Qeios

A new measure of the tail-heaviness of a probability distribution

Hening Huang
Teledyne RD Instruments (retired)
San Diego, California, USA

Abstract This technical note proposes a new index for measuring the tail-heaviness of a probability distribution, named tail-heaviness index. The proposed tail-heaviness index is defined as the ratio between the right-side uncertainty length of the distribution and that of the standard exponential distribution. Tail-heaviness indices for four distributions are presented.

Keywords Informity; tail-heaviness; probability distribution; uncertainty length

1. Introduction

Tail-heaviness is an important characteristic of a probability distribution. An effective measure of tail-heaviness is useful to practitioners and researchers when building statistical models and comparing different distributions.

Kurtosis (or excess kurtosis) is traditionally used as a measure of the tail-heaviness (Brys et al. 2006, Westfall 2014, Ortega 2017). However, kurtosis has several limitations or drawbacks. First, there is no agreement on what kurtosis really measures (Brys et al. 2006). Second, the kurtosis of some well-known heavy-tailed distributions (such as Pareto and Cauchy distributions) does not exist. Third, kurtosis is very sensitive to outliers in the data (Brys et al. 2006).

A number of studies have proposed different measures to quantify the tail-heaviness of probability distributions. Schuster (1984) used the right-tail exponent measure to classify distributions into three classes: short- medium-, and long-tailed. Brys et al. (2006) proposed several tail weight measures based on robust measures of skewness. They considered left and right tail weight measures to make them applicable to asymmetric distributions. Jordanova and Petkova (2017) proposed a measure for heavy-tailedness of the observed distribution and introduced classification of the distributions with respect to their heavy-tailedness. Ortega (2017) proposed a tail weight coefficient based on the tail set that is constructed using a fixed cut-off value. However, the literature reveals considerable uncertainty, controversy, and error in connection with the distinctions between tail weights (Heyde and Kou 2004).

In this technical note, we propose a new index for measuring the tail-heaviness of a probability distribution, named tail-heaviness index. In the following sections, section 2 describes the proposed tail-heaviness index. Section 3 gives examples of the tail-heaviness indices for four distributions. Sections 4 and 5 present discussion and conclusion respectively.

2. The proposed tail-heaviness index

Consider a continuous random variable Y with its probability density function (PDF) $p(y)$. Whether the distribution is considered "heavy-tailed" is inherently relative (Nair et al. 2022). We use the standard exponential distribution (rate parameter $\lambda=1$) as a reference distribution to compare the tail weights of distributions. That is, a distribution is considered heavy-tailed if its tail weight is heavier than the tail weight of the standard exponential distribution. This is consistent with Nair et al. (2022) for defining the class of heavy-tailed distributions. Ortega (2017) also used the

https://doi.org/10.32388/9B8HK9

exponential distribution as a reference to divide distributions into two categories: thin-tailed distributions and heavy-tailed distributions.

We consider unimodal distributions only. We divide the PDF of a unimodal distribution into two parts at the median: the right side and the left side. Our tail-heaviness analysis focuses only on the right side of the distribution (or the downside tail). The study of the downside tail of the distribution of asset returns is a fundamental and essential factor for the analysis of financial market risks (González-Sánchez and Nave Pineda 2023). It is important to note that, there is no rigorous definition for the "tail" (Wicklin 2014). We argue that the tail shape cannot be completely separated from the body shape of the distribution, i.e. the tail shape and the body shape are related. Therefore, it is reasonable to consider the half (the right side) of the distribution in the tail-heaviness analysis.

We adopt the concept of uncertainty length as a measure of the tail weight of a distribution. The uncertainty length of the *Y* distribution is defined as (Huang 20023)

$$\delta(Y) = \frac{1}{\beta(Y)} = \frac{1}{\int_{-\infty}^{\infty}[p(y)]^2 dy}, \tag{1}$$

where $\beta(Y) = \int_{-\infty}^{\infty}[p(y)]^2 dy$ is the continuous informity (Huang 2023).

The right-side uncertainty length of the distribution can be written as

$$\delta_{\text{right}}(Y) = 0.5\frac{1}{\beta_{\text{right}}(Y)}, \tag{2}$$

where $\beta_{\text{right}}(Y)$ is the right-side informity defined as

$$\beta_{\text{right}}(Y) = 2\int_{\text{median}}^{\infty}[p(y)]^2 dy. \tag{3}$$

If the distribution is symmetric (e.g. the normal distribution), $\beta_{\text{right}}(Y) = \beta(Y)$.

The proposed tail-heaviness index, denoted by θ, is defined as

$$\theta = \frac{\delta_{\text{right}}(Y)}{\delta_{\text{right-exp}}}, \tag{4}$$

where $\delta_{\text{right-exp}}$ is the right-side uncertainty length of the standard exponential distribution. The PDF of the standard exponential distribution is $p_{\text{exp}}(y) = e^{-y}$, and its median is ln2. Then, its right-side informity can be calculated as

$$\beta_{\text{right-exp}}(Y) = 2\int_{\ln 2}^{\infty} e^{-2y} dy = e^{-2\ln 2} = 0.25 \tag{5}$$

Then, the right-side uncertainty length of the standard exponential distribution is

$$\delta_{\text{right-exp}} = 0.5\frac{1}{\beta_{\text{right-exp}}(Y)} = 0.5\frac{1}{0.25} = 2 \tag{6}$$

Accordingly, the proposed tail-heaviness index is rewritten as

$$\theta = \frac{1}{2}\delta_{\text{right}}(Y) = \frac{0.5/\beta(Y)}{2}. \tag{7}$$

3. Examples

This section gives examples of the proposed tail-heaviness indices for four distributions.

3.1 The normal distribution

The PDF of the normal distribution is

$$p_{\text{normal}}(y) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right], \tag{8}$$

where μ is the mean and σ is the standard deviation.

The informity of the normal distribution is (Huang 2023)

$$\beta_{\text{normal}}(Y) = \frac{1}{2\sigma\sqrt{\pi}}. \tag{9}$$

Since the normal distribution is symmetric, $\beta_{\text{right-normal}}(Y) = \beta_{\text{normal}}(Y)$. The proposed tail-heaviness index for the normal distribution is

$$\theta_{\text{normal}} = \frac{\delta_{\text{right-normal}}}{\delta_{\text{right-exp}}} = \frac{0.5/\beta_{\text{normal}}(Y)}{2} = \frac{\sqrt{\pi}}{2}\sigma. \tag{10}$$

Note that θ_{normal} increases linearly with σ and does not depend on μ. For the standard normal distribution ($\mu = 0, \sigma = 1$), $\theta_{\text{normal}} = 0.886 < 1$. Therefore, the standard normal distribution is considered light-tailed. This result is consistent with common sense.

3.2 The Pareto distribution

The PDF of the Pareto distribution is

$$p_{\text{Pareto}}(y) = \begin{cases} \frac{\alpha y_m^\alpha}{y^{\alpha+1}} & y \geq y_m, \\ 0 & y < y_m, \end{cases} \tag{11}$$

where α is the shape parameter (also known as tail index) and y_m is the minimum value of the distribution.

The median of the Pareto distribution is $y_m\sqrt[\alpha]{2}$. Then, its right-side informity is

$$\beta_{\text{right-Pareto}}(Y) = 2\alpha^2 y_m^{2\alpha} \int_{y_m \sqrt[\alpha]{2}}^{\infty} \frac{1}{y^{2(\alpha+1)}} dy = \frac{1}{2^{1+1/\alpha}} \frac{\alpha^2}{(2\alpha+1)} \frac{1}{y_m}. \tag{12}$$

The proposed tail-heaviness index for the Pareto distribution is

$$\theta_{\text{Pareto}} = \frac{\delta_{\text{right-Pareto}}}{\delta_{\text{right-exp}}} = \frac{0.5/\beta_{\text{right-Pareto}}(Y)}{2} = \frac{1}{2\alpha^2} 2^{\frac{1}{\alpha}} (2\alpha+1) y_m. \tag{13}$$

Equation (13) indicates that θ_{Pareto} depends on α and y_m. When $\alpha = 1$ and $y_m = 1$, $\theta_{\text{Pareto}} = 3$. Figure 1 shows the plot of θ_{Pareto} as a function of the shape parameter α at $y_m = 1$.

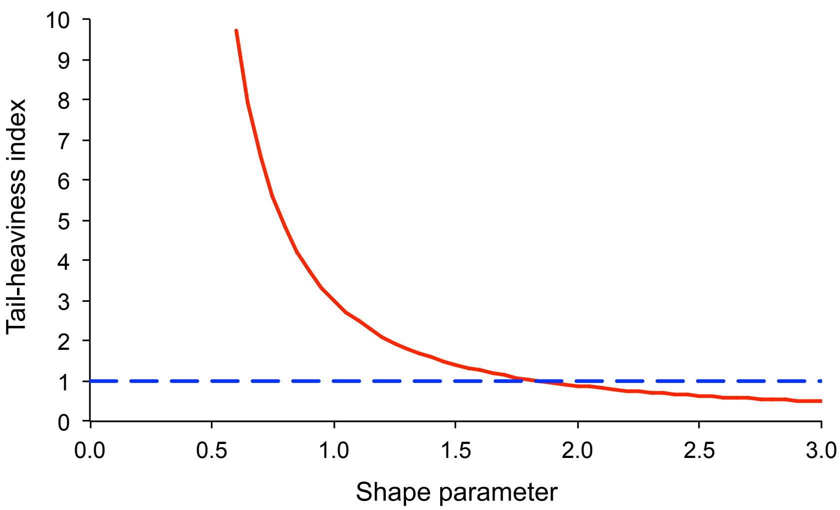

Figure 1. Tail-heaviness index for the Pareto distribution as a function of the shape parameter α ($y_m = 1$)

As can be seen from Figure 1, as α decreases, θ_{Pareto} increases because the tail becomes heavier. On the other hand, as α increases, θ_{Pareto} decreases because the tail becomes lighter.

3.3 The Cauchy distribution

The PDF of the Cauchy distribution is

$$p_{\text{Cauchy}}(y) = \frac{1}{\pi\gamma[1+\frac{1}{\gamma^2}(y-y_0)^2]}, \tag{14}$$

where y_0 is the location parameter and γ is the scale parameter.

The informity of the Cauchy distribution is (Huang 2023)

$$\beta_{\text{Cauchy}}(Y) = \frac{1}{2\pi\gamma}. \tag{15}$$

Since the Cauchy distribution is symmetric, $\beta_{\text{right-Cauchy}}(Y) = \beta_{\text{Cauchy}}(Y)$. The proposed tail-heaviness index for the Cauchy distribution is

$$\theta_{\text{Cauchy}} = \frac{\delta_{\text{right-Cauchy}}}{\delta_{\text{right-exp}}} = \frac{0.5/\beta_{\text{Cauchy}}(Y)}{2} = \frac{\pi}{2}\gamma. \tag{16}$$

Note that θ_{Cauchy} increases linearly with γ and does not depend on y_0.

3.4 The t-distribution

The PDF of the *t*-distribution is written as

$$p_t(y) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{y^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \tag{17}$$

where ν is the degrees of freedom.

The informity of the *t*-distribution can be calculated as (Huang 2024)

$$\beta_t(Y) = \int [p(y)]^2\, dy = \frac{\Gamma^2\left(\frac{\nu+1}{2}\right)}{\nu\pi\Gamma^2\left(\frac{\nu}{2}\right)} \int_{-\infty}^{\infty} \frac{1}{\left(1+\frac{y^2}{\nu}\right)^{\nu+1}}\, dy. \tag{18}$$

Since the *t*-distribution is symmetric, $\beta_{\text{right-}t}(Y) = \beta_t(Y)$. The proposed tail-heaviness index for the *t*-distribution is

$$\theta_t = \frac{\delta_{\text{right-}t}}{\delta_{\text{right-exp}}} = \frac{0.5/\beta_t(Y)}{2} = \frac{\nu\pi\Gamma^2\left(\frac{\nu}{2}\right)}{2\Gamma^2\left(\frac{\nu+1}{2}\right)} \frac{1}{\int_{-\infty}^{\infty} \frac{1}{\left(1+\frac{y^2}{\nu}\right)^{\nu+1}} dy}. \tag{19}$$

The solution for θ_t (or $\beta_t(Y)$) can be obtained by using the following integrals

$$\int_{-\infty}^{\infty} \frac{1}{(1+cy^2)^n}\, dy = \frac{(2n-3)}{2(n-1)} \int_{-\infty}^{\infty} \frac{1}{(1+cy^2)^{n-1}}\, dy, \text{ and}$$

$$\int_{-\infty}^{\infty} \frac{1}{1+cy^2}\, dy = \frac{\pi}{\sqrt{c}},$$

where $n = \nu - 1$ and $c = \frac{1}{\nu}$.

Figure 2 shows θ_t as a function of ν. As can be seen from Figure 2, at $\nu = 1$ (in this case, the Cauchy distribution with $\gamma = 1$), θ_t =1.571. θ_t decreases as ν increases, and as ν goes to infinity, θ_t should approach 0.886, the tail-heaviness index for the standard normal distribution.

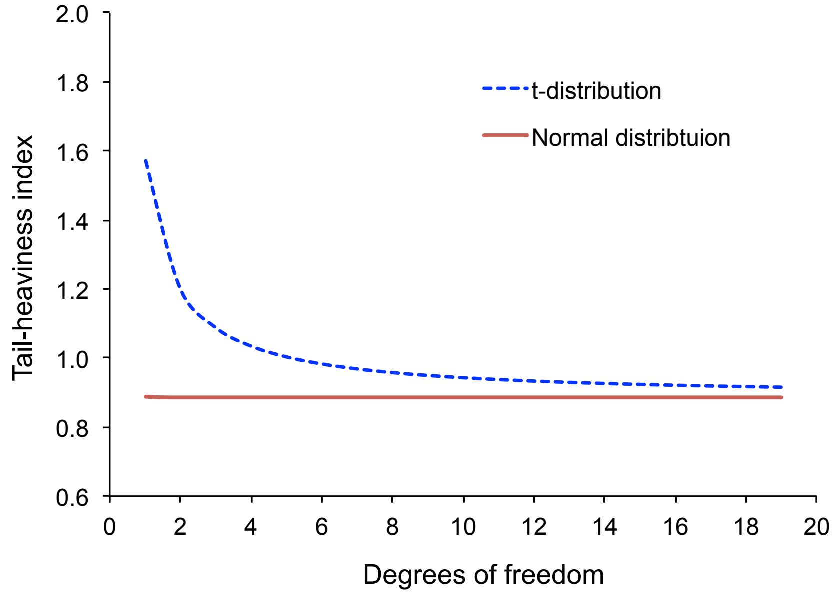

Figure 2. Tail-heaviness index for the *t*-distribution as a function of the degrees of freedom

4. Discussion

The proposed tail-heaviness index θ measures the tail-heaviness of a distribution relative to the tail-heaviness of the standard exponential distribution. If $\theta > 1$, the distribution is considered heavy-tailed; if $\theta < 1$, the distribution is considered light-tailed.

The proposed tail-heaviness index can be calculated for any distribution even if the first, second, or fourth moment of the distribution does not exist or is undefined. This is because the numerical value of the uncertainty length (or informity) of any distribution can always be obtained. In contrast, the usual tail-heaviness measure: kurtosis does not exist for the Pareto distribution or for the *t*-distribution for the degrees of freedom $\nu \leq 4$.

Furthermore, the proposed tail-heaviness index is not limited to distributions with infinite support. It is also applicable to distributions with finite support, such as triangular distributions.

5. Conclusion

The proposed tail-heaviness index θ provides a simple and effective measure of the tail-heaviness of a probability distribution. It is easy to understand and calculate. If $\theta > 1$, the distribution is considered heavy-tailed; if $\theta < 1$, the distribution is considered light-tailed. In principle, the proposed tail-heaviness index is valid for and applicable to any unimodal distribution: symmetric or asymmetric. It can be calculated for any distribution even if the first, second or fourth moment of the distribution does not exist. The tail-heaviness indices for four distributions are presented. Further work is needed to derive tail-heaviness indices for other distributions.

Conflict of Interest

The author reports there are no competing interests to declare.

References

Brys G, Hubert M, and Struyf A 2006 Robust measures of tail weight *Computational Statistics & Data Analysis* **50**(3) 733-759 https://doi.org/10.1016/j.csda.2004.09.012

González-Sánchez M and Nave Pineda J M 2023 Where is the distribution tail threshold? A tale on tail and copulas in financial risk measurement *International Review of Financial Analysis* Elsevier **86**(C)

Heyde C C and Kou S G 2004 On the controversy over tailweight of distributions *Operations Research Letters* **32** 399-408

Huang H 2024 A new index for measuring the peakedness of a probability distribution (preprint) *ResearchGate* https://www.researchgate.net/publication/379239769_A_new_index_for_measuring_the_peakedness_of_a_probability_distribution DOI: 10.13140/RG.2.2.15737.68964

Huang H 2023 The theory of informity (preprint) *ResearchGate* https://www.researchgate.net/publication/376206296_The_theory_of_informity DOI: 10.13140/RG.2.2.28832.97287

Jordanova P K and Petkova M P 2017 Measuring heavy-tailedness of distributions *AIP Conf. Proc.* 1910, 060002 https://doi.org/10.1063/1.5013996

Nair J, Wierman A, and Zwart B 2022 *The Fundamentals of Heavy Tails: Properties, Emergence, and Estimation* Cambridge University Press DOI: https://doi.org/10.1017/9781009053730

Ortega J F 2017 Tail weight measures for distributions A new tail weight coefficient *Estadística Española* **59**(193) 103-114

Schuster E F 1984 Classification of probability laws by tail behavior *Journal of the American Statistical Association* **79** 936-939

Westfall P H 2014 Kurtosis as Peakedness, 1905 - 2014. *R.I.P. Am Stat.* **68**(3) 191-195 doi: 10.1080/00031305.2014.917055. PMID: 25678714; PMCID: PMC4321753

Wicklin R 2014 Fat-tailed and long-tailed distributions *SAS Blogs, The DO Loop*, https://blogs.sas.com/content/iml/2014/10/13/fat-tailed-and-long-tailed-distributions.html accessed April 2024