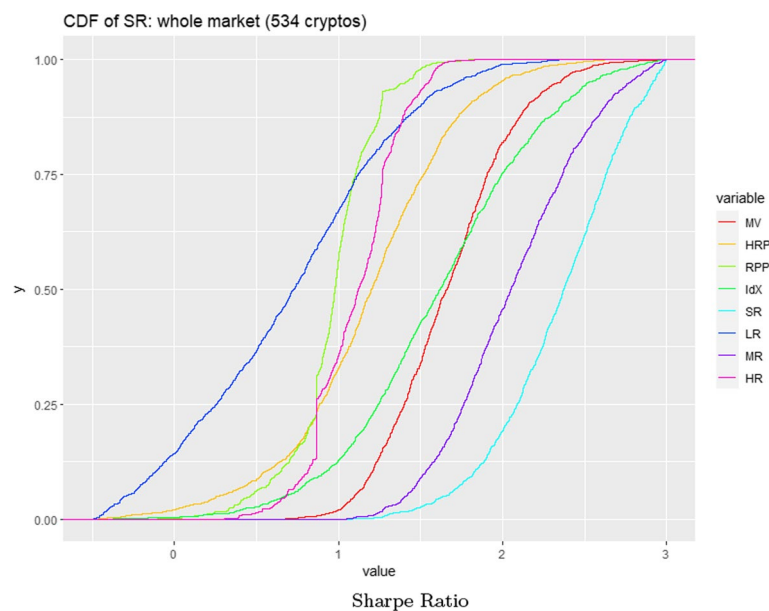
Fig. 5 Cumulative distribution of annualized sharpe ratio (SR_{ann}) of the simulations (1500 random investing paths) for 175 higher marketcap cryptos
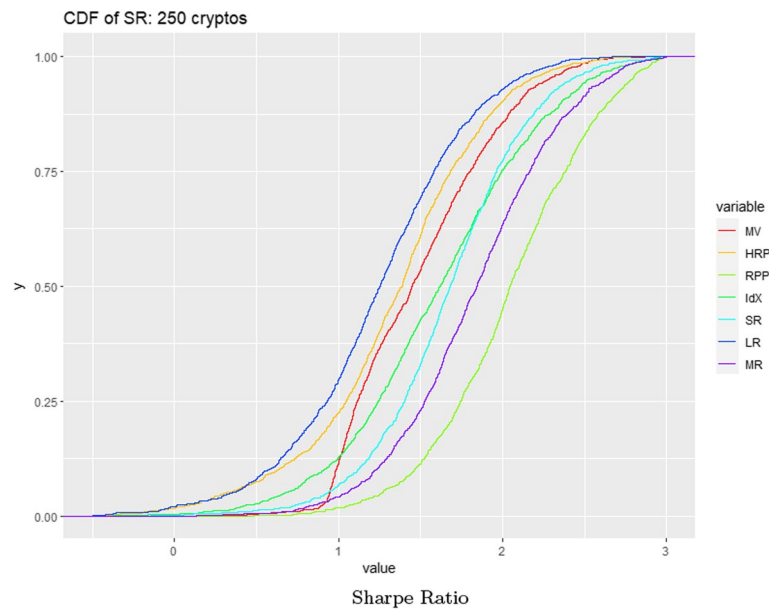
Fig. 6 Cumulative distribution of annualized sharpe ratio (SR_{ann}) of the simulations (1500 random investing paths) for 250 higher marketcap cryptos

For the smaller market size (175 cryptos), Table 5 and Fig. 5 reveal that the RPP method outperforms the rest in terms of annualized return and annualized Sharpe ratio, according to the central tendency measures followed by SR and MR strategies.

In terms of drawdown, the classical MV obtains the lowest median value with 0.043, followed by RPP value in Table 6. However, the other methods obtain similar drawdown values between 0.219 and 0.301.

Fig. 7 Cumulative distribution of annualized sharpe ratio (SR_{ann}) of the simulations (1500 random investing paths) for the whole market

In terms of ETL, MV again clearly outperforms the others with -0.039 ETL value in Table 7. LR and IdX are the riskier models (-0.274 and -0.270) with RPP again performing better than the others but worse than the MV model.

Regarding the Sharpe ratio, a measure that balances risk and returns, RPP clearly outperforms the other models with a Sharpe ratio value of 2.238 in Table 8. MR AND SR models keep a good trade-off between risk and returns with values of 1.637 and 1.601, respectively, which is better than the others except for RPP.

When the market size is increased from 175 to 250 cryptos, we observe a clear impact on the Annualized Return performance of MR that is improved up to 3.815 (median)), while that of the MV is reduced up to 0.826 (see in Table 5). For the 250 market size, MR outperform the other models, followed by SR and the RPP. We observe no relevant differences in terms of Drawdown when the market size increases from 175 to 250. Similarly, we find no relevant differences for ETL when comparing 175 and 250 market sizes (see Table 7). However, for the Sharpe ratio indicator, MV and MR models clearly improve with values 1.449 of and 1.847, respectively, in Table 8 and Fig. 6 when the market size is greater. The RPP method outperforms the others with a value of 2.062 followed by MR according to the statistics presented.

If we investigate the financial performance for the whole market size, we can see that the annualized returns of the MR and SR strategies increase dramatically, while that of the benchmarking approaches is slightly reduced. The proposed clustered methods seem to make profit from those cryptocurrencies with smaller market capitalization, while the benchmarks seem to suffer in correctly computing the estimators when the market size increases. For Annualized Returns, the MR outperforms the other models reaching 88.360 value in Table 5, followed by SR with 40.747. In terms of Drawdown, MV outperforms the others with the lowest value followed by MR, SR and HR. At this point and referring to Drawdown, we have to take the fact

Table 5 Annualized returns (r_G^{ann}) of the simulations (1500 random investing paths) for the different strategies and market sizes

MkSize	Stat	MV	HRP	RPP	IdX	SR	LR	MR	HR
175	Mean	2.808	3.023	4.106	1.779	4.051	1.580	3.883	
	Stdev	2.635	3.839	2.421	1.192	3.347	1.170	3.526	
	Minimum	− 0.020	− 0.645	0.197	− 0.309	− 0.233	− 0.432	− 0.202	
	Median	1.955	2.095	3.603	1.576	3.057	1.344	2.894	
	Maximum	16.268	78.651	16.645	9.674	27.429	8.304	49.893	
250	Mean	1.434	2.546	3.649	1.779	4.045	1.299	4.920	
	Stdev	1.400	2.984	2.113	1.192	3.568	1.108	3.948	
	Minimum	0.035	− 0.510	0.173	− 0.309	− 0.273	− 0.439	− 0.143	
	Median	0.826	1.730	3.270	1.576	3.046	1.075	3.815	
	Maximum	8.455	38.607	19.421	9.674	29.719	8.522	47.773	
Whole market	Mean	0.452	2.014	2.199	1.779	72.908	0.180	252.135	1.432
	Stdev	0.316	2.723	2.375	1.192	93.082	0.293	630.091	0.984
	Minimum	0.055	− 0.493	− 0.096	− 0.309	1.124	− 0.574	1.448	0.032
	Median	0.356	1.235	0.874	1.576	40.747	0.164	88.360	1.176
	Maximum	2.172	29.621	12.365	9.674	956.979	1.626	11283.591	8.654

Table 6 Maximum Drawdown (D_{max}) of the simulations (1500 random investing paths) for the different strategies

MkSize	Stat	MV	HRP	RPP	IdX	SR	LR	MR	HR
175	Mean	0.046	0.226	0.118	0.257	0.238	0.250	0.235	
	Stdev	0.036	0.152	0.143	0.170	0.155	0.160	0.149	
	Minimum	− 0.000	− 0.000	− 0.000	− 0.000	− 0.000	− 0.000	− 0.000	
	Median	0.043	0.219	0.073	0.281	0.239	0.301	0.243	
	Maximum	0.196	0.748	0.556	0.630	0.698	0.638	0.687	
250	Mean	0.052	0.255	0.156	0.257	0.243	0.257	0.254	
	Stdev	0.036	0.149	0.139	0.170	0.152	0.159	0.154	
	Minimum	− 0.000	− 0.000	− 0.000	− 0.000	− 0.000	− 0.000	− 0.000	
	Median	0.049	0.262	0.143	0.281	0.257	0.288	0.262	
	Maximum	0.177	0.715	0.545	0.630	0.666	0.636	0.650	
Whole market	Mean	0.008	0.220	0.133	0.257	0.065	0.206	0.071	0.074
	Stdev	0.012	0.142	0.148	0.170	0.108	0.146	0.132	0.114
	Minimum	− 0.000	− 0.000	− 0.000	− 0.000	− 0.000	− 0.000	− 0.000	− 0.000
	Median	0.000	0.207	0.042	0.281	0.000	0.172	0.000	0.000
	Maximum	0.048	0.796	0.477	0.630	0.403	0.710	0.469	0.403

that for many iterations of MV and RPP, models are not able to converge on the due training window into account; the outcome of the model is zero and that explains the medium value zero. The worst performance for the Drawdown corresponds to IdX, HRP, and LR with 0.281, 0.207, and 0.172 of median values, respectively. The riskier model according to ETL is the IdX benchmark and the riskless corresponds to MV model with a value of 0.0 in Table 7. In terms of Sharpe ratio, the SR strategy again outperforms the others with a median value of 2.469 in Table 8 followed by MR with a value of 2.078. The worst Sharpe ratio performance corresponds to LR and RPP with values of 0.680 and 0.979, respectively.

Table 7 Conditional VaR (CVaR or ETL) of the simulations (1500 random investing paths) for the different strategies

MkSize	Stat	MV	HRP	RPP	IdX	SR	LR	MR	HR
175	Mean	− 0.041	− 0.207	− 0.117	− 0.251	− 0.224	− 0.237	− 0.221	
	Stdev	0.029	0.140	0.142	0.168	0.150	0.156	0.144	
	Minimum	− 0.101	− 0.606	− 0.550	− 0.590	− 0.619	− 0.592	− 0.564	
	Median	− 0.039	− 0.200	− 0.073	− 0.270	− 0.215	− 0.274	− 0.215	
	Maximum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
250	Mean	− 0.043	− 0.228	− 0.155	− 0.251	− 0.239	− 0.242	− 0.249	
	Stdev	0.028	0.133	0.138	0.168	0.148	0.154	0.151	
	Minimum	− 0.094	− 0.556	− 0.545	− 0.590	− 0.520	− 0.559	− 0.508	
	Median	− 0.042	− 0.219	− 0.143	− 0.270	− 0.257	− 0.272	− 0.258	
	Maximum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
Whole market	Mean	− 0.007	− 0.182	− 0.133	− 0.251	− 0.065	− 0.182	− 0.071	− 0.074
	Stdev	0.010	0.115	0.148	0.168	0.108	0.137	0.132	0.114
	Minimum	− 0.035	− 0.480	− 0.477	− 0.590	− 0.403	− 0.609	− 0.456	− 0.403
	Median	0.000	− 0.164	− 0.042	− 0.270	0.000	− 0.161	0.000	0.000
	Maximum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 8 Annualized sharpe ratio (SR_{ann}) of the simulations (1500 random investing paths) for the different strategies

MkSize	Stat	MV	HRP	RPP	IdX	SR	LR	MR	HR
175	Mean	1.398	1.370	2.264	1.624	1.594	1.424	1.617	
	Stdev	0.351	0.510	0.548	0.572	0.418	0.530	0.451	
	Minimum	− 0.103	− 1.991	0.590	− 0.378	0.180	− 0.842	− 0.282	
	Median	1.349	1.389	2.238	1.623	1.601	1.440	1.637	
	Maximum	2.758	2.793	4.638	3.408	2.836	3.370	3.226	
250	Mean	1.496	1.327	2.082	1.624	1.684	1.232	1.850	
	Stdev	0.441	0.570	0.494	0.572	0.468	0.537	0.493	
	Mininimum	0.287	− 1.731	0.483	− 0.378	− 0.146	− 1.134	0.131	
	Median	1.449	1.379	2.062	1.623	1.681	1.251	1.847	
	Maximum	3.489	3.524	4.033	3.408	3.712	2.752	3.590	
Whole market	Mean	1.672	1.163	0.965	1.624	2.488	0.628	2.116	1.105
	Stdev	0.376	0.554	0.262	0.572	0.508	0.675	0.467	0.260
	Minimum	0.678	− 1.859	− 0.101	− 0.378	1.085	− 1.527	1.051	0.303
	Median	1.662	1.194	0.979	1.623	2.469	0.680	2.078	1.121
	Maximum	3.168	3.131	1.800	3.408	4.340	2.395	3.553	1.865

In general, independently of the method or strategy, we observe a change in the performance ratios when we compare 175/250 market sizes with the whole market in Tables 5, 6, 7, and 8. The portfolio models behave depending on the market size and we see that the standard methods misbehave for the higher market size where we find room for our proposal. If we exclusively take simulation results for SR_{ann} for the whole market represented in Fig. 7, we find that the SR strategy outperforms the others followed by MR.

In terms of the centroid strategies, we observe that MR and SR strategies over-perform compared to the others (LR, HR). Thinking exclusively in terms of returns, MR over-performs compared to SR for similar Max. Drawdown and CVaR. However, in

terms of the Annualized Sharpe Ratio, the SR strategy (median value 2.469) beats all the other strategies and models.

For the MR strategy, via the simulations, we confirm the same results as for the monthly portfolios, that is, a lack of centroids for lower market sizes so no simulations either. In general HR when centroids exist perform better than LR.

Conclusion

We propose a methodology that, combined with the assumptions required for a good performance of MV models, allows us to extend the use of portfolio optimization models to the cryptocurrency market regardless of its size and volatility. Its usefulness extends to any portfolio management model that requires a covariance-based measure of market risk, and it is particularly suitable for managing streaming market data flows.

Our methodology proposes a clustering stage to reduce the problem of the dimensionality. It reveals how the performance of the model can be improved by reducing the number of assets considered and focusing on those that best fit the investor criteria. The results are similar to those obtained by an appropriate feature selection in prediction problems (Kou et al. 2021). Clustering reduces the space where the optimization models work, creating more accurate estimations of the different factors and mitigating possible errors. This methodology can be applied to other financial markets and other portfolio optimization models. It is specially indicated when both a large number of assets and a long time window are considered.

Based on the results, we draw the following conclusions:

- First, we present a smart way to use the prototypes from clusters to automatize the selection of the more suitable partitioning of the market. The proposed methodology works dynamically with streaming price data of cryptos that, in our case, change on a daily basis although it can be easily adapted to other data periodicities. In this way, market partitions and portfolio generations work in concurrent mode autonomously once we set the criteria for cluster pre-selection based on the risk-aversion profiles of the investors.
- The range of performance values when applied to cryptocurrency portfolios exceeds any comparison with traditional markets and this becomes more evident as the size of the market itself grows. Cumulative returns, risk, and drawdowns are significantly higher in cryptocurrency markets. For example, the steep upward trend of the cumulative yield curves means it is not comparable to any growth in traditional financial markets.
- We demonstrate that the performance of the standard model Mean-Variance (MV) applied to the whole market with no partitions is not very different in magnitude to other results derived by other research (Petukhina et al. 2021; Liu 2019; Culjak et al. 2022) which make us confident regarding the results.
  In general,
- Sharpe ratio (Strategy 1) and Mean-Risk (Strategy 2b) outperform all the other strategies in terms of Cumulative, Annualized Returns and Annualized Sharpe Ratio as the market size increases. At this point, we have to discard High-Risk (Strategy 2c) for standard investors as most of the time, there are no centroids available on the

higher rank of volatility, which means this strategy has low suitability for the investor that frequently needs to take a position on the market.

- We observe that strategies based on extreme values centroids (LR and HR) underperform the others along the out-of-sample window. In other words, centroids allocated into the interquartile distribution for estimation windows behave much better along the holding periods. The explanation is aligned with the 2nd and 3rd MV assumptions , which means that in general, risk-based models will perform better when we choose cryptos allocated in the mean region closer to the center of a normal distribution.
- One of the drawbacks of the proposed strategies is the higher drawdown and risk compared with the classical MV, which is the more evident weakness of the proposed strategies, so it is comparable to the HRP model. We consider that SR and MR strategies independent of market size offer a good trade-off between returns, risk, and drawdown.

As we demonstrate, the results are sensitive to the number of cryptocurrencies considered. The smaller the space of cryptocurrencies with more restrictive thresholds during data pre-processing, the higher the chances to exclude the cryptocurrencies with explosive behavior, and the financial performance indicators that are obtained in these cases are more similar to those of *traditional* markets.

Finally, we have to consider that not all 534 cryptocurrencies considered in this research can be directly traded on the market. Depending on the selected exchange, some can be traded and for others, we should find on other exchanges; for instance, we can trade up to 124 crypto assets (we use the terms crypto asset and cryptocurrency interchangeably) in Coinbase[6] or 380 crypto assets in Binance[7] In other cases, some cryptocurrencies could require some days to reach a consensus[8] before incorporating to the portfolio, which introduces additional frictions on the market not considered in our model. *Liquidity pools* bring a solution to such frictions in creating a decentralized finance (DeFi) facilitating the turning of assets into cash and vice versa by application of smart contracts. The counterpart is an increase in the transaction complexity and associated risk together with the high turnover of investors providing that liquidity.[9] In addition, there are different liquidity approaches depending, for instance, on whether there are centralized or decentralized exchanges with different spread mechanisms that impact the performance of the trading models. Further work could consider adding a liquidity criterion to the Prototype Selection Strategy stage in Fig. 1. For instance, this could be based on a spread measure together with the risk criteria that we have already used in the cryptocurrency pre-selection stage to improve the performance of the Mean-Variance portfolios considering the conditions of crypto markets in real-life.

---

[6] https://www.coinbase.com/price/s/listed.

[7] https://www.binance.com/en.

[8] https://blog.nano.org/cryptocurrency-fee-comparison-which-crypto-has-the-lowest-fees-4e9118590e1f.

[9] https://www.nansen.ai/research/all-hail-masterchef-analysing-yield-farming-activity?roistat_visit=1018855.

Abbreviations

A-FCM	Autocorrelation-based fuzzy C-means
CAGR	Compound annual growth rate
CAL	Calmar ratio
CLARA	Clustering large applications
CNN	Convolutional neural network
CVaR	Conditioned vale at risk
CVI	Cluster Validity Index
DBHT	Directed bubble hierarchical tree
ED	Euclidean distance
EMH	Efficient market hypothesis
ERC	Equally-weighted risk contribution
ETL	Expected short fall
EW	Equally-weighted
GARCH	Generalized autoregressive conditional heteroskedasticity
HCAA	Hierarchical clustering based asset allocation
HR	High risk strategy, risk-seeking investor strategy
HRP	Hierarchical risk parity
LR	Low risk strategy, risk-averse investor strategy
MAR	Moving average reversion
MDD	Maximum draw down
MPT	Modern portfolio theory
MR	Mean risk strategy, risk-neutral investor strategy
MST	Minimum spanning tree
MV	Mean-variance
MVO	Mean-variance optimization
OLPS	Online portfolio selection model
OLMAR	Online portfolio selection model average reversion
PAM	Partition around medoids
PMFG	Planar maximally filtered graph
QP	Quadratic program
RPP	Risk parity portfolio
RIP	Random investment path
SOM	Self-organizing maps
SR	Sharpe ratio or sharpe ratio strategy
VaR	Value at risk

Declarations

Competing interest
The authors declare that they have no competing interests.

References
Alessandretti L, ElBahrawy A, Aiello LM, Baronchelli A (2018) Anticipating cryptocurrency prices using machine learning. Complexity

Apopo N, Phiri A (2021) On the (in) efficiency of cryptocurrencies: Have they taken daily or weekly random walks? Heliyon 7(4):e06685

Arbelaitz O, Gurrutxaga I, Muguerza J, Pérez JM, Perona I (2013) An extensive comparative study of cluster validity indices. Pattern Recogn 46:243–256

Bonanno G, Lillo F, Mantegna RN (2001) High-frequency cross-correlation in a set of stocks

Bonanno G, Caldarelli G, Lillo F (2004) Networks of equities in financial markets. Eur Phys J B Condens Matter 38(2):363–371. https://doi.org/10.1140/epjb/e2004-00129-6

Borysov P, Hannig J, Marron J (2014) Asymptotics of hierarchical clustering for growing dimension. J Multivar Anal 124:465–479

Breiman L (1960) Investment policies for expanding businesses optimal in a long-run sense. Naval Res Logist Q 7(4):647–651

Brida J, Risso W (2009) Dynamics and structure of the 30 largest north American companies. Soc Comput Econ 35(1):85–99

Burggraf T (2019) Risk-based portfolio optimization in the cryptocurrency world. Inf Syst Econ J. https://ssrn.com/abstract=3454764

Cardoso P (2021) riskParityPortfolio: design of risk parity portfolios. https://CRAN.R-project.org/package=riskParityPortfolio, r package version 0.2.2

Cerqueti R, Giacalone M, Mattera R (2021) Model-based fuzzy time series clustering of conditional higher moments. Int J Approx Reason 134:34–52. https://doi.org/10.1016/j.ijar.2021.03.011

Cerqueti R, D'Urso P, De Giovanni L, Giacalone M, Mattera R (2022) Weighted score-driven fuzzy clustering of time series with a financial application. Expert Syst Appl 198(116):752. https://doi.org/10.1016/j.eswa.2022.116752

Chuen DLK, Guo LM, Wang Y (2017) Cryptocurrency: a new investment opportunity?

Cornuejols G, Tütüncü R (2006) Optimization methods in finance, vol 5. Cambridge University Press

Culjak M, Tomić B, Žiković S (2022) Benefits of sectoral cryptocurrency portfolio optimization. Res Int Bus Financ 60(101):615. https://doi.org/10.1016/j.ribaf.2022.101615

Dai M, Xu ZQ, Zhou XY (2010) Continuous-time Markowitz's model with transaction costs. SIAM J Financ Math 1(1):96–125. https://doi.org/10.1137/080742889

Datta T, Ghosh I (2015) Using clustering method to understand Indian stock market volatility. Commun Appl Electron 2(6):35–44. https://doi.org/10.5120/cae2015651793

De Prado ML (2016) Building diversified portfolios that outperform out of sample. J Portf Manag 42(4):59–69

Duarte FG, De Castro LN (2020) A framework to perform asset allocation based on partitional clustering. IEEE Access 8(110775–110):88. https://doi.org/10.1109/ACCESS.2020.3001944

D'Urso P, Cappelli C, Di Lallo D, Massari R (2013) Clustering of financial time series. Physica A Stat Mech Appl 392(9):2114–2129

D'Urso P, De Giovanni L, Massari R (2016) Garch-based robust clustering of time series. Fuzzy Sets Syst 305:1–28. https://doi.org/10.1016/j.fss.2016.01.010

D'Urso P, Giovanni LD, Massari R, D'Ecclesia RL, Maharaj EA (2020) Cepstral-based clustering of financial time series. Expert Syst Appl 161(113):705. https://doi.org/10.1016/j.eswa.2020.113705

Eisl A, Gasser SM, Weinmayer K (2015) Caveat emptor: Does bitcoin improve portfolio diversification? Available at SSRN 2408997

Fama EF (1965) The behavior of stock-market prices. J Bus 38(1):34–105

Fang F, Chung W, Ventre C, Basios M, Kanthan L, Li L, Wu F (2021) Ascertaining price formation in cryptocurrency markets with machine learning. Eur J Finance. https://doi.org/10.1080/1351847X.2021.1908390

Fang F, Ventre C, Basios M, Kanthan L, Martinez-Rego D, Wu F, Li L (2022) Cryptocurrency trading: a comprehensive survey. Financ Innov 8(1):1–59

Feng Y, Palomar DP (2015) Scrip: successive convex optimization methods for risk parity portfolio design. IEEE Trans Signal Process 63(19):5285–5300. https://doi.org/10.1109/TSP.2015.2452219

Finkelstein M, Whitley R (1981) Optimal strategies for repeated games. Adv Appl Probab 13(2):415–428

Gatheral J (2008) Random matrix theory and covariance estimation. New York

Gilli M, Maringer D, Schumann E (2019) Numerical methods and optimization in finance. Academic Press, New York

Goetzmann WN, Brown SJ, Gruber MJ, Elton EJ (2014) Modern portfolio theory and investment analysis. Wiley, New York

Goldfarb D, Idnani AU (1983) A numerically stable dual method for solving strictly convex quadratic programs. Math Program 27:1–33

Gubu L, Rosadi D, Abdurakhman YB (2020) Robust mean variance portfolio selection using cluster analysis: a comparison between Kamila and weighted K-mean clustering. Asian Econ Financ Rev 10(10):1169–1186

Guris B (2021) NonlinearTSA: nonlinear time series analysis. R package version 3.5.0

Henning C, Meila M, Murtagh F, Rocci R (2016) Handbook of cluster analysis. CRC Press, Boca Raton

Iorio C, Frasso G, D'Ambrosio A, Siciliano R (2018) A p-spline based clustering approach for portfolio selection. Expert Syst Appl 95:88–103. https://doi.org/10.1016/j.eswa.2017.11.031

Jiang Z, Liang J (2017) Cryptocurrency portfolio management with deep reinforcement learning. In: 2017 Intelligent systems conference (IntelliSys), pp 905–913. https://doi.org/10.1109/IntelliSys.2017.8324237

Kalayci CB, Ertenlice O, Akbay MA (2019) A comprehensive review of deterministic models and applications for mean-variance portfolio optimization. Expert Syst Appl 125:345–368. https://doi.org/10.1016/j.eswa.2019.02.011

Kapetanios G, Shin Y, Snell A (2003) Testing for a unit root in the nonlinear star framework. J Econom 112(2):359–379

Kaufman L (1986) Clustering large data sets. Pattern Recognit Practice. https://doi.org/10.1016/B978-0-444-87877-9.50039-X

Kaufman L, Rousseeuw P (1990) Finding groups in data: an introduction to cluster analysis. https://doi.org/10.2307/2532178

Kelly J (1956) A new interpretation of information rate. Bell Syst Tech J

Khedmati M, Azin P (2020) An online portfolio selection algorithm using clustering approaches and considering transaction costs. Expert Syst Appl 159(113):546. https://doi.org/10.1016/j.eswa.2020.113546

Kolm PN, Tütüncü R, Fabozzi FJ (2014) 60 Years of portfolio optimization: practical challenges and current trends. Eur J Oper Res 234(2):356–371. https://doi.org/10.1016/j.ejor.2013.10.06

Kou G, Peng Y, Wang G (2014) Evaluation of clustering algorithms for financial risk analysis using MCDM methods. Inf Sci 275:1–12. https://doi.org/10.1016/j.ins.2014.02.137

Kou G, Xu Y, Peng Y, Shen F, Chen Y, Chang K, Kou S (2021) Bankruptcy prediction for SMES using transactional data and two-stage multiobjective feature selection. Decis Support Syst 140(113):429. https://doi.org/10.1016/j.dss.2020.113429

Kyriazis NA (2019) A survey on efficiency and profitable trading opportunities in cryptocurrency markets. J Risk Financ Manag. https://doi.org/10.3390/jrfm12020067

Ledoit O, Wolf M (2003) Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. J Empir Financ 10:603–621

Ledoit O, Wolf M (2004) Honey, i shrunk the sample covariance matrix. J Portf Manag 30(4):110–119

Leland HE (1999) Beyond mean-variance: performance measurement in a nonsymmetrical world. Financ Anal J 55(1):27–36

Li B, Hoi SCH (2012) On-line portfolio selection with moving average reversion. In: Proceedings of the 29th international conference on international conference on machine learning, Omnipress, Madison, WI, USA, ICML'12, pp 563–570

Li B, Hoi SCH (2014) Online portfolio selection: a survey. ACM Comput Surv. https://doi.org/10.1145/2512962

Li B, Hoi SC, Sahoo D, Liu ZY (2015) Moving average reversion strategy for on-line portfolio selection. Artif Intell 222:104–123

Li B, Sahoo D, Hoi SC (2016) Olps: a toolbox for on-line portfolio selection. J Mach Learn Res 17(1):1242–1246

Li D, Ng WL (2000) Optimal dynamic portfolio selection: multiperiod mean-variance formulation. Math Financ 10(3):387–406

Li T, Kou G, Peng Y, Yu PS (2021) An integrated cluster detection, optimization, and interpretation approach for financial data. IEEE Trans Cyberne. https://doi.org/10.1109/TCYB.2021.3109066

Lintner J (1965) The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Rev Econom Stat 47(1):13–37

Liu W (2019) Portfolio diversification across cryptocurrencies. Financ Res Lett 29:200–205

Livieris IE, Pintelas E, Stavroyiannis S, Pintelas P (2020) Ensemble deep learning models for forecasting cryptocurrency time-series. Algorithms. https://doi.org/10.3390/a13050121

Lohre H, Rother C, Schäfer KA (2020) Hierarchical risk parity: accounting for tail dependencies in multi-asset multi-factor allocations. New developments and financial applications, machine learning for asset management, pp 329–368

Lorenzo L, Arroyo J (2022) Analysis of the cryptocurrency market using different prototype-based clustering techniques. Financ Innov. https://doi.org/10.1186/s40854-021-00310-9

Luca GD, Zuccolotto P (2017) Dynamic tail dependence clustering of financial time series. Stat Pap 58(3):641–657

Maillard S, Roncalli T, Teïletche J (2010) The properties of equally weighted risk contribution portfolios. J Portf Manag 36(4):60–70

Makarov I, Schoar A (2020) Trading and arbitrage in cryptocurrency markets. J Financ Econ 135(2):293–319

Mantegna R (1999) Hierarchical structure in financial markets. Eur Phys J B 11(1):193–197

Marcus G (2018) Deep learning: a critical appraisal. CoRR abs/1801.00631. arXiv:1801.00631

Markowitz H (1952a) Portfolio selection*. J Financ 7(1):77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x

Markowitz H (1952b) The utility of wealth. J Polit Econ 60. https://EconPapers.repec.org/RePEc:ucp:jpolec:v:60:y:1952:p:151

Markowitz HM (1959) Portfolio selection. Yale University Press, Yale

Marti G, Nielsen F, Bi'nkowski M, Donnat P (2017) A review of two decades of correlations, hierarchies, networks and clustering in financial markets. Papers 1703.00485. arXiv:1703.00485

Mattera R, Giacalone M, Gibert K (2021) Distribution-based entropy weighting clustering of skewed and heavy tailed time series. Symmetry 13(6):959

Michaud RO (1989) The Markowitz optimization enigma: Is 'optimized' optimal? Financ Anal J 45(1):31–42. https://doi.org/10.2469/faj.v45.n1.31

Molyboga M (2020) A modified hierarchical risk parity framework for portfolio management. J Financ Data Sci 2(3):128–139

Musmeci N, Aste T, di Matteo T (2014) Clustering and hierarchy of financial markets data: advantages of the DBHT. arXiv:q-fin.ST/1406.0496v1

Nanda S, Mahanty B, Tiwari M (2010) Clustering Indian stock market data for portfolio management. Expert Syst Appl 37:8793–8798. https://doi.org/10.1016/j.eswa.2010.06.026

Newman M (2005) Power laws, pareto distributions and zipf's law. Contemp Phys 46(5):323–351. https://doi.org/10.1080/00107510500052444

Nguyen Cong L, Wisitpongphan N, Meesad P, Unger H (2014) Clustering stock data for multi-objective portfolio optimization. Int J Comput Intell Appl. https://doi.org/10.1142/S1469026814500114

Onnela JP, Chakraborti A, Kaski K, Kertész J, Kanto A (2003) Dynamics of market correlations: taxonomy and portfolio analysis. Phys Rev E. https://doi.org/10.1103/physreve.68.056110

Otranto E (2008) Clustering heteroskedastic time series by model-based procedures. Comput Stat Data Anal 52(10):4685–4698. https://doi.org/10.1016/j.csda.2008.03.020

Pafka S, Kondor I (2003) Noisy covariance matrices and portfolio optimization ii. Physica A Stat Mech Appl 319:487–494

Palamalai S, Kumar KK, Maity B (2021) Testing the random walk hypothesis for leading cryptocurrencies. Borsa Istanbul Rev 21(3):256–268

Petukhina A, Trimborn S, Härdle WK, Elendner H (2021) Investing with cryptocurrencies: evaluating their potential for portfolio allocation strategies. Quant Finance 21(11):1825–1853. https://doi.org/10.1080/14697688.2021.1880023

Piccolo D (1990) A distance measure for classifying Arima models. J Time Ser Anal 11(2):153–164

de Prado L (2016) Building diversified portfolios that outperform out of sample. J Portf Manag 42(4):59–69. https://doi.org/10.3905/jpm.2016.42.4.059

Qian EE (2016) Risk parity fundamentals. CRC Press, Boca Raton

Raffinot T (2017) Hierarchical clustering-based asset allocation. J Portf Manag 44(2):89–99

Raffinot T (2018) The hierarchical equal risk contribution portfolio. Available at SSRN 3237540

Roncalli T (2013) Introduction to risk parity and budgeting. CRC Press, Boca Raton

Samuelson PA (1965) Proof that properly anticipated prices fluctuate randomly. Manag Rev 6(2)

Schubert E, Rousseeuw PJ (2019) Faster k-medoids clustering: improving the pam, CLARA, and CLARANS algorithms. In: Amato G, Gennaro C, Oria V, Radovanović M (eds) Similarity search and applications. Springer International Publishing, Cham, pp 171–187

Sebastião H, Godinho P (2021) Forecasting and trading cryptocurrencies with machine learning under changing market conditions. Financ Innov 7(1):1–30

Sharpe W (1964) Capital asset prices: a theory of market equilibrium under conditions of risk. J Finance 19(3):425–442

Soleymani F, Vasighi M (2020) Efficient portfolio construction by means of CVAR and k-means++ clustering analysis: evidence from the NYSE. Int J Finance Econ. https://doi.org/10.1002/ijfe.2344

Song JY, Chang W, Song JW (2019) Cluster analysis on the structure of the cryptocurrency market via bitcoin-ethereum filtering. Physica A 527(121):339

Song WM, di Matteo T, Aste T (2012) Hierarchical information clustering by means of topologically embedded graphs. PLoS ONE

Steinbach MC (2001) Markowitz revisited: mean-variance models in financial portfolio analysis. SIAM Rev 43(1):31–85

Stosic D, Stosic D, Ludermir TB, Stosic T (2018) Collective behavior of cryptocurrency price changes. Physica A 507:499–509

Thorp EO (1975) Portfolio choice and the Kelly criterion. In: Stochastic optimization models in finance, Elsevier, pp 599–619

Tumminello M, Di Matteo T, Aste T, Mantegna RN (2006) Correlation based networks of equity returns sampled at different time horizons. Eur Phys J B 55(2):209–217. https://doi.org/10.1140/epjb/e2006-00414-4

Umino K, Kikuchi T, Kunigami M, Yamada T, Terano T (2022) Empirical analyses of Olmar method for financial portfolio selection in stock markets. J Adv Comput Intell Inform 26(4):451–460

Venables WN, Ripley BD (2002) Modern applied statistics with S, 4th edn. Springer, New York. https://www.stats.ox.ac.uk/pub/MASS4/. ISBN 0-387-95457-0

Wang M, Li C, Xue H, Xu F (2014) A new portfolio rebalancing model with transaction costs. J Appl Math

Watorek M, Drozdz S, Kwapien J, Minati L, Oswiecimka P, Stanuszek M (2020) Multiscale characteristics of the emerging global cryptocurrency market. Phys Rep. https://doi.org/10.1016/j.physrep.2020.10.005

Xu M, Chen X, Kou G (2019) A systematic review of blockchain. Financ Innov. https://doi.org/10.1186/s40854-019-0147-z

Zhao JL, Fan S, Yan J (2016) Overview of business innovations and research opportunities in blockchain and introduction to the special issue. Financ Innov 2(1):1–7. https://doi.org/10.1186/s40854-016-0049-2

Zhou L, Zhang L, Zhao Y, Zheng R, Song K (2021) A scientometric review of blockchain research. IseB 19(3):757–787

Publisher's Note