A NOTE ON THE CONSTRUCTION OF GENERALIZED TUKEY-TYPE TRANSFORMATIONS.

Matthias Fischer
Department of Statistics and Econometrics
University of Erlangen-Nürnberg, Germany
Email: Matthias.Fischer@wiso.uni-erlangen.de

SUMMARY

One possibility to construct heavy tail distributions is to directly manipulate a standard Gaussian random variable by means of transformations which satisfy certain conditions. This approach dates back to Tukey (1960) who introduces the popular H-transformation. Alternatively, the K-transformation of MacGillivray & Cannon (1997) or the J-transformation of Fischer & Klein (2004) may be used. Recently, Klein & Fischer (2006) proposed a very general power kurtosis transformation which includes the above-mentioned transformations as special cases. Unfortunately, their transformation requires an infinite number of unknown parameters to be estimated. In contrast, we introduce a very simple method to construct flexible kurtosis transformations. In particular, manageable "superstructures" are suggested in order to statistically discriminate between H-, J- and K-distributions (associated to H-, J- and K-transformations).

Keywords and phrases: Generalized kurtosis transformation, H-transformation

AMS Classification: 62E15

1 Introduction

Assume that Z denotes a standard Gaussian random variable. In order to derive random variables with heavy-tailed distributions, Tukey (1960) suggested to directly transform Z via

$$Y = Z \cdot T(Z)^{\theta}, \tag{1.1}$$

where $\theta \geq 0$ is a kurtosis parameter[1] and $T : \mathbb{R}_+ \to \mathbb{R}$ a suitable positive function ("transformation") which is symmetric around 0 and strictly monotone increasing on $\mathbb{R}_+$. Tukey (1960) discussed the H-transformation

$$T(z) = \exp(0.5z^2) \tag{1.2}$$

which guarantees the existence of moments of Y up to order $1/\theta$ which in turn coincides with the (asymptotic) tail index of Y. Parametric alternatives with existing moments but still heavy tails followed up by MacGillivray & Cannon's (1997) K-transformation, i.e.

$$T(z) = (1 + z^2), \tag{1.3}$$

[1] In order to increase the tail length of Z which we focus on with this work.

by Fischer & Klein (2004) who discussed the J-transformation

$$T(z) = \cosh(z) = 0.5(\exp(z) + \exp(-z)) \tag{1.4}$$

and by Fischer, Horn & Klein's (2006) L-transformation

$$T(z) = \frac{\sinh(z)}{z} = \frac{0.5(e^z - e^{-z})}{z}. \tag{1.5}$$

Until Klein & Fischer (2006), these transformations only "exist in parallel" and no "super-structure" was available. Note that the transformations given in (1.2) to (1.5) are special cases of the very general power series representation

$$T(z) = \sum_{i=0}^{\infty} a_i z^{2i} \tag{1.6}$$

with certain weights a_i, $i \geq 0$ which guarantee that the power series in (1.6) has a finite limit. In particular, the coefficients of Tukey's H-transformation are $a_i = 1/(2^i i!) = 1/(2i)!$, $i \in \mathbb{N}$, the coefficients of the K-transformation are simply $a_0 = a_1 = 1$ and $a_i = 0$, $i > 1$, the coefficients of the J-transformation are given by $a_i = 1/(2i!)$ and that of the L-transformation by $a_i = 1/(2i+1)!$ Unfortunately, the representation in (1.6) is not very operational as we have to estimate an infinite number of unknown parameters $a_0, a_1, \ldots$ to reveal the "data-generating transformation". This motivates the need of alternative flexible transformations which include the above-mentioned transformations, or at least some of them, as special cases.

2 A simple method to construct transformations

Given a symmetric probability density $f \in C^2(\mathbb{R})$ with $f(x) > 0$ and $f'(x) \leq 0$ for $x \in \mathbb{R}$ (briefly $f \in \mathcal{F}$), define

$$T(x; f) \equiv \frac{f(0)}{f(x)}. \tag{2.1}$$

We next show that $T(x; f)$ is actually a Tukey-type transformation which can be used to construct heavy-tailed distributions.

Lemma 1. Assume that $f \in \mathcal{F}$. It follows that $T(0, f) = 1$, $T(-x, f) = T(x, f)$ and that $T(x, f)$ is strictly monotone increasing on $(0, \infty)$. Moreover,

$$T''(x, f) \geq 0 \quad \Longleftrightarrow \quad 2f'(x)^2 \geq f''(x)f(x).$$

Proof: The assertions follow from the first and second derivative of $T(x, f)$,

$$T'(x) = -f(0)\frac{f'(x)}{f(x)^2} \quad \text{and} \quad T''(x) = f(0)\,\frac{2f'(x)^2 - f''(x)f(x)}{f(x)^3}. \quad \square$$

Applying a Taylor series expansion to $T(x, f)$ around $x_0 = 0$ and using the symmetry of $T(x, f)$, we obtain the coefficients of the power series representation from above which are completely determined by the density and its derivatives (provided that all derivatives exist).

Lemma 2. The power series representation of $T(x,f)$ is given by

$$T(x,f) = \sum_{i=0}^{\infty} a_i z^{2i} \quad \text{with} \quad a_i = \frac{1}{(2i)!} \frac{d^{(2i)}}{dx^{(2i)}} \left(\frac{f(0)}{f(x)} \right) \bigg|_{x=0}. \tag{2.2}$$

In particular, with $\psi(x) \equiv -f'(x)/f(x)$, the first three coefficients are

$$a_0 = 1, \quad a_1 = \frac{1}{2}\left(\psi'(0) - \psi(0)^2\right),$$

$$a_2 = \frac{1}{24}\left(\psi'''(0) + 2\psi'(0)\psi''(0) - \psi'(0)\psi(0)^2 - \psi'(0)^2 + \psi(0)\psi'(0)^2\right).$$

The kurtosis transformations stated in equation (1.2), (1.3), (1.4) and (1.5) correspond to well-known probability densities, as the next example shows. Additionally, a new transformation similar to the J-transformation is obtained.

Example 2.1 (H-/K-/J-transformation)**.**

1. Assume that $f(x) = \varphi(x)$, the standard Gaussian density. Obviously, $f(0) = (\sqrt{2\pi})^{-1}$ and the H-transformation from (1.2) is recovered.
2. Plugging the Cauchy density into (2.1), the K-transformation is obtained:
$$f(x) = \frac{1}{\pi(1+x^2)} \quad \Longleftrightarrow \quad T(x,f) = \frac{1/\pi}{1/(\pi(1+x^2))} = 1 + x^2.$$
3. Similarly, the hyperbolic secant distribution is associated to the J-transformation of Fischer & Klein (2004):
$$f(x) = \frac{1}{\pi\cosh(x)} \quad \Longleftrightarrow \quad T(x,f) = \frac{1/\pi}{1/(\pi\cosh(x))} = \cosh(x).$$
4. Plugging the logistic density into (2.1) reveals a "new" transformation which is asymptotically equivalent to the J-transformation:
$$f_4(x) = \frac{\exp(x)}{(1+\exp(x))^2} \quad \Longleftrightarrow \quad T(x,f_4) = \frac{1/4}{\frac{\exp(x)}{(1+\exp(x))^2}} = \frac{(1+\exp(x))^2}{4\exp(x)}.$$

Conversely, given a transformation $T(x)$, the associated "generating density" boils down to

$$f(x;T) = \left(\int_{-\infty}^{\infty} \frac{dx}{T(x)} \right) 1/T(x).$$

Example 2.2 (L-transformation)**.** Starting from the L-transformation $T(x) = \frac{\sinh(x)}{x}$ and noting that

$$\int_{-\infty}^{\infty} \frac{x}{\sinh(x)} dx = \frac{\pi^2}{2},$$

the resulting distribution is the convolution of two hyperbolic secant with density given by

$$f_5(x) = \frac{2}{\pi^2} \frac{x}{\sinh(x)} = \frac{4}{\pi^2} \frac{x}{e^x - e^{-x}}.$$

3 Generalized families of transformations

In order to obtain generalized transformations which include some of these transformations as special cases we simply have to select generalized symmetric distributions which include the corresponding distributions (i.e. Gaussian, Cauchy, logistic, hyperbolic secant distribution) as special cases and apply equation (1.6). We focus on four transformation families, henceforth:

1. The *Student-t transformation family* which nests both H- and K-transformation,
2. the *GED transformation family* which generalizes Tukey's H-transformation,
3. the *GSH transformation family* which generalizes Fischer & Klein's J-transformation,
4. the *Meixner transformation family* which nests both J- and L-transformation.

3.1 The Student-t transformation family

A popular distribution which includes the Cauchy distribution ($\nu = 1$) as well as the normal distribution ($\nu \to \infty$) is the Student-t distribution with density

$$f(x;\nu) = \frac{1}{\sqrt{\nu\pi}} \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} \left(1 + x^2/\nu\right)^{-\frac{\nu+1}{2}}.$$

Consequently, the associated transformation (including H- and K-transformation) is

$$T_S(x;\nu) = \left(1 + x^2/\nu\right)^{\frac{\nu+1}{2}} > 0, \quad \nu \in \mathbb{R}. \tag{3.1}$$

Obviously, $f'(x;\nu) \leq 0$ for $x \in \mathbb{R}$ and convexity of $T(x;\nu)$ holds because

$$T_S''(x;\nu) = T_S(x)\nu(\nu+1)\,\frac{1+x^2}{(\nu+x^2)^2} > 0.$$

3.2 The GED transformation family

A flexible parametric density connecting both normal and Laplace distribution is given by the generalized error (GED) distribution (see Subbotin, 1923 or Box & Tiao, 1962) with pdf

$$f(x;\nu) = \frac{\nu}{2^{1+1/\nu}\Gamma(1/\nu)} \exp(-0.5|x|^\nu), \quad \nu > 0,$$

also known as power exponential density under a different parameterization. The GED is still symmetric, but quite flexible in the tails through the parameter ν: when $\nu < 2$, tails are heavier than the normal ones; when $\nu > 2$, tails are thinner than the corresponding normal tails. The associated GED-transformation generalizes the $H-$transformation (which itself is recovered for $\nu = 2$) and is given by

$$T_{GED}(x;\nu) = \exp(0.5|x|^\nu), \quad \nu > 0.$$

Assuming $x > 0$ and $\nu \geq 1$ makes sure that the transformation is convex because then

$$T''_{GED}(x;\nu) = 0.5\nu x^{\nu-2} T_{GED}(x;\nu)((\nu-1) + 0.5x^\nu) > 0.$$

3.3 The GSH transformation family

The *generalized secant hyperbolic* (GSH) distribution – which is able to model both thin and fat tails – was introduced by Vaughan (2002) and has density

$$f_{GSH}(x;t) = \frac{c_1(t)\exp(x)}{\exp(2x) + 2a(t)\exp(x) + 1} = \frac{c_1(t)}{2\cosh(x) + 2a(t)}, \qquad x \in \mathbb{R} \tag{3.2}$$

$$\text{with} \begin{cases} a(t) = \cos(t), & c_1(t) = \frac{\sin(t)}{t}, \quad \text{for } -\pi < t \leq 0, \\ a(t) = \cosh(t), & c_1(t) = \frac{\sinh(t)}{t}, \quad \text{for } t > 0. \end{cases}$$

Setting $t = 0$ results in the logistic distribution, $t = -\pi/2$ corresponds to the hyperbolic secant distribution. Application of (1.6) to the GSH density provides another family of transformation indexed by the parameter $t \in (-\pi, \infty)$, i.e.

$$T_{GSH}(x;t) = \frac{\cosh(x) + a(t)}{1 + a(t)}, \quad t \in (-\pi, \infty).$$

The GSH-transformation is strictly monotone and convex noting that for $x > 0$

$$T'_{GSH}(x;t) = \frac{\sinh(x)}{1 + a(t)} > 0 \quad \text{and} \quad T''_{GSH}(x;t) = \frac{\cosh(x)}{1 + a(t)} > 0.$$

3.4 The Meixner transformation family

A flexible but not so prominent distribution family is the GHS or symmetric Meixner distribution family. Originally, Meixner (1934) introduced this families based on certain polynomials. Many properties are discussed by Harkness and Harkness (1968). The GHS distribution is obtained as λ^{th}-convolution of the hyperbolic secant family. Its density with kurtosis parameter $\lambda > 0$ is

$$f(x;\lambda) = \frac{2^{2d}}{2\pi\Gamma(2d)} |\Gamma(\lambda + \mathbf{i}x)|^2 = \frac{2^{2d}}{2\pi\Gamma(2d)} \Gamma(\lambda + \mathbf{i}x)\Gamma(\lambda - \mathbf{i}x), \tag{3.3}$$

where $\mathbf{i}$ denotes the imaginary number. Using the relationship

$$|\Gamma(n + \mathbf{i}x)|^2 = \frac{\pi P_n}{x\sinh(\pi x)} \quad \text{for} \quad n = 1, 2, \ldots,$$

with $P_1 = (0^2 + x^2)$, $P_2 = (0^2 + x^2)(1^2 + x^2)$, $P_3 = (0^2 + x^2)(1^2 + x^2)(2^2 + x^2), \ldots$

$$\text{and } |\Gamma(0.5 + ix)|^2 = \frac{\pi}{\cosh(\pi x)}$$

it is straightforward to derive the hyperbolic secant distribution ($\lambda = 0.5$) and the density of the convolution of two hyperbolic secant distributions ($\lambda = 1$), among others:

$$f(x;0.5) = \frac{1}{\cosh(\pi x)}, \quad f(x;1) = \frac{2x}{\sinh(\pi x)}, \quad f(x;3) = \frac{4}{15}\frac{x(1+x^2)}{\sinh(\pi x)}, \quad \ldots$$

Scaling the density by π and combining (1.6) and (3.3) finally leads to the Meixner transformations

$$T_M(x;\lambda) = \frac{\Gamma(\lambda)^2}{\Gamma(\lambda + \mathbf{i}x/\pi)\Gamma(\lambda - \mathbf{i}x/\pi)} = \frac{\Gamma(\lambda)^2}{|\Gamma(\lambda + \mathbf{i}x/\pi)|^2}, \quad \lambda > 0. \tag{3.4}$$

Using the digamma function $\psi(x) = \Gamma'(x)/\Gamma(x)$, the first derivative of the Meixner transformation is given by

$$T'_M(x;\lambda) = T_M(x;\lambda) \cdot \frac{\mathbf{i}}{\pi}\left[\psi\left(\lambda - \frac{\mathbf{i}x}{\pi}\right) - \psi\left(\lambda + \frac{\mathbf{i}x}{\pi}\right)\right] = T_M(x;\lambda) \cdot H(x;\lambda)$$

Using the series representation of the digamma function we obtain

$$H(x;\lambda) \quad = \quad -\frac{\mathbf{i}}{\pi}\sum_{k=0}^{\infty} \frac{2x/\pi\mathbf{i}}{x^2/\pi^2 + (\lambda+k)^2} = \frac{2}{\pi}\sum_{k=0}^{\infty} \frac{x/\pi}{x^2/\pi^2 + (\lambda+k)^2} > 0.$$

Exemplarily, different Meixner, GED, GSH and Student-t transformations are plotted in figure 1, below.

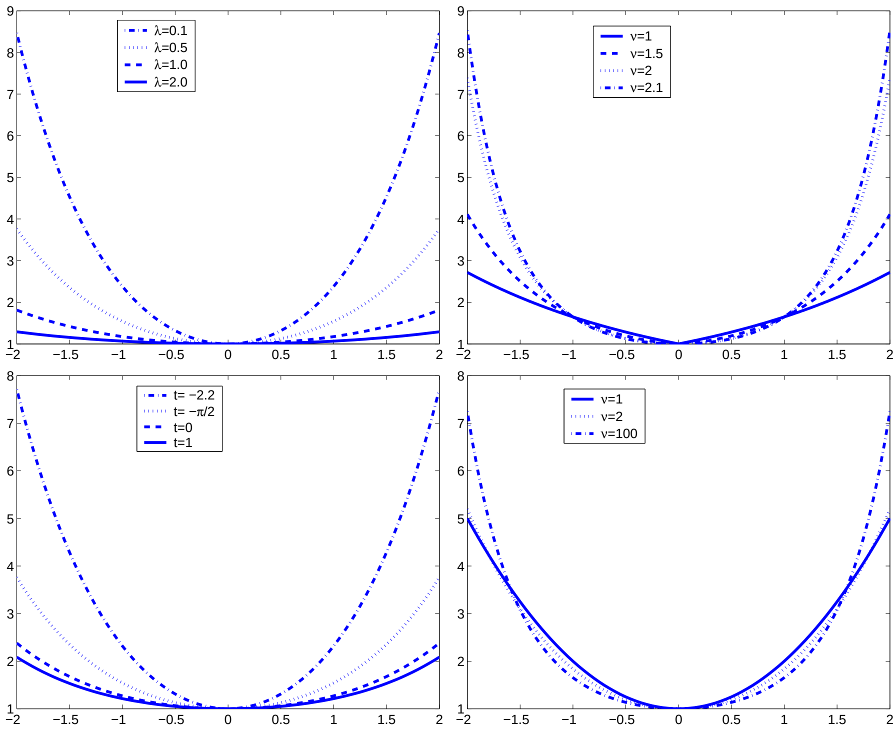

Figure 1: Selected Tukey-type transformations.

4 Application of generalized Tukey-type transformations

4.1 Generalized Tukey-type distributions and its estimation

Starting with a standard normal variable Z, we now focus on the distribution of $Y = K(Z) = \mu + \delta Z \cdot T(Z)^\theta$ from (1.1), where T is one of the generalized Tukey-type transformations considered before and $\mu \in \mathbb{R}$, $\delta > 0$ denote the location and scale parameter, respectively. Applying standard methods of variable transformation, the density of Y requires the inverse transformation of T – which is typically not available in closed form – and is given by

$$f_Y(y) = \frac{f_Z(K^{-1}((y-\mu)/\delta))/\delta}{K'(K^{-1}((y-\mu)/\delta))} \quad \text{with} \quad K'(z) = T(z)^{\theta-1}(T(z) + \theta z T'(z)).$$

Traditionally, quantile-based methods are applied to obtain estimates of the unknown parameters (see, for instance, Tukey, 1960). Due to the increasing computing power, maximum likelihood estimation (MLE) which had been thought intractable can now be tackled numerically. Refering to Rayner & MacGillivray (2002) for both theoretical and computational details, MLE maximizes the logarithm of the likelihood (as a function of the unknown parameters represented by the vector Θ) for a simple random sample $y_1, \ldots, y_n$, given by

$$\mathcal{LL}(\Theta; y_1, \ldots, y_n) = \sum_{i=1}^{n} \ln \left(\frac{f_Z(K^{-1}((y_i-\mu)/\delta))/\delta}{K'(K^{-1}((y_i-\mu)/\delta))} \right).$$

4.2 Modelling financial return distributions

We focus on the continuously compounded returns (e.g. differences of consecutive log prices) of ALLIANZ AG over the period 1 January 1990 to 31 December 2003 (3485 observations). The (sample) mean of the log-returns (which are depicted in figure 2, below) is -0.00002 with a (sample) standard deviation of 0.0221. Moreover, the data set exhibits only a small amount of skewness (the skewness coefficient – measured by the third standardized moments – is given by by -0.069), whereas the kurtosis coefficient – in terms of the fourth standardized moments – is 5.362, reflecting the remarkable leptokurtosis.
The results for the ALLIANZ returns arising from maximum likelihood estimation of the parameters from different Tukey-type distributions are summarized in table 1, below.
Obviously, focussing on the log likelihood value $\mathcal{LL}$, the return data under consideration are closer to the H-distribution than to the K-distribution, but closer to the J-distribution than to the H-distribution. Concerning the estimation results based on the generalized families from section 3, we can state the following observations.

1. Within the GSH- and Meixner transformation family, the fit of the J-distribution cannot be improved and the J-distribution is essentially recovered ($\widehat{t} = -1.45$, exact: $t = -\pi/2$ and $\widehat{\lambda} = 0.53$, exact: 0.5).

2. The estimation results for the distribution family derived from the Student-t density indicate that neither the K-distribution ($\nu = 1$) nor the H-distribution ($\nu = \infty$) are optimal within this family and that $\nu = 5.5$ might be a better choice.

3. The GED-family and its associated distributions provide an additional improvement regarding the log-likelihood.

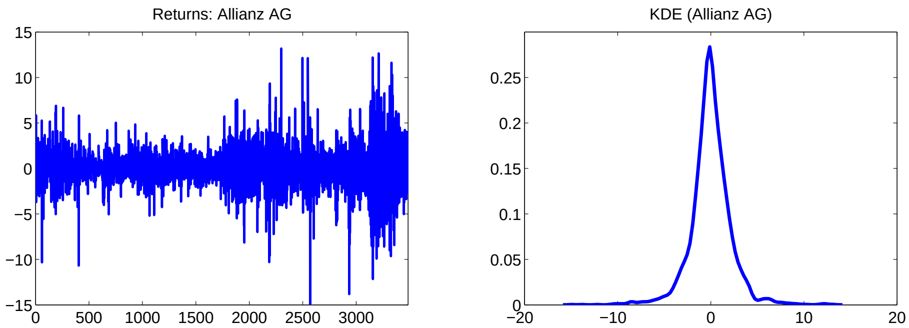

Figure 2: Series of returns and kernel density estimation.

Table 1: Estimation results

T	$\widehat{\mu}$	$\widehat{\delta}$	$\widehat{\theta}$		$\mathcal{LL}$
H	-0.0084	1.465	0.2319		-7315.7
K	-0.0264	1.285	0.3918		-7320.1
J	-0.0201	1.385	0.4059		-7311.2
Student-t	-0.0172	1.403	0.3502	$\widehat{\nu} = 5.5$	-7311.2
GED	-0.0200	1.311	0.4732	$\widehat{\nu} = 1.31$	-7309.3
GSH	-0.0187	1.394	0.4266	$\widehat{t} = -1.45$	-7311.2
Meixner	-0.0188	1.393	0.4339	$\widehat{\lambda} = 0.53$	-7311.2

References

[1] Box, G. E. P. and G. C. Tiao (1962). A further Look at Robustness via Bayes Theorem. *Biometrika*, **3/4**(49), 419-432.

[2] Fischer, M., Horn, A. and I. Klein (2006). Tukey-type Distributions in the context of Financial Data. *Communications in Statistics: Theory and Methods*, accepted for publication.

[3] Fischer, M. and I. Klein (2004). Kurtosis Modelling by means of the J-Transformation. *Allgemeines Statistisches Archiv* **88**(1), 35-50.

[4] Harkness, W. L. and M. L. Harkness (1968). Generalized Hyperbolic Secant Distributionss. *Journal of the American Statistical Association* **63**, 329-337.

[5] Klein, I. and M. Fischer (2006). Power Kurtosis Transformations: Definition, Properties and Ordering. *Allgemeines Statistisches Archiv* **90**(3), 395-402.

[6] MacGillivray, H. L. and W. Cannon (1997). Generalizations of the g-and-h distributions and their uses. *Unpublished*.

[7] Meixner, J. (1934). Orthogonale Polynomsysteme mit einer besonderen Gestalt der erzeugenden Funktion. *Journal of the London Mathematical Society* **9**, 6-13.

[8] Rayner, G. D. and H. L. MacGillivray (2002). Numerical maximum likelihood estimation for the g-and k and generalized g-and h distributions. *Statistics and Computing* **12**, 57-75.

[9] Subbotin, M. T. (1923). On the Law of Frequency of Errors. *Mathematicheskii Sbornik* **31**, 296-301.

[10] Tukey, J.W. (1960). *The Rractical Relationship between the common Transformations of Counts of Amounts*. Princeton University Statistical Techniques Research Group, Technical Report No. 36.

[11] Vaughan, D. (2002). The Generalized Secant Hyperbolic Distribution and its Properties. *Communications in Statistics: Theory and Methods* **31**(2), 219-238.

5 Supplements

Vgl. Klein & Fischer (2006):

$$W(z) + zW'(z) \geq 0, \quad z \geq 0 \text{ proper distribution}$$

$$2W'(z) + zW''(z) \geq 0, \quad z \geq 0 \text{ more kurtosis}$$

5.1 Kurtosis Ordering

Due to theorem 1 of Klein & Fischer (2006), the parameter of the generalized kurtosis transformation, say θ, can be interpreted as kurtosis parameter, if

$$\frac{T''(x;\theta_2)}{T''(x;\theta_1)} \geq \frac{T'(x;\theta_2)}{T'(x;\theta_1)} \geq \frac{T(x;\theta_2)}{T(x;\theta_1)} \tag{5.1}$$

provided that $T(x;\theta) > 0$, $T'(x;\theta) > 0$ and $T''(x;\theta) > 0$.

Plugging T_{GSH} into (5.1), we have to show that

$$1 \geq 1 \geq \frac{\cosh(x) + a(t_1)}{\cosh(x) + a(t_2)}$$

This is true because $a(t) = \cos(t)$ is strictly monotone increasing on $(-\pi, 0]$ and $a(t) = \cosh(t)$ is also strictly monotone increasing on $(0, \infty]$.

Lemma 5.1. *The parameter t of the GSH-transformation is kurtosis parameter in the sense of van Zwet (????). The higher t, the higher is the kurtosis.*

Note: GED and Student-t cannot be ordered according to van Zwet.

5.2 Psi-functions

$$\psi_t(x) = \frac{x\,(\nu + 1)}{\nu + x^2}$$

$$\psi_{GSH}(x) = \frac{\sinh(x)}{\cosh(x) + a(t)}$$

Beachte für die erste bzw. zweite Ableitung gilt

$$T'(x) = T(x)\psi(x)$$

$$T''(x) = T(x)(\psi(x)^2 + \psi'(x)).$$

For the Meixner transformation we have $\psi_{Meixner}(x) = H(x)$.