Finance Research Letters 47 (2022) 102574

Contents lists available at ScienceDirect

Finance Research Letters

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

A novel heavy tail distribution of logarithmic returns of cryptocurrencies

Quang Van Tran [a,b,*], Jaromir Kukal [a]

[a] *Czech Technical University in Prague, Faculty of Nuclear Sciences and Physical Engineering, Department of Software Engineering, Trojanova 13, 120 00 Prague 2, Czech Republic*

[b] *University of Economics and Business in Prague, Faculty of Finance and Accounting, Department of Monetary Theory and Policy, nám W. Churchilla 4, 130 67 Prague 3, Czech Republic*

A R T I C L E I N F O

JEL classification:
C51
C52
C58

Keywords:
Generalized gamma distribution
Regularization
Parameter estimation
Logarithmic returns
Cryptocurrencies

A B S T R A C T

We propose a novel distribution derived from the generalized gamma distribution by symmetrization and regularization around the mean. Besides location and scale parameters, the distribution has three shape parameters with many sub-models as special cases. Its parameters can be estimated by non-linear regression with parameter significance verification and sub-model testing. The applicability of this family of novel distributions is verified on returns of three cryptocurrencies and its suitability is tested by χ^2 goodness of fit testing. The obtained results show that this novel distribution and its sub-models can be viable candidates for modeling the returns of cryptocurrencies.

1. Introduction

Cryptocurrencies have become an unignorable phenomenon as they are a subject of enormous investment interest with immense market capitalization, see Akyildirim et al. (2021). Their trading has also been accompanied by excessive fluctuations whose cause was documented and explained in Bouri et al. (2019), Long et al. (2020) and Vidal-Tomás et al. (2019) and a great amount of research has been devoted to their price dynamics and their impact on financial management, see Akhtaruzzaman et al. (2020), Sensoy (2019), Aslan and Sensoy (2020) and Corbet et al. (2019). Like with returns of all financial assets, a proper model of their returns is of great importance for financial engineering, see Silahli et al. (2019). So far, standard heavy-tail distributions (like the generalized normal distribution in Nadarajah, 2005, the generalized t-distribution family in Theodossiou (1998), the generalized hyperbolic distribution family in Prause (1999) and others, for example in Nolan (1997) and Kukal and Tran (2019)) are unable to accomplish this task and the demand for a proper distribution able to capture this property is very high. With the excessive volatility, the heavy tail property of cryptocurrencies' returns appears to be much more severe than the one of traditional assets. However, there has not been a great deal of attempts to deal with this problem in the literature. We have noticed the use of Laplace distribution, extreme value distribution, generalized error distribution in Osterrieder and Lorenz (2017) and Szczygielski et al. (2020) or a mixture of Laplace distributions in Punzo and Bagnato (2021). Drozdz et al. (2018) have documented several non-random patterns of returns of Bitcoin. However, in these studies there is a lack of proper testing whether data actually comes from these distributions.

* Corresponding author at: Czech Technical University in Prague, Faculty of Nuclear Sciences and Physical Engineering, Department of Software Engineering, Trojanova 13, 120 00 Prague 2, Czech Republic.

E-mail addresses: tranvqua@fjfi.cvut.cz (Q.V. Tran), jaromir.kukal@fjfi.cvut.cz (J. Kukal).

https://doi.org/10.1016/j.frl.2021.102574
Received 10 August 2021; Received in revised form 18 October 2021; Accepted 17 November 2021
Available online 2 December 2021
1544-6123/

To fill the gap in the literature, the objective of this paper is to propose a distribution able to model heavy tail returns of cryptocurrencies. We derive it from the generalized gamma distribution (GGD) first proposed by Stacy (1962). Unfortunately, the GGD is defined only for real positive values, which makes it inapplicable for modeling assets' returns. The novelty of our work is the introduction of modifications that make it usable for this purpose. The first modification is the symmetrization of GGD resulting in a distribution for the whole real domain. The second modification is to smooth it in area around its peak. It is done by regularization, e.i., replacing a part of the original curve by a polynomial smoothing the symmetrized PDF while preserving its other properties.

We use this distribution to model logarithmic returns of three most often encountered cryptocurrencies: Bitcoin, Ethereum, and Ripple. First, with their publicly available data, the parameters of this distribution and its important special cases are estimated by a more effective non-linear regression estimation technique. Then the validity of special cases is tested by using likelihood ratio test and Schwartz information criterion. Finally, the suitability of our novel distribution for modeling returns of three cryptocurrencies is tested using χ squared goodness of fit test. We also compare the applicability of out novel distribution with those often used recently as Laplace distribution, generalized extreme value distribution (GEV) and generalized error distribution (GED). The results of our research show that our model and some of its limiting cases are stronger and more viable candidates for modeling returns of cryptocurrencies.

2. A general approach to distribution of logarithmic returns

The logarithmic return of a given currency pair is defined as random variable $X = \log(R_t/R_{t-1})$ where $R_t, R_{t-1} > 0$ are the exchange rates at time t and $t-1$, respectively. We impose these requirements on a distribution of returns:

- its expected value $\mu = \mathrm{E}X$ exists,
- is must be symmetric around μ,
- its moments $\mathrm{E}(X-\mu)^k$ of any order $k \in \mathbb{N}$ exist,
- its must be smooth for all its parameters,
- it should be able to capture the heavy tail property.

Let $\mathrm{g}(y, \mu, \sigma, \mathbf{q})$ be a probability density function (PDF) of a non-negative random variable Y. For a convex open set $\mathcal{D}$ and $\mu \in \mathbb{R}, \sigma > 0, \mathbf{q} \in \mathcal{D}$ as the mean, the scaling values, and a vector of shape parameters $\mathbf{q}$, the PDF of X can be expressed for $x \in \mathbb{R}$ as

$$\mathrm{f}(x; \mathbf{q}, \mu, \sigma) = \frac{\mathrm{g}\left(\frac{|x-\mu|}{\sigma}, \mathbf{q}\right)}{2\sigma}. \tag{1}$$

The corresponding cumulative distribution function (CDF) of X is

$$\mathrm{F}(x; \mathbf{q}, \mu, \sigma) = \frac{1 + \mathrm{sign}(x-\mu) \cdot \mathrm{G}\left(\frac{|x-\mu|}{\sigma}, \mathbf{q}\right)}{2}, \tag{2}$$

where $\mathrm{G}(y, \mathbf{q})$ is the CDF of Y.

The required properties of $\mathrm{f}(x, \mathbf{q}, \mu, \sigma)$ imply the necessary conditions for $\mathrm{g}(y, \mathbf{q})$ for $y > 0, \mathbf{q} \in (\mathcal{D})$:

- $\mathrm{E}Y^k = \int_0^\infty y^k \mathrm{g}(y, \mathbf{q}) \mathrm{d}y < +\infty$ of any order $k \in \mathbb{N}$ exist,
- $\mathrm{g}(y, \mathbf{q})$ is smooth with respect to $y, \mathbf{q}$,
- $\lim_{y \to 0^+} \mathrm{g}(y, \mathbf{q}) < +\infty$ exists,
- $\lim_{y \to 0^+} \frac{\partial \mathrm{g}(y, \mathbf{q})}{\partial y} = 0$,
- $\mathbf{q}$ is able to model the heavy tail property.

Our novel distribution of X is derived from GGD for Y which satisfies several conditions described above. A more detailed characterization of GGD distribution is given in Appendix A.

3. A regularized distribution of logarithmic returns

To achieve our goal, we split domain Y into two intervals $[0, s]$ and $[s, \infty)$, where $s > 0$ is the regularization parameter and $a < 3 + bs^b$. The density of the novel distribution is proportional to the density of the original GGD defined in (A.1) for $y \in [s, \infty)$. However, for $y \in [0, s]$ we replace the original density by a parabola $P + Qy^2$, where P, Q are parameters such that the new PDF at point s is smooth. The replacement is positive only when $a < 3 + bs^b$. The novel distribution has a vector of shape parameters $\mathbf{q} = (a, b, s)$ and it satisfies all our requirements and its PDF is

$$\mathrm{g}(y, a, b, s) = \begin{cases} (P + Qy^2)/R & \text{for} \quad 0 \le y \le s, \\ y^{a-1} \exp(-y^b)/R & \text{for} \quad y > s, \end{cases} \tag{3}$$

where

$$\begin{aligned} P &= (3 - a + bs^b) s^{a-1} \exp(-s^b)/2, \\ Q &= (a - 1 - bs^b) s^{a-3} \exp(-s^b)/2, \\ R &= Ps + Qs^3/3 + \Gamma(a/b, s^b)/b, \end{aligned} \tag{4}$$

Table 1
The family of TRGG distribution and its sub-models.

Model	Distribution	Fixed values
TRGG	Two-sided Regularized Generalized Gamma	
TGG	Two-sided Generalized Gamma	$s \to 0_+$
TRD	Two-sided Regularized Degraded	$b = 2$
RGN	Regularized Generalized Normal	$a = 1$
TRG	Two-sided Regularized Gamma	$b = 1$
TRW	Two-sided Regularized Weibull	$a = b$

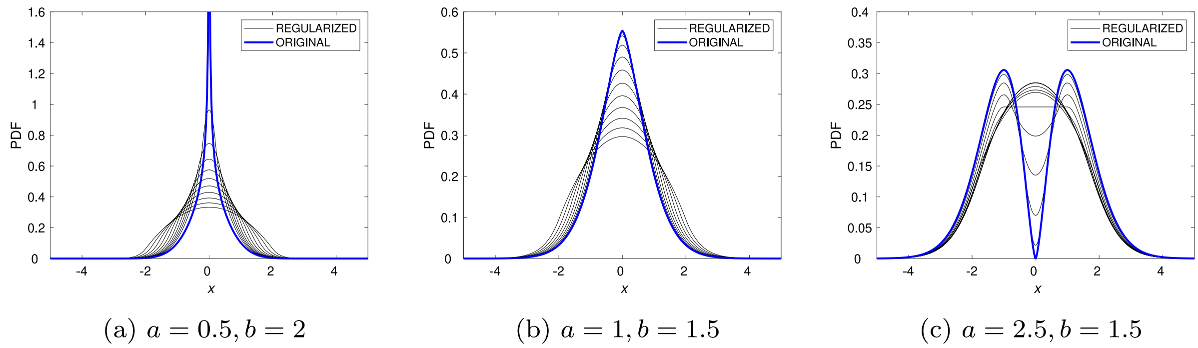

Fig. 1. The regularized PDF and unregularized PDF of a standard TRGG distribution.

where R is the normalization constant. The CDF is

$$\mathrm{G}(y, a, b, s) = \begin{cases} (Py + Qy^3/3)/R & \text{for} \quad 0 \le y \le s, \\ 1 - \Gamma(a/b, y^b)/(bR) & \text{for} \quad y > s. \end{cases} \tag{5}$$

The moments of random variable Y are

$$\mathrm{E}Y^k = \frac{\frac{Ps^{k+1}}{(k+1)} + \frac{Qs^{k+3}}{(k+3)} + \Gamma((a+k)/b, s^b)/b}{Ps + Qs^3/3 + \Gamma(a/b, s^b)/b} < +\infty \tag{6}$$

for any $k > 0$, heavy tail occurs for $0 < b < 1$ again, and the mode $\hat{Y} = 0$ not only for $0 < a \le 1$ but also for $a > 1 \wedge s \ge \left(\frac{a-1}{b}\right)^{1/b}$. Finally, $\lim_{y \to 0^+} \mathrm{g}(y, a, b, s) = \mathrm{g}(0, a, b, s) = P \in [0, +\infty)$ and $\lim_{y \to 0^+} \frac{\partial \mathrm{g}(y,a,b,s)}{\partial y} = 0$.

This distribution can be called as Regularized Generalized Gamma distribution (RGG). It can be directly used for the construction of two-sided distribution of random variable X by substituting (3) and (5) into (1) and (2), respectively. The PDF and CDF of this distribution for a standardized random variable $x = \frac{x-\mu}{\sigma}$ are

$$\mathrm{f}(x, a, b, s) = \frac{1}{2} \begin{cases} (P + Q|x|^2)/R & \text{for} \quad |x| \le s, \\ |x|^{a-1} \exp(-|x|^b)/R & \text{for} \quad |x| > s \end{cases} \tag{7}$$

$$\mathrm{F}(x, a, b, s) = \frac{1}{2} + \frac{\mathrm{sign}(x)}{2} \begin{cases} (P|x| + Q|x|^3/3)/R & \text{for} \quad |x| \le s, \\ 1 - \Gamma(a/b, |x|^b)/(bR) & \text{for} \quad |x| > s, \end{cases} \tag{8}$$

where μ, σ are mean and standard deviation of x, respectively and P, Q, R are defined in (4). This novel distribution is called Two-sided Regularized Generalized Gamma distribution (TRGG). The first moment and median of TRGG distribution are μ. Other odd central moments are zero while the absolute moments of TRGG distribution are

$$\mathrm{E}|X - \mu|^k = \sigma^k \mathrm{E}Y^k < +\infty \tag{9}$$

for any $k > 0$ and even. The PDF of TRGG is uni-modal with $\hat{X} = \mu$ when $0 < a \le 1$ or $a > 1 \wedge s \ge \left(\frac{a-1}{b}\right)^{1/b}$ and bi-modal with $\hat{X} = \mu \pm \left(\frac{a-1}{b}\right)^{1/b}$. Finally, the PDF is smooth in its parameters. The distinction between a regularized and unregularized version of this distribution is visualized and displayed in Fig. 1. In the figure, the left panel shows the case of two-sided generalized gamma distribution (both regularized and unregularized) with $a < 1$, the middle panel displays the case $a = 1$ and the right panel is when $a > 1$. In all three cases, the black lines represent the result of regularization for different values of s.

The TRGG distribution like the GGD has many limiting cases when the shape parameters attain certain values. The most important cases of the TRGG family are summarized in Table 1.

4. Parameter identification and model validation

The TRGG model has five parameters a, b, s, μ, σ. These parameters or parameters of any TRGG sub-model form a vector of parameters $\mathbf{w} \in \mathbb{R}^d$, where $d \leq 5$ is the number of free parameters. A sub-model is obtained when some of parameters of the TRGG model are assigned to a fixed value. The PDF and CDF of TRGG distribution or any sub-model can be formally written as $\mathrm{f}(x, \mathbf{w})$ and $\mathrm{F}(x, \mathbf{w})$, respectively. Instead of the common MLE technique, we estimate them from the CDF with non-linear regression technique. Its advantage is that it does not require the second derivative in the numerical procedure and the derivative of the CDF is the PDF. The objective function is

$$S(\mathbf{w}) = \sum_{i=1}^{n} r_i^2, \tag{10}$$

where $r_i = y_i - \mathrm{F}(x_i, \mathbf{w})$, n is the number of observations, y_i is the empirical CDF of i observations. The point estimates can be obtained numerically with Levenberg–Marquardt algorithm, see Wooldridge (2010) as

$$\mathbf{w}^* = \operatorname*{argmin}_{\mathbf{w} \in \mathbf{W}} S(\mathbf{w}), \tag{11}$$

where $\mathbf{W}$ is the set of all admissible $\mathbf{w}$. The accuracy of the estimates is derived from the so called Jacobian $\mathbf{J} = \frac{\partial r}{\partial \mathbf{w}}$. The estimates of the covariance matrix is:

$$\hat{\Sigma} = \hat{\sigma}^2 (\mathbf{J}' * \mathbf{J})^{-1}, \tag{12}$$

where $\hat{\sigma}^2$ is the estimated variance of r. Subsequently, a vector of asymptotic standard deviations, a vector of z-score and a vector of p-values can be computed.

A model of size d is said to be valid at $\alpha = 0.05$ when the z-score satisfies condition $|z_k| \geq 1.96$ for all estimated parameters. To validify a restriction of a sub-model of size d_{sub}, we use the Likelihood Ratio test (LR test), see Rossi (2018). A valid model of size d is said to be stable when

- all its sub-models of size $d - 1$ are significantly worse according to the LR test at level 0.05,
- all models of size $d + 1$ which consists it as sub-model is not significantly better by the LR test at level 0.05.

Therefore, the significant difference between the log-likelihoods must be greater than 1.9207 for one parameter restriction.

We also employ the Schwartz Information Criterion (BIC) to verify the justifiability of the addition of extra parameters in the case when models are not nested. The value of the criterion is computed as follows

$$BIC = 2LL - k \log(N), \tag{13}$$

where LL is the log-likelihood, k is the number of parameters in the model, and N is the number of observations. The addition of an extra parameter is justifiable if the subsequent increase of LL is greater than $\log(N)/2$.

5. The usability of the proposed distribution for cryptocurrencies

The verification of the suitability of TRGG and its special cases for modeling logarithmic returns of cryptocurrencies is performed with publicly available data from CoinMarketCap website. Three most liquid cryptocurrencies are chosen for this objective. They are Bitcoin, Ethereum and Ripple (XRP). Daily price time series for Bitcoin is available from 26th of April 2013 to 29th of April 2019, for Ethereum and Ripple they are from 7th of August 2015 to 29th of April 2019 and from August 4th of 2013 to 29th of April 2019, respectively. The original price time series are transformed into series of logarithmic returns.

First, all the return series of three cryptocurrencies are used to estimate parameters of the TRGG distribution and its special cases shown in Table 1. As the scaling parameter σ and the regularization parameter s are required to be positive to secure the their non-degeneration, instead of their direct estimation, their corresponding $\log_{10} \sigma$ and $\log_{10} s$ are introduced. The advantage of this approach is that no additional restriction needs to be imposed. The estimation is performed in MATLAB with programs exclusively written for this purpose. Numerical estimation procedure works reliably and provides highly stable results. According to our criterion, all models of our consideration are valid as their estimated parameters always are statistically significant. Due to limited space of this paper, we show the values of loglikelihood function of each sub-model in Table 2 and three control distributions: Laplace, GEV and GED (high values of log-likelihood are in bold). These values are used for testing the validity of sub-models of TRGG by LR test and BIC criterion. The values of BIC criterion are displayed in Table 3.

The values of log-likelihood in Table 2 show that control distributions GEV and Laplace[1] attain too low log-likelihoods and they seem to be unfit for modeling returns of cryptocurrencies and will not be subject of any further comparison. The results of LR test (usable in nested series TRGG, RGN, GED) clearly show that GED is a stable sub-models for Bitcoin. For Ripple, TSGG is best model by testing TRGG vs. any four-parameter distribution and TRGG vs. GED by LR test. It means that symmetrization and regularization bring a statistically significant increase in log-likelihood. For Ethereum, TGG is a stable model therefore the symmetrization of GGD does help. For all three currencies, TRD and TRG are unstable sub-models of TRGG. The BIC criterion allows to directly compare

[1] Laplace distribution is also a special case of the TRGG.

Table 2
The maximum log-likelihood of TRGG family and some alternatives.

Distribution	Bitcoin	XRP	Ethereum
TRGG	**4262.38**	**3221.44**	**1931.34**
TGG	**4261.38**	3218.59	**1930.34**
TRD	4245.02	3154.08	1886.02
RGN	**4261.75**	3219.46	1928.69
TRG	4254.65	3212.22	1924.41
TRW	**4260.91**	3218.93	1927.48
Laplace	4203.81	3092.53	1904.34
GEV	3826.12	2817.98	1673.72
GED (GN)	**4261.71**	3207.50	1926.51

Table 3
The values of BIC criterion of TRGG family and some alternatives.

Distribution	Bitcoin	XRP	Ethereum
TRGG	8486.30	6406.81	3824.45
TGG	8491.99	6406.60	**3831.82**
TRD	8459.28	6277.58	3743.18
RGN	8492.74	**6408.34**	3828.52
TRG	8478.54	6393.86	3819.96
TRW	8491.05	**6407.28**	3826.10
Laplace	8392.24	6170.63	3793.39
GEV	7629.17	5614.32	3324.50
GED (GN)	**8500.35**	6393.36	**3830.08**

Table 4
Estimated values of parameters of TRGG.

Currency	Parameter	Value	Std. error	*z*-score
Bitcoin	a	1.2111	0.0211	10.01
	b	0.5329	0.0094	−49.50
	$\log_{10} s$	−0.5166	0.0810	–
	μ	0.0021	$3.85*10^{-6}$	557.99
	$\log_{10}\sigma$	−2.3588	0.0420	–
XRP	a	2.0341	0.0597	17.33
	b	0.3153	0.0086	−79.63
	$\log_{10} s$	1.5805	0.1843	–
	μ	−0.0022	$6.76*10^{-6}$	−318.07
	$\log_{10}\sigma$	−4.1491	0.1494	–
Ethereum	a	1.4382	0.1010	4.33
	b	0.5204	0.0324	−14.78
	$\log_{10} s$	0.0879	0.0214	–
	μ	$6.63*10^{-5}$	$1.96*10^{-5}$	3.38
	$\log_{10}\sigma$	−2.3018	0.1669	–

all models without being nested condition. As GEV and Laplace distribution are excluded, it will be used to compare GED with distributions with four parameters. The numbers in Table 3 show that GED model dominates other models for Bitcoin and TGG for Ethereum (this is in line with the result of LR test for the nested sequence: TRGG, TGG). For Ripple, the best model is RGN indicating that the regularization helps. The second best sub-model is TRW by BIC criterion.

Next, we display the estimation results in Tables 4–6 only for three models that pass the goodness if fit test best. In the tables, the values for z-score shown in these tables are for the null hypothesis H_0: $a = 1$, H_0: $b = 1$ and H_0: $\mu = 0$. The estimation results reported in Tables 4–6 show that the estimated value of parameter a always is significantly higher than one indicating the existence of a bi-modal distribution. As the TRGG distribution has three shape parameters, when a is fixed to 1, the two remaining parameters can adjust which makes the RGN a good sub-model for Bitcoin. This inference is in line with the results of LR test. However, it does not hold in the case of Ripple and Ethereum, when estimated values of parameter a are too far away from 1. It may work in the similar fashion for s. Though the regularization we have introduced always increases the values of the log-likelihood function. However, it is statistically significant only in the case of XRP when s is large. Otherwise, the two shape parameters a, b can help to carry the effect of regularization. The estimation results are consistent with the fact that GED is a good model for Bitcoin as the estimated value of s for Bitcoin is relatively small and the regularization may not be significant. The estimated values of parameter b in all displayed cases are lower than one. As shown in Appendix B, this means that all series of returns exhibit heavy-tail property.

To investigate the applicability of distributions from TRGG family for modeling returns of cryptocurrencies, we perform the χ squared goodness of fit test. In this test, the theoretical frequencies are compared to the observed ones (for more details on this test, see Huber-Carol et al., 2012). We proceed the test as follows. Taking into consideration the available quantity of data, we decide

Table 5
Estimated values of parameters of TGG.

Currency	Parameter	Value	Std. error	z-score
Bitcoin	a	1.1504	0.0068	22.15
	b	0.5628	0.0046	−95.14
	μ	0.0021	$3.75*10^{-6}$	572.31
	$\log_{10}\sigma$	−2.2357	0.0162	–
XRP	a	1.5454	0.0129	42.18
	b	0.4052	0.0038	−156.91
	μ	−0.0022	$6.55*10^{-6}$	−331.07
	$\log_{10}\sigma$	−3.0009	0.0334	–
Ethereum	a	1.2294	0.0158	14.47
	b	0.6026	0.0104	−39.28
	μ	$6.47*10^{-5}$	$1.93*10^{-5}$	3.35
	$\log_{10}\sigma$	−1.9560	0.0325	–

Table 6
Estimated values of parameters of RGN.

Currency	Parameter	Value	Std. error	z-score
Bitcoin	b	0.6229	0.0025	−148.58
	$\log_{10} s$	−0.3667	0.0140	–
	μ	0.0021	$3.79*10^{-6}$	565.52
	$\log_{10}\sigma$	−2.0111	0.051	–
XRP	b	0.5383	0.0028	−165.00
	$\log_{10} s$	0.1342	0.0136	–
	μ	−0.0021	$7.07*10^{-6}$	−303.38
	$\log_{10}\sigma$	−2.0802	0.0089	–
Ethereum	b	0.6994	0.0070	−42.74
	$\log_{10} s$	0.1324	0.0258	–
	μ	$6.60*10^{-5}$	$1.97*10^{-5}$	3.35
	$\log_{10}\sigma$	−1.6615	0.0112	–

Table 7
χ^2 test results of TRGG family for $N = 100$.

Distribution		Bitcoin	XRP	Ethereum
TRGG	Test statistic	103.32	93.26	111.21
	p-value	0.2398	0.5020	0.1085
TGG	TS	103.32	92.21	111.21
	p-v	0.2626	0.5618	0.1223
TRD	TS	126.34	160.50	129.48
	p-v	0.0174	$< 10^{-16}$	0.0108
RGN	TS	105.15	108.18	108.86
	p-v	0.2236	0.1676	0.2053
TRG	TS	124.15	98.90	102.08
	p-v	0.0240	0.3714	0.2911
TRW	TS	120.31	100.91	119.46
	p-v	0.0407	0.3196	0.0456
Laplace	TS	$4.89 * 10^3$	$4.53 * 10^4$	$8.10 * 10^3$
	p-v	$< 10^{-16}$	$< 10^{-16}$	$< 10^{-16}$
GEV	TS	$1.44 * 10^4$	$4.39 * 10^3$	$2.15 * 10^4$
	p-v	$< 10^{-16}$	$< 10^{-16}$	$< 10^{-16}$
GED	TS	105.12	128.36	122.26
	p-v	0.2463	0.0153	0.0365

to divide the whole domain into 100 and 200 equidistant bins for each distribution which secure the same number of theoretical frequencies computed according to Eq. (8). Also, the two different numbers of bins are chosen to give a sense of generality. The test is applied to the whole TRGG family as well as on those control distributions and the results containing both test statistics and the corresponding p-values are presented in Tables 7 and 8.

The results of the testing displayed in these two tables show that Laplace and GEV distributions are not good for modeling returns of cryptocurrencies. The result is consistent with their low log-likelihood values. Regarding the remaining models, there may exist more distributions which can pass the goodness of fit test for some of three cryptocurrencies. However, there are only three distributions that can reliably be used to model returns of all three cryptocurrencies with the two selected numbers of bins. They are TRGG, TGG and RGN distributions. This inference is consistent with the computed values of log-likelihood displayed in Table 2. Though GED works well for Bitcoin, it is a special case and it does not do well for the other two cryptocurrencies. Hence,

Table 8
χ^2 test results of TRGG family for $N = 200$.

Distribution		Bitcoin	XRP	Ethereum
TRGG	Test Statistic	218.21	200.43	202.82
	p-value	0.1121	0.3605	0.3174
TGG	TS	218.21	189.53	202.82
	p-v	0.1219	0.5970	0.3355
TRD	TS	229.73	251.13	218.43
	p-v	0.0448	0.0041	0.1199
RGN	TS	186.05	200.82	203.70
	p-v	0.6649	0.3724	0.3302
TRG	TS	233.93	175.37	201.35
	p-v	0.0296	0.8401	0.3625
TRW	TS	218.76	170.02	226.68
	p-v	0.1168	0.9013	0.0596
Laplace	TS	$2.31 * 10^3$	$1.78 * 10^3$	$2.20 * 10^3$
	p-v	$< 10^{-16}$	$< 10^{-16}$	$< 10^{-16}$
GEV	TS	$2.15 * 10^3$	$5.61 * 10^3$	$4.17 * 10^3$
	p-v	$< 10^{-16}$	$< 10^{-16}$	$< 10^{-16}$
GED	TS	185.92	243.83	231.75
	p-v	0.6859	0.0114	0.0410

our model and its special cases as such can be a superior alternative to previously known distributions for modeling returns of three chosen cryptocurrencies. Our distributions exhibit heavy tail nature but have finite variance as shown in Section 4, hence, they can be a beneficial boost for risk management and other applications in financial engineering.

6. Conclusions

With respect to the current position of cryptocurrencies in the financial market, a proper model for returns of cryptocurrencies is important. We propose the novel two-sided regularized TRGG and its sub-models as alternatives to solve this problem. The generalized gamma distribution has been symmetrized and regularized to obtain a fully differentiable PDF with respect to both parameters and the independent variable. Then, the basic statistical properties have been explicitly derived from its density. This novel model and a family of its sub-models are applied to model returns of three most liquid cryptocurrencies. The parameters of the full model and sub-models have been estimated from data with an inventive nonlinear least squares technique with consequent parameter and model significance testing.

Despite the fact that GED can be used to model the returns of bitcoins, our results show that TRGG and its two sub-models can reliably model returns of three cryptocurrencies. As these sub-models can be as good as their parent, the TRGG model, hence, they should be prioritized due to their lower number of free parameters. However, the main contribution of the TRGG model is its generalization ability and it can be useful for modeling the returns of other financial assets as well as for applications in other fields. We believe that our work will help to improve the results in the works like in Silahli et al. (2019).

Acknowledgments

The authors thank for financial support of grant OP VVV MEYS, Czechia funded project CZ.02.1.01/0.0/0.0/16019/0000765 Research Center for Informatics as well as grant SGS20/190/OHK4/3T/14 of CTU in Prague and the Institutional Support IP 100040 of University of Economics and Business in Prague.

Appendix A. A summary of standardized GG distribution

Let Y be a positive random variable and $a, b > 0$ be parameters of the GGD also known as Amoroso (1925) distribution with $a = \alpha\beta, b = \beta$. The Amoroso distribution was originally developed for survival analysis and later it was used in many social and technical applications as shown in Kleiber and Kotz (2003) and Pham and Almhana (1995). Its standardized PDF is

$$g(y, a, b) = \frac{b y^{a-1} \exp(-y^b)}{\Gamma(a/b)} \tag{A.1}$$

for $y > 0$. The corresponding CDF is

$$G(y, a, b) = 1 - \frac{\Gamma(a/b, y^b)}{\Gamma(a/b)}, \tag{A.2}$$

where

$$\Gamma(p, \xi) = \int_{\xi}^{\infty} t^{p-1} \exp(-t) \mathrm{d}t, \quad \Gamma(p) = \Gamma(p, 0). \tag{A.3}$$

A GGD has many limiting cases, but it also is a special case of other distributions, see Crooks (2010). The moments of the GGD are finite

$$\mathrm{E}Y^k = \frac{\Gamma((a+k)/b)}{\Gamma(a/b)} < +\infty \tag{A.4}$$

for any $k > 0$. The second useful feature of this distribution is its heavy tail property for $0 < b < 1$ (see the Appendix) which means for $\lambda > 0$

$$\lim_{y\to+\infty} \exp(\lambda y)(1 - \mathrm{G}(y, a, b)) = +\infty. \tag{A.5}$$

The mode $\hat{Y} = 0$ only for $0 < a \le 1$, but otherwise $\hat{Y} = \left(\frac{a-1}{b}\right)^{1/b} > 0$. But the GGD also has undesirable properties:

- $\lim_{y\to 0^+} \mathrm{g}(y, a, b) < +\infty$ only for $a \ge 1$,
- $\lim_{y\to 0^+} \frac{\partial \mathrm{g}(y,a,b)}{\partial y} = 0$ only for either $a > 2$ or $a = 1 \wedge b > 1$.

The novel distribution of Y will be designed as a locally modified GGD which eliminates the above mentioned deficiencies.

Appendix B. Heavy-tail property of TRGG

We will show the heavy tail property of RGG and TRGG automatically retains this property from RGG. According to Bryson (1974) we have to prove

$$\Psi = \exp(\lambda x)(1 - \mathrm{G}(x, a, b, s)) \to +\infty \tag{B.1}$$

for $\lambda > 0$ and where $\mathrm{G}(x, a, b, s)$ is the CDF of the RGG distribution. In order to do so, we use the asymptotic formula for Gamma function for $0 < b < 1$ as

$$\Gamma(p, \xi) = \xi^{p-1} \exp(-\xi) \cdot (1 + \mathrm{O}(\xi^{-1})) \tag{B.2}$$

for $\xi \to +\infty$. When $x \to +\infty$, then $x > s$, inserting (5) into (B.1), we get

$$\Psi = \exp(\lambda x)\Gamma(a/b, x^b)/bR. \tag{B.3}$$

Substituting (B.2) into (B.3) and taking logarithm of it, we obtain

$$\ln \Psi = \lambda x - x^b + (a - b)\ln x + \ln(1 + \mathrm{O}(x^{-b})) - \ln(bR). \tag{B.4}$$

Term λx is dominant for large x as $0 < b < 1$ which implies

$$\lim_{x\to+\infty} \ln \Psi = +\infty \tag{B.5}$$

for all $\lambda > 0$. Q.E.D.

References

Akhtaruzzaman, M., Sensoy, A., Corbet, S., 2020. The influence of bitcoin on portfolio diversification and design. Finance Res. Lett. 37, 101344.
Akyildirim, E., Goncu, A., Sensoy, A., 2021. Prediction of cryptocurrency returns using machine learning. Ann. Oper. Res. 297, 3–36.
Amoroso, L., 1925. Ricerche intorno alla curva dei redditi. Ann. Mat. Pura Appl. (4) 2 (1), 123–159.
Aslan, A., Sensoy, A., 2020. Intraday efficiency-frequency nexus in the cryptocurrency markets. Finance Res. Lett. 35, 101298.
Bouri, E., Shahzad, S.J.H., Roubaud, D., 2019. Co-explosivity in the cryptocurrency market. Finance Res. Lett. 29, 178–183.
Bryson, M., 1974. Heavy tailed distributions: Properties and tests. Technometrics 16 (1), 61–68.
Corbet, S., Eraslan, V., Lucey, B., Sensoy, A., 2019. The effectiveness of technical trading rules in cryptocurrency markets. Finance Res. Lett. 31, 32–37.
Crooks, G.E., 2010. The amoroso distribution. arXiv, 1005.3274.
Drozdz, S., Gebarowski, R., Minati, L., Oswiecimka, P., Watorek, M., 2018. Bitcoin market route to maturity? Evidence from return fluctuations, temporal correlations and multiscaling effects. Chaos 28 (7), 071101.
Huber-Carol, C., Balakrishnan, N., Nikulin, M., Mesbah, M., 2012. Goodness-of-Fit Tests and Model Validity. Springer Science & Business Media.
Kleiber, C., Kotz, S., 2003. Statistical Size Distributions in Economics and Actuarial Sciences. John Wiley & Sons.
Kukal, J., Tran, Q.V., 2019. Voigt distribution and its heavy-tail modeling ability for cryptocurrencies. In: Houda, M., Remes, R. (Eds.), Proceedings of 37th International Conference on Mathematical Methods in Economics, 2019. University of South Bohemia in Ceske Budejovice, pp. 139—144.
Long, H., Zaremba, A., Demir, E., Szczygielski, J.J., Vasenin, M., 2020. Seasonality in the cross-section of cryptocurrency returns. Finance Res. Lett. 35, 101566.
Nadarajah, S., 2005. A generalized normal distribution. J. Appl. Stat. 32 (7), 685–694.
Nolan, J.P., 1997. Numerical calculation of stable densities and distribution functions. Commun. Stat. Stoch. Model. 13, 759—774.
Osterrieder, J., Lorenz, J., 2017. A statistical risk assessment of bitcoin and its extreme tail behavior. Ann. Financial Econ. 12 (01), 1750003.
Pham, T., Almhana, J., 1995. The generalized gamma distribution: Its hazard rate and stress-strength model. IEEE Trans. Reliab. 44 (3), 392—397.
Prause, K., 1999. The Generalized Hyperbolic Model: Estimation, Financial Derivatives, and Risk Measures (Ph.D. Thesis). Freiburg University.
Punzo, A., Bagnato, L., 2021. Modeling the cryptocurrency return distribution via Laplace scale mixtures. Physica A 563, 125354.
Rossi, R.J., 2018. Mathematical Statistics: An Introduction to Likelihood Based Inference. John Wiley & Sons, New York.
Sensoy, A., 2019. The inefficiency of Bitcoin revisited: A high-frequency analysis with alternative currencies. Finance Res. Lett. 28, 68–73.
Silahli, B., Dingec, K.D., Cifter, A., Aydin, N., 2019. Portfolio value-at-risk with two-sided Weibull distribution: Evidence from cryptocurrency markets. Finance Res. Lett. 101425.
Stacy, E.W., 1962. A generalization of the gamma distribution. Ann. Math. Stat. 33 (3), 1187–1192.
Szczygielski, J.J., Karathanasopoulos, A., Zaremba, A., 2020. One shape fits all? A comprehensive examination of cryptocurrency return distributions. Appl. Econ. Lett. 27 (19), 1567–1573.
Theodossiou, P., 1998. Financial data and the skewed generalized T distribution. Manage. Sci. 44 (12), 1650–1661.
Vidal-Tomás, D.A.M., Ibáñez, A.M., Farinós, J.E., 2019. Herding in the cryptocurrency market: CSSD and CSAD approaches. Finance Res. Lett. 30, 181–186.
Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data, 2e. MIT Press.