Check for updates

Received: 12 December 2022 | Accepted: 20 April 2023

DOI: 10.1002/sta4.580

ORIGINAL ARTICLE

WILEY

Deep variance gamma processes

Caitlin M. Berry | **William Kleiber**

Department of Applied Mathematics, University of Colorado-Boulder, Boulder, Colorado, 80309, USA

Correspondence
William Kleiber, Department of Applied Mathematics, University of Colorado-Boulder, Boulder, CO 80309, USA.
Email: william.kleiber@colorado.edu

Funding information
Division of Mathematical Sciences; National Science Foundation

Lévy processes are useful tools for analysis and modeling of jump-diffusion processes. Such processes are commonly used in the financial and physical sciences. One approach to building new Lévy processes is through subordination, or a random time change. In this work, we discuss and examine a type of multiply subordinated Lévy process model that we term a deep variance gamma (DVG) process, including estimation and inspection methods for selecting the appropriate level of subordination given data. We perform an extensive simulation study to identify situations in which different subordination depths are identifiable and provide a rigorous theoretical result detailing the behavior of a DVG process as the levels of subordination tend to infinity. We test the model and estimation approach on a data set of intraday 1-min cryptocurrency returns and show that our approach outperforms other state-of-the-art subordinated Lévy process models.

KEYWORDS
Bitcoin, cryptocurrency, Ethereum, Lévy process, subordination

1 | INTRODUCTION

Modern financial datasets are often modeled using Lévy processes due to the non-Gaussian and uncorrelated price fluctuations, as well as apparent jumps. A Lévy process is a process with stationary and independent increments that is stochastically continuous. Such processes share an intimate relationship with infinitely divisible distributions, which can lead to fruitful theoretical investigations (Sato, 1999). Although Lévy processes have been explored since the 1930s (Sato, 2001), new application areas such as high frequency cryptocurrency returns can challenge existing frameworks (Shirvani et al., 2022). One popular method for building new Lévy processes is through subordination, or stochastic time change. Subordination leads to popular classes of Lévy models such as the variance gamma, α-stable, and the normal inverse Gaussian process, details of which can be found in Cont and Tankov (2004). The motivation for the use of stochastic time change in the context of financial modeling comes from its ability to capture certain characteristics of returns like their non-Gaussianity and correlation with volatility (Carr & Wu, 2004).

The class of variance gamma models has been of interest for decades; Madan and Seneta (1990) introduced the process for modeling uncertainty of security prices and generalized this idea in Madan et al. (1998) with a skewness parameter. Barndorff-Nielsen and Shepard (2006) studied the inconsistency of realized variance as an estimator for the time-change in time-deformed Lévy processes. More recently, Aguilar et al. (2020) compared subordinated models like the variance gamma and normal inverse Gaussian process to an alternative approach, which they term a fractional diffusion model, in calculating risk sensitivities and profit-and-loss explanation, a slightly different pursuit than we investigate here in modeling cryptocurrency returns.

In this work, we begin with the idea of a time-changed Lévy process and explore a particular multiply-subordinated process that we term the deep variance gamma (DVG) process (DVG), in which a variance gamma process is subordinated iteratively with further gamma processes. Recently, Shirvani et al. (2021) proposed this idea of building new Lévy processes by using multiple levels of subordination in which the stochastic time change process itself experiences a stochastic time change. Starting with their idea, we investigate further and analytically show that, under certain conditions on the subordinators, infinite depth of subordination leads to a degenerate process. For moderately deep subordination, the DVG allows for flexibility in having highly non-normal densities that can capture very small values, while allowing for heavy tails, making this a good candidate model for cryptocurrency return data. We provide new discussion on estimation and identification of DVG processes, including a

 © 2023 John Wiley & Sons Ltd.

visual tool for selecting the appropriate depth of subordination given data, as well as detailed simulation study illustrating that moderately deep DVGs can be distinguished from standard variance gamma processes. In the vein of multiple subordination, Shirvani et al. (2022) had success developing a doubly subordinated Lévy process for Bitcoin volatility data; we consider an application of the DVG to a 1-min cryptocurrency returns dataset and show that the model with our approach better fits the distribution of returns than the recently proposed doubly subordinated Lévy process (Shirvani et al., 2022).

2 | LÉVY PROCESS BACKGROUND

Lévy processes are a wide class of models for stochastic processes with stationary and independent increments. A stochastic process $Z(t), t \geq 0$, is a Lévy process if it satisfies the following properties:

1. $Z(0) = 0$ almost surely.
2. For any choices of $0 \leq t_1 < t_2 < \ldots < t_n$, the random variables $Z(t_1), Z(t_2) - Z(t_1), \ldots, Z(t_n) - Z(t_{n-1})$ are independent.
3. The distribution of $Z(t+h) - Z(t)$ does not depend on h.
4. Z is stochastically continuous.

Lévy processes share an equivalence to the class of infinitely divisible distributions in that $Z(t)$ is an infinitely divisible random variable for any t.

The Lévy–Khintchine representation is a celebrated result giving the general form of the characteristic function for a Lévy process.

Theorem 1 Lévy–Khintchine. *Let $Z(t)$ be a Lévy process on $\mathbb{R}$, then*

$$\begin{aligned}\mathbb{E}\exp(i\omega Z(t)) &= \exp\left(-\frac{t}{2}\sigma^2\omega^2 + it\gamma\omega + t\int\left[e^{i\omega x} - 1 - i\omega x \mathbb{1}_{[|x|\leq 1]}(x)\right]\nu(\mathrm{d}x)\right)\\ &= \exp(t\Psi(\omega))\end{aligned}$$

for $\omega \in \mathbb{R}$, where $\gamma \in \mathbb{R}, \sigma \geq 0$, and ν is a measure on $\mathbb{R}$ such that $\nu(\{0\}) = 0$ and $\int(|x|^2 \wedge 1)\nu(\mathrm{d}x) < \infty$.

In this theorem, ν is the Lévy measure, Ψ is the characteristic exponent, and (γ, σ, ν) is often referred to as the characteristic triplet of Z. Here, γ is a drift term and σ^2 represents the variance of the Brownian motion component (Cont & Tankov, 2004) while ν controls the jump behavior.

There are many types of Lévy processes, but one particular class we will utilize throughout this work is the class of gamma processes. Consider a gamma random variable with density $\mu(x) = \lambda^\alpha \Gamma(\alpha)^{-1} x^{\alpha-1}\exp(-\lambda x)$ for $x > 0$; we refer to this as a Gamma(α, λ) distribution with shape parameter $\alpha > 0$ and rate parameter $\lambda > 0$. A gamma process, $Z(t)$, is a Lévy process such that $Z(1)$ is Gamma(α, λ); thus, the distribution of $Z(t)$ is Gamma$(t\alpha, \lambda)$.

One important type of Lévy process is a subordinator; a Lévy process $S(t)$ is a subordinator if it is nonnegative and increasing in t almost surely. Subordinators can be used to create new Lévy processes through a random time change: If $Z(t)$ is a Lévy process and $S(t)$ a subordinator, then $X(t) = Z(S(t))$ is also a Lévy process. Subordinated Lévy processes can be interpreted as time-changed processes where the time change is random; in the financial literature, such random time changes are often interpreted as market time, which speeds up and slows down according to the stochastic subordinator. For example, the gamma process described above is a subordinator.

A convenient representation exists for a nonnegative Lévy process using the Laplace transform. The Laplace transform of a nonnegative measure μ is $L(u) = \int_0^\infty e^{-ux}\mu(\mathrm{d}x)$ for $u \geq 0$. For a nonnegative Lévy process $S(t)$, the Laplace transform can be written $\mathbb{E}\exp(-uS(t)) = \exp(t\ell(u))$, where we refer to $\ell(u)$ as the Laplace exponent of $S(t)$. The Laplace transform of a gamma random variable is

$$L(u) = \left(\frac{1}{1+\lambda^{-1}u}\right)^\alpha$$

and so the Laplace exponent of a gamma process is

$$\ell(u) = -\alpha\log(1+\lambda^{-1}u) \tag{1}$$

which will be useful in the next section.

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.580 by University Of Colorado Librari, Wiley Online Library on [28/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3 | DEEP VARIANCE GAMMA PROCESSES

A favored way of generating new types of Lévy processes is through subordination. The following proposition, the proof of which appears in Cont and Tankov (2004), tells us the effect of subordination on the Lévy exponent.

Proposition 1. *If $Z(t)$ is a Lévy process with characteristic exponent $\Psi(\omega)$, and $S(t)$ is a nonnegative subordinator with Laplace exponent $\ell(u)$ independent of $Z(t)$, then $Y(t) = Z(S(t))$ is a Lévy process. Moreover, $Y(t)$ has characteristic function*

$$\mathbb{E}\exp(i\omega Y(t)) = \exp(t\ell(-\Psi(\omega))). \tag{2}$$

Under the conditions of Proposition 1, suppose $S(t)$ is a gamma process with parameters (α,λ) and $Z(t)$ is a Brownian motion with zero drift and variance σ^2; then $Y(t) = Z(S(t))$ is termed a *variance gamma process*. A variance gamma is a time changed Brownian motion, where the time randomly dilates or contracts according to the gamma process; in the financial literature, the gamma process is sometimes referred to as "market time." The variance gamma has characteristic function

$$\mathbb{E}\exp(i\omega Y(t)) = \exp\left(-t\alpha\log\left(1+\frac{\sigma^2}{2\lambda}\omega^2\right)\right). \tag{3}$$

We can readily see from Equation (3) that the variance gamma is overparameterized in that changes in the rate λ of $S(t)$ are equivalent to changes in the variance of the driving Brownian motion. Thus, we follow Madan and Seneta (1990) and Madan et al. (1998) by setting $\lambda = \alpha$ so that $\mathbb{E}S(t) = t$ for all $t \geq 0$.

Proposition 2 can readily be generalized to produce a large class of useful new Lévy processes, contained in the following theorem.

Theorem 2. *If $Z(t)$ is a Lévy process with characteristic exponent $\Psi(\omega)$ and $S_1(t),\ldots,S_L(t)$ nonnegative subordinators that are mutually independent and independent of $Z(t)$ with Laplace exponents $\ell_1,\ldots,\ell_L$, respectively, then $Y(t) = Z(S_1(S_2(\ldots S_L(t))))$ is a Lévy process with characteristic function*

$$\mathbb{E}\exp(i\omega Y(t)) = \exp(t\ell_L(-\ell_{L-1}(\cdots(-\ell_1(-\Psi(\omega)))))) \tag{4}$$

for all ω and t.

Proof of this theorem can be found in Appendix A.

We are now set up to define the main object of this study. Suppose $Z(t)$ is a Brownian motion with no drift and variance σ^2 at $t = 1$ and let $S_1(t),\ldots,S_L(t)$ be independent gamma subordinators where $S_k(1)$ has parameter α_k for $k = 1,\ldots,L$. We define the *deep variance gamma process of order L* to be $Y(t) = Z(S_1(S_2(\ldots(S_L(t)))))$. This multiply-subordinated process allows for great flexibility in density function shapes including heavy tails and peaked density around zero that commonly occur in financial data.

By Theorem 2, the characteristic exponent of $Y(t)$ is straightforward,

$$-\alpha_L\log\left(1+\frac{\alpha_{L-1}}{\alpha_L}\log\left(1+\frac{\alpha_{L-2}}{\alpha_{L-1}}\log\left(\cdots\frac{\alpha_1}{\alpha_2}\log\left(1+\frac{\sigma^2}{2\alpha_1}\omega^2\right)\right)\right)\right). \tag{5}$$

Unfortunately, the Fourier inverse of (4) is not available for $L \geq 2$ for the deep variance gamma process, so likelihood-based estimation methods are challenging.

Figure 1 shows a set of characteristic function (chf) and probability density function (pdf) pairs for $L = 2,3$, varying some possible combinations of α_k parameters. Some interesting patterns emerge; for the $L = 2$ case, smaller pairs of α lead to a peaked pdf with heavier tails in the chf, while larger values dampen the tails of the chf, leading to a broader, lighter-tailed pdf. In the $L = 3$ case, a similar story can be seen with large and small values of the αs corresponding to peaked/spread out pdfs; here, we can see that changes in α_1 tend to have a greater effect on the shape of the chf/pdf than individually changing the higher-level α_2 or α_3.

A natural inquiry is to consider the behavior of a deep variance gamma process as $L \to \infty$. For ease of exposition, we consider the special case when $\alpha_k = \alpha$ for all $k = 1,\ldots,L$, and behavior of (5) as $L \to \infty$. Shirvani et al. (2021) hypothesized that the distribution of $Y(t)$ converges to a point mass but did not provide a rigorous proof; we will establish a similar result but with proof (see Appendix A), in the following theorem.

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.580 by University Of Colorado Librari, Wiley Online Library on [28/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

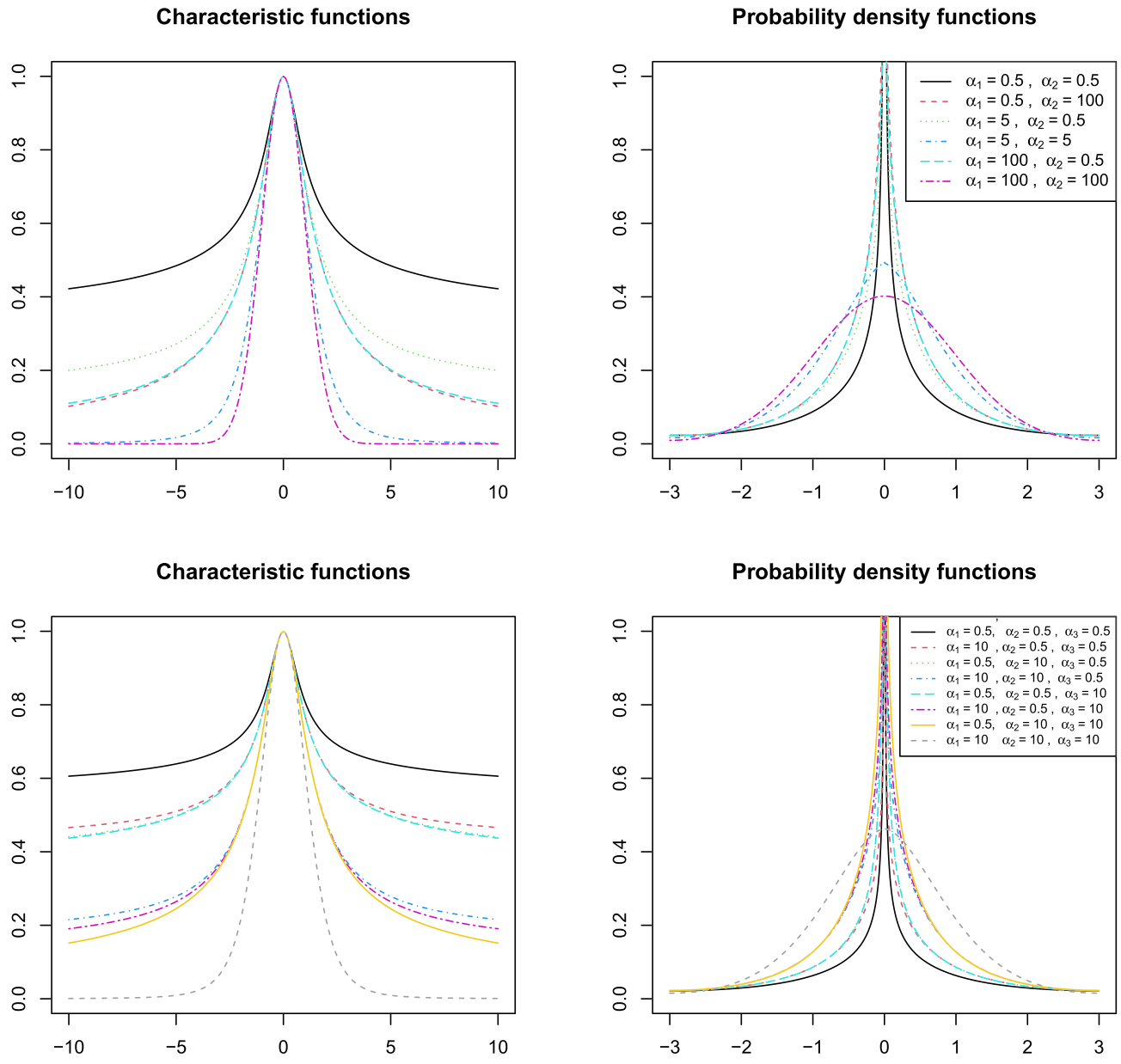

FIGURE 1 Examples of characteristic function and probability density function pairs for the deep variance gamma process with $\sigma = 1$ for depths $L = 2,3$, varying values of $\{\alpha_k\}$

Theorem 3. *Suppose $Y(t)$ is a deep variance gamma process of order L with $\alpha_k \le \alpha_{k+1} \le \Lambda$ for all k and some finite Λ. As $L \to \infty$, the characteristic function of $Y(t)$, $\mathbb{E}\exp(i\omega Y(t))$, converges to unity for all ω and all t.*

According to Theorem 3, as $L \to \infty$ the characteristic function of a DVG converges to a constant everywhere as the levels of composition increase; thus, the distribution of $Y(t)$ converges to degenerate Dirac delta point mass at zero in the limit of infinite subordination. Figure 1 illustrates this claim to some extent with the $L = 3$ pdfs becoming more peaked around zero with the addition of greater levels of depth.

4 | ESTIMATION AND SIMULATION STUDIES

In this section, we discuss an estimation approach for DVG processes and present a detailed set of simulation studies designed to assess our ability to recover DVG parameters from data.

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.580 by University Of Colorado Librari, Wiley Online Library on [28/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4.1 | Estimation

Estimation of Lévy process parameters from data is a topic of much discussion. When probability density functions (pdfs) are analytically available, such as for the variance gamma process, standard maximum likelihood techniques can be used (Barndorff-Nielsen & Blaesild, 1981; Cont & Tankov, 2004; Sueishi & Nishiyama, 2005), although Honore (1998) warns of potential difficulties for unbounded likelihoods. Unfortunately, for DVG of orders $L > 1$, analytic pdfs are unavailable. In such cases, other estimation approaches have been proposed, such as maximum empirical likelihood (Elgin, 2011), and nonparametric methods (Figueroa-López, 2009, 2011).

For the DVG, the characteristic function is available, so we propose following a least squares characteristic function estimation approach that has seen success for Lévy and other types of distributions (Kappus & Reiß, 2011; Xu & Darve, 2020; Yu, 2007). Let $c(\omega,\boldsymbol{\theta}_L,L)$ be a theoretical chf of the DVG with depth L and corresponding parameters $\boldsymbol{\theta}_L$ at frequency $\omega \in \mathbb{R}$, and $c_n(\omega)$ be an empirical chf defined by

$$c_n(\omega) = \frac{1}{n}\sum_{j=1}^{n} \exp(i\omega X_j)$$

where $X_1,\ldots,X_n$ are independent and identically distributed random variables and i is the imaginary number. The least squares chf method uses $\hat{\boldsymbol{\theta}}_L$ that minimizes

$$\ell(\boldsymbol{\theta}_L,L) = \sum_{k=1}^{K} |c(\omega_k,\boldsymbol{\theta}_L,L) - c_n(\omega_k)|^2 \tag{6}$$

where $|\cdot|$ is the complex modulus and $\{\omega_k\}_{k=1}^{K}$ is a set of user-specified frequencies.

A separate issue in estimating DVG parameters is in choosing the appropriate depth level, L. We propose a heuristic approach (seen to work well in the simulation studies and examples below) in which models are fit sequentially with $L = 1,2,\ldots$, each of which results in a minimizing value of (6), $\ell(\hat{\boldsymbol{\theta}}_L,L)$. Adding in additional levels of depth at small L, we expect to gain additional flexibility in the probability model, however, according to Theorem 3, there will be a point of diminishing returns at which $\ell(\hat{\boldsymbol{\theta}}_L,L)$ grows with L. We propose a visual approach by plotting values of $\ell(\hat{\boldsymbol{\theta}}_1,1),\ell(\hat{\boldsymbol{\theta}}_2,2),\ldots$, and identifying the inflection point at which increasing L either offers no improvement in model fit, or becomes detrimental to model fit. Another approach is to examine the relative improvement as levels are increased, $1-\ell(\hat{\boldsymbol{\theta}}_{L+1},L+1)/\ell(\hat{\boldsymbol{\theta}}_L,L)$, choosing an L such that relative improvement is minimized, or becomes negative.

4.2 | Simulation studies

It is not clear a priori that the least squares chf approach will yield reasonable estimates of DVG process parameters, or to what extent sample sizes have an effect on quality of estimates. It is also unclear that, given data, we can easily distinguish between different potential subordination depths. In this section, we investigate a detailed set of simulation studies to first assess our ability to recover parameters, and second to distinguish between different levels of depth for the DVG process.

To assess estimation efficiency, we consider different sample sizes and parameter values for DVG processes with depths $L = 2$ or 3. In particular, consider sample sizes $n \in \{2000,5000,10000\}$ and $\alpha_k \in \{0.5,1,2,5,10,100\}$ for $k = 1,2,3$. For the purposes of this simulation study, we assume $\sigma^2 = 1$ is fixed and known. We set a grid of frequencies at $\omega \in \{-40 + i/2\}_{i=0,\ldots,160}$; in exploratory analyses, we considered other grid resolutions and domains but found that the results were not manifestly affected. Each experiment with a possible combination of sample size and αs is repeated 20 times.

We begin by assessing the ability of the least squares chf method to recover parameter values. Figures 2 and 3 show logarithmic scatterplots of least squares estimates of the α parameters over all experiments colored by sample size. The least squares chf method produces reasonable estimates of DVG parameters at both levels, with greater variability and uncertainty for higher values of α, regardless of level. In view of Figure 1, the variability at higher values of α is due to the flattening of the chf tails for large frequencies, requiring more data to tease apart possible chf shapes at lower frequencies.

In the depth $L = 3$ DVG experiment, Figure 4 considers which pairs of values for $(\alpha_1,\alpha_2,\alpha_3)$ the correct depth model of $L = 3$ has a lower minimizing value of (6) over the best models from depths 1, 2, and 4. The goal is to identify parameter combinations for which it is easier to distinguish between the correct and incorrect models. Generally, most situations correctly identify that $L = 1$ is inappropriate except for the largest pairs of α, due to the flattening of the chf tails for these values. The second row indicates that any parameter combinations with α_1 small are difficult to disentangle from $L = 4$, along with combinations having large values of α_2 and α_3. However, for larger values of α_1, and small-to-moderate values of α_2 and α_3, there is good ability to identify $L = 3$ as the correct depth compared to $L = 4$. The final row suggests is it generally more difficult to distinguish between $L = 3$ and $L = 4$.

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.580 by University Of Colorado Librari, Wiley Online Library on [28/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

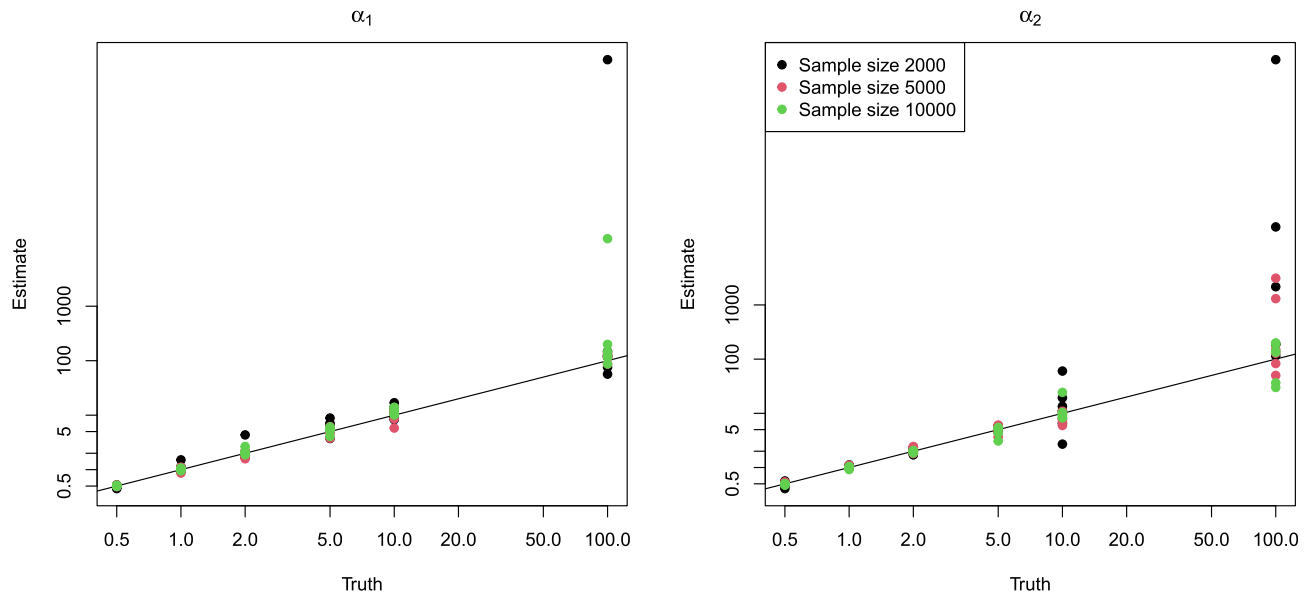

FIGURE 2 Least squares chf estimates of α_1 and α_2 for the first simulation study with $L = 2$, colored by sample size.

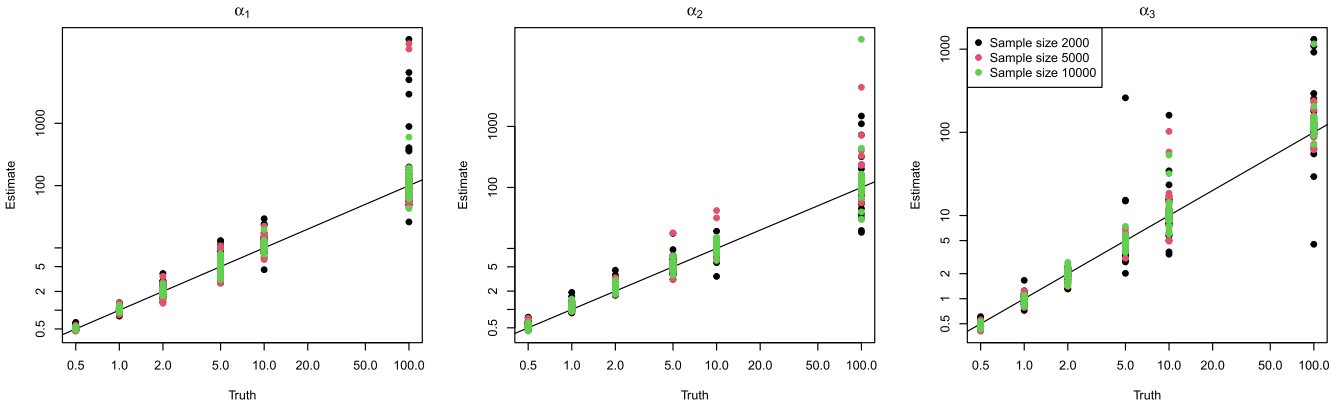

FIGURE 3 Least squares chf estimates of α_1, α_2, and α_3 for the first simulation study with $L = 3$, colored by sample size.

5 | APPLICATION TO HIGH FREQUENCY CRYPTOCURRENCY RETURNS

Cryptocurrencies have become highly popular within the last decade. Led by the popularity of Bitcoin, hundreds of new blockchain currencies have been introduced, some of the most common include Ethereum, Ripple, and Litecoin. Although each has its own unique features, these currencies share the power of blockchain's peer-to-peer distributed ledger technology, in which public records of each transaction are maintained in a chain. Bitcoin, invented by Satoshi Nakamoto, began trading in 2009 while Ethereum began public transactions in 2015. Although Ethereum has smart contract functionality, the ability to execute contracts in an automated way, its associated currency, Ether, is often referred to as Ethereum, the parent blockchain.

Cryptocurrencies are not beholden to traditional stock market business hours and are traded continuously 24 h a day. There is interest in the high-frequency behavior of cryptocurrency returns: Cocco et al. (2017) develop a statistical model to simulate artificial daily Bitcoin prices; Chu et al. (2020) examine high frequency momentum trading using hourly cryptocurrency prices; Celeste et al. (2020) fit wavelet models and examine volatility of Ethereum and Bitcoin daily returns. Challenges include the highly non-normal behavior of high-frequency returns often with heavy tails and a sharp peak around zero (Petukhina et al., 2021; Zhang et al., 2019).

In spite of their clear utility for cryptocurrency modeling, Lévy processes have only been recently explored (Philippas et al., 2019). Shirvani et al. (2022) propose using a doubly subordinated normal inverse Gaussian Lévy process for daily Bitcoin return data. We push this idea further and show that deeper subordination using the DVG provides better fits of the small log-returns, but also captures tail behavior better than the model of Shirvani et al. (2022).

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.580 by University Of Colorado Librari, Wiley Online Library on [28/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

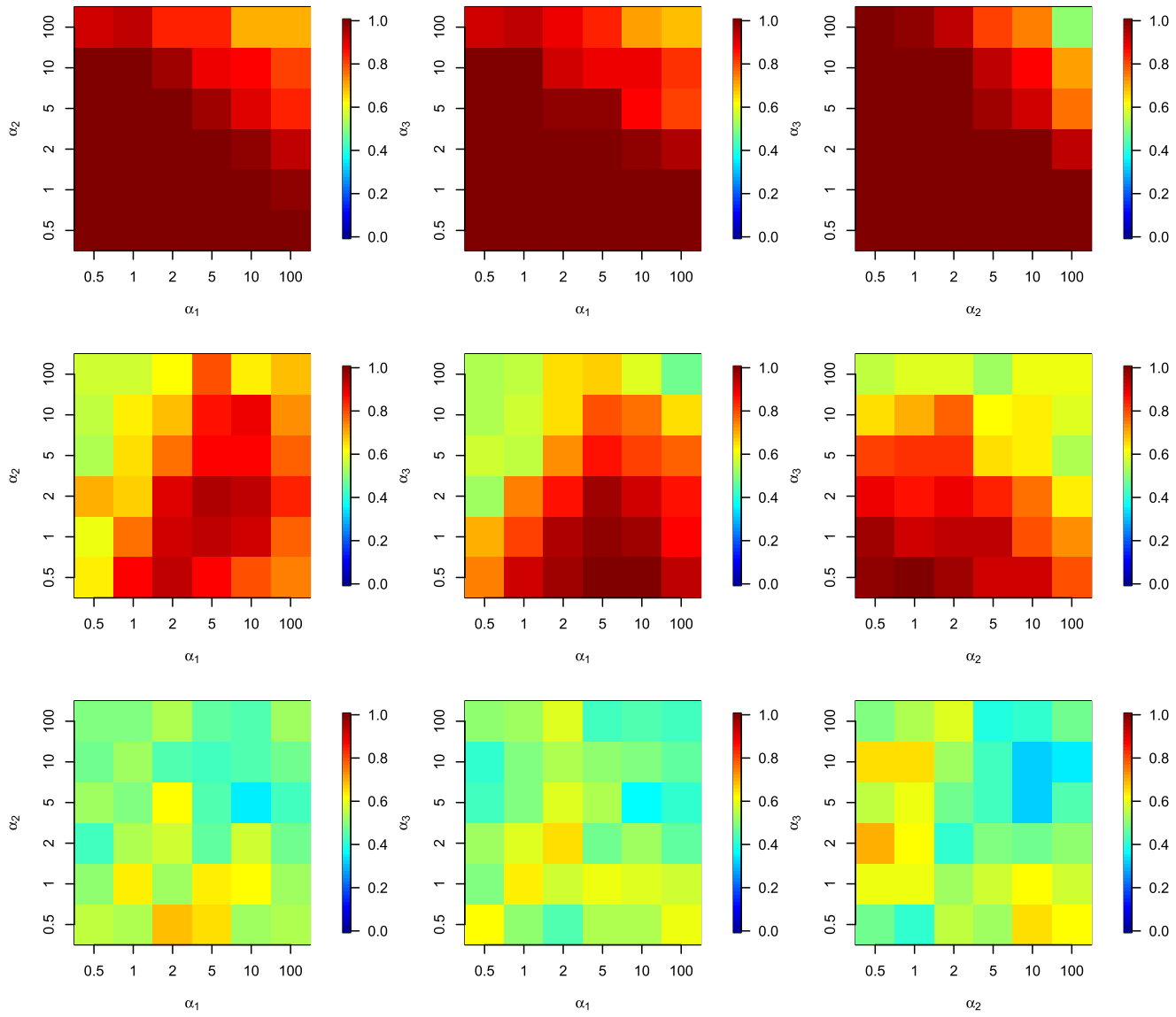

FIGURE 4 Percent of simulations in which the correct 3-depth model is favored over the competing depths of 1 (top row), 2 (middle row), or 4 (bottom row) in terms of minimizing value of (6), split by pairs of $(\alpha_1, \alpha_2, \alpha_3)$

We consider two of the most popular cryptocurrencies: Ethereum and Bitcoin. High-frequency 1-min price data were downloaded from Kaggle (Klein, 2022); for both currencies, we consider a time period of 10,000 values covering approximately one week between October 31, 2021, to November 7, 2021. We model log-returns; that is, if $P(t)$ is the price at time point t, we model $Z(t) = \log(P(t)) - \log(P(t-1))$. As is common with investment return data, exploratory analyses suggest minimal autocorrelation between returns, and highly non-Gaussian behavior (see Figure 5). We thus model the log-return data as a Lévy process, and consider both the proposed DVG model, as well as a recent proposal by Shirvani et al. (2022), the normal double inverse Gaussian (NDIG) Lévy process. Both models are fit using the least-squares chf approach of Section 4. We consider depths $L = 1,\ldots,6$ of the DVG. Parameter estimates are contained in Table 1; we note that values of α_L for small L tend to grow as higher depths are considered, but then stabilize.

Figure 5 shows loss values as a function of depth of the DVG subordinators, along with the chf loss for the NDIG model. All depths of the DVG are superior to the NDIG model, with a depth of $L = 5$ minimizing the chf loss for both cryptocurrencies. Histograms of the log-return values with the best-fitting DVG and NDIG models illustrate some interesting points: The NDIG model fails to capture the frequent small changes in price while the DVG adequately captures this spike in probability at small values. The DVG model also captures the heavy tails of the high-frequency returns better. This is apparent in Figure 6, which shows probability integral transform (PIT) histograms for each currency and model fit. An appropriate statistical model will have nearly uniform PIT histograms; we see the DVG showing good uniformity for both cryptocurrencies,

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.580 by University Of Colorado Librari, Wiley Online Library on [28/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

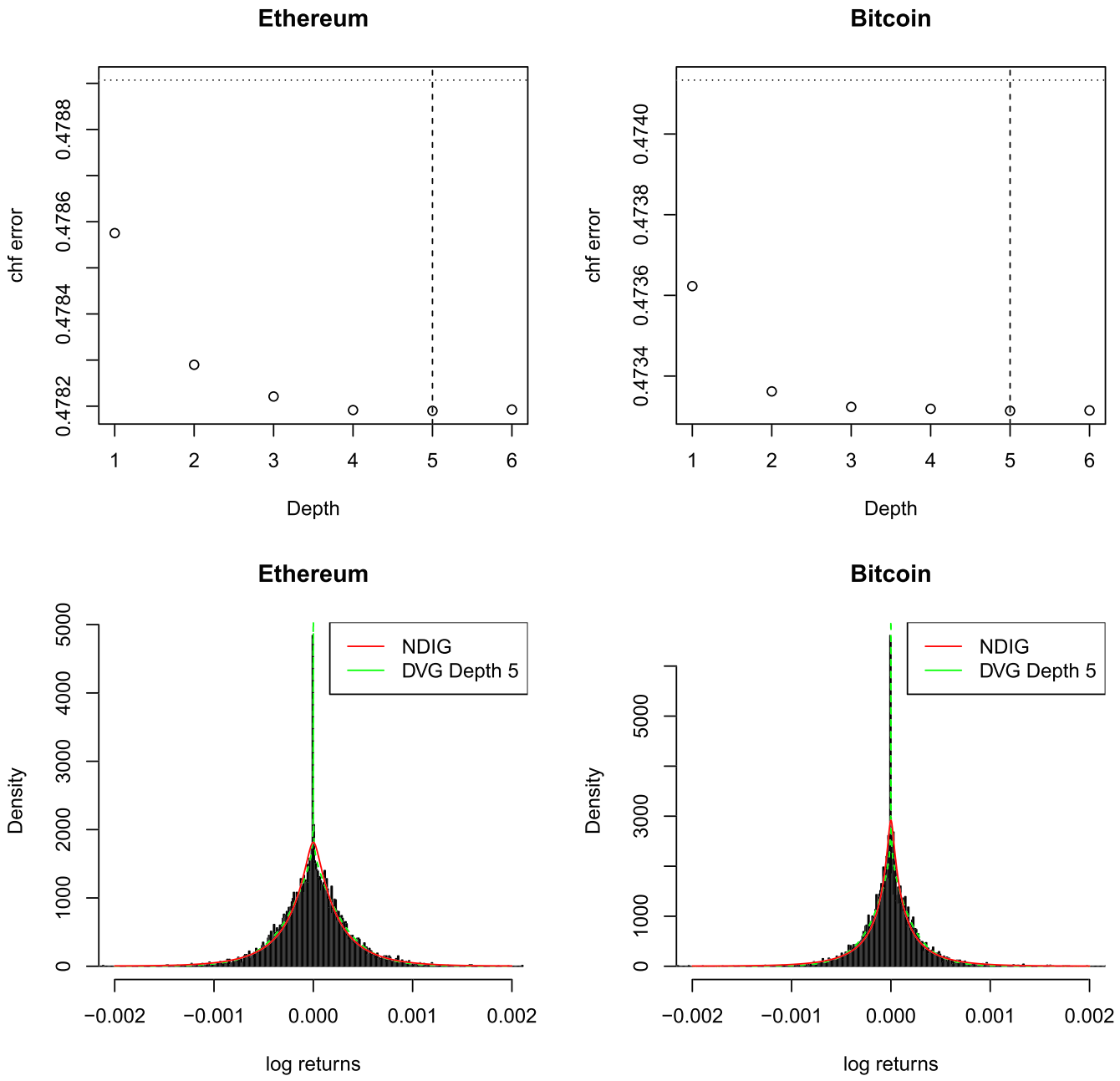

FIGURE 5 Loss values as a function of depth of the deep variance gamma process fits for depths $L = 1,\ldots,6$; the dashed line indicates the best-fitting depth ($L = 5$ in both cases), while the horizontal line indicates the loss value for the NDIG model. Histograms of log returns with estimated NDIG and proposed best-fitting DVG overlaid. The DVG model captures small returns as well as heavy tails better than NDIG.

while the NDIG exhibits underdispersion and strong modes on both sides of the PIT, suggesting it has tails that are too heavy for these data. Note this is in contrast to the results of Shirvani et al. (2022)—however, these authors use daily Bitcoin returns, while we focus on intraday high-frequency one minute behavior.

6 | DISCUSSION

We discuss a deep variance gamma (DVG) Lévy process in which a Brownian motion is subordinated multiple times with independent gamma processes. We investigate theoretical properties of the model, and analytically show that, under certain conditions, as the depth increases without bound, the Lévy process becomes degenerate in the limit. We perform a detailed simulation study and find that, even with moderately-sized data

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.580 by University Of Colorado Librari, Wiley Online Library on [28/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Parameter estimates for different DVG model depths for Ethereum and Bitcoin high-frequency return data.

	Ethereum						Bitcoin					
Depth	1	2	3	4	5	6	1	2	3	4	5	6
σ	3.9e−4	3.7e−4	3.6e−4	3.5e−4	3.5e−4	3.5e−4	3.1e−4	2.8e−4	2.7e−04	2.7e−04	2.7e−4	2.7e−4
α_1	0.8	4.8	8.0	16.5	17.1	8.9	0.6	4.5	11.2	15.8	20.8	13.2
α_2	-	1.5	3.8	9.2	10.3	9.8	-	1.2	3.7	3.6	12.3	93.5
α_3	-	-	2.7	3.8	5.8	3.4	-	-	1.7	3.3	63.4	9.8
α_4	-	-	-	3.5	2.6	37.7	-	-	-	4.4	2.6	3.9
α_5	-	-	-	-	228.8	11.5	-	-	-	-	2.9	2.8
α_6	-	-	-	-	-	11.9	-	-	-	-	-	23.3

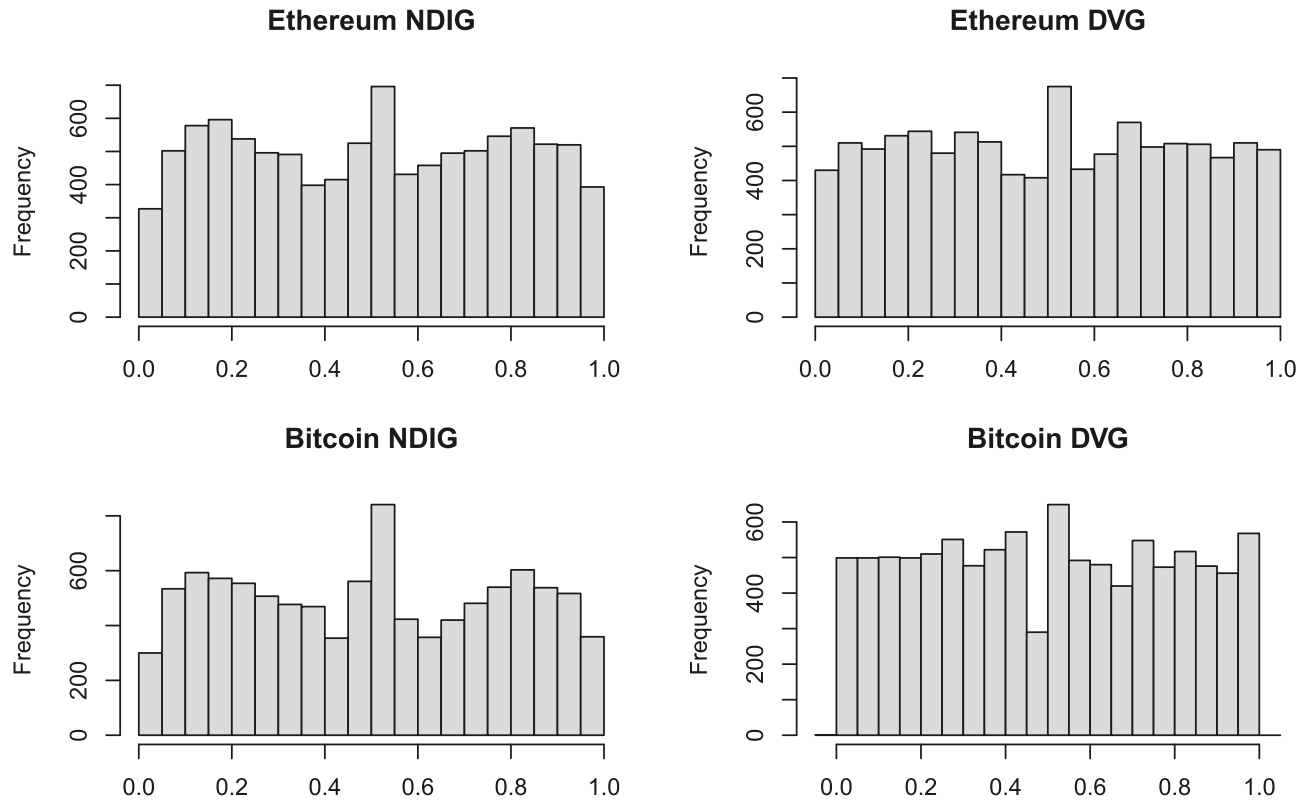

FIGURE 6 PIT histograms for the Ethereum and Bitcoin examples with the normal double inverse Gaussian (NDIG) model and the proposed deep variance gamma (DVG) model.

sets, parameters of the DVG can be well estimated, and appropriate depths can be identified given data. Our findings show that distinguishing the optimal depth is generally most difficult for large gamma parameter values, and we propose a diagnostic approach for choosing an appropriate depth given data. We illustrate the approach on two high-frequency cryptocurrency data sets using Ethereum and Bitcoin one minute returns. We show that a rather deep DVG fits both data sets better than a shallow DVG, or a competing model from the recent literature. The proposed model captures both the tails and the strong peak around zero better than extant models. Future research may be directed toward investigating other deep Lévy processes, some of which have been introduced in Shirvani et al. (2021), as well as developing theoretical criteria for choosing a particular model depth given data.

ACKNOWLEDGEMENTS

This research was supported by NSF DMS-1923062.

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.580 by University Of Colorado Librari, Wiley Online Library on [28/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

William Kleiber https://orcid.org/0000-0003-0411-9108

REFERENCES

Aguilar, J. P., Kirkby, J. L., & Korbel, J. (2020). Pricing, risk and volatility in subordinated market models. *Risks*, *8*, 124.

Barndorff-Nielsen, O. E., & Blaesild, P. (1981). Hyperbolic distributions and ramifications: Contributions to theory and application. In Taillie, C., Patil, G., & Baldessari, B. (Eds.), *Statistical distributions in scientific work* (pp. 19–44). Dordrecht: Reidel.

Barndorff-Nielsen, O. E., & Shepard, N. (2006). Impact of jumps on returns and realised variances: Econometric analysis of time-deformed Lévy processes. *Journal of Econometrics*, *131*, 217–252.

Carr, P., & Wu, L. (2004). Time-changed Lévy processes and option pricing. *Journal of Financial Economics*, *71*, 113–141.

Celeste, V., Corbet, S., & Gurdgiev, C. (2020). Fractal dynamics and wavelet analysis: Deep volatility and return properties of Bitcoin, Ethereum and Ripple. *The Quarterly Review of Economics and Finance*, *76*, 310–324.

Chu, J., Chan, S., & Zhang, Y. (2020). High frequency momentum trading with cryptocurrencies. *Research in International Business and Finance*, *52*, 101176.

Cocco, L., Concas, G., & Marchesi, M. (2017). Using an artificial financial market for studying a cryptocurrency market. *Journal of Economic Interaction and Coordination*, *12*, 345–365.

Cont, R., & Tankov, P. (2004). *Financial modelling with jump processes*. Chapman & Hall/CRC.

Elgin, H. (2011). Lévy processes and parameter estimation by maximum empirical likelihood. Unpublished manuscript.

Figueroa-López, J. E. (2009). Nonparametric estimation for Lévy models based on discrete-sampling, *Optimality: The Third Erich L. Lehmann Symposium*, Lecture Notes – Monograph Series (pp. 117–146). Beachwood, OH: IMS.

Figueroa-López, J. E. (2011). Sieve-based confidence intervals and bands for Lévy densities. *Bernoulli*, *17*, 643–670.

Honore, P. (1998). Pitfalls in estimating jump-diffusion models. Available at SSRN 61998.

Kappus, J., & Reiß, M. (2011). Estimation of the characteristics of a Lévy process observed at arbitrary frequency. (*SFB 649 Discussion Paper, No. 2011-027*): Humboldt University of Berlin, Collaborative Research Center 649 - Economic Risk, Berlin.

Klein, C. (2022). 400+ crypto currency pairs at 1-minute resolution. https://www.kaggle.com/datasets/tencars/392-crypto-currency-pairs-at-minute-resolution. Accessed: 2022-12-01.

Madan, D. B., Carr, P. P., & Chang, E. C. (1998). The variance gamma process and option pricing. *European Finance Review*, *2*, 79–105.

Madan, D. B., & Seneta, E. (1990). The variance gamma model for share market returns. *Journal of Business*, *63*, 511–524.

Petukhina, A. A., Reule, R. C. G., & Härdle, W. F. (2021). Rise of the machines? Intraday high-frequency trading patterns of cryptocurrencies. *The European Journal of Finance*, *27*, 8–30.

Philippas, D., Rjiba, H., Guesmi, K., & Goutte, S. (2019). Media attention and Bitcoin prices. *Finance Research Letters*, *30*, 37–43.

Sato, K. (1999). *Lévy processes and Infinitely Divisible Distributions*: Cambridge University Press.

Sato, K. (2001). Basic results on Lévy processes. In Barndorff-Nielsen, O. E., Resnick, S. I., & Mikosch, T. (Eds.), *Lévy processes: Theory and applications* (pp. 3–37). Birkhäuser.

Shirvani, A., Mittnik, S., Lindquist, W. B., & Rachev, S. T. (2022). Bitcoin volatility and intrinsic time using double subordinated Lévy processes. arXiv: 2109.15051v1.

Shirvani, A., Rachev, S. T., & Fabozzi, F. J. (2021). Multiple subordinated modeling of asset returns Implications for option pricing. *Econometric Reviews*, *40*, 290–319.

Sueishi, N., & Nishiyama, Y. (2005). Estimation of Lévy processes in mathematical finance: A comparative study. In *MODSIM 2005 International Congress on Modelling and Simulation* (pp. 953–959).

Xu, K., & Darve, E. (2020). Calibrating multivariate Lévy processes with neural networks. *Proceedings of Machine Learning Research*, *107*, 207–220.

Yu, J. (2007). Empirical characteristic function estimation and its applications. *Econometric Reviews*, *23*, 93–123.

Zhang, Y., Chan, S., Chu, J., & Nadarajah, S. (2019). Stylised facts for high frequency cryptocurrency data. *Physica A*, *513*, 598–612.

How to cite this article: Berry, C. M., & Kleiber, W. (2023). Deep variance gamma processes. *Stat*, *12*(1), e580. https://doi.org/10.1002/sta4.580

APPENDIX A

This appendix contains proofs the main theorems.

Proof of Theorem 2. We show this by induction. Let $Z(t)$ be a Lévy Process on $\mathbb{R}$ with characteristic exponent $\Psi(\omega)$, and $S_1(t), S_2(t), \ldots, S_L(t)$ be subordinators with Laplace exponents of $\ell_1, \ell_2, \ldots, \ell_L$, respectively. Finally, denote $\mathcal{F}^S$ to be the filtration of $S(t)$. The base case for $L = 1$ is Proposition 1, a proof of which can be found in Cont and Tankov (2004). Now assume the characteristic function for the subordinated process $Y(t) = Z(S_1(S_2(\ldots S_L(t))))$ is

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.580 by University Of Colorado Librari, Wiley Online Library on [28/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\mathbb{E}\exp(i\omega Y(t)) = \exp(t\ell_L(-\ell_{L-1}(...(-\ell_1(-\Psi(\omega)))))) = \exp(t\Psi_Y(\omega))$$

where we define $\Psi_Y(\omega) = \ell_L(-\ell_{L-1}(...(-\ell_1(-\Psi(\omega)))))$. Consider $V(t) = Y(S_{L+1}(t))$, where S_{L+1} is a subordinator with Laplace exponent ℓ_{L+1}. Then we have

$$\begin{aligned}
\mathbb{E}[e^{iuV(t)}] &= \mathbb{E}[e^{iuY(S_{L+1}(t))}] \\
&= \mathbb{E}\left[\mathbb{E}[e^{iuY(S_{L+1}(t))}|\mathcal{F}^{S_{L+1}}]\right] \\
&= \mathbb{E}[e^{\Psi_Y(\omega)S_{L+1}(t)}] \\
&= \mathbb{E}[e^{-(-\Psi_Y(\omega))S_{L+1}(t)}] \\
&= e^{t\ell_{L+1}(-\Psi_Y(\omega))} \\
&= e^{t\ell_{L+1}(-\ell_L(...(-\ell_1(-\Psi(\omega)))))}
\end{aligned}$$

where we use the induction step in the third equality, and form of the Laplace transform of a subordinated process in the fifth equality. ∎

Proof of Theorem 3. The characteristic exponent of a deep variance gamma process of order L is

$$-\alpha_L \log\left(1 + \frac{\alpha_{L-1}}{\alpha_L}\log\left(1 + ...\frac{\alpha_1}{\alpha_2}A(\omega)\right)\right)$$

where $A(\omega) = \log\left(1 + \frac{\sigma^2}{2\alpha_1}\omega^2\right)$. Define

$$\gamma_L = \log\left(1 + \frac{\alpha_{L-1}}{\alpha_L}\log\left(1 + ...\frac{\alpha_1}{\alpha_2}A(\omega)\right)\right)$$

where we suppress dependence of γ_L on ω for ease of notation. Note γ_L forms a decreasing sequence because

$$\gamma_{L+1} = \log\left(1 + \frac{\alpha_L}{\alpha_{L+1}}\gamma_L\right) < \frac{\alpha_L}{\alpha_{L+1}}\gamma_L < \gamma_L$$

by the standard inequality for logarithms that $\log(1+x) < x$ for any positive x and that $\{\alpha_L\}_L$ is an increasing sequence. Additionally, note that $\{\gamma_L\}_L$ is bounded below by zero, thus it converges.

Note that $\{\alpha_L\}_L$ converges as it forms a monotonically increasing sequence that is bounded above by Λ, let α be this limit, and M be the limit of $\{\gamma_L\}_L$. Then at the limit of $\{\gamma_L\}$ we have $M = \log(1 + \frac{a}{\alpha}M)$ which implies $M = 0$. Thus, for any choice of ω, we have the characteristic exponent in (5) converges to zero as $L \to \infty$, and the chf converges to unity. ∎

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.580 by University Of Colorado Librari, Wiley Online Library on [28/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License