

Article

Clustering and Modelling of the Top 30 Cryptocurrency Prices Using Dynamic Time Warping and Machine Learning Methods

Tomáš Šťastný , Jiří Koudelka , Diana Bílková  and Luboš Marek 

Faculty of Informatics and Statistics, Prague University of Economics and Business, W. Churchill Sq. 1938/4, 13067 Prague, Czech Republic

* Correspondence: xstat07@vse.cz (T.Š.); kouj04@vse.cz (J.K.); diana.bilkova@vse.cz (D.B.); marek@vse.cz (L.M.)

Abstract: Cryptocurrencies are a new field of investment opportunities that has experienced a significant growth in the last decade. The crypto market was capitalized at more than USD 3000 bn, having grown from USD 10 m over the period 2011–2021. Generating high returns, investments in cryptocurrencies have also shown high levels of price volatility. By comparing the performance of cryptocurrencies (measured by the crypto index) and standard equities (included in the S&P 500 index), we found that the former has outperformed the latter 14 times over the last two years. In the present paper, we analyzed the 2012–2022 global crypto market developments and main constituents. With a focus on the top 30 cryptocurrencies and their prices, as of 9 April 2022, covering data of the two major market stress events—outbreaks of the COVID-19 pandemic (February 2020) and the Russian invasion of Ukraine (February 2022). We applied the dynamic time warping method including barycentre averaging and k-Shape clustering of time series. The use of the dynamic time warping has been essential for the preparation of data for subsequent clustering and forecasting. In addition, we compared performance of cryptocurrencies and equities. Cryptocurrency time series are rather short, sometimes involving high levels of volatility and including multiple data gaps, whereas equity time series are much longer and well-established. Identifying similarities between them allows analysts to predict crypto prices by considering the evolution of similar equity instruments and their responses to historical events and stress periods. Moreover, we tested various forecasting methods on the 30 cryptocurrencies to compare traditional econometric methods with machine learning approaches.

Keywords: cryptocurrency; dynamic time warping; machine learning; cluster analysis; ARIMA**MSC:** 62M10; 62H30; 68T07**JEL Classification:** C55; G17; C32; C38

Citation: Šťastný, T.; Koudelka, J.; Bílková, D.; Marek, L. Clustering and Modelling of the Top 30 Cryptocurrency Prices Using Dynamic Time Warping and Machine Learning Methods. *Mathematics* **2022**, *10*, 3672. <https://doi.org/10.3390/math10193672>

Academic Editor: Jan Lansky

Received: 18 August 2022

Accepted: 21 September 2022

Published: 7 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction and Paper Objectives

A cryptocurrency is a kind of digital currency which is based on cryptographic proofs that are required for confirmation of each transaction. Cryptocurrencies may be described by a special combination of properties: they are independent from central authorities (e.g., from central banks), they ensure some level of pseudo-anonymity, and they possess a double spending attack protection [1].

Cryptocurrencies belong to decentralized digital currencies where this decentralization is implemented by the p2p architecture. For many types of cryptocurrencies, their new units are often created as a form of reward for solving complex mathematical problems. And usually these problems get more complex and thus more demanding, the more coins of a certain cryptocurrency already exist (e.g., Bitcoin). The speed of these new coins generation is defined for each cryptocurrency upon its creation, but there can be some exemptions [1].

Under normal circumstances, cryptocurrencies ensure some level of pseudo-anonymity if their users follow some basic rules and don't disclose their ownership to public. All transactions are stored as blocks in a big decentralized database called a blockchain. This blockchain database contains all the details about all the transactions, but without the names of counterparties. The cryptocurrencies that are stored in the blockchains are independent from monetary decisions of central banks and public authorities. The generation system, as mentioned, is defined when they are created and remains stable during their life. Some cryptocurrencies provide a rare option for its change given a consensus that is reached by some pre-defined majority of their operators (e.g., 75–95%). However, the cryptocurrencies are still a rather new area of financial instruments. Future years and their long-term usage show more lessons that are learned from their behaviour. The third important feature of the cryptocurrencies is the double spending attack protection. An owner of a certain cryptocurrency (and its coin) cannot use this coin for payment to two or more different counterparties. Each coin can be used just once [1].

Cryptocurrencies are a new area of investment opportunities that has grown significantly over the last decade [2]. The noticeable growth of cryptocurrencies and their prices has been recorded since 2017, accelerating in last two years (see the Figure 1). Investments into cryptocurrencies could and can generate high returns, but they have also shown a high level of price volatility, as discussed below. Cryptocurrencies undoubtedly represent an alternative type of investment opportunities for potential investors with a significant risk appetite [3].

Market capitalization of the global cryptocurrencies market 2015–2022 (in billion USD)

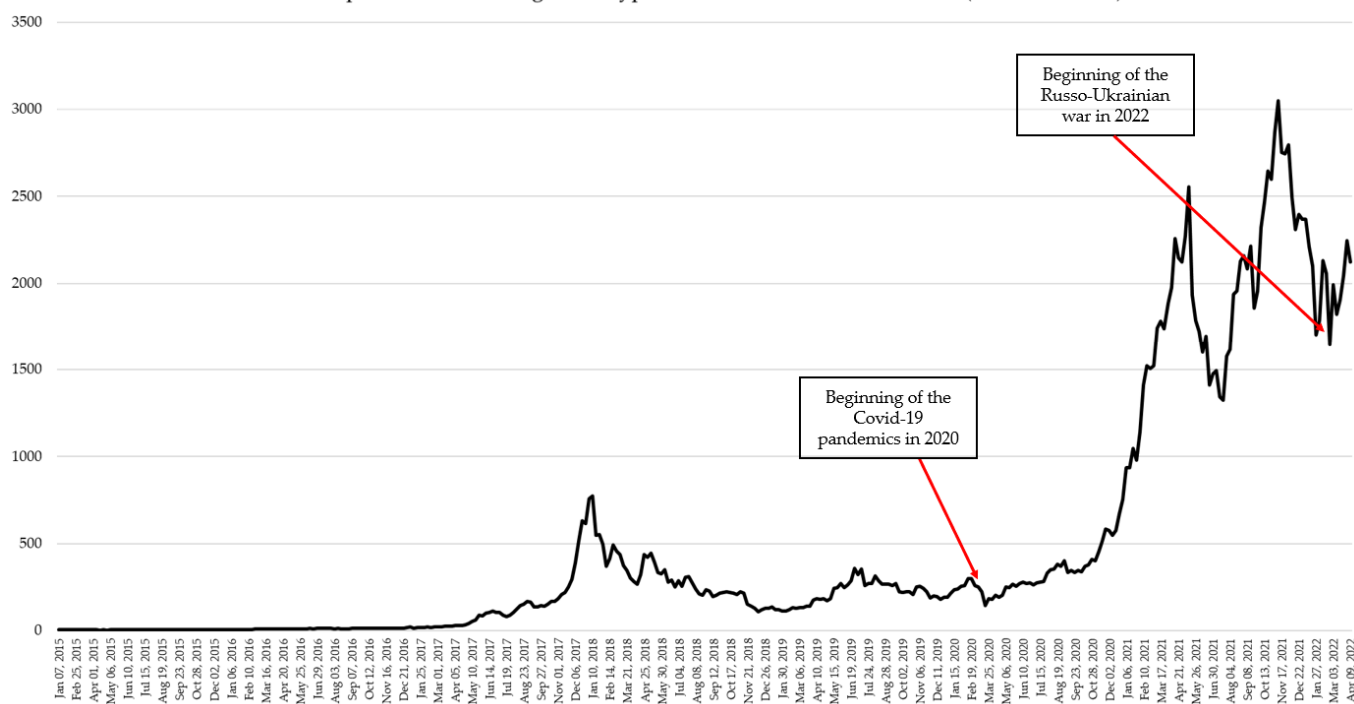


Figure 1. Market capitalization of the global cryptocurrencies market 2015–2020 (n billion USD of yearly return). Source: Own calculation, CoinMarketCap data.

In the first section of the paper, we examined the constituents and developments of the global crypto market since 2012, analyzing the top 30 cryptocurrencies as of 9 April 2022 [4]. Specifically, we included cryptocurrencies data covering the two main stressful events on the market over the last ten years, namely the outbreaks of the COVID-19 pandemic in February 2020 and the Russo-Ukrainian war in February 2022.

In the second section of the paper, we applied the dynamic time warping method on the loaded data in order to standardize them and thus to get a dataset with time series of

the same length and frequency. Subsequently, we applied various clustering methods for time series data of cryptocurrencies and equities considering various levels of their price volatility. We created a new crypto index that allows for measuring the performance of the global crypto market taking into account developments of the wider market and not just the developments of the two most famous cryptocurrencies which are Bitcoin (BTC) and Ethereum (ETH). Furthermore, we prepared a comparison of cryptocurrencies and their price developments with other equities (represented in the S&P 500 index).

In the third section, we focused on the forecasting of cryptocurrencies prices and compared various standard econometrical and machine learning methods. Several forecasting performance metrics were tested in order to determine the most suitable forecasting method. In summary, the main goals of the paper are as follows:

1. The description of the global crypto market including the creation of the crypto index that can be used for cryptocurrencies market monitoring;
2. Transformation of the data using the dynamic time warping;
3. Clustering of time series of cryptocurrencies and standard equities and their comparison;
4. The identification of the most suitable forecasting methods with regards to cryptocurrencies.

1.1. Current Research with Regards to This Topic

As described in the paper, the crypto market experienced its main growth in 2020–2022, especially in terms of various new cryptocurrencies that joined the market in this period. Hence, during the paper preparations, we considered primarily academic papers that were published in these years with particular focus on cryptocurrency properties and their forecasting. Reference [5] analyzed investments into Bitcoin and gold and tried to identify the best available forecasting model. The authors constructed an ARIMA model through differential stationarity processing and found out that the ARIMA model was the best available forecasting model by comparing it with various machine learning models for this problem [5].

Reference [6] created and calibrated the Long short-term memory (LSMT) model for the Bitcoin and gold forecasting. The LSMT algorithm that they utilized is a neural network algorithm which was supposed to be suitable for various long-term processes. The LSMT model led to mixed outcomes and didn't prove to be an ideal tool for Bitcoin forecasting. Reference [7] forecasted Bitcoin's returns and their jumps using a self-exciting process that was embedded in a stochastic volatility model. The authors found out that high and low values of the differences in predicted probabilities of positive and negative jumps in Bitcoin's returns were sufficient for its forecasting. Reference [8] proposed a two-stage approach to parametric nonlinear time series modelling in discrete time with the an objective of incorporating uncertainty or misspecification in the conditional mean and volatility. The author applied the model on Bitcoin data also targeting some COVID-19 periods.

The current academic research primarily focuses on Bitcoin or Ethereum modelling and forecasting, but a limited analysis was conducted on other younger cryptocurrencies that joined the crypto market just in last 3–4 years and they are gaining more and more of market capitalization.

1.2. Data Loading and Data Sources

We downloaded data on the top 30 cryptocurrencies covering approximately 83% of the global crypto market as of 9 April 2022 [2]. The data were collected from the online trading platform CoinMarketCap [9]. Data on each of the selected cryptocurrencies are available after their public trading launch—from before 2014 to the last two years, but their availability and time series length vary. The data suggest that by the spring 2017, activity in the global crypto market was rather low, the cryptocurrencies having boomed from 2020, when some of them entered the market. Therefore, we have limited the research to the period from 31 December 2019 to 9 April 2022, allowing us to capture both recent major global market shocks, i.e., the COVID-19 pandemic and the Russian invasion of Ukraine.

Unfortunately, it is not possible to analyze price volatility during the 2008–2009 financial crisis period because there was no full-fledged crypto market in that period.

Cryptocurrencies are traded in different price buckets (some of them close to 1 USD and some of them more expensive than 10000 USD), which required some form of data standardization for clustering and modelling purposes; modelling and forecasting being considered synonymous in this paper [10]. A few options have been tested and the best outcomes were achieved by transforming cryptocurrency price data to indexes with a starting point as of 31 December 2019. This indexing means that all the data points were divided by the value as of this date, see the Figure 2.

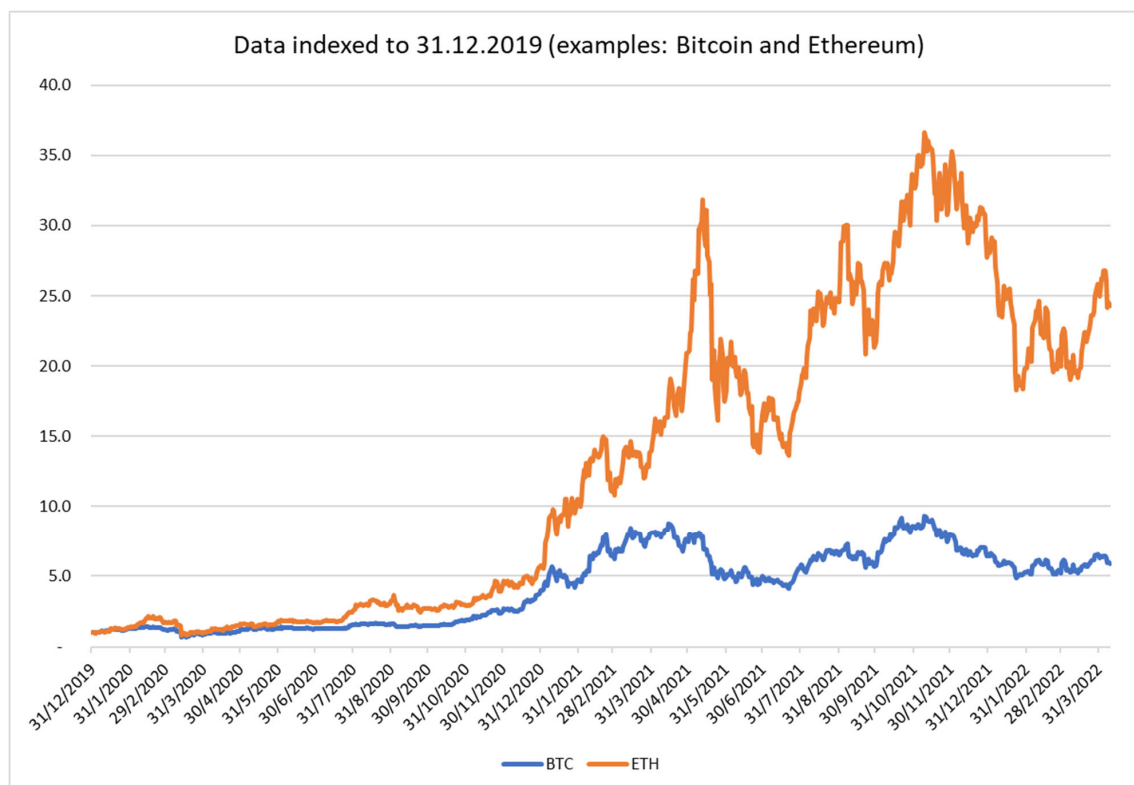


Figure 2. Transformed cryptocurrency (Bitcoin & Ethereum) price data to indexes with a starting point as of 31 December 2019. Source: Own calculation, CoinMarketCap data.

The chart above gives an example of the indexed data for the two cryptocurrencies with the highest global market capitalization, i.e., Bitcoin and Ethereum [11]. The high level of price volatility in the top 30 cryptocurrencies (measured by the standard deviation and maximum/minimum changes) is noticeable, especially compared to other equity assets. Table 1 below shows the average comparison of the top 29 cryptocurrencies (after excluding Shiba Inu as a significant outlier) and the S&P 500 index constituents covering various sectors of the economy (it includes price changes between 31 December 2019 and 9 April 2022).

Table 1. Average comparison of indexed data for the top 29 cryptocurrencies and the S&P 500 index constituents covering various sectors of the economy.

	Median	Max	Min	Standard Deviation
Top 29 cryptocurrencies	6.3	57.5	0.8	13.8
S&P 500 equities	1.2	1.8	0.6	0.3

Source: Own calculation, Bloomberg and CoinMarketCap indexed data.

A single comparison of the average statistics of both cryptocurrencies and equities indicates that the former were materially more volatile, also offering a higher growth potential. The comparison was conducted on the indexed data as described above. The Table 1 outcomes mean that the median cryptocurrencies prices increase was +530% since 31 December 2019 (as data is indexed to this date so the value on 31 December 2019 is equal to 100%).

1.3. Outliers and Exploratory Data Analysis

We conducted an exploratory data analysis of all 30 cryptocurrencies that were examined, which cover $\approx 83\%$ of the global crypto market as of 9 April 2022. Data on other cryptos (top 31–100) were also explored but not used in the analysis as they contained many data gaps and short time series, their global crypto market shares being limited (cf. the Table 2).

Table 2. Exploratory data analysis of theselected 30 cryptocurrencies.

Order	Cryptocurrencies	Crypto Code	Market Capitalization in USD	Global Market Share in %	Median	MAX	MIN	St. Deviation
1	Bitcoin	BTC	804,272,478,981	37.9%	4.8	9.3	0.7	2.6
2	Ethereum	ETH	382,662,893,229	18.0%	12.4	36.6	0.8	10.9
3	Tether	USDT	82,582,177,981	3.9%	1.0	1.0	1.0	0.0
4	BNB	BNB	70,595,674,729	3.3%	11.8	48.7	0.7	15.4
5	USD Coin	USDC	50,844,820,961	2.4%	1.0	1.0	1.0	0.0
6	XRP	XRP	36,222,968,801	1.7%	2.7	9.5	0.7	2.0
7	Solana	SOL	35,663,008,070	1.7%	9.6	270.7	0.5	70.3
8	Terra	LUNA	33,326,242,255	1.6%	18.9	404.1	0.4	99.1
9	Cardano	ADA	32,831,866,551	1.5%	24.2	88.8	0.7	23.6
10	Avalanche	AVAX	22,280,026,674	1.0%	2.3	25.4	0.6	6.4
11	Polkadot	DOT	21,245,308,736	1.0%	5.4	18.4	1.0	4.8
12	Dogecoin	DOGE	18,909,602,131	0.9%	24.9	333.1	0.7	62.3
13	Binance USD	BUSD	17,627,673,803	0.8%	1.0	1.0	0.9	0.0
14	TerraUSD	UST	16,792,352,852	0.8%	1.0	1.0	1.0	0.0
15	Shiba Inu	SHIB	13,000,941,819	0.6%	8.6	84,243.0	0.1	14,962.4
16	Wrapped Bitcoin	WBTC	11,642,832,020	0.5%	4.8	9.3	0.7	2.6
17	NEAR Protocol	NEAR	11,597,979,759	0.5%	1.8	17.1	0.5	3.7
18	Cronos	CRO	11,008,076,857	0.5%	4.2	26.5	0.9	4.7
19	Lido Staked	STETH	10,108,353,766	0.5%	2.6	7.7	1.0	2.1
20	Polygon	MATIC	9,813,486,207	0.5%	8.6	201.9	0.6	55.2
21	Dai	DAI	9,039,566,073	0.4%	1.0	1.1	1.0	0.0
22	Cosmos Hub	ATOM	7,826,780,882	0.4%	2.7	10.6	0.4	2.8
23	Litecoin	LTC	7,756,150,301	0.4%	2.8	9.1	0.7	1.7
24	Chainlink	LINK	7,050,426,203	0.3%	8.7	28.5	1.0	5.9
25	TRON	TRX	6,342,603,677	0.3%	3.6	12.5	0.6	2.6
26	Bitcoin Cash	BCH	6,126,103,334	0.3%	1.8	7.4	0.7	1.0
27	FTX Token	FTT	6,117,493,303	0.3%	10.5	37.3	0.9	10.1
28	Ethereum Classic	ETC	5,428,612,836	0.3%	2.5	29.5	0.8	5.1
29	LEO Token	LEO	5,500,882,010	0.3%	1.8	9.3	1.0	1.7
30	Algorand	ALGO	5,070,856,490	0.2%	3.3	10.6	0.6	2.5
Top 30 cryptocurrencies in total				82.8%				

Source: Own calculation, CoinMarketCap data.

Different levels of volatility were identified in the data, with one extraordinary outlier being detected. Having experienced a peak index increase of more than 8,424,000% over the last two years, the Shiba Inu crypto was removed from the analysis as it might bias the statistics. Moreover, its development has been substantially affected by the activities of some investors, exhibiting high levels of idiosyncratic risk. Hence, a total of 29 cryptocurrencies were included in the clustering and modelling analysis.

2. Dynamic Time Warping

Econometric, machine learning, and clustering methods of time series analysis are dependent on their time alignment. Any time or frequency differences that are related to the data can lead to biased models, which in turn may fail to correctly capture the analyzed variables. There are two most sophisticated algorithms for time series alignment/standardization:

- Dynamic Time Warping (DTW), see [12]
- Canonical Time Warping (CTW), see [13]

In this paper, we focus on the dynamic time warping technique, utilizing its properties in both clustering and forecasting analysis. DTW allows us to match and compare two (or more) time series of different lengths or frequencies, even from different time periods. It is a method that exactly corresponds to the 30 selected cryptocurrencies since some of them were introduced to the crypto market only after 31 December 2019. This method was used in this paper to prepare timely aligned time series used in the clustering and forecasting methods. DTW usage in the cryptocurrencies modelling is essential as crypto data contain various data gaps and many of them have short time series. No research has been published on the DTW application on cryptocurrency prices. In the Figure 3, we see the difference in the time series alignment between DTW and the traditional Euclidean distance (ED) approach.

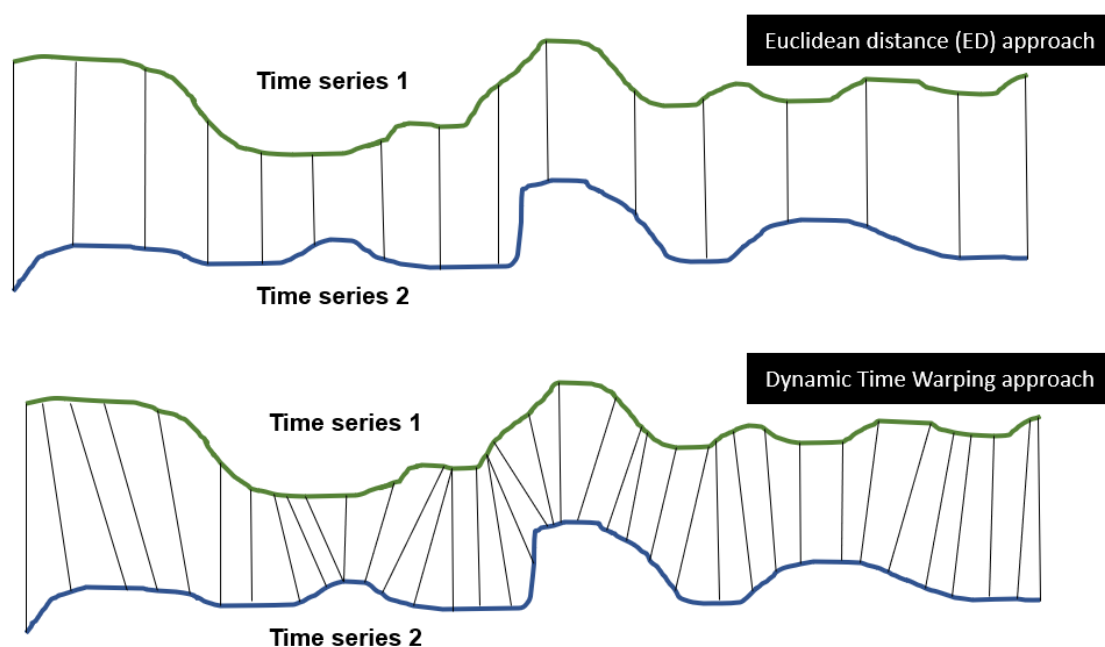


Figure 3. Time series alignment between dynamic time warping and the traditional Euclidean distance approach. Source: (Own research).

The Euclidean distance is described by the following formula, where $\vec{x} = (x_1, x_2, \dots, x_m)$ and $\vec{y} = (y_1, y_2, \dots, y_m)$ are two time series that are analyzed, and whose similarity is measured as:

$$ED(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2}$$

Dynamic time warping [14] is expressed by the following formula, where $W = \{w_1, w_2, \dots, w_k\}$ is a warping path with $k \geq m$, which is a contiguous set of matrix elements that defines a mapping between the time series $\vec{x} = (x_1, x_2, \dots, x_m)$ and $\vec{y} = (y_1, y_2, \dots, y_m)$:

$$DTW(\vec{x}, \vec{y}) = \min \sqrt{\sum_{i=1}^k w_i}$$

A warping path is a list of connections between two time series. The points of the warping path may be related to the same or similar properties of the time series, regardless of the time when they occurred. The DTW algorithm focuses on two aspects:

- The calculation of the best warping path between two time series and their points
- The length (or cost) of an optimal path, i.e., a special metric covering the whole universe of time series with their associated space

2.1. Clustering of Time Series

In order to better understand the development of cryptocurrencies and to discover the similarities between them, we conducted a cluster analysis. Cryptos are considered a rather new investment class, which was almost negligible before 2017, with some cryptocurrencies entering the market as late as in 2020 or 2021. Hence, their time series are very short and often highly volatile. The similarities between them and data-driven clustering outcomes can be utilized in forecasting and investment decisions. The outcomes may help extend the time series of newer cryptos by aligning their indexes to those of similar cryptocurrencies or even of other financial instruments such as equities. Equities have long series and their data are considered to be more stable and robust. Taking into account the clustering results (as described in this paper), an investor will be then able to extend time series of certain “rather new” cryptocurrencies by their similarity to other cryptos or equities. This approach enables an investor to better estimate a reaction of a cryptocurrency to events that happened before the cryptocurrency was even created because the similar time series (e.g., of another similar equity) may have experienced those events 10–20 years ago.

For the purposes of this paper, we focused on the time period between 31 December 2019 and April 2022. The data that were transformed by the DTW were used in the clustering. Moreover, DTW may be used not only for the data transformation, but also for the clustering purposes. In this paper section, we tested various clustering methods including two methods that are based on the DTW features.

2.2. Barycentre Averaging

In terms of clustering methodology, we adapted the current state-of-the-art DTW barycentre averaging (DBA) method [15], having tested the following three averaging approaches:

- Euclidean barycentre (without DTW)
- DBA (subgradient descent algorithm)
- Soft-DTW barycentre (with a gamma parameter)

There are many ways to cluster time series. The most common is the weighted average approach, which involves a lot of noise and unwanted elements that are gathered from clustered time series [16]. Another frequently used tool is a simple time series average, leading, however, to biased results as the weights and contributions of time series may differ. DTW barycentre averaging (DBA) is a novel non-parametric method for time

series classification combining the K-nearest neighbours (KNN) method and dynamic time warping [17]. It is a heuristic global averaging strategy to iteratively improve the initial average estimate in order to minimize its squared distance (calculated using DTW) to average estimates, employing the expectation-maximization algorithm.

All three charts within Figure 4 represent various approaches to the time series clustering. The red line is the cluster center when one cluster is considered. The x -axis presents time count as a number of days since 31 December 2019. The y -axis shows the indexed values for the underlying time series as described above.

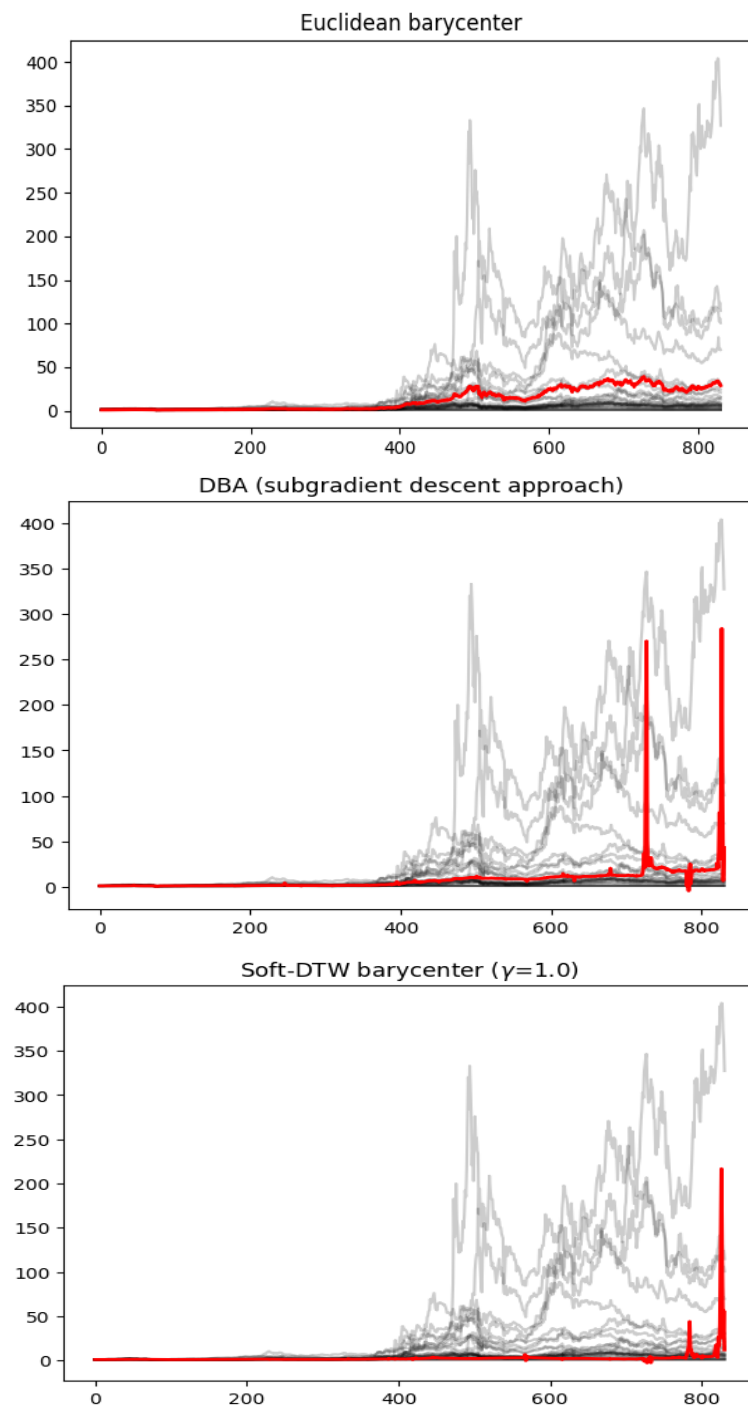


Figure 4. Different approaches to clustering time series. Source: Own calculation, CoinMarketCap data.

In the analysis of crypto data, both DBA-based methods produced very volatile results that were quite significantly influenced by outliers, especially in comparison with the Euclidean barycentre approach. Figure 4 shows differences in clustering outcomes given the chosen method. Despite the original intention to use the DBA methods for clustering and crypto indexing purposes, due to their volatile results, we decided to employ Euclidean barycentric, hierarchical, and k-Shape clustering methods on the crypto data that had already being timely aligned by the DTW.

2.3. Hierarchical and k-Shape Clustering

We applied and tested the following hierarchical clustering methods, mostly based on the linkage criteria:

- Weighted method
- Ward linkage
- Complete linkage
- Average linkage
- Centroid linkage

The results of the weighted hierarchical clustering of cryptocurrencies (using the Euclidean distance measure) are illustrated by the dendrogram below (see the Figure 5), providing a first insight into the relationships between various cryptos.

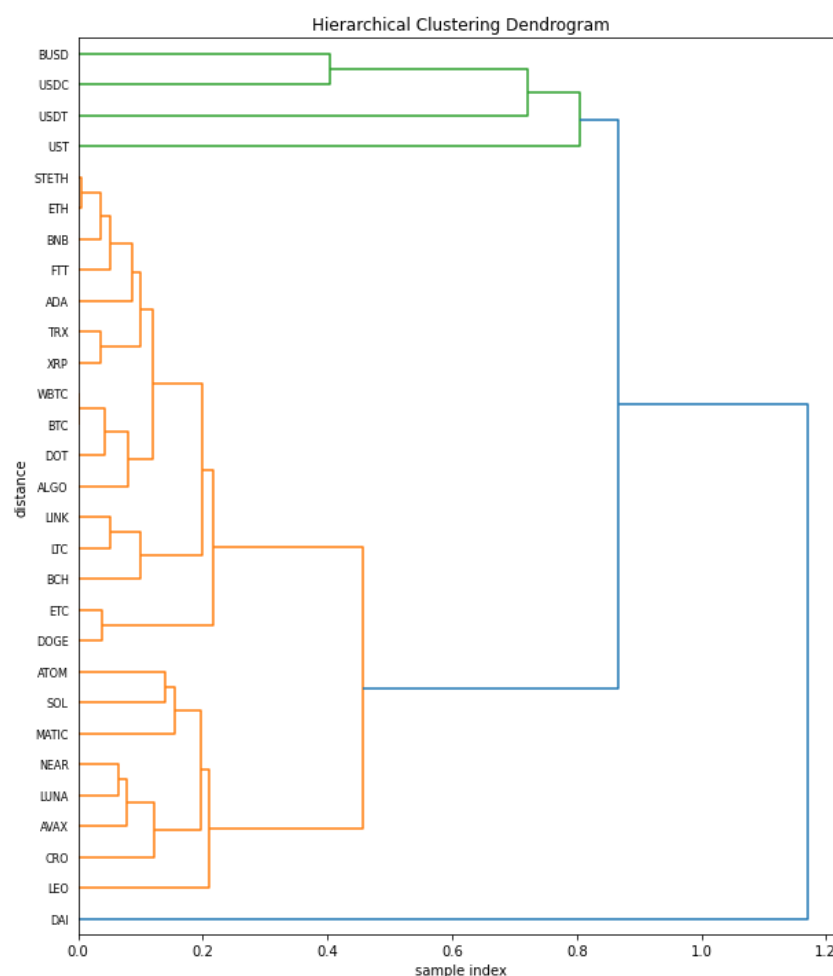


Figure 5. Weighted hierarchical clustering of cryptocurrencies (Euclidean distance measure). Source: Own calculation, CoinMarketCap data.

Hierarchical clustering is a useful tool for exploratory data analysis, but it also has some drawbacks. It does not support dynamic time warping transformation and requires

some important arbitrary decisions, such as the specification of distance metrics and linkage criteria. However, we have not found a clear theoretical basis for those decisions.

To address the above drawbacks, we applied one of the advanced algorithms for time series clustering, namely k-Shape [18] as a partitioning clustering method [19] which protects the shapes of the underlying time series. This algorithm efficiently compares time series and calculates the centroids, while considering various scaling and shifting invariances. It is a centroid-based clustering approach that is grounded on the cross-correlation measure, relying on an iterative improvement process that scales linearly in the number of sequences to generate homogeneous and well-separated clusters [19]. Our time series were transformed to the same length and frequency, utilizing DTW but omitting the DBA method.

Resting on the shape-based distance (SBD) measure and the shape extraction centroid method, the k-Shape algorithm efficiently produces time series clusters. The SBD between two time series is defined by the following formula which results in values ranging from 0 to 2 (0 indicating a perfect similarity between two time series):

$$SBD(\vec{x}, \vec{y}) = 1 - \max_w \left(\frac{CC_w(\vec{x}, \vec{y})}{\sqrt{R_0(\vec{x}, \vec{x}) \cdot R_0(\vec{y}, \vec{y})}} \right),$$

where \vec{x}, \vec{y} are the two analyzed time series and $CC_w(\vec{x}, \vec{y})$ is a cross-correlation between them, w denoting the position where $CC_w(\vec{x}, \vec{y})$ is maximized. The results of the k-Shape clustering may be found in the Table 3 below.

Such a cryptocurrency segmentation should be considered when deciding on the number of clusters to be identified. In order to capture as much information and heterogeneity as possible, the within-cluster sum of squares (WCSS) [20] is calculated. The so-called Elbow method suggested that only two clusters should be selected for the given dataset (see the sharpest point of the curve in the Figure 6); however, it needs to be considered how much information the selected clusters collect.

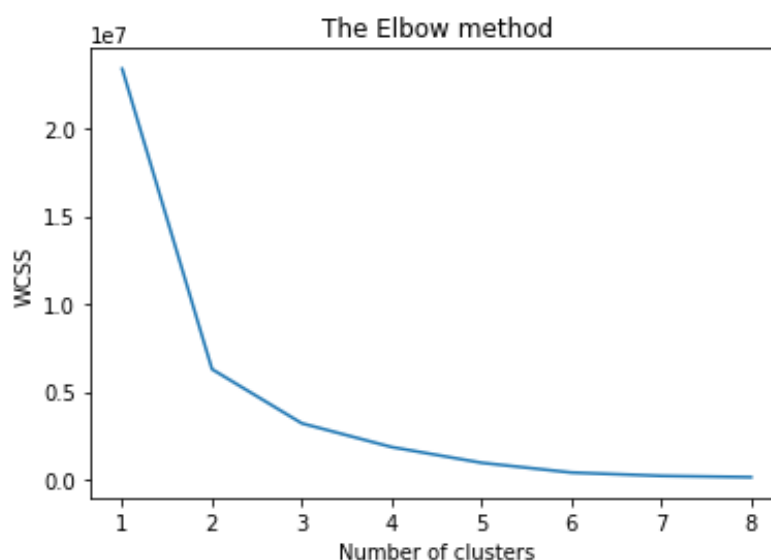


Figure 6. Elbow method. Source: Own calculation, CoinMarketCap data.

WCSS is calculated as the sum of the squared distance between each point and the centroid in a cluster. Based on the data, we chose five clusters to cover more than 90% of the total WCSS. The k-Shape clustering analysis of 29 cryptocurrencies yielded the following outcomes:

Table 3. The k-Shape clustering results of 29 cryptocurrencies.

Order	Cryptocurrencies	Crypto Code	Clusters
1	USD Coin	USDC	1
2	Binance USD	BUSD	1
3	Solana	SOL	2
4	Terra	LUNA	2
5	Avalanche	AVAX	2
6	NEAR Protocol	NEAR	2
7	Cronos	CRO	2
8	Dai	DAI	2
9	Cosmos Hub	ATOM	2
10	LEO Token	LEO	2
11	Litecoin	LTC	3
12	Chainlink	LINK	3
13	Bitcoin Cash	BCH	3
14	Bitcoin	BTC	4
15	Polkadot	DOT	4
16	TerraUSD	UST	4
17	Wrapped Bitcoin	WBTC	4
18	Algorand	ALGO	4
19	Ethereum	ETH	5
20	Tether	USDT	5
21	BNB	BNB	5
22	XRP	XRP	5
23	Cardano	ADA	5
24	Dogecoin	DOGE	5
25	Lido Staked	STETH	5
26	Polygon	MATIC	5
27	TRON	TRX	5
28	FTX Token	FTT	5
29	Ethereum Classic	ETC	5

Source: Own calculation, CoinMarketCap data.

We then applied the Euclidean barycentre method to the five calculated clusters, the results of which are plotted in the Figure 7. The *x*-axis shows the day numbers between 31 December 2019 and 9 April 2022, with the outbreaks of the pandemic and the Russo-Ukrainian war falling on days 60 and 787, respectively.

Cluster 1 includes USD Coin and Binance USD cryptocurrencies which oscillated around the same price since 31 December 2019, four other clusters showing a different price development. There is no immediate cryptocurrency response to the onset of the COVID-19 pandemic. Over the following months, however, the prices of various cryptocurrencies went up as a possible secondary effect of the pandemic and increased inflation expectations. Clusters differ primarily in the speed of the response to the pandemic and its impact on the economy (i.e., the fiscal and monetary easing [21]).

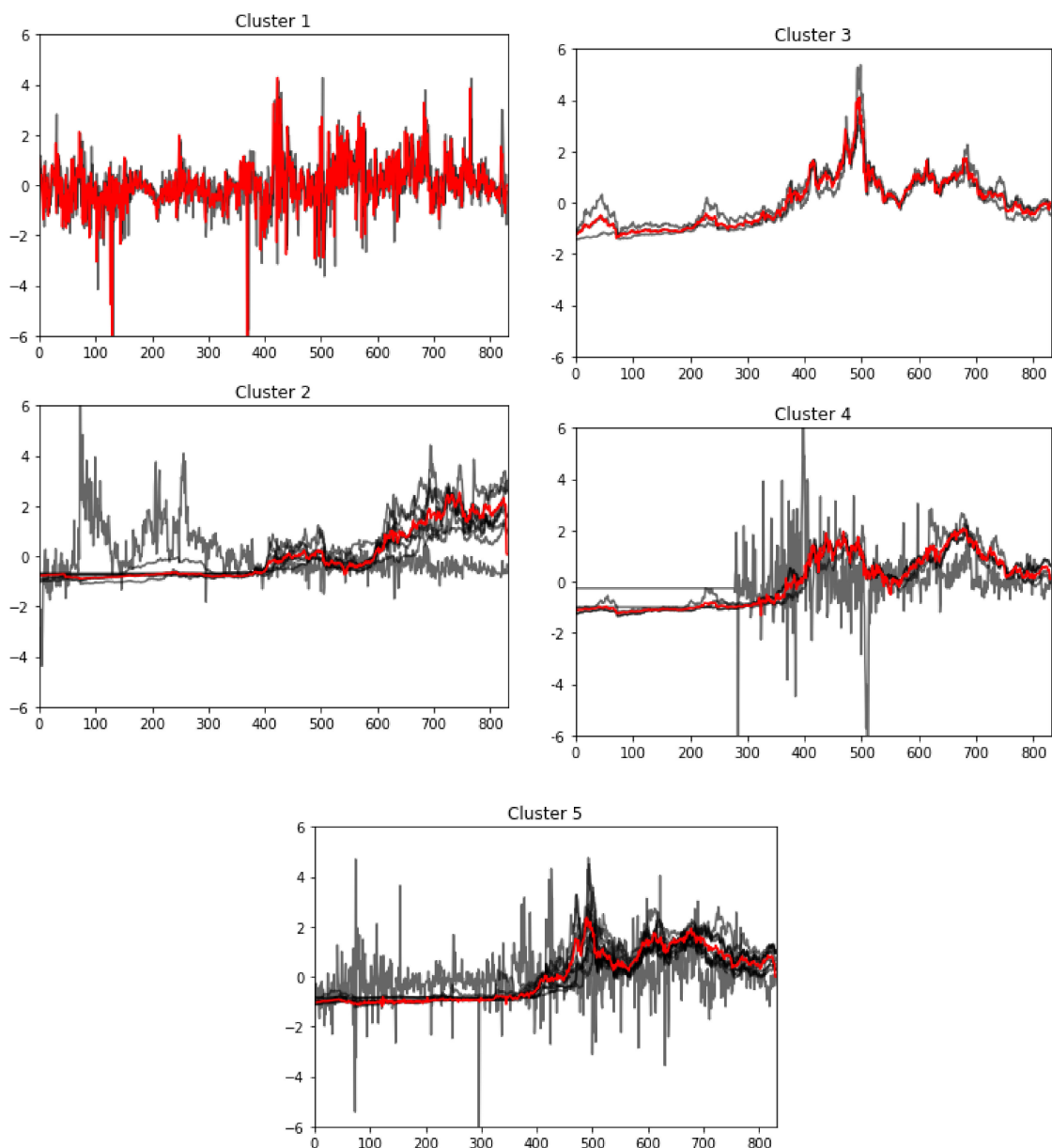


Figure 7. Euclidean barycentre method of the five calculated clusters. Source: Own calculation, CoinMarketCap data.

2.4. Crypto Index

The current global crypto market includes dozens of cryptocurrencies with varying market capitalization and volatility. To track and measure the crypto market performance over a given period or in comparison with the stock market, a benchmark crypto index can be constructed, allowing to gather information on various cryptocurrencies and their prices. Based on the previous analysis, we built this index on the data of the top 30 cryptos in terms of their market cap over the last two years, excluding the only outlier (Shiba Inu).

Then, as with each index, we set the crypto index weights, several variants (more than 10) were tested and some of them are shown in the Figure 8 and in the Table 4.

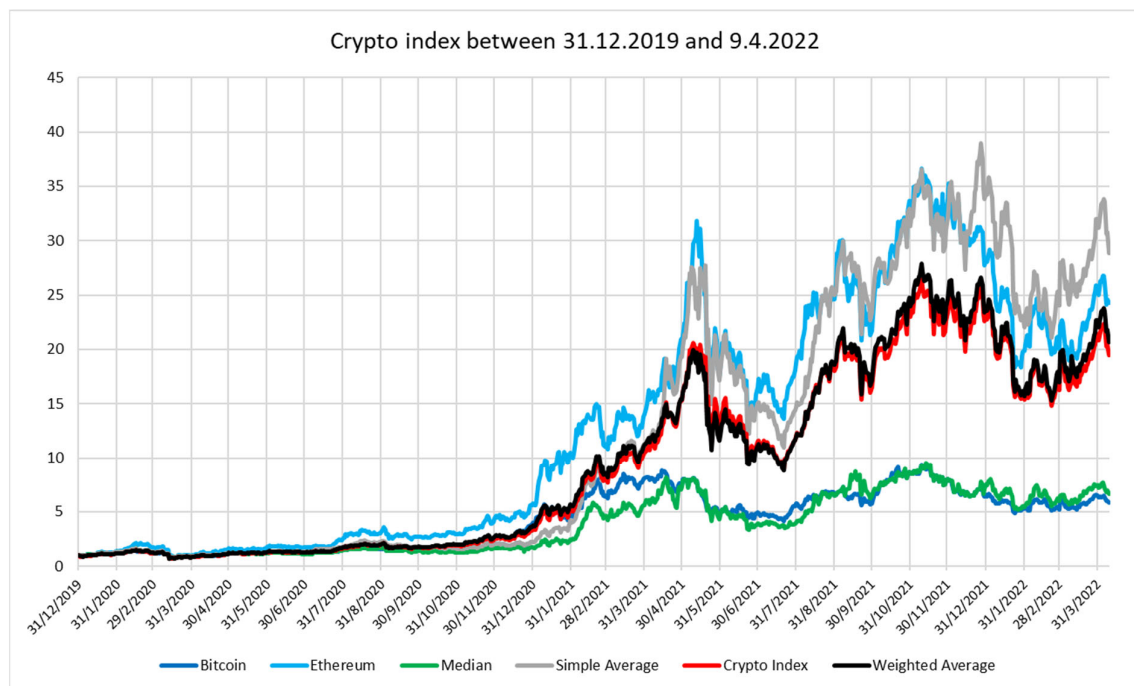


Figure 8. Crypto index between 31 December 2019 and 9 April 2022. Source: Own calculation, CoinMarketCap data.

Table 4. The benchmark crypto index. Source: Own calculation, CoinMarketCap data.

Order	Cryptocurrencies	Crypto Code	Global Market Share in %	Weighted Average Contribution in %	Crypto Index Contributions in %
1	Bitcoin	BTC	37.9%	46.1%	36.0%
2	Ethereum	ETH	18.0%	21.9%	17.2%
3	Tether	USDT	3.9%	4.7%	3.7%
4	BNB	BNB	3.3%	4.0%	3.2%
5	USD Coin	USDC	2.4%	2.9%	2.3%
6	XRP	XRP	1.7%	2.1%	1.6%
7	Solana	SOL	1.7%	2.0%	1.6%
8	Terra	LUNA	1.6%	1.9%	1.6%
9	Cardano	ADA	1.5%	1.9%	1.6%
10	Avalanche	AVAX	1.0%	1.3%	1.6%
11	Polkadot	DOT	1.0%	1.2%	1.6%
12	Dogecoin	DOGE	0.9%	1.1%	1.6%
13	Binance USD	BUSD	0.8%	1.0%	1.6%
14	TerraUSD	UST	0.8%	1.0%	1.6%
15	Shiba Inu	SHIB	0.6%	0.0%	0.0%
16	Wrapped Bitcoin	WBTC	0.5%	0.7%	1.6%
17	NEAR Protocol	NEAR	0.5%	0.7%	1.6%
18	Cronos	CRO	0.5%	0.6%	1.6%
19	Lido Staked	STETH	0.5%	0.6%	1.6%

Table 4. Cont.

Order	Cryptocurrencies	Crypto Code	Global Market Share in %	Weighted Average Contribution in %	Crypto Index Contributions in %
20	Polygon	MATIC	0.5%	0.6%	1.6%
21	Dai	DAI	0.4%	0.5%	1.6%
22	Cosmos Hub	ATOM	0.4%	0.4%	1.6%
23	Litecoin	LTC	0.4%	0.4%	1.6%
24	Chainlink	LINK	0.3%	0.4%	1.6%
25	TRON	TRX	0.3%	0.4%	1.6%
26	Bitcoin Cash	BCH	0.3%	0.4%	1.6%
27	FTX Token	FTT	0.3%	0.4%	1.6%
28	Ethereum Classic	ETC	0.3%	0.3%	1.6%
29	LEO Token	LEO	0.3%	0.3%	1.6%
30	Algorand	ALGO	0.2%	0.3%	1.6%
Top 30 cryptocurrencies in total			82.8%	100.0%	100.0%

The crypto index weights (i.e., contributions) were primarily based on the market capitalization of the top 30 cryptocurrencies, which was heavily influenced by the two best-known cryptos—Bitcoin and Ethereum. With the contributions of other cryptocurrencies being rather limited, we adopted a more structured approach consisting of the following steps:

- The introduction of a new floor for each cryptocurrency weight that was equal to 2%. More than 10 various floors were tested and used for the floor calibration. A preference was given to round numbers and floors leading to results that were similar to the weighted average approach trying to avoid overestimation of the Bitcoin and Ethereum impact. Robustness checks were conducted in the calibration process. The 2% floor led to the most stable results.
- Aggregation of the updated contributions after the floor implementation, their sum being 128%
- Rescaling the updated values by 1.28, obtaining a sum of weights that was equal to 100%
- Listing the final crypto index weights/contributions in a table (see the Table 4).

2.5. Clustering of Equities Included in the S&P 500 Index

In the next step, the k-Shape clustering algorithm was applied to a combined pool of equities that were included in the S&P 500 index and the top 29 cryptocurrencies (top 30 cryptocurrencies after the exclusion of an outlier Shiba Inu) and this combined pool was subsequently analyzed. The results of this clustering may be found in the Figure 9. Cryptocurrencies are a relatively novel type of investment not experiencing many market shocks that equities have already absorbed in the past. For example, cryptocurrencies did not experience the 2008 financial crisis, let alone the upheavals of the 20th century. Clustering of equities with cryptocurrencies can thus provide more penetrating insights into the behaviour of cryptocurrencies, including more accurate predictions. The following results were yielded by applying a five-cluster set-up:

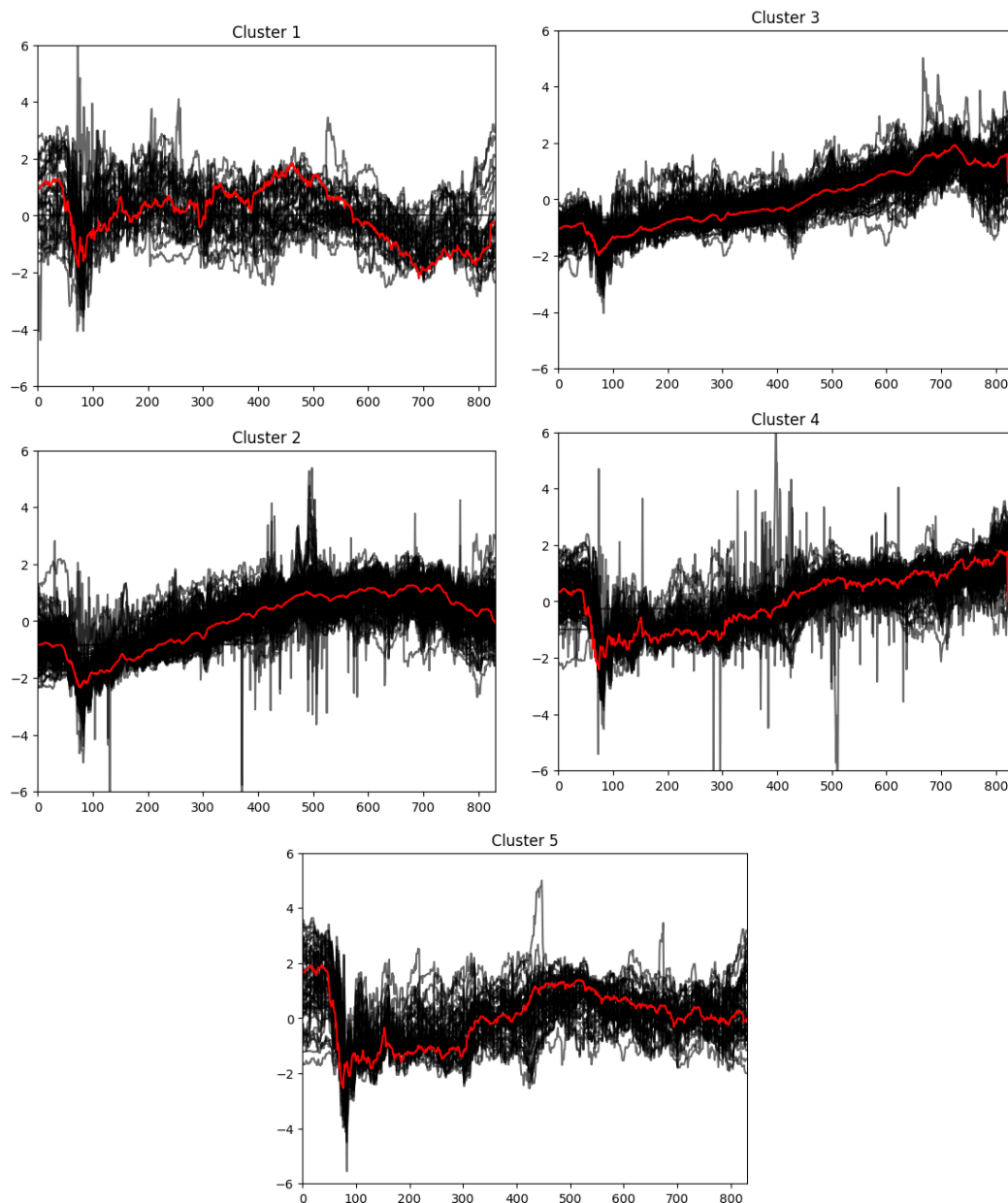


Figure 9. The k-Shape clustering algorithm applied on five clusters. Source: Own calculation, Bloomberg and CoinMarketCap data.

A total of 529 financial instruments (either equities or cryptocurrencies) were almost evenly split into five clusters. One of the advantages of combined clustering is the identification of similarities and the potential extension of cryptocurrency time series. As already described, the crypto market was rather irrelevant (in terms of market capitalization) before 2017. Therefore, the crypto market has limited empirical experience with macroeconomic or other emergencies. For example, investors can only anticipate a potential Bitcoin response to monetary interventions, rising inflation, or commodity shocks. K-Shape clustering of equities that have longer time series allows to extend cryptocurrency indices using those of similar equities from “the same cluster”.

Finally, while comparing equities and cryptocurrencies, we benchmarked the performance of the newly designed crypto index against the standardized and well-known S&P 500. The results of this comparison are displayed in Figure 10 (including the development of Bitcoin as the cryptocurrency with the highest market capitalization). Compared to 500 equity titles that were included in the S&P 500 index, cryptocurrencies showed higher

volatility, but also higher growth potential (and thus higher profitability); the crypto index having grown 14 times faster than the S&P 500 over the last two years [22].

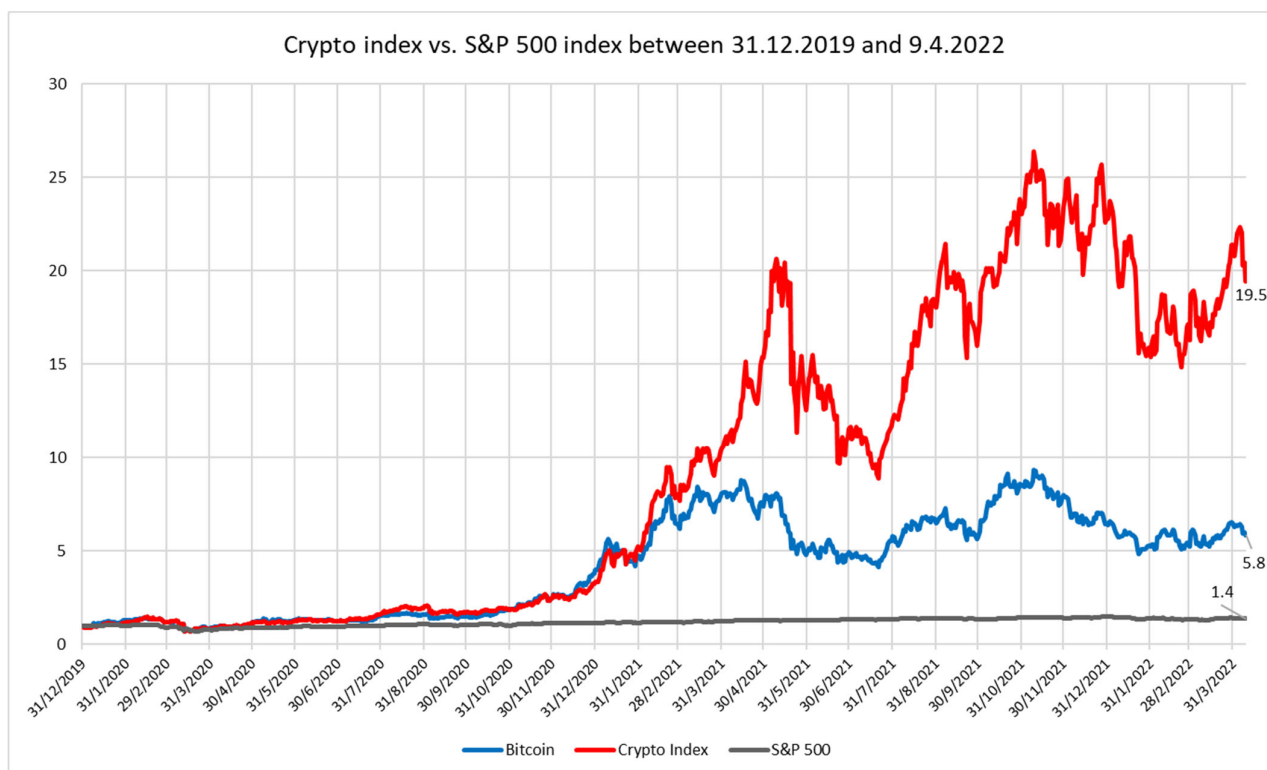


Figure 10. Crypto index vs. S&P 500 index. Source: Own calculation, Bloomberg and CoinMarketCap data.

3. Modelling and Forecasting of Cryptocurrency Prices

Finally, the aim of the paper was to compare the performance of various modelling and forecasting methods, benchmarking the methodologies of standard econometric and machine learning techniques that were applied to cryptocurrency time series. The high growth potential of cryptocurrencies (indicated by their past development outlined above) made them attractive to investors seeking revenue benefits that can curb current inflationary pressures. However, as is well known, historical profits are no guarantee of future profits. Many investors, therefore, pay attention to the future price development of cryptocurrencies. There are several approaches to price forecasting, the present paper focuses on the following categories of time series prediction methods:

1. MA, AR, ARMA, ARIMA
2. Exponential smoothing (ETS)
3. BATS model
4. K-nearest neighbour regression (KNN) ML method
5. Random forest ML method

All five methods were tested on 30 cryptocurrencies and the newly constructed crypto index. For the above listed techniques and all the cryptos, three different forecasting performance metrics were comparatively applied. The results are presented in the following section of the paper.

3.1. Stationarity and Augmented Dickey-Fuller Test

Prior to forecasting, the time series had to be examined to decide which methods were applicable, some of them (e.g., ARIMA or ARMA) requiring stationary time series. Stationarity assumes that each point of a time series is independent of the other points [23]. Time series are stationary unless a change in the time dimension leads to a change in the

shape of the underlying statistical distribution. This means that distribution properties such as mean, covariance, and variance remain constant over time. When applying autocovariance and autocorrelation-based forecasting methods to non-stationary time series, the results may be biased and unreliable. Although many time series are non-stationary in real life, this does not mean that they are unpredictable. They can be standardized and transformed into stationary time series carrying out, for example, a differencing, power, or log transformation. In this study, we applied first-order differencing, which is described by the following formula:

$$\Delta Y_t = Y_t - Y_{t-1} \quad (1)$$

Time series were examined using the following tools:

- Time series visualization
- Autocorrelation function (ACF) analysis
- Partial autocorrelation function (PACF) analysis
- Unit root testing, see [18]

In econometrics, unit root tests are used to check whether a time series is non-stationary (and has a unit root) or whether it follows a stationary process. The null hypothesis is usually defined as the existence of a unit root, an alternative hypothesis being stationarity (with more detailed options such as a trend stationarity or an explosive root).

There are several reliable unit root tests, the augmented Dickey–Fuller (ADF) test being among the most frequently used ones, indicating how significantly the time series is affected by a trend. The ADF test is based on an autoregressive model, optimizing the information criterion across multiple lag values. The ADF null hypothesis assumes the existence of a unit root and thus the non-stationarity of the time series. An alternative hypothesis rejects the null hypothesis, suggesting a stationary process.

The test's null hypothesis is that the time series is not stationary (has a time-dependent structure) and can be represented by a unit root. The alternative hypothesis (rejecting the null one), on the other hand, predicts that the time series is stationary. The ADF results highlight the test p-value. If it exceeds 5%, the null hypothesis cannot be rejected, indicating that the time series is non-stationary and it must either be transformed into a stationary one or methods not requiring stationarity (e.g., ETS or machine learning models) have to be employed.

We conducted an ADF test for the time series of 30 cryptocurrencies (with a time range from 31 December 2019 to 9 April 2022) and for the crypto index. As listed in the Table 5, most of the time series that were analyzed were non-stationary. For methods that require stationarity, we transformed the data using the first-order differentiation, all differenced time series thus becoming stationary.

Table 5. The ADF test for the time series of 30 cryptocurrencies.

Cryptocurrency/Crypto Index	Actual Time Series		First Difference	
	ADF Test Statistics	p-Value	ADF Test Statistics	p-Value
BTC	−1.322	>5%	−29.475	<5%
ETH	−1.139	>5%	−10.945	<5%
USDT	−5.735	<5%	−10.613	<5%
BNB	−1.183	>5%	−11.365	<5%
USDC	−7.773	<5%	−11.141	<5%
XRP	−1.979	>5%	−6.756	<5%
SOL	−1.128	>5%	−5.039	<5%
LUNA	0.917	>5%	−5.714	<5%

Table 5. Cont.

Cryptocurrency/Crypto Index	Actual Time Series		First Difference	
	ADF Test Statistics	p-Value	ADF Test Statistics	p-Value
ADA	−1.483	>5%	−7.664	<5%
AVAX	−0.978	>5%	−9.923	<5%
DOT	−1.603	>5%	−9.327	<5%
DOGE	−2.176	>5%	−5.368	<5%
BUSD	−8.705	<5%	−12.308	<5%
UST	−7.197	<5%	−11.264	<5%
SHIB	−1.787	>5%	−5.701	<5%
WBTC	−1.321	>5%	−29.445	<5%
NEAR	0.130	>5%	−4.847	<5%
CRO	−1.451	>5%	−7.006	<5%
STETH	−0.948	>5%	−7.640	<5%
MATIC	−1.067	>5%	−13.341	<5%
DAI	−5.400	<5%	−12.161	<5%
ATOM	−1.137	>5%	−7.816	<5%
LTC	−1.618	>5%	−9.319	<5%
LINK	−1.787	>5%	−7.918	<5%
TRX	−1.596	>5%	−7.118	<5%
BCH	−2.217	>5%	−6.507	<5%
FTT	−1.057	>5%	−7.718	<5%
ETC	−1.898	>5%	−6.045	<5%
LEO	1.046	>5%	−9.013	<5%
ALGO	−1.436	>5%	−8.887	<5%
Crypto index	−0.614	>5%	−7.759	<5%

Source: Own calculation in Python.

3.2. Forecasting Methods

The indexed data were transformed by dynamic time warping (DTW) to match the time range that was analyzed. For practical purposes, the database was divided into training and testing sets. The mean absolute percentage error (MAPE) was then calculated for all cryptocurrency time series and all methods. The testing part of the data included the last 36 observations that were used to validate the results (cf. the Figure 11). The x-axis shows the day numbers between 31 December 2019 and 9 April 2022, the outbreaks of the pandemic and the Russo-Ukrainian war falling on days 60 and 787, respectively.

ARIMA is a class of forecasting methods that is used for time series analysis, containing related techniques such as ARMA as well as simple AR, I, and MA models. Simpler tools can be derived using a suitable combination of parameters. The ARIMA model consists of three components:

- AR standing for “autoregression”—the dependence between current and previous (lagged) observations
- I standing for “integration”—the differencing of actual observations to make a time series stationary (e.g., by power transformation or by subtracting the observation in the previous time step from the raw observation)
- MA standing for “moving average”—the dependence between true observations and residual errors from a moving average model (applied to lagged observations)

As for observations, the general forecasting equation is as follows:

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \cdots - \theta_p e_{t-p}$$

ARIMA(1,0,0) is an example of a first-order autoregressive model with a time series that is stationary and autocorrelated, including a constant. Its forecasting equation would then be:

$$\hat{Y}_t = \mu + \phi_1 Y_{t-1}.$$

Each of the three ARIMA components is explicitly defined by a model parameter. The usual notation is ARIMA(p, d, q), the parameters clearly indicating the setting of the model:

- $p \Rightarrow$ the number of lagged observations, indicating the “lag order”
- $d \Rightarrow$ the degree of differencing, indicating which observations were subtracted from the previous ones
- $q \Rightarrow$ the size of the moving average window, indicating the number of observations for each moving average calculation

In the present paper, we considered all potential combinations of the ARIMA model (including simple ARMA, AR, I, or MA models) calibrating forecasts for p , d , and q parameters ranging from 0 to 12 (including all their combinations). Only the best MAPE-based calibrated ARIMA model was included in the comparative analysis along with other methods [5].

Exponential smoothing (ETS) methods calculate forecasts as weighted averages of past observations, with weights decreasing exponentially as the observed values become outdated [24]. The more recent the observation, the higher the ETS weight. There are both simple and complex ETS forecasting models that are based on the calculation of weights. The simplest form of the exponential smoothing is given by the following formula:

$$s_t = \alpha x_t + (1 - \alpha)s_{t-1} = s_{t-1} + \alpha(x_t - s_{t-1})$$

where α is a smoothing factor that can take values between 0 and 1, those close to 1 giving more weight to the more recent observation and reducing the smoothing effect, and s_t is the weighted average of the current observation x_t and the previous smoothed ones.

BATS is a forecasting method combining four components—Box–Cox transformation, ARMA errors, and trend and seasonal components [24]. The Box–Cox transformation is applied to the original time series, then modelled as a linear combination of an exponentially smoothed trend, ARMA and seasonal components. The BATS models that were used in this paper employed the AIC-based hyperparameter tuning method determining which of the four components to include or exclude.

K-nearest neighbour regression (KNN) is a machine-learning (ML) algorithm that predicts a given variable based on a selected similarity criterion, allowing for various distance functions to be utilized. KNN is a non-parametric method that has been applied in forecasting, statistical estimation, and pattern recognition over the last 50 years. The KNN algorithm is a type of supervised ML procedure that is used to solve classification and regression tasks [25]. The distance functions vary for different KNN models, and in this case, we applied the Euclidean distance (ED) of the form:

$$ED = \sqrt{\sum_{i=1}^k (x_i - y_i)^2}.$$

The KNN results for the crypto index are shown in the Figure 12. The difference between the orange line (testing data) and the green line (calculated prediction) is a measure of the model/method’s accuracy.

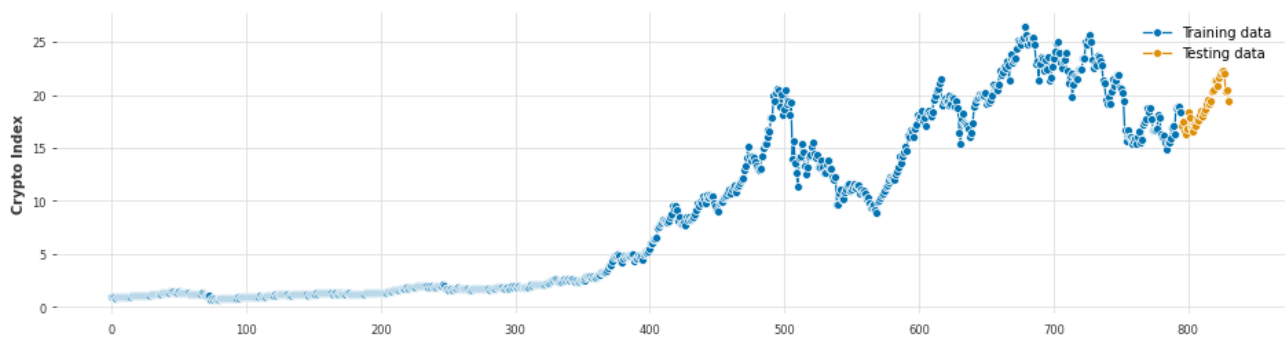


Figure 11. The indexed data transformed by dynamic time warping (training and test data). Source: Own calculation, CoinMarketCap data.

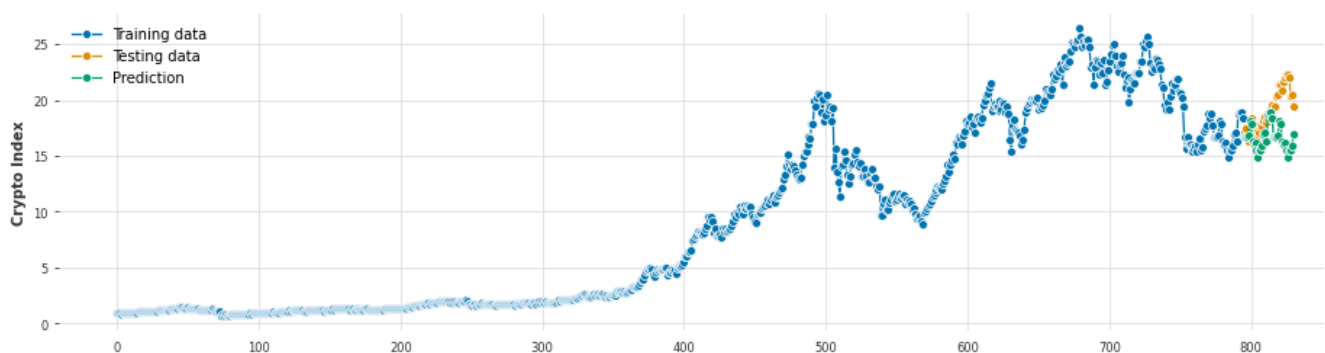


Figure 12. K-nearest neighbour regression result. Source: Own calculation, CoinMarketCap data.

Another ML method is the Random Forest model. It is an aggregation of many decision trees which make primary predictions. Random Forest models can be applied to classification and forecasting tasks [26]. The model itself contains the following steps:

- Dataset splitting: The model randomly selects the features and observations. Different features become responsible for creating different decision trees. Moreover, observations that are divided into training and testing ones can be used to assess the model's accuracy and precision.
- Decision-making process: Each decision tree makes its own decisions based on its features and data.
- Aggregation of the outcomes from various decision trees: Multiple individual decisions are combined to build the final random forest model. This leads to more robust and accurate forecasts in comparison to simpler ML methods.

3.3. Forecast Results and Performance Metrics

We used the following three performance metrics to evaluate the forecast results:

- Mean absolute percentage error (MAPE)
- Mean absolute error (MAE)
- Root mean square error (RMSE)

The mean absolute percentage error (MAPE), also called the mean absolute percentage deviation (MAPD), is a metric of prediction accuracy for predictive statistical methods. It is expressed as a ratio that is defined by the formula:

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|,$$

where A_t is the actual observation of each cryptocurrency, and F_t is the predicted value for the given variable and time step. Their difference is then divided by the actual observation

A_t . Subsequently, the absolute value in this ratio is aggregated for each forecasted time step and divided by the number of the predicted values n .

The mean absolute error (MAE) calculates the average magnitude of absolute errors in a set of the predicted values (see the formula below). All individual errors (differences between forecasts and actual values) belonging to different time steps have equal weights.

$$MAE = \frac{1}{n} \sum_{t=1}^n |A_t - F_t|$$

The root mean square error (RMSE), the last performance benchmark that was tested in this paper, uses a quadratic approach, also measuring the average magnitude of the forecast error. RMSE is a square root of the mean of the squares of the differences between the predictions and the actual values of the cryptocurrencies (cf. the formula below).

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (A_t - F_t)^2}$$

MAE and RMSE measures bear certain similarities. They do not express the average model prediction error in percentage terms, but in units of the predicted variable. Both metrics can range from 0 to infinity regardless of the direction of errors. MAPE, on the other hand, is standardized in percentages, being indifferent to the units of the measured variable. It ranges between 0 and 1 and is more suitable for comparison purposes, especially when the levels of the variables (as in our case various cryptocurrencies) differ significantly. All three metrics are negatively-oriented scores, indicating that their lower values correspond to more accurate forecasting. After unit scaling, all three metrics led to comparable conclusions regardless of their different formulas. For the sake of brevity, we further focus only on MAPE as the most appropriate comparative benchmark (unrelated to underlying time series units). Its results are plotted in Figure 13.

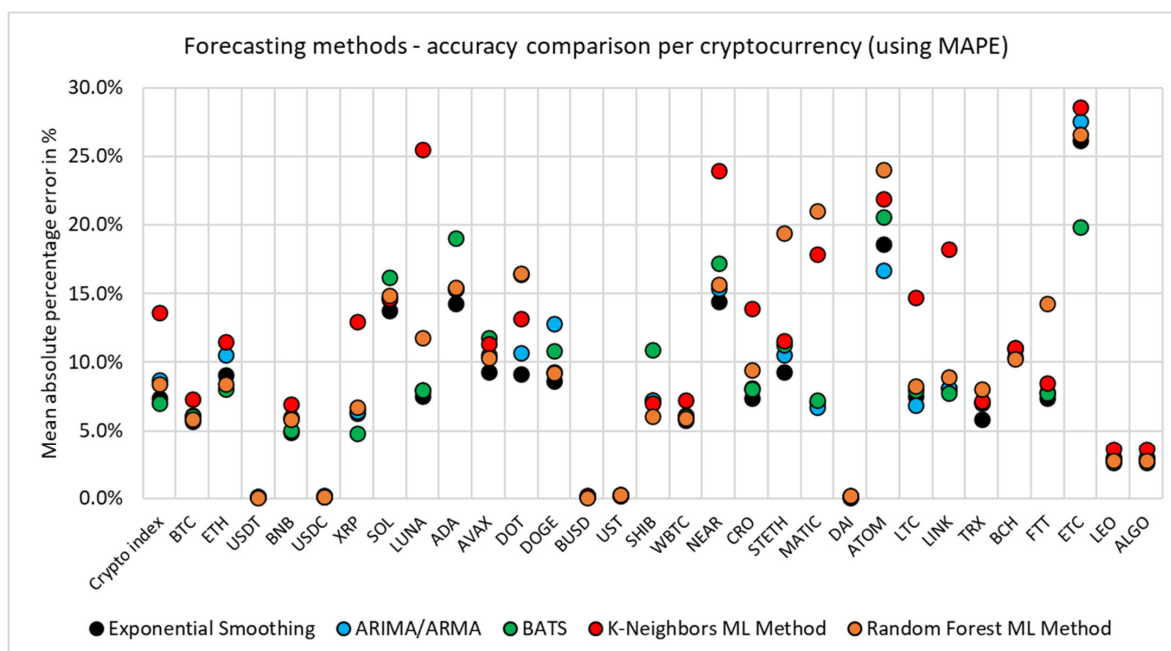


Figure 13. Forecasting method—accuracy comparison per cryptocurrency (using MAPE). Source: Own calculation, CoinMarketCap data.

The chart above shows the MAPE results of five forecasting methods for the 30 selected cryptocurrencies and the newly constructed crypto index. All forecasting methods that

were used were briefly described above. The ARIMA method produces the most optimal potential ARIMA/ARMA model whose parameters capture all combinations from 0 to 12. Thus, traditional econometric forecasting methods have surprisingly led to better results than the machine learning ones (such as the K-nearest neighbours or Random Forest regression techniques), identifying differences in the accuracy of predictions of different cryptocurrencies.

Some cryptocurrencies (e.g., USD Coin, Tether, Binance USD, Terra USD, or Dai) were easier to predict (their models generally resulting in lower MAPE metric values), while other cryptos (e.g., Ethereum Classic, Luna, or NEAR Protocol) indicated more data problems and volatility, leading to worse forecast results.

In addition, for some cryptocurrencies (e.g., Bitcoin or Tether), the choice of the forecasting method did not significantly affect the accuracy of the results, whereas for others (e.g., Polkadot or Cosmos Hub), some modelling methods were more suitable. The accuracy of forecasts may be taken into account in strategic investment decisions. Cryptocurrency forecasting with lower accuracy, i.e., with a higher MAPE/MAE/RMSE score, can be enhanced by the outcomes of the combined clustering with equities. This means to extend the indexed crypto time series by indices of the most statistically similar equities with longer time series available.

The MAPE results were compared across the analyzed sample to reveal their distributions for different prediction methods and for the crypto index, the latter showing the results that were similar to the 50th (median) and 75th percentiles of the sample distribution (cf. the Figure 14). Comparison of the methods indicated bigger differences with increasing percentile levels, which means that outlier cryptocurrencies caused the largest differences across the methods.

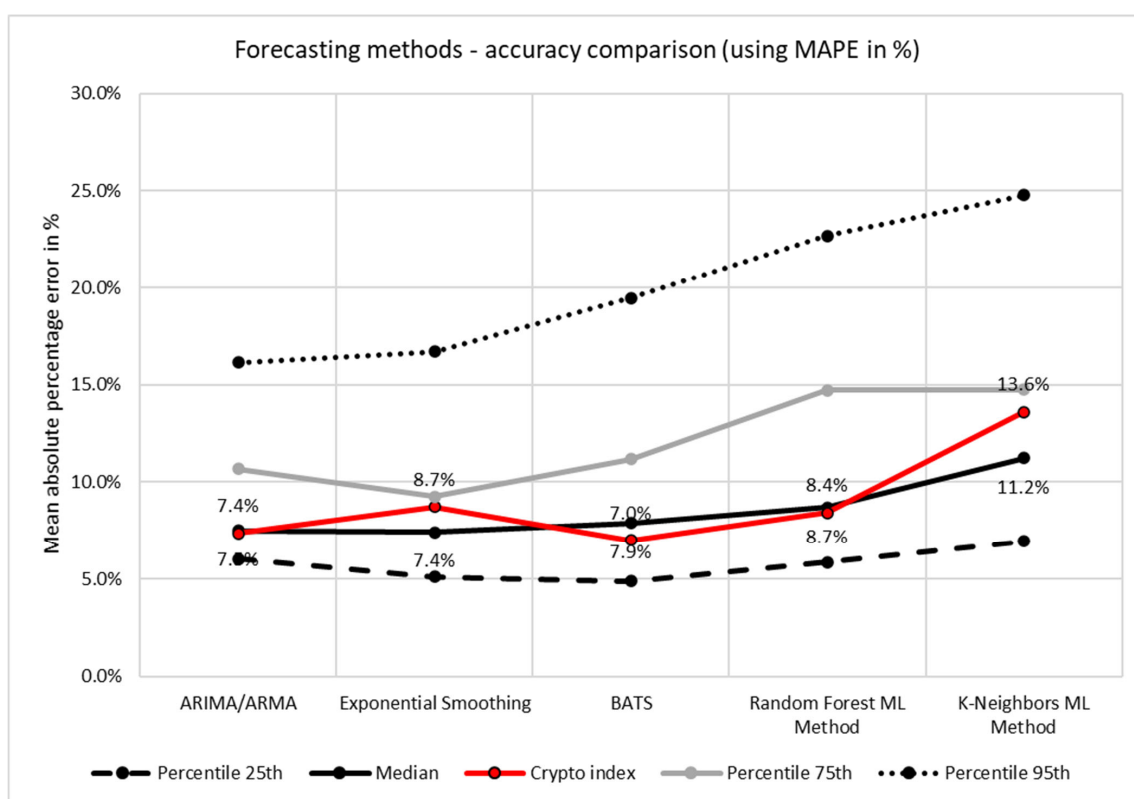


Figure 14. Forecasting methods—accuracy comparison (using MAPE in %). Source: Own calculation, CoinMarketCap data.

4. Conclusions

The paper's goals have been accomplished. The analysis demonstrated the growing importance of the global crypto market in terms of its capitalization which increased from USD 10 m to over 3000 bn between February 2011 and November 2021, respectively. We analyzed the growth developments of the 30 cryptocurrencies with the highest market capitalization in the last two years, designing our own aggregate crypto index, which reflects not only the development of Bitcoin, but also the other 28 top cryptos. The results were reflected in the analysis that was carried out as of 9 April 2022, i.e., the deadline for data retrieval. In addition, we provided a definition of cryptocurrencies and their main properties in line with previous academic works [1].

In exploratory analysis, we processed the data (of cryptocurrencies and equities) using the dynamic time warping (DTW) method to obtain the time series of the same length and frequency. Furthermore, various clustering methods were tested and analyzed on crypto data including two methods that were based on the DTW features. The most robust results were achieved by the k-Shape clustering method, which was also used for combined clustering, merging equities, and cryptocurrencies. The similarities between cryptocurrencies and clustering outcomes can be utilized in forecasting and investment decisions. Combined clustering between equities and cryptocurrencies has a potential to improve the accuracy of cryptocurrency forecasting. The time series of “new” (“new cryptocurrency” means a cryptocurrency with a short time series which was introduced to the market in 2020–2022) cryptocurrencies may be enhanced by aligning their indexes to those of similar cryptocurrencies or even of other financial instruments such as equities. Equities have long series and their data are easily available. This approach enables an investor to better estimate a reaction of a certain cryptocurrency to events that happened before the cryptocurrency was even created. The investor may thus predict a stress impact on its portfolio, despite the fact, that the portfolio contains many “new” cryptocurrencies with short time series. This approach can be used in an investment strategy setting or in decisions regarding the portfolio composition.

As the current academic research primarily focused on Bitcoin or Ethereum modelling and forecasting with limited analysis conducted on other newer cryptocurrencies that entered the crypto markets just in last 3–4 years despite their increasing market capitalization, we tested various forecasting methods on the top 30 cryptocurrencies, not limiting our analysis to Bitcoin and Ethereum only, comparing traditional econometric techniques with machine learning approaches. Traditional methods, ARIMA and ETS, have led to slightly more accurate results. For the ARIMA, we considered all potential combinations of the underlying parameters (including simpler ARMA, AR, I, or MA models) calibrating forecasts for p , d , and q parameters ranging from 0 to 12 (including all their combinations). Only the most accurate ARIMA model was used for each cryptocurrency in the comparison. The implemented machine learning techniques may be further upgraded and more complex machine learning and artificial intelligence methods may be adopted in the follow-up research; more specifically, Deep Learning, recurrent neural network, and long short-term memory (LSTM) methods.

Finally, the data showed that cryptocurrencies may generate high returns, but also high volatility (of returns which is a sign of higher risk). While comparing equities and cryptocurrencies, we benchmarked the performance of the newly designed crypto index against the standardized S&P 500. Compared to 500 equity titles that were included in the S&P 500 index, cryptocurrencies showed higher volatility (and thus higher risk) than equities, but also higher growth potential (and thus higher profitability), the crypto index grew 14 times faster than the S&P 500 over the last two years. A limited short-term response that was also detected of the cryptocurrencies to the COVID-19 pandemic and to the Russo-Ukrainian war beginning over the 2020–2022 period. However, cryptocurrency prices showed a strong sensitivity to the secondary effects of the pandemic, i.e., rising inflation and its expectations. Some investors that were concerned about a potential depreciation of standard currencies and investments, have become more focused on cryptocurrencies

that are not easily influenced by central bank monetary interventions and are independent from central authorities [1], and are depending only on a decentralized, distributed ledger blockchain technology as described in the first section of this paper.

Author Contributions: Conceptualization, T.Š., D.B. and L.M.; methodology, T.Š.; software, T.Š.; validation, J.K., T.Š. and D.B.; formal analysis, T.Š.; investigation, T.Š.; writing—original draft preparation, T.Š.; writing—review and editing, J.K. and L.M.; visualization, T.Š.; supervision, D.B. and L.M.; project administration, J.K.; funding acquisition, J.K. All authors have read and agreed to the published version of the manuscript.

Funding: The authors appreciate the funding support received from the Internal Grant Agency of the Prague University of Economics and Business within the IGA/A Grant Competition for the project [Machine-learning and artificial intelligence based methods to model blockchain generation process and prices of the top 30 most influential cryptocurrencies using dynamic and canonical time warping], no. [ES410022] (project OP RDE IGA/A, CZ.02.2.69/0.0/0.0/19_073/0016936).

Data Availability Statement: The data used to support the findings of this study are available from the corresponding author upon reasonable request only for research purposes.

Acknowledgments: The authors would like to thank the editorial board of the Mathematics Journal for the possibility to publish this article in a special issue: Advances in Blockchain Technology. This article is fully thematically relevant for a special issue and offers a beneficial thematic extension of the journal, which provides an advanced forum for studies related to mathematical sciences, as the article focuses on the use of cryptography-based technologies.

Conflicts of Interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

1. Lánský, J. Possible State Approaches to Cryptocurrencies. *J. Syst. Integr.* **2018**, *9*, 19–31. [CrossRef]
2. CoinMarketCap. Crypto Market Cap Charts. (n.d.). Retrieved Available online: https://www.coingecko.com/en/global_charts (accessed on 18 August 2022).
3. Lánský, J. Analysis of Cryptocurrencies Price Development. *Acta Inform. Pragensia* **2016**, *5*, 118–137. [CrossRef]
4. Motsi-Omojiade, I.D. *Cryptocurrency Regulation: A Reflexive Law Approach*; Routledge: London, UK, 2022.
5. Zhou, Q.; Chen, Z.; Cai, Z.; Xia, Z. Prediction of the Best Portfolio for Bitcoin and Gold based on the ARIMA Model. *Front. Bus. Econ. Manag.* **2022**, *4*, 141–149. [CrossRef]
6. Han, D.; He, M.; Wang, L. Bitcoin or Gold? A Financial Investment Model Based on LSTM. *Front. Bus. Econ. Manag.* **2022**, *4*, 72–77. [CrossRef]
7. Chen, J.; Clements, M.; Urquhart, A. Forecasting Bitcoin. *SSRN Electronic J.* **2022**. [CrossRef]
8. Siu, T. Bayesian nonlinear expectation for time series modelling and its application to Bitcoin. *Empir. Econ.* **2022**, 2–26. [CrossRef] [PubMed]
9. CoinMarketCap. Cryptocurrency Prices, Charts and Market Capitalizations. (n.d.) Available online: <https://coinmarketcap.com/> (accessed on 18 August 2022).
10. Arowolo, M.O.; Ayegba, P.; Yusuf, S.R.; Misra, S. A Prediction Model for Bitcoin Cryptocurrency Prices. In *Blockchain Applications in the Smart Era*; Springer: Cham, Switzerland, 2022; pp. 127–146.
11. Kim, H.M.; Bock, G.W.; Lee, G. Predicting Ethereum prices with machine learning based on Blockchain information. *Expert Syst. Appl.* **2021**, *184*, 115480. [CrossRef]
12. Koschke, R.; Steinbeck, M. Clustering paths with dynamic time warping. In Proceedings of the 2020 Working Conference on Software Visualization (VISOFT), Adelaide, Australia, 28–29 September 2020; pp. 89–99.
13. Zhou, F.; De la Torre, F. Generalized canonical time warping. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 279–294. [CrossRef] [PubMed]
14. Deller, J.R.; Hansen, J.H.; Proakis, J.G. Dynamic Time Warping. In *Discrete-Time Processing of Speech Signals*; IEEE: Piscataway, NJ, USA, 2000; pp. 623–676. [CrossRef]
15. Shi, K.; Qin, H.; Sima, C.; Li, S.; Shen, L.; Ma, Q. Dynamic barycenter averaging kernel in RBF networks for time series classification. *IEEE Access* **2019**, *7*, 47564–47576. [CrossRef]
16. Tran, T.M.; Le XM, T.; Nguyen, H.T.; Huynh, V.N. A novel non-parametric method for time series classification based on k-Nearest Neighbors and Dynamic Time Warping Barycenter Averaging. *Eng. Appl. Artif. Intell.* **2019**, *78*, 173–185. [CrossRef]
17. Shukla, A.K.; Janmajaya, M.; Abraham, A.; Muhuri, P.K. Engineering applications of artificial intelligence: A bibliometric analysis of 30 years (1988–2018). *Eng. Appl. Artif. Intell.* **2019**, *85*, 517–532. [CrossRef]

18. Yang, J.; Ning, C.; Deb, C.; Zhang, F.; Cheong, D.; Lee, S.E.; Sekhar, C.; Tham, K.W. k-Shape clustering algorithm for building energy usage patterns analysis and forecasting model accuracy improvement. *Energy Build.* **2017**, *146*, 27–37. [\[CrossRef\]](#)
19. Paparrizos, J.; Gravano, L. k-shape: Efficient and accurate clustering of time series. In Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Victoria, Australia, 31 May–4 June 2015; pp. 1855–1870.
20. Brusco, M.J.; Stahl, S. *Branch-and-Bound Applications in Combinatorial Data Analysis*; Springer: New York, NY, USA, 2005; pp. 59–76. [\[CrossRef\]](#)
21. Cui, W.; Sterk, V. *Quantitative Easing*; SAR China Research Paper WP; Hong Kong Institute for Monetary and Financial Research (HKIMR): Hong Kong, China, 2019.
22. Yilmazkuday, H. COVID-19 effects on the S&P 500 index. *Appl. Econ. Lett.* **2021**, 1–7. Available online: <https://economics.fiu.edu/research/working-papers/2021/2117/2117.pdf> (accessed on 18 August 2022).
23. Carrion, J.L.; Sansó, A. A guide to the computation of stationarity tests. *Empir. Econ.* **2006**, *31*, 433–448. [\[CrossRef\]](#)
24. De Livera, A.M.; Hyndman, R.J.; Snyder, R.D. Forecasting time series with complex seasonal patterns using exponential smoothing. *J. Am. Stat. Assoc.* **2011**, *106*, 1513–1527. [\[CrossRef\]](#)
25. Jivani, A.G.; Shah, K.; Koul, S.; Naik, V. The adept K-nearest neighbour algorithm-an optimization to the conventional K-nearest neighbour algorithm. *Trans. Mach. Learn. Artif. Intell.* **2016**, *4*, 52. [\[CrossRef\]](#)
26. Genuer, R.; Poggi, J.M. Random forests. In *Random Forests with R*; Springer: Cham, Switzerland, 2020; pp. 33–55.