Hindawi Publishing Corporation
The Scientific World Journal
Volume 2015, Article ID 909231, 16 pages
http://dx.doi.org/10.1155/2015/909231

Research Article

The Lambert Way to Gaussianize Heavy-Tailed Data with the Inverse of Tukey's *h* Transformation as a Special Case

Georg M. Goerg

Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Correspondence should be addressed to Georg M. Goerg; im@gmge.org

Received 25 July 2014; Accepted 1 October 2014

Academic Editor: Taizhong Hu

Copyright © 2015 Georg M. Goerg. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

I present a parametric, bijective transformation to generate heavy tail versions of arbitrary random variables. The tail behavior of this *heavy tail Lambert* W $\times F_X$ random variable depends on a tail parameter $\delta \geq 0$: for $\delta = 0$, $Y \equiv X$, for $\delta > 0$ Y has heavier tails than X. For X being Gaussian it reduces to Tukey's h distribution. The Lambert W function provides an explicit inverse transformation, which can thus remove heavy tails from observed data. It also provides closed-form expressions for the cumulative distribution (cdf) and probability density function (pdf). As a special case, these yield analytic expression for Tukey's h pdf and cdf. Parameters can be estimated by maximum likelihood and applications to S&P 500 log-returns demonstrate the usefulness of the presented methodology. The R package `LambertW` implements most of the introduced methodology and is publicly available on `CRAN`.

1. Introduction

Statistical theory and practice are both tightly linked to Normality. In theory, many methods require Gaussian data or noise: (i) regression often assumes Gaussian errors; (ii) many time series models are based on Gaussian white noise [1–3]. In such cases, a model $\mathcal{M}_{\mathcal{N}}$, parameter estimates and their standard errors, and other properties are then studied, all based on the ideal(istic) assumption of Normality.

In practice, however, data/noise often exhibits asymmetry and heavy tails, for example, wind speed data [4], human dynamics [5], or Internet traffic data [6]. Particularly notable examples are financial data [7, 8] and speech signals [9], which almost exclusively exhibit heavy tails. Thus a model $\mathcal{M}_{\mathcal{N}}$ developed for the Gaussian case does not necessarily provide accurate inference anymore.

One way to overcome this shortcoming is to replace $\mathcal{M}_{\mathcal{N}}$ with a new model $\mathcal{M}_G$, where G is a heavy tail distribution: (i) regression with Cauchy errors [10]; (ii) forecasting long memory processes with heavy tail innovations [11, 12], or ARMA modeling of electricity loads with hyperbolic noise [13]. See also Adler et al. [14] for a wide range of statistical applications and methodology for heavy-tailed data.

While such fundamental approaches are attractive from a theoretical perspective, they can become unsatisfactory from a practical point of view. Many successful statistical models and techniques assume Normality, their theory is very well understood, and many algorithms are implemented for the simple and often much faster Gaussian case. Thus developing models based on an entirely unrelated distribution G is like throwing out the (Gaussian) baby with the bathwater.

It would be very useful to transform a Gaussian random variable X to a heavy-tailed random variable Y and vice versa and thus rely on knowledge and algorithms for the well-understood Gaussian case, while still capturing heavy tails in the data. Optimally such a transformation should (a) be bijective, (b) include Normality as a special case for hypothesis testing, and (c) be parametric so the optimal transformation can be estimated efficiently.

Figure 1 illustrates this pragmatic approach: researchers can make their observations $\mathbf{y}$ as Gaussian as possible ($\mathbf{x}_\tau$) before making inference based on their favorite Gaussian model $\mathcal{M}_{\mathcal{N}}$. This avoids the development of, or the data analysts waiting for, a whole new theory of $\mathcal{M}_G$ and new implementations based on a particular heavy-tailed distribution G, while still improving statistical inference from

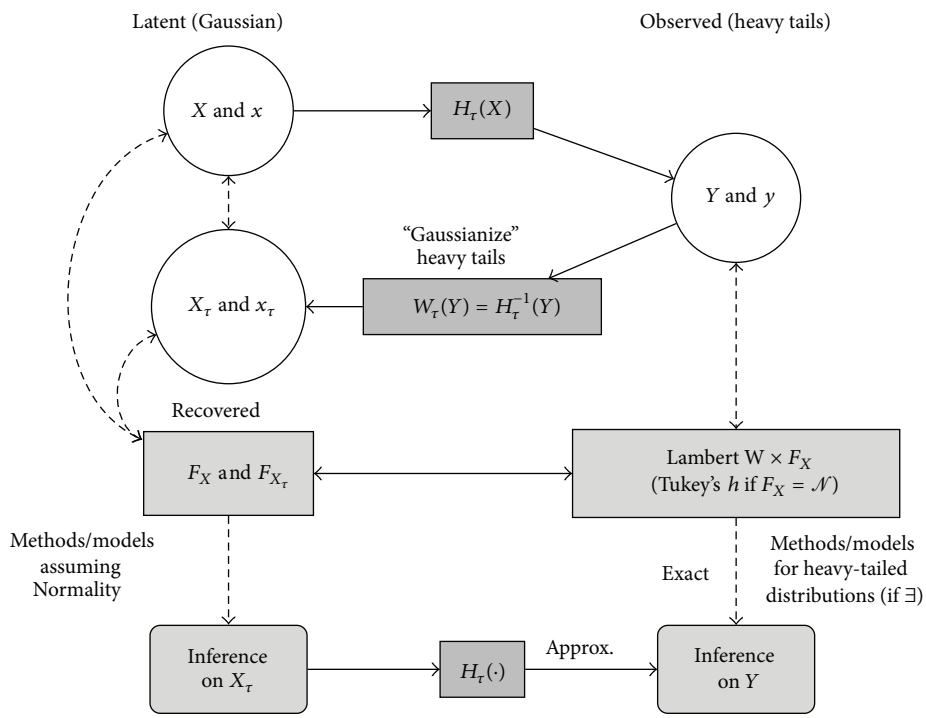


Figure 1: Schematic view of the heavy tail Lambert W × F_X framework. Latent input $X \sim F_X$: $H_\tau(X)$ from (6) transforms (solid arrows) X to $Y \sim$ Lambert W × F_X and generates heavy tails (right) Observed heavy-tailed Y and $\mathbf{y}$: (1) use $W_\tau(\cdot)$ to back-transform $\mathbf{y}$ to latent "Normal" $\mathbf{x}_\tau$, (2) use model $\mathcal{M}_\mathcal{N}$ of your choice (regression, time series models, hypothesis testing, etc.) for inference on $\mathbf{x}_\tau$, and (3) convert results back to the original "heavy-tailed world" of $\mathbf{y}$ (right).

heavy-tailed data $\mathbf{y}$. For example, consider $\mathbf{y} = (y_1, \ldots, y_{500})$ from a standard Cauchy distribution $\mathcal{C}(0, 1)$ in Figure 2(a): modeling heavy tails by a transformation makes it even possible to Gaussianize this Cauchy sample (Figure 2(c)). This "nice" data $\mathbf{x}_\tau$ can then be subsequently analyzed with common techniques. For example, the location can now be estimated using the sample average (Figure 2(d)). For details see Section 6.1.

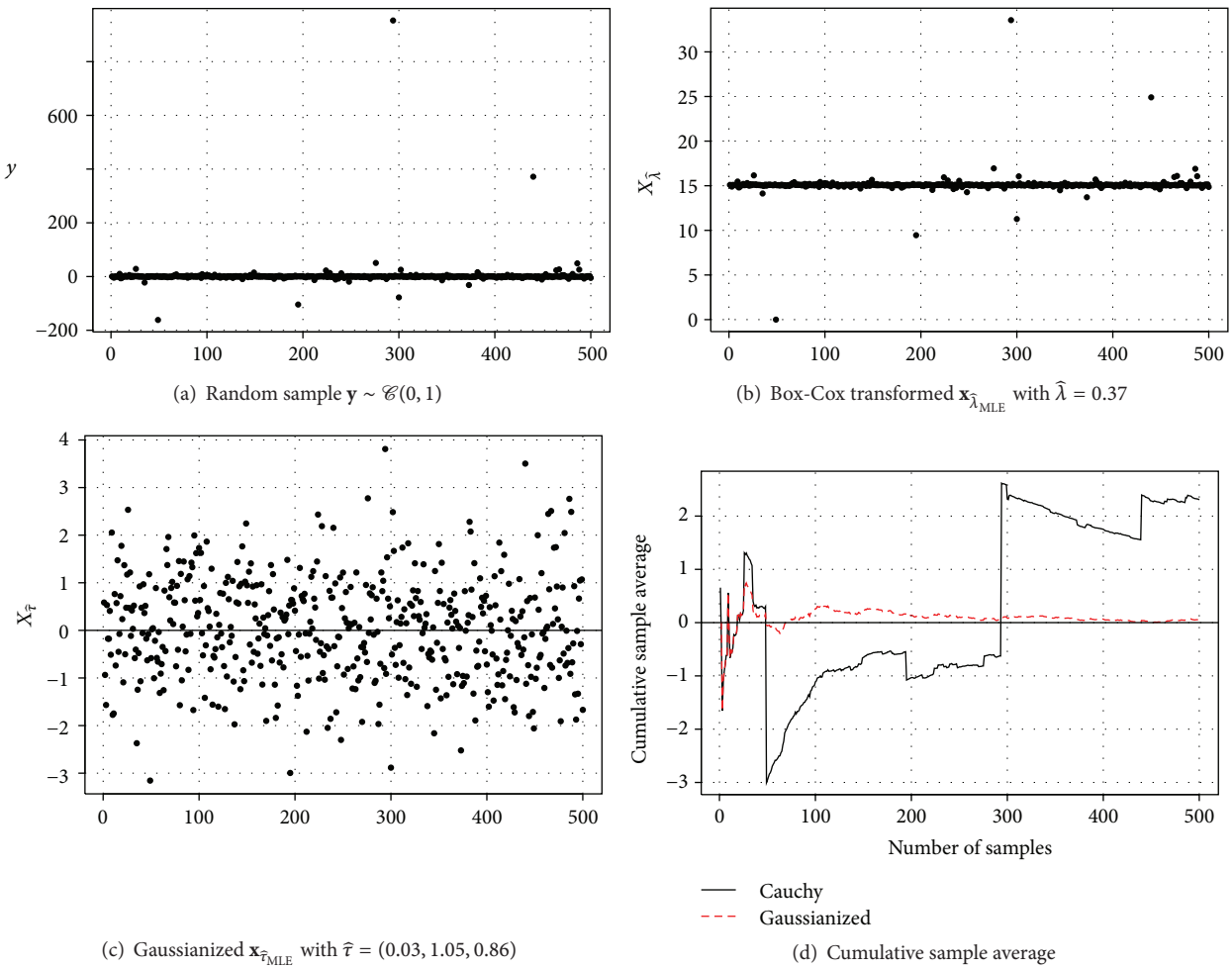
Liu et al. [15] use a semiparametric approach, where Y has a *nonparanormal* distribution if $f(Y) \sim \mathcal{N}(\mu, \sigma^2)$ and $f(\cdot)$ is an increasing smooth function; they estimate $f(\cdot)$ using nonparametric methods. This leads to a greater flexibility in the distribution of Y, but it suffers from two drawbacks: (i) nonparametric methods have slower convergence rates and thus need large samples, and (ii) for identifiability of $f(\cdot)$, $\mathbb{E}f(Y) \equiv \mathbb{E}Y$ and $\mathrm{Var}(f(Y)) \equiv \mathrm{Var}(Y)$ must hold. While (i) is inherent to nonparametric methods, point (ii) requires Y to have finite mean and variance, which is often not met for heavy-tailed data. Thus here we use parametric transformations which do not rely on restrictive identifiability conditions and also work well for small sample sizes.

The main contributions of this work are threefold: (a) a metafamily of heavy tail Lambert W × F_X distributions (see also [16]) with Tukey's h distribution [17] as a special case, (b) a bijective transformation to "Gaussianize" heavy-tailed data (Section 2), and (c) simple expressions for the cumulative distribution function (cdf) $G_Y(y)$ and probability density function (pdf) $g_Y(y)$ (Section 2.4). In particular, analytic expressions for the pdf and cdf for Tukey's h (Section 3) are presented here, to the best of the author's knowledge, for the first time in the literature.

Section 4 introduces a method of moments estimator and studies the maximum likelihood estimator (MLE). Section 5 shows their finite sample properties. As has been shown in many case studies, Tukey's h distribution (heavy tail Lambert W × Gaussian) is useful to model data with unimodal, heavy-tailed densities. Applications to S&P 500 log-returns confirm the usefulness of the Lambert W framework (Section 6). Finally, we discuss the new methodology and future work in Section 7. Detailed derivations and proofs are given in the Supplementary Material available online at http://dx.doi.org/10.1155/2015/909231.

Computations, figures, and simulations were done in R [18]. The R package `LambertW` implements most of the presented methodology and is publicly available on CRAN.

1.1. Multivariate Extensions. While this work focuses on the univariate case, multivariate extensions of the presented methods can be defined component-wise, analogously to the multivariate version of Tukey's h distribution [19]. While this may not make the transformed random variables jointly Gaussian, it still provides a good starting point for more well-behaved multivariate estimation.



(a) Random sample $\mathbf{y} \sim \mathscr{C}(0,1)$

(b) Box-Cox transformed $\mathbf{x}_{\widehat{\lambda}_{\mathrm{MLE}}}$ with $\widehat{\lambda} = 0.37$

(c) Gaussianized $\mathbf{x}_{\widehat{\tau}_{\mathrm{MLE}}}$ with $\widehat{\tau} = (0.03, 1.05, 0.86)$

(d) Cumulative sample average

Figure 2: Gaussianizing a standard Cauchy sample. For (d) $\tau^{(n)}$ was estimated for each fixed $n = 5, \ldots, 500$, before Gaussianizing $(y_1, \ldots, y_n)$.

1.2. Box-Cox Transformation. A popular method to deal with skewed, high variance data is the Box-Cox transformation

$$\mathbf{x}_\lambda = \begin{cases} \dfrac{\mathbf{y}^\lambda - 1}{\lambda}, & \text{if } \lambda > 0, \\ \log \mathbf{y}, & \text{if } \lambda = 0. \end{cases} \tag{1}$$

A major limitation of (1) is the nonnegativity constraint on $\mathbf{y}$, which prohibits its use in many applications. To avoid this limitation it is common to shift the data, $\widetilde{\mathbf{y}} = \mathbf{y} + |\min(\mathbf{y})| \geq 0$, which restricts Y to a half-open interval. If, however, the underlying process can occur on the entire real line, such a shift undermines statistical inference for yet unobserved data (see [20]). Even if out-of-sample prediction is not important for the practitioner, Figure 2(b) shows that the Box-Cox transformation in fact fails to remove heavy tails from the Cauchy sample. (We use $\widetilde{\mathbf{y}} = \mathbf{y} + |\min(\mathbf{y})| + 1$ and use `boxcox` from the `MASS` R package; $\widehat{\lambda} = 0.37$.)

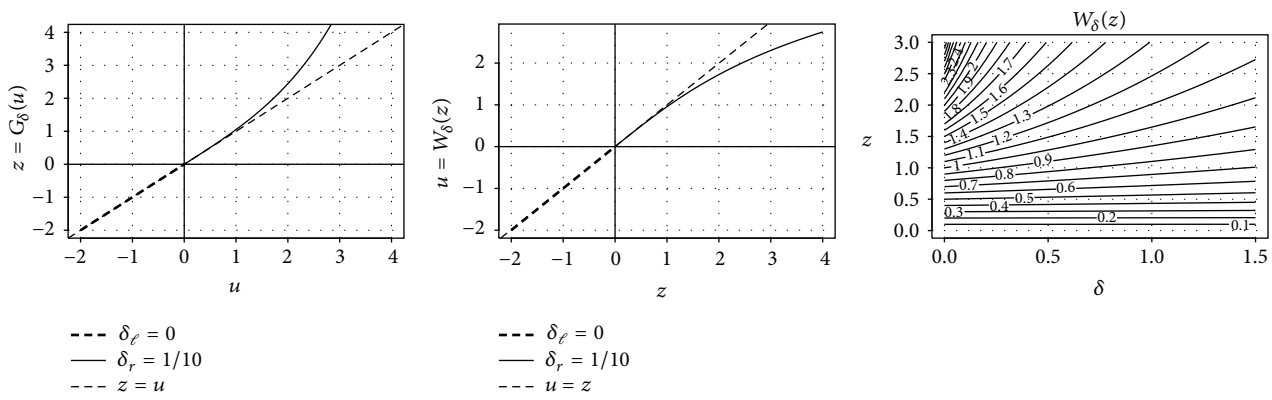
Moreover, the main purpose of the Box-Cox transformation is to stabilize variance [21–23] and remove right tail skewness [24]; a lower empirical kurtosis is merely a by-result of the variance stabilization. In contrast, the Lambert W framework is designed primarily to model heavy-tailed random variables and remove heavy tails from data and has no difficulties with negative values.

2. Generating Heavy Tails Using Transformations

Random variables exhibit heavy tails if more mass than for a Gaussian random variable lies at the outer end of the density support. A random variable Z has a tail index a if its cdf satisfies $1 - F_Z(z) \sim L(z)z^{-a}$, where $L(z)$ is a slowly varying function at infinity, that is, $\lim_{z\to\infty} L(tz)/L(z) = 1$ for all $t > 0$ [25]. (There are various similar definitions of heavy, fat, or long tails; for this work these differences are not essential.) The heavy tail index a is an important characteristic of Z; for example, only moments up to order a can exist.

2.1. Tukey's h Distribution. A parametric transformation is the basis of Tukey's h random variables [17]

$$Z = U \exp\left(\frac{h}{2}U^2\right), \quad h \geq 0, \tag{2}$$



(a) $\delta_\ell\delta_r$ transformation (3)

(b) Inverse $W_{\delta_\ell,\delta_r}(z)$ in (80) in the supplementary material

(c) Inverse $W_\delta(z)$(9) as a function of δ and z

Figure 3: Transformation and inverse transformation for $\delta_\ell = 0$ and $\delta_r = 1/10$: identity on the left (same tail behavior) and a heavy-tailed transformation in the right tail of input U.

where U is standard Normal random variable and h is the heavy tail parameter. The random variable Z has tail parameter $a = 1/h$ [17] and reduces to the Gaussian for $h = 0$. Morgenthaler and Tukey [26] extend the h distribution to the skewed, heavy-tailed family of hh random variables

$$Z = \begin{cases} U \exp\left(\frac{\delta_\ell}{2}U^2\right), & \text{if } U \le 0, \\ U \exp\left(\frac{\delta_r}{2}U^2\right), & \text{if } U > 0, \end{cases} \tag{3}$$

where again $U \sim \mathcal{N}(0,1)$. Here $\delta_\ell \ge 0$ and $\delta_r \ge 0$ shape the left and right tail of Z, respectively; thus transformation (3) can model skewed and heavy-tailed data; see Figure 3(a). For the sake of brevity let $H_\delta(u) := u\exp((\delta/2)u^2)$.

However, despite their great flexibility, Tukey's h and hh distributions are not very popular in statistical practice, because expressions for the cdf or pdf have not been available in closed form. Although Morgenthaler and Tukey [26] express the pdf of (2) as ($h \equiv \delta$)

$$g_Z(z) = \frac{f_U\left(H_\delta^{-1}(z)\right)}{H'_\delta\left(H_\delta^{-1}(z)\right)}, \tag{4}$$

they fall short of making $H_\delta^{-1}(z)$ explicit. So far the inverse of (2) or (3) has been considered analytically intractable [4, 27]. Thus parameter inference relied on matching empirical and theoretical quantiles [4, 17, 26], or by the method of moments [28]. Only recently Headrick et al. [28] provided numerical approximations to the inverse. However, numerical approximations can be slow and prohibit analytical derivations. Thus a closed form, analytically tractable pdf, which can be computed efficiently, is essential for a widespread use of Tukey's h (and variants).

In this work I present this long sought closed-form inverse, which has a large body of literature in mathematics and is readily available in standard statistics software. For ease of notation and concision main results are shown for $\delta_\ell = \delta_r = \delta$; analogous results for $\delta_\ell \ne \delta_r$ will be stated without details.

2.2. Heavy Tail Lambert W Random Variables. Tukey's h transformation (2) is strongly related to the approach taken by Goerg [16] to introduce skewness in continuous random variables $X \sim F_X(x)$. In particular, if $Z \sim$ Tukey's h, then $Z^2 \sim$ skewed Lambert W $\times\ \chi_1^2$ with skew parameter $\gamma = h$.

Adapting the skew Lambert W $\times\ F_X$ input/output idea (see Figure 1), Tukey's h random variables can be generalized to *heavy-tailed Lambert* W $\times\ F_X$ *random variables*. (Most concepts and methods from the skew Lambert W $\times\ F_X$ case transfer one-to-one to the heavy tail Lambert W random variables presented here. Thus for the sake of concision I refer to Goerg [16] for details of the Lambert W framework.)

Definition 1. Let U be a continuous random variable with cdf $F_U(u \mid \boldsymbol{\beta})$, pdf $f_U(u \mid \boldsymbol{\beta})$, and parameter vector $\boldsymbol{\beta}$. Then,

$$Z = U \exp\left(\frac{\delta}{2}U^2\right), \quad \delta \in \mathbb{R}, \tag{5}$$

is a *noncentral, nonscaled heavy tail Lambert* W $\times\ F_U$ random variable with parameter vector $\theta = (\boldsymbol{\beta}, \delta)$, where δ is the tail parameter.

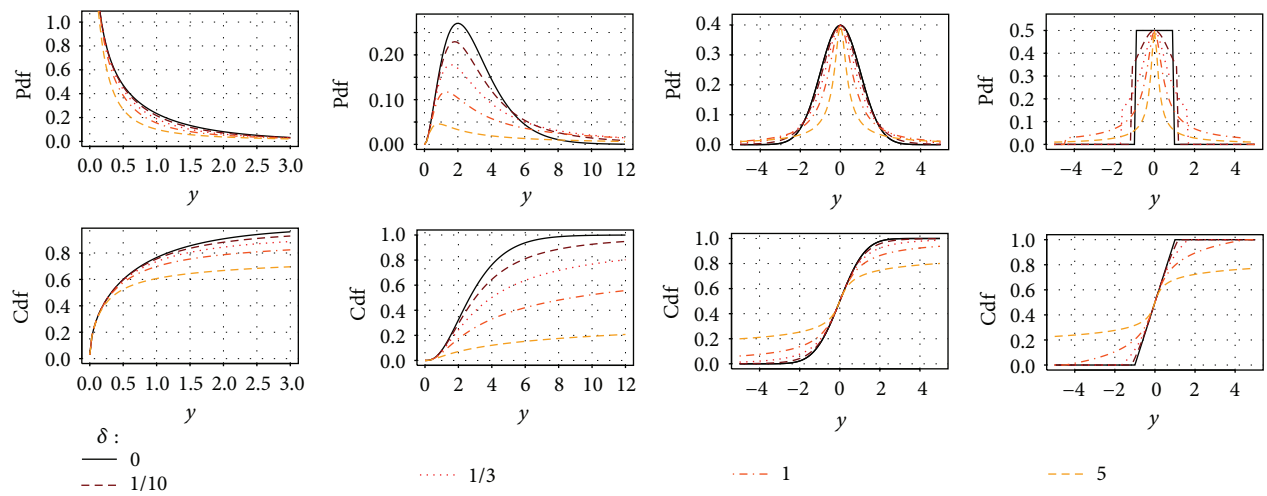
Tukey's h distribution results for U being a standard Gaussian $\mathcal{N}(0,1)$.

Definition 2. For a continuous location-scale family random variable $X \sim F_X(x \mid \boldsymbol{\beta})$ define a *location-scale heavy-tailed Lambert* W $\times\ F_X$ random variable

$$Y = \left\{U \exp\left(\frac{\delta}{2}U^2\right)\right\}\sigma_X + \mu_X, \quad \delta \in \mathbb{R}, \tag{6}$$

with parameter vector $\theta = (\boldsymbol{\beta}, \delta)$, where $U = (X - \mu_X)/\sigma_X$ and μ_X and σ_X are mean and standard deviation of X, respectively.

The input is not necessarily Gaussian (Tukey's h) but can be any other location-scale continuous random variable, for example, from a uniform distribution, $X \sim U(a,b)$ (see Figure 4).



(a) Lambert W× χ^2_k with $\boldsymbol{\beta}=k=1$ (b) Lambert W × $\Gamma(s,r)$ with $\boldsymbol{\beta}=(s,r)=(3,1)$ (c) Lambert W × $\mathcal{N}(\mu,\sigma^2)$ with $\boldsymbol{\beta}=(\mu,\sigma)=(0,1)$ (d) Lambert W × $U(a,b)$ with $\boldsymbol{\beta}=(a,b)=(-1,1)$

FIGURE 4: Pdf (left) and cdf (right) of a heavy tail (a) "noncentral, nonscaled," (b) "scale," and (c and d) "location-scale" Lambert W × F_X random variable Y for various degrees of heavy tails (color, dashed lines).

Definition 3. Let $X \sim F_X(x/s \mid \boldsymbol{\beta})$ be a continuous scale-family random variable, with scale parameter s and standard deviation σ_X; let $U = X/\sigma_X$. Then,

$$Y = X \exp\left(\frac{\delta}{2}U^2\right), \quad \delta \in \mathbb{R}, \tag{7}$$

is a scaled heavy-tailed Lambert W × F_X random variable with parameter $\theta = (\boldsymbol{\beta}, \delta)$.

Let $\tau := (\mu_X(\boldsymbol{\beta}), \sigma_X(\boldsymbol{\beta}), \delta)$ define transformation (6). (For noncentral, nonscale input set $\tau = (0, 1, \delta)$; for scale-family input $\tau = (0, \sigma_X, \delta)$.) The shape parameter δ(= Tukey's h) governs the tail behavior of Y: for $\delta > 0$ values further away from μ_X are increasingly emphasized, leading to a heavy-tailed version of X; for $\delta = 0$, $Y \equiv X$, and for $\delta < 0$ values far away from the mean are mapped back again closer to μ_X. For $\delta \geq 0$ and $X \in (-\infty, \infty)$, also $Y \in (-\infty, \infty)$. For $\delta \geq 0$ and $X \in [0, \infty)$, also $Y \in [0, \infty)$.

Remark 4 (only nonnegative δ). Although $\delta < 0$ gives interesting properties for Y, it defines a nonbijective transformation and leads to parameter-dependent support and nonunique input. Thus for the remainder of this work assume $\delta \geq 0$, unless stated otherwise.

2.3. Inverse Transformation: "Gaussianize" Heavy-Tailed Data. Transformation (6) is bijective and its inverse can be obtained via the Lambert W function, which is the inverse of $z = u\exp(u)$, that is, that function which satisfies $W(z)\exp(W(z)) = z$. It has been studied extensively in mathematics, physics, and other areas of science [29–31] and is implemented in the GNU Scientific Library (GSL) [32]. Only recently the Lambert W function received attention in the statistics literature [16, 33–35]. It has many useful properties (see Appendix A in the supplementary material and Corless et al. [29]), in particular, $W(z)$ is bijective for $z \geq 0$.

Lemma 5. *The inverse transformation of (6) is*

$$W_\tau(Y) := W_\delta\left(\frac{Y-\mu_X}{\sigma_X}\right)\sigma_X + \mu_X = U\sigma_X + \mu_X = X, \tag{8}$$

where

$$W_\delta(z) := \operatorname{sgn}(z)\left(\frac{W\left(\delta z^2\right)}{\delta}\right)^{1/2}, \tag{9}$$

and $\operatorname{sgn}(z)$ *is the sign of* z. $W_\delta(z)$ *is bijective for all* $\delta \geq 0$ *and all* $z \in \mathbb{R}$.

Lemma 5 gives for the first time an analytic, bijective inverse of Tukey's h transformation: $H_\delta^{-1}(y)$ of Morgenthaler and Tukey [26] is now analytically available as (8). Bijectivity implies that for any data $\mathbf{y}$ and parameter τ, the exact input $\mathbf{x}_\tau = W_\tau(\mathbf{y}) \sim F_X(x)$ can be obtained.

In view of the importance and popularity of Normality, we clearly want to back-transform heavy-tailed data to data from a Gaussian rather than yet another heavy-tailed distribution. Tail behavior of random variables is typically compared by their kurtosis $\gamma_2(X) = \mathbb{E}(X-\mu_X)^4/\sigma_X^4$, which equals 3 if X is Gaussian. Hence for the future when we "normalize $\mathbf{y}$" we cannot only center and scale but also transform it to $\mathbf{x}_\tau$ with $\widehat{\gamma}_2(\mathbf{x}_\tau) = 3$ (see Figure 2(c)). While $\gamma_2(X) = 3$ does not guarantee that X is Gaussian, it is a good baseline for a Gaussian sample. Furthermore, it puts different data not only on the same scale, but also on the same *tail*.

This data-driven view of the Lambert W framework can also be useful for kernel density estimation (KDE), where multivariate data is often prescaled to unit variance, so the same bandwidth can be used in each dimension [36, 37]. Thus

"normalizing" the Lambert Way can also improve KDE for heavy-tailed data (see also [38, 39]).

Remark 6 (generalized transformation). Transformation (2) can be generalized to

$$Z = U \exp\left(\frac{\delta}{2}\left(U^2\right)^{\alpha}\right), \quad \alpha > 0. \tag{10}$$

The inner term U^2 guarantees bijectivity for all $\alpha > 0$. The inverse is

$$W_{\delta,\alpha}(z) := \operatorname{sgn}(z)\left(\frac{W\left(\alpha\delta\left(z^2\right)^{\alpha}\right)}{\alpha\delta}\right)^{1/2\alpha}. \tag{11}$$

For comparison with Tukey's h I consider $\alpha = 1$ only. For $\alpha = 1/2$ transformation (10) is closely related to skewed Lambert W $\times$ F_X distributions.

2.4. Distribution and Density. For ease of notation let

$$z = \frac{y-\mu_X}{\sigma_X}, \qquad u = W_\delta(z), \qquad x = W_\tau(y) = u\sigma_X + \mu_X. \tag{12}$$

Theorem 7. *The cdf and pdf of a location-scale heavy tail Lambert W $\times$ F_X random variable Y equal*

$$G_Y(y \mid \boldsymbol{\beta}, \delta) = F_X\left(W_\delta\left(\frac{y-\mu_x}{\sigma_x}\right)\sigma_X + \mu_X \mid \boldsymbol{\beta}\right), \tag{13}$$

$$\begin{aligned} g_Y(y \mid \boldsymbol{\beta}, \delta) &= f_X\left(W_\delta\left(\frac{y-\mu_X}{\sigma_X}\right)\sigma_X + \mu_X \mid \boldsymbol{\beta}\right) \\ &\quad \cdot \frac{W_\delta\left((y-\mu_X)/\sigma_X\right)}{\left((y-\mu_X)/\sigma_X\right)\left[1 + W\left(\delta\left((y-\mu_X)/\sigma_X\right)^2\right)\right]}. \end{aligned} \tag{14}$$

Clearly, $G_Y(y \mid \boldsymbol{\beta}, \delta = 0) = F_X(y \mid \boldsymbol{\beta})$ and $g_Y(y \mid \boldsymbol{\beta}, \delta = 0) = f_X(y \mid \boldsymbol{\beta})$, since $\lim_{\delta\to 0} W_\delta(z) = z$ and $\lim_{\delta\to 0} W(\delta z^2) = 0$ for all $z \in \mathbb{R}$.

For scale family or noncentral, nonscale input set $\mu_X = 0$ or $\mu_X = 0$, $\sigma_X = 1$.

The explicit formula (14) allows a fast computation and theoretical analysis of the likelihood, which is essential for statistical inference. Detailed properties of (14) are given in Section 4.1.

Figure 4 shows (13) and (14) for various $\delta \geq 0$ with four different input distributions: for $\delta = h = 0$ the input equals the output (solid black); for larger δ the tails of $G_Y(y \mid \theta)$ and $g_Y(y \mid \theta)$ get heavier (dashed colored).

2.5. Quantile Function. Quantile fitting has been the standard technique to estimate μ_X, σ_X, and δ of Tukey's h. In particular, the medians of Y and X are equal. Thus for symmetric, location-scale family input the sample median of $\mathbf{y}$ is a robust estimate of μ_X for any $\delta \geq 0$ (see also Section 5). General quantiles can be computed via [17]

$$y_\alpha = u_\alpha \exp\left(\frac{\delta}{2}u_\alpha^2\right)\sigma_X + \mu_X, \tag{15}$$

where $u_\alpha = W_\delta(z_\alpha)$ are the α-quantiles of $F_U(u)$.

3. Tukey's h Distribution: Gaussian Input

For Gaussian input Lambert W $\times$ F_X equals Tukey's h, which has been studied extensively. Dutta and Babbel [40] show that

$$\mathbb{E}Z^n = \begin{cases} 0, & \text{if } n \text{ is odd}, n < \dfrac{1}{\delta}, \\ \dfrac{n!\,(1-n\delta)^{-(n+1)/2}}{2^{n/2}\,(n/2)!}, & \text{if } n \text{ is even}, n < \dfrac{1}{\delta}, \\ \nexists, & \text{if } n \text{ is odd}, n > \dfrac{1}{\delta}, \\ \infty, & \text{if } n \text{ is even}, n > \dfrac{1}{\delta}, \end{cases} \tag{16}$$

which, in particular, implies that [28]

$$\mathbb{E}Z = \mathbb{E}Z^3 = 0, \quad \text{if } \delta < 1 \text{ and } \frac{1}{3}, \text{ respectively}, \tag{17}$$

$$\begin{aligned} \mathbb{E}Z^2 &= \frac{1}{(1-2\delta)^{3/2}}, \quad \text{if } \delta < \frac{1}{2}, \\ \mathbb{E}Z^4 &= 3\frac{1}{(1-4\delta)^{5/2}}, \quad \text{if } \delta < \frac{1}{4}. \end{aligned} \tag{18}$$

Thus the kurtosis of Y equals (see Figure 5)

$$\gamma_2(\delta) = 3\frac{(1-2\delta)^3}{(1-4\delta)^{5/2}} \quad \text{for } \delta < 1/4. \tag{19}$$

For $\delta = 0$, (18) and (19) reduce to the familiar Gaussian values.

Expanding (19) around $\delta = 0$ yields

$$\gamma_2(\delta) = 3 + 12\delta + 66\delta^2 + \mathcal{O}\left(\delta^3\right). \tag{20}$$

Dropping $\mathcal{O}(\delta^3)$ and solving (20) gives a rule of thumb estimate

$$\widehat{\delta}_{\text{Taylor}} = \frac{1}{66}\left[\sqrt{66\,\widehat{\gamma}_2(\mathbf{y}) - 162} - 6\right]_+, \tag{21}$$

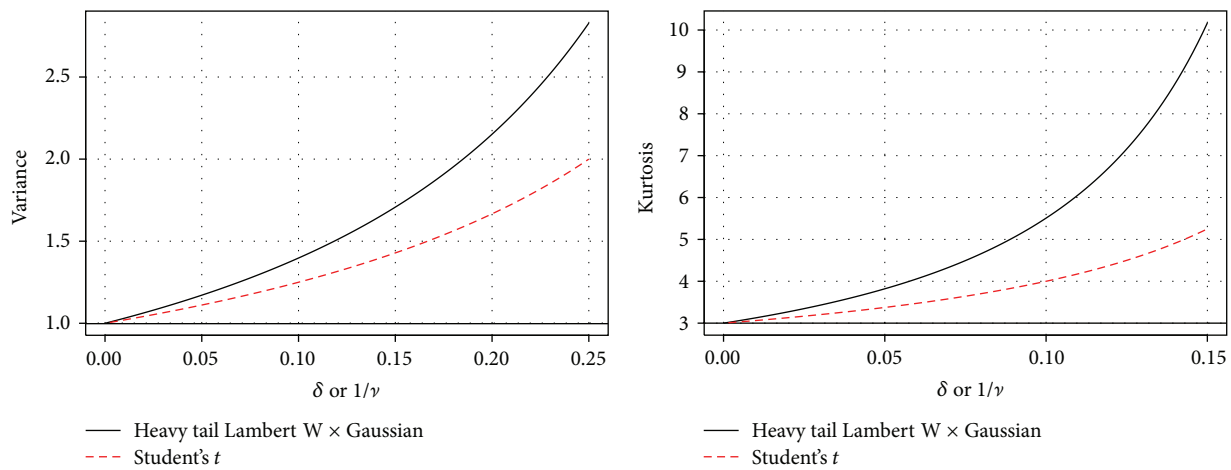
where $\widehat{\gamma}_2(\mathbf{y})$ is the sample kurtosis and $[a]_+ = \max(a, 0)$; that is, $\widehat{\delta}_{\text{Taylor}} > 0$ if $\widehat{\gamma}_2(\mathbf{y}) > 3$; otherwise, set $\widehat{\delta}_{\text{Taylor}} = 0$.

Corollary 8. *The cdf of Tukey's h equals*

$$G_Y(y \mid \mu_X, \sigma_X, \delta) = \Phi\left(\frac{W_\tau(y) - \mu_X}{\sigma_X}\right), \tag{22}$$

where $\Phi(u)$ is the cdf of a standard Normal. The pdf equals (for $\delta > 0$)

$$\begin{aligned} g_Y(y \mid \mu_X, \sigma_X, \delta) &= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1+\delta}{2}W_\delta\left(\frac{y-\mu_X}{\sigma_X}\right)^2\right) \\ &\quad \cdot \frac{1}{1 + W\left(\delta\left((y-\mu_X)/\sigma_X\right)^2\right)}. \end{aligned} \tag{23}$$



(a) Variance (b) Kurtosis

Figure 5: Comparing moments of Lambert W × Gaussian and Student's t.

Proof. Take $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ in Theorem 7. □

Section 4.1 studies functional properties of (23) in more detail.

3.1. Tukey's h versus Student's t. Student's t_ν-distribution with ν degrees of freedom is often used to model heavy-tailed data [41, 42], as its tail index equals ν. Thus the nth moment of a Student's t random variable T exists if $n < \nu$. In particular,

$$\begin{aligned} \mathbb{E}T &= \mathbb{E}T^3 = 0 \quad \text{if } \nu > 1 \text{ or } > 3, \\ \mathbb{E}T^2 &= \frac{\nu}{\nu - 2} = \frac{1}{1 - (2/\nu)} \quad \text{if } \nu > 2, \end{aligned} \tag{24}$$

and kurtosis

$$\gamma_2(\nu) = 3\frac{\nu - 2}{\nu - 4} = 3\frac{1 - 2(1/\nu)}{1 - 4(1/\nu)} \quad \text{if } \nu > 4. \tag{25}$$

Comparing (24) and (25) with (18) and (19) shows a natural association between $1/\nu$ and δ and a close similarity between the first four moments of Student's t and Tukey's h (Figure 5). By continuity and monotonicity, the first four moments of a location-scale t-distribution can always be exactly matched by a corresponding location-scale Lambert W × Gaussian. Thus if Student's t is used to model heavy tails and not as the true distribution of a test statistic it might be worthwhile to also fit heavy tail Lambert W × Gaussian distributions for an equally valuable "second opinion." For example, a parallel analysis on S&P 500 log-returns in Section 6.2 leads to divergent inference regarding the existence of fourth moments.

4. Parameter Estimation

Due to the lack of a closed form pdf of Y, $\theta = (\boldsymbol{\beta}, \delta)$ has typically been estimated by matching quantiles or a method of moments estimator [4, 26, 28]. These methods can now be replaced by the, fast and usually efficient, maximum likelihood estimator (MLE). Rayner and MacGillivray [43] introduce a numerical MLE procedure based on quantile functions, but they conclude that "sample sizes significantly larger than 100 should be used to obtain reliable estimates." Simulations in Section 5 show that the MLE using the closed form Lambert W × F_X distribution converges quickly and is accurate even for sample sizes as small as $N = 10$.

4.1. Maximum Likelihood Estimation (MLE). For an i.i.d. sample $(y_1, \ldots, y_N) = \mathbf{y} \sim g_Y(y \mid \boldsymbol{\beta}, \delta)$ the log-likelihood function equals

$$\ell(\theta; \mathbf{y}) = \sum_{i=1}^{N} \log g_Y(y_i \mid \boldsymbol{\beta}, \delta). \tag{26}$$

The MLE is that $\theta = (\boldsymbol{\beta}, \delta)$ which maximizes (26); that is,

$$\widehat{\theta}_{\text{MLE}} = \left(\widehat{\boldsymbol{\beta}}, \widehat{\delta}\right)_{\text{MLE}} = \arg\max_{\boldsymbol{\beta}, \delta} \ell(\boldsymbol{\beta}, \delta; \mathbf{y}). \tag{27}$$

Since $g_Y(y_i \mid \boldsymbol{\beta}, \delta)$ is a function of $f_X(x_i \mid \boldsymbol{\beta})$, the MLE depends on the specification of the input density. Equation (26) can be decomposed as

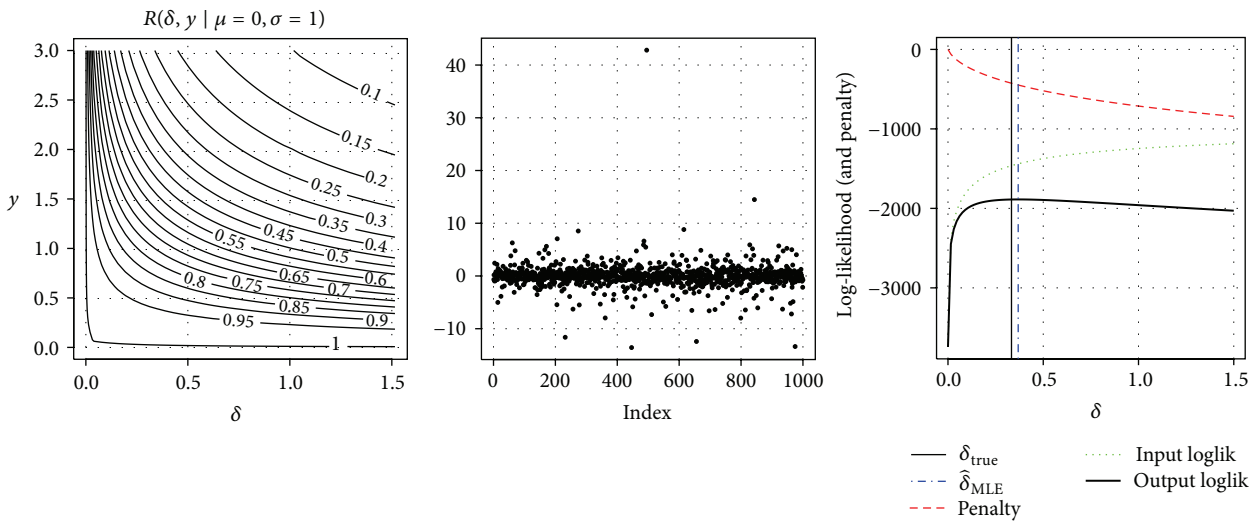
$$\ell(\boldsymbol{\beta}, \delta; \mathbf{y}) = \ell(\boldsymbol{\beta}; \mathbf{x}_\tau) + \mathcal{R}(\tau; \mathbf{y}), \tag{28}$$

where

$$\begin{aligned} \ell(\boldsymbol{\beta}; \mathbf{x}_\tau) &= \sum_{i=1}^{N} \log f_X\left(W_\delta\left(\frac{y_i - \mu_X}{\sigma_X}\right)\sigma_X + \mu_X \mid \boldsymbol{\beta}\right) \\ &= \sum_{i=1}^{N} \log f_X(x_{\tau,i} \mid \boldsymbol{\beta}) \end{aligned} \tag{29}$$

is the log-likelihood of the back-transformed data $\mathbf{x}_\tau = (x_{\tau,1}, \ldots, x_{\tau,N})$ (via (8)) and

$$\mathcal{R}(\tau; \mathbf{y}) = \sum_{i=1}^{n} \log R(\mu_X, \sigma_X, \delta; y_i), \tag{30}$$



(a) Penalty $R(\mu_X, \sigma_X, \delta; y_i)$(31) as a function of δ and y ($\mu_X = 0$ and $\sigma_X = 1$)

(b) Sample $\mathbf{z}$ of Lambert W × Gaussian with $\delta = 1/3$ (left); log-likelihood $\ell(\theta; \mathbf{y})$ (solid, black) decomposes in input log-likelihood (dotted, green) and penalty (dashed, red)

Figure 6: Log-likelihood decomposition for Lambert W × F_X distributions.

where

$$R(\mu_X, \sigma_X, \delta; y_i) = \frac{W_\delta((y_i - \mu_X)/\sigma_X)}{((y_i - \mu_X)/\sigma_X)\left[1 + \delta\left(W_\delta((y_i - \mu_X)/\sigma_X)\right)^2\right]}. \quad (31)$$

Note that $R(\mu_X, \sigma_X, \delta; y_i)$ only depends on $\mu_X(\boldsymbol{\beta})$ and $\sigma_X(\boldsymbol{\beta})$ (and δ), but not necessarily on every coordinate of $\boldsymbol{\beta}$.

Decomposition (28) shows the difference between the exact MLE $(\widehat{\boldsymbol{\beta}}, \widehat{\delta})$ based on $\mathbf{y}$ and the approximate MLE $\widehat{\boldsymbol{\beta}}_{\mathbf{x}_\tau}$ based on $\mathbf{x}_\tau$ alone: if we knew $\tau = (\mu_X, \sigma_X, \delta)$ beforehand, then we could back-transform $\mathbf{y}$ to $\mathbf{x}_\tau$ and estimate $\widehat{\boldsymbol{\beta}}_{\mathbf{x}_\tau}$ from $\mathbf{x}_\tau$ (maximize (29) with respect to $\boldsymbol{\beta}$). In practice, however, τ must also be estimated and this enters the likelihood via the additive term $\mathscr{R}(\tau; \mathbf{y})$. A little calculation shows that for any $y_i \in \mathbb{R}$, $\log R(\mu_X, \sigma_X, \delta; y_i) \leq 0$ if $\delta \geq 0$, with equality if and only if $\delta = 0$. Thus $\mathscr{R}(\tau; \mathbf{y})$ can be interpreted as a penalty for transforming the data. Maximizing (28) faces a trade-off between transforming the data to follow $f_X(x \mid \boldsymbol{\beta})$ (and increasing $\ell(\boldsymbol{\beta}; \mathbf{x}_{\widehat{\tau}})$) and the penalty of a more extreme transformation (and decreasing $\mathscr{R}(\tau; \mathbf{y})$); see Figure 6(b).

Figure 6(a) shows a contour plot of $R(\mu_X = 0, \sigma_X = 1, \delta; y)$ as a function of δ and $y = z$: it increases (in absolute value) either if δ gets larger (for fixed y) or for larger y (for fixed δ). In both cases, increasing δ makes the transformed data $W_\delta(z)$ get closer to $0 = \mu_X$, which in turn increases its input likelihood. For $\delta = 0$, the penalty disappears since input equals output; for $y = 0$ there is no penalty since $W_\delta(0) = 0$ for all δ.

Figure 6(b) shows an i.i.d. sample ($N = 1000$) $\mathbf{z} \sim$ Lambert W × Gaussian with $\delta = 1/3$ and the decomposition of the log-likelihood as in (28). Since $\boldsymbol{\beta} = (0, 1)$ is known, the likelihood and penalty are only functions of δ. Theorem 9 shows that the convexity of the penalty (decreasing, red) and concavity of the input likelihood (increasing, green) as a function of δ holds true in general for any data $\mathbf{z}$, and their sum (solid, black) has a unique maximum; here $\widehat{\delta}_{\text{MLE}} = 0.37$ (blue, dashed vertical line). See Theorem 9 below for details.

The maximization of (28) can be carried out numerically. Here I show existence and uniqueness of $\widehat{\delta}_{\text{MLE}}$ assuming that μ_X and σ_X are known. Further theoretical results for $\widehat{\theta}_{\text{MLE}}$ remain for future work. Given the "nice" form of $g_Y(y)$, which is continuous, twice differentiable (under the assumption that $fx(\cdot)$ is twice differentiable), the MLE for $\theta = (\boldsymbol{\beta}, \delta)$ should have the usual optimality properties, such as being consistent and efficient [44].

4.1.1. Properties of the MLE for the Heavy Tail Parameter δ. Without loss of generality let $\mu_X = 0$ and $\sigma_X = 1$. In this case

$$\ell(\delta; \mathbf{z}) \propto -\frac{1}{2}\sum_{i=1}^{N}\left[W_\delta(z_i)\right]^2 + \sum_{i=1}^{N}\log\frac{W_\delta(z_i)}{z_i} - \sum_{i=1}^{N}\log\left(1 + \delta\left[W_\delta(z_i)\right]^2\right) \quad (32)$$

$$= -\frac{1+\delta}{2}\sum_{i=1}^{N}\left[W_\delta(z_i)\right]^2 - \sum_{i=1}^{N}\log\left(1 + \delta\left[W_\delta(z_i)\right]^2\right). \quad (33)$$

Theorem 9 (see [14]). *Let Z have a Lambert W × Gaussian distribution, where $\mu_X = 0$ and $\sigma_X = 1$ are assumed to be known and fixed. Also consider only $\delta \in [0, \infty)$. (While for some samples $\mathbf{z}$ the MLE also exists for $\delta < 0$, it cannot be guaranteed for all $\mathbf{z}$. If $\delta < 0$ (and $z \neq 0$), then $W_\delta(z)$ is either*

not unique in $\mathbb{R}$ (principal and nonprincipal branch) or does not have a real-valued solution in $\mathbb{R}$ if $\delta z^2 < e^{-1}$.)

(a) *If*

$$\frac{\sum_{i=1}^{n} z_i^4}{\sum_{i=1}^{n} z_i^2} \leq 3, \tag{34}$$

then $\widehat{\delta}_{MLE} = 0$.

If (34) does not hold, then

(b) *$\widehat{\delta}_{MLE} > 0$ exists and is a positive solution to*

$$\sum_{i=1}^{N} z_i^2 W'\left(\delta z_i^2\right)\left(\frac{1}{2}W_\delta\left(z_i\right)^2 - \left(\frac{1}{2} + \frac{1}{1+W\left(\delta z_i^2\right)}\right)\right) = 0. \tag{35}$$

(c) *There is only one such δ satisfying (35); that is, $\widehat{\delta}_{MLE}$ is unique.*

Proof. See Appendix B in the supplementary material. □

Condition (34) says that $\widehat{\delta}_{\text{MLE}} > 0$ only if the data is heavy-tailed enough. Points (b) and (c) guarantee that there is no ambiguity in the heavy tail estimate. This is an advantage over Student's t-distribution, for example, which has numerical problems and local maxima for unknown (and small) ν [45, 46]. On the contrary, $\widehat{\delta}_{\text{MLE}}$ is always a global maximum.

Given the heavy tails in $\mathbf{z}$ one might expect convergence issues for larger δ. However, $W_\delta(Z) \sim \mathcal{N}(0,1)$ for the true $\delta \geq 0$ and close to a standard Gaussian if $\widehat{\delta}_{\text{MLE}} \approx \delta$. Since the log-likelihood and its gradient depend on δ and $\mathbf{z}$ only via $W_\delta(\mathbf{z})$ the performance of the MLE should thus not get worse for large δ as long as the initial estimate is close enough to the truth. Simulations in Section 5 support this conjecture, even for $\widehat{\theta}_{\text{MLE}}$.

4.2. Iterative Generalized Method of Moments (IGMM). A disadvantage of the MLE is the mandatory a priori specification of the input distribution. Especially for heavy-tailed data the eye is a bad judge to choose a particular parametric $f_X(x \mid \boldsymbol{\beta})$. It would be useful to directly estimate τ without the intermediate step of estimating θ first.

Goerg [16] presented an estimator for τ based on iterative generalized methods of moments (IGMM). The idea of IGMM is to find a τ such that the back-transformed data $\mathbf{x}_\tau$ has desired properties, for example, being symmetric or having kurtosis 3. An estimator for μ_X, σ_X, and δ can be constructed entirely analogous to the IGMM estimator for the skewed Lambert W $\times$ F case. See the Supplementary Material, Appendix C for details.

IGMM requires less specific knowledge about the input distribution and is usually also faster than the MLE. Once $\widehat{\tau}_{\text{IGMM}}$ has been obtained, the back-transformed $\mathbf{x}_{\widehat{\tau}_{\text{IGMM}}}$ can be used to check if X has characteristics of a known parametric distribution $F_X(x \mid \boldsymbol{\beta})$. It must be noted though that testing for a particular distribution F_X is too optimistic as $\mathbf{x}_{\widehat{\tau}}$ has "nicer" properties regarding F_X than the true $\mathbf{x}$ would have. However, estimating the transformation requires only three parameters and for a large enough sample, losing three degrees of freedom should not matter in practice.

5. Simulations

This section explores finite sample properties of estimators for $\theta = (\mu_X, \sigma_X, \delta)$ and (μ_Y, σ_Y) under Gaussian input $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$. In particular, it compares Gaussian MLE (estimation of μ_Y and σ_Y only), IGMM and Lambert W $\times$ Gaussian MLE, and, for a heavy tail competitor, the median. (For IGMM, optimization was restricted to $\delta \in [0, 10]$.) All results below are based on $n = 1,000$ replications.

5.1. Estimating δ Only. Here I show finite sample properties of $\widehat{\delta}_{\text{MLE}}$ for $U \sim \mathcal{N}(0,1)$, where $\mu_X = 0$ and $\sigma_X = 1$ are known and fixed. Theorem 9 shows that $\widehat{\delta}_{\text{MLE}}$ is unique: either at the boundary $\delta = 0$ or at the globally optimal solution to (35). Results in Table 1 were obtained by numerical optimization restricted to $\delta \geq 0$ ($\Leftrightarrow \log\delta \in \mathbb{R}$) using the `nlm` function in *R*.

Table 1 suggests that the MLE is asymptotically unbiased for every δ and converges quickly (at about $N = 400$) to its asymptotic variance, which is increasing with δ. Assuming μ_X and σ_X to be known is unrealistic and thus these finite sample properties are only an indication of the behavior of the joint MLE, $\widehat{\theta}_{\text{MLE}}$. Nevertheless they are very remarkable for extremely heavy-tailed data ($\delta > 1$), where classic average-based methods typically break down. One reason lies in the particular form of the likelihood (32) and its gradient (35) (Theorem 9): although both depend on $\mathbf{z}$, they only do so through $W_\delta(\mathbf{z}) = \mathbf{u} \sim \mathcal{N}(0,1)$. Hence as long as $\widehat{\delta}_{\text{MLE}}$ is sufficiently close to the true δ, (32) and (35) are functions of almost Gaussian random variables and standard asymptotic results should still apply.

5.2. Estimating All Parameters Jointly. Here we consider the realistic scenario where μ_X and σ_X are also unknown. We consider various sample sizes ($N = 50$, 100, and 1000) and different degrees of heavy tails, $\delta \in \{0, 1/3, 1, 1.5\}$, each one representing a particularly interesting situation: (i) Gaussian data (does additional, superfluous, estimation of δ affect other estimates?), (ii) fourth moments do not exist, (iii) nonexisting mean, and (iv) extremely heavy-tailed data (can we get useful estimates at all?)

The convergence tolerance for IGMM was set to tol $= 1.22 \cdot 10^{-4}$. Table 2 summarizes the simulation.

The Gaussian MLE estimates σ_Y directly, while IGMM and the Lambert W $\times$ Gaussian MLE estimate δ and σ_X, which implicitly give $\widehat{\sigma}_Y$ through $\sigma_Y(\delta, \sigma_X) = \sigma_X \cdot (1/\sqrt{(1-2\delta)^{3/2}})$ if $\delta < 1/2$ (see (18)). For a fair comparison each subtable also includes a column for $\widehat{\sigma}_Y = \widehat{\sigma}_X \cdot (1/\sqrt{(1-2\widehat{\delta})^{3/2}})$. Some of these entries contain "∞," even for $\delta < 1/2$; this occurs if at least one $\widehat{\delta} \geq 1/2$. For any $\delta < 1$, $\mu_X = \mu_Y$, thus $\widehat{\mu}_X$ and $\widehat{\mu}_Y$ can be compared directly. For $\delta \geq 1$,

Table 1: Finite sample properties of $\widehat{\delta}_{\mathrm{MLE}}$. For each N, δ was estimated $n = 1{,}000$ times from a random sample $\mathbf{z}\sim$Tukey's h. The left column for each δ shows bias, $\overline{\widehat{\delta}}_{\mathrm{MLE}} - \delta$; each right column shows the root mean square error (RMSE) times $\sqrt{N}$.

N	$\delta = 0$		$\delta = 1/10$		$\delta = 1/3$		$\delta = 1/2$	
10	0.025	0.191	−0.017	0.394	−0.042	0.915	−0.082	1.167
50	0.013	0.187	−0.010	0.492	−0.018	0.931	−0.016	1.156
100	0.010	0.200	−0.010	0.513	−0.009	0.914	−0.006	1.225
400	0.005	0.186	−0.003	0.528	0.000	0.927	−0.004	1.211
1000	0.003	0.197	0.000	0.532	−0.001	0.928	−0.001	1.203
2000	0.003	0.217	−0.001	0.523	0.000	0.935	−0.001	1.130
N	$\delta = 1$		$\delta = 2$		$\delta = 5$			
10	−0.054	1.987	−0.104	3.384	−0.050	7.601		
50	−0.017	1.948	−0.009	3.529	0.014	7.942		
100	−0.014	2.024	−0.001	3.294	0.011	7.798		
400	0.001	1.919	−0.002	3.433	0.001	7.855		
1000	0.001	1.955	0.001	3.553	−0.001	7.409		
2000	0.001	1.896	0.000	3.508	−0.001	7.578		

the mean does not exist; each subtable for these δ interprets μ_Y as the median.

Gaussian Data ($\delta = 0$). This setting checks if imposing the Lambert W framework, even though it is superfluous, causes a quality loss in the estimation of $\mu_Y = \mu_X$ or $\sigma_Y = \sigma_X$. Furthermore, critical values for $H_0 : \delta = 0$ (Gaussian) can be obtained. As in the δ-only case above, Table 2(a) suggests that estimators are asymptotically unbiased and quickly tend to a large-sample variance. Additional estimation of δ does not affect the efficiency of $\widehat{\mu}_X$ compared to estimating solely μ. Estimating σ_Y directly by Gaussian MLE does not give better results than the Lambert W × Gaussian MLE.

No Fourth Moment ($\delta = 1/3$). Here $\sigma_Y(\delta, \sigma_X = 1) = 2.28$, but fourth moments do not exist. This results in an increasing empirical standard deviation of $\widehat{\sigma}_y$ as N grows. In contrast, estimates for σ_X are not drifting off. In the presence of these large heavy tails the median is much less variable than Gaussian MLE and IGMM. Yet, Lambert W × Gaussian MLE for μ_X even outperforms the median.

Nonexisting Mean ($\delta = 1$). Both sample moments diverge, and their standard errors are also growing quickly. The median still provides a very good estimate for the location but is again inferior to both Lambert W estimators, which are closer to the true values and appear to converge to an asymptotic variance at rate $\sqrt{N}$.

Extreme Heavy Tails ($\delta = 1.5$). As in Section 5.1, IGMM and Lambert W MLE continue to provide excellent estimates even though the data is extremely heavy tailed. Moreover, Lambert W MLE also has the smallest empirical standard deviation overall. In particular, the Lambert W MLE for μ_X has an approximately 15% lower standard deviation than the median.

The last column shows that for some N about 1% of the $n = 1{,}000$ simulations generated invalid likelihood values (NA and ∞). Here the search for the optimal δ led into regions with a numerical overflow in the evaluation of $W_\delta(z)$. For a comparable summary, these few cases were omitted and new simulations added until full $n = 1{,}000$ finite estimates were found. Since this only happened in 1% of the cases and also such heavy-tailed data is rarely encountered in practice, this numerical issue is not a limitation in statistical practice.

5.3. Discussion of the Simulations. IGMM performs well independent of the magnitude of δ. As expected the Lambert W MLE for θ has the best properties: it can recover the truth for all δ, and for $\delta = 0$ it performs as well as the classic sample mean and standard deviation. For small δ it has the same empirical standard deviation as the Gaussian MLE, but a lower one than the median for large δ.

Hence the only advantage of estimating μ_Y and σ_Y by sample moments of $\mathbf{y}$ is speed; otherwise the Lambert W × Gaussian MLE is at least as good as the Gaussian MLE and clearly outperforms it in presence of heavy tails.

6. Applications

Tukey's h distribution has already proven useful to model heavy-tailed data, but parametric inference was limited to quantile fitting or methods of moments estimation [4, 27, 28]. Theorem 7 allows us to estimate θ by ML.

This section applies the presented methodology on simulated as well as real world data: (i) Section 6.1 demonstrates Gaussianizing on the Cauchy sample from the Introduction, and (ii) Section 6.2 shows that heavy tail Lambert W × Gaussian distributions provide an excellent fit to daily S&P 500 log-return series.

6.1. Estimating Location of a Cauchy with the Sample Mean. It is well known that the sample mean $\overline{\mathbf{y}}$ is a poor estimate of the location parameter of a Cauchy distribution, since the sampling distribution of $\overline{\mathbf{y}}$ is again a Cauchy (see [47] for a recent overview); in particular, its variance does not go to 0 for $N \to \infty$.

Table 2: In each subtable: (first rows) average, (middle rows) proportion of estimates below truth, and (bottom rows) empirical standard deviation times $\sqrt{N}$. True $\mu_X = 0$ and $\sigma_X = 1$.

(a) Truly Gaussian data: $\delta = 0$

$\delta = 0$	Median	Gaussian MLE		IGMM				Lambert W MLE				NA
N		μ_Y	σ_Y	μ_X	σ_X	δ	σ_Y	μ_X	σ_X	δ	σ_Y	Ratio
50	0.00	0.00	0.98	0.00	0.97	0.02	0.99	0.00	0.96	0.02	0.98	0
100	0.00	0.00	0.99	0.00	0.98	0.01	1.00	0.00	0.97	0.01	0.99	0
1000	0.00	0.00	1.00	0.00	0.99	0.00	1.00	0.00	0.99	0.00	1.00	0
50	0.50	0.50	0.57	0.51	0.60	0.66	0.54	0.51	0.65	0.66	0.56	0
100	0.50	0.51	0.56	0.51	0.62	0.62	0.53	0.52	0.65	0.62	0.56	0
1000	0.50	0.49	0.52	0.49	0.62	0.56	0.52	0.49	0.63	0.56	0.52	0
50	1.24	1.01	0.72	1.01	0.76	0.21	0.73	1.02	0.78	0.26	0.72	0
100	1.25	1.02	0.70	1.02	0.76	0.23	0.70	1.03	0.78	0.26	0.70	0
1000	1.26	0.98	0.73	0.98	0.79	0.22	0.73	0.98	0.79	0.22	0.73	0

(b) No fourth moments: $\delta = 1/3$

$\delta = 1/3$	Median	Gaussian MLE		IGMM				Lambert W MLE				NA
N		μ_Y	σ_Y	μ_X	σ_X	δ	σ_Y	μ_X	σ_X	δ	σ_Y	Ratio
50	0.00	0.00	1.98	0.00	1.07	0.29	∞	0.00	1.01	0.33	∞	0
100	0.00	0.00	2.03	0.00	1.04	0.31	∞	0.00	1.00	0.33	∞	0
1000	0.00	0.00	2.18	0.00	1.00	0.33	2.34	0.00	1.00	0.33	2.34	0
50	0.50	0.51	0.78	0.50	0.38	0.63	0.60	0.50	0.52	0.54	0.54	0
100	0.50	0.51	0.78	0.51	0.42	0.61	0.60	0.50	0.51	0.54	0.54	0
1000	0.48	0.51	0.77	0.51	0.47	0.56	0.55	0.51	0.50	0.53	0.52	0
50	1.27	2.21	6.56	1.44	1.45	1.10	NA	1.23	1.35	1.14	NA	0
100	1.30	2.33	11.28	1.43	1.42	1.12	NA	1.19	1.34	1.09	NA	0
1000	1.23	2.25	16.76	1.39	1.45	1.20	15.97	1.17	1.33	1.08	12.30	0

(c) Non-existing mean: $\delta = 1$

$\delta = 1$	Median	Gaussian MLE		IGMM				Lambert W MLE				NA
N		μ_Y	σ_Y	μ_X	σ_X	δ	σ_Y	μ_X	σ_X	δ	σ_Y	Ratio
50	0.00	−0.10	24.6	−0.01	1.18	0.90	∞	0.00	1.01	0.99	∞	0
100	0.00	0.74	72.4	0.00	1.09	0.95	∞	0.00	1.01	0.99	∞	0
1000	0.00	3.84	348.1	0.00	1.01	1.00	∞	0.00	1.00	1.00	∞	0
50	0.53	0.52	1.0	0.51	0.34	0.65	1	0.51	0.52	0.52	1	0
100	0.50	0.52	1.0	0.51	0.38	0.63	1	0.50	0.53	0.53	1	0
1000	0.49	0.52	1.0	0.51	0.48	0.53	1	0.49	0.51	0.51	1	0
50	1.27	65.85	424.3	2.10	2.50	2.32	NA	1.19	1.70	2.16	NA	0
100	1.30	410.75	4050.2	2.01	2.28	2.59	NA	1.17	1.74	2.25	NA	0
1000	1.26	3307.58	104052.7	1.93	2.21	2.81	NA	1.11	1.64	2.18	NA	0

(d) Extreme heavy tails: $\delta = 1.5$

$\delta = 1.5$	Median	Gaussian MLE		IGMM				Lambert W MLE				NA
N		μ_Y	σ_Y	μ_X	σ_X	δ	σ_Y	μ_X	σ_X	δ	σ_Y	Ratio
50	−0.02	6.84	309	−0.02	1.23	1.37	∞	−0.01	1.00	1.49	∞	0.01
100	0.00	−51.16	3080	−0.01	1.12	1.44	∞	0.00	1.01	1.50	∞	0.00
1000	0.00	176.13	14251	0.00	1.01	1.49	∞	0.00	1.00	1.50	∞	0.00
50	0.53	0.48	1	0.51	0.34	0.64	1	0.53	0.53	0.54	1	0.01
100	0.51	0.53	1	0.54	0.37	0.61	1	0.52	0.51	0.51	1	0.00
1000	0.50	0.50	1	0.50	0.47	0.54	1	0.49	0.53	0.52	1	0.00
50	1.32	1347.71	9261	2.57	3.20	3.12	NA	1.15	1.86	2.76	NA	0.01
100	1.33	42156.28	418435	2.39	2.87	3.44	NA	1.12	1.78	2.84	NA	0.00
1000	1.26	124462.82	3903629	2.18	2.66	3.67	NA	1.11	1.80	2.85	NA	0.00

Table 3: Summary statistics for observed (heavy-tailed) $\mathbf{y}$ and back-transformed (Gaussianized) data $\mathbf{x}_{\widehat{\tau}_{\text{MLE}}}$. $**$ stands for $<10^{-16}$ and $*$ for $< 2.2 \cdot 10^{-16}$.

	$\mathbf{y} \sim \mathscr{C}(0,1)$ (Section 6.1)			$\mathbf{y}$ = S&P 500 (Section 6.2)	
	$\mathbf{y}$	$\mathbf{x}_{\widehat{\tau}}$	$\mathbf{x}_{\widehat{\lambda}}$	$\mathbf{y}$	$\mathbf{x}_{\widehat{\tau}}$
Min	−161.59	−3.16	0	−7.11	−2.42
Max	952.95	3.81	33.18	4.99	2.23
Mean	2.30	0.03	14.98	0.05	0.05
Median	0.04	0.04	14.96	0.04	0.04
Standard Deviation	46.980	1.06	1.20	0.95	0.71
Skewness	17.43	0.12	3.90	−0.30	−0.04
Kurtosis	343.34	3.21	161.75	7.70	2.93
Shapiro-Wilk	$*$	0.71	$**$	$*$	0.24
Anderson-Darling	$**$	0.51	$**$	$*$	0.18

Heavy-tailed Lambert W × Gaussian distributions have similar properties to a Cauchy for $\delta \approx 1$. The mean of X (μ_X) equals the location of Y (c) due to symmetry around μ_X and c, respectively. Thus we can estimate τ from the Cauchy sample $\mathbf{y}$, transform $\mathbf{y}$ to $\mathbf{x}_{\widehat{\tau}}$, get $\widehat{\mu}_X$ from $\mathbf{x}_{\widehat{\tau}} = W_{\widehat{\tau}}(\mathbf{y})$, and thus obtain an estimate for c by $\widehat{c} = \widehat{\mu}_X$.

The random sample $\mathbf{y} \sim \mathscr{C}(0,1)$, with pdf $f(y) = 1/\pi(1+y^2)$, in Figure 2(a) has heavy tails with two extreme (positive) observations. A Cauchy ML fit gives $\widehat{c} = 0.03(0.055)$ and $\widehat{s} = 0.86(0.053)$ (standard errors in parenthesis). A Lambert W × Gaussian MLE gives $\widehat{\mu}_X = 0.03(0.055)$, $\widehat{\sigma}_X = 1.05(0.072)$, and $\widehat{\delta} = 0.86(0.082)$. Thus both fits correctly fail to reject $\mu_X = c = 0$. Table 3(a) shows summary statistics on both samples. Since the Cauchy distribution does not have a well-defined mean, $\overline{\mathbf{y}} = 2.304(2.101)$ is not meaningful. However, $\mathbf{x}_{\widehat{\tau}_{\text{MLE}}}$ is approximately Gaussian and we can use the sample average for inference: $\overline{\mathbf{x}}_{\widehat{\tau}_{\text{MLE}}} = 0.033(0.0472)$ correctly fails to reject a zero location for $\mathbf{y}$. The transformed $\mathbf{x}_{\widehat{\tau}_{\text{MLE}}}$ features additional Gaussian characteristics (symmetric, no excess kurtosis), and even the null hypothesis of Normality cannot be rejected (P-value ≥ 0.5). Note, however, that Normality for the transformed data is only an empirical approximation; the random variable $W_\tau((Y-\mu_X)/\sigma_X)$, where Y is Cauchy, does *not* have a Normal distribution.

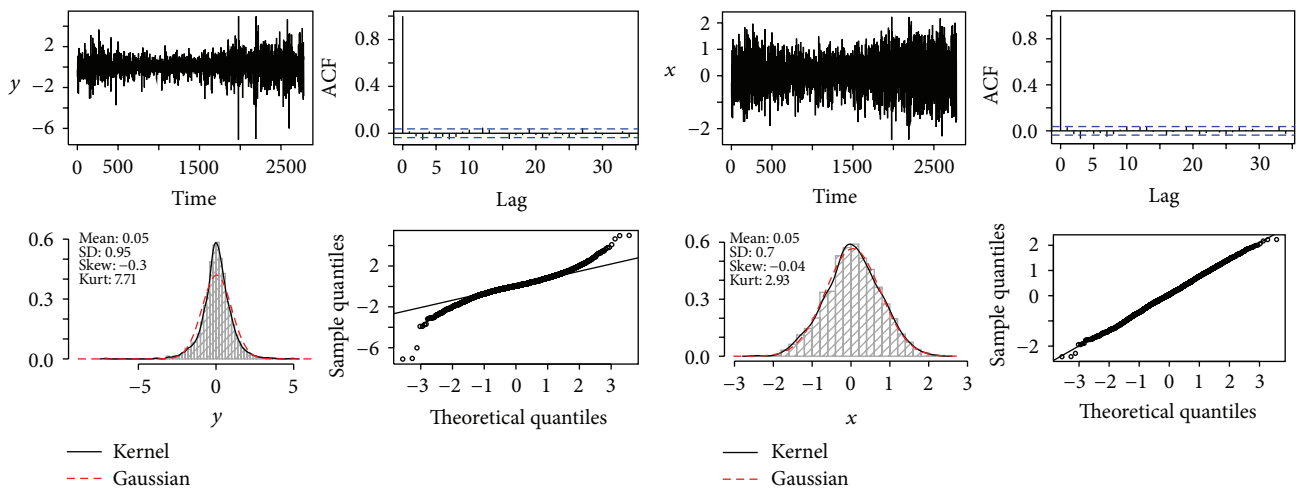
Figure 2(d) shows the cumulative sample average for the original sample and its Gaussianized version. For a fair comparison $\widehat{\tau}^{(n)}_{\text{MLE}}$ was reestimated cumulatively for each $n = 5, \ldots, 500$ and then used to compute $(x_1, \ldots, x_n)$. The transformation works extremely well: data point y_{49} is highly influential for $\overline{\mathbf{y}}$ but has no relevant effect on $\overline{\mathbf{x}}_{\widehat{\tau}^{(n)}_{\text{MLE}}}$. Even for small n it is already clear that the location of the underlying Cauchy distribution is approximately zero.

Although it is a simulated example, it demonstrates that removing (strong) heavy tails from data works well and provides "nice" data that can then be analyzed with more refined Gaussian methods.

6.2. Heavy Tails in Finance: S&P 500 Case Study. A lot of financial data displays negative skewness and excess kurtosis. Since financial data in general is not i.i.d., it is often modeled with a (skew) Student's t-distribution underlying a (generalized) autoregressive conditional heteroskedastic (GARCH) [2, 48] or a stochastic volatility (SV) [49, 50] model. Using the Lambert W approach we can build upon the knowledge and implications of Normality (and avoid deriving properties of a GARCH or SV model with heavy-tailed innovations) and simply "Gaussianize" the returns before fitting more complex, GARCH or SV, models.

Remark 10. Time series models with Lambert W × Gaussian white noise are far beyond the scope of this work but can be a direction of future research. Here I only consider the unconditional distribution.

Figure 7(a) shows the S&P 500 log-returns with a total of $N = 2,780$ daily observations (R package `MASS`, dataset `SP500`). Table 3(b) confirms the heavy tails (sample kurtosis 7.70) but also indicates negative skewness (−0.296). As the sample skewness $\widehat{\gamma}_1(\mathbf{y})$ is very sensitive to outliers, we fit a distribution which allows skewness and test for symmetry. In case of the double-tail Lambert W × Gaussian this means testing $H_0 : \delta_\ell = \delta_r = \delta$ versus $H_1 : \delta_\ell \neq \delta_r$. Using the likelihood expression in (28), we can use a likelihood ratio test with one degree of freedom (3 versus 4 parameters). The log-likelihood of the double-tail fit (Table 4(a)) equals $-3606.0 = -2972.27 + (-633.73)$ (input log-likelihood + penalty), while the symmetric δ fit gives $-3606.56 = -2971.47 + (-635.09)$. Here the symmetric fit gives a transformed sample that is more Gaussian, but it pays a greater penalty for transforming the data. Comparing twice their difference to a χ^2_1 distribution gives a P-value of 0.29. For comparison, a skew-t fit [51], with location c, scale s, shape α, and ν degrees of freedom, also yields (Function `st.mle` in the R package `sn`.) a nonsignificant $\widehat{\alpha}$ (Table 4(b)). Thus both fits cannot reject symmetry.



(a) Observed heavy tail returns $\mathbf{y}$

(b) Gaussianized returns $\mathbf{x}_{\widehat{\tau}_{\text{MLE}}}$

Figure 7: Lambert W Gaussianization of S&P 500 log-returns: $\widehat{\tau} = (0.05, 0.70, 0.17)$. In (a) and (b): data (top left); autocorrelation function (ACF) (top right); histogram, Gaussian fit, and KDE (bottom left); Normal QQ plot (bottom right).

Assume we have to make a decision if we should trade a certificate replicating the S&P 500. Since we can either buy or sell, it is not important if the average return is positive or negative, as long as it is significantly different from zero.

6.2.1. Gaussian Fit to Returns. Estimating (μ_Y, σ_Y) by Gaussian MLE and thus ignoring the heavy tails, $\widehat{\mu}_Y = 0$ cannot be rejected on a $\alpha = 1\%$ level (Table 4(e)). Ignoring heavy tails we would thus decide to not trade a replicating certificate at all.

6.2.2. Heavy Tail Fit to Returns. Both a heavy tail Lambert W × Gaussian (Table 4(c)) and Student t-fit (Table 4(d)) reject the zero mean null (P-values: 10^{-4} and $3 \cdot 10^{-5}$, resp.).

Location and scale estimates are almost identical, but tail estimates lead to different conclusions: while for $\widehat{\nu} = 3.71$ only moments up to order 3 exist, in the Lambert W × Gaussian case moments up to order 5 exist ($1/0.172 = 5.81$). This is especially noteworthy as many theoretical results in the (financial) time series literature rely on finite fourth moments [52, 53]; consequently many empirical studies test this assumption [7, 54]. Here Student's t and a Lambert W × Gaussian fit give different conclusions. Since previous empirical studies often use Student's t as a baseline [41], it might be worthwhile to reexamine their findings in light of heavy tail Lambert W × Gaussian distributions.

6.2.3. "Gaussianizing" Financial Returns. The back-transformed $\mathbf{x}_{\widehat{\tau}_{\text{MLE}}}$ is indistinguishable from a Gaussian sample (Figure 7(b)) and thus demonstrates that a Lambert W × Gaussian distribution is indeed appropriate for $\mathbf{x}_{\widehat{\tau}_{\text{MLE}}}$. Not even one test can reject Normality: P-values are 0.18, 0.18, 0.31, and 0.24, respectively (Anderson-Darling, Cramer-von-Mises, Shapiro-Francia, Shapiro-Wilk; see Thode [55]). Table 3(b) confirms that Lambert W "Gaussianiziation" was successful: $\widehat{\gamma}_1(\mathbf{x}_{\widehat{\tau}}) = -0.039$ and $\widehat{\gamma}_2(\mathbf{x}_{\widehat{\tau}}) = 2.93$ are within the typical variation for a Gaussian sample of size $N = 2780$. Thus

$$Y = \left(Ue^{(0.172/2)U^2}\right) 0.705 + 0.055, \quad U = \frac{X - 0.055}{0.705}, \quad U \sim \mathcal{N}(0,1) \tag{36}$$

is an adequate (unconditional) Lambert W × Gaussian model for the S&P 500 log-returns $\mathbf{y}$. For trading, this means that the expected return is significantly larger than zero ($\widehat{\mu}_X = 0.055 > 0$) and thus replicating certificates should be bought.

6.2.4. Gaussian MLE for Gaussianized Data. For $\delta_l = \delta_r \equiv \delta < 1$, also $\mu_X \equiv \mu_Y$. We can therefore replace testing $\mu_y = 0$ versus $\mu_y \neq 0$ for a non-Gaussian $\mathbf{y}$, with the very well-understood hypothesis test $\mu_x = 0$ versus $\mu_x \neq 0$ for the Gaussian $\mathbf{x}_{\widehat{\tau}_{\text{MLE}}}$. In particular, standard errors based on $\widehat{\sigma}/\sqrt{N}$ and thus t and P-values should be closer to the "truth" (Tables 4(c) and 4(d)) than a Gaussian MLE on the non-Gaussian $\mathbf{y}$ (Table 4(e)). Table 4(f) shows that standard errors for $\widehat{\mu}_{\mathbf{x}}$ are even a bit too small compared to the heavy-tailed versions. Since the "Gaussianizing" transformation was estimated, treating $\mathbf{x}_{\widehat{\tau}_{\text{MLE}}}$ as if it was original data is too optimistic regarding its Normality (recall the penalty (30) in the total likelihood (28)).

This example confirms that if a model and its theoretical properties are based on Normality, but the observed data is heavy-tailed, then Gaussianizing the data first gives more reliable inference than applying Gaussian methods to the original, heavy-tailed data (Figure 1). Clearly, a joint estimation of the model parameters based on Lambert W × Gaussian random variables (or any other heavy-tailed distribution) would be optimal. However, theoretical properties and estimation techniques may not be available or well understood. The Lambert Way to Gaussianize data is

Table 4: MLE fits to S&P 500 $\mathbf{y}$ Tables 4(a), 4(b), 4(c), 4(d), and 4(e) and the Gaussianized data $\mathbf{x}_{\hat{\tau}_{\text{MLE}}}$ Table 4(f).

(a) Double-tail Lambert W × Gaussian = Tukey's *hh* (S&P 500)

	Est.	se	t	Pr (> |t|)
μ_X	0.06	0.015	3.66	0.00
σ_X	0.71	0.016	44.00	0.00
δ_ℓ	0.19	0.021	8.99	0.00
δ_r	0.16	0.019	8.24	0.00

(b) Skew t (S&P 500)

	Est.	se	t	Pr (> |t|)
c	0.10	0.061	1.65	0.10
s	0.67	0.017	38.47	0.00
α	−0.08	0.101	−0.77	0.44
ν	3.73	0.297	12.57	0.00

(c) Lambert W × Gaussian = Tukey's h (S&P 500)

	Est.	se	t	Pr (> |t|)
μ_X	0.06	0.015	3.65	0.000
σ_X	0.71	0.016	43.95	0.000
δ	0.17	0.016	11.05	0.000

(d) Student's t (S&P 500)

	Est.	se	t	Pr (> |t|)
c	0.06	0.015	3.65	0.00
s	0.67	0.017	39.51	0.00
ν	3.72	0.295	12.61	0.00

(e) Gaussian (S&P 500)

	Est.	se	t	Pr (> |t|)
μ_Y	0.05	0.018	2.55	0.01
σ_Y	0.95	0.013	74.57	0.00

(f) Gaussian ($\mathbf{x}_{\hat{\tau}_{\text{MLE}}}$)

	Est.	se	t	Pr (> |t|)
$\mu_{x_{\hat{\tau}}}$	0.05	0.013	3.81	0.00
$\sigma_{x_{\hat{\tau}}}$	0.71	0.009	74.57	0.00

thus a pragmatic method to improve statistical inference on heavy-tailed data, while preserving the applicability and interpretation of Gaussian models.

7. Discussion and Outlook

In this work I use the Lambert W function to model and remove heavy tails from continuous random variables using a data-transformation approach. For Gaussian random variables this not only contributes to existing work on Tukey's h distribution but also gives convincing empirical results: unimodal data with heavy tails can be transformed to Gaussian data. Properties of a Gaussian model $\mathcal{M}_{\mathcal{N}}$ on the back-transformed data mimic the features of the "true" heavy-tailed model $\mathcal{M}_G$ very closely.

Since Normality is the single most typical and often required assumption in many areas of statistics, machine learning, and signal processing, future research can take many directions. From a theoretical perspective properties of Lambert W × F_X distributions viewed as a generalization of already well-known distributions F_X can be studied. This area will profit from existing literature on the Lambert W function, which has been discovered only recently by the statistics community. Empirical work can focus on transforming the data and comparing approximate Gaussian with joint heavy tail analyses. The comparisons in this work showed that approximate inference for Gaussianized data is comparable with the direct heavy tail modeling and so provides a simple tool to improve inference for heavy-tailed data in statistical practice.

I also provide the R package Lambert W, publicly available at `CRAN`, to facilitate the use of heavy tail Lambert W × F_X distributions in practice.

Disclosure

The first version of this article entitled "Closed-form cdf and pdf of Tukey's h-distribution, the heavy-tail Lambert W approach, and how to bijectively 'Gaussianize' heavy-tailed data" has appeared online [56].This research was performed while the author was at Carnegie Mellon University. Currently, Georg M. Goerg works at Google Inc., New York, NY 10011, USA.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

References

[1] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, Springer Series in Statistics, 1998.

[2] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, vol. 50, no. 4, pp. 987–1007, 1982.

[3] C. W. J. Granger and R. Joyeux, "An introduction to long-memory time series models and fractional differencing," *Journal of Time Series Analysis*, vol. 1, pp. 15–30, 2001.

[4] C. A. Field, "Using the *gh* distribution to model extreme wind speeds," *Journal of Statistical Planning and Inference*, vol. 122, no. 1-2, pp. 15–22, 2004.

[5] A. Vazquez, J. G. Oliveira, Z. Dezso, K.-I. Goh, I. Kondor, and A.-L. Barabasi, "Modeling bursts and heavy tails in human dynamics," *Physical Review E*, vol. 73, no. 3, Article ID 036127, 2006.

[6] M. Gidlund and N. Debernardi, "Scheduling performance of heavy-tailed data traffic in wireless high-speed shared channels," in *Proceedings of the IEEE Wireless Communications & Networking Conference (WCNC '09)*, pp. 1818–1823, IEEE Press, Budapest, Hungary, April 2009.

[7] R. Cont, "Empirical properties of asset returns: stylized facts and statistical issues," *Quantitative Finance*, vol. 1, pp. 223–236, 2001.

[8] T.-H. Kim and H. White, "On more robust estim ation of skewness and kurtosis: simulation and application to the SP500 index," Tech. Rep., Department of Economics, UCSD, 2003.

[9] T. C. Aysal and K. E. Barner, "Second-order heavy-tailed distributions and tail analysis," *IEEE Transactions on Signal Processing*, vol. 54, no. 7, pp. 2827–2832, 2006.

[10] V. K. Smith, "Least squares regression with Cauchy errors," *Oxford Bulletin of Economics and Statistics*, vol. 35, no. 3, pp. 223–231, 1973.

[11] J. Ilow, "Forecasting network traffic using FARIMA models with heavy tailed innovations," in *Proceedings of the IEEE Interntional Conference on Acoustics, Speech, and Signal Processing*, pp. 3814–3817, June 2000.

[12] W. Palma and M. Zevallos, "Fitting non-Gaussian persistent data," *Applied Stochastic Models in Business and Industry*, vol. 27, no. 1, pp. 23–36, 2011.

[13] J. Nowicka-Zagrajek and R. Weron, "Modeling electricity loads in California: ARMA models with hyperbolic noise," *Signal Processing*, vol. 82, no. 12, pp. 1903–1915, 2002.

[14] R. J. Adler, R. E. Feldman, and M. S. Taqqu, Eds., *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, Birkhauser, Cambridge, Mass, USA, 1998.

[15] H. Liu, J. Lafferty, and L. Wasserman, "The nonparanormal: semiparametric estimation of high dimensional undirected graphs," *Journal of Machine Learning Research*, vol. 10, pp. 2295–2328, 2009.

[16] G. M. Goerg, "Lambert W random variables—a new family of generalized skewed distributions with applications to risk estimation," *The Annals of Applied Statistics*, vol. 5, no. 3, pp. 2197–2230, 2011.

[17] D. C. Hoaglin, "Summarizing shape numerically: the g-and-h distributions," in *Exploring Data Tables, Trends, and Shapes*, D. C. Hoaglin, F. Mosteller, and J. W. Tukey, Eds., chapter 11, pp. 461–513, John Wiley & Sons, Hoboken, NJ, USA, 1985.

[18] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2010.

[19] C. Field and M. G. Genton, "The Multivariate g-and-h Distribution," *Technometrics*, vol. 48, no. 1, pp. 104–111, 2006.

[20] R. M. Sakia, "The Box-Cox transformation technique: a review," *Journal of the Royal Statistical Society, Series D: The Statistician*, vol. 41, no. 2, pp. 169–178, 1992.

[21] J. R. Blaylock, L. E. Salathe, and R. D. Green, "A note on the Box-Cox transformation under heteroskedasticity," *Western Journal of Agricultural Economics*, vol. 5, no. 2, pp. 129–136, 1980.

[22] A. J. Lawrance, "A note on the variance of the Box-Cox regression transformation estimate," *Journal of the Royal Statistical Society C*, vol. 36, no. 2, pp. 221–223, 1987.

[23] G. Tsiotas, "On the use of the Box-Cox transformation on conditional variance models," *Finance Research Letters*, vol. 4, no. 1, pp. 28–32, 2007.

[24] S. Gonçalves and N. Meddahi, "Box-Cox transforms for realized volatility," *Journal of Econometrics*, vol. 160, no. 1, pp. 129–144, 2011.

[25] N. H. Bingham, C. M. Goldie, and J. L. Teugels, *Regular Variation. Encyclopedia of Mathematics and its Applications*, Cambridge University Press, Cambridge, UK, 1989.

[26] S. Morgenthaler and J. W. Tukey, "Fitting quantiles: doubling, *HR*, *HQ*, and *HHH* distributions," *Journal of Computational and Graphical Statistics*, vol. 9, no. 1, pp. 180–195, 2000.

[27] M. Fischer, "Generalized Tukey-type distributions with application to financial and teletraffic data," *Statistical Papers*, vol. 51, no. 1, pp. 41–56, 2010.

[28] T. C. Headrick, R. K. Kowalchuk, and Y. Sheng, "Parametric probability densities and distribution functions for Tukey g-and-h transform ations and their use for fi tting data," *Applied Mathematical Sciences*, vol. 2, no. 9, pp. 449–462, 2008.

[29] R. M. Corless, G. H. Gonnet, D. E. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert *W* function," *Advances in Computational Mathematics*, vol. 5, no. 4, pp. 329–359, 1996.

[30] M. Rosenlicht, "On the explicit solvability of certain transcendental equations," *Publications Mathématiques de l'Institut des Hautes Études Scientifiques*, no. 36, pp. 15–22, 1969.

[31] S. R. Valluri, D. J. Jeffrey, and R. M. Corless, "Some applications of the Lambert W function to physics," *Canadian Journal of Physics*, vol. 78, no. 9, pp. 823–831, 2000.

[32] M. Galassi, J. Davies, J. Theiler et al., *GNU Scientific Library Reference Manual*, 3rd edition, 2011, http://www.gnu.org/software/gsl/.

[33] P. Jodrá, "A closed-form expression for the quantile function of the Gompertz-Makeham distribution," *Mathematics and Computers in Simulation*, vol. 79, no. 10, pp. 3069–3075, 2009.

[34] A. G. Pakes, "Lambert's W, infinite divisibility and poisson mixtures," *Journal of Mathematical Analysis and Applications*, vol. 378, no. 2, pp. 480–492, 2011.

[35] M. Stehlík, "Distributions of exact tests in the exponential family," *Metrika*, vol. 57, no. 2, pp. 145–164, 2003.

[36] J.-N. Hwang, S.-R. Lay, and A. Lippman, "Nonparametric multivariate density estimation: a comparative study," *IEEE Transactions on Signal Processing*, vol. 42, no. 10, pp. 2795–2810, 1994.

[37] L. Wasserman, *All of Nonparametric Statistics*, Springer Texts in Statistics, 2007.

[38] R. E. Maiboroda and N. M. Markovich, "Estimation of heavy-tailed probability density function with application to web data," *Computational Statistics*, vol. 19, no. 4, pp. 569–592, 2004.

[39] N. M. Markovich, "Accuracy of transformed kernel density estimates for a heavy-tailed distribution," *Automation and Remote Control*, vol. 66, no. 2, pp. 217–232, 2005.

[40] K. K. Dutta and D. Babbel, "On measuring skewness and kurtosis in short rate distributions: the case of the US dollar London inter bank offer rates," Tech. Rep., Wharton School Center for Financial Institutions, University of Pennsylvania, 2002.

[41] C. S. Wong, W. S. Chan, and P. L. Kam, "A Student t-mixture autoregressive model with applications to heavy-tailed financial data," *Biometrika*, vol. 96, no. 3, pp. 751–760, 2009.

[42] J. Yan, "Asymmetry, fat-tail, and autoregressive conditional density in financial return data with systems of frequency curves," 2005, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.76.2741.

[43] G. D. Rayner and H. L. MacGillivray, "Numerical maximum likelihood estimation for the g-and-k and generalized g-and-h distributions," *Statistics and Computing*, vol. 12, no. 1, pp. 57–75, 2002.

[44] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, Springer Texts in Statistics, Springer, New York, NY, USA, 2nd edition, 1998.

[45] C. Fernandez and M. F. Steel, "Multivariate Student-t regression models: pitfalls and inference," *Biometrika*, vol. 86, no. 1, pp. 153–167, 1999.

[46] C. Liu and D. B. Rubin, "ML estimation of the t distribution using EM and its extensions, ECM and ECME," *Statistica Sinica*, vol. 5, pp. 19–39, 1995.

[47] M. P. Cohen, "Sample means of independent standard CAUchy random variables are standard cauchy: a new approach," *The American Mathematical Monthly*, vol. 119, no. 3, pp. 240–244, 2012.

[48] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, vol. 31, no. 3, pp. 307–327, 1986.

[49] R. Deo, C. Hurvich, and Y. Lu, "Forecasting realized volatility using a long-memory stochastic volatility model: estimation, prediction and seasonal adjustment," *Journal of Econometrics*, vol. 131, no. 1-2, pp. 29–58, 2006.

[50] A. Melino and S. M. Turnbull, "Pricing foreign currency options with stochastic volatility," *Journal of Econometrics*, vol. 45, no. 1-2, pp. 239–265, 1990.

[51] A. Azzalini and A. Capitanio, "Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 65, no. 2, pp. 367–389, 2003.

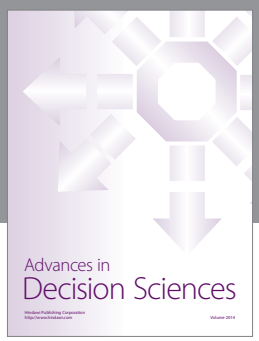
[52] R. N. Mantegna and H. E. Stanley, "Modeling of financial data: comparison of the truncated Lévy flight and the ARCH(1) and GARCH(1,1) processes," *Physica A: Statistical Mechanics and its Applications*, vol. 254, no. 1-2, pp. 77–84, 1998.

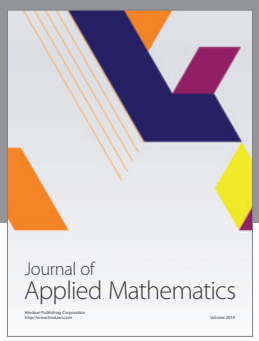
[53] P. A. Zadrozny, *Necessary and Sufficient Restrictions for Existence of a Unique Fourth Moment of a Univariate GARCH(p,q) Process*, vol. 20 of *Advances in Econometrics*, chapter 13, Emerald Group Publishing, 2005, https://ideas.repec.org/p/ces/ceswps/_1505.html.

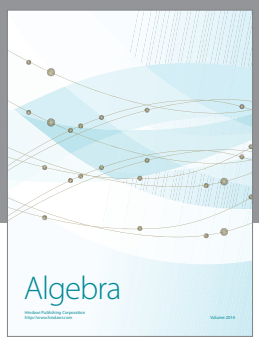
[54] R. Huisman, K. G. Koedijk, C. J. Kool, and F. Palm, "Tail-index estimates in small samples," *Journal of Business & Economic Statistics*, vol. 19, no. 2, pp. 208–216, 2001.

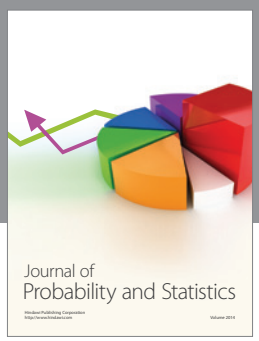
[55] H. C. Thode Jr., *Testing for Normality*, CRC Press, New York, NY, USA, 2002.

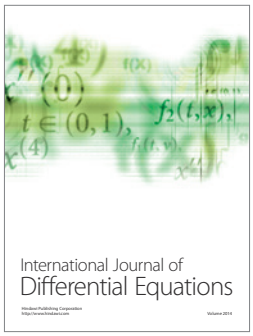
[56] G. M. Goerg, "Closed-form cdf and pdf of Tukey's h-distribution, the heavy-tail Lambert W approach, and how to bijectively "Gaussianize" heavy-tailed data," Tech. Rep., Carnegie Mellon University, 2010, http://arxiv-web3.library.cornell.edu/abs/1010.2265v1.

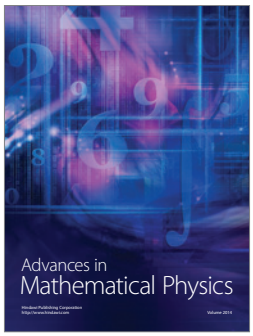
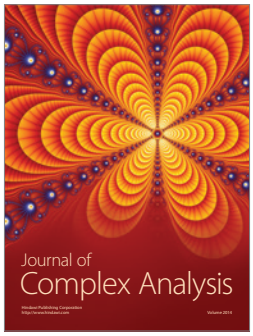
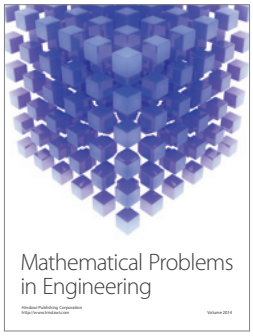
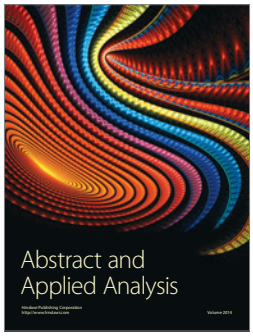
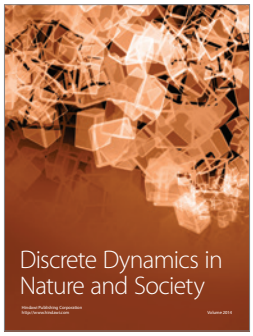
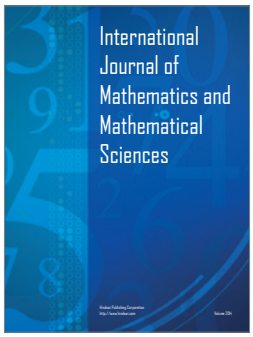
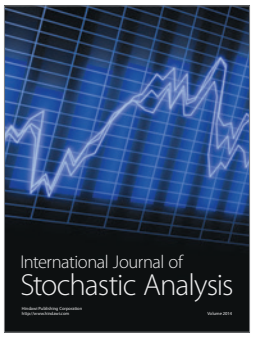










Hindawi

Submit your manuscripts at
http://www.hindawi.com