UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Center For International Economics

Working Paper Series

Working Paper No. 2021-04

Uni- and multivariate extensions of the sinh-arcsinh normal distribution applied to distributional regression

Yuanhua Feng and Wolfgang Karl Härdle

May 2021

Center For International Economics

Uni- and multivariate extensions of the sinh-arcsinh normal distribution applied to distributional regression

Yuanhua Feng[a] and Wolfgang Karl Härdle[b]
[a]Department of Economics, Paderborn University
[b]School of Business and Economics, Humboldt University Berlin

September 21, 2021

Abstract

This paper introduces first an extended SAN (sinh-arcsinh normal) family of distributions by allowing the transformed normal random variable to be unstandardized. A Log-SAN transformation for non-negative random variables and the associate Log-SAN family of distributions are then proposed. Properties of those distributions are investigated. A maximum likelihood estimation procedure is proposed. A chain mixed multivariate extension of the SAN distributions and a corresponding distributional regression model are then defined. Those approaches can help us to discover possible spurious or hidden bimodal property of a multivariate distribution. The proposals are illustrated by different examples.

Keywords: Extended SAN distributions, Log-SAN distribution, MLE, chain mixed multivariate distributions, distributional regression, spurious and hidden bimodality

1 Introduction

Consider first a univariate real-valued random variable Y. According to the SAN (sinh-arcsinh normal) distribution proposed by Jones and Pewsey (2009, 2019), it is assumed that there exist four transformation parameters $\xi, \beta \in \mathbb{R}$ and $\eta, \alpha > 0$ such that

$$Z = S[(Y - \xi)/\eta] = \sinh[\alpha \sinh^{-1}[(Y - \xi)/\eta] - \beta] \tag{1}$$

is $N(\mu, \sigma^2)$ distributed, where ξ, η, β and α are the location, scale, skewness and shape parameters, respectively, and $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ are two nuisance parameters. A six parameter SAN family of distributions is defined, including normal distributions as special cases with $\alpha = 1$ and $\beta = 0$. The original proposal of Jones and Pewsey (2009), hereafter JP09, is a four parameter subfamily assuming $Z \sim N(0, 1)$, which will be called the standard SAN family of distributions. Those distributions are widely employed in theory and application (see e.g. Duerinckx et al., 2014; Stasinopoulos et al., 2017; Hothorn et al., 2018; Jones et al., 2019; Fasiolo et al., 2020). Advantages of SAN distributions compared to related proposals are indicated in JP09. Now, skewness, shape and heavy- or light-tails can be modeled simultaneously. A six parameter SAN distribution can also be bimodal. Those distributions can be estimated by maximum likelihood. See Risco et al. (2011), and Pewsey and Abe (2015) for further extensions of the SAN distributions.

Our main purpose is the modeling of a non-negative random variable $X \geq 0$ with $X > 0$ a.s. In light of the transformation models (e.g. Box and Cox, 1964; Collins, 1991; Cheng et al., 1997; Foster et al., 2001; Hothorn et al., 2014) we focus on random variables, which can be transformed monotonically to a (nearly) normal distribution by combining a pilot transformation for non-negative random variables with the SAN approach. In particular, the application of the SAN transformation to the log-data and the associate distributions will be studied in detail. Let $Y = \ln(X)$, this leads to a Log-SAN transformation for $X \geq 0$:

$$Z = T(X) = S\{[Y - \xi]/\eta\} = \sinh\{\alpha \sinh^{-1}[(\ln(X) - \xi)/\eta] - \beta\}. \tag{2}$$

A Log-SAN distribution is defined assuming $Z \sim N(\mu, \sigma^2)$. It is a wide extension of the log-normal distribution, including the latter as a special case with $\alpha = 1$ and $\beta = 0$. A closely related proposal is the Birnbaum-Saunders (BS) distribution (Birnbaum-Saunders, 1969), i.e. a log-sinh normal distribution (Rieck and Nedelman, 1991). A log-archsinh normal distribution is defined by applying Johnson's S_U distribution (Johnson, 1949) to the log-data. The former has extremely lighter tails with existing mgf (moment generating function), while the latter is extremely heavy-tailed without any finite moment. The Log-SAN distributions have clear advantages compared to them. If $\alpha > 0.5$, their properties are similar to those of the log-normal distributions. They have finite moments of all orders, but their mgf does not exist. Their tailwights are controlled by α and are the same as those of a log-normal distribution, if $\alpha = 1$.

Consider now a p-dimensional real-valued random vector $\mathbf{Y} = (Y_1, ..., Y_p)'$, $p > 1$. Some of its elements may be obtained from non-negative random variables through a transformation. A multivariate SAN distribution is defined in JP09 assuming that the elementwise transformaed random variables $Z_i = S(Y_i)$, $i = 1, ..., p$, are multivariate normal random variables. It is found that this definition should not be used, when the marginal and conditional distributions of the elements are of different shapes. Consider a bivariate bimodal distribution, where Y_1 and Y_2 are linearly dependent. Now, both marginal distributions will be bimodal, but the distribution of e.g. $Y_2|Y_1$ is unimodal. The bimodality in Y_2 is a spurious distributional property caused by Y_1 or vice versa. Now, the elementwise transformation will introduce two artificial modes into the joint distribution, so that it is no longer bimodal. To overcome this problem, a chain mixed multivariate distribution (CMMD) is proposed. The joint density is defined as a chain product of the first marginal and subsequent chain conditional distributions, assuming that the relationships among all elements are of known forms, where the distributions of the elements may come from different families. This definition is based on a dimension reduction rule and can be easily applied to high dimensional cases. Moreover, parametric conditional distributions can be investigated following this definition, if $\mathbf{Y}$ is partitioned into two parts. In particular, the last step in this approach defines an explicit distributional regression of Y_p on Y_1, ..., Y_{p-1}. See e.g. Chernozhukov et al. (2013), and Rothe and Wied (2013) for recent studies on distributional regression. A fixed design distributional regression with SAN errors is also proposed. It is shown that this model can help us to discover hidden bimodality, which cannot be observed from the marginal distribution. These proposals provide parametric alternatives to the commonly used nonparametric approaches for conditional distributions (Hall and Müller, 2003; Hall et al., 2004) or to quantile regression (Koenker and Bassett 1978; Koenker, 2005; Wei and He, 2006). See Koenker et al. (2013) for a detailed discussion on the relationship between distributional regression and quantile regression.

Maximum likelihood procedures for estimating the SAN and Log-SAN distributions are developed. An ad hoc multistep procedure is proposed for practical implementation of the multivariate extensions. The proposals are applied to different examples and compared to related approaches. In particular, spurious bimodality is shown by the Old Faithful Geyser data in Härdle (1991) and hidden bimodality is illustrated by the Italian GDP growth panel from 1951 to 1998 (Baiocchi, 2006). Finally, the SAN and Log-SAN transformations can be applied to define new subordinated long memory Gaussian processes (Beran et al., 2013; Papiras and Taqqu, 2017). Discussion on those topics will be carried out elsewhere.

The paper is organized as follows. The SAN and Log-SAN distributions are studied in Sections 2 and 3. Maximum likelihood estimation is discussed in Section 4. Multivariate extensions of the proposals are introduced in Section 5. Section 6 reports the application. Final remarks in Section 7 close the paper. Proofs of results are put in the appendix.

2 Properties of the canonical families

The canonical SAN- and Log-SAN distributions with $\xi = 0$ and $\eta = 1$ will be discussed in detail. Now, the inverse of $S(Y)$, $Y = K(Z) = S^{-1}(Z)$ with $K(Z) = \sinh[\alpha^{-1}\sinh^{-1}(Z)+\beta/\alpha] = S(Z, \alpha^{-1}, -\beta/\alpha)$ is another SAN transformation with parameters α^{-1} and $-\beta/\alpha$. The canonical Log-SAN transformation has the following equivalent forms

$$\begin{aligned} T(X) &= \frac{1}{2}\left[e^{-\beta}\{[1+\ln^2(X)]^{1/2}+\ln(X)\}^{\alpha} - e^{\beta}\{[1+\ln^2(X)]^{1/2}+\ln(X)\}^{-\alpha}\right] && (3) \\ &= \frac{1}{2}\left[e^{-\beta}\{[1+\ln^2(X)]^{1/2}+\ln(X)\}^{\alpha} - e^{\beta}\{[1+\ln^2(X)]^{1/2}-\ln(X)\}^{\alpha}\right]. && (4) \end{aligned}$$

The inverse canonical Log-SAN transformation $X = G(Z) = \exp[K(Z)]$ is given by

$$\begin{aligned} G(Z) &= \exp\{\sinh[\alpha^{-1}\sinh^{-1}(Z)+\beta/\alpha]\} \\ &= \exp\left(\frac{1}{2}\left\{e^{\beta/\alpha}[(1+Z^2)^{1/2}+Z]^{1/\alpha} - e^{-\beta/\alpha}[(1+Z^2)^{1/2}+Z]^{-1/\alpha}\right\}\right) && (5) \\ &= \exp\left(\frac{1}{2}\left\{e^{\beta/\alpha}[(1+Z^2)^{1/2}+Z]^{1/\alpha} - e^{-\beta/\alpha}[(1+Z^2)^{1/2}-Z]^{1/\alpha}\right\}\right). && (6) \end{aligned}$$

Those equivalent formulas are useful for simulation and further statistical inferences.

The density function of the canonical SAN family is

$$f_{CS}(y) = \frac{\alpha}{\sigma\sqrt{2\pi(1+y^2)}}C(y)\exp\left\{-\frac{[S(y)-\mu]^2}{2\sigma^2}\right\} \qquad (7)$$

with $C(y) = \cosh[\alpha \sinh^{-1}(y) - \beta]$, which reduces to that given in (2) of JP09, if $\mu = 0$ and $\sigma^2 = 1$. It is easy to see that f_{CS} is symmetric, if $\beta = \mu = 0$. For $\beta \neq 0$ or $\mu \neq 0$, the two densities $f_{CS,\beta,\mu}(y)$ and $f_{CS,-\beta,-\mu}(y)$ are such that $f_{CS,-\beta,-\mu}(y) = f_{CS,\beta,\mu}(-y)$. The shape of the SAN distributions will be strongly affected by σ^2. Moreover, μ will affect the skewness and the shape of f_{CS}, too. The skewness parameter β can also have some effect on the shape of this distribution. The distribution function associated with f_{CS} in (7) is $F_{CS}(y) = \Phi\{[S(y) - \mu]/\sigma\}$ with the quantile function $F_{CS}^{-1}(u) = K[\Phi^{-1}(u) * \sigma + \mu]$, $0 < u < 1$, and the median $y_M = K(\mu) = \sinh\{[\sinh^{-1}(\mu) + \beta]/\alpha\}$.

Let $X \geq 0$ be a non-negative random variable and $Y = \ln(X)$ follow a canonical SAN distribution, X has a canonical Log-SAN distribution with the density

$$f_{CL}(x) = \frac{\alpha}{\sigma\sqrt{2\pi}x\sqrt{1+\ln^2 x}} C(\ln x) \exp\left\{-\frac{[S(\ln x) - \mu)^2}{2\sigma^2}\right\}, \tag{8}$$

which reduces to a log-normal density, if $\beta = 0$ and $\alpha = 1$. The distribution and quantile functions are now $F_{CL}(x) = \Phi\{[S(\ln(x)) - \mu]/\sigma\}$ and $F_{CL}^{-1}(u) = G[\Phi^{-1}(u) * \sigma + \mu]$, $0 < u < 1$, with the median $x_M = G(\mu) = \exp\big(\sinh\{[\sinh^{-1}(\mu) + \beta]/\alpha\}\big)$.

The canonical SAN distribution shares some properties of its two-parameter standard counterpart. However, the former can be bimodal, while the latter is always unimodal. Consider first the tail behaviors of the proposed distributions. As for the original proposal in JP09, the tailweights of f_{CS} are of the order $O[\exp(-|y|^{2\alpha})]$, which are not affected by μ and σ. Hence, the tailweights of those distributions are the same as those of the generalized normal distribution with a tail-thickness 2α. Compared to the normal distribution, it is with heavier tails, if $\alpha < 1$ and lighter tails, if $\alpha > 1$. In particular, we have

Theorem 1. *The tailweights and moment properties of a SAN or a Log-SAN distributed random variable, denoted by Y and X, respectively, are quantified as follows*

i) *For any $\alpha > 0$, Y has finite moments of all orders with lighter tails, if $\alpha > 1$ and heavier tails, if $\alpha < 1$, compared to those of a normal distribution.*

ii) *If $\alpha < 0.5$ and $u > 0$, $E(X^u) = \infty$. For $\alpha > 0.5$, the mgf of Y exists on $I\!R$. Now, X has finite moments of all orders, but its mgf does not exist at any $u > 0$.*

For the SAN distributions, any $\alpha > 0$ is meaningful. If $\alpha = 1$, the tailweight of the SAN distributions is the same as that of a normal distribution. For a Log-SAN distribution we assume that $\alpha > 0.5$, implying $E(X^k) < \infty$ for any $k \in \mathbb{R}$. Formulas of the moments of the standard canonical SAN family are obtained by JP09. Their results can be extended to the case with $\mu = 0$ and $\sigma^2 > 0$.

Proposition 1. *The moments of $f_{CS}(y)$ with $\mu = 0$ are*

$$E(Y^k_{\alpha,\beta,\sigma}) = \frac{1}{2^k}\sum_{r=0}^{k}\binom{k}{r}(-1)^r \exp\left\{(k-2r)\frac{\beta}{\alpha}\right\} Q_{(k-2r)/\alpha}, \tag{9}$$

where

$$Q_\delta = \frac{e^{1/(4\sigma^2)}}{\sqrt{8\pi\sigma^2}}\{K_{(\delta+1)/2}[1/(4\sigma^2)] + K_{(\delta-1)/2}[1/(4\sigma^2)]\} \tag{10}$$

and K_ν is the modified Bessel function of the second order with $K_\nu(u) = K_{-\nu}(u)$.

It is easy to see that $Q_\delta = Q_{-\delta}$ and $E(Y^k_{\alpha,0,\sigma}) = 0$, if k is odd. Results for the SAN family with $\mu \neq 0$ or for the Log-SAN family are too complex and are omitted.

Furthermore, β acts as a skewness parameter in a SAN or Log-SAN distribution.

Theorem 2. *If other parameters are fixed, in the SAN and Log-SAN distributions β acts as a skewness parameter in the sense of van Zwet's (1964) skewness ordering.*

This fact is not affected by μ and σ^2. That is, if the other parameters are fixed, a SAN distribution with a bigger β is more positively skewed. This is also true for the Log-SAN distributions, although it is always positively skewed, its level of skewness increases with β. It can be shown that μ will also affect the skewness of f_{CS} and f_{CL}. However, it is not a skewness parameter. Some properties of the symmetric SAN family with $\beta = \mu = 0$, $f_{CS0}(y)$ say, including the special case with $\sigma^2 = 1$ studied in JP09, are now stated.

Theorem 3. *For the symmetric subfamily f_{CS0} with $\beta = 0$ and $\mu = 0$ we have*

i) *α acts as a kurtosis parameter in the sense of van Zwet's (1964) kurtosis ordering.*

ii) *f_{CS0} is unimodal, if $\sigma^2 \leq 1$ or $\alpha \leq 1$. A necessary and sufficient condition for the unimodal property of f_{CS0} is $\alpha^2(1 - \sigma^{-2}) \leq 1$.*

The result in i) is an extension of the finding in JP09, which indicates that the kurtosis of a distribution in this subfamily decreases with α. The kurtosis parameter is defined for a symmetric distribution and is irrelevant for the Log-SAN distribution. Without the restriction $\sigma^2 = 1$, f_{CS0} can be sometimes bimodal. We see, $\alpha \leq 1$ or $\sigma^2 \leq 1$ is sufficient for the unimodality of f_{CS0}. The necessary and sufficient condition for this is $\alpha^2(1-\sigma^{-2}) \leq 1$. Note that $f'_{CS0}(0) = 0$ is always true. The sign of $f''_{CS0}(0)$ is the same as that of $\alpha^2(1-\sigma^{-2}) - 1$. Hence, f_{CS0} has a single mode at $y = 0$, if $\alpha^2(1-\sigma^{-2}) < 1$. And, it is with a flat peak, if $\alpha^2(1-\sigma^{-2}) = 1$. If $\alpha^2(1-\sigma^{-2}) > 1$, we have $f''_{CS0}(0) > 0$ so that f_{CS0} has a local minimum at $y = 0$ with two symmetric modes around zero. A heavy-tailed distribution in this subfamily with $\alpha < 1$ can only be unimodal.

Finally, consider the asymmetric subfamily with $\mu = 0$ but $\beta \neq 0$, denoted by $f_{CSb}(y)$ with $f_{CSb,-\beta}(y) = f_{CSb,\beta}(-y)$, which also includes the standard SAN family as a special case. Let $y_\beta = \sinh(\beta/\alpha)$, we have $y_\beta < 0$ for $\beta < 0$ and $y_\beta > 0$ for $\beta > 0$. For this subfamily only the following fact is stated.

Proposition 2. *If $\sigma^2 \leq 1$, the asymmetric density function $f_{CSb}(y)$ is always unimodal with a mode at y_P say, which lies between zero and y_β.*

It confirms the finding in JP09 that a standard SAN distribution can only be unimodal. Now, $\alpha^2(1-\sigma^{-2}) \leq 1$ is still sufficient but not necessary for the unimodality of $f_{CSb}(y)$. A necessary and sufficient condition for this property seems to be $(1+y_\beta^2)\alpha^2(1-\sigma^{-2}) \leq 1$. Further discussion on those topics is omitted.

Selected SAN (left) and corresponding Log-SAN (right) densities are shown in Figure 1, where each window is with four densities displayed in solid, dashed, dashed and dotted and long-dashed lines. Those in Figure 1(a) are symmetric with $\alpha = 0.65$, 0.80, 3 and 6, and $\sigma = 0.5$, 0.4, 1.1 and 1.8. The first two are heavy-tailed. The third is light-tailed with $\alpha = 3$, but still unimodal. The last is strongly light-tailed and bimodal. The bimodal feature of a SAN density can be taken over by its log-counterpart. The densities in (c) are asymmetric with $\mu = 0$ and similar tail properties as those in 1(a), where the parameters $\alpha = 0.65$, 0.8, 3 and 7, $\beta = 0.5$, 0.5, 2 and 6, and $\sigma = 0.6$, 0.4, 1 and 2, are used. Those in (e) are all heavy-tailed with $\alpha = 0.75$, $\beta = 0.75$ and $\sigma = 0.25$ fixed, where $\mu =$-0.4, -0.2, 0 and 0.2 are used to show the effect of μ on the skewness and scale of those densities. Now, the scale of a distribution increases and its skewness decreases with μ.

3 Complete SAN and Log-SAN families

In practice, six parameter SAN and Log-SAN distributions obtained by the canonical SAN transformation of $\tilde{Y} = (Y - \xi)/\eta$ with $\xi \in \mathbb{R}$ and $\eta > 0$ should be used. Those two families are closed for location-scale or scale-power changes, respectively. Their properties will be discussed briefly. The density function of the six-parameter SAN family is

$$\begin{aligned} f_S(y) &= \frac{\alpha}{\eta\sigma\sqrt{2\pi}\sqrt{1+[(y-\xi)/\eta]^2}} C[(y-\xi)/\eta] \\ & \quad \exp\left(-\frac{\{S[(y-\xi)/\eta]-\mu\}^2}{2\sigma^2}\right). \end{aligned} \tag{11}$$

That is $f_S(y) = \eta^{-1} f_{CS}(\tilde{y})$ with $\tilde{y} = (y - \xi)/\eta$. The distribution function associated with the density in (11) is $F_S(y) = \Phi\{[S((y-\xi)/\eta)-\mu]/\sigma\}$ with the quantile function $F_S^{-1}(u) = \eta K[\Phi^{-1}(u) * \sigma + \mu] + \xi$, $0 < u < 1$, and median $y_M = \eta K(\mu) + \xi = \eta \sinh\{[\sinh^{-1}(\mu) + \beta]/\alpha\} + \xi$. The density function of the corresponding Log-SAN family is

$$\begin{aligned} f_L(x) &= \frac{\alpha}{\eta\sigma\sqrt{2\pi}x\sqrt{1+[(\ln x-\xi)/\eta]^2}} C[(\ln x-\xi)/\eta] \\ & \quad \exp\left(-\frac{\{S[(\ln x-\xi)/\eta]-\mu\}^2}{2\sigma^2}\right). \end{aligned} \tag{12}$$

Again, $f_L(x)$ is obtained from $f_{CL}(x)$ with $\ln x$ being replaced by $(\ln x - \xi)/\eta$ and rescaled by η^{-1}. The distribution function of the Log-SAN family is $F_L(x) = \Phi\{[S((\ln(x)-\xi)/\eta)-\mu]/\sigma\}$ with the quantile function $F_L^{-1}(u) = e^{\xi}\{G[\Phi^{-1}(u)*\sigma+\mu]\}^{\eta}$, $0 < u < 1$, and median $x_M = e^{\xi}[G(\mu)]^{\eta} = \exp\left(\eta \sinh\{[\sinh^{-1}(\mu) + \beta]/\alpha\} + \xi\right)$. Assuming $Z \sim N(0,1)$, the above density functions reduce to those of the standard SAN and Log-SAN families. The former was that used in JP09.

It is easy to see that $f_S(y)$ is symmetric, if ξ, μ and β are all zero. Although $f_S(y)$ is a wide extension of $f_{CS}(y)$, its shape, skewness, modal property and its kurtosis in the symmetric subfamily are all not affected by ξ and η. Moreover, the main results in Section 2 on the canonical SAN and Log-SAN families hold for the current cases. The tailweights of those distributions are still determined by α, such that all results in Theorem 1 hold for the complete SAN and Log-SAN distributions. Moreover, if other parameters in the SAN or Log-SAN distributions are fixed, β acts as a skewness parameter. For the symmetric SAN subfamily with $\beta = 0$, $\mu = 0$ and $\xi = 0$, α is still a kurtosis parameter.

4 Maximum likelihood estimation

For given observations y_i, $i = 1, ..., n$, let $\tilde{y}_i = (y_i - \xi)/\eta$ and $\boldsymbol{\theta} = (\xi, \eta, \alpha, \beta, \mu, \sigma^2)'$. The log-likelihood function for estimating a SAN distribution is

$$L(\boldsymbol{\theta}) = n \ln(\alpha/\eta/\sqrt{2\pi}/\sigma) + \sum_{i=1}^{n} \{\ln[C(\tilde{y}_i)/\sqrt{1+\tilde{y}_i^2}] - [S(\tilde{y}_i) - \mu]^2/2/\sigma^2\}. \qquad (13)$$

Set $\mu = 0$ and $\sigma^2 = 1$, (13) reduces to that for estimating a standard SAN distribution with the parameter vector $\boldsymbol{\theta}_4 = (\xi, \eta, \alpha, \beta)$. For a positive random variable with observations $x_1, ..., x_n$, a Log-SAN distribution can be fitted. Let $y_i = \ln(x_i)$, the log-likelihood function for a Log-SAN distribution is given by

$$L_L(\boldsymbol{\theta}) = n \ln(\alpha/\eta/\sqrt{2\pi}/\sigma) + \sum_{i=1}^{n} \{\ln[C(\tilde{y}_i)/\sqrt{1+\tilde{y}_i^2}] - [S(\tilde{y}_i) - \mu]^2/2/\sigma^2\} - \sum_{i=1}^{n} \ln(x_i), \quad (14)$$

where the notations are as defined before. The log-likelihood $L_L(\boldsymbol{\theta})$ in (14) differs to that in (13) in two ways: 1) The term $y_{x,i}$ is now defined based on $\ln(x_i)$. And 2) There is an additional component in this function reflecting the contribution of x_i to the log-likelihood, which is a constant given the data and will not affect the solutions. However, the log-likelihood function of the Log-SAN distribution should be calculated using $L_L(\boldsymbol{\theta})$ so that it is comparable to that of a SAN distribution fitted to x_i. The elements of the observed information matrix are given in an online supplement. In this paper σ instead of σ^2 is treated as the targeted parameter. It should be indicated that although $\hat{\mu}$ and $\hat{\sigma}$ are still asymptotically independent, both are correlated to the other estimators.

For practical implementation, the optimization is done by the default option of the R function 'optim', i.e. by a direct search using the simplex algorithm of Nelder-Mead (Nelder and Mead, 1965). A standard SAN distribution will be fitted using the naive initial vector $(\bar{y}, \hat{\sigma}_y, 1, 0)$. For the complete SAN family the score function given in the supplement shows that for given $\boldsymbol{\theta}_4$ the tow nuisance parameters μ and σ can be estimated from the transformed data. This fact is applied in the developed algorithm so that the number of directly searched parameters in this case is still 4. It is shown that this simplified optimization algorithm runs more stably than a full search procedure. The two algorithms will result in the same estimates, up to some negligible numerical differences, provided that both procedures converge. The above-mentioned naive initial vector can usually be used for fitting a six-parameter SAN distribution. If this fails to work, one can for instance try to use the estimates of a standard SAN distribution as the initial values in this stage. An auxiliary procedure is also developed for finding suitable initial values by maximizing the p-value of the Shapiro and Wilk (1965) normality test.

5 Multivariate and conditional extensions

Now, we will extend the SAN distributions to a random vector $\mathbf{Y} = (Y_1, ..., Y_p)'$. For simplicity, it is assumed that suitable transformations of all or some elements are done beforehand, if necessary. A multivariate SAN (MSAN) distribution can be defined by assuming that $\mathbf{Z} = \mathbf{S} = (S_1, ..., S_p)'$ is jointly normally distributed with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{V}$, where S_i is the SAN transformation of Y_i. This is an extension of the proposal in JP09. However, this definition should not be used, if the marginal and conditional distributions of the elements are with different shapes, e.g. if one is unimodal and the other is bimodal. Now, the correct joint distribution cannot be obtained based on elementwise transformations. Hence, we propose an alternative multivariate extension of the SAN distributions defined based on the following chain rule

$$f_{\mathbf{Y}}(y_1, ..., y_p) = f_{Y_1}(y_1) f_{Y_2|Y_1}(y_2|y_1)...f_{Y_p|Y_1,...,Y_{p-1}}(y_p|y_1, ..., y_{p-1}), \tag{15}$$

provided that the order is known and that the first marginal distribution, the subsequent conditional distributions and the required relationships are all of known forms. In this paper, we assume that the univariate distributions are members of the SAN family or of its subfamilies. A multivariate distribution defined in this way is called a CMMD (chain mixed multivariate distribution). Moreover, we assume that only the conditional mean $\mu_{Y_i}|Y_1, ..., Y_{i-1}$, for $1 < i \leq p$, depends on the previous elements, i.e. $f_{Y_i|Y_1,...,Y_{i-1}}(y_p - \mu_i|y_1, ..., y_{p-1})$, and that $\mu_i|Y_1, ..., Y_{p-1}$ is of a linear form. Now, the covariance matrix of the marginal distributions is preserved. The definitions of the CMMD and the MSAN distributions are equivalent, if all elements are independent SAN random variables.

Definition (15) can be used to study the conditional distributions, if $\mathbf{Y}$ is partitioned into two parts. In particular, the last step on the right-hand side defines a special random design distributional regression. Moreover, we also propose the use of the following fixed design distributional regression

$$Y_i = a_0 + a_1 x_{1i} + ... + a_k x_{ki} + \varepsilon_i, i = 1, ..., n, \tag{16}$$

with iid SAN errors ε_i. Thus, the error distribution is assumed to be known up to some unknown parameters. We will see that Model (15) can help us to discover hidden bimodal property, where the unconditional distribution is unimodal but the conditional distribution is bimodal. An ad hoc multistep procedure is used for practical implementation, where the mean function(s), the marginal and conditional distributions are estimated separately using existing approaches. The development of joint maximum likelihood estimation and further studies on related topics are beyond the aim of this paper.

6 Application

In the sequel, the standard SAN distribution and its extension will be denoted by SA4 and SA6, respectively. We first applied the proposals to the strengths of $n = 63$ glass fibres used in JP09. The data were originally published in Smith and Naylor (1987) and is available in the R package 'ismev'. Now, the SA6 cannot be well estimated and the Log-SA6 has a bigger BIC. The best distribution selected is the Log-SA4 with a log-likelihood $ll = -9.16$. But the 95%-confidence interval of α is around 0.5 so that the log-transformation should not be used. Therefore, the SA4 density with $ll = -10$ as proposed by JP09b remains to be the most reasonable choice. For further application the miscellaneous share (Wmisc) of the Budget Shares for Italian Households from 1973 to 1992 with 1729 observations (Bollino et al., 2000), the Old Faithful Geyser eruption and waiting times (in minutes, called OFGE and OFGW) used in Härdle (1991) with 272 observations, the numbers of applications (Apps), acceptances (Acce) and enrollments (Enro) for 777 US Colleges from the 1995 issue of US News and World Report, see James et al. (2021), as well as the Italian GDP growth panel in 21 regions from 1951 to 1998 (IGDP, millions of Lire, 1990=base) in Baiocchi (2006) are selected. Those datasets are available in the R packages 'np', 'Ecdat' as well as 'ISLR'.

The two examples Wmisc and OFGE were used to show some details in the univariate case, for which normal, SA4 and SA6 distributions are fitted to the original and log-data as well. The estimates, ll, AIC, BIC and p-values of the Shapiro-Wilk (sw-) test of the transformed data are listed in Table 1. For Wmisc, the log-transformation is helpful and the best distribution chosen by the BIC is the Log-SA6. For OFGE, all uni-modal distributions are clearly not suitable, because its distribution is bimodal. Now, the SA6 and Log-SA6 distributions have almost the same ll and BIC. Therefore, we propose to use the SA6 distribution without the log-transformation. The p-value of the normality test for both selected distributions (bold marked) is much bigger than 5%. Thus, those distributions fit the data well, for which the standard deviations of the estimated parameters are also given. We see, the errors of the estimates are sometimes very large, in particular for OFGE. The estimation quality of μ and σ is strongly affected by the other parameters. The extended BS distribution proposed by Vilca et al. (2011) based on the skewed normal distribution of Azzalini (1985) was fitted using the R package 'bssn' as a comparison. The achieved log-likelihoods are $ll = -2323.8$ and -438.2 for Wmisc and OFGE, respectively, indicating that this distribution should not be used for those examples. For OFGE, a two component normal mixture distribution is further fitted, which achieves $ll = -276.4$ and is not a reasonable choice. Histograms for the log-data of Wmisc and the original data of OFGE together with the corresponding densities are displayed in Figures 2(a) and (b).

The joint distribution of OFGE (x) and OFGW (y) is bimodal. Furthermore, the simple linear regression $\hat{y}_i = 33.47 + 10.73x_i$ with $r^2 = 0.812$ fits the relationship between them very well. The observations of OFGW and the residuals are displayed in Figures 2 (c) and (d). Although the marginal histogram of OFGW is also bimodal, the residuals are clearly unimodal. This indicates that the bimodality of OFGW can be thought of as a spurious distributional property caused by that of OFGE. The best distribution selected for the residuals is just a zero mean normal one with $\hat{\sigma}^2 = 34.72$. The finally fitted bivariate SA6-normal distribution with $ll = -1127$ is displayed in Figure 2(e), which provided a satisfactory parametric alternative to the nonparametric results for this dataset as given in Härdle (1991), and Hafield and Racine (2008). The log-likelihood of the bivariate normal distribution fitted to those data is $ll = -1290$, which is much smaller. A bivariate SAN distribution defined by elementwise transformations should not be used for this example, because it will result in either a uni-modal estimate or an estimate with four modes.

Now, we will illustrate the application of the CMMD in high dimensional case using the log-data of Apps (y_1), Acce (y_2) and Enro (y_3) following the logical order. It is found that the relationships between thosee variables are roughly linear, for which we have

$$\underset{sd\ \ (0.055)\ \ (0.007)}{\hat{y}_{2i} = 0.429 + 0.990y_{1i}} \qquad \text{and} \qquad \underset{(0.079)\ \ (0.046)\ \ \ \ (0.050)}{\hat{y}_{3i} = -0.315 + 0.014y_{1i} + 0.898y_{2i}},$$

where y_1 in the second regression is irrelevant but is left in the model. The selected distributions in the three steps are: 1) An SA4 for y_1 with coefficients 7.12, 1.65, 1.40, and 0.228; 2) An SA4 for the residuals of y_2 with coefficients 0.0999, 0.0927, 0.684 and -0.385; and 3) A normal distribution with zero mean and $\hat{\sigma}^2 = 0.0923$ for the residuals of y_3. The scatter plot of y_1, y_2 and y_3 is shown in Figure 2(f), where the values of the joint density are indicated by colors. Further details for this example are omitted.

Finally, the IGDP panel data (x_i) is used to show the application of the fixed design distributional regression with SAN errors and hidden bimodal property. The serial correlation is ignored in the current paper. It is found that the square root transformation is better than the log-one, because the errors in $y_i = \sqrt{x_i}$ are roughly identically distributed, for which the following polynomial regression is selected by the BIC:

$$\begin{array}{llll} \hat{y}_i = 2.479 + 5.60 \cdot 10^{-3} t^2 & - 1.69 \cdot 10^4 t^3 & + 1.55 \cdot 10^{-6} t^4, & t = 1, ..., 48. \\ sd \ \ (0.042) \ \ (4.9 \cdot 10^{-4}) & (2.4 \cdot 10^{-5}) & (3.1 \cdot 10^{-7}) & \end{array}$$

Note that the marginal histogram of y_i were unimodal. The figure for this is omitted. Both of the scatter plot and the histogram of the residuals given in Figures 3(a) and (b) indicate that the conditional distributions are bimodal, which is a hidden distributional property and cannot be discovered without taking the regression function into account. For the residuals, an SA6 distribution with coefficients $\hat{\xi} = -0.201$, $\hat{\eta} = 0.229$, $\hat{\alpha} = 1.732$, $\hat{\beta} = 0.139$, $\hat{\mu} = 2.711$, 8.695 and $ll = -647.7$ was selected. The conditional densities are displayed in Figure 3(c), which are much better than the nonparametric results for this example as shown in Figure 4 of Hafield and Racine (2008). The estimated quantile curves for the original data with $p = 0.05$, 0.25, 0.50, 0.75 and 0.95 are illustrated in Figure 3(d). We see those parametric quantile curves are quite reasonable. Moreover, a two component normal mixture distribution is also fitted to the residuals with $\hat{\mu}_1 = -0.57$, $\hat{\sigma}_1 = 0.232$, $\hat{\mu}_2 = 0.341$, $\hat{\sigma}_2 = 0.282$, $\hat{w} = 0.374$ and $ll = -648.2$. This distribution has one parameter fewer and hence a smaller BIC, which provides another reasonable distributional regression for this example. Further details on this alternative model are omitted to save space. This finding and the related results fitted to OFGE indicate that whether an SA6 distribution performs better than a normal mixture one, depends on the data. If both perform similarly, the former is more preferable, because statistical inferences based on it are more simple.

7 Concluding remarks

In this paper the SAN family of distributions is extended and a novel Log-SAN family of distributions is introduced. Properties of those distributions are investigated. It is proposed to fit such a distribution by maximum likelihood. The proposals are further extended to multivariate cases and distributional regression. An ad hoc multistep estimation procedure is proposed for the practical implementation of those proposals. Application to different examples shows that the proposals are very attractive in theory and practice. In particular, they provide useful alternatives to some well-known nonparametric approaches and quantile regression. They can also help us to discover possible spurious or hidden distributional properties. There are many open questions in this context. In particular, theoretical study on the proposed multivariate distribution and distributional regression should be carried out. Application of the SAN and Log-SAN distributions to long memory time series is also of great interest.

Acknowledgment: This paper was supported by the German DFG Project FE 1500/2-1. Financial support of the European Unions Horizon 2020 research and innovation program FIN- TECH: A Financial supervision and Technology compliance training programme under the grant agreement No 825215 (Topic: ICT-35-2018, Type of action: CSA), the European Cooperation in Science & Technology COST Action grant CA19130 - Fintech and Artificial Intelligence in Finance - Towards a transparent financial industry, the Deutsche Forschungsgemeinschafts IRTG 1792 grant, the Yushan Scholar Program of Taiwan and the Czech Science Foundations grant no. 19-28231X / CAS: XDA 23020303 are greatly acknowledged. The data sources are acknowledged in the context. We are grateful to Dr. Xuehai Zhang, Mr. Sebastian Letmathe, Ms. Shujie Li, Mr. Bastian Schäfer and Mr. Dominik Schulz at the Paderborn University, Germany, for helpful discussions.

Table 1: Estimated parameters and some statistics for selected cases

Data	dis.	$\hat{\xi}$	$\hat{\eta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\mu}$	$\hat{\sigma}$	ll	AIC	BIC	$\hat{p}$
Wmisc	N	—	—	1	0	1.87	1.46	-3102	6208	6219	0.00
(orig.)	SA4	1.18	0.27	0.55	0.49	0	1	-2238	4484	4506	0.00
	SA6	1.26	0.72	0.69	1.13	-1.07	0.89	-2218	4449	4482	0.14
Wmisc	N	—	—	1	0	0.44	0.59	-2293	4590	4601	0.00
(log)	SA4	0.34	0.28	0.66	0.11	0	1	-2220	4448	4470	0.012
	SA6	0.14	0.15	0.69	-0.11	1.13	1.75	-2211	4435	**4467**	0.48
	(sd)	(0.036)	(0.057)	(0.044)	(0.055)	(0.42)	(0.66)				
OFGE	N	—	—	1	0	3.49	1.14	-421.4	846.8	854.0	0.00
(orig.)	SA4	6.04	0.05	1.93	-8.77	0	1	-364.0	735.9	750.4	0.00
	SA6	2.18	1.03	4.82	2.89	16.3	42.8	-258.4	528.8	**550.4**	0.93
	(sd)	(0.44)	(0.34)	(1.02)	(1.53)	(10.4)	(26.6)				
OFGE	N	—	—	1	0	1.19	0.37	-440.9	885.8	893.0	0.00
(log)	SA4	1.72	0.01	1.16	-5.80	0	1	-351.4	710.9	725.3	0.00
	SA6	1.08	0.42	5.10	-0.26	19.8	52.4	-258.0	528.0	549.7	0.93

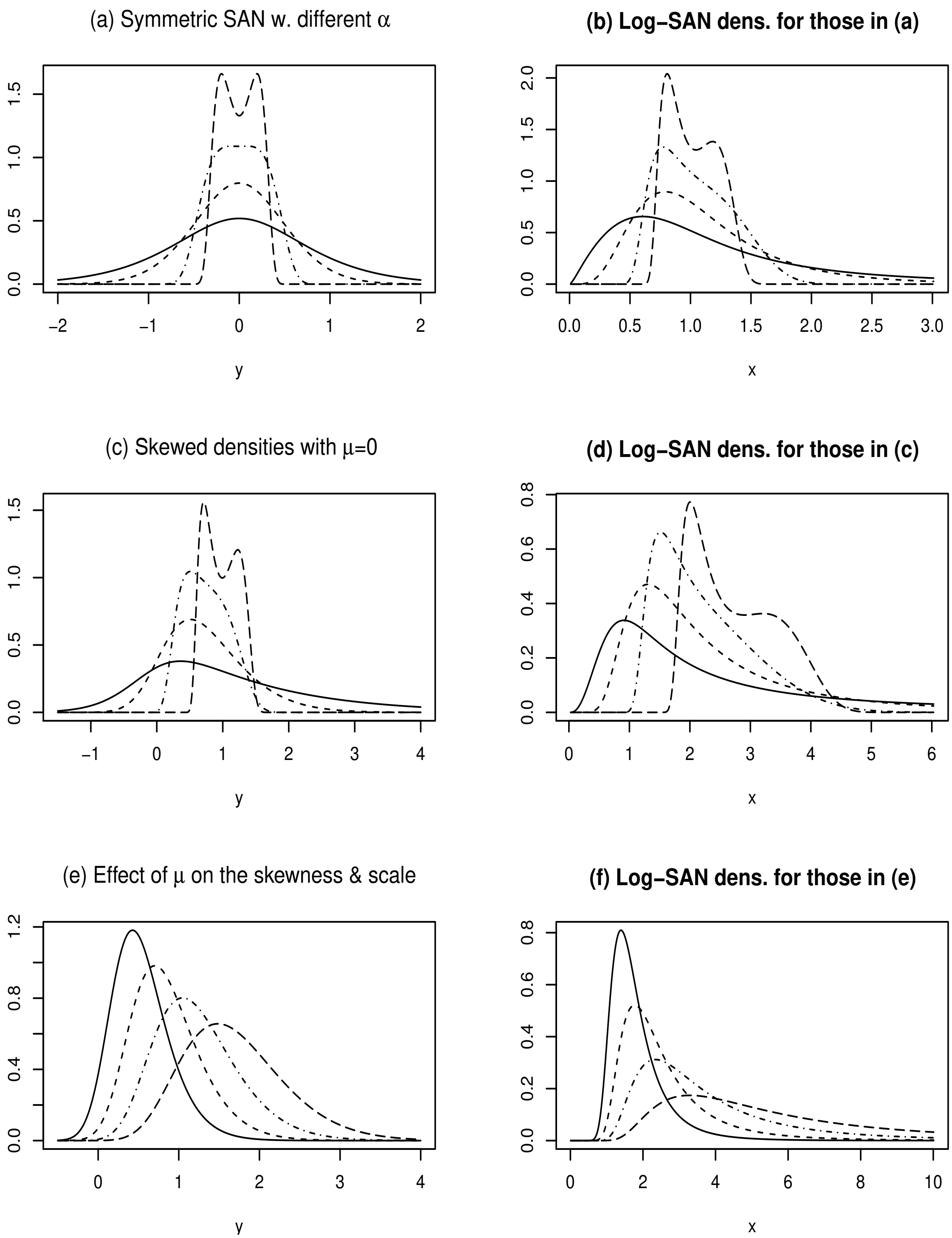

Figure 1: Examples of different SAN (left) and Log-SAN (right)

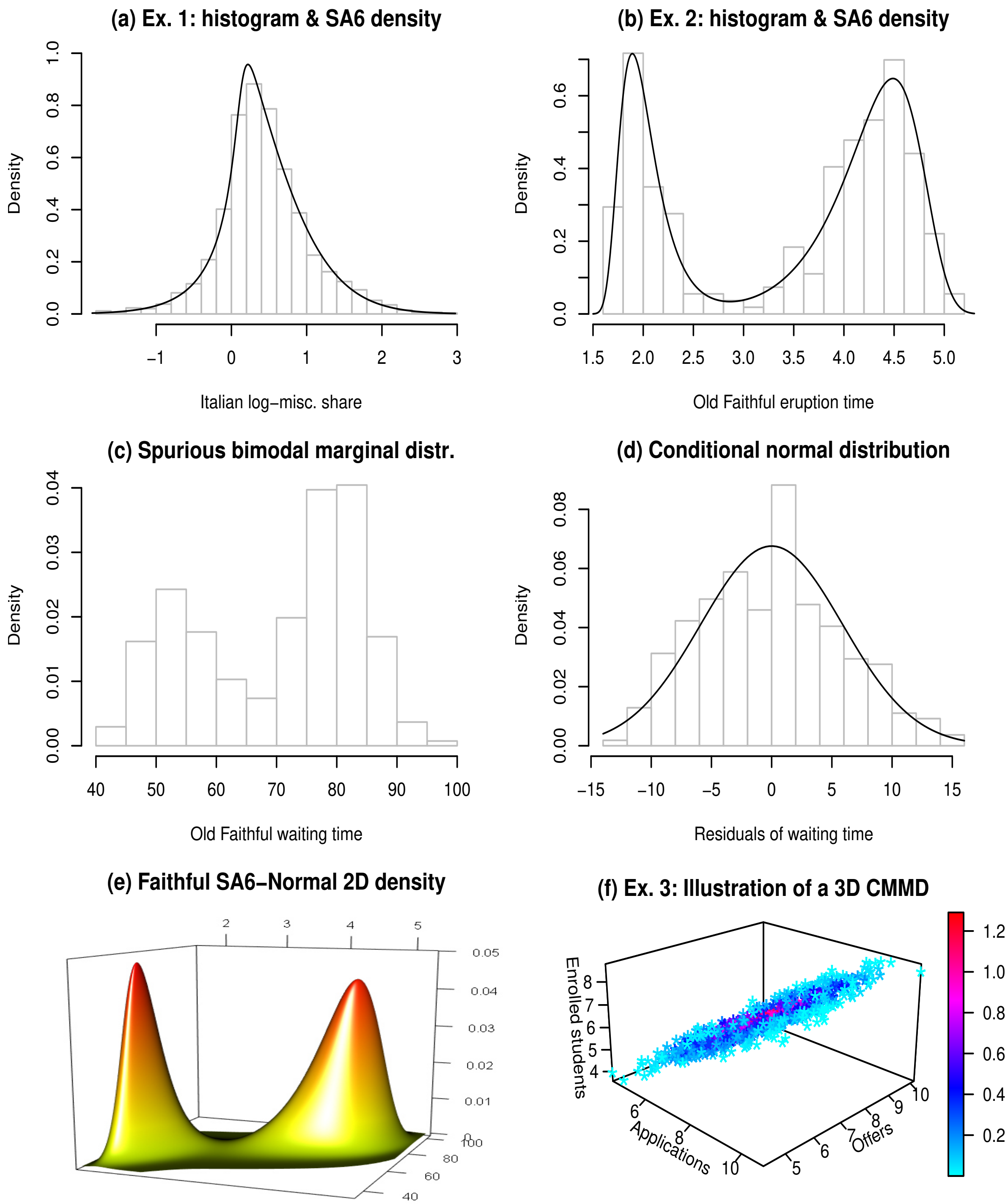

Figure 2: Results for the first two examples

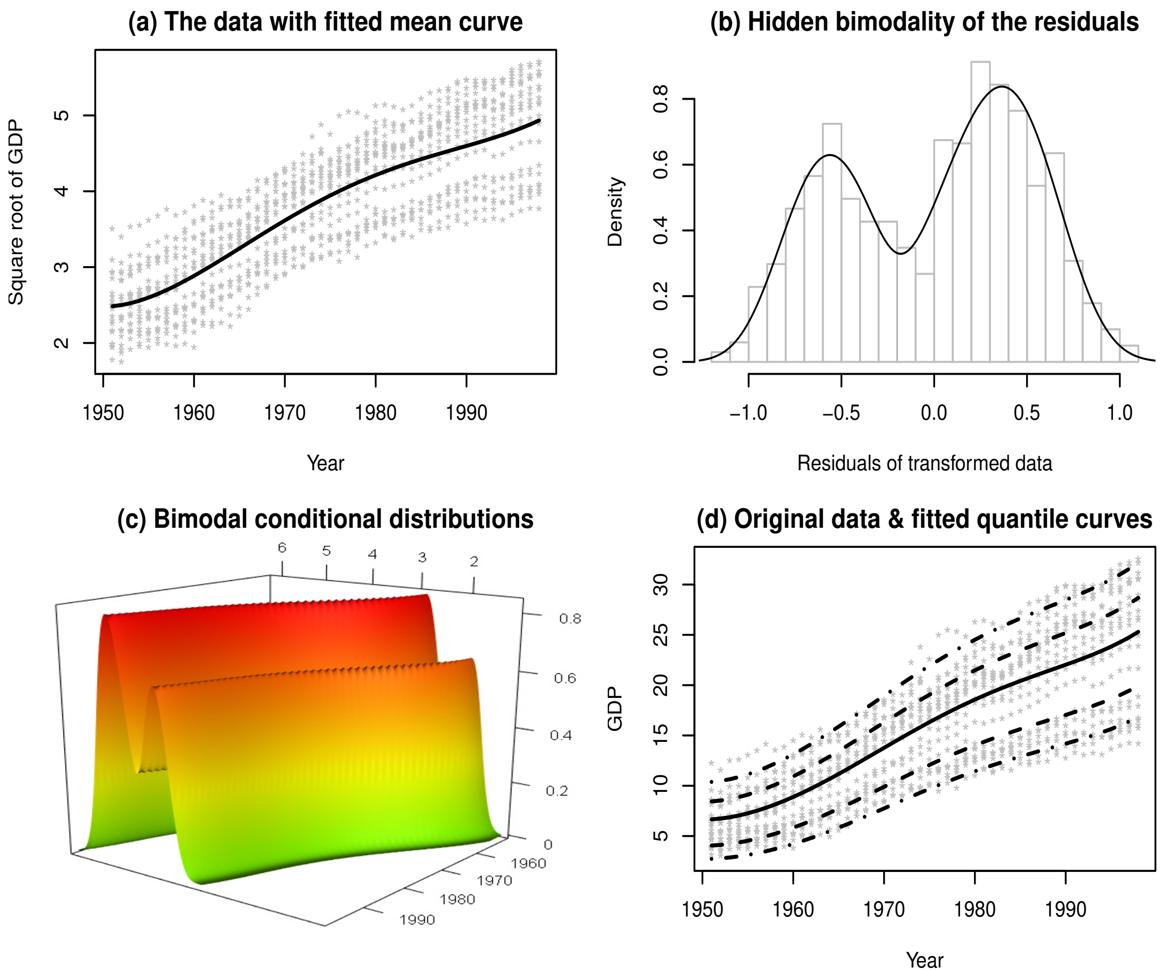

Figure 3: Results for the first two examples

References

Azzalini, A. (1985). A class of distributions which includes the normal ones. Scand. J. Statist. 12, 171-178.

Baiocchi, G. (2006). *Economic Applications of Nonparametric Methods*. Ph.D. thesis, University of York. https://core.ac.uk/download/pdf/77022754.pdf

Beran, J., Feng, Y., Ghosh, S. & Kulik, R. (2013). *Long memory processes - probabilistic properties and statistical methods*, Berlin: Springer.

Birnbaum, Z. W. and Saunders, S. C. (1969). A new family of life distributions. Journal of Applied Probability 6, 319-327.

Bollino, C. A., Perali, F. and Rossi, N. (2000). Linear household technologies. *J. Appl. Econom.*, 15, 253-274.

Box, G. and Cox, D. (1964). An analysis of transformations. Journal of the Royal Statistical Society, Series B, 26, 211-252.

Cheng, S., Wei, L., Ying, Z. (1997). Prediction of survival probabilities with semiparametric transformationmodels. J. Am. Statist. Assoc., 92, 227-235.

Chernozhukov, V., Fernndez-Val, I. and Melly, B. (2013). Inference on counterfactual distributions. Econometrica 81, 2205-2268.

Collins, S. (1991). Prediction Techniques for BoxCox Regression Models. *J. Busin. & Econ. Statist.*, 9, 267-277.

Duerinckx, M., Ley, C. and Swan, Y. (2014). Maximum likelihood characterization of distributions. Bernoulli, Bernoulli 20, 775-802.

Foster, A., Tian, L. and Wei, L. (2001). Estimation for the Box-Cox transformationmodel without assuming parametric error distribution. *J. Amer. Statist. Assoc.*, 96, 1097-1101.

Fasiolo, M, Wood, S. N., Zaffran, M. Nedellec, R. and Goude, Y. (2020). Fast Calibrated Additive Quantile Regression. *J. Amer. Statist. Assoc. - Theory and Methods*, forthcoming, https://doi.org/10.1080/01621459.2020.1725521

Härdle, W. (1991). S̑moothing Techniques with Implementation in S. New York: Springer.

Hafield, T. and Racine, J.S. (2008). Nonparametric Econometrics: The np Package. *J. Statist. Software*, 27, Issue 5. http://www.jstatsoft.org/

Hall, P. and Müller, H. G. (2003) Order-preserving nonparametric regression, with applications to conditional distribution and quantile function estimation. J. Am. Statist. Ass., 98, 598-608.

Hall, P., Racine, J. S., Li, Q. (2004). Cross-Validation and the Estimation of Conditional Probability Densities. *J. Amer. Statist. Assoc.*, 99, 1015-1026.

Hothorn, T., Möst, L. and Bühlmann, P. (2018). Most Likely Transformations. *Scand. J. Statist.*, 45, 110-134.

Hothorn, T., Kneib, T. and Bühlmann, P. (2014). Conditional transformation models. *J. R. Stat. Soc. Ser. B.* (Stat. Methodol.), 76, 3-27.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2021). *An Introduction to Statistical Learning with Applications in R* (2nd ed.). Springer, New York.

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation". *Biometrika*, 36, 149-176.

Jones, M. C and Pewsey, A. (2009). Sinh-arcsinh distributions. Biometrika, 96, 4, 761-780.

Jones, M. C and Pewsey, A. (2019). The sinharcsinh normal distribution. *Significance*, 16, 6-7.

Jones, M. C., Marchand, E. and Strawderman, W. E. (2019). On An Intriguing Distributional Identity. *Amer. Statist.*, 73, 16-21.

Koenker, R. (2005). *Quantile Regression.* Cambridge University Press, Cambridge.

Koenker, R. and Koenker; Bassett, G., Jr. (1978). Regression QuantilesRoger. *Econometrica*, 46, 33-50.

Koenker, R., Leorato, S. and Peracchi, F. (2013). Distributional vs. Quantile Regression. No. 1329, EIEF Working Papers Series. http://www.eief.it/files/2013/12/wp-29-distributional-vs-quantile-regression.pdf

Nelder, J.A. and Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal*, 7, 308-313

Pewsey, A. and Abe, T. (2015). The sinh-arcsinhed logistic family of distributions: properties and inference. *Ann Inst Stat Math*, 67, 573-594.

Pipiras, V., and Taqqu, M. S. (2017). *Long-range dependence and self-similarity*. Cambridge university press.

Rieck, J. R. and Nedelman, J. R. (1991). A log-linear model for the Birnbaum-Saunders distribution. *Technometrics*, 33, 51-60.

Rothe, C. and Wied, D. (2013). Misspecification Testing in a Class of Conditional Distributional Models. *J. Amer. Statist. Assoc.*, 108, 314-324.

Rosco, J. F., Jones, M. C., Pewsey, A. (2011). Skew t distributions via the sinh-arcsinh transformation. *TEST*, 20, 630-652.

Shapiro, S. S. and Wilk, M.B (1965). An analysis of variance Test for Normality (complete samples). Biometrika, 52, 591-611.

Smith, R. L. and Naylor, J. C. (1987). A comparison of maximum likelihood and Bayesian estimators for the threeparameter Weibull distribution. *Appl. Statist.* 36, 358-369.

Stasinopoulos, M. D., Rigby, R. A., Heller, G. Z. and Voudouris, V. (2017). Flexible Regression and Smoothing: Using GAMLSS in R. CRC Press.

Wei, Y. and He, X. (2006). Conditional Growth Charts (with discussions). *Ann. Statist.*. 34, 20692097 and 2126-2131.

Van Zwet, W. R. (1964). Convex transformations of random variables. Amsterdam: Mathematisch Centrum.

Vilca, F., Santana, L. R., Leiva, V. and Balakrishnan, N. (2011). Estimation of extreme percentiles in Birnbaum Saunders distributions. *Comput. Statist. & Data Anal.*, 55, 1665-1678.

Appendix. Proofs of the main results

We begin by stating some basic facts of the SAN transformation. Several equivalent formulas of $S(y)$ are given in Equations (3) and (4) in JP09. Similar equivalent formulas also hold for $C(y)$:

$$\begin{aligned} C(y) &= \frac{1}{2}\left[e^{-\beta}\exp\{\alpha\sinh^{-1}(y)\} + e^{\beta}\exp\{-\alpha\sinh^{-1}(y)\}\right] \\ &= \frac{1}{2}\left[e^{-\beta}\{(1+y^2)^{1/2}+y\}^{\alpha} + e^{\beta}\{(1+y^2)^{1/2}+y\}^{-\alpha}\right] \\ &= \frac{1}{2}\left[e^{-\beta}\{(1+y^2)^{1/2}+y\}^{\alpha} + e^{\beta}\{(1+y^2)^{1/2}-y\}^{\alpha}\right]. \end{aligned} \quad (A.1)$$

A few useful facts about $S(y)$ and $C(y)$ are e.g. 1) $C(y) = \sqrt{1+S^2(y)}$; 2) $S(y) = y$ and $C(y) = \sqrt{1+y^2}$, if $\beta = 0$ and $\alpha = 1$; 3) For given y, $C(y)$ increases monotonically over α, $S(y)$ decreases over α, if $y < 0$, and increases over α, if $y > 0$; 4) $|S(y)| > |y|$ and $C(y) > \sqrt{1+y^2}$, if $\alpha > 1$, and $|S(y)| < |y|$ and $C(y) < \sqrt{1+y^2}$, if $\alpha < 1$; 5) $S'(y) = \alpha C(y)/\sqrt{1+y^2}$ and $C'(y) = \alpha S(y)/\sqrt{1+y^2}$.

A sketched proof of Theorem 1. Following Equation (4) in JP09 and (A.1), it is easy to show that, for large $|y|$, $S(y) \approx 2^{\alpha-1}\text{sgn}(y)\exp[-\text{sgn}(y)\beta]|y|^{\alpha}$ and $C(y) \approx 2^{\alpha-1}\exp[-\text{sgn}(y)\beta]|y|^{\alpha}$. This results in

$$f_{CS}(y) = O\left(|y|^{\alpha-1}\exp\left\{-2^{2\alpha-3}\sigma^{-2}\exp[-2\text{sgn}(y)\beta]|y|^{2\alpha}\right\}\right). \quad (A.2)$$

The tails of f_{CS} are of the Weibull type and decay in an exponential-power rate. The proof of further results in i) to ii) based on (A.2) is straightforward. A detailed proof is omitted. We will only check the results on the moments and mgf X briefly. Note that, for any $k > 0$, if $E(X^k) < \infty$ or not is only determined by the right tail of $f_{CS}(y)$. It is easy to show that $E(X^k) = E(e^{ky}) = \infty$, if $\alpha < 0.5$ and $E(X^k) = E(e^{ky}) < \infty$, if $\alpha > 0.5$, because $ye^{ky}f_{CS}(y) \to \infty$, as $y \to \infty$, if $\alpha < 0.5$, and $ye^{ky}f_{CS}(y) \to 0$, as $y \to \infty$, if $\alpha > 0.5$. That is, $\alpha > 0.5$ is necessary and sufficient so that $E(X^k) < \infty$ for any $k > 0$. The fact that the mgf of X for any $u > 0$ does never exist is also easy to check, because for $\alpha > 0$, $E(e^{uX}) = E(e^{ue^y}) = \infty$, if $u > 0$. Theorem 1 is proved. $\diamond$

Proof of Proposition 1. Note that the re-transformation of $S(y)$ is another sinh-arcsinh transformation $K(Z)$ with parameters $1/\alpha$ and $-\beta/\alpha$. By means of the second equivalent form of $K(Z)$ given in JP09 we obtain the expansion of $E(Y^k_{\alpha,\beta,\sigma^2})$ in (9), which is the same as that shown on Page 764 in JP09. The only difference is that now $Z \sim (0, \sigma^2)$

with a general variance $\sigma^2 > 0$. This leads to a generalization of the components Q_δ in $E(Y^k_{\alpha,\beta,\sigma^2})$. Set $\zeta = z + (1+z^2)^{1/2}$ and $x = \zeta^2/(8\sigma^2)$, we have

$$\begin{aligned}
Q_\delta &= E\left[\{(1+z^2)^{1/2}+z\}^\delta\right] \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\left[\{(1+z^2)^{1/2}+z\}^\delta\right] e^{-\frac{1}{2\sigma^2}z^2}dz \\
&= \frac{1}{\sqrt{8\pi\sigma^2}}\int_0^\infty \zeta^\delta\left(1+\frac{1}{\zeta^2}\right)\exp\left\{-\frac{1}{8\sigma^2}\left(\zeta-\frac{1}{\zeta}\right)^2\right\}d\zeta \\
&= \frac{e^{1/(4\sigma^2)}(8\sigma^2)^{(\delta+1)/2}}{\sqrt{32\pi\sigma^2}}\int_0^\infty x^{(\delta-1)/2}\left(1+\frac{1}{8\sigma^2 x}\right)\exp\left(-x-\frac{1}{64\sigma^2 x}\right)dx \\
&= \frac{e^{1/(4\sigma^2)}}{\sqrt{8\pi\sigma^2}}\{K_{(\delta+1)/2}[1/(4\sigma^2)]+K_{(\delta-1)/2}[1/(4\sigma^2)]\} \qquad \text{(A.3)}
\end{aligned}$$

by (3.471.12) of Gradshteyn and Ryzhik (2007), where K is the modified Bessel function of the second kind with $K_{-\nu}(x) = K_\nu(x)$. $\diamond$

Remark A.1. *It is easy to see that $Q_{-\delta} = Q_\delta$. Like in the results in JP09, the above extensions only involve the application of the well-studied modified Bessel function of the second order. However, the moments of a SAN distribution with $\mu \neq 0$ cannot be stated based on those special functions. Now, the resulting special integrals are of a form, which is not yet studied in the literature. The formulas of the moments of a Log-SAN distribution will involve other unknown special integrals. Detail discussion on those topics is omitted.*

Proof of Theorem 2.

For the SAN distribution, consider the case with fixed parameters except for β and let the two distribution functions with β_1 and β_2, where $\beta_2 > \beta_1$, be $F_{CS,1}$ and $F_{CS,2}$, respectively. According to van Zwet's (1964) skewness ordering, β is a skewness parameter such that $F_{CS,1}$ is less positively skewed than $F_{CS,2}$, denoted by $F_{CS,1} \leq_S F_{CS,2}$, if $F^{-1}_{CS,2}(F_{CS,1})$ is convex for all y. Using the formulas given in the context after (7) we have

$$\begin{aligned}
F^{-1}_{CS,2}[F_{CS,1}(y)] &= \sinh\{[\sinh^{-1}(S[y;\alpha,\beta_1])+\beta_2]/\alpha\} \\
&= \sinh\{[\alpha\sinh^{-1}(y)-\beta_1+\beta_2]/\alpha\} \qquad \text{(A.4)} \\
&= S(y;1,d_\beta)
\end{aligned}$$

with $d_\beta = (\beta_1-\beta_2)/\alpha < 0$. This fact is not affected by the additional parameters μ and σ^2 and is the same as that obtained in Section 2.2 in JP09. Further proof of the result in this case follows from their argument, who showed that because $[S(y;1,d_\beta)]'_y > 0$.

For the Log-SAN case let $F_{CL,1}(x)$ and $F_{CL,2}(x)$ denote the two distribution functions with β_1 and β_2. Based on the formulas given in the context after (8) and similar derivation as above we have

$$\begin{aligned} F^{-1}_{CL,2}[F_{CL,1}(x)] &= \exp(\sinh\{[\sinh^{-1}(S[\ln(x);\alpha,\beta_1]) + \beta_2]/\alpha\}) \\ &= \exp(\sinh\{[\alpha\sinh^{-1}(\ln(x)) - \beta_1 + \beta_2]/\alpha\}) \qquad \text{(A.5)} \\ &= \exp\{S[\ln(x);1,d_\beta]\}, \end{aligned}$$

where d_β is as defined before. We have

$$\{F^{-1}_{CL,2}[F_{CL,1}(x)]\}'_x = F^{-1}_{CL,2}[F_{CL,1}(x)][S(y;1,d_\beta)]'_y \frac{1}{x} > 0, \qquad \text{(A.6)}$$

where $y = \ln(x)$ and the three components on the right-hand side are all positive. ◇

Proof of Theorem 3.

i) For $\beta = \mu = 0$, denote the simplified form of $S(y)$ and $C(y)$ by $S_0(y) = \sinh[\alpha\sinh^{-1}(y)]$ and $C_0(y) = \cosh[\alpha\sinh^{-1}(y)]$, respectively. The density function $f_{CS0}(y)$ in this case can be deduced from $f_{CS}(y)$ with $S(y)$ and $C(y)$ being replaced by $S_0(y)$ and $C_0(y)$, and putting $\mu = 0$, which is given by

$$f_{CS0}(y) = \frac{\alpha}{\sigma\sqrt{2\pi(1+y^2)}} C_0(y) \exp\left\{-\frac{[S_0(y)]^2}{2\sigma^2}\right\}. \qquad \text{(A.7)}$$

The distribution and quantile functions of this subfamily are $F_{CS0}(y) = \Phi[S_0(y)/\sigma]$ and $F^{-1}_{CS0}(u) = \sinh\{\sinh^{-1}[\sigma\Phi^{-1}(u)]/\alpha\}$, respectively. Let $\alpha_1 > \alpha_2$. Denote the two distribution functions with α_1 and α_2 by $F_{CS0,1}$ and $F_{CS0,2}$. According to Zwet's ordering for kurtosis of symmetric distributions around zero we need to show that $F^{-1}_{CS0,2}[F_{CS0,1}(y)]$ is convex for $y > 0$. Insert the formulas of $F_{CS0,1}(y)$ and $F^{-1}_{CS0,2}(u)$ into $F^{-1}_{CS0,2}[F_{CS0,1}(y)]$, we can obtain

$$\begin{aligned} F^{-1}_{CS0,2}[F_{CS0,1}(y)] &= \sinh\{\sinh^{-1}[\sigma\Phi^{-1}(\Phi[S_{0,1}(y)/\sigma])]/\alpha_2\} \\ &= \sinh\{\sinh^{-1}[S_{0,1}(y)]/\alpha_2\} \qquad \text{(A.8)} \\ &= \sinh[\alpha_1\sinh^{-1}(y)/\alpha_2]. \end{aligned}$$

That is $F^{-1}_{CS0,2}[F_{CS0,1}(y)] = S_0(y;c_\alpha)$ with $c_\alpha = \alpha_1/\alpha_2 > 1$. This is again the same as that obtained in Section 3.1 of JP09. The authors showed that $F^{-1}_{CS0,2}[f_{CS0,1}(y)]$ is convex for $y > 0$. That is α is a kurtosis parameter according to Zwet's ordering for kurtosis of symmetric distributions around zero and this property is not affected by the additional parameter σ^2.

ii) The key point to find the modes of the density $f(y)$ is to discuss the properties of $f'(y)$ or $(\ln f)'(y)$, and the sign of the second derivative at their zeros. This analysis can also be equivalently carried out by analyzing those properties of $f'(y)h(y)$ or $(\ln f)'(y)h(y)$, where $h(y)$ is some positive function. This idea will be employed in the following to obtain a formula, which is easy to use in the case under consideration.

Note that $S_0'(y) = \alpha C_0(y)/\sqrt{1+y^2}$ and $C_0'(y) = \alpha S_0(y)/\sqrt{1+y^2}$. Now, the discussion on the zeros of $f_{CS0}'(y)$ or $[\ln(f_{CS0})(y)]'$ is equivalent to analyze the zeros of

$$d_{CS0}(y) \;=\; \frac{\alpha S_0(y)}{C_0(y)}\left[1-\frac{C_0^2(y)}{\sigma^2}\right]-\frac{y}{\sqrt{1+y^2}}, \tag{A.9}$$

which is anti-symmetric with $d_{CS0}(0)\equiv 0$, $\lim_{y\to-\infty} d_{CS0}(y) = \infty$ and $\lim_{y\to\infty} d_{CS0}(y) = -\infty$. The two terms $y/\sqrt{1+y^2}$ and $S_0(y)/C_0(y) = S_0(y)/\sqrt{1+S_0^2(y)}$ are monotonically increasing. Consider first the special case with $\sigma^2 \le 1$. This means that $[1-C_0^2(y)/\sigma^2] \le 0$, which increases monotonically for $y<0$, and decreases monotonically for $y>0$. Hence, the product $\alpha S_0(y)/C_0(y)[1-C_0^2(y)/\sigma^2]$ is positive for $y<0$, negative for $y>0$ and decreases monotonically from ∞ to $-\infty$. We see d_{CS0} decreases monotonicaly from ∞ to $-\infty$, provided that $\sigma^2\le 1$. Thus it has only one zero at $y=0$ with $d_{CS0}'(0)<0$. This leads to $f_{CS}'(0)=0$ and $f_{CS}''(0)<0$. That is $f_{CS0}(y)$ has now a unique mode at $y=0$.

In the special case with $\alpha=1$, we have $S_0(y)=y$ and $C_0(y)=\sqrt{1+y^2}$. This results in $d_{CS0}(y) = -\sigma^{-2}S_0(y)C_0(y)$. It is easy to show that now $d_{CS0}(y)$ decreases monotonically from ∞ to $-\infty$ with a single zero at $y=0$, which corresponds again to the unique mode of $f_{CS0}(y)$.

Now, consider the case with $\alpha<1$. We first state some results on $S_0(y)$ and $C_0(y)$ without proof: $S_0(y;\alpha)$ is an increasing function of α for $y<0$ and a decreasing function of α for $y>0$. For any $y\neq 0$, $C_0(y;\alpha)$ is an increasing function of α. In particular we have $|S_0(y)| < |y| = |S_0(y,1)|$, if $\alpha<1$. This results in $S_0(y)/C_0(y) < y/\sqrt{1+y^2}$, for $y>0$, and $S_0(y)/C_0(y) > y/\sqrt{1+y^2}$, for $y<0$. Moreover, for $y>0$ we have

$$d_{CS0}(y) \;<\; \frac{\alpha S_0(y)}{C_0(y)}-\frac{y}{\sqrt{1+y^2}}<0.$$

Analogously, it can be shown that $d_{CS0}(y) > 0$, for $y < 0$. That is d_{CS0} has only one zero at $y = 0$ with $d'_{CS0}(0) < 0$. Note that we have not shown that $d_{CS0}(y)$ is monotonically decreasing in this case. In summary, $f_{CS0}(y)$ is unimodal, if $\alpha \leq 1$ or $\sigma^2 \leq 1$.

Now, consider $\sigma^2 > 1$. Let $y = -\sinh[\cosh^{-1}(\sigma)/\alpha] < 0$ and $y_U = \sinh[\cosh^{-1}(\sigma)/\alpha] = |y| > 0$. Note that y and y_U are not well defined for $\sigma^2 < 1$ and they are $y = y_U = 0$ in the trivial case with $\sigma^2 = 1$. It is easy to see $d_{CS0}(y)$ decreases monotonically in the interval $(-\infty, y)$ from ∞ to $y_U/\sqrt{1+y_U^2}$ and it decreases monotonically again in the interval $[y_U, \infty)$ from $-y_U/\sqrt{1+y_U^2}$ to $-\infty$. That is all possible zeros of d_{CS0} lie within the symmetric interval (y, y_U) around zero. In the sequel we will only discuss the behavior of d_{CS0} within $(0, y_U)$ by means of the follwing ratio between the two part on the right-hand side of (A.9)

$$\begin{aligned} R_0(y) &= \frac{\alpha S_0(y)\sqrt{1+y^2}}{yC_0(y)} - \frac{\alpha S_0(y)C_0(y)\sqrt{1+y^2}}{\sigma^2 y} \\ &= R_{01}(y) - R_{02}(y), \end{aligned} \tag{A.10}$$

where $R_{01}(y)$ and $R_{02}(y)$ stand for the first and the second parts on the right-hand side of (A.10), respectively. A zero of d_{CS0} at some $y > 0$ should satisfy $R_0(y) = 1$. Furthermore, we have $d_{CS0}(y) < 0$, if $R_0(y) < 1$ and $d_{CS0}(y) > 0$, if $R_0(y) > 1$. The dominating term of the Taylor expansion of $S_0(y)$ is αy. This leads to $\lim_{y \to 0^+} R_{01}(y) = \alpha^2$, $\lim_{y \to 0^+} R_{02}(y) = \alpha^2\sigma^{-2}$ and $\lim_{y \to 0^+} R_0(y) = \alpha^2(1-\sigma^{-2})$. And we also have $R_0(y_U) \equiv 0$. Straightforward calculation results in

$$R'_{01}(y) = \frac{\alpha[\alpha y\sqrt{1+y^2} - S_{A0}(y)C_{A0}(y)]}{C_{A0}^2(y)y^2\sqrt{1+y^2}}. \tag{A.11}$$

Based on (A.1) and the basic facts on S_A as given in JP09 it can be shown that, for $y \geq 0$, $C_{A0}(y) > \sqrt{1+y^2}$ and $S_{A0}(y) > \alpha y$. This shows that $R'_{01}(y) < 0$ and $R_{01}(y)$ decreases monotonically. Similarly, it can be shown that $S_{A0}(y)/[\alpha y]$ increases monotonically. This implies that $R_{02}(y)$ is monotonically increasing and $R_0(y)$ decreases monotonically from $\alpha^2(1-\sigma^{-2})$ to zero. This results in $d_{CS0}(y) < 0$ for $y \in (0, y_U)$, if $\alpha^2(1-\sigma^{-2}) \leq 1$ and hence it does not have any zero within this interval.

For $\alpha^2(1-\sigma^{-2}) > 1$ however, there is one (and only one) solution of $R_0(y) = 1$, denoted by y_S. We have $0 < y_S < y_U$ and it can be further shown that y_S corresponds to a mode of $f_{CS0}(y)$. Due to the symmetry of $f_{CS0}(y)$, it has another mode at $y_P = -y_S < 0$. And now $y = 0$ becomes a local minimal point of $f_{CS0}(y)$. Theorem 3 is proved. $\diamond$

Proof of Proposition 2.

Now, we choose to use the following determinant term of f'_{CSb}:

$$d_{CSb}(y) = \alpha\frac{S_A(y)}{C_A(y)}\left[1-\frac{C_A^2(y)}{\sigma^2}\right]-\frac{y}{\sqrt{1+y^2}}, \tag{A.12}$$

which is of the same form of d_{CS0}, but with S_{A0} and C_{A0} being replaced by S_A and C_A. Let $S_\beta = S_A(0) = \sinh(-\beta) < 0$ and $C_\beta = C_A(0) = \cosh(-\beta) > 1$. We have $S_A(y_\beta) = 0$ and $C_A(y_\beta) = 1$. Now, $S_A(y)$ is an increasing function but not anti-symmetric about y_β and $C_A(y)$ is decreasing for $y < y_\beta$ and increasing for $y > y_\beta$, but not symmetric about y_β. The first derivative of d_{CSb} is given by

$$\begin{aligned} d'_{CSb}(y) &= \frac{\alpha^2 C_A(y)}{\sqrt{1+y^2}C_A(y)}-\frac{\alpha^2 S_A^2(y)}{\sqrt{1+y^2}C_A^2(y)}-\frac{\alpha^2 C_A^2(y)}{\sqrt{1+y^2}\sigma^2}-\frac{\alpha^2 S_A^2(y)}{\sqrt{1+y^2}\sigma^2} \\ &\quad -\frac{1}{\sqrt{1+y^2}}+\frac{y^2}{(1+y^2)^{3/2}} \\ &= \frac{1}{\sqrt{1+y^2}}\left\{\alpha^2[C_A^{-2}(y)-1/\sigma^2-2S_A^2(y)/\sigma^2]-\frac{1}{(1+y^2)}\right\}. \end{aligned} \tag{A.13}$$

From (A.13) it is easy to see that $d'_{CSb} < 0$, if $\sigma^2 \leq 1$. And, d_{CSb} takes values from $-\infty$ to ∞, which hence only has one zero, denoted by y_P Moreover, we have $0 < y_P < y_\beta$, because $d_{CSb}(0) = \alpha S_\beta(1-C_\beta^2/\sigma^2)C_\beta^{-1} > 0$ and $d_{CSb}(y_\beta) = -y_\beta/\sqrt{1+y_\beta^2} < 0$. Hence, y_P corresponds to a unimode of f_{CSb} in this subfamily. Proposition 2 is proved. ◇

Elements of the observed information matrix:

Those details can be put into an online supplement

Elements of the observed information matrix for the SAN distribution will be given in Equations (S.1) to (S.21), respectively. The related results of JP09 can be obtained from the formulas for the first four parameters by setting $\mu = 0$ and $\sigma^2 = 1$. This fact will be shown using the first element. Formulas for the elements of the observed information matrix for the Log-SAN distribution are the same, but calculated with the log-data.

Let y_i, $i = 1, ..., n$, be the observations and $\tilde{y}_i = (y_i - \xi)/\eta$. Recall that

$$L(\boldsymbol{\theta}) = n \ln(\alpha/\eta/\sqrt{2\pi}/\sigma) + \sum_{i=1}^{n} \{\ln[C(\tilde{y}_i)/\sqrt{1+\tilde{y}_i^2}] - [S(\tilde{y}_i) - \mu]^2/2/\sigma^2\}.$$

Some basic formulas we need for the derivation are

$S'_\xi(\tilde{y}_i) = -\alpha\eta^{-1}C(\tilde{y}_i)/\sqrt{1+\tilde{y}_i^2}$, $C'_\xi(\tilde{y}_i) = -\alpha\eta^{-1}S(\tilde{y}_i)/\sqrt{1+\tilde{y}_i^2}$,
$[\sqrt{1+\tilde{y}_i^2}]'_\xi = -\eta^{-1}\tilde{y}_i/\sqrt{1+\tilde{y}_i^2}$, $(\tilde{y}_i)'_\xi = -\eta^{-1}$,

$S'_\eta(\tilde{y}_i) = -\alpha\eta^{-1}\tilde{y}_i C(\tilde{y}_i)/\sqrt{1+\tilde{y}_i^2}$, $C'_\eta(\tilde{y}_i) = -\alpha\eta^{-1}\tilde{y}_i S(\tilde{y}_i)/\sqrt{1+\tilde{y}_i^2}$,
$[\sqrt{1+\tilde{y}_i^2}]'_\eta = -\eta^{-1}\tilde{y}_i^2/\sqrt{1+\tilde{y}_i^2}$, $(\tilde{y}_i)'_\eta = -\eta^{-1}\tilde{y}_i$,

$S'_\alpha(\tilde{y}_i) = \sinh^{-1}(\tilde{y}_i)C(\tilde{y}_i)$ and $C'_\alpha(\tilde{y}_i) = \sinh^{-1}(\tilde{y}_i)S(\tilde{y}_i)$,

$S'_\beta(\tilde{y}_i) = -C(\tilde{y}_i)$ and $C'_\beta(\tilde{y}_i) = -S(\tilde{y}_i)$ and $C^2(\tilde{y}_i) = 1 + S^2(\tilde{y}_i)$.

The score functions are

$$\begin{aligned}
\frac{\partial L(\tilde{y}_i|\boldsymbol{\theta})}{\partial \xi} &= -\frac{\alpha S(\tilde{y}_i)}{\eta C(\tilde{y}_i)\sqrt{1+\tilde{y}_i^2}} + \frac{\tilde{y}_i}{\eta(1+\tilde{y}_i^2)} + \frac{\alpha C(\tilde{y}_i)[S(\tilde{y}_i) - \mu]}{\eta\sigma^2\sqrt{1+\tilde{y}_i^2}}, \\
\frac{\partial L(\tilde{y}_i|\boldsymbol{\theta})}{\partial \eta} &= -\frac{1}{\eta} - \frac{\alpha\tilde{y}_i S(\tilde{y}_i)}{\eta C(\tilde{y}_i)\sqrt{1+\tilde{y}_i^2}} + \frac{\tilde{y}_i^2}{\eta(1+\tilde{y}_i^2)} + \frac{\alpha\tilde{y}_i C(\tilde{y}_i)[S(\tilde{y}_i) - \mu]}{\eta\sigma^2\sqrt{1+\tilde{y}_i^2}}, \\
\frac{\partial L(\tilde{y}_i|\boldsymbol{\theta})}{\partial \alpha} &= \frac{1}{\alpha} + \frac{\sinh^{-1}(\tilde{y}_i)S(\tilde{y}_i)}{C(\tilde{y}_i)} - \sinh^{-1}(\tilde{y}_i)C(\tilde{y}_i)[S(\tilde{y}_i) - \mu]/\sigma^2, \\
\frac{\partial L(\tilde{y}_i|\boldsymbol{\theta})}{\partial \beta} &= -\frac{S(\tilde{y}_i)}{C(\tilde{y}_i)} + C(\tilde{y}_i)[S(\tilde{y}_i) - \mu]/\sigma^2, \\
\frac{\partial L(\tilde{y}_i|\boldsymbol{\theta})}{\partial \mu} &= [S(\tilde{y}_i) - \mu]/\sigma^2, \\
\frac{\partial L(\tilde{y}_i|\boldsymbol{\theta})}{\partial \sigma} &= -\frac{1}{\sigma} + [S(\tilde{y}_i) - \mu]^2/\sigma^3.
\end{aligned}$$

The term involving the second derivative with respect to ξ is

$$\begin{aligned}
-\frac{\partial L^2}{\partial \xi^2} &= \frac{1}{\eta^2}\sum_{i=1}^{n}\left\{\left[\frac{-\alpha^2}{1+\tilde{y}_i^2}+\frac{\alpha^2 S^2(\tilde{y}_i)}{C^2(\tilde{y}_i)(1+\tilde{y}_i^2)}+\frac{\alpha\tilde{y}_i S(\tilde{y}_i)}{C(\tilde{y}_i)(1+\tilde{y}_i^2)^{3/2}}\right]\right.\\
&\quad+\left[\frac{1}{1+\tilde{y}_i^2}-\frac{2\tilde{y}_i^2}{(1+\tilde{y}_i^2)^2}\right]\\
&\quad\left.+\frac{1}{\sigma^2}\left[\frac{\alpha^2 S(\tilde{y}_i)[S(\tilde{y}_i)-\mu]}{1+\tilde{y}_i^2}+\frac{\alpha^2 C^2(\tilde{y}_i)}{1+\tilde{y}_i^2}-\frac{\alpha\tilde{y}_i C(\tilde{y}_i)[S(\tilde{y}_i)-\mu]}{(1+\tilde{y}_i^2)^{3/2}}\right]\right\}\\
&= \frac{1}{\eta^2}\sum_{i=1}^{n}\left\{\frac{\alpha[\tilde{y}_i S(\tilde{y}_i)C(\tilde{y}_i)-\alpha\sqrt{1+\tilde{y}_i^2}]}{C^2(\tilde{y}_i)(1+\tilde{y}_i^2)^{3/2}}+\frac{1-\tilde{y}_i^2}{(1+\tilde{y}_i^2)^2}\right.\\
&\quad\left.+\frac{1}{\sigma^2}\left[\frac{\alpha^2[1+2S^2(\tilde{y}_i)-\mu S(\tilde{y}_i)]}{1+\tilde{y}_i^2}-\frac{\alpha\tilde{y}_i C(\tilde{y}_i)[S(\tilde{y}_i)-\mu]}{(1+\tilde{y}_i^2)^{3/2}}\right]\right\}. \qquad \text{(S.1)}
\end{aligned}$$

Put $\mu = 0$ and $\sigma^2 = 1$, we obtain the corresponding formula as given in JP09

$$-\frac{\partial L^2}{\partial \xi^2} = \frac{1}{\eta^2}\sum_{i=1}^{n}\left\{\frac{1-\tilde{y}_i^2}{(1+\tilde{y}_i^2)^2}+\alpha S^2(\tilde{y}_i)\frac{\alpha\sqrt{1+\tilde{y}_i^2}[3+2S^2(\tilde{y}_i)]-\tilde{y}_i S(\tilde{y}_i)C(\tilde{y}_i)}{C^2(\tilde{y}_i)(1+\tilde{y}_i^2)^{3/2}}\right\}.$$

All other elements will be given directly with out additional description.

$$\begin{aligned}
-\frac{\partial L^2}{\partial \xi\partial\eta} &= \frac{1}{\eta^2}\sum_{i=1}^{n}\left\{\left[\frac{-\alpha^2\tilde{y}_i}{1+\tilde{y}_i^2}+\frac{\alpha^2\tilde{y}_i S^2(\tilde{y}_i)}{C^2(\tilde{y}_i)(1+\tilde{y}_i^2)}+\frac{\alpha\tilde{y}_i^2 S(\tilde{y}_i)}{C(\tilde{y}_i)(1+\tilde{y}_i^2)^{3/2}}\right]\right.\\
&\quad+\left[\frac{\tilde{y}_i}{1+\tilde{y}_i^2}-\frac{2\tilde{y}_i^3}{(1+\tilde{y}_i^2)^2}\right]\\
&\quad\left.+\frac{1}{\sigma^2}\left[\frac{\alpha^2\tilde{y}_i S(\tilde{y}_i)[S(\tilde{y}_i)-\mu]}{1+\tilde{y}_i^2}+\frac{\alpha^2\tilde{y}_i C^2(\tilde{y}_i)}{1+\tilde{y}_i^2}-\frac{\alpha\tilde{y}_i^2 C(\tilde{y}_i)[S(\tilde{y}_i)-\mu]}{(1+\tilde{y}_i^2)^{3/2}}\right]\right\}\\
&= \frac{1}{\eta^2}\sum_{i=1}^{n}\tilde{y}_i\left\{\frac{\alpha[\tilde{y}_i S(\tilde{y}_i)C(\tilde{y}_i)-\alpha\sqrt{1+\tilde{y}_i^2}]}{C^2(\tilde{y}_i)(1+\tilde{y}_i^2)^{3/2}}+\frac{1-\tilde{y}_i^2}{(1+\tilde{y}_i^2)^2}\right.\\
&\quad\left.+\frac{1}{\sigma^2}\left[\frac{\alpha^2[1+2S^2(\tilde{y}_i)-\mu S(\tilde{y}_i)]}{1+\tilde{y}_i^2}-\frac{\alpha\tilde{y}_i C(\tilde{y}_i)[S(\tilde{y}_i)-\mu]}{(1+\tilde{y}_i^2)^{3/2}}\right]\right\}. \qquad \text{(S.2)}
\end{aligned}$$

$$\begin{aligned}
-\frac{\partial L^2}{\partial \xi\partial\alpha} &= \frac{1}{\eta}\sum_{i=1}^{n}\left\{\left[\frac{S(\tilde{y}_i)}{C(\tilde{y}_i)\sqrt{1+\tilde{y}_i^2}}+\frac{\alpha\sinh^{-1}(\tilde{y}_i)}{\sqrt{1+\tilde{y}_i^2}}-\frac{\alpha\sinh^{-1}(\tilde{y}_i)S^2(\tilde{y}_i)}{C^2(\tilde{y}_i)\sqrt{1+\tilde{y}_i^2}}\right]\right.\\
&\quad\left.-\frac{1}{\sigma^2}\left[\frac{C(\tilde{y}_i)[S(\tilde{y}_i)-\mu]}{\sqrt{1+\tilde{y}_i^2}}+\frac{\alpha\sinh^{-1}(\tilde{y}_i)S(\tilde{y}_i)[S(\tilde{y}_i)-\mu]}{\sqrt{1+\tilde{y}_i^2}}+\frac{\alpha\sinh^{-1}(\tilde{y}_i)C^2(\tilde{y}_i)}{\sqrt{1+\tilde{y}_i^2}}\right]\right\}\\
&= -\frac{1}{\eta\sigma^2}\sum_{i=1}^{n}\left\{\frac{C^3(\tilde{y}_i)[S(\tilde{y}_i)-\mu]-\sigma^2 S(\tilde{y}_i)C(\tilde{y}_i)}{C^2(\tilde{y}_i)\sqrt{1+\tilde{y}_i^2}}\right.\\
&\quad\left.+\frac{\alpha\{1+3S^2(\tilde{y}_i)+2S^4(\tilde{y}_i)-\mu S(\tilde{y}_i)[1+S^2(\tilde{y}_i)]-\sigma^2\}\sinh^{-1}(\tilde{y}_i)}{C^2(\tilde{y}_i)\sqrt{1+\tilde{y}_i^2}}\right\}. \qquad \text{(S.3)}
\end{aligned}$$

$$
\begin{aligned}
-\frac{\partial L^2}{\partial \xi \partial \beta} &= \frac{1}{\eta}\sum_{i=1}^{n}\left\{\left[\frac{-\alpha}{\sqrt{1+\tilde{y}_i^2}} + \frac{\alpha S^2(\tilde{y}_i)}{C^2(\tilde{y}_i)\sqrt{1+\tilde{y}_i^2}}\right]\right. \\
&\quad \left. + \frac{1}{\sigma^2}\left[\frac{\alpha S(\tilde{y}_i)[S(\tilde{y}_i)-\mu]}{\sqrt{1+\tilde{y}_i^2}} + \frac{\alpha C^2(\tilde{y}_i)}{\sqrt{1+\tilde{y}_i^2}}\right]\right\} \\
&= \frac{\alpha}{\eta\sigma^2}\sum_{i=1}^{n}\left[\frac{C^2(\tilde{y}_i)S(\tilde{y}_i)[S(\tilde{y}_i)-\mu] + C^4(\tilde{y}_i) - \sigma^2}{C^2(\tilde{y}_i)\sqrt{1+\tilde{y}_i^2}}\right]. \qquad \text{(S.4)}
\end{aligned}
$$

$$
-\frac{\partial L^2}{\partial \xi \partial \mu} = \frac{\alpha}{\eta\sigma^2}\sum_{i=1}^{n}\frac{C(\tilde{y}_i)}{\sqrt{1+\tilde{y}_i^2}}. \qquad \text{(S.5)}
$$

$$
-\frac{\partial L^2}{\partial \xi \partial \sigma} = \frac{2\alpha}{\eta\sigma^3}\sum_{i=1}^{n}\frac{C(\tilde{y}_i)[S(\tilde{y}_i)-\mu]}{\sqrt{1+\tilde{y}_i^2}}. \qquad \text{(S.6)}
$$

$$
\begin{aligned}
-\frac{\partial L^2}{\partial \eta^2} &= \frac{1}{\eta^2}\sum_{i=1}^{n}\left\{\left[\frac{-\alpha^2\tilde{y}_i^2}{1+\tilde{y}_i^2} + \frac{\alpha^2\tilde{y}_i^2 S^2(\tilde{y}_i)}{C^2(\tilde{y}_i)(1+\tilde{y}_i^2)} + \frac{\alpha\tilde{y}_i^3 S(\tilde{y}_i)}{C(\tilde{y}_i)(1+\tilde{y}_i^2)^{3/2}}\right.\right. \\
&\quad \left. - \frac{\alpha\tilde{y}_i S(\tilde{y}_i)}{C(\tilde{y}_i)\sqrt{1+\tilde{y}_i^2}}\right] + \left[\frac{2\tilde{y}_i^2}{1+\tilde{y}_i^2} - \frac{2\tilde{y}_i^4}{(1+\tilde{y}_i^2)^2}\right] \\
&\quad + \frac{1}{\sigma^2}\left[\frac{\alpha^2\tilde{y}_i^2 S(\tilde{y}_i)[S(\tilde{y}_i)-\mu]}{1+\tilde{y}_i^2} + \frac{\alpha^2\tilde{y}_i^2 C^2(\tilde{y}_i)}{1+\tilde{y}_i^2}\right. \\
&\quad \left.\left. - \frac{\alpha\tilde{y}_i^3 C(\tilde{y}_i)[S(\tilde{y}_i)-\mu]}{(1+\tilde{y}_i^2)^{3/2}} + \frac{\alpha\tilde{y}_i C(\tilde{y}_i)[S(\tilde{y}_i)-\mu]}{\sqrt{1+\tilde{y}_i^2}}\right]\right\} \\
&= \frac{1}{\eta^2}\sum_{i=1}^{n}\left\{\frac{-\alpha^2\tilde{y}_i^2}{C^2(\tilde{y}_i)(1+\tilde{y}_i^2)} - \frac{\alpha\tilde{y}_i S(\tilde{y}_i)}{C(\tilde{y}_i)(1+\tilde{y}_i^2)^{3/2}} + \frac{2\tilde{y}_i^2}{(1+\tilde{y}_i^2)^2}\right. \\
&\quad \left. + \frac{\alpha}{\sigma^2}\left[\frac{\alpha[1+2S^2(\tilde{y}_i)-\mu S(\tilde{y}_i)]\tilde{y}_i^2}{1+\tilde{y}_i^2} + \frac{\tilde{y}_i C(\tilde{y}_i)[S(\tilde{y}_i)-\mu]}{(1+\tilde{y}_i^2)^{3/2}}\right]\right\}. \qquad \text{(S.7)}
\end{aligned}
$$

$$
\begin{aligned}
-\frac{\partial L^2}{\partial \eta \partial \alpha} &= \frac{1}{\eta}\sum_{i=1}^{n}\left\{\left[\frac{\tilde{y}_i S(\tilde{y}_i)}{C(\tilde{y}_i)\sqrt{1+\tilde{y}_i^2}} + \frac{\alpha\tilde{y}_i \sinh^{-1}(\tilde{y}_i)}{\sqrt{1+\tilde{y}_i^2}} - \frac{\alpha\tilde{y}_i \sinh^{-1}(\tilde{y}_i) S^2(\tilde{y}_i)}{C^2(\tilde{y}_i)\sqrt{1+\tilde{y}_i^2}}\right]\right. \\
&\quad \left. - \frac{1}{\sigma^2}\left[\frac{\tilde{y}_i C(\tilde{y}_i)[S(\tilde{y}_i)-\mu]}{\sqrt{1+\tilde{y}_i^2}} + \frac{\alpha\tilde{y}_i \sinh(\tilde{y}_i) S(\tilde{y}_i)[S(\tilde{y}_i)-\mu]}{\sqrt{1+\tilde{y}_i^2}} + \frac{\alpha\tilde{y}_i \sinh(\tilde{y}_i) C^2(\tilde{y}_i)}{\sqrt{1+\tilde{y}_i^2}}\right]\right\} \\
&= -\frac{1}{\eta\sigma^2}\sum_{i=1}^{n}\tilde{y}_i\left\{\frac{C^3(\tilde{y}_i)[S(\tilde{y}_i)-\mu] - \sigma^2 S(\tilde{y}_i)C(\tilde{y}_i)}{C^2(\tilde{y}_i)\sqrt{1+\tilde{y}_i^2}}\right. \\
&\quad \left. + \frac{\alpha\{C^2(\tilde{y}_i)S(\tilde{y}_i)[S(\tilde{y}_i)-\mu] + C^4(\tilde{y}_i) - \sigma^2\}\sinh^{-1}(\tilde{y}_i)}{C^2(\tilde{y}_i)\sqrt{1+\tilde{y}_i^2}}\right\}. \qquad \text{(S.8)}
\end{aligned}
$$

$$
\begin{aligned}
-\frac{\partial L^2}{\partial\eta\partial\beta} &= \frac{1}{\eta}\sum_{i=1}^{n}\left\{\left[\frac{-\alpha\tilde{y}_i}{\sqrt{1+\tilde{y}_i^2}}+\frac{\alpha\tilde{y}_i S^2(\tilde{y}_i)}{C^2(\tilde{y}_i)\sqrt{1+\tilde{y}_i^2}}\right]\right. \\
&\quad \left. +\frac{1}{\sigma^2}\left[\frac{\alpha\tilde{y}_i S(\tilde{y}_i)[S(\tilde{y}_i)-\mu]}{\sqrt{1+\tilde{y}_i^2}}+\frac{\alpha\tilde{y}_i C^2(\tilde{y}_i)}{\sqrt{1+\tilde{y}_i^2}}\right]\right\} \\
&= \frac{\alpha}{\eta\sigma^2}\sum_{i=1}^{n}\tilde{y}_i\left[\frac{C^2(\tilde{y}_i)S(\tilde{y}_i)[S(\tilde{y}_i)-\mu]+C^4(\tilde{y}_i)-\sigma^2}{C^2(\tilde{y}_i)\sqrt{1+\tilde{y}_i^2}}\right]. \qquad \text{(S.9)}
\end{aligned}
$$

$$
-\frac{\partial L^2}{\partial\eta\partial\mu} = \frac{\alpha}{\eta\sigma^2}\sum_{i=1}^{n}\frac{\tilde{y}_i C(\tilde{y}_i)}{\sqrt{1+\tilde{y}_i^2}}. \qquad \text{(S.10)}
$$

$$
-\frac{\partial L^2}{\partial\eta\partial\sigma} = \frac{2\alpha}{\eta\sigma^3}\sum_{i=1}^{n}\frac{\tilde{y}_i C(\tilde{y}_i)[S(\tilde{y}_i)-\mu]}{\sqrt{1+\tilde{y}_i^2}}. \qquad \text{(S.11)}
$$

$$
\begin{aligned}
-\frac{\partial L^2}{\partial\alpha^2} &= \frac{n}{\alpha^2}+\sum_{i=1}^{n}\left\{\left[\frac{[\sinh^{-1}(\tilde{y}_i)]^2 S^2(\tilde{y}_i)}{C^2(\tilde{y}_i)}-[\sinh^{-1}(\tilde{y}_i)]^2\right]\right. \\
&\quad \left. +\frac{1}{\sigma^2}\left[[\sinh^{-1}(\tilde{y}_i)]^2 S(\tilde{y}_i)[S(\tilde{y}_i)-\mu]+[\sinh^{-1}(\tilde{y}_i)]^2 C^2(\tilde{y}_i)\right]\right\} \\
&= \frac{n}{\alpha^2}+\sum_{i=1}^{n}[\sinh^{-1}(\tilde{y}_i)]^2\{[1+2S^2(\tilde{y}_i)-S(\tilde{y}_i)\mu]/\sigma^2-C^{-2}(\tilde{y}_i)\}. \qquad \text{(S.12)}
\end{aligned}
$$

$$
\begin{aligned}
-\frac{\partial L^2}{\partial\alpha\partial\beta} &= \sum_{i=1}^{n}\left\{\sinh^{-1}(\tilde{y}_i)-\sinh^{-1}(\tilde{y}_i)S^2(\tilde{y}_i)C^{-2}(\tilde{y}_i)\right. \\
&\quad \left. -[\sinh^{-1}(\tilde{y}_i)S(\tilde{y}_i)[S(\tilde{y}_i)-\mu]+\sinh^{-1}(\tilde{y}_i)C^2(\tilde{y}_i)]/\sigma^2\right\} \\
&= \sum_{i=1}^{n}\sinh^{-1}(\tilde{y}_i)\left\{C^{-2}(\tilde{y}_i)-\sigma^{-2}[1+2S^2(\tilde{y}_i)-\mu S(\tilde{y}_i)]\right\}. \qquad \text{(S.13)}
\end{aligned}
$$

$$
-\frac{\partial L^2}{\partial\alpha\partial\mu} = -\sum_{i=1}^{n}\sinh^{-1}(\tilde{y}_i)C(\tilde{y}_i)/\sigma^2. \qquad \text{(S.14)}
$$

$$
-\frac{\partial L^2}{\partial\alpha\partial\sigma} = -\sum_{i=1}^{n}2\sinh^{-1}(\tilde{y}_i)C(\tilde{y}_i)[S(\tilde{y}_i)-\mu]/\sigma^3. \qquad \text{(S.15)}
$$

$$-\frac{\partial L^2}{\partial \beta^2} \quad = \quad \sum_{i=1}^{n} \left\{ \frac{S^2(\tilde{y}_i)}{C^2(\tilde{y}_i)} - 1 + S(\tilde{y}_i)[S(\tilde{y}_i - \mu]/\sigma^2 + C^2(\tilde{y}_i)/\sigma^2 \right\}. \tag{S.16}$$

$$-\frac{\partial L^2}{\partial \beta \partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} C(\tilde{y}_i). \tag{S.17}$$

$$-\frac{\partial L^2}{\partial \beta \partial \sigma} = \sum_{i=1}^{n} 2C(\tilde{y}_i)[S(\tilde{y}_i) - \mu]/\sigma^3. \tag{S.18}$$

$$-\frac{\partial L^2}{\partial \mu^2} = n/\sigma^2. \tag{S.19}$$

$$-\frac{\partial L^2}{\partial \mu \partial \sigma} = \sum_{i=1}^{n} 2[S(\tilde{y}_i) - \mu]/\sigma^3. \tag{S.20}$$

$$-\frac{\partial L^2}{\partial \sigma^2} = \frac{1}{\sigma^2} \left\{ \sum_{i=1}^{n} \frac{3[S(\tilde{y}_i) - \mu]^2}{\sigma^2} - n \right\}. \tag{S.21}$$

Recent discussion papers

2021-04	Yuanhua Feng	Uni- and multivariate extensions of the sinh-arcsinh normal distribution applied to distributional regression
2021-03	Sebastian Letmathe Yuanhua Feng André Uhde	Semiparametric GARCH models with long memory applied to Value at Risk and Expected Shortfall
2021-02	Claus-Jochen Haake Walter Trockel	Socio-legal Systems and Implementation of the Nash Solution in Debreu-Hurwicz Equilibrium
2021-01	Thomas Gries Paul Welfens	Testing as an Approach to Control the Corona Epidemic Dynamics and Avoid Lockdowns
2020-10	Fatma Aslan Papatya Duman Walter Trockel	Non-cohesive TU-games: Efficiency and Duality
2020-09	Yuanhua Feng Jan Beran Sebastian Letmathe Sucharita Ghosh	Fractionally integrated Log-GARCH with application to value at risk and expected shortfall
2020-08	Fatma Aslan Papatya Duman Walter Trockel	Removed from the CIE series
2020-07	Thomas Gries Veronika Müller	Conflict Economics and Psychological Human Needs
2020-06	Thomas Mayrhofer Hendrik Schmitz	Prudence and prevention – Empirical evidence*
2020-05	Daniel A. Kamhöfer Hendrik Schmitz Matthias Westphal	Marginal College Wage Premiums under Selection into Employment*
2020-04	Thomas Gries	Income polarization and stagnation in a stochastic model of growth: When the demand side matters
2020-03	Nico Stoeckmann	A note on Bergstrand's 1985 gravity model
2020-02	Papatya Duman Walter Trockel	Nash Smoothing on the Test Bench: Hα -Essential Equilibria
2020-01	Joachim Heinzel Simon Hoof	Oligopolistic Upstream Competition with Differentiated Inputs
2019-11	Claus-Jochen Haake Thorsten Upmann Papatya Duman	The Decomposability of the Nash Bargaining Solution in Labor Markets
2019-10	Angelika Endres Joachim Heinzel	The Impact of Product Qualities on Downstream Bundling in a Distribution Channel
2019-09	Xuehai Zhang	Value at Risk and Expected Shortfall under General Semi-parametric GARCH models
2019-08	Xuehai Zhang	A Box-Cox semiparametric multiplicative error model
2019-07	Angelika Endres Joachim Heinzel	Republished as CIE Working Paper 2019-10