On the return distributions of a basket of cryptocurrencies and subsequent implications

Christoph J. Börner, Ingo Hoffmann, Lars M. Kürzinger, Tim Schmitz

Article - Version of Record

Suggested Citation:
Börner, C. J., Hoffmann, I., Kürzinger, L., & Schmitz, T. (2025). On the return distributions of a basket of cryptocurrencies and subsequent implications. Research in Economics, 79(1), Article 101028. https://doi.org/10.1016/j.rie.2025.101028

Research in Economics 79 (2025) 101028

Contents lists available at ScienceDirect

Research in Economics

journal homepage: www.elsevier.com/locate/rie

Research paper

On the return distributions of a basket of cryptocurrencies and subsequent implications☆

Christoph J. Börner, Ingo Hoffmann, Lars M. Kürzinger *, Tim Schmitz

Chair of Financial Services, Faculty of Business Administration and Economics, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

ARTICLE INFO

JEL classification:
C12
C13
C43
E22

Keywords:
Body-/Tail-models
Cryptocurrencies
Index construction
Market segmentation
Statistical tests

ABSTRACT

This study evaluates the risk associated with capital allocation in cryptocurrencies (CCs) using a basket of 27 CCs and the CC index EWCI-. We apply basic statistical tests to model the body distribution of CC returns. Consistent with prior research, the stable distribution (SDI) is the most suitable model for the body distribution. However, due to less favorable properties in the tail area for high quantiles, the generalized Pareto distribution is employed. A combination of both distributions is utilized to calculate Value at Risk and Conditional Value at Risk, revealing distinct risk characteristics in two subgroups of CCs.

1. Introduction

Following the financial crisis of 2007 and the following period of extreme uncertainty and volatility, trust in the financial system and its institutions, such as central banks and their monetary policies, were shattered (Kaya Soylu et al., 2020; Bouri et al., 2017a).

Against this background, the market for cryptocurrencies (CCs) started to emerge in 2009 with the development of Bitcoin by Nakamoto (2008). This innovative peer-to-peer electronic cash system is not accountable to any subordinating institution, but is managed and controlled by its own community using the blockchain technology. Furthermore, the anonymity and security of transactions represent another noteworthy feature Bitcoin promises its users (Kakinaka and Umeno, 2020), resulting in increasing trading volumes and prices (Corbet et al., 2019). This development raises questions for both investors and regulators alike regarding the CC market's characteristics and risk profile, which need to be answered for CCs to become an investable asset class for a wide range of investors (Gkillas and Katsiampa, 2018; Majoros and Zempléni, 2018; Osterrieder et al., 2017) and to provide guidance for risk management in general (Kakinaka and Umeno, 2020). In this context, this study investigates the question of which family of distribution functions suitably and most accurately models the returns of CCs. We answer this question using a novel approach to separate a distribution's body from its tail proposed by Hoffmann and Börner (2021). By doing so, we are able determine the risk and statistical properties associated with CCs and provide valuable implications for portfolio management and regulators alike.

Although the technical properties of CCs are well understood, CCs' behavior remains to be fully comprehended and analyzed. Thus far, many economic studies have focused on Bitcoin and other prominent CCs such as Ethereum and Ripple since they dominate the

☆ We thank Coinmarketcap.com for generously providing the cryptocurrency time series data for our research.

* Corresponding author.
E-mail addresses: Christoph.Boerner@hhu.de (C.J. Börner), Ingo.Hoffmann@hhu.de (I. Hoffmann), Lars.Kuerzinger@hhu.de (L.M. Kürzinger), Tim.Schmitz@hhu.de (T. Schmitz).

https://doi.org/10.1016/j.rie.2025.101028
Received 11 June 2024; Accepted 6 January 2025
Available online 11 January 2025

CC market due to their high proportion of total market capitalization (Glas, 2019). In this regard, Baur et al. (2018) and Glas (2019) find Bitcoin and other CCs to be uncorrelated with traditional assets in times of financial distress. Additionally, Gkillas et al. (2018), demonstrate that CC's behavior also differs from that of fiat currencies. Furthermore, various studies analyze certain characteristics of CCs, e.g., volatility (Polasik et al., 2015; Balcilar et al., 2017), diversification issues (see, i.a., Brière et al., 2015; Selgin, 2015; Corbet et al., 2018; Schmitz and Hoffmann, 2020) and safe haven properties (Bouri et al., 2017b; Urquhart, 2018).[1] However, to evaluate the corresponding market risk and to completely understand the whole CC market (with its typical features), we follow a literature strand of studies analyzing return distributions of different selected CCs. As numerous studies observe nongaussian behavior and heavy tails in return distributions (Osterrieder et al., 2017; Gkillas et al., 2018; Gkillas and Katsiampa, 2018), a distribution model accounting for these observed characteristics ought to be implemented. To account for such characteristics (Majoros and Zempléni, 2018; Kakinaka and Umeno, 2020) use stable distributions (SDIs) in their recent studies. This paper is based on these results. The findings there are reproduced here in an expanded database and possibilities are shown to mitigate the weaknesses of the concept in the risk assessment, especially in the tail area. However, Kakinaka and Umeno (2020) observe the SDI to be unable to efficiently grasp the heavy tails of the analyzed return distributions in all scenarios in comparison to other possible distributions. Given this background, the generalized Pareto distribution (GPD), a statistical distribution that appears to more accurately model heavy tail properties, is used in further studies (Gkillas et al., 2018; Gkillas and Katsiampa, 2018). Following the approach presented in Hoffmann and Börner (2021), we therefore attempt to use a combination of both described distributions. Analytically discovering the beginning of the tail of the analyzed return distributions enables us to divide the given data into a body and a tail, and we implement a different distribution for each. For the first sample, containing the tail of potential losses, we implement the GPD, since the literature and our tests show its goodness of fit for estimating tail values. For the remaining body of our data, we apply the SDI because its fit outperforms that of other possible distributions.

The aim of our study is to add to the existing literature by implementing a novel approach intended to achieve higher quality modeling of the return distributions observed in the CC market. Furthermore, thus far, most studies have merely been concerned with analyzing characteristics of Bitcoin or the most prominent CCs, e.g., Ethereum, Ripple and Litecoin (Baur et al., 2018; Bouri et al., 2017b; Osterrieder et al., 2017; Gkillas et al., 2018; Gkillas and Katsiampa, 2018; Majoros and Zempléni, 2018; Kakinaka and Umeno, 2020). Therefore, most studies do not consider the entirety of the CC market, which, as Glas (2019) notes, might lead to potential bias. Hence, we join (Glas, 2019; ElBahrawy et al., 2017; Schmitz and Hoffmann, 2020) in an attempt to provide a broader overview of the CC market. By doing so, we address two existing gaps in literature identified by Corbet et al. (2019) in their extensive literature review. Namely, we extend the number and size of the analyzed data in an attempt to analyze CCs as an asset class. Furthermore, our research provides practical relevance in the form of an improved risk assessment. By analytically separating a distribution's body from its tail and implementing different distributions, we are able to more precisely estimate risk measures in form of the value at risk and the conditional value at risk, both of which are important for regulators and investors alike.

The remainder of this paper is structured as follows: In Section 2, we present and describe the data used in our following analysis. In Section 3, we perform a series of statistical analyses and tests that lead us to the family of SDIs as the best model choice for the body of the CC return distribution. Based on these findings, Section 4 is concerned with the assessment of the tail risk inherent in an investment in a single CC or the basket aggregated in the $EWCI^{-}$. We compare both methods of risk assessment at high quantiles. We first use the body model and then adapt the GPD as a tail model for risk assessment in terms of the value at risk and the conditional value at risk. The last section summarizes our most import results and provides an overview of future research topics.

2. Data

As a starting point of our data collection, we follow different studies in the literature by extracting daily prices of selected CCs. Central studies were, e.g., Fry and Cheah (2016), Hayes (2017), Brauneis and Mestel (2018), Caporale et al. (2018), Gandal et al. (2018) and Glas (2019). The relevant data originates from the website coinmarketcap.com (URL: https://coinmarketcap.com) and is downloaded for each of the $n = 66$ CCs from the Coinmarketcap Market Cap Ranking (reference date: 2014-01-01), see Table 1, assuming an observation period from 2014-01-01 to 2019-06-01, as it was originally done by Schmitz and Hoffmann (2020) and – based on this study – also in later derivative works (e.g. by Börner et al. (2022)). We follow both studies and replicate their way of sample selection (as described above) and data editing (as described below) and thus end up with a replicated version of their original dataset:

We aim to consider as many CCs as possible from this original sample for our final analyses in order to illustrate a preferably high share of the CC market. Nonetheless, we need to exclude all those CCs in the dataset with longer data gaps (here: with five or more consecutive missing observations). By using a Last Observation Carried Forward (LOCF) procedure, as in Trimborn et al. (2020) and Schmitz and Hoffmann (2020), we are then able to consider all the remaining CCs (means: those CCs with smaller data gaps) in our final dataset.

After conducting these steps, we also end up with 27 remaining CCs, (e.g. as in Börner et al. (2022) and Schmitz and Hoffmann (2020)). These CCs are depicted in Table 1. In the next steps, we are again guided by the procedure described in Schmitz and Hoffmann (2020): Using the daily USD–EUR exchange rates collected from Thomson Reuters Eikon, the CC price data (originally denoted in USD) is converted to EUR. Furthermore, the resulting daily prices are converted to weekly prices in a subsequent step to

[1] For a more extensive literature review, see Corbet et al. (2019).

Table 1
Derivation of the dataset under study.
Source: The table is based on an original table by Schmitz and Hoffmann (2020) for an identical market snapshot (full sample and reduced subsamples), which was also created with cryptocurrency market data by CoinMarketCap.com. The subsampled version of this original dataset by Schmitz and Hoffmann (2020) was also used in the derivative work of Börner et al. (2022).

Cryptocurrency Sample at 2014-01-01: $n = 66$ CCs
(as in: Schmitz and Hoffmann (2020))

→ **Thereof:** $n = 27$ CCs considered in the final dataset (printed bold)
(as in: Schmitz and Hoffmann (2020) and Börner et al. (2022))

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

→ **Thereof:** $n = 39$ CCs excluded from the final dataset due to the data gap critereon
(as in: Schmitz and Hoffmann (2020) and Börner et al. (2022))

CC	ID	CC	ID	CC	ID
Argentum	ARG	Elacoin	ELC	Luckycoin	LKY
AsicCoin	ASC	EZCoin	EZC	MemoryCoin	MMC
BBQCoin	BQC	FastCoin	FST	MinCoin	MNC
BetaCoin	BET	Fedoracoin	TIPS	NetCoin	NET
BitShares PTS	PTS	Franko	FRK	Noirbits	NRB
Bullion	CBX	Globalcoin	GLC	Orbitcoin	ORB
ByteCoin	BTE	GrandCoin	GDC	Philosopher Stones	PHS
CatCoin	CAT	HoboNickels	HBN	Phoenixcoin	PXC
Copperlark	CLR	I0Coin	IOC	SexCoin	SXC
CraftCoin	CRC	Ixcoin	IXC	Spots	SPT
Datacoin	DTC	Joulecoin	XJO	StableCoin	SBC
Devcoin	DVC	Junkcoin	JKC	Tickets	TIX
Earthcoin	EAC	LottoCoin	LOT	TigerCoin	TGC

avoid any possible weekday biases. Moreover, we do not only use our weekly CC prices on an individual CC level, but also calculate an equally weighted CC index (EWCI), following Schmitz and Hoffmann (2020) to get an aggregated perspective. As we exclude more CCs than the beforementioned study, we will call this index $EWCI^{-}$ for a more precise distinction.

Using their respective price data, we follow Börner et al. (2022) and finally calculate logarithmic returns (simply abbreviated as 'returns' in the remainder of this study) for all the considered individual CCs and the aggregated $EWCI^{-}$ index.

3. The return distribution of cryptocurrencies

For an initial classification of CCs, key simple statistics from the standard repertoire of empirical statistics are used below. The description and evaluation of additional statistical properties of CCs, for example value at risk or lower partial moments, are carried out using a suitable distribution function.

Our results show that the family of SDIs is a suitable model for the examined returns of CCs. Hence, this family of distributions is used in Section 4 for a more in-depth analysis of the statistical properties of CCs.

3.1. Determination of basic key statistical figures of the cryptocurrencies

Standard procedures lead us to estimates of the set of basic key statistical figures: mean $\hat{\mu}$, variance $\hat{\sigma}^2$ and bandwidth (Table 2). The returns scatter strongly around a center close to zero. While the variance and thus the standard deviation indicate leptokurtic behavior and therefore a concentration of returns, the sometimes considerable bandwidth indicates a strong blur of returns over a wide measurement range. This leads to the preliminary conclusion that the returns of CCs in the middle value range follow a concentrated distribution that has pronounced fat tails in the outer areas. In particular, the large variance (∼7%) and the bandwidth (∼300%) of CCs clearly show the completely different character of CCs compared to traditional asset classes. Comparable values on a weekly basis for the traditional asset classes (stocks, bonds, real estate, etc.) fall in the range of ∼0.02% (variance) or ∼0.1% (bandwidth) on average. Hence, in comparison there is considerable risk associated with CCs.

Additionally, in this step, we use the Hartigan dip test (Hartigan and Hartigan, 1985) to test the null hypothesis H0 that the empirical distribution is unimodal and symmetric. Thus, for each dataset, the Hartigan dip statistics (HDS) are calculated and

Table 2
Basic key statistical figures and tests on unimodality and symmetry. Units in percent and boolean; see text. Note: Although Schmitz and Hoffmann (2020) calculate similar descriptive statistics for the same cryptocurrency sample, different results occur due to the usage logarithmic returns in this work.

Crypto	*Mean*	*Variance*	Bandwidth		HDS test		SIG test	
ID	$\hat{\mu}$	$\hat{\sigma}^2$	min	max	H0	*p*-value	H0	*p*-value
EWCI$^-$	0.3	1.6	−40.8	37.7	0	79.2	0	10.8
ANC	−1.0	14.3	−241.4	162.9	0	99.0	0	17.1
BTB	−0.7	11.7	−201.2	150.4	0	97.4	0	37.2
BTC	0.9	1.0	−30.4	42.2	0	99.8	0	59.2
CSC	−1.2	30.5	−697.9	169.0	0	92.2	0	51.2
DEM	−1.4	9.3	−100.8	145.8	0	99.6	0	85.8
DMD	0.0	4.1	−84.7	102.1	0	99.0	0	43.9
DGC	−1.7	9.0	−217.4	116.2	0	96.6	0	31.1
DOGE	0.9	3.7	−60.8	144.9	0	97.6	1	0.1
FTC	−0.9	7.6	−144.7	171.7	0	83.2	0	10.8
FLO	0.7	6.8	−67.9	162.2	0	84.8	0	43.9
FRC	−0.5	21.3	−332.1	338.9	0	100.0	0	95.3
GLC	0.2	5.6	−74.1	91.0	0	100.0	0	51.2
IFC	−0.6	10.6	−126.4	296.1	0	50.0	0	51.2
LTC	0.6	2.1	−34.2	87.5	0	99.0	0	21.1
MEC	−1.6	5.6	−112.9	133.1	0	96.6	0	85.8
NMC	−0.9	3.0	−110.2	82.4	0	100.0	0	25.8
NVC	−1.0	5.5	−235.7	128.2	0	99.6	0	8.4
NXT	−0.1	4.2	−83.9	106.5	0	93.8	0	8.4
OMNI	−1.4	5.4	−73.1	116.9	0	82.6	0	43.9
PPC	−0.9	2.7	−60.2	73.8	0	86.8	0	21.1
XPM	−1.0	4.1	−67.7	117.7	0	91.2	1	4.9
ORK	−1.1	7.2	−94.1	137.9	0	96.8	0	37.2
XRP	1.1	3.8	−72.9	109.7	0	100.0	1	0.0
TAG	−1.1	5.3	−63.6	136.3	0	100.0	0	59.2
TRC	−1.0	6.7	−72.7	162.6	0	94.6	0	17.1
WDC	−1.6	6.8	−121.7	110.3	0	99.8	0	51.2
ZET	−1.0	6.6	−97.7	131.4	0	94.8	0	95.3

evaluated. To test for symmetry, the simple sign test, cf., e.g., Gibbons and Chakraborti (2011), is carried out for each dataset. The last four columns in Table 2 show the results of both tests.[2]

The results and especially the high *p*-values of the HDS test strongly suggest that all datasets obey a unimodal distribution. The CC Infinitecoin (IFC) shows the lowest *p*-value. This indicates that the empirical distribution could be multimodal. In fact, the histogram of returns for the IFC suggests a multipeaked nature. There are no indications of a fundamental structural break, and hence, this tends to be more of a random nature and is due to insufficient statistics (cf. the theorem of Glivenko (1933) and Cantelli (1933)), which might occur with short samples in particular. When adjusting a unimodal distribution function later in Section 3.3, we expect a lower quality of the distribution model for the returns of this currency.

Apart from three exceptions, a clear result can also be seen in the SIG test for symmetry of the empirical distribution of returns. For the vast majority of CCs, the assumption of a symmetrical distribution of returns at a moderately high level of significance cannot be rejected. While the assumption of a symmetrical distribution of the returns is narrowly rejected for the CC Primecoin (XPM), the rejection for the CCs Dogecoin (DOGE) and Ripple (XRP) is almost clear. We therefore assume that the distributions are slightly skewed. Going forward, we assume the returns of the CCs to be concentrated around zero and the empirical distribution to have a fat tail due to the large bandwidth. Furthermore, we expect the empirical distribution to have a unimodal and essentially symmetrical shape. We cannot rule out that the datasets in question may be (slightly) skewed. We will take this into account when selecting and adapting a suitable distribution function in Section 3.3.

3.2. Statistical tests to further reduce the variety of possible distributions

A number of mathematical models are available for the statistical description of CC returns. On the basis of some characteristic features of the dataset, the family of models can be narrowed down, and a suitable family of functions for representing the distribution can be deduced. In the following, a series of statistical tests are conducted to infer possible function families for the description of our datasets. The same tests are also carried out for the EWCI$^-$ defined in Section 2. In total, $N = 28$ time series are considered in the tests described below.

Overall, the statistical tests in Section 3.1 and the following are used to examine whether the combined hypothesis that the datasets have a unimodal, symmetrical and stationary distribution must be rejected. Furthermore, we assess whether the hypothesis

[2] Note that for all tests performed in this paper, the following applies: The boolean "0" indicates that the null hypotheses cannot be rejected at the 5% level, and alternatively the boolean "1" indicates a rejection.

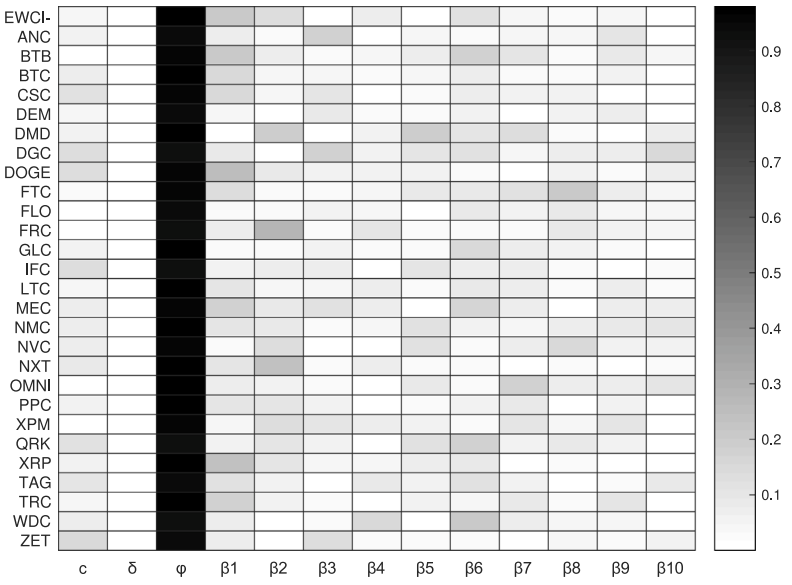

Fig. 1. The heatmap reflects the structure of the autoregressive model with drift coefficient c, deterministic trend coefficient δ, AR(1) coefficient φ and coefficients β_i for the lag terms up to ten time shifts for each CC.

of an independent, identical distribution of the individual returns must be rejected, which is an important property required to specify a distribution function.

The augmented Dickey–Fuller (ADF) test (Dickey and Fuller, 1979; Wooldridge, 2020) is used to test a possible rejection of the stationarity hypothesis. Finally, the autoregressive conditional heteroskedasticity (ARCH) test according to Engle (Engle, 1982, 2002) is used to check whether the hypothesis of homoskedasticity of the innovation process ϵ_t for the individual CCs and the EWCI$^-$ must be rejected.

Augmented Dickey–Fuller test

The ADF test is performed considering the autoregressive model for the CC return time series, y_t, of each CC[3]:

$$y_t = c + \delta t + \varphi y_{t-1} + \beta_1 \Delta y_{t-1} + \beta_2 \Delta y_{t-2} + \cdots + \beta_p \Delta y_{t-p} + \epsilon_t. \tag{1}$$

with a drift coefficient c, a deterministic trend coefficient δ, an AR(1) coefficient φ and the coefficients β_i for the lag terms $i = 1, \ldots, p$ up to the order $p = 10$. In Eq. (1) ϵ_t denotes the innovation process. The aim of the test is to examine the hypothesis of trend stationarity, i.e., $\delta = 0$ is the null hypothesis, in the tables denoted by H0; see, e.g., Wooldridge (2020).

The heatmap in Fig. 1 visualizes the structure of the fitted autoregressive model Eq. (1).

We find the deterministic trend coefficient δ to be comparable to zero in all CCs considered. The results in Table 3 in the first four columns provide deeper insight. For the vast majority of CCs, the p-values are comfortably high that a rejection of the null hypothesis of trend stationarity is not indicated here. However, as can be seen in the table, the ADF test rejects the null hypothesis for the CCs FLO (FLO), Quark (QRK) and Worldcoin (WDC) at the 5% confidence level. Next, we more closely examine the corresponding trend coefficients: $\delta_{\text{FLO}} = 1.6\text{e}{-03}$, $\delta_{\text{QRK}} = 1.5\text{e}{-04}$ and $\delta_{\text{WDC}} = 1.0\text{e}{-04}$. Since all the coefficients δ are close to zero, the influence of a possible trend is likely to be of significantly less importance. Therefore, in the following, we assume trend stationarity ($\delta = 0$) for the time series of CCs.

The third column in the heatmap in Fig. 1 illustrates the value of φ. We find the parameter φ to be greater than 0.9 for all CCs and, generally, clearly close to 1. The latter is an important condition to be fulfilled for the assumption of a random walk ($\varphi = 1$).

The ADF test was performed for lags p up to order ten. More complicated dynamics with serial correlation are made apparent in the analysis by the fact that the coefficients of the terms corresponding to the lags are clearly different from zero.

We find the absolute values of the coefficients (that means $|\beta_i|$) to be close to zero. In fact, the majority of the absolute coefficients $|\beta_i|$ are clearly smaller than 0.2 as an upper limit[4] and a basic model $y_t = c + y_{t-1} + \epsilon_t$ can be assumed (in Table 3, third column denoted by "G"). Only a few single coefficients exceed the value 0.2 do so by a small margin, and in these cases, a model with a slightly influential lag structure $y_t = c + y_{t-1} + \bar{\beta} \times (\text{Lag-Structure}) + \epsilon_t$ with an average coefficient $\bar{\beta} = 0.03$ could be considered. By marking the corresponding CCs with "L", Table 3 shows for which CCs this is the case. Due to the observed insignificance of the

[3] Note that the returns r_t are calculated in this notation according to $r_t = y_t - y_{t-1}$.

[4] By similar argumentation as in Wooldridge (2020, Chapter 11 therein), a repeated substitution causes the effectiveness of the corresponding terms to fall below the 5% mark in the next time step and thus become largely insignificant.

Table 3
Results of the statistical tests. Units in percent and boolean; see text.

CC	ADF test			ARCH test		Distribution
ID	H0	Model	*p*-value	H0	*p*-value	
EWCI⁻	0	L	29.4	1	0.0	–
ANC	0	G	11.7	0	6.6	IID
BTB	0	L	22.9	1	0.0	–
BTC	0	G	48.7	1	3.1	≈IID
CSC	0	G	49.2	0	50.8	IID
DEM	0	G	24.0	1	0.0	–
DMD	0	G	75.8	0	18.0	IID
DGC	0	G	11.0	1	0.0	–
DOGE	0	L	14.7	0	69.8	IID
FTC	0	L	14.9	1	0.0	–
FLO	1	L	4.0	0	63.2	≈IID
FRC	0	L	14.3	0	71.6	IID
GLC	0	G	65.2	0	16.6	IID
IFC	0	G	6.4	0	38.3	IID
LTC	0	G	34.2	0	25.2	IID
MEC	0	G	17.7	1	1.5	≈IID
NMC	0	G	42.0	0	38.1	IID
NVC	0	G	22.8	0	92.6	IID
NXT	0	L	50.4	1	0.0	–
OMNI	0	G	35.7	0	70.2	IID
PPC	0	G	41.5	1	0.4	≈IID
XPM	0	G	12.7	0	11.1	IID
QRK	1	L	2.0	1	0.3	≈IID
XRP	0	L	35.1	1	0.0	–
TAG	0	G	6.0	0	60.4	IID
TRC	0	G	46.0	1	1.9	≈IID
WDC	1	L	2.5	1	0.0	–
ZET	0	G	22.6	0	6.0	IID

lag structure, we assume a basic model "G" in these cases and expect statistical inaccuracies to distort the result due to the limited length of the time series. We therefore postulate possible serial correlation in the datasets to be of minor importance. Thus, the detailed analysis of the results of the ADF test suggest that the model $r_t = c + \epsilon_t$ cannot not be rejected for the returns r_t. Here, c denotes the individual time-constant drift term for each CC, and as above, ϵ_t denotes the innovation process.

Test of autoregressive conditional heteroskedasticity

The ARCH test is performed for time shifts q up to the order of ten. Up to this lag, the null hypothesis that the innovation process of returns is homoskedastic could not be rejected for most CCs (see Table 3); i.e., the basic model $\epsilon_t = \sigma z_t$ with constant volatility σ and z_t being an independent and identically (IID) distributed process with mean 0 and variance 1 could not be rejected for the majority of CCs. Note that the results found in literature show a heterogeneous picture, and we do not find ARCH effects in contrast to other related studies (Peng et al., 2018; Dyhrberg, 2016; Avital et al., 2014). Ultimately, only 13 of 28 time series examined show ARCH effects. The differences across studies may derive from different sampling frequencies or the differently chosen time period or its length and show that the design of the data collection may influence the results. The aforementioned results would justify the following calculation: $\mathrm{E}[r_t] = \mu = c$ and $\mathrm{Var}[r_t] = \sigma^2$ for the corresponding CCs. Estimates for the mean and the variance of the individual returns are noted in Table 2. Their calculation is also justified with the combined consideration of the test results described above.

When combining the two tests, we noted a characterization of the return distribution in the last column of Table 3. For the majority of CCs, IID returns can be assumed. Another part is approximately independent and identical distributed (≈IID), because either the lag structure is less important in the ADF model or the rejection of homoskedasticity based on the *p*-value is only weakly justified. For eight CCs, the assumption of IID returns is clearly rejected.

Note that the test for IID returns could have been performed with a turning point test (Bienaymé, 1874; Kendall and Stuart, 1977). However, as we are interested in a deeper analysis of the possible serial correlation in our datasets, we use a combination of the ADF test and the ARCH test instead.

All tests conducted thus far do not reject the assumption that the returns obey an essentially symmetrical, unimodal distribution. Furthermore, for the majority of CCs, the assumption of IID or nearly IID returns holds. In the first case (IID), the modeling of the empirical distribution with a distribution function is justified. In the second case (≈IID), the model represents a coarser approximation. In the latter case, if the assumption of IID returns were to be rejected, the distribution function could only be used as a rough approximation and must be – as in case (≈IID) – examined more precisely and critically in individual cases, as we show in the following section.

3.3. Determination of the appropriate return distribution function

When modeling the empirical distribution of returns, we focus on families of unimodal distribution functions that are defined over the entire axis (infinite support). Hence, a more detailed investigation of the following distribution functions suggests itself: normal distribution (N), the generalized extreme value distribution (GED), the generalized logistic distribution type 0 and type 3 (GLD0, GLD3) and the SDI.

The analysis below proves that the family of SDIs is the most suitable alternative for modeling the distributions of the CC returns under study. The analyses performed thus confirm the results found in the literature (Majoros and Zempléni, 2018; Kakinaka and Umeno, 2020). Consequently, we present this family of functions afterwards in more detail.

Following Nolan (2020, Def. 1.4 therein) the SDIs represent a family of distributions appropriate for modeling heavy-tailed and skewed data. In this context, it is noteworthy, that the linear combination of two IID and stably distributed random variables follows the same distributional characteristics as both individual variables. A random variable X follows the SDI $S(\alpha, \beta, \gamma, \delta)$ if its characteristic function can be defined as follows:

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

The first parameter of the distribution, $0 < \alpha \leq 2$ (named: *shape parameter*), is used to model the tail of the distribution. The second parameter of the distribution, $-1 \leq \beta \leq +1$, is used as a *skewness parameter*: For $\beta < 0$ ($\beta > 0$) the distribution is left-skewed (right-skewed). The distribution is symmetric, if $\beta = 0$. When α is small, the skewness of β is significant. As α increases, the effect of β decreases. Furthermore, $\gamma \in \mathbb{R}^{+}$ is used as a *scale parameter*, and $\delta \in \mathbb{R}$ is a *location parameter*.

For the special case of $\alpha = 2$, the SDI's characteristic function, see Eq. (2), reduces to $\mathrm{E}\left[\exp\left(\mathrm{i}tX\right)\right] = \exp\left(\mathrm{i}\delta t - (\gamma t)^2\right)$ and therefore becomes independent of β, so that the SDI becomes equal to N with mean δ and standard deviation $\sigma = \sqrt{2}\gamma$. For a more detailed description, compare Nolan (2020). For other applications of SDIs in the context of CCs, see, e.g., Börner et al. (2022).

Evaluation of distance measurements to compare model quality

In the following, we use standard distance measures to determine and compare the model qualities of N, GED, GLD0, GLD3 and SDI for the individual CCs. There are several distance measures available that are suitable to measure the potential differences between an empirical distribution function and a modeled distribution function. The distance measures from Cramér (1928) and von Mises (1931) (W^2-Distance), Anderson and Darling (1952, 1954) (A^2-Distance) and Kolmogorov (1933) and Smirnov (1936, 1948) (KS-Distance) are widely used in the literature. A brief summary of the distance measures employed here is given in Appendix A.

In Table 4, we summarize the results for the Anderson–Darling (AD) distance (A^2). The results for the Kolmogorov–Smirnov (KS) distance and the Cramér von Mises (CvM) distance (W^2) are compiled in Tables A.10 and A.11 in Appendix A.

The values of the various distance measures show that the GED is least suitable to model the empirical distribution function. This may derive from the fact that the GED contains a fundamental skewness, which can only be slightly influenced via parameter selection. Furthermore, once the shape parameter becomes different from zero, a fundamental change in the distribution model occurs, and the definition interval on the x-axis becomes restricted. Additionally, associated with a change in the sign of the shape parameter is a fundamental change in the distribution model and an abrupt change in the sign of the upper (or lower) bound on the x-axis; see, e.g., Embrechts et al. (1997). In the present case, these properties of the GED make it difficult to precisely adapt the distribution to the dataset.

On closer inspection of the calculated distances, the N also does not appear to be suitable as a model, since it is neither suitable for the modeling of empirical distribution functions with fat tails nor for those with a slight skew. Overall, the GLD0 shows significantly smaller distances across all CCs but is also not ideally suited, as it is completely symmetrical and is therefore not able to model any slight skewness. The best results in terms of the smallest distance can be achieved with the GLD3 and the SDI. When comparing all CCs, the corresponding distances are very close to one another. If only the Cramér von Mises distance and KS distance are considered, see Appendix A, we find that approximately half of the empirical distributions of CC returns can be modeled with one or the other distribution. However, once more attention is paid to the tail, i.e., if deviations in the tail area should receive comparably higher weightings to account for tail risks, and the AD distance A^2 is considered, the share of CCs for which the SDI is the most suitable model predominates.

This result ties in with those of Majoros and Zempléni (2018) and Kakinaka and Umeno (2020). Using intraday and daily time series for a small sample of CCs, they show the SDI family to be the best choice for modeling the empirical distribution function of intraday and daily returns of CCs.

For a broader sample of CCs, we found that the SDI is, on average, much better suited to model both slight skewness and pronounced tails in the empirical distribution function. Therefore, we use the SDI for all CCs to model the distribution function of returns.

Table 4
Anderson–Darling Distance for different body model distributions.

CC	Anderson–Darling Distance A^2					Best choice
ID	N	GED	GLD0	GLD3	SDI	
EWCI⁻	3.41	147.4	1.76	0.81	0.79	SDI
ANC	8.65	22.4	2.53	1.61	0.39	SDI
BTB	3.36	12.3	0.76	0.48	0.22	SDI
BTC	2.07	224.8	0.88	0.60	0.97	GLD3
CSC	21.80	42.6	3.90	n.d.	0.43	SDI
DEM	2.64	6.7	0.54	0.17	0.20	GLD3
DMD	3.80	23.7	1.25	0.29	0.54	GLD3
DGC	6.84	26.6	2.05	0.65	0.78	GLD3
DOGE	9.56	9.1	4.10	1.77	0.53	SDI
FTC	9.41	14.2	1.77	1.05	0.17	SDI
FLO	1.74	1.5	0.54	0.54	0.32	SDI
FRC	17.36	n.d.	5.21	1.98	0.39	SDI
GLC	1.53	16.8	0.40	0.19	0.42	GLD3
IFC	n.d.	n.d.	4.79	3.43	2.57	SDI
LTC	7.10	13.7	2.72	0.72	0.71	SDI
MEC	9.19	18.8	3.71	1.25	0.45	SDI
NMC	6.02	60.5	1.73	0.48	0.47	SDI
NVC	17.24	51.3	4.08	1.47	0.47	SDI
NXT	6.69	17.6	2.39	0.96	0.41	SDI
OMNI	1.25	4.4	0.25	0.24	0.24	SDI
PPC	4.93	31.9	1.63	0.61	0.48	SDI
XPM	5.66	5.0	1.53	0.67	0.27	SDI
QRK	6.32	9.5	2.12	0.53	0.52	SDI
XRP	13.90	13.0	5.71	2.79	0.62	SDI
TAG	4.85	5.4	1.23	0.30	0.64	GLD3
TRC	4.36	5.5	0.80	0.45	0.13	SDI
WDC	9.68	24.1	3.95	1.35	0.57	SDI
ZET	5.23	12.3	1.58	0.41	0.48	GLD3

Table 5
Parameters of the SDI and goodness of fit test.

CC	Parameter of the SDI $S(\alpha,\beta,\delta,\gamma)$								AD test	
ID	$\hat{\alpha}$	$\pm\Delta\alpha$	$\hat{\beta}$	$\pm\Delta\beta$	$\hat{\gamma}$	$\pm\Delta\gamma$	$\hat{\delta}$	$\pm\Delta\delta$	H0	p-value
EWCI⁻	1.62	0.18	0.46	0.39	0.07	0.01	−0.01	0.02	0	48.9
ANC	1.41	0.17	0.26	0.31	0.16	0.02	−0.03	0.03	0	85.8
BTB	1.67	0.18	0.38	0.45	0.18	0.02	−0.03	0.04	0	98.3
BTC	1.78	0.16	−0.21	0.61	0.06	0.01	0.01	0.01	0	37.5
CSC	1.45	0.18	0.27	0.32	0.16	0.02	−0.04	0.03	0	81.4
DEM	1.65	0.18	−0.04	0.45	0.17	0.02	−0.01	0.03	0	99.0
DMD	1.56	0.18	0.10	0.39	0.11	0.01	−0.01	0.02	0	70.7
DGC	1.57	0.18	0.31	0.38	0.14	0.02	−0.04	0.03	0	49.7
DOGE	1.33	0.17	0.35	0.27	0.08	0.01	−0.02	0.02	0	72.0
FTC	1.63	0.18	0.54	0.38	0.12	0.01	−0.05	0.03	0	99.7
FLO	1.88	n.d.	1.00	n.d.	0.16	n.d.	−0.02	n.d.	0	92.3
FRC	1.24	0.16	0.06	0.26	0.13	0.02	−0.02	0.03	0	86.0
GLC	1.80	0.16	0.43	0.63	0.15	0.02	−0.02	0.03	0	82.5
IFC	1.47	0.18	0.25	0.33	0.12	0.01	−0.04	0.02	1	4.6
LTC	1.37	0.17	0.06	0.30	0.06	0.01	−0.01	0.01	0	55.2
MEC	1.25	0.16	−0.01	0.26	0.09	0.01	−0.02	0.02	0	79.9
NMC	1.48	0.18	0.07	0.34	0.08	0.01	−0.01	0.02	0	78.0
NVC	1.42	0.17	0.28	0.31	0.08	0.01	−0.03	0.02	0	77.5
NXT	1.48	0.18	0.37	0.32	0.10	0.01	−0.03	0.02	0	83.8
OMNI	1.82	0.16	0.26	0.71	0.14	0.01	−0.03	0.03	0	97.7
PPC	1.50	0.18	0.10	0.35	0.08	0.01	−0.02	0.02	0	76.8
XPM	1.58	0.18	0.42	0.37	0.10	0.01	−0.04	0.02	0	95.6
QRK	1.45	0.18	0.17	0.33	0.12	0.02	−0.03	0.02	0	72.6
XRP	1.33	0.17	0.35	0.27	0.07	0.01	−0.03	0.01	0	63.2
TAG	1.63	0.18	0.10	0.44	0.12	0.01	−0.02	0.02	0	61.1
TRC	1.64	0.18	0.29	0.43	0.13	0.02	−0.04	0.03	0	99.9
WDC	1.26	0.16	0.08	0.26	0.10	0.01	−0.03	0.02	0	67.2
ZET	1.53	0.18	0.15	0.36	0.13	0.02	−0.02	0.03	0	76.5

The stable distribution family as a model for cryptocurrencies

Table 5 shows the results of the parameter estimation for the SDI $S(\alpha,\beta,\delta,\gamma)$.

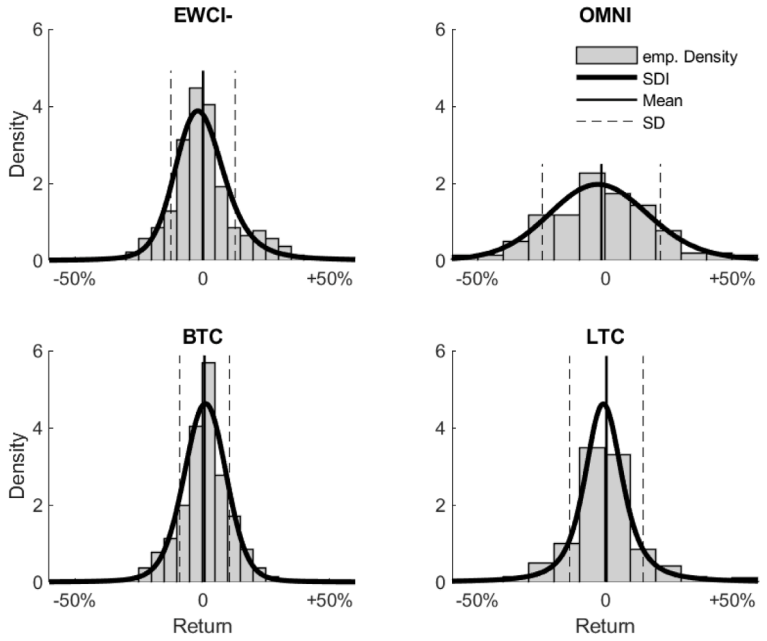

Fig. 2. The empirical densities and the results of the modeling of the distributions with the family of SDIs can be seen for the EWCI⁻ and selected CCs.

For the individual parameters of the SDI, the 95% scatter intervals are also provided. The latter can be determined from the covariance matrix of the estimated parameters. The covariance matrix of the parameter estimates is a matrix in which the off-diagonal element (i, j) resembles the covariance between the estimates of the ith parameter and the jth parameter. For the CC FLO, these scatter intervals cannot be determined numerically using the dataset at hand. This is because the corresponding empirical distribution on the far right shows a very long, pronounced tail and is strongly skewed to the right. The estimated parameter $\hat{\beta}$ of the SDI accordingly takes a value of 1; see Table 5. It may be the case that the single right tail data point, i.e., the return in period 62, represents an outlier, which is difficult to determine and correct afterwards without any further knowledge.

On average, the estimated parameter $\hat{\alpha}$ exceeds 1.5. With parameter α increasing to its limit value of 2.0, it can be seen that the distribution function becomes similar to the N and the skewness parameter ($\beta \neq 0$) becomes increasingly insignificant. The parameters β and α of the SDI are mutually dependent, and as described above, the meaning of β decreases when α increases. Thus, it is generally difficult to infer the skewness from the value β alone. A relative comparison of the distributions with respect to skewness is only possible if α has the same value. Hence, for a more precise analysis of a distribution's skewness, other methods are necessary. In this manner, we used the SIG test as an example and noted the results in Table 2.

For some CCs and the EWCI⁻ index, Fig. 2 shows the empirical densities and the density of the corresponding SDI in comparison.

In addition, the complete AD goodness-of-fit test was performed. The last two columns of Table 5 show that for almost all CCs with very high significance (high p-values), the null hypothesis that the adjusted SDI models the dataset cannot be rejected. Only for the CC IFC this assumption is rejected at the 5% level. This is probably because the empirical distribution suggests a slight bimodal distribution. This peculiarity of the CC IFC is also indicated in the results of the HDS test in Table 2.

Overall, the SDI family represents a suitable framework for modeling the distribution function of CC returns. We will exploit this finding for the assessment of tail risks and the comparison of our results with other modeling approaches.

4. Assessment of tail risks

4.1. Modeling of cryptocurrencies' tail risks

Especially when considering high quantiles in the risk assessment process, we follow e.g. Hoffmann and Börner (2020a, 2021) and make use of a separated modeling of the parent distribution's tail. In practice, the GPD is used predominantly as a tail model for such a modeling (Basel Commitee on Banking Supervision, 2009). Henceforth, we briefly discuss the main steps of the tail modeling using the GPD in this section.

It is well known and studied that for a wide class of distribution functions, GPD is suitable as a model for the limiting distribution in the tail region if the tail truncating threshold u is large enough (Gnedenko, 1943; Balkema and de Haan, 1974; Pickands, 1975).

The GPD is a distribution with (typically) two input parameters and follows the distribution function (Embrechts et al., 1997; McNeil et al., 2015):

$$F(x) = 1 - \left(1 + \xi \frac{x}{\sigma}\right)^{-\frac{1}{\xi}}, \tag{3}$$

Table 6
Parameters of the GPD and goodness of fit test for the loss tail. Units in percent.

CC	*Prop.*	GPD Parameter			Goodness of Fit		
ID		$\hat{u}$	$\hat{\xi}$	$\hat{\sigma}$	*p*-values		
					LT	CvM	AD
EWCI$^-$	45.7	−1.7	−0.09	0.09	88.0	84.9	92.6
ANC	4.3	−52.0	−0.08	0.61	98.5	98.1	93.9
BTB	4.3	−51.0	0.63	0.13	99.0	96.1	98.1
BTC	24.1	−3.9	−0.30	0.10	93.0	97.5	98.1
CSC	9.9	−35.5	0.82	0.11	96.7	99.7	94.3
DEM	51.4	−1.3	−0.05	0.22	54.2	58.2	60.8
DMD	4.3	−32.5	0.04	0.13	94.5	89.0	93.2
DGC	10.3	−31.7	0.61	0.08	99.8	99.6	99.5
DOGE	21.6	−9.2	0.13	0.09	99.1	99.4	98.9
FTC	3.5	−37.0	0.99	0.05	99.7	99.4	93.7
FLO	16.3	−20.3	−0.25	0.17	93.9	91.5	89.2
FRC	11.0	−28.8	0.28	0.31	98.8	98.3	99.6
GLC	51.8	0.2	−0.19	0.20	99.9	99.9	100.0
IFC	20.9	−18.2	0.55	0.07	55.0	70.4	52.8
LTC	44.3	−0.7	−0.18	0.11	57.5	67.7	54.1
MEC	56.4	0.0	0.16	0.12	99.8	99.2	93.3
NMC	46.1	−1.9	0.15	0.09	78.0	95.7	94.0
NVC	9.2	−19.3	0.72	0.05	77.2	93.6	86.0
NXT	2.8	−34.6	1.27	0.03	60.1	76.7	67.7
OMNI	15.2	−23.0	−0.11	0.14	95.9	96.2	97.1
PPC	8.5	21.3	0.00	0.09	96.7	97.1	98.8
XPM	4.6	29.5	−0.08	0.12	91.4	93.1	96.7
QRK	63.5	−1.6	0.05	0.16	47.7	60.8	61.1
XRP	2.8	−25.6	0.76	0.04	99.2	98.8	99.1
TAG	53.2	−0.5	−0.13	0.17	91.2	92.7	96.8
TRC	35.5	−9.9	−0.06	0.15	64.2	70.9	81.0
WDC	62.8	0.8	0.18	0.13	94.8	94.4	88.9
ZET	20.9	−17.8	0.27	0.11	59.8	83.6	82.5

where $\sigma > 0$ is the scale parameter and ξ is the shape parameter (a.k.a. tail parameter). The density function is described as

$$f(x) = \frac{1}{\sigma}\left(1 + \xi\frac{x}{\sigma}\right)^{-\frac{1+\xi}{\xi}}, \tag{4}$$

with $0 \leq x < \infty$ for $\xi \geq 0$ and $0 \leq x \leq -\frac{\sigma}{\xi}$ when $\xi < 0$. The mean and variance are depicted as $\mathrm{E}[x] = \frac{\sigma}{1-\xi}$ and $\mathrm{Var}[x] = \frac{\sigma^2}{(1-\xi)^2(1-2\xi)}$, respectively.

After the foundations of the GPD were introduced by Pickands (1975), not only theoretical advancements, but also practical applications were built on this work (Davison, 1984; Smith, 1984, 1985; van Montfort and Witter, 1985; Hosking and Wallis, 1987; Davison and Smith, 1990; Embrechts et al., 1997; Choulakian and Stephens, 2001; McNeil et al., 2015; Hoffmann and Börner, 2020a,b, 2021). Besides applications in engineering, the GPD is also the most widely used and recommended distribution function in finance for risk assessment at high quantiles (Embrechts et al., 1997; McNeil et al., 2015; Basel Commitee on Banking Supervision, 2009).

For the means of parameter estimation, the standard maximum likelihood method is the standard approach in the related literature (Davison, 1984; Smith, 1984, 1985; Hosking and Wallis, 1987; Embrechts et al., 1997). This method is also used here to separately estimate the parameters of the tail distributions of all CC return series under study (belonging to the 27 single CCs and the EWCI$^-$ index). Nevertheless, a plausibility check of the results is highly recommended. As we show in Section 4.2, when assessing tail risks, it is advisable to evaluate the results to avoid possible misinterpretations in individual cases.

In the course of a separated tail modeling, we now only need to consider data points, which belong to the tail area of the underlying empirical distribution function, i.e., the data that belongs to the area below a threshold u in case of the loss tail, for the parameter estimation of the GPD. The correct determination of the threshold u is of crucial importance. Following Hoffmann and Börner (2020a, 2021), the recently developed fully automated process that does not require any user intervention or additional parameters is used, to determine the threshold u. A brief description of the procedure is given in Appendix B.

Table 6 depicts the estimated parameters of the GPD for the different CCs. The second column reports the proportion of the whole dataset belonging to the loss tail. The proportion of the return data below the threshold $\hat{u}$ is used to fit the parameters ξ, σ of the GPD. The threshold value lies within the bandwidth shown in Table 2 and when considering the loss tail closer to the lower interval limit of the bandwidth. Due to the results of different standard goodness of fit tests (here: CvM and AD), the null hypothesis that the GPD is a suitable model for the tail of the CC return distribution cannot be rejected at any significance level. In addition, the *p*-values for the lower tail (LT) statistics according to Ahmad et al. (1988) are given in Table 6. The corresponding statistic AL^2 is defined in Appendix B and used here to determine the threshold value u. As can be seen in column six of Table 6, the LT statistics also present high confidence levels.

Table 7
Value at risk of the CCs for different confidence levels and different calculation methods. Losses with a positive sign and units in percent.

CC	Value at risk											
	empirical				Tail model (GPD)				Body model (SDI)			
ID	95%	97%	99%	99.9%	95%	97%	99%	99.9%	95%	97%	99%	99.9%
EWCI-	26	29	38	41	19	23	30	43	18	22	33	116
ANC	101	133	212	241	42	73	136	252	46	60	119	582
BTB	75	91	154	201	49	56	82	251	47	55	82	276
BTC	21	24	28	30	16	19	24	31	15	18	28	91
CSC	117	166	424	698	46	58	110	595	47	61	116	535
DEM	72	84	100	101	51	61	82	122	48	60	100	367
DMD	44	52	72	85	30	37	52	85	30	38	69	280
DGC	65	82	153	217	39	46	71	229	40	49	81	314
DOGE	36	43	56	61	23	29	42	77	23	31	64	343
FTC	50	60	109	145	36	38	50	205	32	37	53	177
FLO	49	54	66	68	38	44	55	70	37	42	52	68
FRC	110	142	256	332	57	78	136	333	54	79	185	1159
GLC	49	55	68	74	38	44	56	74	36	42	56	140
IFC	61	77	114	126	34	43	75	251	35	45	82	365
LTC	27	31	34	34	20	24	31	41	21	30	62	321
MEC	59	71	99	113	37	46	70	137	38	56	128	787
NMC	42	50	92	110	27	34	51	97	25	33	63	284
NVC	47	63	141	236	23	28	46	188	24	31	59	278
NXT	42	48	78	84	33	34	42	210	27	33	59	258
OMNI	48	55	67	73	37	43	55	76	37	42	56	142
PPC	35	40	55	60	26	31	41	61	25	33	60	258
XPM	40	46	62	68	29	35	47	70	27	33	51	191
QRK	62	71	91	94	41	50	70	117	38	50	96	440
XRP	30	36	57	73	24	25	32	85	22	30	61	330
TAG	47	53	62	64	35	41	53	73	34	42	69	252
TRC	51	57	69	73	37	44	57	83	36	44	68	238
WDC	64	79	109	122	39	50	76	150	40	57	128	769
ZET	59	73	91	98	37	46	70	151	37	47	85	357

Table 8
Average deviation from the empirical value at risk and scattering.

	ΔVaR			Confidence levels			
				95%	97%	99%	99.9%
Mean	GPD	./.	Emp.	0.5	−0.2	−4.6	15.5
	SDI	./.	Emp.	−0.3	0.4	10.5	214.1
SD	GPD	./.	Emp.	2.4	2.1	7.3	43.1
	SDI	./.	Emp.	1.8	4.1	19.3	211.7

4.2. Risk assessment at high quantiles

In this section, we use the SDI as the body model and the GPD as the tail model to determine the risk parameters of value at risk (as a quantile) and the conditional value at risk (as a weighted loss when the loss threshold is exceeded); see, i.a., Embrechts et al. (1997) and Hull (2018). The dataset includes $T = 282$ return observations for each CC, so that a comparison of the results with the empirically determined values is possible for moderately high confidence levels (≈99%). The corresponding values for the confidence level of 99.9%, which is important for regulatory purposes (Basel Commitee on Banking Supervision, 2004; European Parliament, 2009, 2013a,b), can only be estimated for data records of this length using a previously fitted body or tail model. The calculation of the quantiles is also subject to a statistical spread, and the estimation error increases the fatter the tail is and the higher the confidence level selected; see, e.g., Hoffmann and Börner (2020b).

Value at risk

Table 7 illustrates the results of the risk assessment for the most common confidence levels found in literature and regulatory requirements. The value at risk for the observed CCs is shown for the various models. Overall, in the overwhelming number of individual cases, the assessment of risk with the tail model (GPD) is closer to the empirical value at risk values. This applies to the lower confidence levels in particular, but even a high confidence level of 99.9%, better estimates are possible in individual cases than with the body model. This becomes apparent from the statistical parameters of the deviation analysis shown in Table 8. The mean value and standard deviation over the set of CCs are shown. For this purpose, the deviation between the modeled variable and the corresponding empirical value at risk was determined. On average, adopting the GPD as the tail model leads to better risk estimates.

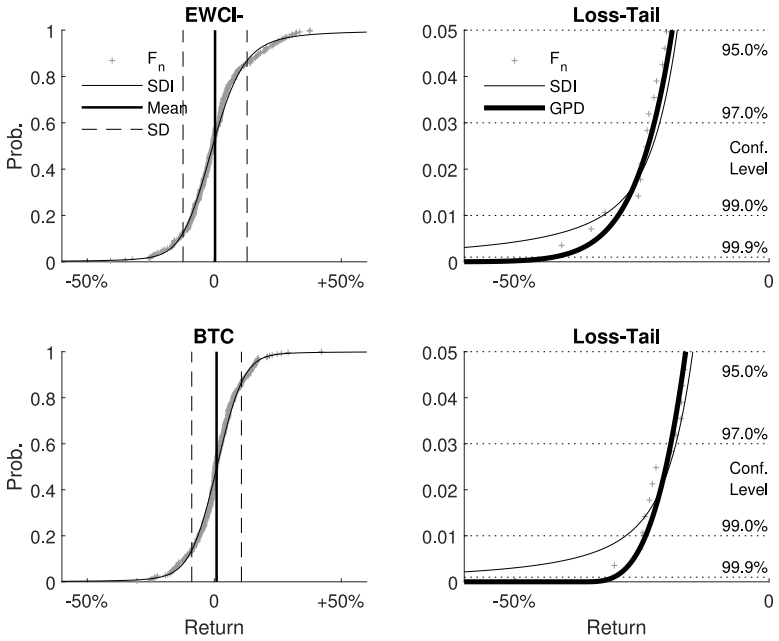

Fig. 3. The empirical distribution function and the distribution function modeled with the SDI can be seen for the EWCI⁻ and an example CC (left panels). The right graphics focus on the loss tail. The GPD adopted as a tail model and the confidence levels that are important for the regulator are also shown.

When comparing CCs with one another, a heterogeneous picture emerges; see Table 7. If the empirical value at risk for the 99.9% confidence interval is taken as a measure, two subgroups can be defined for the cutoff value $\text{VaR}_{99.9\%} \approx 100\%$. One group possesses a significant LT risk ($\text{VaR}_{99.9\%} < 100\%$) as the corresponding $\hat{\xi}$ values in Table 6 indicate. The other group ($\text{VaR}_{99.9\%} > 100\%$) partly embodies a significantly higher tail risk. Correspondingly, large $\hat{\xi}$ values can be determined for these CCs.

Fig. 3 shows the empirical distribution function for the EWCI⁻ and Bitcoin (BTC), the SDI as a body model and the GPD as a tail model in comparison. The graphics on the right portray an enlargement of the loss area. Particularly in this region, the GPD models the empirical distribution function very well. Considering the analysis above, we find that the GPD is ideally suited to conduct risk assessment at high quantiles. Therefore, we exclusively consider the GPD to estimate the conditional value at risk as a further risk indicator in the following.

Conditional value at risk

The following Table 9 shows the conditional value at risk calculated with the tail model (see Table 6) for the individual CCs. The calculation of the conditional value at risk can be conducted using the mean excess function of the GPD:

$$e(v) = \frac{\sigma + \xi v}{1 - \xi}, \tag{5}$$

with v being greater than the lower bound of the definition interval of the GPD to consider the loss tail and the parameters σ, ξ of the GPD. The mean excess function is bound to the restrictions $\xi < 1$ and $\sigma + \xi v > 0$; see, e.g., Embrechts et al. (1997, Theorem 3.4.13). In Table 9, we used Eq. (5) to estimate the conditional value at risk for each CC, setting $v = \text{VaR}_{p\%}$.

Again, the grouping of CCs described above can be seen. A group of CCs with a fat tail and therefore higher tail risk can be distinguished from a group with moderate risk; see Table 9. Furthermore, two peculiarities are noticeable concerning the CCs Feathercoin (FTC) and Nxt (NXT). For both CCs, the tail is modeled on a small number of data points that have been assigned to the tail. This individual property of the dataset deriving from the random distribution of the data in the tail area is assumed to be given and, as noted above, is not corrected. In particular, when estimating the parameter ξ, small samples lead to large statistical errors. Regrading the conspicuous CCs, the parameter is very close to 1 in one case (FTC) and even higher in the other case (NXT), see Table 6. As a result, the calculation of the conditional value at risk for the CC NXT is not possible and must be discarded; cf. Eq. (5) and the restriction $\xi < 1$. On the other hand, the calculation of the FTC with $\xi \approx 1$ has to be questioned critically. Hence, the conditional value at risk may only be a rough estimate in this case.

5. Conclusion

The aim of this study is to find a distribution that most accurately models CC returns and does not suffer from restrictions in specific parts of the distribution. In former research, the SDI and GPD have been found to adequately model the body and the tail of the CC return distributions, respectively. Nevertheless both distributions prove to be unsuitable to appropriately model the

Table 9
Conditional value at risk of CCs for different confidence levels calculated with the corresponding tail model (GPD). Losses with a positive sign and units in percent.

CC	Conditional value at risk			
ID	95%	97%	99%	99.9%
EWCI-	25	29	35	47
ANC	96	125	184	292
BTB	169	188	258	717
BTC	20	23	26	31
CSC	308	374	655	3 294
DEM	70	79	99	138
DMD	45	52	68	102
DGC	118	136	200	600
DOGE	37	43	59	98
FTC	3219	3408	4343	16 648
FLO	44	49	58	69
FRC	122	151	231	504
GLC	49	54	64	79
IFC	91	112	181	571
LTC	26	29	35	44
MEC	59	70	99	178
NMC	43	51	71	125
NVC	99	115	180	679
NXT	−134	−138	−165	−784
OMNI	46	51	62	80
PPC	35	40	49	70
XPM	38	43	55	76
QRK	59	69	90	140
XRP	114	121	147	367
TAG	46	52	62	80
TRC	49	55	68	92
WDC	63	76	107	197
ZET	66	79	112	223

entirety of the distribution. Therefore, using a novel approach to separate the distribution's tail from its body, we model the entire distribution by combining the model abilities of the SDI for the body and the GPD for the tail.

We select 27 CCs from the broad market of CCs according to predefined criteria and construct the representative index EWCI^{-}. Overall, we find independent, identical distributions such as the GPD and the SDI to be well suited for the most part and the family of SDIs in particular to be able to model the slightly skewed empirical distributions, especially in the body region. A comparison between different distribution functions shows that the SDI has outstanding modeling properties across the entire dataset. However, we show that the assessment of risks associated with fat tails can be performed more precisely with the GPD. The analysis of tail risks in the CC market using the GPD further hints at a certain internal structure of the CC market. The CC market can roughly be divided into two sets: CCs with moderate risk and CCs with high risk. This finding provides valuable information for both investors and regulators alike. Hence, our results are not only relevant for scientific applications and extensions but also for conceivable future regulation if the CC asset class is to become a permanent, noteworthy component of institutional investors' portfolios in the financial sector in the future. In this regard, numerous future extensions and research topics are conceivable. On the one hand, the SDI's suitability to model CC returns in portfolio optimization remains to be investigated. On the other hand, further research considering the segmentation of the CC market could enrich the understanding of CCs and improve forecasts concerned with the fundamental behavior of different CCs.

CRediT authorship contribution statement

Christoph J. Börner: Conceptualization, Supervision. **Ingo Hoffmann:** Conceptualization, Formal analysis, Methodology. **Lars M. Kürzinger:** Conceptualization, Methodology, Project administration, Validation, Writing – original draft, Writing – review & editing. **Tim Schmitz:** Data curation, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We thank Coinmarketcap.com for generously providing the cryptocurrency time series data for our research.

Table A.10
Cramér von Mises Distance for different body model distributions.

CC	Cramér von Mises Distance W^2					Best choice
ID	N	GED	GLD0	GLD3	SDI	
EWCI⁻	0.62	29.89	0.23	0.07	0.12	GLD3
ANC	1.51	4.25	0.41	0.24	0.06	SDI
BTB	0.55	2.22	0.11	0.06	0.04	SDI
BTC	0.40	41.81	0.16	0.11	0.19	GLD3
CSC	3.94	8.22	0.60	0.15	0.09	SDI
DEM	0.43	1.20	0.08	0.02	0.04	GLD3
DMD	0.70	5.19	0.22	0.05	0.10	GLD3
DGC	1.21	5.33	0.33	0.06	0.14	GLD3
DOGE	1.86	1.70	0.64	0.20	0.07	SDI
FTC	1.54	2.37	0.21	0.09	0.03	SDI
FLO	0.29	0.23	0.07	0.07	0.05	SDI
FRC	3.26	6.35	0.94	0.33	0.07	SDI
GLC	0.27	3.66	0.06	0.02	0.08	GLD3
IFC	2.92	4.43	0.91	0.72	0.58	SDI
LTC	1.35	2.98	0.46	0.11	0.12	GLD3
MEC	1.75	3.78	0.67	0.22	0.06	SDI
NMC	1.11	13.40	0.32	0.08	0.09	GLD3
NVC	3.09	10.46	0.65	0.17	0.09	SDI
NXT	1.20	3.57	0.33	0.09	0.07	SDI
OMNI	0.18	0.84	0.03	0.04	0.04	GLD0
PPC	0.88	7.07	0.26	0.09	0.08	SDI
XPM	0.99	0.86	0.20	0.05	0.05	SDI
QRK	1.15	1.74	0.36	0.08	0.10	GLD3
XRP	2.57	2.44	0.79	0.28	0.09	SDI
TAG	0.87	0.94	0.22	0.04	0.13	GLD3
TRC	0.71	0.96	0.09	0.04	0.02	SDI
WDC	1.80	5.03	0.70	0.23	0.10	SDI
ZET	0.92	2.33	0.25	0.05	0.09	GLD3

Appendix A. Distance measures – Tables of results

A.1. Distance measures

In what follows, a brief summary of the used distance measures is given.

Cramér von Mises and Anderson–Darling distance measures

Following Hoffmann and Börner (2020a, 2021), our first choice to measure the distance between the empirical distribution functions $F_n(x)$ (Kolmogorov, 1933) and a model $F(x)$, is a weighted mean square error calculated as

$$\hat{R}_n = n \int_{-\infty}^{+\infty} \left(F_n(x) - F(x)\right)^2 \; w(F(x)) \, \mathrm{d}F(x) \tag{A.1}$$

and originally introduced by Cramér (1928), von Mises (1931) and Smirnov (1936) in the context of statistical (hypothesis) testing, cf. also Shorack and Wellner (2009). From a more decision-theoretical point of view (Ferguson, 1967), numerous studies also used the weighted mean square error as an application to determine distribution parameters by using minimum distance approaches (Wolfowitz, 1957; Blyth, 1970; Parr and Schucany, 1980; Boos, 1982). This measure of error is also used when adapting tail models (Hoffmann and Börner, 2020a, 2021). Therefore, in Section 4.1, we applied this distance measure in connection with the adaption of a suitable tail model for CC returns. A brief overview of the procedure to fit a suitable tail model is given in Appendix B.

Using a (non-negative) weight function $w(t)$, the formula in Eq. (A.1) is able to consider the differences between the different distribution functions more accentuated in those areas, where the respective distance measure should be particularly sensitive (Hoffmann and Börner, 2020a, 2021). Usually the weight function

$$w(t) = \frac{1}{t^a(1-t)^b} \tag{A.2}$$

with parameters $a, b \geq 0$ and $t \in [0,1]$ is considered. Here, a affects the weight at the lower tail and b at the upper tail. For $a = b = 0$, Eq. (A.1) provides the CvM distance W^2 used in the corresponding statistic (Cramér, 1928; von Mises, 1931). For the case of $a = b = 1$ (means: a heavy weighting of the tail area), the resulting expression becomes equal to the AD distance A^2, which is used in the corresponding statistic by Anderson and Darling (1952, 1954). Thus, potential differences between the two distributions in the upper and lower tails of the distribution $F(x)$ have a higher weighting in the calculation of the AD distance.

The results for the AD distance are shown in Table 4 in the main text in Section 3.3.

Table A.11
Kolmogorov–Smirnov distance for different body model distributions.

CC	Kolmogorov–Smirnov Distances KS					Best choice
ID	N	GED	GLD0	GLD3	SDI	
EWCI⁻	0.100	0.509	0.063	0.043	0.051	GLD3
ANC	0.132	0.233	0.069	0.070	0.047	SDI
BTB	0.087	0.156	0.056	0.037	0.042	GLD3
BTC	0.077	0.591	0.058	0.047	0.060	GLD3
CSC	0.179	0.296	0.085	0.053	0.047	SDI
DEM	0.071	0.127	0.048	0.027	0.039	GLD3
DMD	0.094	0.233	0.056	0.039	0.050	GLD3
DGC	0.117	0.241	0.066	0.034	0.045	GLD3
DOGE	0.159	0.132	0.099	0.053	0.042	SDI
FTC	0.134	0.141	0.061	0.045	0.031	SDI
FLO	0.069	0.060	0.036	0.035	0.038	GLD3
FRC	0.166	0.250	0.097	0.061	0.047	SDI
GLC	0.072	0.190	0.038	0.025	0.040	GLD3
IFC	0.194	0.251	0.142	0.144	0.106	SDI
LTC	0.133	0.184	0.077	0.043	0.058	GLD3
MEC	0.144	0.206	0.096	0.070	0.041	SDI
NMC	0.095	0.368	0.056	0.040	0.037	SDI
NVC	0.167	0.338	0.092	0.052	0.042	SDI
NXT	0.134	0.196	0.077	0.051	0.039	SDI
OMNI	0.052	0.113	0.028	0.031	0.036	GLD0
PPC	0.109	0.262	0.063	0.048	0.053	GLD3
XPM	0.116	0.095	0.055	0.048	0.035	SDI
QRK	0.108	0.139	0.073	0.048	0.048	SDI
XRP	0.171	0.168	0.101	0.061	0.039	SDI
TAG	0.101	0.103	0.056	0.037	0.048	GLD3
TRC	0.099	0.120	0.039	0.033	0.023	SDI
WDC	0.144	0.236	0.103	0.075	0.052	SDI
ZET	0.123	0.154	0.077	0.049	0.047	SDI

Kolmogorov–Smirnov distance measure

Furthermore, we also determine the well-known distance between the empirical distribution function and the distribution model of Kolmogorov (1933) and Smirnov (1936, 1948). The KS distance calculates the supremum of the absolute difference between the empirical and the estimated distribution functions. Hence, the KS distance quantifies possible differences between both the theoretically assumed and the empirically observed distribution functions of the CC returns under study. A more theoretical overview and comparisons to other distance measures can be found in, e.g. Stephens (1974), Shorack and Wellner (2009).

Appendix B. *FindTheTail* – Determining threshold u

As a foundation of our CC tail modeling application, we start with the common assumption, that there is a threshold u, which divides the underlying (parent) distribution into a body and a tail as separately modeled areas (Embrechts et al., 1997; McNeil et al., 2015; Hoffmann and Börner, 2021). This separation is a common approach to capture high quantiles of distributions more accurately (European Parliament, 2009).

Various authors have proposed methods for determining the appropriate threshold u and subsequently the GPD as a model for the tail from empirical data. Most methods require the setting of parameters, which often requires experience and hinders full automation of the modeling process. We follow Hoffmann and Börner (2020a, 2021) and use their full automated process for the determination of the threshold u and the parametrization of the tail model.

Starting with a suitable distance measure $\hat{R}_n = \hat{R}_n(F_n, \hat{F})$ as a function of the estimated GPD $\hat{F}(x)$ and the empirical distribution function F_n (Kolmogorov, 1933), an automated modeling process can be constructed using the following pseudo algorithm (Hoffmann and Börner, 2020a, 2021):

1. We arrange the (random) sample data, which is assumed to be drawn from an unknown (parent) distribution, in a descending order: $x_{(1)} \geq x_{(2)} \geq \cdots \geq x_{(n)}$.
2. Assuming $k = 2, \ldots, n$, we now estimate the parameters of the GPD for each k. (Note: For numerical reasons, the process starts at $k = 2$).
3. We then calculate the probabilities $\hat{F}(x_{(i)})$ for $i = 1, \ldots, k$ with the estimated GPD, and determine the distance $\hat{R}_k$ for $k = 2, \ldots, n$.
4. At last, we identify the index k^*, which is relevant for the minimum distance $\hat{R}_k$.

Building on this beforementioned algorithm, we can now estimate the optimal threshold ($\hat{u} = x_{(k^*)}$), and finalize the tail modeling of our unknown (parent) distribution, which is here proxied by the estimated GPD $\hat{F}(x)$ derived from the abovementioned subset $x_{(1)} \geq x_{(2)} \geq \cdots \geq x_{(k^*)}$.

As proposed by Hoffmann and Börner (2020a, 2021) the distance measure defined by Ahmad et al. (1988) is used in the algorithm above. This distance measure is also based on the weighted mean square error, Eq. (A.1), and can be noted in two variants. The two variants of the distance measure of Ahmad et al. (1988) are derived from Eq. (A.1) when the integral is calculated with the weight functions Eq. (A.2) and $(a, b) = (1, 0)$ for the lower tail ($= AL^2$) and $(a, b) = (0, 1)$ for the upper tail ($= AU^2$). With the asymmetrical weight function defined so far the distance measure $\hat{R}_n$ take more account of the difference between the measured and the modeled data, especially in the tail region.

References

Ahmad, M.I., Sinclair, C.D., Spurr, B.D., 1988. Assessment of flood frequency models using empirical distribution function statistics. Water Resour. Res. 24, 1323–1328. http://dx.doi.org/10.1029/WR024i008p01323.

Anderson, T.W., Darling, D.A., 1952. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. Ann. Math. Stat. 23, 193–212. http://dx.doi.org/10.1214/aoms/1177729437.

Anderson, T.W., Darling, D.A., 1954. A test of goodness of fit. J. Amer. Statist. Assoc. 49, 765–769. http://dx.doi.org/10.2307/2281537.

Avital, M., Leimeister, J.M., Schultze, U., 2014. ECIS 2014 Proceedings: 22th European Conference on Information Systems; Tel Aviv, Israel, June 9-11, 2014. AIS Electronic Library, URL http://aisel.aisnet.org/ecis2014/.

Balcilar, M., Bouri, E., Gupta, R., Roubaud, D., 2017. Can volume predict bitcoin returns and volatility? a quantiles-based approach. Econ. Model. 64, 74–81. http://dx.doi.org/10.1016/J.ECONMOD.2017.03.019.

Balkema, A.A., de Haan, L., 1974. Residual life time at great age. Ann. Probab. 2, 792–804. http://dx.doi.org/10.1214/aop/1176996548.

Basel Commitee on Banking Supervision, 2004. International convergence of capital measurement and capital standards - a revised framework.

Basel Commitee on Banking Supervision, 2009. Observed range of practice in key elements of advanced measurement approaches (ama).

Baur, D.G., Dimpfl, T., Kuck, K., 2018. Bitcoin, gold and the us dollar–a replication and extension. Finance Res. Lett. 25, 103–110.

Bienaymé, I.J., 1874. Sur une question de probabilités. Bull. Soc. Math. France 2, 153–154.

Blyth, C.R., 1970. On the inference and decision models of statistics. Ann. Math. Stat. 41, 1034–1058.

Boos, D.D., 1982. Minimum anderson-darling estimation. Comm. Statist. Theory Methods 11, 2747–2774.

Börner, C.J., Hoffmann, I., Krettek, J., Schmitz, T., 2022. Bitcoin: like a satellite or always hardcore? A core–satellite identification in the cryptocurrency market. J. Asset Manag. 23, 310–321. http://dx.doi.org/10.1057/s41260-022-00267-z.

Bouri, E., Gupta, R., Tiwari, A.K., Roubaud, D., 2017a. Does bitcoin hedge global uncertainty? evidence from wavelet-based quantile-in-quantile regressions. Finance Res. Lett. 23, 87–95. http://dx.doi.org/10.1016/j.frl.2017.02.009.

Bouri, E., Molnár, P., Azzi, G., Roubaud, D., Hagfors, L.I., 2017b. On the hedge and safe haven properties of bitcoin: Is it really more than a diversifier? Finance Res. Lett. 20, 192–198. http://dx.doi.org/10.1016/j.frl.2016.09.025.

Brauneis, A., Mestel, R., 2018. Price discovery of cryptocurrencies: Bitcoin and beyond. Econom. Lett. 165, 58–61.

Brière, M., Oosterlinck, K., Szafarz, A., 2015. Virtual currency, tangible return: Porfolio diversification with bitcoin. J. Asset Manag. 16, 365–373.

Cantelli, F.P., 1933. Sulla determinazione empirica delle leggi di probabilità. Giornale Inst. Ital. degli Attuari 1933, 421–424.

Caporale, G.M., Gil-Alana, L., Plastun, A., 2018. Persistence in the cryptocurrency market. Res. Int. Bus. Finance 46, 141–148. http://dx.doi.org/10.1016/j.ribaf.2018.01.002.

Choulakian, V., Stephens, M.A., 2001. Goodness-of-fit tests for the generalized pareto distribution. Technometrics 43, 478–484.

Corbet, S., Lucey, B.M., Peat, M., Vigne, S., 2018. Bitcoin futures - what use are they? Econom. Lett. 172, 23–27.

Corbet, S., Lucey, B., Urquhart, A., Yarovaya, L., 2019. Cryptocurrencies as a financial asset: A systematic analysis. Int. Rev. Financ. Anal. 62, 182–199. http://dx.doi.org/10.1016/j.irfa.2018.09.003.

Cramér, H., 1928. On the composition of elementary errors: Second paper: Statistical applications. Scand. Actuar. J. 1, 141–180.

Davison, A.C., 1984. Modelling Excess over High Threshold, with an Application: Statistical Extremes and Applications. Reidel Publishing Company, Dordrecht, Netherlands.

Davison, A.C., Smith, R.L., 1990. Models for exceedances over high thresholds (with comments). J. R. Stat. Soc. Ser. B Stat. Methodol. 52, 393–442.

Dickey, D.A., Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. J. Amer. Statist. Assoc. 74, 427–431.

Dyhrberg, A.H., 2016. Hedging capabilities of bitcoin. is it the virtual gold? Finance Res. Lett. 16, 139–144. http://dx.doi.org/10.1016/j.frl.2015.10.025, URL https://www.sciencedirect.com/science/article/pii/S1544612315001208.

ElBahrawy, A., Alessandretti, L., Kandler, A., Pastor-Satorras, R., Baronchelli, A., 2017. Evolutionary dynamics of the cryptocurrency market. Royal Soc. Open Sci. 4.

Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events: For Insurance and Finance. In: Applications of Mathematics, Stochastic Modelling and Applied Probability, Springer Berlin Heidelberg, http://dx.doi.org/10.1007/978-3-642-33483-2.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica 50, 987–1007. http://dx.doi.org/10.2307/1912773.

Engle, R.F., 2002. A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. J. Bus. Econom. Statist. 20, 339–350.

European Parliament, 2009. Directive 2009/138/ec of the european parliament and of the council of 25 november 2009 on the taking-up and pursuit of the business of insurance and reinsurance (solvency ii). Off. J. Eur. Union 52, 1–155.

European Parliament, 2013a. Directive 2013/36/eu of the european parliament and of the council of 26 june 2013 on access to the activity of credit institutions and the prudential supervision of credit institutions and investment firms, amending directive 2002/87/ec and repealing directives 2006/48/ec and 2006/49/ecand investment firms, amending directive 2002/87/ec and repealing directives 2006/48/ec and 2006/49/ec. Off. J. Eur. Union 56, 338–436.

European Parliament, 2013b. Regulation (eu) no 575/2013 of the european parliament and of the council of 26 june 2013 on prudential requirements for credit institutions and investment firms and amending regulation (eu) no 648/2012. Off. J. Eur. Union 56, 1–337.

Ferguson, T.S., 1967. Mathematical Statistics: A Decision Theoretic Approach. In: Probability and Mathematical Statistics a Series of Monographs and Textbooks, vol. 1, Academic Press, New York.

Fry, J., Cheah, E.T., 2016. Negative bubbles and shocks in cryptocurrency markets. Int. Rev. Financ. Anal. 47, 343–352. http://dx.doi.org/10.1016/j.irfa.2016.02.008.

Gandal, N., Hamrick, J.T., Moore, T., Oberman, T., 2018. Price manipulation in the bitcoin ecosystem. J. Monetary Econ. 95, 86–96.

Gibbons, J.D., Chakraborti, S., 2011. Nonparametric Statistical Inference, fifth ed. In: Statistics, vol. 198, Chapman & Hall/CRC Press, Boca Raton, Fla.

Gkillas, K., Bekiros, S., Siriopoulos, C., 2018. Extreme correlation in cryptocurrency markets. SSRN Electron. J. http://dx.doi.org/10.2139/ssrn.3180934.

Gkillas, K., Katsiampa, P., 2018. An application of extreme value theory to cryptocurrencies. Econom. Lett. 164, 109–111. http://dx.doi.org/10.1016/j.econlet.2018.01.020.

Glas, T.N., 2019. Investments in cryptocurrencies: Handle with care!. J. Altern. Invest. 22, 96–113.

Glivenko, V., 1933. Sulla determinazione empirica delle leggi di probabilità. Giornale Inst. Ital. degli Attuari 1933, 92–99.

Gnedenko, B.V., 1943. Sur la distribution limite du terme maximum d'une série aléatoire. Ann. of Math. 44, 423–453.

Hartigan, J.A., Hartigan, P.M., 1985. The dip test of unimodality. Ann. Statist. 13, 70–84.
Hayes, A.S., 2017. Cryptocurrency value formation: An empirical study leading to a cost of production model for valuing bitcoin. Telemat. Inform. 34, 1308–1321. http://dx.doi.org/10.1016/j.tele.2016.05.005.
Hoffmann, I., Börner, C.J., 2020a. The risk function of the goodness-of-fit tests for tail models. Statist. Pap. 1–17.
Hoffmann, I., Börner, C.J., 2020b. Tail models and the statistical limit of accuracy in risk assessment. J. Risk Financ. 21, 201–216. http://dx.doi.org/10.1108/JRF-11-2019-0217.
Hoffmann, I., Börner, C., 2021. Body and tail: an automated tail-detecting procedure. J. Risk 23, 43–69. http://dx.doi.org/10.21314/JOR.2020.447.
Hosking, J.R.M., Wallis, J.R., 1987. Parameter and quantile estimation for the generalized pareto distribution. Technometrics 29, 339–349. http://dx.doi.org/10.2307/1260343.
Hull, J.C., 2018. Options, Futures, and other Derivatives, ninth ed. Pearson Education, Harlow, global edition ed..
Kakinaka, S., Umeno, K., 2020. Characterizing cryptocurrency market with Lévy's stable distributions. J. Phys. Soc. Japan 89, 024802. http://dx.doi.org/10.7566/JPSJ.89.024802.
Kaya Soylu, P., Okur, M., Çatıkkaş, Ö., Altintig, Z.A., 2020. Long memory in the volatility of selected cryptocurrencies: Bitcoin, ethereum and ripple. J. Risk Financ. Manag. 13, 107. http://dx.doi.org/10.3390/jrfm13060107.
Kendall, M.G., Stuart, A., 1977. The Advanced Theory of Statistics: in Three Volumes, fourth ed. Charles Griffin & Co Ltd, London.
Kolmogorov, A.N., 1933. Sulla determinazione empirica di una legge di distribuzione. Giornale Inst. Ital. degli Attuari 8, 3–91.
Majoros, S., Zempléni, A., 2018. Multivariate Stable Distributions and their Applications for Modelling Cryptocurrency-Returns. Working Paper 2018, URL http://arxiv.org/pdf/1810.09521v1.
McNeil, A.J., Frey, R., Embrechts, P., 2015. Quantitative Risk Management: Concepts, Techniques and Tools, Revised edition ed. Princeton University Press, Princeton and Oxford.
Nakamoto, S., 2008. Bitcoin: A peer-to-peer electronic cash system. Consulted 1, 2012.
Nolan, J.P., 2020. Univariate Stable Distributions: Models for Heavy Tailed Data, first ed. In: Springer Series in Operations Research and Financial Engineering, Springer International Publishing and Imprint: Springer, Cham, http://dx.doi.org/10.1007/978-3-030-52915-4, 2020 ed..
Osterrieder, J., Lorenz, J., Strika, M., 2017. Bitcoin and cryptocurrencies - not for the faint-hearted. Int. Finance Bank. 13, 145–193.
Parr, W.C., Schucany, W.R., 1980. Minimum distance and robust estimation. J. Amer. Statist. Assoc. 75, 616–624.
Peng, Y., Albuquerque, P.H.M., Camboim de Sá, J.M., Padula, A.J.A., Montenegro, M.R., 2018. The best of two worlds: Forecasting high frequency volatility for cryptocurrencies and traditional currencies with support vector regression. Expert Syst. Appl. 97, 177–192. http://dx.doi.org/10.1016/j.eswa.2017.12.004.
Pickands, III, J., 1975. Statistical inference using extreme order statistics. Ann. Statist. 3, 119–131. http://dx.doi.org/10.1214/aos/1176343003.
Polasik, M., Piotrowska, A.I., Wisniewski, T.P., Kotkowski, R., Lightfoot, G., 2015. Price fluctuations and the use of bitcoin: An empirical inquiry. Int. J. Electron. Commerce 20, 9–49.
Schmitz, T., Hoffmann, I., 2020. Re-Evaluating Cryptocurrencies' Contribution to Portfolio Diversification – A Portfolio Analysis with Special Focus on German Investors. Working Paper, URL http://arxiv.org/pdf/2006.06237v2.
Selgin, G., 2015. Synthetic commodity money. J. Financ. Stab. 9, 2–99.
Shorack, G.R., Wellner, J.A., 2009. Empirical Processes with Applications to Statistics. In: Classics in Applied Mathematics, vol. 59, Society for Industrial and Applied Mathematics, Philadelphia, PA.
Smirnov, N.V., 1936. Sur la distribution de w^2-criterion (critérion de r. von mises). C. R. Acad. Sci. Paris 202, 449–452.
Smirnov, N.V., 1948. Table for estimating the goodness of fit of empirical distributions. Ann. Math. Stat. 19, 279–281. http://dx.doi.org/10.1214/aoms/1177730256.
Smith, R.L., 1984. Threshold methods for sample extremes: Statistical extremes and applications 1984. pp. 621–638.
Smith, R.L., 1985. Maximum likelihood estimations in a class of nonregular cases. Biometrika 72, 67–90.
Stephens, M.A., 1974. Edf statistics for goodness of fit and some comparisons. J. Amer. Statist. Assoc. 69, 730–737. http://dx.doi.org/10.2307/2286009.
Trimborn, S., Li, M., Härdle, W.K., 2020. Investing with cryptocurrencies - a liquidity constrained investments approach: Working paper. J. Financ. Econom. 18, 280–306.
Urquhart, A., 2018. What causes the attention of bitcoin? Econom. Lett. 166, 40–44.
van Montfort, M.A.J., Witter, J.V., 1985. Testing exponentiality against generalized pareto distribution. J. Hydrol. 78, 305–315.
von Mises, R.E., 1931. Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik. Franz Deuticke, Leipzig und Wien.
Wolfowitz, J., 1957. The minimum distance method. Ann. Appl. Stat. 28, 75–88.
Wooldridge, J.M., 2020. Introductory Econometrics: A Modern Approach, seventh ed. CENGAGE, Boston, MA.

Further reading

Bader, B., Yan, J., Zhang, X., 2018. Automated threshold selection for extreme value analysis via goodness-of-fit tests with application to batched return level mapping. Ann. Appl. Stat. 12, 310–329.
Chapelle, A., Crama, Y., Hübner, G., Peters, J., 2005. Measuring and managing operational risk in financial sector: An integrated framework. Soc. Sci. Res. Netw. Electron. J. http://dx.doi.org/10.2139/ssrn.675186.
Crama, Y., Hübner, G., Peters, J. (Eds.), 2007. Impact of the Collection Threshold on the Determination of the Capital Charge for Operational Risk: Advances in Risk Management. Palgrave Macmillan, New York.
Hoffmann, I., Börner, C.J., 2018. Body and Tail: Separating the Distribution Function by an Efficient Tail-Detecting Procedure in Risk Management. Working Paper.
Huisman, R., Koedijk, K.G., Kool, C.J.M., Palm, F., 2001. Tail-index estimates in small samples. J. Bus. Econom. Statist. 19, 208–216.
Longin, F., Solnik, B., 2001. Extreme correlation of international equity markets. J. Finance 56, 649–676. http://dx.doi.org/10.1111/0022-1082.00340.