Börner, Christoph J.; Hoffmann, Ingo; Krettek, Jonas; Schmitz, Tim

Bitcoin: like a satellite or always hardcore? A core–satellite identification in the cryptocurrency market

Journal of Asset Management

Provided in Cooperation with:
Springer Nature

Suggested Citation: Börner, Christoph J.; Hoffmann, Ingo; Krettek, Jonas; Schmitz, Tim (2022) : Bitcoin: like a satellite or always hardcore? A core–satellite identification in the cryptocurrency market, Journal of Asset Management, ISSN 1479-179X, Palgrave Macmillan, London, Vol. 23, Iss. 4, pp. 310-321, https://doi.org/10.1057/s41260-022-00267-z

Journal of Asset Management (2022) 23:310–321
https://doi.org/10.1057/s41260-022-00267-z

ORIGINAL ARTICLE

Bitcoin: like a satellite or always hardcore? A core–satellite identification in the cryptocurrency market

Christoph J. Börner[1] · Ingo Hoffmann[1] · Jonas Krettek[1] · Tim Schmitz[1]

Revised: 1 April 2022 / Accepted: 20 April 2022 / Published online: 21 May 2022

Abstract

Cryptocurrencies (CCs) have become increasingly interesting for institutional investors' strategic asset allocation and will therefore be a fixed component of professional portfolios in the future. However, this asset class differs from established assets primarily in that it has a higher standard deviation and tail risk. The question then arises whether CCs with similar statistical key figures exist. On this basis, a core market incorporating CCs with comparable properties enables the implementation of a tracking error approach. A prerequisite for this is the segmentation of the CC market into a core and a satellite, with the latter comprising the accumulation of the residual CCs remaining in the complement. Using a concrete example, we segment the CC market into these components based on modern methods from image/pattern recognition.

Keywords Cryptocurrencies · Core–satellite identification · Market segmentation · Pattern recognition

JEL Classification C14 · C46 · C55 · E22 · G10

Introduction

Cryptocurrencies (CCs) have gained tremendous attention and popularity in media and society in recent years, not least because of their high market volatility. Due to their nature, CCs are seen more as an investment object than as a currency (Baur et al. 2018) in the classical sense. The development of rising investment volumes has been a trend for years, and it can be assumed that CCs are gradually on their way to becoming an established asset class. Against this background, it seems plausible that CCs will become a fixed component of institutional investors' portfolios in the future.

In professional portfolio management, one approach is to segment the investment universe into a core of assets with homogenous statistical properties and assets that differ significantly from these properties—the so-called satellite. The core market can then be tracked using adequate asset picks with a tracking error approach. The satellite investments represent only a small proportion of the total portfolio and are mostly composed of actively managed sub-portfolios covering selected areas that are meant to deliver above-average returns and have a diversifying effect due to their low correlation with the core investment (Amenc et al. 2012).

In standard portfolios, for example, satellite investments such as geographical regions, asset classes different from the core investment and the purchase of portfolios with different management styles or strategies are suitable for enriching or diversifying the core portfolio. It is also possible to consider a certain asset class and differentiate between core investments and satellite investments. Examples include the sector selection of corporate bonds or the segmentation of stocks into "with the market" (core) and "high beta stocks" (satellite).

In this paper, we investigate specifically the CC asset class and propose a method to segment the core market

✉ Jonas Krettek
Jonas.Krettek@hhu.de

Christoph J. Börner
Christoph.Boerner@hhu.de

Ingo Hoffmann
Ingo.Hoffmann@hhu.de

Tim Schmitz
Tim.Schmitz@hhu.de

[1] Chair of Business Administration, esp. Financial Services, Faculty of Business Administration and Economics, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

from the satellite based on the development of key statistical parameters.

Attempts to depict the CC market holistically for reasons of portfolio and risk management have already been investigated in the literature. A prominent series of studies have addressed the construction of an index for CCs. In this context, the CC market index CRyptocurrency IndeX (CRIX) proposed by Trimborn and Härdle (2018) represents a well-known example and is intended to serve as a starting point to address these economic questions. A similar top-down approach based on the 30 largest CCs by market capitalization has been used to calculate the CC index CCi30 (Rivin et al. 2017).

However, instead of focusing on market capitalization and trading volume and thus prioritizing larger CCs, we identify the core market by applying a core–satellite approach based on the individual risk–return profile. Our approach has some potential advantages compared to a top-down constructed index. While such indices take only the largest CCs into account and may suffer from survivorship bias, the core–satellite approach identifies the core of the market, i.e., those CCs that behave similarly in statistical terms. Although we currently exclusively consider 27 CCs due to data gaps, regarding shifts in perspective as the market grows, it might become possible to use our method to identify the core market from a large number of CCs. To build a portfolio, investors would then no longer need to replicate the indices but could deliberately buy fewer individual assets of the core of the CC universe and combine them with those of the satellite. Since the core market could be represented with fewer assets in the portfolio, the monitoring costs for the portfolio would decrease. Moreover, potential problems in the portfolio, such as price collapses, operational risk (Trimborn et al. 2020) or the extinction of entire CCs, could be countered more quickly. This is a decisive advantage, especially due to the dynamics and speed of the market.

To identify CCs showing a comparable performance from 2014 to 2019, we consider returns as well as standard deviations proposed by modern portfolio theory (Markowitz 1952). One general problem is that CCs are different from traditional asset classes, especially in terms of extreme tails and corresponding tail risk. Against this background, Majoros and Zempléni (2018) and Börner et al. (2021) show that the stable distribution (SDI) is well suited to statistically model the returns of CCs overall, especially in the tail area. Thus, we extend our database by including the tail parameter α of the SDI to specifically consider the tail risk. To identify similar patterns in the development of statistical parameters, we use the dynamic time warping (DTW) algorithm. This algorithm was originally developed for speech recognition (Sakoe and Chiba 1978) but is widely used for clustering and classification in various application fields today (Giorgino 2009).

The DTW analysis leads to DTW distances that are defined in pairs. The question arises if assets—CCs in the present case—can be grouped together in such a way that they are similar in the sense of short DTW distances according to specified criteria, namely, the aforementioned statistical indicators. This would allow assets to be divided into a core and a satellite. The particular difficulty lies in the fact that sorting individual DTW distances becomes a monotonically increasing function over natural numbers, and the possible value range $[0, \infty]$ is almost continuously covered in many cases. On this basis, it must be examined whether a specific DTW distance can be derived purely from the data, acting in further steps as a boundary to divide the investment universe into a core and a satellite. In the following, we present a general procedure that is based on modern methods of pattern recognition and precisely answer these questions, and we systematically show the separation of core assets within an investment universe. The process is not restricted to a specific asset class and can be used wherever it is important to separate similar from dissimilar assets.

Using the development of statistical parameters, we show that segmenting the CC market into a core and a satellite succeeds when applying our method. Furthermore, we answer the question of whether Bitcoin is indeed part of the hard core of the CC market or just a satellite. As the CC market becomes more professional, that is, as market capitalization, liquidity, and market depth increase, our method might become an indispensable tool for professional asset management.

The remainder of this paper is structured as follows: In Sect. 2, we describe the data used for our analysis. In Sect. 3, a brief overview of the DTW methodology is given. In the main part of our study, Sect. 4, we develop the identification method that separates the CC universe into a core market segment and a complementary segment that is an accumulation of the remaining residual CCs—the satellite. The separation procedure is shown using real data. Section 5 presents some robustness checks. The last section summarizes our most important results and gives an overview of further research topics.

Data

For the foundation of our analysis, we follow various studies by extracting the daily prices of CCs from the website coinmarketcap.com (Fry and Cheah 2016; Hayes 2017; Brauneis and Mestel 2018; Caporale et al. 2018; Gandal et al. 2018; Glas 2019). To depict the CC market as a whole, we aim to include as many CCs in our analysis as possible. However, there is a trade-off between having the longest time series possible and the number of CCs in the sample because, on average, seven CCs die out per week (ElBahrawy et al.

Table 1 Considered CCs, data source: CoinMarketCap

CC	ID	CC	ID	CC	ID
Anoncoin	ANC	BitBar	BTB	Bitcoin	BTC
CasinoCoin	CSC	Deutsche eMark	DEM	Diamond	DMD
Digitalcoin	DGC	Dogecoin	DOGE	Feathercoin	FTC
FLO	FLO	Freicoin	FRC	GoldCoin	GLC
Infinitecoin	IFC	Litecoin	LTC	Megacoin	MEC
Namecoin	NMC	Novacoin	NVC	Nxt	NXT
Omni	OMNI	Peercoin	PPC	Primecoin	XPM
Quark	QRK	Ripple	XRP	TagCoin	TAG
Terracoin	TRC	WorldCoin	WDC	Zetacoin	ZET

2017). Against this background, we end up with an observation period from 2014-01-01 to 2019-06-01 taking 66 potential CCs from the CoinMarketCap Market Cap Ranking at the reference date of 2014-01-01 into consideration, as these CCs have been present throughout the entire timeframe.

As data gaps appear in the time series of most CCs, we exclude all CCs with five or more consecutive missing observations. By utilizing the last observation carried forward (LOCF) approach, as previously done in Schmitz and Hoffmann (2021), Trimborn et al. (2020), Börner et al. (2021), we are able to include all CCs with smaller data gaps. Hence, $N = 27$ CCs remain in our dataset, as depicted in Table 1.

In a next step, we convert the CC closing prices denoted in USD to EUR prices using the daily USD–EUR exchange rates retrieved from Thomson Reuters Eikon. To prevent potential weekday biases, the resulting (daily) observations are converted into weekly observations (Dorfleitner and Lung 2018; Aslanidis et al. 2020). Intraday data are not considered to further avoid biases, e.g., through pump and dump schemes.

As a starting point, we compute logarithmized weekly returns, which are referred to as returns for the sake of simplicity in the following. On this basis, we calculate the average weekly returns per year as well as the standard deviations and fit the tail parameter α of the SDI. Our longitudinal analysis from 2014 to 2019 allows us to examine the market dynamics of the statistical parameters herein.

Dynamic time warping

As we use three variables simultaneously, a simple correlation analysis is not adequate since it would provide only information about the similarity of a certain statistical variable that has no valid significance for an $N \times N$ matrix with $N = 27$, with a given total of only six annual measuring points. Furthermore, a simple correlation analysis is unable to capture the dynamic character of the development of the statistical parameters. Therefore, we use the DTW algorithm to segment the market into CCs that show similar behavior over the investigated period.

Sigaki et al. (2019) employ this methodology, thereby revealing clusters of CCs with similar informational efficiency. However, they consider only returns as a variable on a smaller time period. Although pursuing a different goal, we also select a DTW distance matrix, but instead of analyzing clusters, we implement pattern recognition to identify core and satellite CCs. Because the DTW algorithm is well known and widely used, only a cursory overview of this method is given in the following, focusing on the features relevant to our analysis.

The DTW distance can be used as a shape-based dissimilarity measure that finds the optimum warping path between two time series by minimizing a cost function (Sakoe and Chiba 1978; Aghabozorgi et al. 2015). By following the definition and notation of the main literature, in the first step, a so-called distance matrix between each pair of time series compared needs to be calculated. This distance matrix can be based on various metrics. For our analysis and for reasons of robustness, we compute Manhattan, Euclidean and squared Euclidean distance matrices. As explained in Sect. 4 in more detail, we use three variables per CC to determine the distance matrices between each pair of (multivariate) time series over the course of 2014–2019. We end up with a distance matrix for each of the three metrics and each pair of time series. Note that the distance matrix described thus far is made up of a scheme of six rows and six columns due to the six discrete points in time. For a specific currency pair, each cell in the scheme contains the distance in the respective metric for a specific point in time. In the literature, the latter is also referred to as the *local cost matrix*. The above scheme must be carefully distinguished from the distance matrix **D** defined in Sect. 4.

Given the distance matrix, i.e., the 6×6 scheme of each CC pair, the DTW algorithm finds the optimal alignment through it starting in each distance matrix at (2014, 2014) and finishing at (2019, 2019) (Sakoe and Chiba 1978). This implies that the time differences between the time series are eliminated by warping the time axis of one series so that the maximum coincidence is attained with the other (Sakoe and Chiba 1978). The individual distances of the DTW path are aggregated to total costs using a cost function. The total costs, referred to as the DTW distance d, reflect the minimum costs between the time series compared. For clarity, it should be noted that the DTW distance between the same objects equals 0 since there is no dissimilarity. The upper-left part of Fig. 1 shows the DTW distance d_{mn} for each pair $m, n = 1, ..., N$ of the $N = 27$ CCs.

We outline the underlying methodology only briefly, but there are several restrictions and setting options for

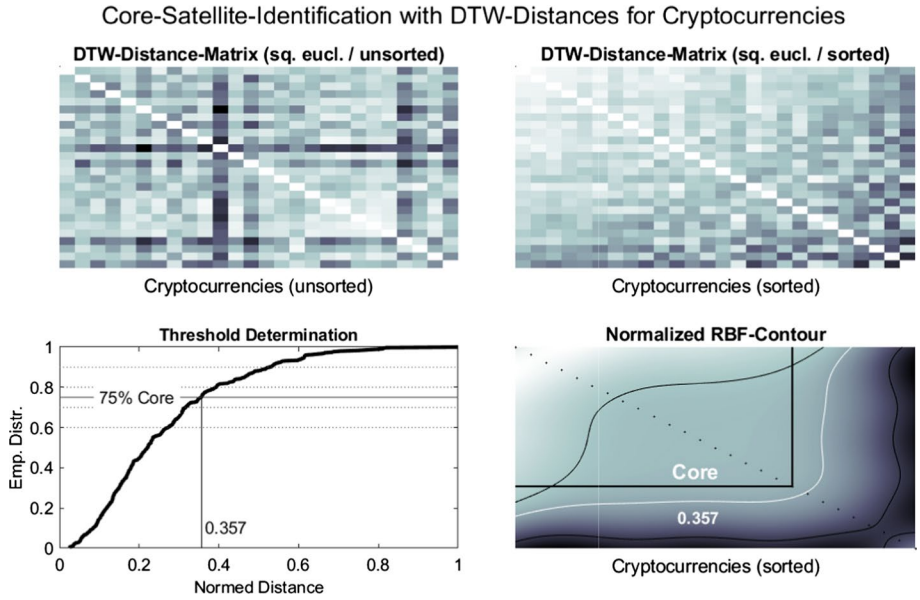

Fig. 1 Shown is the procedure to identify and separate a core of some CCs within the whole market. The remaining CCs belong to a set that encloses the satellite

Table 2 Input data for the DTW distance analyses (exemplary excerpt)

CC		Value	Sample vector						DTW dist.	
No.	ID		2014	2015	2016	2017	2018	2019	Metric	d_{mn}
6	DMD	$\langle r \rangle$	− 4.52	1.67	− 0.73	8.46	− 5.50	0.85	Manh.	
		s	28.12	22.96	9.65	20.00	14.37	16.48		6.01
		α	1.81	1.18	2.00	2.00	1.30	2.00	Eucl.	
11	FRC	$\langle r \rangle$	− 7.10	− 1.62	− 0.25	5.92	− 1.97	5.05		3.10
		s	18.74	20.27	59.19	44.84	29.43	104.04	sq. Eucl.	
		α	2.00	1.44	0.63	1.18	1.51	0.90		5.03
21	XPM	$\langle r \rangle$	− 7.27	− 0.38	− 0.58	5.38	− 2.98	0.70	Manh.	
		s	15.58	24.28	8.66	26.47	22.84	13.35		1.06
		α	1.67	1.65	1.79	1.63	1.76	1.57	Eucl.	
27	ZET	$\langle r \rangle$	− 5.74	0.16	0.51	2.88	− 3.73	1.11		0.53
		s	28.72	24.63	16.39	35.35	20.14	24.01	sq. Eucl.	
		α	1.65	1.68	1.72	1.74	1.83	1.80		0.13

Return $\langle r \rangle$ and standard deviation s in percent per week

the algorithm and the cost function. For a more detailed overview, see, e.g., Sakoe and Chiba (1978) and Giorgino (2009).

Core–satellite identification

In strategic asset allocation, a core–satellite strategy is the division of an investment into a portfolio consisting of a broadly diversified core investment that is intended to offer a basic return with moderate risk and several individual investments (satellites) with higher risk and higher earnings potential. The latter serves to increase the return of the overall investment (Methling and von Nitzsch 2019).

The returns, sample average $\langle r \rangle$, standard deviation s and tail parameter α are examined as the essential statistical parameters for CCs.

A brief overview of the implemented parameterization and the SDI's main features are given in Appendix 1. The tail parameter α plays a significant role in differentiating between CCs in which the returns almost obey a normal distribution (i.e., $\alpha \to 2$) or possess a long tail (i.e., $\alpha \ll 2$) with correspondingly high tail risks.

Overall, we consider the dynamics of the sample vector $(\langle r \rangle, s, \alpha)'$ over time from 2014 to 2019 in our analysis.

Table 2 shows an exemplary excerpt of four CCs from the whole dataset. The aim is to use the temporal development of the statistical parameters to infer CCs that can be assigned to a market core due to their similar statistical behavior.

Table 3 Analysis of the DTW distances for the different metrics

CC		Name	Metric			Core
No.	ID		Manhattan	Euclidean	Sq. Eucl.	Intersection
		Threshold	0.554	0.564	0.357	
1	ANC	Anoncoin	1	1	1	**C**
2	BTB	BitBar	1	1	1	**C**
3	BTC	Bitcoin	1	1	1	**C**
4	CSC	CasinoCoin	0	0	0	S
5	DEM	Deutsche eMark	1	1	1	**C**
6	DMD	Diamond	0	1	0	S
7	DGC	Digitalcoin	1	1	1	**C**
8	DOGE	Dogecoin	1	1	0	S
9	FTC	Feathercoin	1	1	1	**C**
10	FLO	FLO	1	1	1	**C**
11	FRC	Freicoin	0	0	0	S
12	GLC	GoldCoin	1	1	1	**C**
13	IFC	Infinitecoin	0	0	0	S
14	LTC	Litecoin	1	1	1	**C**
15	MEC	Megacoin	1	1	1	**C**
16	NMC	Namecoin	1	1	1	**C**
17	NVC	Novacoin	1	1	1	**C**
18	NXT	Nxt	1	1	1	**C**
19	OMNI	Omni	1	1	1	**C**
20	PPC	Peercoin	1	1	1	**C**
21	XPM	Primecoin	1	1	1	**C**
22	QRK	Quark	1	1	1	**C**
23	XRP	Ripple	0	0	0	S
24	TAG	TagCoin	1	1	0	S
25	TRC	Terracoin	1	1	1	**C**
26	WDC	WorldCoin	0	0	0	S
27	ZET	Zetacoin	1	1	1	**C**

Identification of the core (C) and the satellites (S)

The three-dimensional vector $(\langle r \rangle, s, \alpha)'$ is examined over the course of six years, and the DTW distance d_{mn} is determined in pairs, with $m, n = 1, ..., N$. The DTW distance is calculated in three different metrics: Manhattan, Euclidean and squared Euclidean. This allows three matrices $\mathbf{D}_{\text{Metric}} \in \mathbb{R}^{N \times N}$ for the CCs to be determined, each for a specific metric, with elements $d_{mn;\text{Metric}}$. These square matrices are symmetrical $d_{nm} = d_{mn}$, the entries on the diagonal are zero $d_{mm} = 0$, and the off-diagonal elements are all positive.

For the two pairs of CCs in the last column of Table 2, the calculated DTW distances in each metric can be compared. Note that the number in the first column corresponds to the numbering in Table 3.

The first pair (Diamond (DMD) and Freicoin (FRC)) admirably exhibits a considerable distance in each metric. A detailed analysis of the vectors over time shows the reason for this great distance. On the one hand, clear differences can be observed with regard to the absolute level and the sign (same year) of the returns. On the other hand, there are strong differences in the standard deviation and in the tail parameter (same year). While the standard deviation of the CC Diamond remains almost the same at a high level, the scattering of the returns of the FRC increases dramatically in 2019. For both CCs, it is remarkable that the underlying return distribution changes from almost normal to a heavy-tailed distribution. This can be clearly seen in the year-to-year change in the tail parameter α. Since this change does not occur at the same time (same year) for the two CCs, the DTW analysis results in large distances. Furthermore, we observe this changing distribution behavior with other CCs. At this point, we recommend that a potentially time-varying and hence nonstationary distribution warrants closer examination to control the risk for institutional investors if CCs represent a significant component of the assets being allocated.

The second pair of CCs (Primecoin (XPM) and Zetacoin (ZET)) illustrates two CCs behaving very similarly. Overall, comparatively small DTW distances can be observed here. The returns, the standard deviations and the tail parameters are closely related. It is also worth noting that the form of the underlying distribution of returns hardly changes; the variability of the tail parameter is likely to derive from statistical errors based on the small database.

The upper-left part of Fig. 1 shows the distance matrix $\mathbf{D}_{\mathrm{SE}}$ for the DTW distances in the squared Euclidean metric as a surface plot for all CCs. The ordering descends from the numbers given in the first column of Table 3. The colors indicate the value of the DTW distance from small (white) to large (black), i.e., in this example, from 0 to approximately 5. The entries with zero DTW distance are marked white.

The problem of identifying groupings in the set of CCs leads to the problem of finding structures in the distance matrix $\mathbf{D}_{\mathrm{Metric}}$. One possibility to carry out this structural analysis of the distance matrices is to apply methods that have been used for a long time in the investigation of hierarchical matrices (Liu et al. 2012; Hackbusch 2015). Similar to image recognition, these methods aim to recognize patterns in matrices.

In the first step, the CCs are rearranged in such a way that the CC displaying the greatest distance to all others on average is depicted on the right. In a descending order, the CCs with successively smaller distances are arranged to the left.[1] As a result, we gain an ordered set of CCs, and the resulting surface plot changes, as shown in the upper-right part of Fig. 1. The similarity of different CCs with regard to the dynamics of the statistical key figures is given when the DTW distance is small and tends toward zero. This is the case for CCs in the upper-left white corner in the sorted matrix. Starting from the top-left corner in the direction of the main diagonal up to a certain distance d_{bound}, the CCs thus delimited represent the market core of similar CCs.

Since the height profile above the sorted distance matrix has a peaked, rough structure, cf. Fig. 3 in the Appendix, the set of CCs belonging to the core cannot be delimited reliably. Therefore, modeling is carried out first, yielding a smoother height profile. We use a modeling method comparable to the analysis of hierarchical matrices, cf., e.g., Hackbusch (2015) and the corresponding literature cited therein.

There is a certain basic structure of the matrix that simplifies the modeling problem. The sorted distance matrix is square, symmetrical and has only positive elements. The entries become larger on average toward the right and lower edge of the sorted distance matrix, essentially revealing a concave structure. In addition, we look only for a certain block in the sorted matrix, which starts in the upper-left corner and is itself square.

The surface's concave structure can be modeled well with radial basis functions, which have their center points—similar to a frame—in the outer area of the edges of the sorted distance matrix. A brief overview of the radial basis function model class used is given in Appendix 2. If the individual elements of the distance matrix are normalized between 0 and 1, the modeling leads to an area whose height profile can be seen in the lower-right part of Fig. 1.

In the next step, we define the boundary condition d_{bound}, which delimits the set of CCs belonging to the core. The distance matrix incorporates solely positive elements, but it is not positive definite in all cases; thus, some eigenvalues can be negative, and an analysis of the eigenvalue spectrum does not lead to the definition of a suitable threshold d_{bound}.

Instead of this consideration, we analyze the empirical distribution function over the elements of the upper triangular matrix normalized between 0 and 1 and delimit an area that contains p percent of the smallest distances, cf. the lower-left part of Fig. 1. In practice, the p value will be somewhere between 60 and 90% (dotted lines), where the steep slope of the empirical distribution function merges into the flatter area. This represents the tail area of the empirical distribution function in which the empirically measured DTW distances increase rapidly. If small values are found for d_{bound}, the associated core is more homogeneous. The larger the values are, the more heterogeneous the core becomes with regard to the statistical parameters. In our example, we find the inflection point at $p \approx 75\%$. The associated threshold $d_{\mathrm{bound}} = 0.357$ is used to draw a contour (white) in the modeled height profile on the right, which delimits an upper block matrix.

This block matrix describes the market core if the squared Euclidean metric is used. In Table 3, the penultimate column shows the CCs belonging to the core if this metric is utilized (a Boolean value of 1 indicates belonging to the core).

In our analysis, we examine the DTW matrices for all metrics in the same way. The CCs belonging to the core according to the respective metric are shown in Table 3. Note that this method can also be used to find very similar CCs with almost the same statistical behavior if the lower inflection point is identified in the empirical distribution function and the bound d_{bound} is thus determined. We have also examined this path of segmentation (not explicitly shown here). This segmentation leads to the delimitation of five CCs in the white upper-left corner of the lower-right part in Fig. 1, which behave almost identically considering the statistical key figures over a long period of time.

Table 3 shows that the number of CCs belonging to the core depends on the metric. In portfolio management, it might be predicated on a decision of practicability regarding which metric to use and which dependency to accept.

[1] This form of ordering is the same as sorting according to maximum rows or the column total.

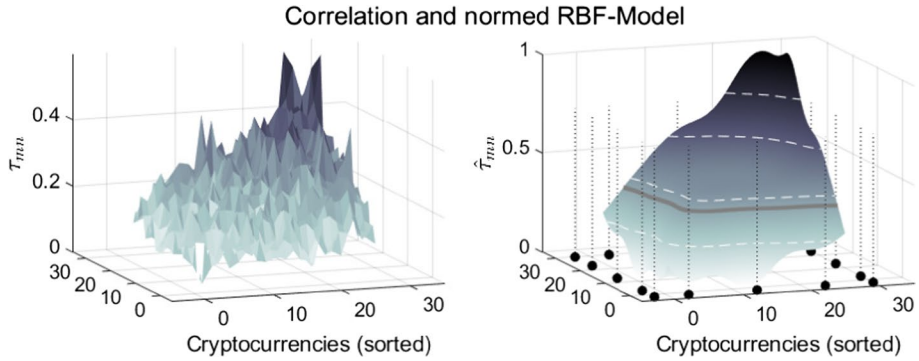

Fig. 2 Critical step of the modeling process for the segmentation of the CC market using correlations

However, this dependency can be avoided by considering all metrics and selecting those CCs as the market core that are contained in the intersection of all metrics. This approach is illustrated in the last column of Table 3. In this column, all CCs belonging to the core according to all metrics are marked with **C**. This knowledge provides a decisive advantage in asset management when integrating a certain share of CCs in a portfolio.

Table 2 shows examples of CCs as representatives of the identified core CCs (XPM and ZET) and the identified satellite CCs (DMD and FRC). In practice, a simple portfolio could be constructed as follows. For example, 5–10 CCs with high liquidity and market depth similar to XPM/ZET that belong to the core are selected from the entire dataset. These CCs form the core investment. Individual CCs can then be selected from among the satellites, which can be expected to offer a higher return if the risk is higher. In the first case, the tracking error can be determined in relation to the core, and that in the second case can be determined in relation to the overall market. In any case, the composition is optimized taking a specified limitation of the tracking error into account. Continuous control of the tracking error and tactical readjustment of the weights leads to tracking of the core (first case) or the overall market (second case), whereby the tracking error specified by the institutional investor is followed.

Robustness

Core–satellite identification using correlations

Our developed segmentation process takes into account the specifics of the CC market. This is achieved by explicitly considering the dynamics of the statistical parameters and by simultaneous utilizing the return, the standard deviation and especially the tail parameter.

Although correlation analysis is a common tool for determining similar and dissimilar assets (see, e.g., Heaney and Hooper (2001), Bekaert et al. (2011) and the corresponding literature cited therein), it is not possible to use several parameters simultaneously (e.g., the three-dimensional vector $(\langle r\rangle, s, \alpha)'$) and to consider their dynamics. This is exactly the problem: we lose important information that is necessary for the successful segmentation process, which we illustrate below.

At the beginning, we calculate the correlations of the returns over the entire period and end up with a 27×27 matrix. Here, we use Kendall's correlation coefficient because it is robust to a violation of the normal distribution assumption. Based on the correlation matrix, the steps of the segmentation process described in Sect. 4 are carried out. The unsorted and sorted correlation matrices (not shown here) entail the problem that the pairwise correlations with values of one are located on the diagonal. First, we perform the segmentation process without preprocessing the diagonals and find that no analysis is possible due to their highly peaked structure. Nevertheless, to enable carrying out the procedure, for the second step, we replace the entries on the diagonal by the mean values of the neighboring cells (a kind of presmoothing step). This at least succeeds in smoothing the peaked structure by the radial basis functions so that an evaluation is possible. Figure 2 shows the smoothing of the modified correlation matrix, with the correlations ranging from 0.0491 to 0.5981 in the left part, whereby they are normalized for modeling reasons in the right part. A brighter area indicates a lower correlation and vice versa.

Note that higher DTW distances are associated with greater dissimilarity, whereby a higher correlation indicates greater similarity. In this respect, the core of the CC market is now located in the corner opposite to the "DTW core," i.e., in the upper-right corner of the right part of Fig. 2. However, Fig. 2 clearly shows that the identification of a core is not possible at high correlations because no "plateau" of similar CCs emerges. Rather, the smoothed area falls continuously, so a suitable threshold d_{bound} cannot be reliably determined.

Table 4 Correlation analysis of the entire period

	CCs	Range	Mean	Median	Mean p value	Median p value
Complete	27	0.0491–0.5981	0.2223	0.2180	0.0227	0.0002
Core	19	0.0780–0.5981	0.2416	0.2308	0.0097	0.0001
Satellites	8	0.0491–0.3731	0.1751	0.1480	0.0638	0.0129

Kendall's correlation coefficient is used, and the calculated p values are based on a two-sided significance test

Return correlations of the core and satellites

Another issue to be discussed comprises the correlations of CCs within the core, the correlations of CCs among the satellites and the potential differences between them. These correlations can provide important insights into whether the intended market segmentation has been achieved. Note that we analyze the correlations of the core and satellite CCs determined using our segmentation process with DTW distances.

Although we consider three parameters and their dynamics over six years and do not base the market segmentation on correlations, the correlations of the core CCs should be higher than those of the satellite CCs. This is the intended result, and it would provide evidence for successful market segmentation because we expect statistically homogeneous CCs in the core that are heterogeneous to the satellites.

While CCs are less correlated with traditional asset classes (Brière et al. 2015; Kuo Chuen et al. 2017; Liu and Tsyvinski 2021), the correlation of CCs is more pronounced (Dorfleitner and Lung 2018), which is also confirmed by our correlation analyses, as shown in Table 4.

For the whole sample, we report a range of correlations from 0.0491 to 0.5981 with a mean of 0.2223 and a median of 0.218. The average p value is 0.0227, and the median is 0.0002, which implies statistical significance at least at the 5% significance level.

With regard to considering the core and satellite CCs, the calculated correlations provide evidence for the intended segmentation of the CC market. The core CC correlations range from 0.078 to 0.5981 with a mean of 0.2416 and a median of 0.2308. The p value is 0.0097 on average and 0.0001 at the median, so it is statistically significant at least at the 1% significance level. In contrast, the correlations of the satellite CCs show a range from 0.0491 to 0.3731 with a mean of 0.1751 and a median of 0.148. The p value is 0.0638 at the mean and 0.0129 at the median, which indicates statistical significance at least at the 10% level. The initial assessment indicates a higher correlation range of CCs in the core compared to those among the satellites, which is our desired result.

Furthermore, we check whether the correlations between the core CCs differ significantly from those between the satellite CCs. This is accomplished by using Welch's parametric t test and the nonparametric Mann–Whitney U test. We perform one-sided tests because we assume the correlations between the core CCs are higher than those between the satellite CCs. The first test proves the null hypothesis that the means of the correlations of the core CCs and satellite CCs are equal, whereas the second test examines this for the central tendencies. The alternative hypothesis states that the mean or central tendencies of the correlations of the core CCs are higher than those of the satellite CCs. The respective null hypothesis is rejected at the 1% significance level for both tests, with a p value of 0.0007 for Welch's t test and 0.0002 for the Mann–Whitney U test.

Thus, we can conclude that the correlations of the core CCs are higher than those of the satellite CCs. This finding provides evidence for the intended segmentation of the CC market.

Conclusion

In our study, we show how a general, purely data-driven process can be utilized successfully to separate an investment universe into similar assets (core) and dissimilar assets (satellites). We verify the feasibility of this approach and outline the necessary sequence of steps for the segmentation in detail. Using the example of the modern CC asset class, we separate the investment universe into similar CCs (core) and dissimilar CCs (satellites) as the residual share. In addition, we discern interesting results specifically concerning the CCs.

The results in Table 3 show that of the 27 CCs studied, 19 CCs belong to the core, whereas 8 CCs represent the satellites. The question raised at the beginning of whether Bitcoin actually represents the "hard core" of the CC market can be answered in a differentiated manner. It turns out that Bitcoin is part of the core, but not in its center. Recall that the threshold $d_{\text{bound}} = 0.357$ is used to draw a contour (white) in the modeled height profile on the lower-right part of Fig. 1, which delimits an upper block matrix. This block matrix represents the core of the CC market, wherein Bitcoin lies just inside and thus narrowly belongs to the core. Therefore, it is not part of the "hard core" of very similar CCs with low normalized DTW distances but is comparatively dissimilar to other core CCs. Thus, a dominant role

within the CC market, as appears in other analyses, cannot be confirmed by our findings.

Although our paper presents a methodological approach that aims to explain our proposed segmentation process, we nevertheless briefly discuss an exemplary use case of the results. In general, the market segmentation result can be used in portfolio management by institutional investors to track the core market with a few selected CCs in a tracking error approach, as described in more detail in Sect. 4. For example, 5–10 CCs of the core with high liquidity and market depth are selected to form the core investment. To increase returns, a higher-level management approach can then be used to build up individual positions in CCs that belong to the satellite, thus implementing a core–satellite portfolio.

One potential challenge for this approach might lie in liquidity problems, especially in the case of smaller altcoins (other than Bitcoin). However, studies indicate that CCs typically make up a smaller component of a portfolio of traditional assets, mitigating this issue (Dorfleitner and Lung 2018; Schmitz and Hoffmann 2021). In addition, methods such as the liquidity bounded risk–return optimization (LIBRO) approach by Trimborn et al. (2020) exist that can be used to optimize portfolios under liquidity constraints. Furthermore, it is conceivable that liquid CCs are incorporated into the core so that they can be purchased anyway without the fear of liquidity restrictions. Beyond that, it can be assumed that the development of the CC market will make it suitable for larger investment volumes in the future.

As already mentioned, the proposed method is not limited to CCs. A suitable market segmentation in other asset classes is conceivable, as well. The advantages of the product-based implementation of a topic-centered, combined 'core–satellite & tracking-error' strategy in the private or institutional investor segment is reserved for further studies.

Appendix 1: Stable distribution—the tail parameter α

The analyses in Börner et al. (2021) show that a family of SDIs is the most promising option for modeling the distribution of the returns of CCs. Therefore, this family of functions is also used in the present study and is introduced here in detail. Several different parametrizations exist for the SDI. In the following formulation, we follow the presentation and the parametrization of the SDI described in Nolan (2020, Def. 1.4 therein). SDIs are a class of probability distributions suitable for modeling heavy-tailed and skewed distributions. A linear combination of two independent, identically distributed stable random variables has the same distribution as the individual variables. A random variable X has the SDI $S(\alpha, \beta, \gamma, \delta)$ if its characteristic function is given by:

$$\mathrm{E}\left[\exp\left(\mathrm{i}tX\right)\right] = \begin{cases} \exp\left(\mathrm{i}\delta t - |\gamma t|^{\alpha}\left[1 + \mathrm{i}\beta\,\mathrm{sign}(t)\,\tan\left(\frac{\pi\alpha}{2}\right)\left(|\gamma t|^{1-\alpha} - 1\right)\right]\right) & \alpha \neq 1 \\ \exp\left(\mathrm{i}\delta t - |\gamma t|\left[1 + \mathrm{i}\beta\,\mathrm{sign}(t)\,\frac{2}{\pi}\ln\left(|\gamma t|\right)\right]\right) & \alpha = 1 \end{cases} \tag{1}$$

The first parameter $0 < \alpha \leq 2$ is called the shape parameter and describes the tail of the distribution. Sometimes this parameter is also denoted as the *tail parameter*, *index of stability* or *characteristic exponent*. The second parameter $-1 \leq \beta \leq +1$ is the skewness parameter. If $\beta = 0$, then the distribution is symmetric; otherwise, it is left-skewed ($\beta < 0$) or right-skewed ($\beta > 0$). When α is small, the skewness of β is significant. As α increases, the effect of β decreases. Further, $\gamma \in \mathbb{R}^{+}$ is called the scale parameter, and $\delta \in \mathbb{R}$ is the location parameter. For the special case of $\alpha = 2$, the characteristic function given by Eq. (1) reduces to $\mathrm{E}\left[\exp\left(\mathrm{i}tX\right)\right] = \exp\left(\mathrm{i}\delta t - (\gamma t)^2\right)$ and becomes independent of the skewness parameter β, and the SDI is equal to a normal distribution with mean δ and standard deviation $\sigma = \sqrt{2}\gamma$. This is an important property for portfolio theory, for example, when considering multivariate distributions. The rationale is that it is basically possible to model the normally distributed components of a random vector with the same function class. In the main part, the tail parameter α is estimated for each year under consideration on a weekly return basis and used as input data for the DTW distance analysis.

Appendix 2: Modeling with radial basis functions

In many scientific fields (Powell 1977; Poggio and Girosi 1990; Sahin 1997; Biancolini 2017), radial basis functions are used to carry out a function approximation of the following form:

$$y(\mathbf{x}) = \sum_{m=1}^{M} \lambda_m \, \phi(\|\mathbf{x} - \mathbf{x}_m\|) \tag{2}$$

where $y(\mathbf{x})$ is a one-dimensional function depending on $\mathbf{x} \in \mathbb{R}^n$. The function $y(\mathbf{x})$ is modeled as a sum of M radial basis functions, each centered at a different center $\mathbf{x}_m$ and weighted with an appropriate coefficient λ_m. The real value of every radial basis function is strictly positive and depends only on the distance between the point $\mathbf{x}$ and the center $\mathbf{x}_m$. The distance $r = \|\mathbf{x} - \mathbf{x}_m\|$ is determined in a previously defined norm. We only use the Euclidean distance as the norm in our analyses. To model and reconstruct the height profile over the distance matrix $\mathbf{D}$ in Sect. 4, we use radial basis functions of Gaussian type:

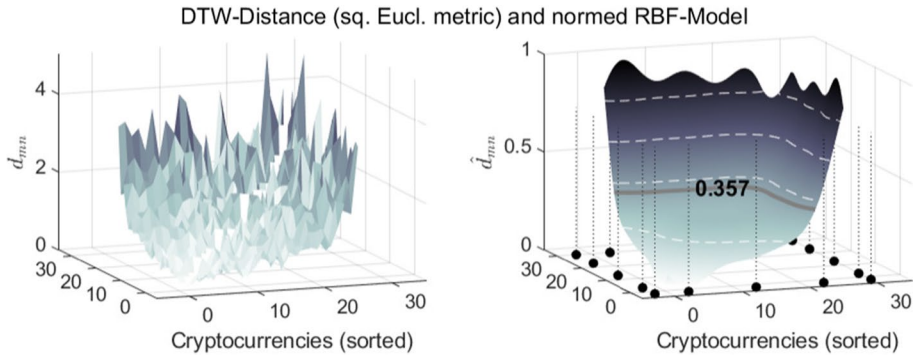

Fig. 3 An explicit and detailed representation of the modeling process leading to the segmentation of the CC market in Sect. 4

$$\phi(r) = \exp\left(-ar^2\right) \tag{3}$$

with infinite support and a positive shape parameter a. The latter can also be interpreted as the effective range of the radial basis function. If R denotes the distance between two different centers and $0 < p < 1$ denotes the desired residual effect at the next center, then the area of effect can be set by a due to $a = -R^{-2} \ln p$. The parameter vector $\boldsymbol{\lambda}$ is determined using a least squares approach. In some applications, we find that least squares fit encounters problems with ill-conditioned matrices. Therefore, we extend our Lagrange function to be minimized by a regularization term. The latter term is also referred to as cost-functional and takes into account the costs of the deviation from a smooth function. The theoretical foundations of this approach can be traced back to earlier work by Tikhonov (1943, 1963). The implemented regularization procedure is currently standard (Poggio and Girosi 1990), cf. also Sahin (1997); Biancolini (2017) and the substantial amount of literature cited therein. Hence, we set the Lagrange function

$$\mathcal{L} = \sigma^2 + \alpha \boldsymbol{\lambda}' \boldsymbol{\lambda} \tag{4}$$

where σ^2 is the squared error between the modeled $(\hat{y}_i)$ and sample (y_i) values for $i = 1, \ldots, N$ with sample length N. Furthermore, α is a positive real number called the regularization parameter. If $\alpha \to 0$, the problem is unconstrained, and the resulting model can be completely determined from the sample. On the other hand, if $\alpha \to \infty$, the a priori desired smoothness of the resulting model dominates and leads to a highly smooth function, nearly flat in the limit and almost independent of the measured sample. Finally, the solution to the minimization problem given by Eq. 4 is

$$\boldsymbol{\lambda} = \left(\boldsymbol{\Phi} + \frac{\alpha}{N}\mathbf{E}\right)^{-1} \mathbf{v}. \tag{5}$$

Abbreviating $\phi_{\text{im}} = \phi(\|\mathbf{x}_i - \mathbf{x}_m\|)$ as the value of the mth radial basis function at the sample point $\mathbf{x}_i$ for $i = 1, \ldots, N$ and given the output y_i, then the vector $\mathbf{v} = \left(\langle y_i \phi_{\text{im}} \rangle\right)_{m=1,\ldots,M}$ and matrix $\boldsymbol{\Phi} = \left(\langle \phi_{ik} \phi_{\text{im}} \rangle\right)_{k,m=1,\ldots,M}$, where $\langle \cdot \rangle$ denotes the sample average. Further, $\mathbf{E}$ denotes the identity matrix in $\mathbb{R}^{M \times M}$. In practice, in very few applications do we assign successively increasing values $0 \leq \alpha < 100$ to the regularization parameter until the observable local roughness or heavily peaked structure of the modeled surface vanishes. We observe that the height profile of the distance matrix $\mathbf{D}$ remains well reconstructed, but modeling the absolute height worsens with the increasing influence of regularization. The modeling properties improve if a constant term is incrementally added to the model given in Eq. (2). The solution expressed by Eq. (5) does not change if the number of radial basis functions is increased only by 1, $M \to M + 1$, and the value identical to 1 is assigned to the first radial basis function, $\phi_1 = 1$ for all $\mathbf{x} \in \mathbb{R}^n$, while the changes are considered in the elements of the vector $\mathbf{v}$ and matrix $\boldsymbol{\Phi}$. The majority of the analyses could be carried out with $\alpha = 0$ and led without regularization procedures to very good results. For the results shown in the main part, we do not apply the regularization procedure. In Fig. 3, an example of the modeling process with radial basis functions and $\alpha = 0$ is shown. The left figure shows the rough and peaked height structure d_{mn} above the DTW distance matrix $\mathbf{D}_{\text{SE}}$ calculated in Sect. 4. It is the three-dimensional counterpart of the upper-right panel of Fig. 1 viewed from the upper-left corner along the main diagonal. The graphic on the right shows the surface of the standardized height structure $\hat{d}_{mn}$ modeled with radial basis functions. Some contour lines (dashed white) are also shown, each with a distance of 0.2 units. The contour of the threshold $d_{\text{bound}} = 0.357$ for the squared Euclidean metric is shown in light gray, cf. Sect. 4. The bullet points and the corresponding vertical dashed lines illustrate the centers and positions of the radial basis functions, respectively.

Acknowledgements We thank CoinMarketCap for generously providing the cryptocurrency time series data for our research.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

Aghabozorgi, S., A. Seyed Shirkhorshidi, and T. Ying Wah. 2015. Time-series clustering—A decade review. *Information Systems* 53: 16–38.

Amenc, N., P. Malaise, and L. Martellini. 2012. Revisiting core-satellite investing—A dynamic model of relative risk management. Working Paper.

Aslanidis, N., A.F. Bariviera, and C.S. Savva. 2020. Weekly dynamic conditional correlations among cryptocurrencies and traditional assets. Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3550879

Baur, D.G., K. Hong, and A.D. Lee. 2018. Bitcoin: Medium of exchange or speculative assets? *Journal of International Financial Markets, Institutions and Money* 54: 177–189.

Bekaert, G., C.R. Harvey, C.T. Lundblad, and S. Siegel. 2011. What segments equity markets? *Review of Financial Studies* 24 (12): 3841–3890.

Biancolini, M.E. 2017. *Fast radial basis functions for engineering applications*. Cham: Springer.

Börner, C.J., I. Hoffmann, L.M. Kürzinger, and T. Schmitz. 2021. On the return distributions of a basket of cryptocurrencies and subsequent implications. Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3851563

Brauneis, A., and R. Mestel. 2018. Price discovery of cryptocurrencies: Bitcoin and beyond. *Economics Letters* 165: 58–61.

Brière, M., K. Oosterlinck, and A. Szafarz. 2015. Virtual currency, tangible return: Porfolio diversification with bitcoin. *Journal of Asset Management* 16 (6): 365–373.

Caporale, G.M., L. Gil-Alana, and A. Plastun. 2018. Persistence in the cryptocurrency market. *Research in International Business and Finance* 46: 141–148.

Dorfleitner, G., and C. Lung. 2018. Cryptocurrencies from the perspective of euro investors: A re-examination of diversification benefits and a new day-of-the-week effect. *Journal of Asset Management* 19 (7): 472–494.

ElBahrawy, A., L. Alessandretti, A. Kandler, R. Pastor-Satorras, and A. Baronchelli. 2017. Evolutionary dynamics of the cryptocurrency market. *The Royal Society Open Science* 4 (11): 170623.

Fry, J., and E.T. Cheah. 2016. Negative bubbles and shocks in cryptocurrency markets. *International Review of Financial Analysis* 47: 343–352.

Gandal, N., J.T. Hamrick, T. Moore, and T. Oberman. 2018. Price manipulation in the bitcoin ecosystem. *Journal of Monetary Economics* 95: 86–96.

Giorgino, T. 2009. Computing and visualizing dynamic time warping alignments in r: The dtw package. *Journal of Statistical Software* 31 (7): 1–24.

Glas, T.N. 2019. Investments in cryptocurrencies: Handle with care! *The Journal of Alternative Investments* 22 (1): 96–113.

Hackbusch, W. 2015. *Hierarchical matrices: Algorithms and analysis*. Vol. 49 of Springer series in computational mathematics. Heidelberg: Springer.

Hayes, A.S. 2017. Cryptocurrency value formation: An empirical study leading to a cost of production model for valuing bitcoin. *Telematics & Informatics* 34 (7): 1308–1321.

Heaney, R., and V. Hooper. 2001. Regionalism, political risk and capital market segmentation in international asset pricing. *Journal of Economic Integration* 16 (3): 299–312.

Kuo Chuen, D.L., L. Guo, and Y. Wang. 2017. Cryptocurrency: A new investment opportunity? *The Journal of Alternative Investments* 20 (3): 16–40.

Liu, X., X.-H. Zhu, P. Qiu, and W. Chen. 2012. A correlation-matrix-based hierarchical clustering method for functional connectivity analysis. *Journal of Neuroscience Methods* 211 (1): 94–102.

Liu, Y., and A. Tsyvinski. 2021. Risks and returns of cryptocurrency. *Review of Financial Studies* 34 (6): 2689–2727.

Majoros, S., and A. Zempléni. 2018. Multivariate stable distributions and their applications for modelling cryptocurrency-returns. Working paper 2018. arXiv:1810.09521v1

Markowitz, H.M. 1952. Portfolio selection. *The Journal of Finance* 7 (1): 77–91.

Methling, F., and R. von Nitzsch. 2019. Thematic portfolio optimization: Challenging the core satellite approach. *Financial Markets and Portfolio Management* 33 (2): 133–154.

Nolan, J.P. 2020. *Univariate stable distributions: Models for heavy tailed data*, 1st ed. Springer series in operations research and financial engineering. Cham: Springer.

Poggio, T., and F. Girosi. 1990. Regularization algorithms for learning that are equivalent to multilayer networks. *Science (New York, NY)* 247 (4945): 978–982.

Powell, M.J.D. 1977. Restart procedures for the conjugate gradient method. *Mathematical Programming* 12 (1): 241–254.

Rivin, I., C. Scevola, and R. Davis. 2017. Cci30—The crypto currencies index. https://cci30.com/

Sahin, F., 1997. A radial basis function approach to a color image classification problem in a real time industrial application. Thesis. Virgina Tech, Blacksburg, VA. http://hdl.handle.net/10919/36847

Sakoe, H., and S. Chiba. 1978. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26 (1): 43–49.

Schmitz, T., and I. Hoffmann. 2021. Re-evaluating cryptocurrencies' contribution to portfolio diversification—A portfolio analysis with special focus on German investors. Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3625458

Sigaki, H.Y.D., M. Perc, and H.V. Ribeiro. 2019. Clustering patterns in efficiency and the coming-of-age of the cryptocurrency market. *Scientific Reports* 9 (1): 1440.

Tikhonov, A.N. 1943. On the stability of inverse problems. *Doklady Akademii Nauk SSSR* 39 (5): 195–198.

Tikhonov, A.N. 1963. Solution of incorrectly formulated problems and the regularization method. *Doklady Akademii Nauk SSSR* 151: 501–504.

Trimborn, S., and W.K. Härdle. 2018. Crix an index for cryptocurrencies. *Journal of Empirical Finance* 49: 107–122.

Trimborn, S., M. Li, and W.K. Härdle. 2020. Investing with cryptocurrencies–A liquidity constrained investments approach. *Journal of Financial Econometrics* 18 (2): 280–306.

Christoph J. Börner graduated in business administration from the University of Cologne, Germany, in 1990. He also received the Ph.D. degree from University of Cologne in 1994. In his doctoral thesis, as in his habilitation (2000), he dealt with a topic from banking. In 2001, he was Visiting Professor at the University of Mannheim, Germany. In 2002, he was appointed to the Chair of Business Administration, esp. Financial Services, at Heinrich Heine University Düsseldorf, Germany. His research interests include the structure and strategies of financial services firms, cryptocurrencies, and risk management.

Ingo Hoffmann graduated in theoretical physics in 1993 from the University of Wuppertal, Germany. In 1994, he moved to the University of Dortmund, Germany, and received the Ph.D. degree in chemical engineering in 1998. He works for DZ BANK AG, Germany, since 1998 and is responsible for capital market topics as well as asset and risk management. Since 2017, he has also been conducting research at the Chair of Business Administration, esp. Financial Services at Heinrich Heine University Düsseldorf, Germany, on asset and risk management topics related to institutional investors.

Jonas Krettek received a Master's degree in business administration from Heinrich Heine University Düsseldorf, Germany, in 2020. During his studies, he completed various internships at banks and a semester abroad at the Università degli studi di Trieste, Italy. Since October 2020, he has been a research associate and doctoral candidate at the Chair of Business Administration, esp. Financial Services at Heinrich Heine University Düsseldorf.

Tim Schmitz received a Master's degree in business administration from Heinrich Heine University Düsseldorf, Germany, in 2017. Since October 2017, he has been a research associate and doctoral candidate at the Chair of Business Administration, esp. Financial Services at Heinrich Heine University Düsseldorf.