SORT 39 (1) January-June 2015, 35-50

A new class of Skew-Normal-Cauchy distribution

Jaime Arrué[1], Héctor W. Gómez[2], Hugo S. Salinas[3] and Heleno Bolfarine[4]

Abstract

In this paper we study a new class of skew-Cauchy distributions inspired on the family extended two-piece skew normal distribution. The new family of distributions encompasses three well known families of distributions, the normal, the two-piece skew-normal and the skew-normal-Cauchy distributions. Some properties of the new distribution are investigated, inference via maximum likelihood estimation is implemented and results of a real data application, which reveal good performance of the new model, are reported.

MSC: 60E05, 62F12

Keywords: Cauchy distribution, kurtosis, maximum likelihood estimation, singular information matrix, skewness, Skew-Normal-Cauchy distribution

1. Introduction

Arnold et al. (2009) introduced a random variable $X \sim ETN(\alpha,\beta)$ with probability density function given by:

$$f_{ETN}(x;\alpha,\beta) = 2c_\alpha \phi(x)\Phi(\alpha|x|)\Phi(\beta x), \quad -\infty < x < \infty, \tag{1}$$

where $\alpha,\beta \in \mathbb{R}$, $c_\alpha = 2\pi/(\pi+2\arctan(\alpha))$, and $\phi(\cdot)$ and $\Phi(\cdot)$ are the density and cumulative distribution functions of the standard $N(0,1)$ distribution, respectively.

[1] Departamento de Matemáticas, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta, Chile. jaime.arrue@uantof.cl

[2] Departamento de Matemáticas, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta, Chile. hector.gomez@uantof.cl

[3] Departamento de Matemáticas, Facultad de Ingeniería, Universidad de Atacama, Copiapó, Chile. hugo.salinas@uda.cl

[4] Departamento de Estatística, IME, Universidad de Sao Paulo, Sao Paulo, Brasil. hbolfar@ime.usp.br
Received: October 2013
Accepted: September 2014

Notice that for the particular case $\alpha = 0$ the well known skew-normal distribution (Azzalini, 1985) with density function given by

$$f_{SN}(x;\beta) = 2\phi(x)\Phi(\beta x), \quad -\infty < x < \infty, \tag{2}$$

is obtained. For $\beta = 0$, one obtains the so called two-piece skew-normal distribution given by Kim (2005), denoted by $\{TN(\alpha) : -\infty < \alpha < \infty\}$ with probability density function given by

$$f_{TN}(x;\alpha) = c_\alpha \phi(x)\Phi(\alpha|x|), \quad -\infty < x < \infty, \tag{3}$$

with c_α as the normalizing constant. Another family of models studied in Nadarajah and Kotz (2003), is generated by using the kernel of the normal distribution, that is,

$$h(x;\lambda) = 2\phi(x)G(\beta x), \quad -\infty < x < \infty, \tag{4}$$

with $\beta \in (-\infty,\infty)$ and $G(\cdot)$ is a symmetric distribution function. A particular case of this class follows by taking $G(\cdot)$ as the CDF of the Cauchy distribution, which as shown by Nadarajah and Kotz (2003), results in a model with the same range of asymmetry, but with greater kurtosis than that of the skew-normal model. The pdf for a random variable X with this distribution, which we denote by $X \sim SNC(\beta)$, can be written as

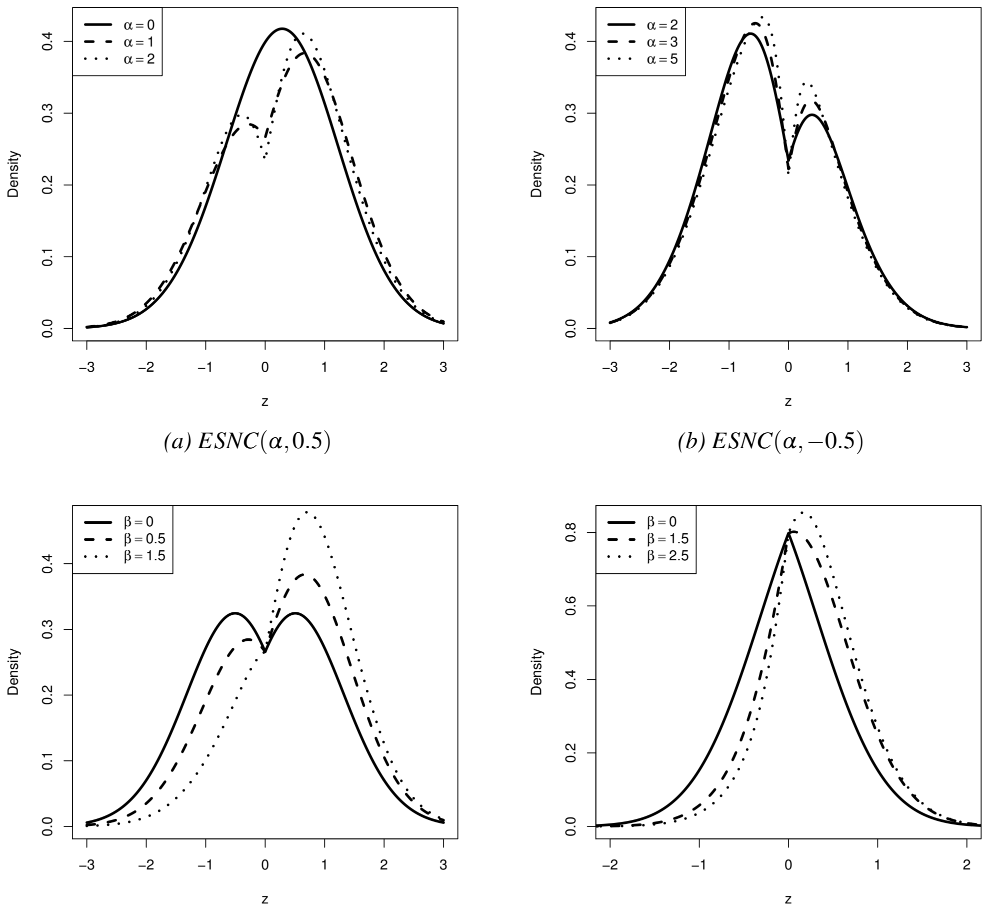
$$f_{SNC}(x;\beta) = 2\phi(x)\left\{\frac{1}{2} + \frac{1}{\pi}\arctan(\beta x)\right\}, \quad -\infty < x < \infty. \tag{5}$$

Arrué, Gómez, Varela and Bolfarine (2010) studied some properties, stochastic representation and information matrix for the model given in (5). A random variable Z has a extended skew-normal-Cauchy random variable with parameter $\alpha, \beta \in (-\infty,\infty)$, denoted $Z \sim ESNC(\alpha,\beta)$, if its probability density function is

$$f(z;\alpha,\beta) = 2c_\alpha \phi(z)\Phi(\alpha|z|)\left\{\frac{1}{2} + \frac{1}{\pi}\arctan(\beta z)\right\}, \quad -\infty < z < \infty. \tag{6}$$

For the rest of the article, Z will denote a random variable with density (6). Figures 1 depicts shapes of density function (6) for different parameter values (continuous and discontinuous lines).

This model is important because it contains strictly (not as limiting cases) the normal, SNC and TN distributions. Moreover, this distribution inherits the bimodal nature of the TN model which is controlled by parameter α, that is, when $\alpha > 0$ the model is bimodal and when $\alpha < 0$ it is unimodal. Since it contains the Cauchy distribution, greater flexibility in the kurtosis is earned and therefore could better fit data sets containing outlying observations.



(a) $ESNC(\alpha, 0.5)$ *(b)* $ESNC(\alpha, -0.5)$

(c) $ESNC(1, \beta)$ *(d)* $ESNC(-1, \beta)$

Figure 1: *Examples of the ESNC density.*

One of the main focus of the paper is to develop a stochastic representation of the ESNC model which allows moments derivation in a simpler way. We derive also the Fisher information matrix for the ESNC model and show that it is singular for $\alpha = \beta = 0$. Using the approach in Rotnitzky, Cox, Bottai and Robins (2000), an alternative parametrization is proposed which makes the Fisher information matrix nonsingular at $\alpha = \beta = 0$.

The paper is organized as follows. Section 2 presents properties of the ESNC model. Section 3 presents a stochastic representation for this model which allows a simple derivation for the moments generating function leading to simple expressions for asymmetry and kurtosis coefficients. The Fisher information matrix is derived in Section 4.2, which turns out to be singular for $\alpha = \beta = 0$. A parametrization is studied which makes it nonsingular for $\alpha = \beta = 0$ and allow an asymptotic study of the MLE properties at this point. In Section 5 we use a data set to illustrate the flexibility of the model ESNC, for this we use the maximum likelihood approach and compare it with the TN and SNC models. The paper is concluded with a discussion section.

2. Distributional properties of the ESNC model

Clearly, density (6) is continuous at $z = 0$ for all α and β, However, it is not differentiable at $z = 0$ for $\alpha \neq 0$. In the following we present uni/bimodal properties possessed by the ESNC family. Notice that this model contains the normal, two-piece skew-normal and skew-normal-cauchy as special cases. The following properties follow immediately from the (6).

Property 1 *The $ESNC(0,0)$ density is the $N(0,1)$ density.*

Property 2 *The $ESNC(0,\beta)$ density is the $SNC(\beta)$ density.*

Property 3 *The $ESNC(\alpha,0)$ density is the $TN(\alpha)$ density.*

Property 4 *As $\alpha \to \infty$, $f(z;\alpha,\beta)$ tends to the $SNC(\beta)$ density. In contrast, as $\alpha \to -\infty$, $f(z;\alpha,\beta)$ degenerates at 0.*

Property 5 *As $\beta \to \infty$, $f(z;\alpha,\beta)$ tends to the $2c_\alpha\phi(z)\Phi(\alpha z)I(z \geq 0)$ density. In contrast, as $\beta \to -\infty$, $f(z;\alpha,\beta)$ tends to the $2c_\alpha\phi(z)\Phi(-\alpha z)I(z < 0)$ density.*

Property 6 *If $Z \sim ESNC(\alpha,\beta)$ random variable, then $-Z \sim ESNC(\alpha,-\beta)$ random variable.*

Property 7 *For $\alpha > 0$, the density (6) is bimodal, i.e. in each region of $z \in (-\infty,0]$ and $z \in [0,\infty)$, $\log f(z;\alpha,\beta)$ is a concave function of z.*

Property 8 *For $\alpha > 0$, the two modes of (6) are located at $z = z_0$ and $z = z_1$ satisfying*

$$z_0 = -\alpha\frac{\phi(\alpha z_0)}{\Phi(-\alpha z_0)} + \frac{\beta}{\pi(1+\beta^2 z_0^2)} \quad \textit{and} \quad z_1 = \alpha\frac{\phi(\alpha z_1)}{\Phi(\alpha z_1)} + \frac{\beta}{\pi(1+\beta^2 z_1^2)},$$

where $z_0 < 0$ and $z_1 > 0$.

Property 9 *For $\alpha < 0$, the single mode of (6) is located at $z = 0$, because $f'(z;\alpha,\beta) < 0$ for $z > 0$ and $f'(z;\alpha,\beta) > 0$ for $z < 0$.*

3. A stochastic representation

The main result states that if $Z \sim ESNC(\alpha, \beta)$ then the distribution of Z can be obtained as a mixture in the asymmetry parameter between the extended two-piece skew-normal and half-normal (HN) distributions. In the following $I(A)$ denotes the indicator function of the set A.

Proposition 1 *If $Z|Y = y \sim ETN(\alpha, \beta y)$ and $Y \sim HN(0,1)$ then $Z \sim ESNC(\alpha, \beta)$.*

Proof. Let $Z|Y = y \sim ETN(\alpha, \beta y)$ and $Y \sim 2\phi(y)I(y \geq 0)$, then

$$
\begin{aligned}
f(z;\alpha,\beta) &= \int_0^\infty 2c_\alpha \phi(z)\Phi(\alpha|z|)\Phi(\beta yz)2\phi(y)dy \\
&= 2c_\alpha \phi(z)\Phi(\alpha|z|)\int_0^\infty 2\Phi(\beta yz)\phi(y)dy \\
&= 4c_\alpha \phi(z)\Phi(\alpha|z|)\int_0^\infty \int_{-\infty}^{\beta z} \phi(t)\phi(y)dtdy \\
&= 4c_\alpha \phi(z)\Phi(\alpha|z|)\left[\int_0^\infty \int_{-\infty}^{0} \phi(t)\phi(y)dtdy + \int_0^\infty \int_{0}^{\beta z} \phi(t)\phi(y)dtdy\right]
\end{aligned}
$$

The terms $\int_0^\infty \int_{-\infty}^{0} \phi(t)\phi(y)dtdy$ and $\int_0^\infty \int_{0}^{\beta x} \phi(t)\phi(y)dtdy$ are the integrals of the bivariate normal distribution. Then, making changes in variables $t = r\cos u$ and $y = r\sin u$ we have

$$
\begin{aligned}
f(z;\alpha,\beta) &= 4c_\alpha \phi(z)\Phi(\alpha|z|)\left[\frac{1}{4} + \frac{1}{2\pi}\int_0^{\arctan(\beta z)} \int_0^{\frac{\pi}{2}} e^{-r^2/2} r dr du\right] \\
&= 2c_\alpha \phi(z)\Phi(\alpha|z|)\left\{\frac{1}{2} + \frac{1}{\pi}\arctan(\beta z)\right\},
\end{aligned}
$$

which concludes the proof. ■

3.1. Location and scale extension

For applications it is convenient to add location and scale parameters to the ESNC distribution. If $Z \sim ESNC(\alpha, \beta)$ and if $X = \mu + \sigma Z$, where $\mu \in (-\infty, \infty)$ and $\sigma > 0$, then we can write $X \sim ESNC(\mu, \sigma, \alpha, \beta)$ or, at times, $X \sim ESNC(\theta)$ where $\theta = (\mu, \sigma, \alpha, \beta)$'. This leads to the following definition.

Definition 1 *A random variable X has a distribution in the ESNC location and scale family if the density is given by*

$$f(x;\theta) = \frac{2c_\alpha}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\alpha\left|\frac{x-\mu}{\sigma}\right|\right)\left\{\frac{1}{2}+\frac{1}{\pi}\arctan\left(\frac{\beta(x-\mu)}{\sigma}\right)\right\}, \quad -\infty < x < \infty. \tag{7}$$

We write $X \sim ESNC(\theta)$ or $X \sim ESNC(\mu,\sigma,\alpha,\beta)$.

3.2. Moments

In order to evaluate moments of the ESNC distribution, the following technical propositions will be useful. In these propositions, we use the notation

$$a_r(\alpha,\lambda) := \int_0^\infty 2c_\alpha t^r\phi(t)\Phi(\lambda t)dt, \tag{8}$$

and

$$d_r(\alpha,\beta) := \int_0^\infty 2c_\alpha t^r\phi(t)\Phi(\alpha t)\Phi(\beta t)dt, \tag{9}$$

where $\alpha,\beta,\lambda \in (-\infty,\infty)$.

We provide next the recursive formulation for computing the functions above for a random variable with density given in (6) which will be fundamental for computing moments of the random variable $X \sim ESNC(\theta)$. The proof is presented in Arnold et al. (2009).

Proposition 2 *According to (8),*

$$a_r(\alpha,\lambda) = \begin{cases} \frac{\pi+2\arctan(\lambda)}{\pi+2\arctan(\alpha)}, & r=0,\\ \frac{c_\alpha}{\sqrt{2\pi}}\left(1+\frac{\lambda}{\sqrt{1+\lambda^2}}\right), & r=1,\\ (r-1)a_{r-2}(\alpha,\lambda)+\frac{2^{r/2-1}\lambda c_\alpha}{\pi(1+\lambda^2)^{r/2}}\Gamma\left(\frac{r}{2}\right), & r\geq 2, \end{cases} \tag{10}$$

Proposition 3 *Let $U \sim TSN(\alpha)$, then*

$$a_r(\alpha) := E(U^r) = \begin{cases} 1, & r=0,\\ \frac{c_\alpha}{\sqrt{2\pi}}\left(1+\frac{\alpha}{\sqrt{1+\alpha^2}}\right), & r=1,\\ (r-1)a_{r-2}(\alpha)+\frac{2^{r/2-1}\alpha c_\alpha}{\pi(1+\alpha^2)^{r/2}}\Gamma\left(\frac{r}{2}\right), & r\geq 2. \end{cases} \tag{11}$$

This result is obtained for $\lambda = \alpha$ in Equation (10).

Proposition 4 *According to (9),*

$$d_r(\alpha,\beta)=\begin{cases}\int_0^\infty 2c_\alpha\phi(t)\Phi(\alpha t)\Phi(\beta t)dt, & r=0,\\ \frac{c_\alpha}{\sqrt{2\pi}}\left[\frac{1}{2}+\frac{\alpha}{\sqrt{1+\alpha^2}}\Psi\left(\frac{\beta}{\sqrt{1+\alpha^2}}\right)+\frac{\beta}{\sqrt{1+\beta^2}}\Psi\left(\frac{\alpha}{\sqrt{1+\beta^2}}\right)\right], & r=1,\\ (r-1)d_{r-2}(\alpha,\beta)+\frac{\alpha}{\sqrt{2\pi}(1+\alpha^2)^{r/2}}a_{r-1}(\alpha,\lambda_1)+\frac{\beta}{\sqrt{2\pi}(1+\beta^2)^{r/2}}a_{r-1}(\alpha,\lambda_2), & r\geq 2,\end{cases} \tag{12}$$

where $\Psi(t)=1/2+\arctan(t)/\pi$ is a CDF of the standard Cauchy distribution, $\lambda_1=\beta/\sqrt{1+\alpha^2}$, $\lambda_2=\alpha/\sqrt{1+\beta^2}$ and $d_0(\alpha,\beta)$ must be evaluated numerically.

Proposition 5 *Let* $Z\sim ESNC(\alpha,\beta)$, $Y\sim 2\phi(y)I(y\geq 0)$ *and* $X=\mu+\sigma Z\sim ESNC(\theta)$ *so that, for* $r=1,2,\ldots$, *we have:*

$$E(Z^r)=(1-(-1)^r)E(d_r(\alpha,\beta Y))+(-1)^r a_r(\alpha) \quad and \quad E(X^r)=\sum_{k=0}^{r}\binom{r}{k}\mu^{r-k}\sigma^k E(Z^k), \tag{13}$$

where $a_r(\alpha)$ *and* $d_r(\alpha,\cdot)$ *are given in (11) and (12), respectively.*

Proof. For computing moments of the random variable $Z\sim ESNC(\alpha,\beta)$ we use conditional expectations and the stochastic representation given in Proposition 1, leading to

$$\begin{aligned}E(Z^r)=E(E(Z^r|Y)) &= \int_0^\infty[(1-(-1)^r)d_r(\alpha,\beta y)+(-1)^r a_r(\alpha)]2\phi(y)dy\\ &= (1-(-1)^r)\int_0^\infty 2d_r(\alpha,\beta y)\phi(y)dy+(-1)^r a_r(\alpha)\int_0^\infty 2\phi(y)dy\\ &= (1-(-1)^r)\int_0^\infty 2d_r(\alpha,\beta y)\phi(y)dy+(-1)^r a_r(\alpha)\\ &= (1-(-1)^r)E(d_r(\alpha,\beta Y))+(-1)^r a_r(\alpha).\end{aligned}$$

∎

Corollary 1 *If* $Z\sim ESNC(\alpha,\beta)$, *then*

$$E(Z^r)=\begin{cases} a_r(\alpha), & r\ even,\\ 2k_r(\alpha,\beta)-a_r(\alpha), & r\ odd,\end{cases} \tag{14}$$

where $k_r(\alpha,\beta):=E(d_r(\alpha,\beta Y))$.

The even moments of the ESNC distribution coincide with the even moments of the ETN distribution given by Arnold et al. (2009).

In the following we present expressions for computing $k_r(\alpha,\beta)$ when r is odd. The proofs for the results presented next follow directly from (10) and (12).

Proposition 6 *Under the conditions in Proposition 5, we have*

$$k_r(\alpha,\beta)=\begin{cases} \frac{2c_\alpha}{\sqrt{2\pi}}\left[\frac{1}{4}+\frac{\alpha}{\sqrt{1+\alpha^2}}\int_0^\infty \phi(y)\Psi\left(\frac{\beta y}{\sqrt{1+\alpha^2}}\right)dy\right. & r=1,\\ \left.+\beta\int_0^\infty \frac{y\phi(y)}{\sqrt{1+\beta^2y^2}}\Psi\left(\frac{\alpha}{\sqrt{1+\beta^2y^2}}\right)dy\right], & \\ (r-1)k_{r-2}(\alpha,\beta)+\frac{\alpha}{\sqrt{2\pi}(1+\alpha^2)^{r/2}}g_{r-1}(\alpha,\beta)+\frac{\beta}{\sqrt{2\pi}}j_{r-1}(\alpha,\beta), & r=3,5,\ldots \end{cases} \tag{15}$$

where

$$g_r(\alpha,\beta)=\begin{cases} 2c_\alpha\int_0^\infty \phi(y)\Psi\left(\frac{\beta y}{\sqrt{1+\alpha^2}}\right)dy, & r=0,\\ (r-1)g_{r-2}(\alpha,\beta)+\frac{\beta^{1-r}c_\alpha e^{\frac{1+\alpha^2}{2\beta^2}}}{\sqrt{2\pi^3}(1+\alpha^2)^{(1-r)/2}}\Gamma\left(\frac{r}{2}\right)\Gamma\left(1-\frac{r}{2},\frac{1+\alpha^2}{2\beta^2}\right), & r=2,4,\ldots \end{cases} \tag{16}$$

where $\Gamma(a,z)=\int_z^\infty e^{-t}t^{a-1}dt$ *is the incomplete Gamma function.*

$$j_r(\alpha,\beta)=\begin{cases} 2c_\alpha\int_0^\infty \frac{y\phi(y)}{(1+\beta^2y^2)}\Psi\left(\frac{\alpha}{\sqrt{1+\beta^2y^2}}\right)dy, & r=0,\\ (r-1)j_{r-2}(\alpha,\beta)+\frac{\alpha c_\alpha\Gamma\left(\frac{r}{2}\right)}{2^{-r/2}\pi}\int_0^\infty \frac{y\phi(y)dy}{(1+\alpha^2+\beta^2y^2)^{r/2}\sqrt{1+\beta^2y^2}}, & r=2,4,\ldots \end{cases} \tag{17}$$

The terms $k_r(\alpha,\beta)$, $g_r(\alpha,\beta)$ and $j_r(\alpha,\beta)$ can be calculated using numerical integration for r, α and β. For reference we list the first four moments of the standard ESNC distribution. If $Z\sim ESNC(\alpha,\beta)$ then

$$\begin{aligned} E(Z) &= \frac{c_\alpha}{\sqrt{2\pi}}\left(-\frac{\alpha}{\sqrt{1+\alpha^2}}+\frac{4\alpha}{\sqrt{1+\alpha^2}}\int_0^\infty \phi(y)\Psi\left(\frac{\beta y}{\sqrt{1+\alpha^2}}\right)dy\right. \qquad (18)\\ &+ \left.4\beta\int_0^\infty \frac{y\phi(y)}{\sqrt{1+\beta^2y^2}}\Psi\left(\frac{\alpha}{\sqrt{1+\beta^2y^2}}\right)dy\right),\\ E(Z^2) &= 1+\frac{\alpha c_\alpha}{\pi(1+\alpha^2)}, \qquad (19)\\ E(Z^3) &= 2k_3(\alpha,\beta)-\frac{c_\alpha}{\sqrt{2\pi}}\left(2+\frac{2\alpha}{\sqrt{1+\alpha^2}}+\frac{\alpha}{\sqrt{(1+\alpha^2)^3}}\right), \qquad (20)\\ E(Z^4) &= 3+\frac{\alpha c_\alpha(5+3\alpha^2)}{\pi(1+\alpha^2)^2}, \qquad (21) \end{aligned}$$

Standard expressions for kurtosis and skewness can then be obtained using Equations $(18)-(21)$.

4. ML estimation

4.1. Likelihood

Suppose that we have available a sample of size n, $X_1, X_2, \ldots, X_n$ from an $ESNC(\theta)$ distribution. In principle, the representation $X_i = \mu + \sigma Z_i$ and the four moment expressions for Z given in Equations $(18)-(21)$ could be used to obtain method of moments estimates of the four parameters. However, the approach is not pursued further. Instead, we will discuss the implementation of the maximum likelihood approach for this distribution given that it is more efficient asymptotically. The log-likelihood function of a random sample $(X_1, X_2, \ldots, X_n)$ from an $ESNC(\theta)$ distribution takes the form

$$l(\theta; X_1, X_2, \ldots, X_n) \propto n \log\left(\frac{c_\alpha}{\sigma}\right) - \frac{1}{2}\sum_{i=1}^{n} Z_i^2 + \sum_{i=1}^{n} \log \Phi(\alpha|Z_i|) + \sum_{i=1}^{n} \log \Psi(\beta Z_i), \quad (22)$$

Table 2 (see Appendix) shows the average MLEs of μ, σ, α and β for 1000 random of size n (SD: standard deviation for the 1000 estimates). We do not consider the case of $\beta < 0$, since by the reflection property 2.6, if $X \sim ESNC(0,1,\alpha,-\beta)$ then $-X \sim ESNC(0,1,\alpha,\beta)$. Several parameter values are considered and moderate and large sample sizes are used. The table shows that for large values of α and β the, MLEs tend to overestimate (if positive) the true values of α and β. This overestimation decreases as the true parameter values decrease and as sample size increases. If one wants to reduce the asymptotic bias of the MLEs one can apply the correction approach in Firth (1993), which amounts to penalize the likelihood for a MLE with less bias value.

4.2. The Fisher information matrix

4.2.1. Special cases

In the special case where $\alpha = 0$ and $\beta = 0$ the information matrix for the ESNC model (see Appendix) is singular, that is,

$$\left|I_{(\mu,\sigma,0,0)}\right| = \begin{vmatrix} \frac{1}{\sigma^2} & 0 & 0 & \frac{2}{\pi\sigma} \\ 0 & \frac{2}{\sigma^2} & \frac{2}{\pi\sigma} & 0 \\ 0 & \frac{2}{\pi\sigma} & \frac{2(\pi-2)}{\pi^2} & 0 \\ \frac{2}{\pi\sigma} & 0 & 0 & \frac{4}{\pi^2} \end{vmatrix} = 0.$$

Comparing the above information matrix with the Fisher information matrix corresponding to model $SNC(\beta)$ given in Arrué et al. (2010) we note that they differ only in the row and column corresponding to the second derivative with respect to the parameter α. The columns corresponding to the parameters μ and β are linearly dependent, so the information matrix is singular. This difficulty has been noticed and investigated in Azzalini (1985) in the context of the skew-normal distribution and was later studied in Chiogna (2005) in some other contexts. DiCiccio and Monti (2004) studied this singularity problem in the context of the skew-exponential power distribution and Salinas, Arellano-Valle and Gómez (2007) studied it in the context of the extended skew-exponential power distribution. In summary, for this special case when the parameters α and β tend to zero, we could not perform asymptotic statistical inference on these parameters, since the information matrix is singular. And to overcome this problem, we will use a reparametrization given by Rotnitzky et al. (2000), which is to transform the score function S_β in one that is linearly independent from the other score functions of score (for the other parameters). With this procedure we obtain a nonsingular information matrix.

4.2.2. Nonsingular Fisher information matrix

As considered in Arrué et al. (2010), we consider next a parameter transformation that makes the information matrix nonsingular. Indeed, after extensive algebraic manipulations, by using the approach in Rotnitzky et al. (2000), it follows that the convenient parametrization is the same as the one derived in Arrué et al. (2010) for the SNC model, namely,

$$\mu^* = \mu + \frac{2}{\pi}\sigma\beta, \quad \sigma^* = \sigma\left(1 - \frac{2}{\pi^2}\beta^2\right), \quad \alpha^* = \alpha, \quad \beta^* = \beta$$

and hence the score vector obtained is $\left(S_\mu, S_\sigma, S_\alpha, S_\beta^3/3!\right)$ where

$$S_\beta^3 = \left.\frac{\partial^3 l(\theta;X^*)}{\partial\beta^3}\right|_{\alpha=\beta=0}.$$

Therefore, to obtain the transformed Fisher information matrix, we have to compute

$$E\left(S_\mu \frac{S_\beta^3}{3!}\right) = \frac{-2}{\pi\sigma}$$
$$E\left(S_\sigma \frac{S_\beta^3}{3!}\right) = 0$$

$$E\left(S_\alpha \frac{S_\beta^3}{3!}\right) = 0$$

$$E\left(\left(\frac{S_\beta^3}{3!}\right)^2\right) = \frac{4}{9\pi^6}\left[96+\pi^2\left(\frac{1}{\sigma^2}-48\right)+15\pi^4\right]$$

leading to the nonsingular Fisher information matrix for the ESNC model

$$I_{(\mu,\sigma,0,0)} = \begin{pmatrix} \frac{1}{\sigma^2} & 0 & 0 & \frac{-2}{\pi\sigma} \\ 0 & \frac{2}{\sigma^2} & \frac{2}{\pi\sigma} & 0 \\ 0 & \frac{2}{\pi\sigma} & \frac{2(\pi-2)}{\pi^2} & 0 \\ \frac{-2}{\pi\sigma} & 0 & 0 & \frac{4}{9\pi^6}\left[96+\pi^2\left(\frac{1}{\sigma^2}-48\right)+15\pi^4\right]. \end{pmatrix}$$

Comparing this information matrix with the one in Arrué et al. (2010) for the $SNC(\beta)$, it follows that they differentiate only on the row and column corresponding to the additional parameter α. Hence, computing the inverse $(I^*)^{-1}$ we have the asymptotic variance of the maximum likelihood estimators for the parameters μ, σ, α and β, respectively.

5. Illustration

To illustrate the estimation procedure discussed in the previous section we consider the variable N-Cream available in the data base Creaminess of cream cheese (see Urlhttp://www.models.kvl.dk/Cream) which was used by Arnold et al. (2009). The corresponding descriptive statistics for this variable are given by the sample size $n = 240$, the mean $\bar{x} = 7.578$ and the variance $s^2 = 2.964$. Quantities $\sqrt{b_1} = -0.551$ and $b_2 = 3.173$ correspond to the sample asymmetry and kurtosis coefficients, respectively. In Table 1, the five models Normal (N), SNC, mixture (MIX), ETN and ESNC with additional location and scale parameters are fitted to the data. MIX is a mixture of two normal distributions represented by $f_Z(z;\mu,\sigma,\mu_1,\sigma_1,p) = p\frac{1}{\sigma}\phi\left(\frac{z-\mu}{\sigma}\right) + (1-p)\frac{1}{\sigma_1}\phi\left(\frac{z-\mu_1}{\sigma_1}\right)$. Notice that the N and SNC models are nested within the ESNC model, so that likelihood ratio tests will provide meaningful comparisons for these models.

In all cases, the parameters are estimated by maximum likelihood using the R-package `optim` (2011). The standard errors of the maximum likelihood estimates are calculated using the information matrix corresponding to each model.

The summaries provided by Table 1 illustrate a key feature of the ESNC model; its flexibility and the wide range of coefficients of skewness and kurtosis that it can adapt to, in contrast to the other models. For example, it is clear that the fit of the normal model is inadequate because of the high degree of skewness of the data. To compare the ESNC model with the normal and SNC models, consider testing the null hypothesis of a normal or a SNC distribution against an ESNC distribution using the likelihood ratio statistics based on the ratios $\Lambda_1 = L_N(\widehat{\mu},\widehat{\sigma},\widehat{\alpha})/L_{ESNC}(\widehat{\mu},\widehat{\sigma},\widehat{\alpha},\widehat{\beta})$ and $\Lambda_2 = L_{SNC}(\widehat{\mu},\widehat{\sigma},\widehat{\alpha})/L_{ESNC}(\widehat{\mu},\widehat{\sigma},\widehat{\alpha},\widehat{\beta})$. Substituting the estimated values, we obtain $-2\log(\Lambda_1) = -2(-469.5862+461.555) = 16.062$ and $-2\log(\Lambda_2) = -2(-466.036+461.555) = 8.962$ which, when compared with the 95% critical value of the $\chi_1^2 = 3.84$, indicate that the null hypotheses are clearly rejected and there is strong indication that the ESNC distribution presents a much better fit than either the N or the SNC distribution to the data set under consideration. In particular, there are significant differences between normal and ESNC models, so not for use reparametrization Rotnitzky et al. (2000). The conclusion of these analysis is that the ESNC model appears to be more appropriate for the particular data set analyzed here. Moreover, using the AIC criterion to MIX, ETN and ESNC models, we can conclude that the ESNC distribution fits better the data. Furthermore, using the delta-method to the information matrix (see Appendix) we have calculated the population estimates of the mean and variance (and their standard deviations), given by $\widehat{E(X)} = 7.596(0.007)$ and $\widehat{V(X)} = 2.897(0.002)$. These points are illustrated in more detail in Figure 2 where the histograms and the fitted curves for the data sets are displayed.

6. Discussion

The paper introduced an extension of the SNC model in Arrué et al. (2010) based on the model defined in Arnold et al. (2009). Some properties of the model are studied and inference is implemented via the maximum likelihood approach. The Fisher information matrix is derived and it is shown to be singular in the vicinity of symmetry. A parameter transformation is presented which contours the singularity problem, and which turn out to be exactly the one derived for the model studied in Arrué et al. (2010). A data set illustration reveals the good performance of the model introduced.

7. Appendix

A. In the formula of Equation (16), use the following integral:

$$\int_0^\infty \frac{y\phi(y)dy}{(1+\alpha^2+\beta^2y^2)^{r/2}} = \frac{e^{\frac{1+\alpha^2}{2\beta^2}}}{2^{\frac{(r+1)}{2}}\sqrt{\pi}|\beta|^r}\Gamma\left(1-\frac{r}{2},\frac{1+\alpha^2}{2\beta^2}\right).$$

Proof. Using the `Integrate[ ]` of Mathematica (2008) we have the result. ■

B. $\delta_k = E\left(\mathrm{sgn}(Z)Z^k\left(\frac{\phi(\alpha Z)}{\Phi(\alpha|Z|)}\right)\right) = \frac{4\sqrt{2}c_\alpha}{\pi^{3/2}}\int_0^\infty z^k\phi(\sqrt{1+\alpha^2}z)\arctan(\beta z)dz$

C. The score functions are given by

$$
\begin{aligned}
\frac{\partial l(\theta;\underline{X})}{\partial\mu} &= \sum_{i=1}^n \frac{Z_i}{\sigma} - \frac{\alpha}{\sigma}\sum_{i=1}^n \frac{\phi(\alpha Z_i)}{\Phi(\alpha|Z_i|)}\mathrm{sgn}(Z_i) - \frac{\beta}{\sigma}\sum_{i=1}^n \frac{\psi(\beta Z_i)}{\Psi(\beta Z_i)},\\
\frac{\partial l(\theta;\underline{X})}{\partial\sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^n Z_i^2 - \frac{\alpha}{\sigma}\sum_{i=1}^n \frac{\phi(\alpha Z_i)}{\Phi(\alpha|Z_i|)}|Z_i| - \frac{\beta}{\sigma}\sum_{i=1}^n \frac{\psi(\beta Z_i)}{\Psi(\beta Z_i)}Z_i,\\
\frac{\partial l(\theta;\underline{X})}{\partial\alpha} &= -\frac{nc_\alpha}{\pi(1+\alpha^2)} + \sum_{i=1}^n \frac{\phi(\alpha Z_i)}{\Phi(\alpha|Z_i|)}|Z_i|,\\
\frac{\partial l(\theta;\underline{X})}{\partial\beta} &= \sum_{i=1}^n \frac{\psi(\beta Z_i)}{\Psi(\beta Z_i)}Z_i.
\end{aligned}
$$

where $\psi(t) = 1/(\pi(1+t^2))$ is a PDF of the standard Cauchy distribution.

D. For one observation $X \sim ESNC(\theta)$, the ij-th element of the information matrix I is given by

$$
I_{\theta_i\theta_j} = -E\left[\frac{\partial^2 l(\theta;X)}{\partial\theta_i\partial\theta_j}\right], \tag{23}
$$

Eventually, one obtains the following expressions for the elements of the information matrix.

$$
\begin{aligned}
I_{\mu\mu} &= \frac{1}{\sigma^2} + \frac{\alpha^3 c_\alpha}{\sigma^2\pi(1+\alpha^2)} - \frac{\alpha^2}{\sigma^2}\eta_0 + \frac{\beta^2}{\sigma^2}\rho_0,\\
I_{\mu\sigma} &= \frac{2}{\sigma^2}E(Z) - \frac{1}{\sigma^2}\alpha\delta_0 + \frac{1}{\sigma^2}\alpha^3\delta_2 + \frac{\alpha^2}{\sigma^2}\eta_1 - \frac{\beta}{\sigma^2}\xi + \frac{2\pi\beta^3}{\sigma^2}\tau + \frac{\beta^2}{\sigma^2}\rho_1,\\
I_{\mu\alpha} &= \frac{1}{\sigma}\delta_0 - \frac{1}{\sigma}\alpha^2\delta_2 - \frac{\alpha}{\sigma}\eta_1,\\
I_{\mu\beta} &= \frac{\xi}{\sigma} - \frac{2\pi\beta^2}{\sigma}\tau - \frac{\beta}{\sigma}\rho_1,
\end{aligned}
$$

$$
\begin{aligned}
I_{\sigma\sigma} &= \frac{2}{\sigma^2} + \frac{\alpha(1+3\alpha^2)c_\alpha}{\sigma^2\pi(1+\alpha^2)^2} + \frac{\alpha^2}{\sigma^2}\eta_2 + \frac{\beta^2}{\sigma^2}\rho_2,\\
I_{\sigma\alpha} &= \frac{c_\alpha(1-\alpha^2)}{\sigma\pi(1+\alpha^2)^2} - \frac{\alpha}{\sigma}\eta_2,\\
I_{\sigma\beta} &= -\frac{\beta}{\sigma}\rho_2,\\
I_{\alpha\alpha} &= -\frac{c_\alpha^2}{\pi^2(1+\alpha^2)^2} + \eta_2,
\end{aligned}
$$

$$I_{\alpha\beta} = 0,$$

$$I_{\beta\beta} = \rho_2,$$

where $\xi = E\left(\frac{\psi(\beta Z)}{\Psi(\beta Z)}\right)$, $\tau = E\left(Z^2\frac{\psi^2(\beta Z)}{\Psi(\beta Z)}\right)$, $\eta_k = E\left(Z^k\left(\frac{\phi(\alpha Z)}{\Phi(\alpha|Z|)}\right)^2\right)$, $\rho_k = = E\left(Z^k\left(\frac{\psi(\beta Z)}{\Psi(\beta Z)}\right)^2\right)$ and $\delta_k = E\left(\mathrm{sgn}(Z)Z^k\left(\frac{\phi(\alpha Z)}{\Phi(\alpha|Z|)}\right)\right)$ must be evaluated numerically, with $Z \sim ESNC(\alpha,\beta)$.

Table 1: *Estimated parameters and log-likelihood values for the models N, SNC, MIX, ETN and ESNC for the N-Cream variable. The corresponding standard errors are in parentheses.*

MLE	N	SNC	MIX	ETN	ESNC
μ	7.577(0.110)	9.142(0.161)	6.082(1.203)	6.712(0.117)	6.717(0.104)
σ	1.712(0.078)	2.320(0.152)	1.558(0.498)	1.783(0.096)	1.781(0.094)
α	–	−4.095(1.155)	–	1.855(0.808)	1.863(0.810)
β	–	–	–	0.590(0.122)	1.062(0.267)
μ_1	–	–	8.435(0.257)	–	–
σ_1	–	–	1.097(0.138)	–	–
p	–	–	0.364(0.245)	–	–
Log-lik	−469.586	−466.036	−461.125	−463.671	−461.555
AIC	943.172	938.072	932.250	935.342	931.110

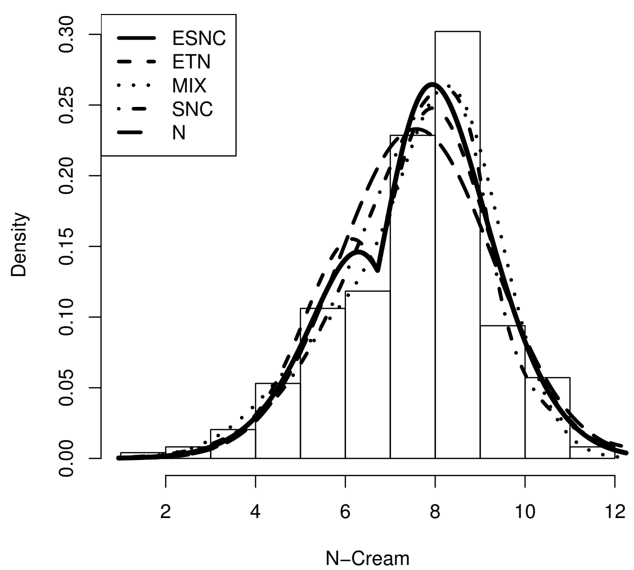


Figure 2: *Histogram for the N-Cream variable. The curves represent densities fitted by maximum likelihood.*

Table 2: *MLEs for the ESNC distribution.*

μ	σ	α	β	n	$\widehat{\mu}$(SD)	$\widehat{\sigma}$(SD)	$\widehat{\alpha}$(SD)	$\widehat{\beta}$(SD)
0	1	4	4	100	0.014 (0.007)	0.983 (0.009)	5.155 (0.507)	4.312 (0.173)
0	1	4	4	300	0.005 (0.002)	0.992 (0.003)	4.830 (0.197)	4.114 (0.053)
0	1	4	4	500	0.006 (0.001)	0.995 (0.002)	4.698 (0.120)	4.038 (0.031)
0	1	4	2	100	0.007 (0.007)	0.991 (0.009)	5.061 (0.494)	2.175 (0.082)
0	1	4	2	300	0.001 (0.002)	0.996 (0.003)	4.886 (0.196)	2.053 (0.025)
0	1	4	2	500	0.001 (0.001)	0.997 (0.002)	4.690 (0.120)	2.033 (0.015)
0	1	4	0	100	-0.011 (0.008)	1.021 (0.008)	4.625 (0.422)	0.031 (0.026)
0	1	4	0	300	-0.005 (0.003)	1.006 (0.003)	5.000 (0.195)	0.015 (0.007)
0	1	4	0	500	-0.003 (0.002)	1.003 (0.002)	4.877 (0.122)	0.009 (0.004)
0	1	2	4	100	0.044 (0.009)	0.977 (0.009)	2.801 (0.268)	4.148 (0.184)
0	1	2	4	300	0.028 (0.003)	0.988 (0.003)	2.540 (0.100)	3.937 (0.056)
0	1	2	4	500	0.021 (0.002)	0.990 (0.002)	2.440 (0.063)	3.910 (0.033)
0	1	2	2	100	0.035 (0.010)	0.985 (0.009)	2.503 (0.222)	2.043 (0.083)
0	1	2	2	300	0.020 (0.003)	0.994 (0.003)	2.366 (0.087)	1.980 (0.026)
0	1	2	2	500	0.014 (0.002)	0.995 (0.002)	2.298 (0.054)	1.978 (0.016)
0	1	2	0	100	-0.006 (0.011)	1.020 (0.008)	2.149 (0.183)	0.016 (0.027)
0	1	2	0	300	-0.003 (0.004)	1.006 (0.003)	2.175 (0.077)	0.004 (0.008)
0	1	2	0	500	-0.002 (0.002)	1.004 (0.002)	2.156 (0.050)	0.002 (0.004)
0	1	0	4	100	0.008 (0.011)	1.207 (0.098)	0.376 (0.196)	5.713 (0.526)
0	1	0	4	300	-0.003 (0.005)	1.149 (0.025)	0.143 (0.056)	4.832 (0.128)
0	1	0	4	500	-0.004 (0.004)	1.118 (0.014)	0.077 (0.033)	4.591 (0.071)
0	1	0	2	100	0.035 (0.012)	1.164 (0.084)	0.283 (0.162)	2.479 (0.228)
0	1	0	2	300	0.031 (0.006)	1.089 (0.019)	0.099 (0.044)	2.139 (0.058)
0	1	0	2	500	0.027 (0.004)	1.068 (0.009)	0.040 (0.024)	2.063 (0.031)
0	1	0	0	100	-0.008 (0.014)	1.125 (0.048)	0.469 (0.123)	0.027 (0.054)
0	1	0	0	300	0.005 (0.007)	1.087 (0.014)	0.107 (0.032)	-0.010 (0.017)
0	1	0	0	500	-0.001 (0.005)	1.062 (0.007)	0.036 (0.016)	-0.001 (0.011)
0	1	-2	0	100	0.005 (0.003)	1.234 (0.348)	-2.446 (0.805)	0.012 (0.352)
0	1	-2	0	300	0.002 (0.002)	1.258 (0.185)	-2.502 (0.422)	0.010 (0.110)
0	1	-2	0	500	0.001 (0.001)	1.248 (0.124)	-2.499 (0.285)	0.002 (0.067)
0	1	-4	0	100	0.002 (0.002)	0.902 (0.301)	-3.539 (1.270)	0.077 (0.465)
0	1	-4	0	300	0.001 (0.001)	0.925 (0.139)	-3.654 (0.587)	0.014 (0.154)
0	1	-4	0	500	0.000 (0.001)	0.938 (0.098)	-3.714 (0.414)	0.003 (0.081)
0	1	-2	4	100	-0.012 (0.003)	1.007 (0.279)	-1.937 (0.679)	5.233 (1.322)
0	1	-2	4	300	-0.010 (0.002)	0.970 (0.112)	-1.858 (0.274)	4.417 (0.467)
0	1	-2	4	500	-0.007 (0.001)	0.969 (0.076)	-1.871 (0.184)	4.171 (0.293)
0	1	-2	2	100	-0.017 (0.003)	1.116 (0.317)	-2.175 (0.750)	3.057 (0.853)
0	1	-2	2	300	-0.015 (0.002)	1.057 (0.124)	-2.036 (0.293)	2.609 (0.304)
0	1	-2	2	500	-0.013 (0.001)	1.074 (0.091)	-2.083 (0.214)	2.489 (0.203)
0	1	-4	4	100	-0.007 (0.002)	0.904 (0.358)	-3.549 (1.520)	4.468 (1.652)
0	1	-4	4	300	-0.005 (0.001)	0.921 (0.164)	-3.629 (0.690)	4.114 (0.699)
0	1	-4	4	500	-0.003 (0.001)	0.939 (0.112)	-3.712 (0.473)	4.052 (0.471)
0	1	-4	2	100	-0.007 (0.002)	0.958 (0.377)	-3.783 (1.579)	2.662 (1.134)
0	1	-4	2	300	-0.005 (0.001)	0.973 (0.167)	-3.847 (0.698)	2.331 (0.431)
0	1	-4	2	500	-0.003 (0.001)	0.965 (0.106)	-3.821 (0.444)	2.168 (0.246)

Acknowledgments

We thank two referees for comments and suggestions that substantially improved the presentation. The research of H. W. Gómez was supported by SEMILLERO UA-2014 (Chile). The research of H. Bolfarine was supported by CNPq and Fapesp (Brasil).

References

Arnold, B.C., Gómez, H.W., Salinas, H.S. (2009). On multiple constraint skewed models. *Statistics: A Journal of Theoretical and Applied Statistics*, 43, 279–293.

Arrué, J.R., Gómez, H.W., Varela H., Bolfarine, H. (2010). On the skew-normal-Cauchy distribution. *Communications in Statistics: Theory and Methods*, 40, 15–27.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12, 171–178.

Chiogna, M. (2005). A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution. *Statistical Methods & Applications*, 14, 331–341.

DiCiccio, T. J. and Monti, A. C. (2004). Inferential aspects of the skew exponential power distribution. *Journal of the American Statistical Association*, 99, 439–450.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80, 27–38.

Kim, H. J. (2005). On the class of two-piece skew-normal distributions. *Statistics*, 39, 537–553.

Nadarajah, S., Kotz, S. (2003). Skewed distributions generated by the normal kernel. *Statistics & Probability Letters*, 65, 269–277.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL :http://www.R-project.org.

Rotnitzky, A. Cox, D.R., Bottai, M., Robins, J (2000). Likelihood-based inference With singular information matrix. *Bernoulli*, 6, 243–284.

Salinas, H. S., Arellano-Valle, R. B., Gómez, H. W. (2007). Skew-exponential power distribution and its derivation. *Communications in Statistics: Theory and Methods*, 36, 1673–1689.

Wolfram Research, Inc., Mathematica, version 7.0, Champaign, IL (2008).