COMMUNICATIONS IN STATISTICS - SIMULATION AND COMPUTATION®
2026, VOL. 55, NO. 4, 1146–1166
https://doi.org/10.1080/03610918.2024.2427225

Computational aspects of likelihood-based inference for the univariate generalized hyperbolic distribution

Arnold van Wyk[a], Adelchi Azzalini[b], and Andriette Bekker[a]

[a]Department of Statistics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria, South Africa; [b]Dipartimento di Scienze Statistiche, Università di Padova, Padova PD, Italy

ABSTRACT
The generalized hyperbolic distribution is among the more often adopted parametric families in a wide range of application areas, thanks to its high flexibility as the parameters vary and also to a plausible stochastic mechanism for its genesis. This high flexibility comes at some cost, however, namely the frequent difficulty of estimating its parameters due to the presence of flat areas of the log-likelihood function, so that selected points of the parameter space, while very distant, can be essentially equivalent as for data fitting. This phenomenon affects not only maximum likelihood estimation, but Bayesian methods too, since the target function is little affected by the introduction of a prior distribution. Our interest focuses in fact on maximum likelihood estimation of the Generalized hyperbolic distribution, working in the univariate case. This paper improves upon currently employed computational techniques by presenting an alternative proposal that works effectively in reaching the global maximum of the likelihood function. The paper further illustrates the above mentioned problems in a number of cases, using both simulated and real data.

ARTICLE HISTORY
Received 4 April 2023
Accepted 4 November 2024

KEYWORDS
EM algorithm; Flexible parametric distributions; Generalized hyperbolic distributions; Maximum likelihood estimation; Nelder-Mead simplex method; Profile likelihood

1. Introduction

The generalized hyperbolic (GH) distribution was first developed by (Barndorff-Nielsen 1977) to model the mass-size distribution of sand particles, with further developments in (Barndorff-Nielsen 1978; Barndorff-Nielsen and Blæsild 1981). This model emanated from a geostatistical study, and results from a normal mean-variance mixture, where the mixture variable emanates from a generalized inverse Gaussian distribution. One of the primary appeals of the class of generalized hyperbolic distributions lies in the semi-heavy tail property possessed by some of its subclasses (notably the hyperbolic and normal inverse Gaussian subclasses), and is one of the reasons for the GH distribution's popularity in modeling financial markets (Eberlein and Keller 1995; Puig and Stephens 2001). Some other applications of the GH distribution include the modeling of turbulent velocity increments (Barndorff-Nielsen, BlæSild, and Schmiegel 2004), the modeling of wavelet image coefficients (Bhuiyan, Omair Ahmad, and Swamy 2007), and in the analysis of the financial risk of cryptocurrencies (Zhang et al. 2019).

There are, however, some issues that present when fitting this distribution to data. There is a recurrent issue of a flat log-likelihood shape near the maximum, and this is exacerbated when the

CONTACT Arnold van Wyk arcvanwyk@gmail.com Department of Statistics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria 0002, South Africa.

parameters of the full GH distribution[1] are estimated simultaneously. The possible occurrence of a non-quadratic shape near the maximum of the log-likelihood invalidates asymptotic theory which generates standard errors, and as such a large flat region near the maximum can be problematic. Besides this flatness of the log-likelihood function, there is also the issue of multiple maxima when estimating the parameters of the GH distribution. The form of the probability density function (pdf) also restricts us to numerical methods, specifically derivative free methods, to estimate the parameters. This, combined with a flat log-likelihood function and the possible existence of multiple maxima, is a definite cause for concern when working with this distribution.

Although the method of maximum likelihood is a strong means of parameter estimation, it is not without its shortfalls. The susceptibility of the standard minimization algorithms to boundaries of the parameter constraints (see (3)), as well as the above-mentioned flatness of the likelihood function, are examples of this (Qiao and Tsokos 1995; Prause 1997; Sundberg 2010). This problem is amplified when dealing with flexible distribution classes (Ley 2015). This creates a situation where two vastly different parameter estimates can result in the same fitted distribution of the data, although the model is formally identifiable. This situation is often denoted as 'near non-identifiability' following the terminology of (Monahan 1983) and a number of other authors. Essentially the same concept is denoted "near parameter redundancy" in the monograph of (Cole 2020) and the connected literature.

There have also been various instances in the literature whereby the GH distribution is fitted to data and results are discussed, but with no discussion of the existence of multiple local maxima and other problematic aspects of the likelihood function of the GH distribution. These papers have been examined for this issue and no reports on computational complication have been mentioned concerning the potentially problematic shape of the log-likelihood function. Examples in the literature include fitting the GH distribution to: daily returns of the PS120 (Rege and Menezes 2012), equity returns (Konlack Socgnia and Wilcox 2014), chinese stock prices (Li and Wu 2007), the Bucharest stock exchange (Baciu 2015), cryptocurrencies (Zhang et al. 2019), commodity futures (Pal 2023), and gold price returns (Chinhamu, Huang, and Chikobvu 2015).

As such, exploration of these problems is a necessity, especially when the underlying distribution possesses properties that may hinder the overall quality of the estimation process, such as the the lack of a closed form derivative or a badly behaved likelihood function. There is clear evidence in literature of potential issues when estimating the parameters of the GH distribution by means of the method of maximum likelihood, but we have not found a detailed numerical investigation of this issue. This exploration represents a key goal of the present paper.

The above issues are not only relevant in this context, but also extend to other domains. Bayesian methods are an example of this. When a lot of data is available, the posterior distribution is largely effected by the likelihood, unless the prior distribution is extremely sharp, which is rarely the case. This puts a lot of emphasis on the role of the log-likelihood function as far as Bayesian methods are concerned, as a poorly-behaving log-likelihood function will have a direct impact on the construction of the posterior distribution. This necessitates the investigation of the behavior of the log-likelihood of the GH distribution in this context as well.

What sets the generalized hyperbolic distribution apart from the hyperbolic distribution, and the flexible distribution framework as a whole, is the introduction of the index parameter λ, giving us five parameters in total, which are denoted GH(λ, α, β, δ, μ) in one of the commonly adopted parameterizations (Barndorff-Nielsen 1978). This results in this distribution being a superclass of numerous flexible distributions, often referred to as subfamilies, which include but are not limited to: the variance-gamma distribution, the Laplace distribution, the Student's t distribution, and naturally, the hyperbolic distribution (Paolella 2007, 317–26). This adds an extra layer of flexibility, and it is of interest to examine whether this flexibility has an impact on the

[1]Full GH distribution refers to the full five parameter model, where λ is included in the estimation process and not fixed.

behavior of the log-likelihood function, and possibly on the quality of the maximum likelihood estimation process as a whole.

Challenges resulting from a flat likelihood function, which is believed to be caused in large part by the index parameter λ, are briefly noted in (Barndorff-Nielsen and Blæsild 1981; Prause 1999), and (Snoussi and Idier 2006). As previously stated, the λ parameter is largely responsible for the existence of the various subclasses, begging the question if there is possible near over-parametrization, or if the nature of flexible distributions has an inherent, negative impact on the maximum likelihood estimation method. The above quoted papers also report a possible near non-identifiability issue, with reference to a specific case where the normal inverse Gaussian subfamily ($\lambda = -0.5$) and the hyperbolic subfamily ($\lambda = 1$) are nearly identical. This is potentially alarming as it could greatly decrease the validity, as well as the inferential power of the resulting estimates.

What is apparent from the literature is that, although there is some discussion, albeit brief, on the impacts of the above issues, there is not a sufficient, in-depth analysis on whether or not the steps taken in any way ensure that the quality of the estimates are of an acceptable standard. Furthermore, the existence of multiple local maxima when fitting the GH distribution does not seem to have been examined in the literature. This opens up the opportunity for further exploration into these behaviors and potential issues and forms the basis of the motivation for this paper.

Consequently, a key aim of this paper is to explore the behavior of the log-likelihood function of the generalized hyperbolic distribution, as well as its impact on the use of the method of maximum likelihood to estimate the GH parameters. It is also of interest to investigate potential solutions to problems stemming from this particular log-likelihood function. The intermediate goal is to undertake a detailed exploratory analysis on the potential problems that may be present when using the maximum likelihood as a means of estimating parameters, especially with flexible distributions. The aim of this paper is thus two-fold. Firstly, we want to explore the behavior of the log-likelihood function and examine the recurrent presence of problematic aspects of the GH log-likelihood. Secondly, we put forward a maximization algorithm, with the aim of illustrating the impact of the log-likelihood function and creating a platform for future work.

2. The generalized hyperbolic distribution

The GH distribution is considered a highly flexible distribution. By convention, this term refers to distributions that allow substantial variation of their behavior when the underlying parameters span their admissible range. Arguably, the most classical instance of a flexible class of distributions are represented by the Pearson system of curves, whereby the pdf is regulated by four parameters, thus allowing for greater variation in terms of measures of skewness and of kurtosis. This naturally provides a greater flexibility, than for example the normal distribution, where only location and scale can be varied.

2.1. Theoretical aspects

In the one-dimensional case the generalized hyperbolic distribution has the following pdf

$$\begin{aligned} f_X(x;\lambda,\alpha,\beta,\delta,\mu) = {} & a(\lambda,\alpha,\beta,\delta)e^{\beta(x-\mu)} \\ & \times K_{\lambda-\frac{1}{2}}\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right) \\ & \times \left(\delta^2+(x-\mu)^2\right)^{\frac{1}{2}\left(\lambda-\frac{1}{2}\right)} \qquad (x \in \mathbb{R}) \end{aligned} \tag{1}$$

Where $\lambda, \alpha, \beta, \delta$ and μ are the parameters, the expression $K(\cdot)$ denotes the modified Bessel function of the third kind (see Abramowitz and Stegun [1972]), and $a(\lambda,\alpha,\beta,\delta)$ is a norming constant given by

$$a(\lambda,\alpha,\beta,\delta)=\frac{(\alpha^2-\beta^2)^{\frac{\lambda}{2}}}{\sqrt{2\pi}\ \alpha^{\lambda-\frac{1}{2}}\ \delta^{\lambda}\ K_{\lambda}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}. \tag{2}$$

The domain of variation of the parameter space is given by

$$\begin{array}{llll} \alpha>0, & |\beta|<\alpha, & \delta\geq 0, & \text{if}\quad \lambda>0,\\ \alpha>0, & |\beta|<\alpha, & \delta>0, & \text{if}\quad \lambda=0,\\ \alpha\geq 0, & |\beta|\leq\alpha, & \delta>0, & \text{if}\quad \lambda<0 \end{array} \tag{3}$$

Where $\mu\in\mathbb{R}$. A more in-depth description of the parameters is now given:

λ : Commonly seen as the index parameter as it gives rise to many distinctions amongst the subfamilies of the generalized hyperbolic distribution. It also influences the shape of the pdf.
α : The tail parameter that regulates the 'fatness' of the tails. The larger the value of α, the lighter the tails of the distribution.
β : The skewness parameter, with $|\beta|<\alpha$. An in increase in β compared to α will result in an increase in the skewness. For $\beta=0$ the distribution is symmetric.
δ : Influences the shape of the pdf near its mode, and as such is often referred to as the 'peakedness' parameter. Larger values of δ will result in an overall flatter peak of the pdf.
μ : The location parameter.

Table 1 contains a list of the relevant subclasses of the GH distribution as well as a breakdown of the parameters and how they relate to the full GH distribution. For typical examples of the GH density for various parameter values see Figure 1; additional examples are provided in (Paolella 2007, 319).

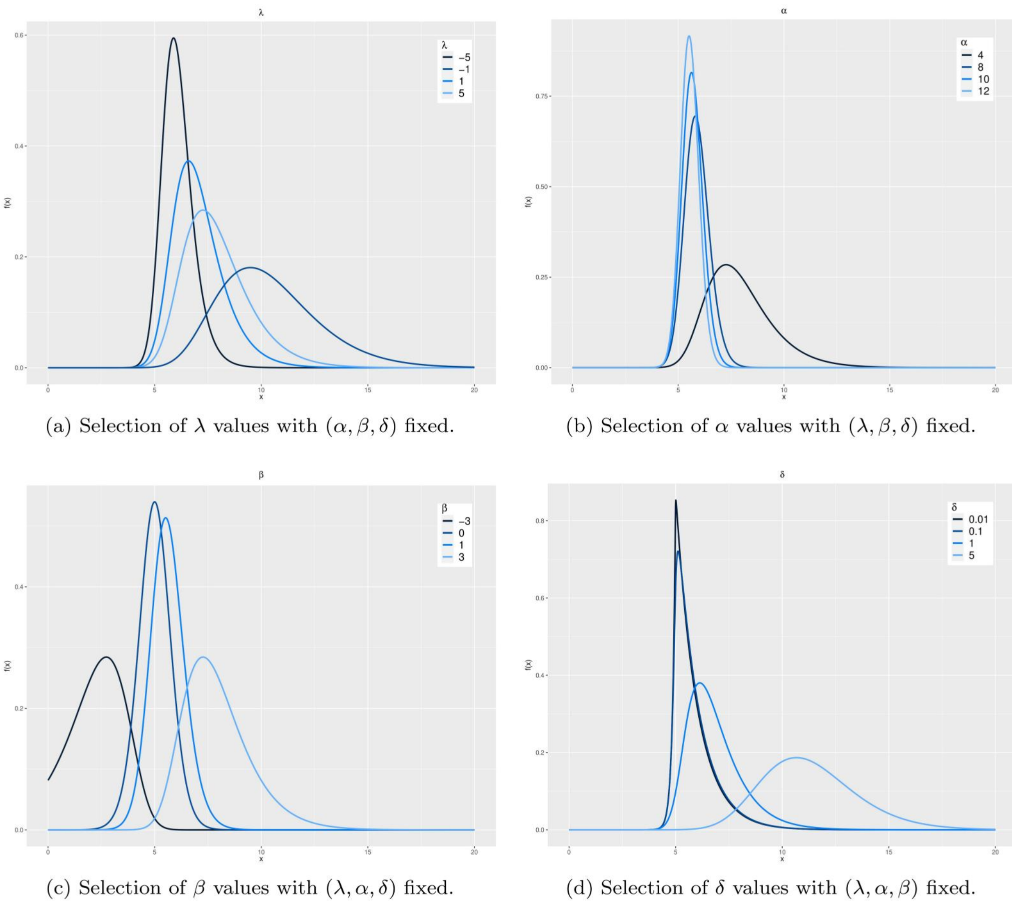
2.2. Inferential tools and algorithms

For the process of fitting the GH distribution, *via* maximization of the log-likelihood function to data, the following numerical algorithms are considered:

1. Nelder-Mead simplex method;
2. EM algorithm;
3. Proposed method: alternating profile-likelihood-based algorithm (APLB).

Table 1. Subfamilies of the GH distribution and corresponding parameter spaces, with location parameter $\mu\in\mathbb{R}$ in each case.

Distribution	Parameter space			
Variance-gamma	$\lambda>0$	$\alpha>0$	$\|\beta\|<\alpha$	$\delta=0$
Asymmetric Laplace	$\lambda=1$	$\alpha>0$	$\|\beta\|<\alpha$	$\delta=0$
Laplace	$\lambda=1$	$\alpha>0$	$\beta=0$	$\delta=0$
Hyperbolic	$\lambda=1$	$\alpha>0$	$\|\beta\|<\alpha$	$\delta>0$
Hyperbolic asymmetric t	$\lambda<0$	$\alpha=\|\beta\|$	$\beta\geq 0$	$\delta>0$
Student's t	$\lambda<0$	$\alpha=0$	$\beta=0$	$\delta>0$
Asymmetric Cauchy	$\lambda=-\frac{1}{2}$	$\alpha=\|\beta\|$	$\beta\in\mathbb{R}$	$\delta>0$
Cauchy	$\lambda=-\frac{1}{2}$	$\alpha=0$	$\beta=0$	$\delta>0$
Normal inverse Gaussian	$\lambda=-\frac{1}{2}$	$\alpha>0$	$\|\beta\|<\alpha$	$\delta>0$



(a) Selection of λ values with (α, β, δ) fixed.

(b) Selection of α values with (λ, β, δ) fixed.

(c) Selection of β values with $(\lambda, \alpha, \delta)$ fixed.

(d) Selection of δ values with (λ, α, β) fixed.

Figure 1. Model representation of GH density for different values of λ, α, β, and δ.

The Nelder-Mead simplex method is a widely used numerical method and is especially useful in that it does not require the calculation or use of derivatives of the target function, that is, the (negative) log-likelihood function. This is very useful when working with the GH distribution as there does not exist a closed form for the derivative of the pdf. It is also one of the method options within the `optim` function, a commonly used function for numerical optimization in `R` (R Core Team 2022) (the computational environment of choice for the present work). The use of this method will also serve as a good baseline for assessing the performance of the ML estimation of the GH parameters in the methods that follow. This is also the methodology followed by some popular `R` packages for the GH distribution, namely the ghyp (Weibel, Luethi, and Breymann 2020), GeneralizedHyperbolic (Scott 2018), and HyperbolicDist (Scott 2023) packages.

An extension of the `optim` function, namely the `constrOptim` function, will also be considered. This function allows for specification of linear inequality constraints while using the Nelder-Mead simplex method to minimize the objective function. This allows us to consider $\alpha > 0$, $\delta > 0$, and especially the bounded relationship $|\beta| < \alpha$.

The EM algorithm is one of the most commonly used numerical methods in the literature used in the estimation of the GH parameters. In (Karlis 2002; Hu 2005; Aas and Haff 2006; Hellmich and Kassberger 2011; McNeil, Frey, and Embrechts 2015; Panahi 2018), and (Prause 1999) we see the EM algorithm used to fit the GH model or one of its subclasses to data.

There exists another algorithm, namely the Differential Evolution algorithm developed by (Mullen et al. 2011). It stems from the idea of differential evolution developed by (Storn and

Price 1997). This evolution strategy finds the global optimum of a real-valued function without requiring the function to be continuous or differentiable. Fortunately, this algorithm was coded into an R package called `DEoptim`. While at first glance this does seem like an attractive solution to estimating the parameters of the GH distribution, there are unfortunately some drawbacks that prevent it from being a reliable solution. In particular, the triangular region of α and β due to the restrictions in (3), coupled with the fact that the bounds of each parameter estimate are not easily known in advance, can lead to a failure to converge in some instances. This is as a result of the `DEoptim` algorithm requiring the specification of an interval for each parameter as opposed to a set of initial values.

As such, the proposed method, namely the APLB method, will be considered. This method aims to exploit the idea of profiling the log-likelihood function in order to create better behaved functions for the minimization algorithms, and thus potentially increase the likelihood of successful convergence to a global maximum. For more on the profile likelihood see (Murphy and Van der Vaart 2000). A detailed description of the method is provided in Sec. 3.1.

2.3. Problematic aspects of the log-likelihood

2.3.1. General points

Multiple issues emerged during the process of estimation, as well as during the general exploration of the behavior of the log-likelihood function of the GH distribution. A flat log-likelihood function seems to be a recurrent feature of the generalized hyperbolic distribution. This is exacerbated when the index parameter λ is included in the estimation process (as opposed to being fixed at the outset).

This leads to optimization methods either stopping or getting stuck before reaching the global maximum, or progressing in the wrong search direction altogether. This issue is not limited to the numerical methods used here, but is a consequence of numerical methods in general when a flat function region is present. This stems from the fact that, whether the method makes use of the function gradient or simply the function values, the steepness of the functions region, especially around a global or local maximum, will have an impact on the methods ability to correctly converge.

In our numerical work, we make use of the 'NYSE composite index' dataset as in (Prause 1999). In Figure 2, we observe the resulting estimates for α and β when a grid of starting values are considered (using the Nelder-Mead simplex method). The red box indicates the bounds of the grid in terms of α and β, and comprises 3125 initial points for these parameters. Figure 3 provides a zoomed in view of Figure 2. While this can be observed in both images, Figure 3 provides a clear example of a large linear region with negligible difference in log-likelihood value. For some context, Table 2 contains a simple breakdown of the distribution of the estimates in terms of final log-likelihood value. From this table, we see that roughly 7% of the initial value choices resulted in estimates with log-likelihood values greater than 6406 (points in red). The significance of the choice of 6406 simply comes from analyzing the resulting estimates and deciding that this was a good cutoff point both visually and in terms of the log-likelihood value (see Figure 3). Given that the initial values were all contained within the red grid in Figure 3, this is not indicative of an ideal estimation outcome.

2.3.2. Effect of initial values

The choice of the initial value also had a large impact on the successful convergence to a maximum. Choosing an initial value too close to the boundary of the parameter space often results in the search algorithm getting stuck on this boundary, unable to progress toward the global maximum. An example of this is choosing values of α_0 and β_0 that are close in magnitude. This problem is due in part to the aforementioned flat likelihood function behavior. It is worth noting that

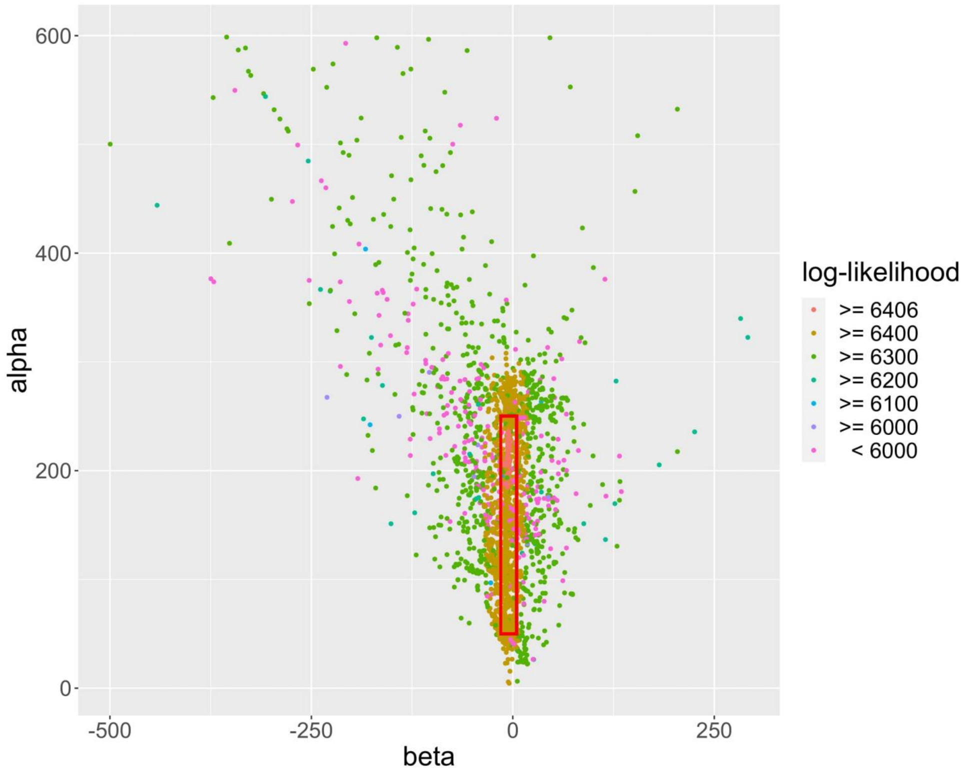


Figure 2. NYSE composite index: Scatterplot of $(\hat{\alpha}, \hat{\beta})$ estimates for GH fit using the Nelder-Mead simplex method across a grid of starting values.

while choosing initial values close to the parameter bounds decreased the chance of successful convergence, there were instances where initial values regressed to the boundary even when they were not close to it. It therefore seems the shape and behavior of the log-likelihood function is responsible for these occurrences. An example of this behavior can be seen in Figure 4, where the above process is applied to year end prices of Standard and Poor's most notable stock market price index, the S&P 500. This data is taken from (Brown, Spears, and Levy 2002), and can also be found in the *GeneralizedHyperbolic* package in the R software environment.

It was also found that when using the full log-likelihood function, even seemingly sensible initial value choices could lead to incorrect, or even non-convergence. The sample size also plays an important role in not only the generation of adequate initial values, but on the rate of convergence to the maximum. Section 3.3 contains a simulation study highlighting the issue of stability of the GH parameter estimates across numerous sample sizes.

Another issue worth noting is that including λ in the estimation process has a clear impact on the quality and outcome of the procedure, and as such, a typical solution has been to fix λ to a value corresponding to one of the subclasses of the GH distribution. The problem with this then lies in selecting an appropriate subclass, and while there are some characteristics inherent to each subclass, it is largely a process of trial and error. While it is sometimes possible to try every possible subclass and see which behaves best, the idea of running through large sets of initial values as well as subclass choices can be rather laborious, and is often not a feasible solution, especially when faced with rather large datasets.

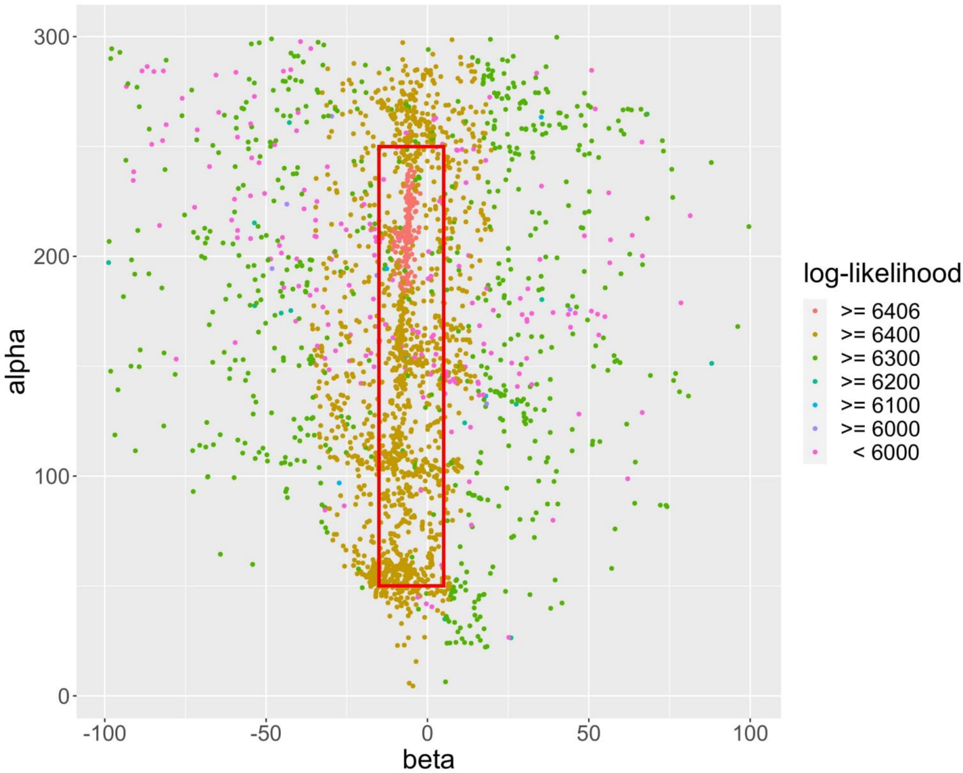


Figure 3. NYSE composite index: Scatterplot of $(\hat{\alpha}, \hat{\beta})$ estimates for GH fit using the Nelder-Mead simplex method across a grid of starting values.

Table 2. NYSE composite index: the number of estimates falling in each log-likelihood bracket corresponding to the GH distribution as in Figure 2.

Log-likelihood Value	Number of Estimates
$\geq$**6406**	240
[6400, 6406)	1611
[6300, 6400)	835
[6200, 6300)	34
[6100, 6200)	7
[6000, 6100)	9
$<$ 6000	258

2.3.3. Multiple local maxima

While the parameterization specified in Eq. (2.1) is typically employed, we will transition to the (ψ, χ) parameterization in the subsequent example. This reparameterization is implemented by setting $\chi = \delta^2$ and $\psi = \alpha^2 - \beta^2$, with pdf given by:

$$\begin{aligned} f_X(x; \lambda, \psi, \beta, \chi, \mu) = {} & a(\lambda, \psi, \beta, \chi) e^{\beta(x-\mu)} \\ & \times K_{\lambda - \frac{1}{2}}\left(\sqrt{(\chi + (x-\mu)^2)(\psi + \beta^2)}\right) \\ & \times \left((\chi + (x-\mu)^2)(\psi + \beta^2)\right)^{\frac{1}{2}\left(\lambda - \frac{1}{2}\right)} \quad (x \in \mathbb{R}) \end{aligned} \tag{4}$$

For more information on this parameterization see (Barndorff-Nielsen 1977). In Figure 5 we observe a contour plot of the log-likelihood in terms of ψ and χ. It is evident here that the log-likelihood is influenced solely by the ratio of ψ and χ as the log-likelihood is essentially constant. This elongation essentially indicates redundancy. Thus, there is a distinct indication of near non-

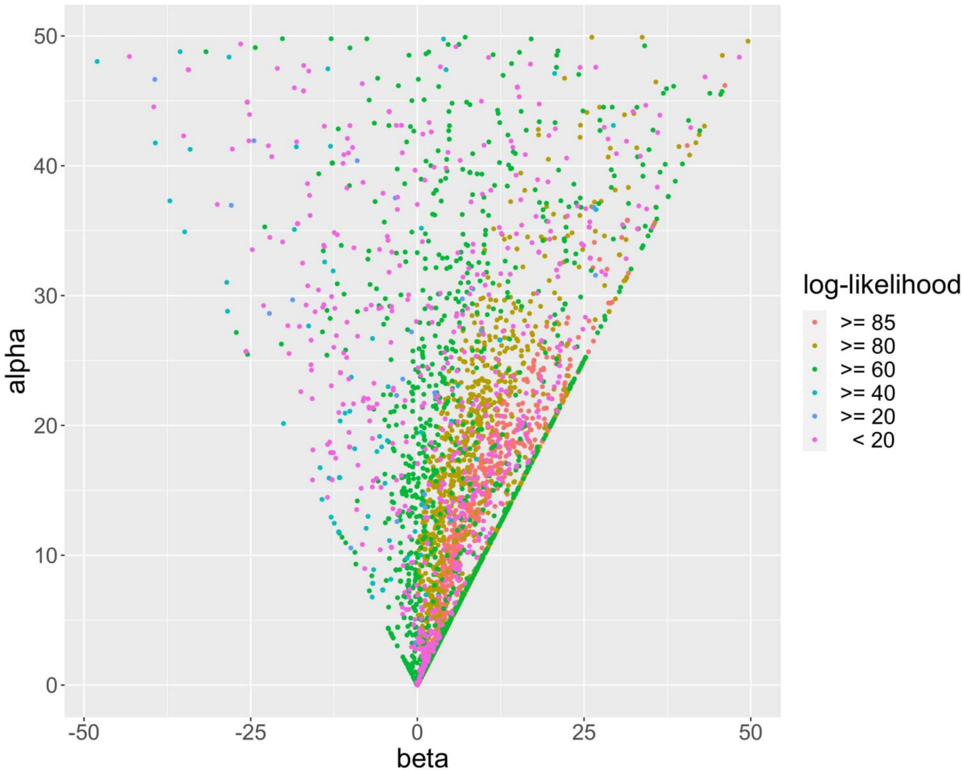


Figure 4. S&P 500 index: Scatterplot of $(\hat{\alpha}, \hat{\beta})$ estimates using the Nelder-Mead simplex method across a grid of starting values when the full GH distribution is fitted.

identifiability issues for certain data sets, confirming the assertion made in the introduction to near over-parameterization.

Another crucial issue is the possible presence of multiple local maxima of the log-likelihood, a problem which does not seem to have been considered in the literature. To examine this, we revert back to the (α, β) parameterization in (2.1). Figure 6 provides a display of the contour plot in terms of the log-likelihood for a set of $n = 1000$ values simulated from a GH(−0.5,100,−50,1,0) distribution. In Figure 7 the respective local maxima are displayed for visual clarity. In Figure 6 we observe two distinct local maxima, with locations $(\beta, \delta) = (-106.05, 0.2753)$ and $(\beta, \delta) = (-28.81, 1.508)$, with log-likelihood values of 667.94 and 718.96 respectively. This is a clear indication that these sort of issues do present when fitting the GH distribution to data. What makes this even more concerning is that the data was simulated using size $n = 1000$. One would normally expect such issues for much smaller sample sizes, but the fact that this persists for this size emphasizes the concerns raised about the shape of the log-likelihood function. In this instance, when the moments estimates are used as the initial values (a common choice in literature) the algorithm converges to the set of estimates $(\lambda, \alpha, \beta, \delta, \mu) = (-8.39, 113.905, -28.81, 1.508, -0.2021)$ with the higher log-likelihood value of 718.96.

This is also not a unique occurrence. In Figures 8 and 9 we observe the same behavior as before. The data in Figures 8 and 9 comes from the daily log returns of the American express stock (AXP) from the Dow Jones index from the years 1993–2000. This results in a dataset of size $n = 2020$ and corresponds to the dataset used in (McNeil, Frey, and Embrechts 2015, 84). For some visual clarity, Figure 10 provides a replication of Figure 9(b) with the δ parameter on the log scale.

What makes this occurrence in a known dataset so relevant is not only the practical nature of the issue, but the sample size. These issues are expected to be more prevalent in smaller samples,

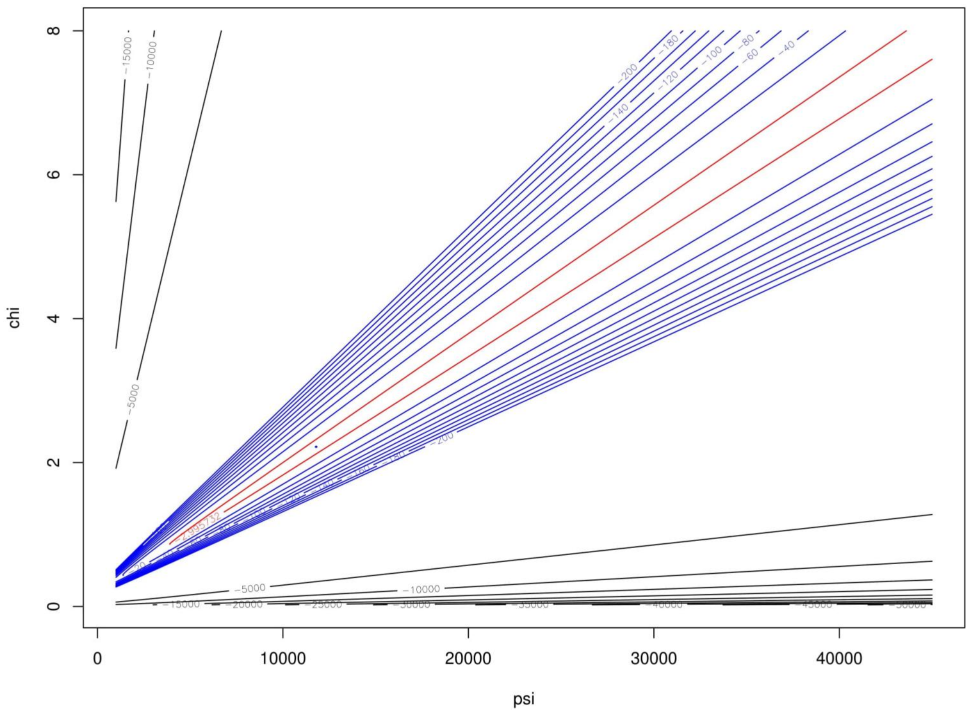


Figure 5. Contour plot in terms of ψ and χ from the simulated GH(−0.5,100, β,δ,0) data when the (ψ,χ) parameterization is considered.

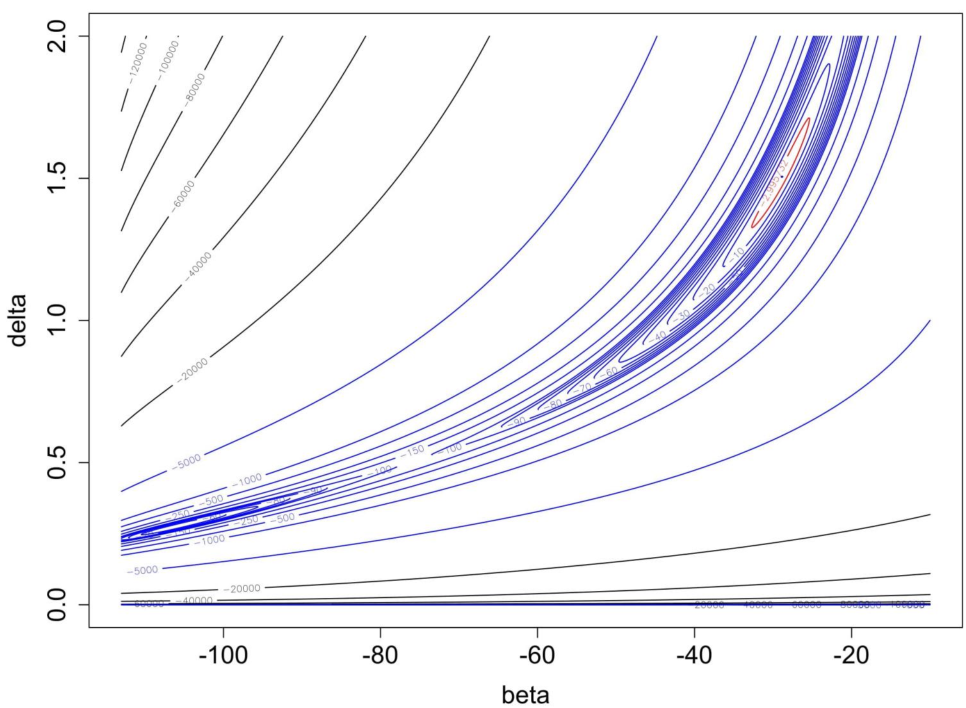
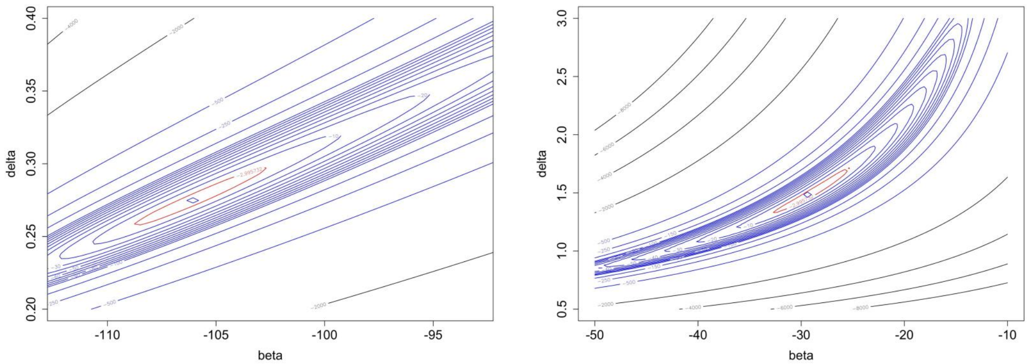


Figure 6. Contour plot in terms of β and δ from GH(−0.5,100, β,δ,0). The above plot refers to a sample of size $n = 1000$ from GH(−0.5,100,−50,1,0).



(a) (b)

Figure 7. Contour plot of (a) the local maximum corresponding to the bottom left of Figure 6, and (b) the local maximum corresponding to the top right of Figure 6. The above plot refers to a sample of size $n = 1000$ from GH(−0.5,100,−50,1,0).

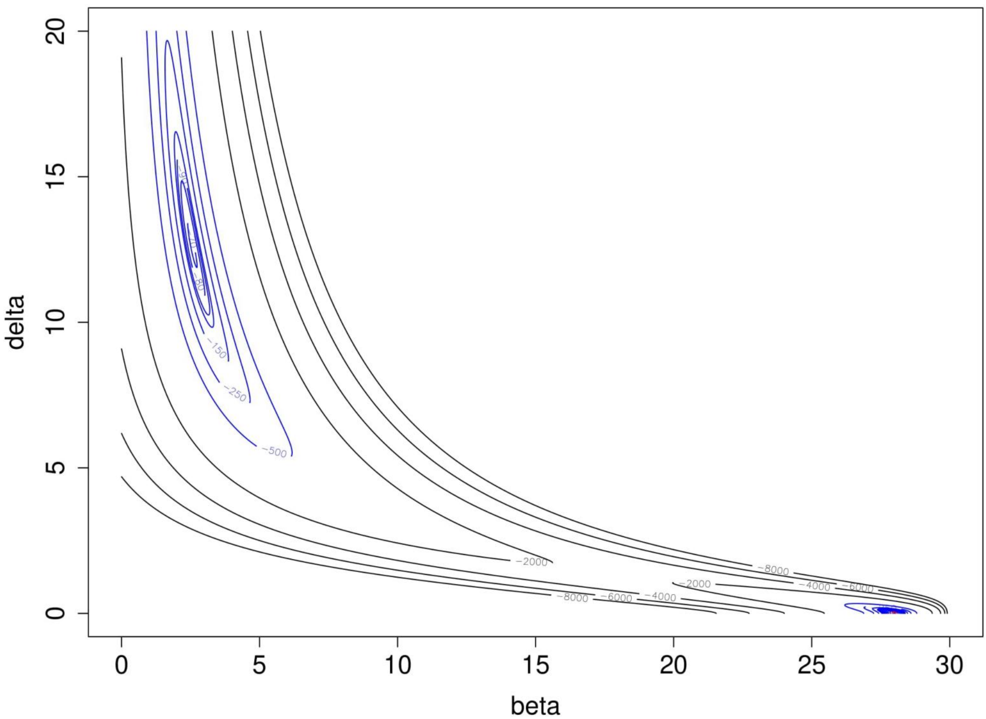


Figure 8. AXP daily returns data: Contour plot in terms of β and δ.

with the expectation that larger samples will result in consistent unimodal log-likelihood functions. In this example we observe two local maxima at $(\beta, \delta) = (-2.56, 16.29)$ and $(\beta, \delta) = (35.96, 0.000598)$, with log-likelihood values of 2021.52 and 1951.81 respectively. What is interesting in this case, is that when the moments estimates are used as the initial values (as before), the algorithm converges to the set of estimates $(\lambda, \alpha, \beta, \delta, \mu) = (2.31, 37.99, 35.96, 0.000598, 1.769)$ with the lower log-likelihood value of 1951.81. This occurrence is a clear cause for concern and is indicative of a broader issue when estimating parameters in this manner.

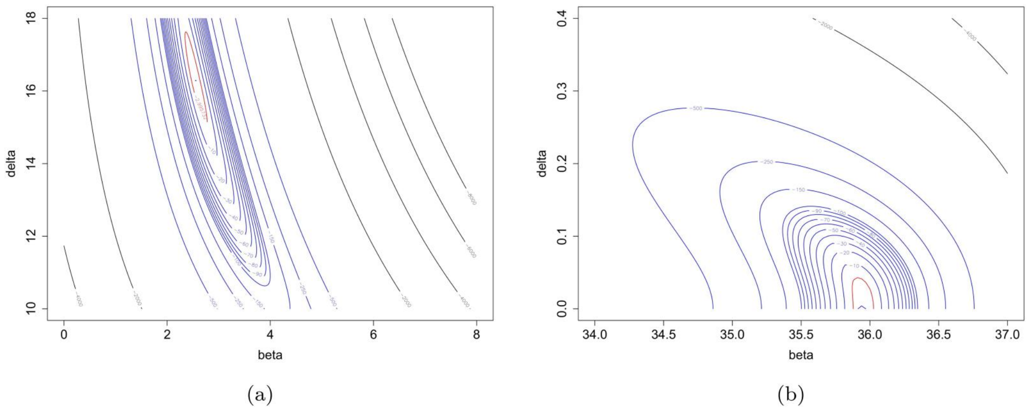


Figure 9. AXP daily returns data: Contour plot of (a) the local maximum corresponding to the top left of Figure 8, and (b) the local maximum corresponding to the bottom right of Figure 8.

3. A proposal for log-likelihood maximization

Here we outline a proposed method for estimating the GH parameters, with the primary aim of alleviating the difficulties in maximization of the log-likelihood function. This kind of "alternating" maximization procedure has been widely applied for parameter estimation in (Jain, Netrapalli, and Sanghavi 2013; Netrapalli, Jain, and Sanghavi 2013; Yi, Caramanis, and Sanghavi 2014), and a detailed description of the methodology, as well as its convergence properties can be found in (Andresen and Spokoiny 2016). It is important to note that, since the root of the problem is in the flatness of the log-likelihood function, it cannot be eliminated by the choice of numerical optimization technique.

3.1. Alternating profile-likelihood-based algorithm

In much the same way we use the profile log-likelihood to allow for interpretable visual results, we can use it to split the estimation process up into two parts instead of simultaneously estimating all the parameters. By analyzing the marginal (profile) log-likelihood functions and splitting the parameter set in two, this creates a desirable situation whereby the functions to be maximized are better behaved than the full log-likelihood. This is due to the lower dimensional space in which the function is minimized which in turn allows for better performance of the numerical methods.

The formulation that follows is initially for the subclasses of the GH distribution that have fixed λ. In a similar fashion to the profile likelihood, the parameter vector $\boldsymbol{\theta} = (\alpha, \beta, \delta, \mu)$ is split into $\boldsymbol{\xi} = (\alpha, \beta)$, and $\boldsymbol{\eta} = (\delta, \mu)$. This choice of split was decided based on trial and error, and it was found that this pairing led to convergence of the algorithm. It is of course possible that other pairings could very well lead to the same outcome, but this pairing seemed to have the fastest and most consistent convergence. The first log-likelihood function will be maximized (using any available derivative free method) on $\boldsymbol{\xi} = (\alpha, \beta)$, with $\boldsymbol{\eta} = (\delta, \mu)$ fixed, and the second will be maximized on $\boldsymbol{\eta} = (\delta, \mu)$, with $\boldsymbol{\xi} = (\alpha, \beta)$ fixed. The equations to be maximized are thus given by

$$L_{p_1}(\boldsymbol{\xi}) = \sup_{\boldsymbol{\eta}} L(\boldsymbol{\xi}, \boldsymbol{\eta}) \tag{5}$$

and

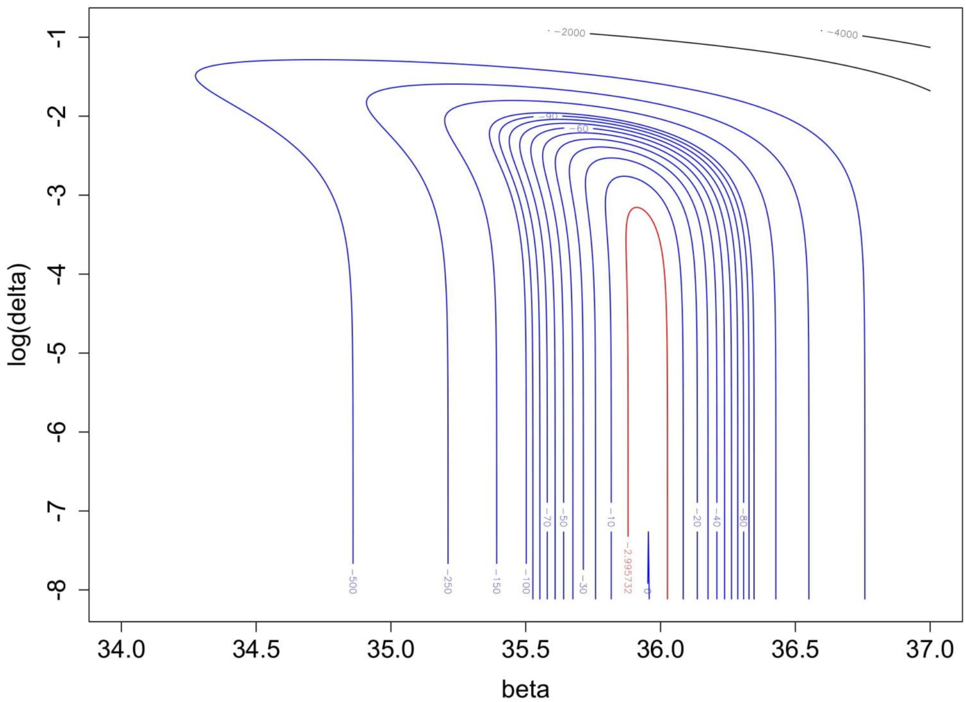


Figure 10. AXP daily returns data: Contour plot of (b) the local maximum corresponding to the bottom right of Figure 8 (in terms of β and $\log(\delta)$).

$$L_{p_2}(\boldsymbol{\eta}) = \sup_{\xi} L(\boldsymbol{\xi}, \boldsymbol{\eta}). \tag{6}$$

A step-by-step on the implementation of this method is given as follows:

Step 1: Set iteration number $k = 0$ and $\epsilon > 0$ (ϵ is typically chosen to be between 10^{-3} and 10^{-6}).
Step 2: Choose initial values $\boldsymbol{\theta_k} = (\alpha_k, \beta_k, \delta_k, \mu_k)$.
Step 3: Maximize $L_{p_2}(\boldsymbol{\eta})$ w.r.t. $\boldsymbol{\eta} = (\delta, \mu)$, yielding parameters μ_{k+1} and δ_{k+1}.
Step 4: Set $\mu_k = \mu_{k+1}$ and $\delta_k = \delta_{k+1}$.
Step 5: Maximize $L_{p_1}(\boldsymbol{\xi})$ w.r.t. $\boldsymbol{\xi} = (\alpha, \beta)$, yielding parameters α_{k+1} and β_{k+1}.
Step 6: Set $\alpha_k = \alpha_{k+1}$ and $\beta_k = \beta_{k+1}$.
Step 7: If $\frac{\left|\theta_i^{(k+1)} - \theta_i^{(k)}\right|}{1 + \left|\theta_i^{(k+1)}\right|} \geq \epsilon$ for any i, set $k = k + 1$ and return to step 2.
Step 8: Otherwise terminate the process.

The stopping criterion in Step 7 of the Algorithm is chosen as it is a good tradeoff between an absolute and relative stopping criterion. For values close to zero, as is often the case with the δ parameter, this criterion will behave like an absolute error, and for values further away from zero it will behave like that of a relative error. This accounts for differing parameter scales and is found to be more consistent. This is by no means the only applicable stop criterion to use. Both parameter based and log-likelihood based stop criterions have been tested, and both result in satisfactory convergence of the algorithm.

It was found that the addition of λ into the estimation algorithm is not only possible, but yielded results corresponding to those of the other methods used. For clarity, λ was introduced into the estimation process by setting $\boldsymbol{\xi} = (\lambda, \alpha, \beta)$, $\boldsymbol{\eta} = (\delta, \mu)$, and simply proceeding as before.

This method is of course not limited to the GH distribution alone, as the core logic can be applied to any distribution with similar parameter type and/or structure. When considering other distributions for this method, it is advised to consider different parameter groupings to set up equations (5) and (6).

There seems to be no clear way to decide on how to split ξ and $\boldsymbol{\eta}$. It is therefore advised that the reader defines ξ and $\boldsymbol{\eta}$ as sensibly as possible.

3.2. Fitting the GH distribution to a real dataset

Since we are working with the generalized hyperbolic distribution, it is natural to make use of any packages that accommodate this distribution and make the process easier. There are three notable packages that deal with the generalized hyperbolic distribution, namely:

The following are some popular/existing packages in R that can be used for fitting the GH distribution to data:

1. The *ghyp* package,
2. The *GeneralizedHyperbolic* package,
3. The *HyperbolicDist* package.

While these packages offer an array of useful tools in the context of the GH distribution, the estimation tools offered make use of the Nelder-Mead simplex method within the `optim()` function, and the evidence provided in Sec. 2.3 indicates that this method does not lead to reliable convergence to a maximum.

In terms of the methods considered for comparison, we focus on the Nelder-Mead simplex method by virtue of its use in most, if not all, of the existing packages or functions for fitting the GH distribution to univariate data. The EM algorithm is also considered and implemented using (McNeil, Frey, and Embrechts 2015) and (Hu 2005). While the EM algorithm does have similiar performance to the existing Nelder-Mead simplex method, it has the added burden of computation time and an increased sensitity to initial value selection. On this point, consider Table 3, comparing the performance of the Nelder-Mead simplex method, EM algorithm, and APLB method when the hyperbolic subclass is fitted to the NYSE dataset. For this comparison each method is fitted across the same grid of starting values, and the resulting log-likelihood values are recorded. From this it is clear that the outcomes are similar, albeit with an increased incidence of failed convergence with the EM algorithm compared to the rest.

In Table 4 we observe the result of fitting the models to the NYSE Composite Index data (Prause 1999) using the APLB method. For both the hyperbolic and normal inverse Gaussian subclasses all the estimation methods performed rather well, in that they were all able to converge to their respective maximum points. In Table 5 we see a comparison of the different methods

Table 3. Comparing estimation results from fitting the hyperbolic subclass using the Nelder-Mead, EM, and APLB methods: NYSE composite index.

	Number of estimates		
Log-likelihood Value	Nelder-Mead	EM	APLB
≥ 6408	117	109	625
[6400, 6408)	202	0	0
[6300, 6400)	172	5	0
[6200, 6300)	35	6	0
[6000, 6200)	44	0	0
[1000, 6000)	55	185	0
≤ 1000	0	45	0
Failed to converge	0	275	0

Table 4. Estimates produced by the APLB method for the GH distribution and relevant subclasses: NYSE composite index.

	λ	α	β	δ	μ	log-L
Hyperbolic	1	225.03	−5.84	0.0016	0.00064	6408.27
Normal inverse Gaussian	−0.5	135.84	−8.80	0.0059	0.00078	6406.74
Generalized hyperbolic	0.81	212.56	−5.93	0.0022	0.00066	6408.31
Hyperbolic asymmetric t	−1.96	9.91	−9.91	0.0094	0.00084	6402.09

Table 5. Comparison of Nelder-Mead and the APLB method for the GH distribution: NYSE composite index.

	λ	α	β	δ	μ	log-L
Nelder-Mead	0.67	203.49	−7.09	0.0027	0.00070	6408.26
Nelder-Mead (constrained)	0.84	215.08	−6.72	0.0022	0.00068	6408.30
APLB method	0.81	212.56	−5.93	0.0022	0.00066	6408.31

Table 6. Computing times for the various estimation methods for the GH distribution and a subset of subclasses: NYSE composite index.

		Time (s)		
	Method	MIN	MAX	AVG
Hyperbolic	Nelder-Mead	0.3531	0.7380	0.4398
	APLB method	1.8381	3.3919	2.8489
Normal inverse Gaussian	Nelder-Mead	0.1995	0.3022	0.2227
	APLB method	3.6183	7.9235	3.9536
Generalized hyperbolic	Nelder-Mead	0.2277	0.2512	0.23417
	APLB method	68.61	69.57	68.76

used to fit the GH distribution to data. The term 'constrained' refers to the fact that the bounds on the parameters are included and accounted for in the estimation process, more specifically that $|\beta| < \alpha$ and $\delta \geq 0$. From Table 5 it is clear that the GH distribution is hard to fit, as we have clearly distinct parameter combinations which are essentially equivalent in terms of log-likelihood. This is clearly a consequence of the flatness of the log-likelihood function, further amplified by the inclusion of λ in the estimation process. The hyberbolic asymmetric t fit had some stability issues when various initial values were tested, but when the moment estimates were used as initial values the process seems to converge to the required maximum.

An observation when fitting the GH model or one of the relevant subclasses, is that it is not always correct to assume a subclass will lead to a good fit. For some of the fitted subclasses, the rate of occurence of stability issues and inadequate convergence was higher. There were also instances of non-convergence, and sometimes even divergence. Another aspect that must be carefully considered is the proximity of the initial values for α and β to the boundary of the constraint $|\beta| < \alpha$. It is found that when α_0 and β_0 are close in magnitude, the estimation algorithm can get 'stuck' on the line that governs the constraint between these parameters, and consequently fail to converge to the global maximum point. Another scenario that can occur is the global maximum point itself being close to this boundary. This is often an indication that the current distribution being fitted is not the most suitable and another subclass should be fitted.

The computation times when using the various estimation methods can be found in Table 6. In each instance the process of estimation is repeated 100 times, and the resulting metrics are shown. It is clear that the APLB method takes an increased amount of time to estimate the parameters in each case, especially in the case when the full GH distribution is fitted to the data. While this is generally not an ideal outcome, there are a few things that need to be considered.

Firstly, the APLB method is significantly more robust against the choice of initial values. This claim is supported by Figure 11, comparing the standard Nelder-Mead algorithm with the APLB method using the same grid of initial values. For this process, both the Nelder-Mead simplex method and the APLB method are applied to the NYSE composite index dataset across the same grid of initial values, and the resulting estimates are displayed. It is clear here that the APLB

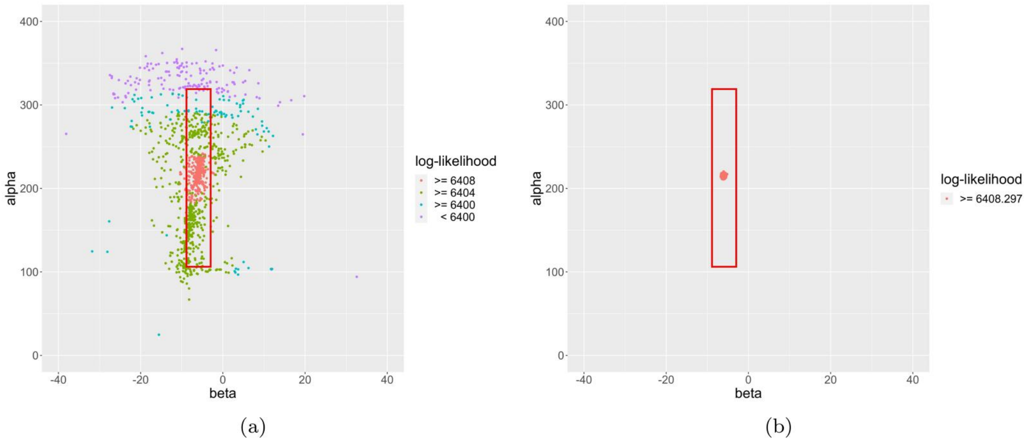


Figure 11. NYSE composite index: Scatterplot of $(\hat{\alpha}, \hat{\beta})$ estimates for GH fit using (a) the Nelder-Mead simplex method, and (b) the APLB method across a grid of starting values.

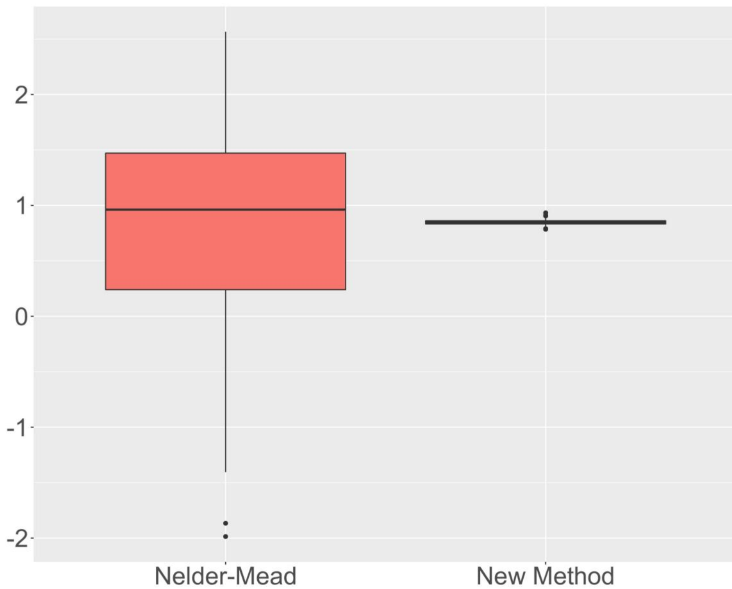
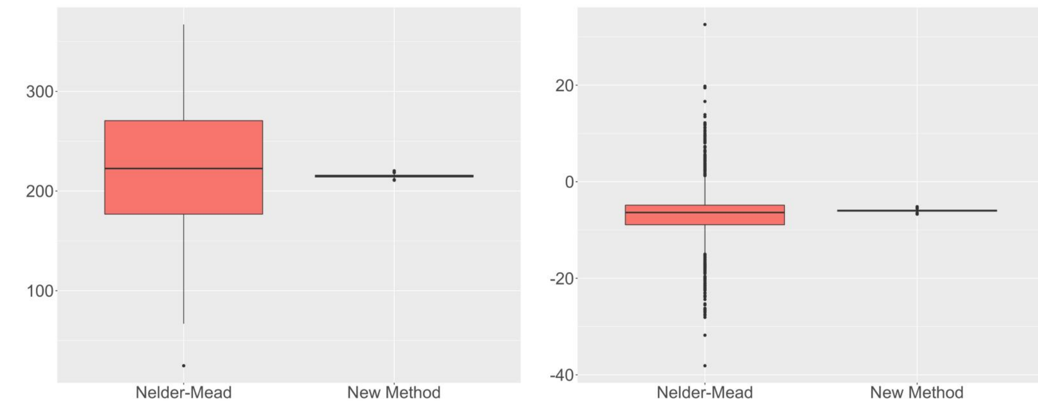


Figure 12. NYSE composite index: Boxplot of resulting λ estimates comparing the Nelder-Mead simplex method and the APLB method for a selection of starting values as in Figure 11.

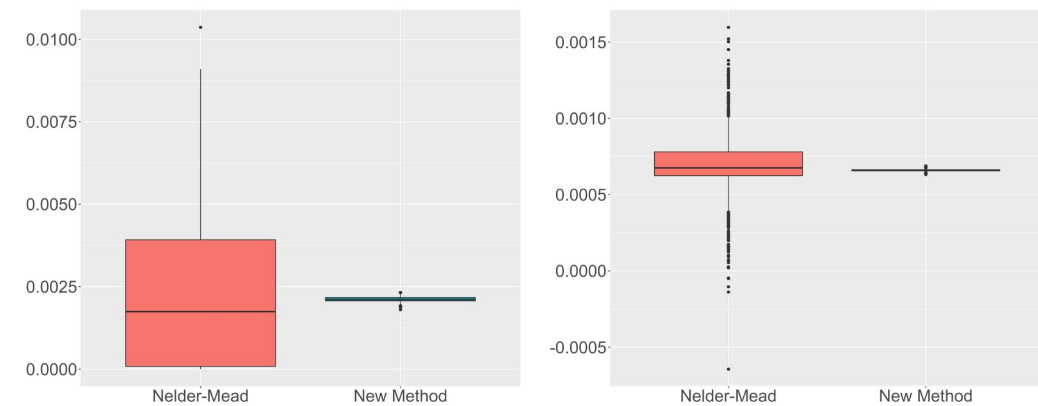
method (Figure 11b) far outperforms standard algorithms (Figure 11a). As an example, when looking at Figures 11, the APLB method (Figure 11b) had consistent convergence to the maximum for the given grid of initial values, whereas the standard Nelder-Mead algorithm had results as in Figure 11a. To expand on this, if the minimum log-likelihood value result of the APLB method is taken as a threshold point, the probability of the APLB method converging when the Nelder-Mead simplex method does not is 0.91083 (91%).

Secondly, the objective functions used in the APLB method have reduced dimension compared to the full log-likelihood. This results in both and increased stability and an increased reliability when estimating the full GH distribution. In practical terms, this translates to less initial value choices causing the algorithm to break down (stability), and less initial value choices failing to adequately reach the maximum (reliability). While the APLB method does take significantly



(a) (b)

Figure 13. NYSE composite index: Boxplot of resulting (a) α, and (b) β estimates comparing the Nelder-Mead simplex method and the APLB method for a selection of starting values as in Figure 11.



(a) (b)

Figure 14. NYSE composite index: Boxplot of resulting (a) δ, and (b) μ estimates comparing the Nelder-Mead simplex method and the APLB method for a selection of starting values as in Figure 11.

longer than the existing numerical methods, these are strong arguments for the use of this method, and should be a clear justification for the increased run time of the algorithm.

Some descriptives are provided in Figures 12–14 to compare the spread of the resulting estimates between the Nelder-Mead simplex method and the APLB method. From this it is clear that there is stability in not only the α and β estimates (as seen in Figure 11b), but in the other parameters as well when the APLB method is considered. So while there is a definite increase in computation time, this increase in stability is a desirable aspect of the method, especially when stability is a known issue when λ is calibrated (see Sec. 3.3).

It should be noted that for the results in Figure 11 the initial values were selected in a uniform fashion around the maximum point. If this region is extended, perhaps to better emulate a scenario where we have no knowledge of a maximum, it was found that the outcome worsens when a standard method such as Nelder-Mead is considered.

Table 7. Mean of the estimates of the GH distribution at different sample sizes: NYSE composite index; 100 iterations.

Sample Size	Mean λ	Mean α	Mean β	Mean δ	Mean μ
50	−9.1541	288.57	−13.46	1.2841E-02	1.1638E-03
100	−0.1347	209.35	−14.36	4.0426E-03	8.0695E-04
250	0.1689	197.23	−8.39	3.5650E-03	7.3587E-04
500	0.5216	200.91	−6.64	2.6174E-03	6.3528E-04
750	0.6236	204.03	−5.80	2.4967E-03	6.5766E-04
1000	0.6417	204.77	−6.01	2.4506E-03	6.5761E-04
2500	0.7224	208.29	−5.42	2.2848E-03	6.3583E-04
5000	0.7600	209.77	−5.43	2.2211E-03	6.3612E-04
10,000	0.7580	209.46	−5.46	2.2866E-03	6.3788E-04

Table 8. Standard error of the estimates of the GH distribution at different sample sizes: NYSE composite index; 100 iterations.

Sample Size	Std error λ	Std error α	Std error β	Std error δ	Std error μ
50	2.5662	21.9518	21.5852	2.1830E-03	4.5120E-04
100	0.2725	11.7707	9.6529	5.8871E-04	2.3578E-04
250	0.2454	7.1192	2.5276	4.7760E-04	8.2383E-05
500	0.0893	4.8919	1.4841	2.8384E-04	4.4778E-05
750	0.0212	1.4321	0.3108	6.7595E-05	1.0925E-05
1000	0.0202	1.3397	0.2571	6.4772E-05	8.9660E-06
2500	0.0140	0.9279	0.1766	4.6941E-05	6.1566E-06
5000	0.0106	0.7107	0.1329	3.6574E-05	4.8699E-06
10,000	0.0086	0.5673	0.1083	2.8246E-05	3.8875E-06.

3.3. Simulation study

In order to get a better understanding of the stability of fitting the GH distribution to data, a simulation study is performed. In this study we simulate datasets of up to size 10,000 from various instances of the GH class, including the hyperbolic distribution. The parameter values in each case are chosen to be the maximum likelihood estimates for each of the subclasses fitted to the NYSE Composite Index data.

Using these estimates as the true parameter values of the simulated data allows for a more comprehensive comparison, as these estimates come from fitting notable subclasses of the GH distribution, as well as the full GH distribution to the same dataset. Simulating in this fashion also ensures that a meaningful set of parameters is chosen to simulate from in each case, and helps paint a picture as to the stability of the fitting the GH model and its subclasses to real world datasets.

A description of the process to generate the ensuing tables will now be given. For varying sample sizes, we generate a random sample using the chosen parameter values after which we fit the relevant GH model or subclass to this sampled dataset. This process is repeated 1000 times for each sample size, thus allowing us to compute the mean and standard errors of the parameters in each case.

In Tables 7 and 8 we observe the means and standard errors when the full GH model is fitted. While the standard errors seem to have increased in stability, the same can not be said for the stability of the λ parameter. The α and β parameters seem to stabilize quite well from sample size $n = 500$ onwards, but λ does not exhibit the same trend at all. It seems that only from $n = 2500$ onwards that we observe stability in terms of the calibration of λ. For smaller sample sizes the average estimate for λ is significantly different from the estimated value of 0.81 in Table 4. Not only this, but at the observed rate of change, it seems that the estimated value of 0.81 for λ will only be achieved for simulated sample sizes larger than 10,000.

This occurrence serves as a reinforcement of the same pattern of behaviors stemming from the estimation of the λ parameter, namely a negative impact not only on the shape and behavior of the log-likelihood function of the GH distribution, but on the estimation process itself and the quality of the resulting parameter estimates. This surely calls into question whether λ should be included in the estimation process, as this analysis has shown that it seems a much more sensible choice to fix λ and fit the relevant subclass/es to the data.

4. Conclusion

As can be seen in the exploratory analysis undertaken, there exist clear problems that need to be navigated when fitting the GH distribution to data by means of the log-likelihood function. This stems not only from the formulation of the GH distribution, but from the behavior of the ensuing log-likelihood function. There is clear evidence that the dimensionality of the parameter space causes behavioral problems in terms of the log-likelihood. This seems to stem not only from a potential near over-parameterization, but from the functional form of the GH distribution as well.

This is not to say that fitting the GH model to data is a fruitless pursuit. What we want to communicate is that a certain degree of care needs to be exercised when fitting the model. It was, for example, found that the inclusion of the λ parameter had a significant impact on not only the likelihood of convergence, but on the likelihood of converging to an adequate estimate. This is in large part due the the flatness of the log-likelihood function, especially that contributed by the λ parameter. There also seems to be some linearity in the relationship between the α and β parameters. This is a possible indication of a redundant parameterization, but can also be an indication of a practical identifiability issue stemming from the data (for a more detailed description of the different kinds of identifiability, please see [Lam, Docherty, and Murray 2022]).

The present contribution has examined problematic aspects of the GH likelihood function in greater detail than elsewhere in the literature. As a specific point, we have highlighted the possible presence of multiple local maxima; this issue has not been flagged before, as far as we are know. Also, we have proposed a technique to improve the GH maximum likelihood search. While this contribution does not claim to represent a definitive word on this theme, we believe that it provides at least a better understanding of the problem and motivates further work in this direction.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was based upon research supported by the South African National Research Foundation [SRUG2204203865 nr. 120839], STATOMET, University of Pretoria, as well as the Centre of Excellence in Mathematical and Statistical Sciences at the University of the Witwatersrand.

ORCID

Andriette Bekker http://orcid.org/0000-0003-4793-5674

Data availability statement

All the relevant datasets can be found at https://github.com/ArnoldvanWyk/likelihood_inference_GHYP

References

Aas, K., and I. H. Haff. 2006. The generalized hyperbolic skew student's t-distribution. *J. Financial Econ* 4 (2):275–309.

Abramowitz, M., and I. A. Stegun. 1972. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. New York: Dover.

Andresen, A., and V. Spokoiny. 2016. Convergence of an alternating maximization procedure. *JMLR* 17 (1):2229–81.

Baciu, O. 2015. Generalized hyperbolic distributions: Empirical evidence on Bucharest stock exchange. *BFR* 7 (1):7–18.

Barndorff-Nielsen, O. E. 1977. Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* 353:401–19.

Barndorff-Nielsen, O. E. 1978. Hyperbolic distributions and distributions on hyperbolae. *Scandinavian Journal of Statistics* 5:151–7.

Barndorff-Nielsen, O. E., and P. Blæsild. 1981. Hyperbolic distributions and ramifications: Contributions to theory and application. In *Statistical Distributions in Scientific Work*, eds. C. Taillie, G. Patil and B. Baldessari. Vol. 4., 19–44. Dordrecht: Reidel.

Barndorff-Nielsen, O. E., P. BlæSild, and J. Schmiegel. 2004. A parsimonious and universal description of turbulent velocity increments. *The European Physical Journal B* 41 (3):345–63. doi:10.1140/epjb/e2004-00328-1.

Bhuiyan, M. I. H., M. Omair Ahmad, and M. N. S. Swamy. 2007. On the modeling of the wavelet image coefficients using a symmetric generalized hyperbolic PDF. Proceedings of the 50th Midwest Symposium on Circuits and Systems, Montreal, Canada.

Brown, B. W., F. M. Spears, and L. B. Levy. 2002. The log F: A distribution for all seasons. *Computational Statistics*. 17 (1):47–58. doi:10.1007/s001800200098.

Chinhamu, K., C. Huang, and D. Chikobvu. 2015. Evaluating risk in gold prices with generalized hyperbolic and stable distributions. *Proceedings of the 57th Annual Conference of SASA* :17–24.

Cole, D. 2020. *Parameter redundancy and identifiability*. New York: CRC Press.

Eberlein, E., and U. Keller. 1995. Hyperbolic distributions in finance. *Bernoulli* 1 (3):281–99. doi:10.2307/3318481.

Hellmich, M., and S. Kassberger. 2011. Efficient and robust portfolio optimization in the multivariate generalized hyperbolic framework. *Quantitative Finance* 11 (10):1503–16. doi:10.1080/14697680903280483.

Hu, W. 2005. Calibration of multivariate generalized hyperbolic distributions using the EM algorithm, with applications in risk management, portfolio optimization and portfolio credit risk. Diss., Florida State University.

Jain, P., P. Netrapalli, and S. Sanghavi. 2013. Low-rank matrix completion using alternating minimization. *STOC* 45:665–74. doi:10.1145/2488608.2488693.

Karlis, D. 2002. An EM type algorithm for maximum likelihood estimation of the normal–inverse Gaussian distribution. *Statistics & Probability Letters* 57 (1):43–52. doi:10.1016/S0167-7152(02)00040-8.

Konlack Socgnia, V., and D. Wilcox. 2014. A comparison of generalized hyperbolic distribution models for equity returns. *Journal of Applied Mathematics* 2014 (1):1–15. doi:10.1155/2014/263465.

Lam, N., P. Docherty, and R. Murray. 2022. Practical identifiability of parametrised models: A review of benefits and limitations of various approaches. *Mathematical and Computers Simulation*. 199:202–16. doi:10.1016/j.matcom.2022.03.020.

Ley, C. 2015. Flexible modelling in statistics: Past, present and future. *Journal of the French Statistical Society* 156 (1):76–96.

Li, H., and L. Wu. 2007. The statistical analysis of the log-return series of the chinese stockprices: An application of the generalized hyperbolic distributions. *Journal of Data Analysis* 2 (2):161–78.

McNeil, A. J., R. Frey, and P. Embrechts. 2015. *Quantitative risk management: Concepts, techniques and tools*. Rev. ed. Princeton: Princeton University Press.

Monahan, J. F. 1983. Fully Bayesian analysis of ARMA time series models. *Journal of Econometrics* 21 (3):307–31. doi:10.1016/0304-4076(83)90048-9.

Mullen, K., D. Ardia, D. Gil, D. Windover, and J. Cline. 2011. DEoptim: An R package for global optimization by differential evolution. *Journal of Statistical Software* 40 (6):1–26. doi:10.18637/jss.v040.i06.

Murphy, S. A., and A. W. Van der Vaart. 2000. On profile likelihood. *Journal of the American Statistical Association* 95 (450):449–65. doi:10.1080/01621459.2000.10474219.

Netrapalli, P., P. Jain, and S. Sanghavi. 2013. Phase retrieval using alternating minimization. *Adv. Neural Inf. Process* 26:2796–804.

Pal, D. 2023. The distribution of commodity futures: A test of the generalized hyperbolic process. *Applied Economics* 56 (15):1763–83. doi:10.1080/00036846.2023.2177598.

Panahi, H. 2018. Discriminating between the normal inverse Gaussian and generalized hyperbolic skew-t distributions with a follow-up the stock exchange data. *Yugoslav Journal of Operations Research* 28 (2):185–99. doi:10.2298/YJOR170815013P.

Paolella, M. S. 2007. *Intermediate probability: A computational approach*, 299–337. Chichester: John Wiley & Sons.

Prause, K. 1997. Modelling financial data using generalized hyperbolic distributions. *FDM Preprint* 48.

Prause, K. 1999. The generalized hyperbolic model: Estimation, financial derivatives and risk measures. Diss., Albert Ludwig University.

Puig, P., and M. A. Stephens. 2001. Goodness-of-fit tests for the hyperbolic distribution. *Canadian Journal of Statistics* 29 (2):309–20. doi:10.2307/3316079.

Qiao, H., and C. P. Tsokos. 1995. Estimation of the three parameter Weibull probability distribution. *Mathematical and Computers Simulation*. 39 (1-2):173–85. doi:10.1016/0378-4754(95)95213-5.

R Core Team. 2022. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from: https://www.R-project.org/.

Rege, S., and A. Menezes. 2012. Comparing the generalised hyperbolic and the normal inverse Gaussian distributions for the daily returns of the PSI20. *CEEAplA* :1–14.

Scott, D. 2018. GeneralizedHyperbolic: The generalized hyperbolic distribution. Version 0.8-4 [software]. https://CRAN.R-project.org/package=GeneralizedHyperbolic.

Scott, D. 2023. HyperbolicDist: The hyperbolic distribution. Version 0.6-2 [software]. https://CRAN.R-project.org/package=HyperbolicDist.

Snoussi, H., and J. Idier. 2006. Bayesian blind separation of generalized hyperbolic processes in noisy and underdeterminate mixtures. *IEEE Transactions on Signal Processing* 54 (9):3257–69. doi:10.1109/TSP.2006.877660.

Storn, R., and K. Price. 1997. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *J Glob Optim* 11:341–59.

Sundberg, R. 2010. Flat and multimodal likelihoods and model lack of fit in curved exponential families. *Scandinavian Journal of Statistics* 37 (4):632–43. doi:10.1111/j.1467-9469.2010.00703.x.

Weibel, M., D. Luethi, and W. Breymann. 2020. ghyp: Generalized hyperbolic distribution and its special cases. Version 1.6.1 [software]. https://CRAN.R-project.org/package=ghyp.

Yi, X., C. Caramanis, and S. Sanghavi. 2014. Alternating minimization for mixed linear regression. *Proceedings of the 31st International Conference on Machine Learning, in Proceedings of Machine Learning Research* 32 (2):613–21. Available from https://proceedings.mlr.press/v32/yia14.html.

Zhang, Y., J. Chu, S. Chan, and B. Chan. 2019. The generalised hyperbolic distribution and its subclass in the analysis of a new era of cryptocurrencies: Ethereum and its financial risk. *Physica A: Statistical Mechanics and Its Applications* 526:120900. doi:10.1016/j.physa.2019.04.136.