SCALING IN MEAN-VARIANCE GAUSSIAN MIXTURES AND ITS IMPACT ON VOLATILITY MODELING

Farrukh Javed* **and Krzysztof Podgórski**†
Department of Statistics, Lund University, Sweden

April 14, 2025

Abstract

The mean-variance Gaussian mixtures randomize scaling parameters of the normal distribution, making them attractive for volatility modeling where stochastic scaling appears both in the mixture and through the volatility equation. The scaling of the Gaussian part can be confounded with that of the mean, and volatility may also influence the return equation through factors beyond the scaling parameter. These complexities can lead to confusion and impact key properties such as variance, asymmetry, and tail behavior. This article examines how mean-variance mixture scaling affects volatility and clarifies the role of different parameterizations of generalized hyperbolic distributions. Using volatility models based on the inverse Gaussian mixing variable, we highlight fundamental differences previously overlooked. For US stock prices, we show that a model where volatility enters only as a scale exhibits more variability than one where it also affects the concentration parameter. This distinction underscores the importance of proper model identification for investors in risk assessment.

Keywords: Generalized hyperbolic distributions, mean-variance Gaussian mixtures, scaling parameters, volatility modeling.

1 Introduction

The variance Gaussian mixtures are stochastic models that randomize scaling parameters of the normal distribution. For this reason, they may be attractive for volatility models in finance, where scaling (volatility) plays a critical and central role. Thus stochastic scaling in this context

*Corresponding author: Farrukh.Javed@stat.lu.se
†Krzysztof.Podgorski@stat.lu.se

is present both in the mixture model and additionally through the volatility equation. It is thus critical to distinguish one from the other and clarify the scaling properties of variance Gaussian mixtures. This becomes even more challenging for the mean-variance Gaussian mixtures which naturally extend variance mixtures to account for asymmetry. In this case, the scaling of the Gaussian part can be confounded with the scaling of the mean part and this confusion may affect the volatility equation by dramatically influencing key properties, such as the distribution's variance, asymmetry, and tail behavior. Our goal is to analyze the problem and explicitly identify how these mean-variance mixtures have been used in the past and thus set the foundation for their proper use in the future.

The importance of clarification of the role of different parameters and their effect on the model properties is especially important for the Gaussian mixture models since they are widely used across various fields due to their flexibility in modeling complex data patterns. One popular class of such mixture models is made of the generalized hyperbolic (GH) distributions and arises in the case when the stochastic mixture has a generalized inversed Gaussian (GIG) distribution. This GH class is a fairly flexible while simple alternative to the Gaussian distributions, see, for instance, Eberlein and Prause (2002). The GH distributions are conveniently expressed as normal variance-mean mixtures and are particularly useful for modeling heavier-than-normal-tailed data. The *five-parameter family* of GH class is flexible enough to capture the classical pentad of distribution characteristics - *center, scale, tail, skewness,* and *peakedness* - while effectively modeling asymmetry and heavy tails through four key parameters corresponding to the fundamental statistical moments: mean, variance, skewness, and kurtosis. These attributes make the GH family particularly valuable in fields like finance, where such features are commonly referred to as the *"stylized facts"*. Due to their ability to simultaneously capture the peaked and skewed behavior often observed in financial return data, GH distributions and some of their subclasses have been extensively utilized for modeling financial data in various studies, including Madan and Seneta (1990), Eberlein and Keller (1995), Barndorff-Nielsen (1997), Madan et al. (1998),

Barndorff-Nielsen and Shephard (2001), and Jensen and Lunde (2001). The GH class includes important subclasses: when a gamma distribution governs the mixing distribution, the resulting model is commonly called the variance gamma model or generalized Laplace (GL) model; when the inverse Gaussian distribution governs the stochastic variance, the distribution is called the Normal Inverse Gaussian (NIG) distribution.

Using Gaussian mixture models has additional benefits in mathematical finance, where random volatility embedded in the time series models is at the center. There the stochastic equation for the volatility can be combined with the stochastic variance mix in the main equation. The clarification of different parameterizations and investigating how they enter different models and their roles in capturing important properties of the financial time series becomes even more important. This is the aim of this work.

In particular, we present and compare different parameterizations of the scale and location parameterization in the normal inverse Gaussian (NIG) distribution. This distribution with a certain parameterization has been introduced in Barndorff-Nielsen (1997). However, the way this model has been used in the volatility modeling requires clarifying how the scale and location parameters influence the mixture representation. We explore various strategies for parameterizing the scale and examine their impact on the behavior and characteristics of the resulting model. By clarifying these differences, we illustrate how various parameterizations influence both the theoretical interpretation and practical application of the NIG distribution by using different scenarios for the volatility models. This distinction is particularly important because the choice of parameterization directly impacts key model behaviors such as skewness, kurtosis, and tail properties, which are critical in financial modeling, especially for volatility estimation. The article highlights how small variations in the parameterization of the scale can profoundly influence the model's behavior, altering its dynamics and, consequently, its ability to accurately reflect real-world phenomena.

We conclude the article with a practical example that uses volatility models to illustrate how

different parameterizations of the NIG distribution can produce distinct volatility dynamics. This case study emphasizes the critical importance of selecting the correct parameterization, as it directly affects the model's volatility behavior.

2 The scaling of the Gaussian mean-variance mixtures

A popular method of extending Gaussian distribution to account for asymmetry and fat tails is mixing normal distributions. Such mixtures are distributed according to a density that is represented as a weighted average of normal densities. In particular, a variance mixture of normal densities is given through

$$f(x) = \int_0^\infty \frac{1}{\sqrt{\gamma}}\, \phi(x/\sqrt{\gamma})\, dF(\gamma),$$

where ϕ is the standard normal density and $dF(\gamma)$ an arbitrary probability distribution on $[0,\infty)$ that serves as weights with which the (normal) densities are mixed together. The variance mixture of normal distributions is equivalently given as the distribution of $X = \sqrt{\Gamma}Z$, where Γ is distributed according to F and independently of a standard normal variable Z. A well-known extension of this idea of mixing of normal densities is given in the following definition.

Definition 1. A random variable X (and also the corresponding distribution) is called a *normal variance-mean mixture* with a non-negative and non-constant *mixing variable* Γ with the *mixing distribution* F, the variance scale $\sigma > 0$, the mean scale $\mu \in \mathbb{R}$, and the center $\lambda \in \mathbb{R}$, if

$$X = \sigma\sqrt{\Gamma}Z + \mu\Gamma + \lambda, \tag{1}$$

where Z is a standard normal variable independent of Γ. We write $M(\mu, \sigma, \lambda; F)$ for the distribution of X.

When the scale parameter is considered in discussions of the volatility models, it is more natural to consider $\nu = \mu/\sigma$, so that σ becomes the scaling parameter. With this adjustment

$$X = \sigma\left(\sqrt{\Gamma}Z + \nu\Gamma\right) + \lambda. \tag{2}$$

This is naturally referred to as a *location-scale mean-variance normal mixture* model, with $\lambda \in \mathbb{R}$ as a location and $\sigma > 0$ as as a scale.

Following this definition, it is natural to consider a family of normal variance-mean mixtures by considering some specific family of mixing variable. A family of distributions $\mathcal{D}$ is called semi-parametric if there is a function D from $P \times \mathcal{F}$ to $\mathcal{D}$, where P is a subset of a finite dimensional space, $\mathcal{F}$ an arbitrary set, and $\mathcal{D} = \{D(p;F); p \in P, F \in \mathcal{F}\}$. The mapping D is a semi-parameterization of $\mathcal{D}$, if it is one-to-one.

Definition 2. Let $\mathcal{S}$ be a family of distributions on the positive half-line. A semi-parametric family of distributions $\mathcal{M} = \{M(\mu,\sigma,\lambda;F); (\mu,\sigma,\lambda) \in \mathbb{R}^2 \times (0,\infty), F \in \mathcal{S}\}$ is called a *normal variance-mean mixture* generated by the *mixing familiy* $\mathcal{S}$, if its members have the stochastic representation (1) in which the distribution F of Γ belongs to $\mathcal{S}$ and the mapping M is its *semi-parametrization*, i.e. $(\mu,\sigma,\lambda;F)$ defines the unique distribution $M(\mu,\sigma,\lambda;F) \in \mathcal{M}$.

For a distribution F on $(0,\infty)$ and $\delta > 0$, let $F_\delta(A) = F(A/\delta)$, where $A \subset (0,\infty)$ is measurable. Consider a family $\mathcal{S}_1$ of distributions on $(0,\infty)$. A family $\mathcal{S}$ is called a *scale family* if it is semi-parameterized by $S(\delta;F) = F_\delta$, with the scale parameter $\delta > 0$, $F \in \mathcal{S}_1 \subset \mathcal{S}$. The following theorem is the basis for our discussion in the subsequent section.

Theorem 1. *Let* $\mathcal{M} = \{M(\mu,\sigma,\lambda;F); \mu \in \mathbb{R}, \sigma > 0, \lambda \in \mathbb{R}, F \in \mathcal{S}\}$ *is a normal variance-mean mixture family, and* $\mathcal{S} = \{S(\delta;F_1) : \delta > 0, F_1 \in \mathcal{S}_1\}$ *is a scale family, then*

$$\mathcal{M} = \{M(\mu,\sigma,\lambda;F); \mu \in \mathbb{R}, \sigma > 0, \lambda \in \mathbb{R}, F \in \mathcal{S}_1\}$$

and $M : \mathbb{R} \times (0,\infty) \times \mathbb{R} \times \mathcal{S}_1 \mapsto \mathcal{M}$ *is its semi-parameterization.*

3 Scaling of the GIG distributions

The GH distributions have been successfully applied to various financial data and are fairly flexible while simple alternative to the Gaussian distributions, see, for instance, Barndorff-Nielsen

(1997), Jensen and Lunde (2001), and Madan and Seneta (1990). They are characterized by their log-density being a hyperbola and considered as a reasonable alternative for modeling heavier than normal tailed data and known to simultaneously capture the peaked and skewed behavior of returns data. The class of generalized hyperbolic (GH) distributions is obtained as normal variance-mean mixtures (1) with Γ having a generalized inverse Gaussian (GIG) distribution. To clarify the relation between the parameters of the mixture and the mixing variable, we start with parametrization of the latter.

Definition 3. The GIG distributions constitute a three parameter family which is defined through the densities

$$\tilde{f}(x;\psi,\chi,\tau) = \frac{(\psi/\chi)^{\tau/2}}{2K_\tau(\sqrt{\psi\chi})} x^{\tau-1} e^{-(\psi x+\chi/x)/2}, \;\; x>0,$$

where K_τ is the Bessel function of the second kind and the parameters satisfy

$$\begin{aligned} &\psi>0\,,\chi\geq 0\,, \text{ if } \tau>0,\\ &\psi>0\,,\chi>0\,, \text{ if } \tau=0,\\ &\psi\geq 0\,,\chi>0\,, \text{ if } \tau<0, \end{aligned} \tag{3}$$

with this parameter domain denoted by $\mathcal{X}$.

Parameterization 1. A member of the so-called parametrized family of distributions is denoted by $\widetilde{GIG}(\psi,\chi,\tau)$, with (ψ,χ,τ) satisfying (3).

Let us recall important special subclasses of the GIG distributions.

Example 1 (*Inverse Gaussian).* The inverse Gaussian distribution (the first passage time by a Brownian motion of a fixed level) is a GIG with $\tau=-1/2$, and we denote it as $\widetilde{IG}(\psi,\chi) = \widetilde{GIG}(\psi,\chi,-1/2)$.

Example 2 (*Gamma and its reciprocal).* The gamma distribution is GIG with $\chi=0$ and $\psi,\tau>0$, i.e. $\widetilde{G}(\psi,\tau) = \widetilde{GIG}(\psi,0,\tau)$. The reciprocal gamma distribution is GIG with $\psi=0$, $\chi>0$, and

$\tau < 0$, i.e. $\widetilde{RG}(\chi,\tau) = \widetilde{GIG}(0,\chi,\tau)$. The standard normalizing constant in these two cases is obtained by the limiting-at-zero properties of the Bessel function.

The three special cases of GIG mentioned above specify the three corresponding GH classes, namely the normal inverse Gaussian (NIG) distributions with the inverse Gaussian mixing, the generalized asymmetric Laplace (GAL) distributions with the gamma mixing, and the inverse gamma normal mean-variance mixtures. For more detailed information we refer to Eberlein and Keller (1995) for the generalized hyperbolic distributions, to Jørgensen (1982) for the generalized inverse Gaussian distributions, and to Kotz et al. (2001) for the generalized Laplace distributions.

It is not difficult to notice that the class of the GIG distributions is invariant on scaling thus, in fact, it is a scale family of distributions. Indeed, if $\Gamma \sim \widetilde{GIG}(\psi,\chi,\tau)$, then

$$\delta\Gamma \sim \widetilde{GIG}(\psi/\delta, \chi\delta, \tau).$$

However, neither of the three parameters $(\psi,\chi,\tau) \in \mathcal{X}$ represents the scale parameter. This has been noticed in Jørgensen (1982), where an alternative parametrization has been presented $(\tilde{\delta},\kappa) = \left(\sqrt{\chi/\psi}, \sqrt{\chi\psi}\right); \psi > 0, \chi > 0, \tau \in \mathbb{R}$, with $\tilde{\delta} > 0$ as a scale parameter. Here it is important to note that this parameterization is incapable of accounting for, in a unified fashion, the gamma and reciprocal gamma distributions. For this reason, we provide the following alternative scale parameterization of the GIG distribution that is free of this flaw.

Parameterization 2. For $\tau \in \mathbb{R}$, define $\rho_\tau = \mathbb{I}_{[0,\infty)}(\tau)$ and the new parameterization, denoted by $GIG(\delta,\kappa,\tau)$, is given by

$$(\delta,\kappa,\tau) = (\rho_\tau/\psi + (1-\rho_\tau)\chi, \sqrt{\chi\psi}, \tau), \tag{4}$$

that maps $\mathcal{X}$ onto the set $\mathcal{Y}$ that is given by

$$\delta > 0, \tau \in \mathbb{R}, \text{ if } \kappa > 0,$$

$$\delta > 0, \tau \neq 0, \text{ if } \kappa = 0,$$

Distribution	**Notation**	**Mean**	**Variance**
Generalized	$\widetilde{GIG}(\psi,\chi,\tau)$	$\sqrt{\frac{\chi}{\psi}}\,R_\tau\left(\sqrt{\chi\psi}\right)$	$\frac{\chi}{\psi}\left(1+2\frac{\tau+1}{\sqrt{\chi\psi}}\,R_\tau\left(\sqrt{\chi\psi}\right)-R_\tau^2\left(\sqrt{\chi\psi}\right)\right)$
inverse Gaussian	$GIG(\delta,\kappa,\tau)$	$\frac{\bar{\kappa}_\tau\delta}{\kappa_\tau}R_\tau(\kappa)$	$\frac{\bar{\kappa}_\tau^2\delta^2}{\kappa_\tau^2}\left(1+2\frac{\tau+1}{\kappa}\,R_\tau(\kappa)-R_\tau^2(\kappa)\right)$
Inverse Gaussian, $\tau=-1/2$	$\widetilde{IG}(\psi,\chi)$	$\sqrt{\frac{\chi}{\psi}}$	$\sqrt{\frac{\chi}{\psi^3}}$
	$IG(\delta,\kappa)$	$\frac{\delta}{\kappa}$	$\frac{\delta^2}{\kappa^3}$
Gamma, $\chi=\kappa=0$, $\tau>0$	$\widetilde{G}(\psi,\tau)$	$\frac{\tau}{\psi}$	$\frac{\tau}{\psi^2}$
	$G(\delta,\tau)$	$\tau\delta$	$\tau\delta^2$
Reciprocal gamma, $\psi=\kappa=0$, $\tau<0$	$\widetilde{RG}(\chi,\tau)$	$\frac{\chi}{\tau-1}$	$\frac{\chi^2}{(\tau-1)^2(\tau-2)}$
	$RG(\delta,\tau)$	$\frac{\delta}{\tau-1}$	$\frac{\delta^2}{(\tau-1)^2(\tau-2)}$

Table 1: Parametrizations of the GIG distributions and their mean and variance. Here $R_\tau(x)=K_{\tau+1}(x)/K_\tau(x)$, for which $R_{-1/2}(x)=1$ and $R_{\tau+1}(x)=2\frac{\tau+1}{x}+1/R_\tau(x)$, see the Appendix of Kotz et al. (2001).

and its inverse mapping from $\mathcal{Y}$ onto $\mathcal{X}$:

$$(\psi,\chi,\tau)=\left(\kappa_\tau^2/\delta,\bar{\kappa}_\tau^2\delta,\tau\right), \tag{5}$$

where $\kappa_\tau^2=\rho_\tau+(1-\rho_\tau)\kappa^2$ and $\bar{\kappa}_\tau^2=\rho_\tau\kappa^2+(1-\rho_\tau)$. The distribution has two shape parameters κ and τ. To distinct between them we call κ the *concentration* parameter and τ simply the *shape* parameter, see also Embrechts (1983) and Whitmore and Yalovsky (1978).

Remark 1. It is straightforward to verify that δ acts as a scaling parameter for the GIG distribution. Therefore, any GIG distribution can be represented in the form $F_\delta(\cdot;\kappa,\tau)$, where $F\sim GIG(1,\kappa,\tau)$. This scaling property can be symbolically described as follows:

$$\delta\cdot GIG(1,\kappa,\tau)=GIG(\delta,\kappa,\tau).$$

The family of distributions, denoted by $\mathcal{S}_1 = \{F_1(\cdot;\kappa,\tau), \kappa \in [0,\infty), \tau \in \mathbb{R}\}$ has distributions $GIG(1,\kappa,\tau)$ expressed by the following densities on the positive half-line

$$f_1(x;\kappa,\tau) = \frac{(\kappa_\tau/\bar{\kappa}_\tau)^\tau}{2K_\tau(\kappa)} x^{\tau-1} e^{-\left(\kappa_\tau^2 x + \bar{\kappa}_\tau^2/x\right)/2}. \tag{6}$$

We conclude the section by presenting important special subclasses of the GIG distributions structured according to the proposed parameterization.

Example 3 (Inverse Gaussian). For the inverse Gaussian distribution, $IG(\delta,\kappa) = \widetilde{IG}(\psi,\chi)$, where $\psi = \kappa^2/\delta$ and $\chi = \delta$ and thus for $\delta = 1$:

$$f(x;\kappa) = \frac{\kappa^{2\tau}}{2K_\tau(\kappa)} x^{-3/2} e^{-\left(\kappa^2 x + 1/x\right)/2}.$$

Example 4 (Gamma and its reciprocal). For the gamma distribution, $\tau > 0$, and $\kappa = 0$, so $G(\delta,\tau) = \widetilde{G}(\psi,\tau)$, where $\psi = 1/\delta$. For the reciprocal gamma distribution, $\tau < 0$, $\kappa = 0$, so that $RG(\delta,\tau) = \widetilde{G}(\chi,\tau)$, where $\chi = \delta$. In the new parameterization, these two classes can be treated as one with $\kappa = 0$ and $\tau \neq 0$, i.e. as $G(\delta,\tau) = GIG(\delta,0,\tau)$ with the sign of τ indicating if it is the gamma or its reciprocal and $\delta > 0$ being a natural scale parameter. This case corresponds to $(0,\infty) \times \{0\} \times \mathbb{R} \setminus \{0\}$ part of the parameter space $\mathcal{Y}$.

4 The scaling parametrization of the GH distributions

As previously noted, the class of GH distributions holds considerable importance in fields such as financial mathematics due to its flexibility in capturing skewness, peakedness, and heavy tails. As a consequence to Theorem 1, we can now rigorously define the parameterization of GH distributions, linking them explicitly to the parameterization of GIG distributions. This connection ensures that the GH distributions are parameterized properly by establishing a direct relationship between the two distribution families.

Definition 4. If a GIG distribution $F_1(\cdot;\kappa,\tau)$ is mixing in (1) that defines X, i.e. we use $\mathcal{S}_1 = \{GIG(1,\kappa,\tau); \kappa \in [0,\infty), \tau \in \mathbb{R}\}$ in Theorem 1. Then the corresponding GH distribution is uniquely

defined by the parameters $(\lambda, \sigma, \mu, \kappa, \tau)$ with the range $\mathcal{Z}$:

$$\begin{aligned} &\lambda \in \mathbb{R}, \sigma > 0, \mu \in \mathbb{R}, \tau \in \mathbb{R}, \text{ if } \kappa > 0, \\ &\lambda \in \mathbb{R}, \sigma > 0, \mu \in \mathbb{R}, \tau \neq 0, \text{ if } \kappa = 0. \end{aligned} \tag{7}$$

Parameterization 3. The parameterization that follows from the above definition symbolically is written as $GH(\lambda, \sigma, \mu, \kappa, \tau)$. In general, if $GIG(\delta, \kappa, \tau)$ is mixing in (1), then $X \sim GH(\lambda, \sigma\sqrt{\delta}, \mu\delta, \kappa, \tau)$.

For the scale-location version of the mixture model given in (2), we consider $\nu = \mu/\sigma$, so that σ acts as the scaling parameter with μ as the *asymmetry* parameter, κ as the *concentration* parameter, and τ as the *shape* parameter. With this adjustment, the GH distributions form a location-scale family, where $\Gamma \sim GIG(1, \kappa, \tau)$ and $X \sim GH(\lambda, \sigma, \nu\sigma, \kappa, \tau)$.

Parameterization 4. The location-scale parameterization is denoted by $\widehat{GH}(\lambda, \sigma, \nu, \kappa, \tau)$. Additionally, if $\Gamma \sim GIG(\delta, \kappa, \tau)$, then the associated X follows $\widehat{GH}(\lambda, \sigma\sqrt{\delta}, \nu\sqrt{\delta}, \kappa, \tau)$, as presented in Parameterization 3.

Now, we present the densities of the considered distribution for completeness. For $\Gamma \sim GIG(1, \kappa, \tau)$, the density of the random variable $\tilde{X} = \sqrt{\Gamma}Z + \nu\Gamma$ at $x \in \mathbb{R}$ is given by the following expression

$$\begin{aligned} f(x; \nu, \kappa, \tau) = &\frac{(\kappa_\tau/\bar{\kappa}_\tau)^\tau}{\sqrt{2\pi}\sqrt{\kappa_\tau^2 + \nu^2}^{\,\tau-\frac{1}{2}} K_\tau(\kappa)} \times \\ &\times \exp(\nu x)\sqrt{\bar{\kappa}_\tau^2 + x^2}^{\,\tau-\frac{1}{2}} K_{\tau-\frac{1}{2}}\left(\sqrt{(\kappa_\tau^2 + \nu^2)(\bar{\kappa}_\tau^2 + x^2)}\right), \end{aligned} \tag{8}$$

see Appendix A for full derivation. This leads to the location-scale family distribution $\widehat{GH}(\lambda, \sigma, \nu, \kappa, \tau)$, which can be expressed by:

$$\begin{aligned} \hat{f}(x; \lambda, \sigma, \nu, \kappa, \tau) = &\frac{(\kappa_\tau/\bar{\kappa}_\tau)^\tau}{\sqrt{2\pi}\sigma\sqrt{\kappa_\tau^2 + \nu^2}^{\,\tau-\frac{1}{2}} K_\tau(\kappa)} \times \\ &\times \exp\left(\nu\frac{x-\lambda}{\sigma}\right)\sqrt{\bar{\kappa}_\tau^2 + \left(\frac{x-\lambda}{\sigma}\right)^2}^{\,\tau-\frac{1}{2}} K_{\tau-\frac{1}{2}}\left(\sqrt{(\kappa_\tau^2 + \nu^2)\left(\bar{\kappa}_\tau^2 + \left(\frac{x-\lambda}{\sigma}\right)^2\right)}\right). \end{aligned} \tag{9}$$

The obtained full probability density function in our parameterization can be compared with the more classical from Barndorff-Nielsen (1978), where the corresponding density is given by

$$\tilde{f}(x;\alpha,\beta,\gamma,\lambda,\tau) = \frac{\sqrt{\alpha^2-\beta^2}^{\tau}}{\sqrt{2\pi}\alpha^{\tau-\frac{1}{2}}\gamma^{\tau}K_{\tau}\left(\gamma\sqrt{\alpha^2-\beta^2}\right)}\times \\ \times\exp\left(\beta(x-\lambda)\right)\sqrt{\gamma^2+(x-\lambda)^2}^{\,\tau-\frac{1}{2}}K_{\tau-\frac{1}{2}}\left(\alpha\sqrt{\gamma^2+(x-\lambda)^2}\right), \quad (10)$$

where the range $\tilde{\mathcal{Z}}$ of the parameters is given by

$$\begin{aligned} &\alpha>0,\ \beta\in(-\alpha,\alpha),\ \gamma>0,\ \tau\in\mathbb{R},\ \lambda\in\mathbb{R},\\ &\alpha>0,\ \beta=\pm\alpha,\ \gamma>0,\ \tau>0,\ \lambda\in\mathbb{R},\\ &\alpha>0,\ \beta\in(-\alpha,\alpha),\ \gamma=0,\ \tau<0,\ \lambda\in\mathbb{R}. \end{aligned} \quad (11)$$

Parameterization 5. The classical parameterization $\widetilde{GH}(\alpha,\beta,\gamma,\lambda,\tau)$ is given by the density (8) and the range (11).

The relations between the parameters (ν,κ,σ) and (α,β,γ) are established by,

$$\alpha=\frac{\sqrt{\kappa_\tau^2+\nu^2}}{\sigma},\ \beta=\frac{\nu}{\sigma},\ \gamma=\sigma\bar{\kappa}_\tau, \quad (12)$$

$$\nu=\beta\begin{cases}\frac{1}{\sqrt{\alpha^2-\beta^2}} & ;\tau\geq 0\\ \gamma & ;\tau<0\end{cases},\ \sigma=\begin{cases}\frac{1}{\sqrt{\alpha^2-\beta^2}} & ;\tau\geq 0\\ \gamma & ;\tau<0\end{cases},\ \kappa=\gamma\sqrt{\alpha^2-\beta^2}, \quad (13)$$

while the parameters (τ,λ) are the same in $GH(\lambda,\sigma,\mu,\kappa,\tau)$ and $\widehat{GH}(\lambda,\sigma,\nu,\kappa,\tau)$.

Remark 2. Since the generalized hyperbolic distribution can be represented as $GH(\lambda,\sigma,\mu,\kappa,\tau)=\widehat{GH}(\lambda,\sigma,\mu/\sigma,\kappa,\tau)$, it is possible to express the parameters of the classical form $\widetilde{GH}(\alpha,\beta,\gamma,\lambda,\tau)$ using the following relations

$$\alpha=\frac{\sqrt{\sigma^2\kappa_\tau^2+\mu^2}}{\sigma^2},\ \beta=\frac{\mu}{\sigma^2},\ \gamma=\sigma\bar{\kappa}_\tau, \quad (14)$$

$$\mu=\beta\begin{cases}\frac{1}{\alpha^2-\beta^2} & ;\tau\geq 0\\ \gamma^2 & ;\tau<0\end{cases},\ \sigma=\begin{cases}\frac{1}{\sqrt{\alpha^2-\beta^2}} & ;\tau\geq 0\\ \gamma & ;\tau<0\end{cases},\ \kappa=\gamma\sqrt{\alpha^2-\beta^2}, \quad (15)$$

With these relations, we can now clearly establish the connection between the classical parameterization of $\widetilde{GIG}(\psi,\chi,\tau)$ and $\widetilde{GH}(\alpha,\beta,\gamma,\lambda,\tau)$, which follows from the mixture representations in (1) and (2). This refinement clarifies the relationships between parameters while simplifying the explanation of the transformations between different parameterizations.

Proposition 1. *Let* $\Gamma \sim \widetilde{GIG}(\psi,\chi,\tau)$ *and X be the mean-variance mixture distribution given in* (1) *with* $\sigma = 1$. *Then*

$$X \sim \widetilde{GH}(\alpha,\beta,\gamma,\lambda,\tau),$$

and

$$(\alpha,\beta,\gamma) = \left(\sqrt{\psi+\mu^2},\mu,\sqrt{\chi}\right),$$

$$(\psi,\chi) = (\alpha^2-\beta^2,\gamma^2).$$

In Barndorff-Nielsen (1997), the scale and location parameterization of the normal inverse gaussian (NIG) distributions has been introduced, though it does not explicitly relate to the scaling of the mixing variable. Using the above notation, we now have the classical parametrization $\widetilde{NIG}(\alpha,\beta,\gamma,\lambda) = \widetilde{GH}(\alpha,\beta,\gamma,\lambda,-1/2)$ and the location-scale family $\widehat{NIG}(\lambda,\sigma,\nu,\kappa) = \widehat{GH}(\lambda,\sigma,\nu,\kappa,-1/2)$. The original location-scale paramaterization has been defined by

$$(\bar{\alpha},\bar{\beta}) = (\alpha\sigma,\beta\sigma).$$

However, in this definition the two parameterizations are used simultaneously to define $\bar{\alpha}$ and $\bar{\beta}$, which can be confusing and making it harder to distinguish between the two. To clarify, we define these parameters separately for each parameterization. First, for the parameterization in $\widetilde{NIG}(\alpha,\beta,\gamma,\lambda)$, we have

$$(\bar{\alpha},\bar{\beta}) = (\alpha\gamma,\beta\gamma), \tag{16}$$

and this location-scale family is defined by

$$\overline{NIG}(\bar{\alpha},\bar{\beta},\gamma,\lambda) = \widetilde{NIG}(\bar{\alpha}/\gamma,\bar{\beta}/\gamma,\gamma,\lambda).$$

Second, for the parametrization in $\widehat{NIG}(\lambda,\sigma,\nu,\kappa)$, we have

$$(\bar{\alpha},\bar{\beta}) = \left(\sqrt{\kappa^2+\nu^2},\nu\right),$$
$$(\nu,\kappa) = \left(\bar{\beta},\sqrt{\bar{\alpha}^2-\bar{\beta}^2}\right),$$

Thus, the relationship between the two location-scale families is expressed as,

$$\widehat{NIG}\left(\lambda,\gamma,\bar{\beta},\sqrt{\bar{\alpha}^2-\bar{\beta}^2}\right) = \overline{NIG}(\bar{\alpha},\bar{\beta},\gamma,\lambda).$$

Extending this to the generalized hyperbolical distribution given in (10), one can use (16) to get the density

$$\bar{f}(x;\bar{\alpha},\bar{\beta},\gamma,\lambda,\tau) = \frac{\sqrt{\bar{\alpha}^2-\bar{\beta}^2}^{\tau}}{\sqrt{2\pi}\bar{\alpha}^{\tau-\frac{1}{2}}\gamma K_\tau\left(\sqrt{\bar{\alpha}^2-\bar{\beta}^2}\right)}\times$$
$$\times\exp\left(\bar{\beta}\frac{x-\lambda}{\gamma}\right)\sqrt{1+\left(\frac{x-\lambda}{\gamma}\right)^2}^{\tau-\frac{1}{2}}K_{\tau-\frac{1}{2}}\left(\bar{\alpha}\sqrt{1+\left(\frac{x-\lambda}{\gamma}\right)^2}\right), \quad (17)$$

which, in turns, leads to another location-scale parameterization.

Parameterization 6. The location-scale parameterization $\overline{GH}(\bar{\alpha},\bar{\beta},\gamma,\lambda,\tau)$ is defined by the density (17) and through the range that is the same as in (11) due to (16).

The following relation between the two location-scale parameterizations $\overline{GH}(\bar{\alpha},\bar{\beta},\gamma,\lambda,\tau)$ and $\widehat{GH}(\lambda,\sigma,\nu,\kappa,\tau)$ holds

$$\sigma = \gamma\begin{cases}\sqrt{\bar{\alpha}^2-\bar{\beta}^2} & ;\tau\geq 0\\ 1 & ;\tau<0\end{cases},\ \nu = \bar{\beta}\begin{cases}\frac{1}{\sqrt{\bar{\alpha}^2-\bar{\beta}^2}} & ;\tau\geq 0\\ 1 & ;\tau<0\end{cases},\ \kappa = \sqrt{\bar{\alpha}^2-\bar{\beta}^2}, \quad (18)$$

$$\bar{\alpha} = \nu\bar{\kappa}_\tau,\ \bar{\beta} = \bar{\kappa}_\tau\sqrt{\kappa_\tau^2+\nu^2},\ \gamma = \sigma/\bar{\kappa}_\tau, \quad (19)$$

while λ and τ are the same in both parameterization.

For referencing convenience, Table 2 summarizes the parameterizations in a concise manner.

5 Stochastic scaling in the volatility modeling

For iid Gaussian innovations e_t's, and iid mixing variables Γ_t's, we consider the following general framework for equity returns

$$y_t = \lambda_t + \eta_t \Gamma_t + r_t \sqrt{\Gamma_t} e_t. \tag{20}$$

Here λ_t is centering and r_t^2 represents time-varying volatility, both depending on the history of returns, i.e. on the filtration $\mathcal{F}_{t-1}$ generated by $\{(\Gamma_s, e_s); s < t\}$. Further, η_t is the skewing parameter that also depends on r_t. Two types of this dependence are considered

- r_t can be viewed as a scaling of $\sqrt{\Gamma_t}$ leading to $\eta_t = \beta r_t^2$,
- r_t can scale the entire skewed term $\beta\Gamma_t + \sqrt{\Gamma_t} e_t$, resulting in $\eta_t = \beta r_t$.

In the domain of volatility models, r_t is often specified as an autoregressive process driven by shocks through

$$r_t^{\vartheta} = c + b r_{t-1}^{\vartheta} + \phi_{t-1},$$

$\vartheta > 0$ (with often recommended choice $\vartheta = 1$ or $\vartheta = 2$), in which ϕ_t is a parametrized function of some random components that involve e_t and, possibly, Γ_t.

The above approaches are natural due to the scaling interpretation of r_t. However, in the case when the distributions of Γ_t's are taken from a parameterized family of distributions, the dependencies between the volatility and return equations can be introduced through other than scaling parameters. In this case

$$y_t = \lambda_t + \mu\Gamma_t + \sigma\sqrt{\Gamma_t} e_t, \tag{21}$$

in which Γ_t has a distribution dependent on r_t. This can be seen in Jensen and Lunde (2001), where r_t enters into the model through the χ parameter that affects the distribution of Γ_t in a manner other than direct scaling of the mixing variable or the entire mixture distribution. These

different forms of dependence on r_t lead to fundamentally different models with distinct properties, a fact which often has been overlooked in the literature leading to misinterpretation of empirical results. The remainder of this section aims to clarify these different model specifications and demonstrate the consequences of the model confusion.

We summarize our discussion by explicitly stating three models exhibiting three types of dependence on the volatilities r_t's:

Mixing variable scaling model: r_t^2 is the scaling parameter in the mixing variable Γ_t. Specifically, $\Gamma_t|\mathcal{F}_{t-1} \sim GIG(r_t^2, \kappa, \tau)$ which leads to

$$y_t|\mathcal{F}_{t-1} \sim GH(\lambda, r_t, \beta r_t^2, \kappa, \tau).$$

From Table 2, the location-scale parameterization becomes

$$y_t|\mathcal{F}_{t-1} \sim \widehat{GH}(\lambda, r_t, \beta, \kappa, \tau).$$

the classical parameterization becomes

$$y_t|\mathcal{F}_{t-1} \sim \widetilde{GH}\left(\sqrt{\frac{\kappa_\tau^2}{r_t^2} + \beta^2}, \beta, r_t\bar{\kappa}_\tau, \lambda, \tau\right).$$

and the classical location-scale parametrization becomes

$$y_t|\mathcal{F}_{t-1} \sim \overline{GH}\left(\beta r_t\bar{\kappa}_\tau, \bar{\kappa}_\tau\sqrt{\kappa_\tau^2 + \beta r_t^2}, \frac{r_t}{\bar{\kappa}_\tau}, \lambda, \tau\right).$$

Mixture scaling model: r_t is scaling to the whole mixture $\beta\Gamma_t + \sqrt{\Gamma_t}e_t$ in which $\Gamma_t|\mathcal{F}_{t-1} \sim GIG(1, \kappa, \tau)$. This leads to

$$y_t|\mathcal{F}_{t-1} \sim GH(\lambda, r_t, \beta r_t, \kappa, \tau)$$

From Table 2, the *Scale-location* parameterization becomes

$$y_t|\mathcal{F}_{t-1} \sim \widehat{GH}\,(\lambda, r_t, \beta, \kappa, \tau)\,.$$

the classical parameterization becomes

$$y_t|\mathcal{F}_{t-1} \sim \widetilde{GH}\left(\frac{\sqrt{\kappa_t^2+\beta^2}}{r_t}, \frac{\beta}{r_t}, r_t\bar{\kappa}_\tau, \lambda, \tau\right).$$

and the classical location-scale parametrization becomes

$$y_t|\mathcal{F}_{t-1} \sim \overline{GH}\left(\beta\bar{\kappa}_\tau, \bar{\kappa}_\tau\sqrt{\kappa_\tau^2+\beta^2}, \frac{r_t}{\bar{\kappa}_\tau}, \lambda, \tau\right)$$

Volatility as concentration parameter: As proposed by Jensen and Lunde (2001), an alternative representation is considered where the volatility r_t enters into the mixing variable as a non-scaling parameter, namely, through the concentration parameter κ. In this case, $\Gamma_t|\mathcal{F}_{t-1} \sim GIG(1, \sqrt{\alpha^2-\mu^2}\, r_t, \tau)$leading to

$$y_t|\mathcal{F}_{t-1} \sim GH\left(\lambda, \sigma, \mu, \sqrt{\alpha^2-\mu^2}\, r_t, \tau\right).$$

and for the *scale-location* parameterization

$$y_t|\mathcal{F}_{t-1} \sim \widehat{GH}\left(\lambda, \sigma, \nu, \sqrt{\alpha^2-\mu^2}\, r_t, \tau\right).$$

The formulas for the parameters in the classical parameterizations become more convoluted and can be obtained from using Table 2 with $\kappa = \sqrt{\alpha^2-\mu^2}\, r_t$.

In Section 4 of Jensen and Lunde (2001), a *respecified* model is considered which, for the case of NIG distribution, is equivalent to our second representation above (mixture scaling model). However, the authors incorrectly stated that the distinction lies only in reparameterization. We emphasize that this is not the case, since the models not only differ in their parameterization but also exhibit distinct statistical properties. How it influences both the theoretical behavior of the models and the empirical insights derived from them is discussed next.

Our goal in the above discussion was to clarify the role of GIG scaling in volatility modeling and to resolve some confusion stemming from past treatments in the literature. We begin with the model presented in Jensen and Lunde (2001), which, using the parameterization $\widetilde{GIG}(\psi, \chi, \tau)$

is stated as follows

$$Y_t = \lambda + \beta\widetilde{\Gamma}_t + \sqrt{\widetilde{\Gamma}_t}\varepsilon_t, \tag{22}$$

where, for a fixed t, $\widetilde{\Gamma}_t|\mathcal{F}_t$ is distributed as $\widetilde{GIG}(\alpha^2-\beta^2, r_t^2, \tau)$, and $r_t = r(\varepsilon_s, s<t)$ is a deterministic function of past i.i.d. standard normal innovations. The specific form of r is irrelevant for our consideration as long as it does not depend on any parameter in (22). According to the Proposition 1, conditionally on $\mathcal{F}_{t-1}$, Y_t in (22) follows $\widetilde{GH}(\alpha, \beta, r_t, \lambda, \tau)$. Further in Jensen and Lunde (2001), following model is later introduced

$$Y_t = \lambda + r_t\left(\nu\widehat{\Gamma}_t + \sqrt{\widehat{\Gamma}_t}\varepsilon_t\right), \tag{23}$$

where $\widehat{\Gamma}_t$'s are iid $GIG(1,\kappa,\tau)$ and Z_t are iid standard normal. In the quoted work, the model is referred to as a simplified parameterization of(22). However, it is important to emphasize that model (23) is not simply a reparameterized version of (22), it represents a fundamentally different model. This distinction is critical, as clarified in the following remark.

Remark 3. In the general GIG case, we have the inequality:

$$\beta\widetilde{\Gamma}_t + \sqrt{\widetilde{\Gamma}_t}\varepsilon_t \overset{d}{\neq} r_t\left(\nu\widehat{\Gamma}_t + \sqrt{\widehat{\Gamma}_t}\varepsilon_t\right),$$

where $\widetilde{\Gamma}_t \sim \widetilde{GIG}(\alpha^2-\beta^2, r_t^2, \tau)$, regardless of the parameter values $\alpha, \beta, \tau, \kappa, \nu$. Here $\overset{d}{=}$ stands for the equivalence of these random variables in distribution. Indeed, as shown in (4), for $\delta = \rho_\tau/(\alpha^2-\beta^2) + (1-\rho_\tau)r_t^2$, we have $\widetilde{\Gamma}_t = \delta\Gamma_t$, where $\Gamma_t \sim GIG(1, r_t\sqrt{\alpha^2-\beta^2}, \tau)$. Thus for the relation

$$\beta\delta\Gamma_t + \sqrt{\delta}\sqrt{\Gamma_t}\varepsilon_t \overset{d}{=} r_t\nu\widehat{\Gamma}_t + r_t\sqrt{\widehat{\Gamma}_t}\varepsilon_t.$$

to hold, Theorem 1 requires $\beta\delta = r_t\nu$, $\delta = r_t^2$, and $\kappa = r_t\sqrt{\alpha^2-\beta^2}$, which yields $\nu = \beta r_t$ and $\kappa = \sqrt{\alpha^2-\beta^2}r_t$. This leads to a contradiction, as α, β, ν, and κ are assumed to be fixed, time-independent parameters.

It should be noted that Jensen and Lunde (2001) only considered the specific case of the NIG distribution, i.e. $\tau = -1/2$. In this case, the first model yields $\delta = r_t^2$, which is true for any $\tau < 0$.

In fact, in the NIG case, a fundamental difference arises between the two models, as described below:

Remark 4. In the case of the NIG (Normal Inverse Gaussian) distributions, consider two models for returns:

$$Y_t = \lambda + \beta r_t^2 \Gamma_t + r_t \sqrt{\Gamma_t}\varepsilon_t \tag{24}$$

$$\widehat{Y}_t = \lambda + \nu r_t \widehat{\Gamma}_t + r_t \sqrt{\widehat{\Gamma}_t}\varepsilon_t, \tag{25}$$

where $\Gamma_t = IG(1, r_t\sqrt{\alpha^2-\beta^2})$ and $\widehat{\Gamma}_t = IG(1,\kappa)$ The key distinction between these models lies in the shape parameter of the Inverse Gaussian distribution. In model (24), the concentration parameter $r_t\sqrt{\alpha^2-\beta^2}$ is time-varying, while in (25), the concentration parameter κ is constant. This time-varying versus constant parameter is critical to how each model behaves, particularly in capturing the volatility dynamics and modelling the asset returns. Importantly, these models cannot be matched in terms of their mean and variance unless the second parameters of the IG distributions in both models are equal, i.e. it must be $\kappa = r_t\sqrt{\alpha^2-\beta^2}$. The reasoning behind this claim is provided in the appendix.

6 Statistical consequences of a model misrepresentation

In this section, we discuss the statistical properties of each representation in greater detail, focusing on key aspects such as volatility clustering, return skewness and kurtosis, and sensitivity to parameter changes. More specifically, the consequences of incorrectly applying model (23) when the fit is actually obtained from the model (22), are presented. To illustrate these consequences we utilize empirical study of Jensen and Lunde (2001) based on the data set of 31 412 daily US stock prices from 2/17/1885 to 12/31/1997. The dataset, hereafter referred to as the LMI series, serves as a robust historical basis for evaluating model performance and understanding the discrepancies that result from incorrect parameterization. By using this data, we are able to

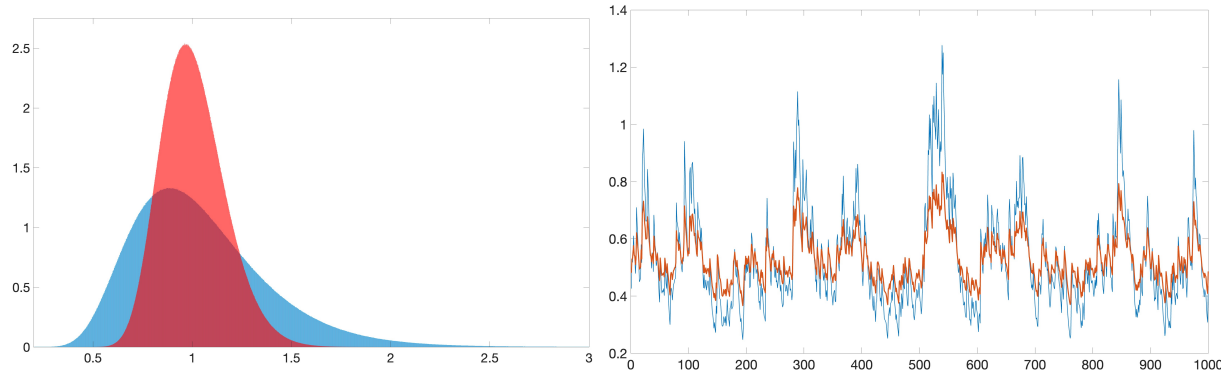

Figure 1: *Left:* The histogram of r_t (red) and r_t^2 (blue) simulated using (26). *Right:* The simulated values of $Var(Y_t|\mathcal{F}_t)$ (red) and $Var(\widehat{Y}_t|\mathcal{F}_t)$ (blue) under the condition that the unconditional variances are matched. One can observe that the model $\hat{Y}_t$ is more volatile. The parameters are taken from the last rows of Tables 3-5.

highlight the differences in model behavior and statistical outcomes, particularly in terms of how volatility and return dynamics are captured. In that work, and following Ding et al. (1993), the volatility r_t is modeled using the asymmetric power ARCH (A-PARCH) model

$$r_t^{\vartheta} = c + b r_{t-1}^{\vartheta} + a\big(|\varepsilon_{t-1}| - \varphi\varepsilon_{t-1}\big)^{\vartheta}. \tag{26}$$

Following the original paper, we report in Table 3 several specifications of the above model based on the fit to the LMI series using the maximum likelihood approach. Using the Monte Carlo simulation, the first two moments of r_t are evaluated and reported them in the table (analytical values are unavailable except in special cases). The histograms of the simulated volatility values and their squares based on the parameters from the last row of Table 3 are shown in Figure 1 *(Left)*. As discussed earlier, Jensen and Lunde (2001) considers two distinct NIG models ($\tau = -1/2$), which are given in Remark 4. In the fit values presented in that work, α and β were used. Therefore, we assume the first model was applied, with the fit to the NIG model displayed in Table 4. Please note that the models in Table 4 correspond to the same order as those in Table 3. Since $\bar{\alpha} = \alpha/\delta$ and $\bar{\beta} = \beta/\delta$ and in the case of (24) $\delta = r_t^2$, $\bar{\alpha}$ and $\bar{\beta}$ are not inherent parameters within the framework of the first model. Consequently, the only way to compare these models

is by using the second model to match certain characteristics of the two models. The simplest method for such comparison is to match the moments of the two models. In this respect, we have the following result.

Proposition 2. *The conditional expectations with respect to the natural filtration* $\mathcal{F}_t = \sigma(\varepsilon_s, s < t)$ *of the models given in* (24) *and* (25) *are linear functions of* r_t*:*

$$E(Y_t|\mathcal{F}_t) = \lambda + \frac{\beta}{\sqrt{\alpha^2 - \beta^2}}\, r_t,$$
$$E(\widehat{Y}_t|\mathcal{F}_t) = \lambda + \frac{\nu}{\kappa}\, r_t.$$

Moreover, the conditional variance of (24) *is a linear function of* r_t*:*

$$Var(Y_t|\mathcal{F}_t) = \frac{\alpha^2}{(\alpha^2 - \beta^2)^{\frac{3}{2}}}\, r_t, \tag{27}$$

while the conditional variance of (25) *is a linear function of* r_t^2*:*

$$Var(\widehat{Y}_t|\mathcal{F}_t) = \frac{\nu^2 + \kappa^2}{\kappa^3}\, r_t^2. \tag{28}$$

As a direct consequence of the above, we have the following corollary.

Corollary 1. *The processes* Y_t *and* $\widehat{Y}_t$ *can be matched such that* $E(Y_t|\mathcal{F}_t) = E(\widehat{Y}_t|\mathcal{F}_t)$ *in which the case*

$$\frac{\beta^2}{\alpha^2 - \beta^2} = \frac{\nu^2}{\kappa^2}.$$

Although it is not possible to match the conditional variances of the two models, it is possible to match their unconditional mean and variance so that the observed values of the returns obtained from the two models will have the sample variances converging in the long run to the same quantities. The conditions of such a match are provided in the next corollary.

Corollary 2. *The processes* Y_t *and* $\widehat{Y}_t$ *cannot be matched such that* $Var(Y_t|\mathcal{F}_t) = Var(\widehat{Y}_t|\mathcal{F}_t)$*. However, if their conditional means are equal, then they can be matched such that* $Var(Y_t) =$

$Var(\widehat{Y}_t)$ in which the case

$$\nu = \frac{M_2}{M_1}\beta, \tag{29}$$

$$\kappa = \sqrt{\alpha^2 - \beta^2}\frac{M_2}{M_1}. \tag{30}$$

These conditions allow the parameters (α, β) to be matched with (ν, κ) facilitating comparisons between the two models. These comparisons are presented in the remaining part of this section and the matched parameters are given in Table 5.

Remark 5. Since the conditional variance of $\widehat{Y}_t$ depends on r_t^2 while that of Y_t on r_t, the $\widehat{Y}_t$ exhibits more volatility, as seen in Figure 1 *(Right)*. This holds despite the fact that their average variances are the same, as indicated in column 5 of Tables 4 and 5.

As we have seen, the differences between the models are evident in their conditional variances, even when the unconditional variances are matched. The higher moments, conditional or otherwise, cannot be fully matched, we turned, therefore, our attention to examining the difference between the third and fourth moments. These findings are summarized in the next result.

Proposition 3. *Let us consider the following conditional central moments of Y_t and $\widehat{Y}_t$*

$$CM_3(r_t) = E\left(\left(Y_t - E\left(Y_t|r_t\right)\right)^3|r_t\right), \widehat{CM}_3(r_t) = E\left(\left(\widehat{Y}_t - E(\widehat{Y}_t|r_t)\right)^3|r_t\right),$$
$$CM_4(r_t) = E\left(\left(Y_t - E(Y_t|r_t)\right)^4|r_t\right), \widehat{CM}_4(r_t) = E\left(\left(\widehat{Y}_t - E(\widehat{Y}_t|r_t)\right)^4|r_t\right).$$

Then

$$CM_3(r_t) = \frac{3\alpha^2\beta}{(\alpha^2-\beta^2)^{5/2}}r_t,$$
$$CM_4(r_t) = 3\left(\alpha^2\frac{\alpha^2+4\beta^2}{(\alpha^2-\beta^2)^{7/2}} + \frac{4\alpha^4-6\alpha^2\beta^2+3\beta^4}{(\alpha^2-\beta^2)^3}r_t\right)r_t,$$
$$\widehat{CM}_3(r_t) = \frac{3\nu}{\kappa^3}\left(\frac{\nu^2}{\kappa^2}+1\right)r_t^3,$$
$$\widehat{CM}_4(r_t) = \frac{3}{\kappa^3}\left(\frac{5+\kappa}{\kappa^4}\nu^4 + 2\frac{3+\kappa}{\kappa^2}\nu^2 + 1 + 4\kappa\right)r_t^4.$$

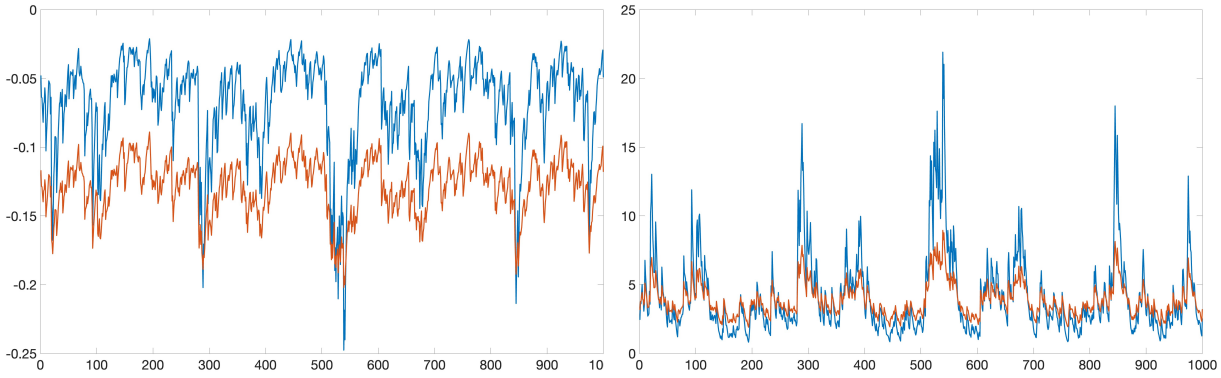

Figure 2: Comparison of the conditional third (left) and fourth moments (right) of the models (24) (red) and (25) (blue) under the parameters from the last row of Tables 4 and 5

The difference between the two models, as formally demonstrated in the above result, is visually illustrated in Figure 2, where samples of the conditional moments are presented using fitted parameter values.

Conclusions

Given the extensive use of the NIG distribution in financial volatility modeling, it is essential for practitioners to understand the subtle differences in parameterization. Choosing the right parameterization allows the model to accurately capture the intended volatility characteristics, thus improving its reliability and effectiveness in practical applications. In this article, we present and compare two distinct parameterizations of the distribution, detailing how variations in scale and location parameters shape each mixture representation. We examine different strategies for parameterizing scale and assess how these choices affect model behavior and characteristics. By clarifying these distinctions, we show how different parameterizations impact both the theoretical interpretation and practical use of the NIG distribution in statistical modeling. An empirical analysis further demonstrates how parameterization choices influence volatility model behavior, highlighting the importance of selecting the right parameterization for accurate and effective volatility modeling.

A Proofs and derivations

Proof of Theorem 1. Following Definition 2, in order to prove that M restricted to $\mathbb{R}^2 \times (0,\infty) \times \mathcal{S}_1$ is a semi-parameterization of $\mathcal{M}$, we need to show two conditions: firstly, that any distribution represented as in (1), with the distribution F of Γ in $\mathcal{S}$ can be also obtained by such a representation with the distribution F of Γ in $\mathcal{S}_1$; secondly, that the restricted mapping M is one-to-one. If $D \in \mathcal{M}$ has the stochastic representation $\sigma\sqrt{\Gamma}Z + \mu\Gamma + \lambda$ with the distribution F of Γ in $\mathcal{S}$, then there are $s > 0$ and a distribution F_1 in $\mathcal{S}_1$ such that $\Gamma/s \sim F_1$. Thus D has also representation $\tilde{\sigma}\sqrt{\tilde{\Gamma}}Z + \tilde{\mu}\tilde{\Gamma} + \lambda$, where $\tilde{\sigma} = \sigma\sqrt{s}$ and $\tilde{\mu} = \mu s$ and $\tilde{\Gamma} = \Gamma/s \in \mathcal{S}_1$. Let us assume that a distribution from $\mathcal{M}$ has the representations: $\sigma\sqrt{\Gamma}Z + \mu\Gamma + \lambda$ and $\tilde{\sigma}\sqrt{\tilde{\Gamma}}Z + \tilde{\mu}\tilde{\Gamma} + \tilde{\lambda}$ with the distributions of Γ and $\tilde{\Gamma}$ both in $\mathcal{S}_1$. First, $\tilde{\lambda} = \lambda$, since Γ cannot be constant. By evaluating the characteristic functions, we obtain

$$\int_0^\infty e^{-\left(\frac{\sigma^2 t^2}{2} + i\mu t\right)\gamma} dF(\gamma) = \int_0^\infty e^{-\left(\frac{\tilde{\sigma}^2 t^2}{2} + i\tilde{\mu} t\right)\gamma} d\tilde{F}(\gamma).$$

By equalling the imaginary parts of the above, we get

$$\int_0^\infty \sin(\gamma t)\, dF_\mu(\gamma) = \int_0^\infty \sin(\gamma t)\, d\tilde{F}_{\tilde{\mu}}(\gamma).$$

By considering Radon measures that are asymmetrizations of F_μ and $\tilde{F}_{\tilde{\mu}}$ and using the fact that two Radon measures with the same Fourier transform are identical, we obtain that $\tilde{\mu}\tilde{\Gamma}$ has the same distribution as $\mu\Gamma$, see Corollary 6.3, Maruyama (2018). Since both $\tilde{F}$ and F belongs to $\mathcal{S}_1$ and the mapping $(\mu; F) \mapsto F_\mu$ constitutes a semi-parametrization, $\mu = \tilde{\mu}$ and $F = \tilde{F}$. By taking the variance of $\sigma\sqrt{\Gamma}Z + \mu\Gamma$ and $\tilde{\sigma}\sqrt{\Gamma}Z + \mu\Gamma$, one gets also $\tilde{\sigma} = \sigma$ (or, if Γ does not have a finite second moment, one can replace it by its truncation). □

Let us consider a variable X as in (2) with $\lambda = 0$, $\sigma = 1$, and the distribution of Γ given by a

density f_Γ. Then its density is given by

$$\begin{aligned} f_X(x) &= \frac{1}{\sqrt{2\pi}} \int_0^\infty \gamma^{-1/2} f_\Gamma(\gamma) e^{-\frac{(x-\nu\gamma)^2}{2\gamma}} d\gamma \\ &= \frac{e^{\nu x}}{\sqrt{2\pi}} \int_0^\infty \gamma^{-1/2} f_\Gamma(\gamma) e^{-\frac{x^2}{2\gamma} - \frac{\nu^2\gamma}{2}} d\gamma. \end{aligned}$$

Consider now f_Γ as given in (6), so we get

$$\begin{aligned} f_X(x) &= \frac{(\kappa_\tau/\bar{\kappa}_\tau)^\tau e^{\nu x}}{2\sqrt{2\pi} K_\tau(\kappa)} \int_0^\infty \gamma^{\tau-3/2} e^{-\left((\kappa_\tau^2+\nu^2)\gamma + (\bar{\kappa}_\tau^2+x^2)/\gamma\right)/2} d\gamma \\ &= \frac{e^{\nu x} \sqrt{\bar{\kappa}_\tau^2 + x^2}^{\,\tau-1/2} K_{\tau-\frac{1}{2}}\left(\sqrt{(\kappa_\tau^2+\nu^2)(\bar{\kappa}_\tau^2+x^2)}\right)}{\sqrt{2\pi}(\bar{\kappa}_\tau/\kappa_\tau)^\tau \sqrt{\kappa_\tau^2+\nu^2}^{\,\tau-1/2} K_\tau(\kappa)}, \end{aligned}$$

where we use the well-known relation for the modified Bessel function of the second type

$$K_{\tau-1/2}(u) = \int_0^\infty t^{-\tau-3/2} e^{-(t+1/t)u/2}\, dt.$$

Proof of Proposition 1. First let $\tau \geq 0$. By (4), $\Gamma \sim \widetilde{GIG}(\psi,\chi,\tau) = GIG\left(1/\psi, \sqrt{\chi\psi}, \tau\right)$. By Parameterization 3, $X \sim GH(\lambda, 1/\sqrt{\psi}, \mu/\psi, \sqrt{\chi\psi}, \tau)$. Then Remark 2 leads to

$$X \sim \widetilde{GH}(\sqrt{\psi+\mu^2}, \mu, \sqrt{\chi}, \lambda, \tau),$$

and $\psi = \alpha^2 - \beta^2$, $\chi = \gamma^2$.

If $\tau < 0$, then $\widetilde{GIG}(\psi,\chi,\tau) = GIG\left(\chi, \sqrt{\chi\psi}, \tau\right)$ and $X \sim GH(\lambda, \sqrt{\chi}, \mu\chi, \sqrt{\chi\psi}, \tau)$. Again by Remark 2, $X \sim \widetilde{GH}(\sqrt{\psi+\mu^2}, \mu, \sqrt{\chi}, \lambda, \tau)$, and again $\psi = \alpha^2 - \beta^2$ and $\chi = \gamma^2$. □

Proof of Proposition 2. The relations for the conditional means follow immediately from the formulas for the means of the IG distribution given in Table 1. Regarding the conditional variances, we use the same table to obtain, for the first model,

$$\begin{aligned} Var(Y_t|r_t) &= Var(E(Y_t|r_t,\Gamma_t)|r_t) + E(Var(Y_t|r_t,\Gamma_t)|r_t) \\ &= \beta^2 r_t^4 Var(\Gamma_t|r_t) + r_t^2 E(\Gamma_t|r_t) \\ &= \frac{\alpha^2}{(\alpha^2-\beta^2)^{\frac{3}{2}}} r_t, \end{aligned}$$

and, for the second model,

$$\begin{aligned} Var(\widehat{Y}_t|r_t) &= Var(E(\widehat{Y}_t|r_t,\widehat{\Gamma}_t)|r_t) + E(Var(\widehat{Y}_t|r_t,\widehat{\Gamma}_t)|r_t) \\ &= \nu^2 r_t^2 Var(\widehat{\Gamma}_t|r_t) + r_t^2 E(\widehat{\Gamma}_t|r_t) \\ &= \frac{\nu^2+\kappa^2}{\kappa^3} r_t^2. \end{aligned}$$

□

Proof of Corollary 2. One cannot match the conditional variances of the two models since the first one is linear in r_t, while the second depends linearly on r_t^2 as shown in Proposition 2. Assuming that the conditional means are matched, one chooses the parameters so that $E(Var(Y_t|r_t)) = E(Var(\widehat{Y}_t|r_t))$, which by Proposition 2 leads to the second matching equation

$$\frac{\alpha^2}{(\alpha^2-\beta^2)^{\frac{3}{2}}} M_1 = \frac{\nu^2+\kappa^2}{\kappa^3} M_2.$$

This together with Corollary 2 yields the result. □

Proof of Proposition 3. We note the following straightforward relation of the first four moments of X given in (1) in terms of the moment $m_1 = E\Gamma$ and the central moments $\bar{m}_k = E(\Gamma - m_1)^k$:

$$\begin{aligned} E(X) &= \mu m_1 + \lambda,\ Var(X) = \mu^2 \bar{m}_2 + \sigma^2 m_1, \\ E(X-EX)^3 &= \mu^3 \bar{m}_3 + 3\mu\sigma^2 \bar{m}_2, \\ E(X-EX)^4 &= \mu^4 \bar{m}_4 + 6\mu^2\sigma^2 (\bar{m}_3 + \bar{m}_2 m_1) + 3\sigma^4 \left(\bar{m}_2 + 4m_1^2\right). \end{aligned} \tag{31}$$

If $\Gamma \sim IG(1,\kappa)$, then the following formulas for the moments hold

$$m_1 = \frac{1}{\kappa},\ \bar{m}_2 = \frac{1}{\kappa^3},\ \bar{m}_3 = \frac{3}{\kappa^5},\ \bar{m}_4 = \frac{3}{\kappa^6}\left(\frac{5}{\kappa}+1\right). \tag{32}$$

By applying these to the model (24), one substitutes μ by βr_t^2, σ by r_t, and κ by $r_t\sqrt{\alpha^2-\beta^2}$, while for the model (25), μ by νr_t, σ by r_t, κ stays. After some elementary while tedious algebra,

the result is obtained from the following

$$
\begin{aligned}
CM_3(r_t) &= \frac{3\beta^3 r_t}{(\alpha^2-\beta^2)^{5/2}} + \frac{3\beta r_t}{(\alpha^2-\beta^2)^{3/2}},\\
\widehat{CM}_3(r_t) &= \frac{3\nu^3 r_t^3}{\kappa^5} + \frac{3\nu r_t^3}{\kappa^3},\\
CM_4(r_t) &= 3\left(\frac{\beta^4\left(5+\sqrt{\alpha^2-\beta^2}r_t\right)}{(\alpha^2-\beta^2)^{7/2}} + \frac{2\beta^2\left(3+\sqrt{\alpha^2-\beta^2}r_t\right)}{(\alpha^2-\beta^2)^{5/2}} + \frac{1+4\sqrt{\alpha^2-\beta^2}r_t}{(\alpha^2-\beta^2)^{3/2}}\right) r_t,\\
\widehat{CM}_4(r_t) &= \frac{3\nu^4 r_t^4}{\kappa^6}\left(\frac{5}{\kappa}+1\right) + \frac{6\nu^2 r_t^4}{\kappa^4}\left(\frac{3}{\kappa}+1\right) + 3r_t^4\left(\frac{1}{\kappa^3}+\frac{4}{\kappa^2}\right).
\end{aligned}
$$

□

Argument behind Remark 4. Matching the two models through the mean and the variance, which by using (31) leads to the following relations

$$
\begin{aligned}
\beta r_t m_1 &= \nu m_1',\\
\beta^2 r_t^2 \bar{m}_2 + m_1 &= \nu^2 \bar{m}_2' + m_1',
\end{aligned}
$$

where the moments of the inverse Gaussian distributions given in (32) satisfy $m_1 = \frac{1}{r_t\sqrt{\alpha^2-\beta^2}}$, $m_1' = \frac{1}{\kappa}$, and $\bar{m}_2 = \frac{1}{r_t^3(\alpha^2-\beta^2)^{3/2}}$, $\bar{m}_2' = \frac{1}{\kappa^3}$. Thus

$$
\begin{aligned}
\frac{\beta^2}{\alpha^2-\beta^2} &= \frac{\nu^2}{\kappa^2},\\
\frac{1}{r_t\sqrt{\alpha^2-\beta^2}}\frac{\alpha^2}{\alpha^2-\beta^2} &= \frac{1}{\kappa}\left(\frac{\nu^2}{\kappa^2}+1\right),
\end{aligned}
$$

so that necessarily $\kappa = r_t\sqrt{\alpha^2-\beta^2}$. □

Acknowledgments

The first author acknowledges partial support from the Torsten Söderberg grant (DNR: E36/29), while the second author acknowledges partial support from the Swedish Research Council (VR) grant (DNR: 2020-05168).

References

Barndorff-Nielsen, O. (1978). Hyperbolic distributions and distributions on hyperbolae. *Scandinavian Journal of Statistics*, 5(3):151–157.

Barndorff-Nielsen, O. (1997). Normal inverse gaussian distributions and stochastic volatility modelling. *Scand. J. Statist.*, 24:1–13.

Barndorff-Nielsen, O. and Shephard, N. (2001). Non-gaussian ornstein-uhlenbeck-based models and some of their uses in financial economics (with discussion). *Journal of the Royal Statistical Society B*, 63:167–241.

Ding, Z., Granger, C., and Engle, R. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1):83–106.

Eberlein, E. and Keller, U. (1995). Hyperbolic distributions in finance. *Bernoulli*, 1:281–299.

Eberlein, E. and Prause, K. (2002). *The generalized hyperbolic model: financial derivatives and risk measures*. Springer.

Embrechts, P. (1983). A property of the generalized inverse gaussian distribution with some applications. *Journal of Applied Probability*, 20(3):537–544.

Jensen, M. B. and Lunde, A. (2001). The NIG-S&ARCH model: A fat-tailed, stochastic, and autoregressive conditional heteroskedastic volatility model. *Econometrics Journal*, 4:167–342.

Jørgensen, B. (1982). *Statistical properties of the generalized inverse Gaussian distribution*. Lecture Notes in Statistics. Springer-Verlag.

Kotz, S., Kozubowski, T., and Podgórski, K. (2001). *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*. Birkhauser.

Madan, D. B., Carr, P., and Chang, E. C. (1998). The variance gamma process and option pricing. *Review of Finance.*, 2:74–105.

Madan, D. B. and Seneta, E. (1990). The variance-gamma (v.g.) model for share markets returns. *Journal of Business.*, 63:511–524.

Maruyama, T. (2018). *Fourier Analysis of Economic Phenomena*, pages 121–163. Springer Singapore, Singapore.

Whitmore, G. A. and Yalovsky, M. (1978). A normalizing logarithmic transformation for inverse gaussian random variables. *Technometrics*, 20(2):207–208.

No.	Description	Notation and range	Relations
1)	*Mean-variance GIG mixture* (1) λ- location μ- asymmetry (through GIG mixing mean) σ- size of GIG variance κ, τ - concentration/shape parameters, controlling mass in tail and center,	$GH(\lambda, \sigma, \mu, \kappa, \tau)$ Parameter range: $\mathcal{Z}$ in (7)	**2)** $\mu = \nu\sigma$ **3)** Eq. (15)
2)	*Location-scale* (2) λ, σ- location/scale ν- asymmetry (through GIG mixing mean) κ, τ- as in **1)**.	$\widehat{GH}(\lambda, \sigma, \nu, \kappa, \tau)$ Parameter range as in **1)**	**1)** $\nu = \mu/\sigma$ **3)** Eq. (13) **4)** Eq. (18)
3)	*Classical* (10) α- steepness β- asymmetry γ, τ - center/shape λ - location	$\widetilde{GH}(\alpha, \beta, \gamma, \lambda, \tau)$ Parameter range: $\widetilde{\mathcal{Z}}$ in (11)	**1)** Eq. (14) **2)** Eq. (12) **4)** $\alpha = \bar{\alpha}/\gamma$, $\beta = \bar{\beta}/\gamma$
4)	*Classical location-scale* (17) $\bar{\alpha}$ - steepness $\bar{\beta}$ - asymmetry λ, γ - location/scale τ - tail shape	$\overline{GH}(\bar{\alpha}, \bar{\beta}, \gamma, \lambda, \tau)$ Parameter range as in **3)**	**2)** Eq. (19) **3)** $\bar{\alpha} = \alpha\gamma$, $\bar{\beta} = \beta\gamma$

Table 2: Parametrizations of generalized hyperbolic distributions.

Asymmetry β	a	b	c	ϑ	φ	M_1	M_2
0	0.089	0.900	0.011	2	0	0.990	1.000
-0.124	0.087	0.902	0.011	2	0	0.990	1.000
0	0.089	0.910	0.014	1.288	0.363	0.989	1.004
-0.154	0.088	0.915	0.012	1.263	0.378	1.006	1.038

Table 3: Parameter estimates and corresponding moments for the volatility model reported from Jensen and Lunde (2001) for different asymmetry parameter β in (24). The moments are numerically evaluated using the Monte Carlo method, with the sample size 10^8 and the standard error not exceeding 10^{-3}.

β	α	λ	$E(Y_t)$	$Var(Y_t)$	$Skew(Y_t)$	$Kurt(Y_t)$
0	1.759	0.060	0.060	0.563	0.001	4.787
-0.124	1.740	0.123	0.052	0.574	-0.169	4.842
0	1.941	0.056	0.056	0.510	-0.001	4.637
-0.154	1.917	0.122	0.041	0.530	-0.183	4.680

Table 4: Parameter estimates of the NIG return models (24) as reported in Jensen and Lunde (2001) together with the first four moment characteristics obtained by the Monte Carlo simulations from the fitted model.

ν	κ	λ	$E(\hat{Y}_t)$	$Var(\hat{Y}_t)$	$Skew(\hat{Y}_t)$	$Kurt(\hat{Y}_t)$
0	1.777	0.060	0.060	0.563	0.001	5.077
-0.125	1.753	0.123	0.052	0.574	-0.178	5.137
0	1.969	0.056	0.056	0.510	-0.001	5.009
-0.159	1.972	0.122	0.041	0.530	-0.196	5.054

Table 5: Parameter estimates of the NIG return models (25) obtained from (24) by the matching equations (29) and (30) together with the first four moment characteristics obtained by the Monte Carlo simulations from the model.