Elsevier Editorial System(tm) for Statistics and Probability Letters
Manuscript Draft

Manuscript Number: STAPRO-D-15-00050R2

Title: On the extended two-parameter generalized skew-normal distribution

Article Type: Letter

Keywords: Hidden truncation; Sample selection; Extended skew-normal distribution

Corresponding Author: Dr. Emmanuel Olusegun Ogundimu, Ph.D

Corresponding Author's Institution: Centre for Statistics in Medicine

First Author: Emmanuel Olusegun Ogundimu, Ph.D

Order of Authors: Emmanuel Olusegun Ogundimu, Ph.D; Jane L Hutton, PhD

Abstract: We propose a three-parameter skew-normal distribution, obtained by using hidden truncation on a skew-normal random variable. The hidden truncation framework permits direct interpretation of the model parameters. A link is established between the model and the closed skew-normal distribution.

Reply to the Associate Editor:

On the extended two-parameter generalized skew-normal distribution.

E.O. Ogundimu and J.L. Hutton

We thank the Associate Editor for the comments provided. We have worked on the references.

On the extended two-parameter generalized skew-normal distribution

Emmanuel O. Ogundimu[a,*], Jane L. Hutton[b]

[a]*Centre for Statistics in Medicine, University of Oxford, UK.*
[b]*Department of Statistics, University of Warwick, UK.*

Abstract

We propose a three-parameter skew-normal distribution, obtained by using hidden truncation on a skew-normal random variable. The hidden truncation framework permits direct interpretation of the model parameters. A link is established between the model and the closed skew-normal distribution.
Keywords: Hidden truncation; Sample selection; Extended skew-normal distribution

1. Introduction

Hidden truncation models have a long history before Azzalini (1985) popularized and studied the skew-normal (SN) distribution. Birnbaum (1950) studied the distribution and its extensions in the context of educational testing where he showed that the SN distribution can result from linear truncation of multivariate normal random variable. Weinstein (1964), using a convolution of normal and truncated normal random variables derived a distribution similar to SN, expressed implicitly. Roberts (1966) considered the distribution resulting from selecting the maximum or minimum value from suitably standardized measurements taken on a pair of twins. The resulting distribution is also similar to the SN distribution. In the Bayesian context, O'Hagan & Leonard (1976) suggested the use of an extended version of SN distribution as a possible prior for a normal mean. The above early studies showed that simple and common nonlinear operations such as truncation, conditioning and censoring carried out on normal random variables lead to versions of SN random variables.

*Corresponding author
Email addresses: emmanuel.ogundimu@csm.ox.ac.uk (Emmanuel O. Ogundimu), J.L.Hutton@warwick.ac.uk (Jane L. Hutton)

A standard skew-normal distribution has a probability density function (PDF) of the form

$$f(z;\lambda) = 2\phi(z)\Phi(\lambda z), \quad z \in \mathbb{R}, \tag{1}$$

where $\lambda \in \mathbb{R}$ is the skewness parameter because it regulates the shape of the density function. An extension of model (1) can be described as follows: Suppose (X, Y) are standard bivariate normal random variables with correlation ρ, and that the values of X are selected only if $Y > c$, a constant. Then, the PDF of $X \mid Y > c$ is

$$f(x;\lambda_0,\lambda_1) = \phi(x)\Phi(\lambda_0 + \lambda_1 x) \Big/ \Phi\left(\lambda_0 \Big/ \sqrt{1+\lambda_1^2}\right), \quad x \in \mathbb{R},\ \lambda_0, \lambda_1 \in \mathbb{R}, \tag{2}$$

where $\lambda_0 = -c/\sqrt{1-\rho^2}$ and $\lambda_1 = \rho/\sqrt{1-\rho^2}$. Model (2) is called an extended skew-normal (ESN) distribution since it extends model (1) by an additional shift (skewness) parameter λ_0. This extension has been examined by Azzalini (1985) and Arnold et al. (1993) while the multivariate case has been considered by Arnold & Beaver (2000a). A link between this model and the continuous component of sample selection density was established in Copas & Li (1997) and further studied in a multilevel sample selection framework in Ogundimu & Hutton (2014).

The hidden truncation model based on the ESN distribution is formulated from the normal distribution. The normal assumption is used for convenience, but is unrealistic in many real applications. The development of more general hidden truncation models with the normal distribution as a particular case is therefore necessary. Arnold & Beaver (2002) proposed the general method for constructing hidden truncation models. As noted by Arnold (2009), these models are difficult to deal with analytically unless the joint density, (X, Y) is a member of some tractable family of multivariate distribution. There are also challenges related to making inference from these models.

We propose a three-parameter skew-normal distribution which arises from hidden truncation on a skew-normal random variable. A skew-normal random variable is chosen because, as noted by Mudholkar & Hutson (2000), of the most common deviations from normality, skewness and heavier tails, the effects of non-normality due to skewness are generally more serious. For example, assuming symmetry when there is asymmetry leads to biased point estimates of location. We show that the resulting distribution extends the ESN distribution with an additional skewness parameter. Equivalently, the distribution extends the two parameter generalized skew-normal (GSN(λ_1, λ_2)) distribution developed by Jamalizadeh et. al. (2008) by an additional shift parameter, λ_0. This formulation also allows us to interpret the parameters in the model in a straightforward manner. In addition, we show the merits of using the proposed model for modelling observational data arising from sample selection.

The rest of the paper is organised as follows. In section 2, we review the basics of hidden truncation models and describe the model for a skew-normal random variable. In section 3, a hidden truncation model arising from skew-normal random variable is introduced and its properties are studied. The model is applied to a data set in Section 4. A link is established between the model and sample selection models in section 5 and conclusions given in section 6.

2. Basis of truncation models

Consider a two dimensional absolutely continuous random vector (X, Y). The conditional distribution of X given $Y \in C$, where C is a Borel set in $\mathbb{R}$, has its PDF given by

$$f(X|Y \in C) = \frac{f(x)P(Y \in C|X = x)}{P(Y \in C)}, \tag{3}$$

using Bayes' rule for the decomposition of the density $f(X|Y \in C)$, (see Arellano-Valle et al., 2006). Selection distributions depend on the subset C. The usual selection subset is a half-line, defined by

$$C(\beta) = \{y \in \mathbb{R}|\, y > \beta\},$$

where β is the truncation point. The hidden truncation equivalent of (3) consists of upper and lower truncation subset defined by

$$C(\alpha, \beta) = \{y \in \mathbb{R}|\, \alpha > y > \beta\}. \tag{4}$$

The use of subset (4) is the basis of the model considered in Arnold et al. (1993).

For this paper we focus on the selection subset $C(0)$ which leads to an extension of the ESN model (2). Note that $C(\beta)$ and $C(0)$ differ only by location change, since no symmetry around 0 is assumed. In this case, (3) can be written as

$$f(x|Y > 0) = \frac{f(x)P(Y > 0|X = x)}{P(Y > 0)}, \tag{5}$$

which corresponds to equation (5.1) of Arellano-Valle et al. (2006). Equation (5) can be described as a weighted version of the original density function of X. A special case of (5) can be derived using the proposition below, which only requires the assumption of independence between X and Y.

Proposition 1. *Suppose X and Y are two independent random variables, with arbitrary and possibly different distribution. The variable X is observed only if Y satisfies the constraints $\lambda_0 + \lambda_1 X > Y$. Assume X has density function ψ_1 with associated distribution function Ψ_1 and Y has density ψ_2 with distribution function Ψ_2. The conditional density of $X|\lambda_0 + \lambda_1 X > Y$ is*

$$f(x|\lambda_0 + \lambda_1 X > Y) = \frac{\psi_1(x)\Psi_2(\lambda_0 + \lambda_1 x)}{P(\lambda_0 + \lambda_1 X > Y)}. \tag{6}$$

Equation (6) is the basis of the ESN density given by (2), in which X and Y are independent normal random variables with X selected when the associated Y exceeds a threshold, which is not necessarily its mean. This density reduces to the density of the random variable X when $\lambda_1 = 0$, regardless of the value of λ_0.

The computation of the denominator, $P(\lambda_0 + \lambda_1 X > Y)$ in (6) may not be available in analytic form unless X and Y are stable random variables such that a tractable expression can be derived for $\lambda_0 + \lambda_1 X - Y$. The case for a Cauchy random variable was discussed in Arnold & Beaver (2000b).

Suppose we apply (6) to independent random variables $X \sim \text{SN}(0, 1, \lambda_x)$ and $Y \sim \text{N}(0, 1)$, where SN and N represent skew-normal and normal distributions respectively, so that the first factor in the numerator of (6) is $2\phi(x)\Phi(\lambda_1 x)$ and the second factor is $\Phi(\lambda_0 + \lambda_2 x)$. Computation of the denominator of (6), $p = P(\lambda_0 > Z)$ where $Z = Y - \lambda_2 X$, amounts to finding the cumulative distribution function (CDF) of Z. What we need is a simple extention of Property I of Azzalini (1985) which gives the distribution of $(Y + X)\big/\sqrt{2}$, while here we have the multiplicative factor $-\lambda_2$; equivalently, we can appeal to Proposition 2.3 of (Azzalini, 2014b, p26) which provides this simple extension. We obtain that $Z \sim \text{SN}(0, 1 + \lambda_2^2, -\lambda_1)$ and consequently $p = 2k(\lambda_0, \lambda_1, \lambda_2)^{-1}$, where $k^{-1} = \frac{1}{2} F_{SN}\left(\frac{\lambda_0}{\sqrt{1+\lambda_2^2}}; \frac{-\lambda_1\lambda_2}{\sqrt{1+\lambda_1^2+\lambda_2^2}}\right)$ and F_{SN} is the standard CDF of the Azzalini's skew-normal distribution (Azzalini, 1985).

This model implies that a skew-normal random variable X is observed only when a concomitant normal random variable Y is greater than zero. The next section provides detailed description of the model.

3. Extended two parameter generalized skew-normal distribution

3.1. Definitions and Simple properties

Definition 1. *A random variable $Z_{\lambda_0,\lambda_1,\lambda_2}$ is said to have an extended two-parameter generalized skew-normal (EGSN) distribution, if its PDF is*

$$f(z;\lambda_0,\lambda_1,\lambda_2) = k(\lambda_0,\lambda_1,\lambda_2)\phi(z)\Phi(\lambda_1 z)\Phi(\lambda_0+\lambda_2 z), \quad z \in \mathbb{R}, \tag{7}$$

where $\lambda_0, \lambda_1, \lambda_2 \in \mathbb{R}$, λ_1 & λ_2 are the skewness parameters and λ_0 is the shift parameter. Since (7) is a PDF, we must have

$$k(\lambda_0,\lambda_1,\lambda_2)^{-1} = \int_{-\infty}^{\infty} \phi(z)\Phi(\lambda_1 z)\Phi(\lambda_0+\lambda_2 z)\,dz = E[\Phi(\lambda_1 X)\Phi(\lambda_0+\lambda_2 X)],$$

where $X \sim N(0,1)$. Direct integration yields

$$k(\lambda_0,\lambda_1,\lambda_2)^{-1} = \Phi_2\Big(0, \frac{\lambda_0}{\sqrt{1+\lambda_2^2}}; \frac{\lambda_1\lambda_2}{\sqrt{1+\lambda_1^2}\sqrt{1+\lambda_2^2}}\Big) = \frac{1}{2}F_{SN}\Big(\frac{\lambda_0}{\sqrt{1+\lambda_2^2}}; \frac{-\lambda_1\lambda_2}{\sqrt{1+\lambda_1^2+\lambda_2^2}}\Big),$$

where Φ_2 is a standard bivariate normal CDF and F_{SN} is as defined in section 2. The evaluation of F_{SN} can be obtained from the 'psn' function in Azzalini's skew-normal package in R (Azzalini, 2014a).

Thus, the extended two-parameter generalized skew-normal density in (7) becomes

$$f(z;\lambda_0,\lambda_1,\lambda_2) = \frac{2}{F_{SN}\Big(\frac{\lambda_0}{\sqrt{1+\lambda_2^2}}; \frac{-\lambda_1\lambda_2}{\sqrt{1+\lambda_1^2+\lambda_2^2}}\Big)}\phi(z)\Phi(\lambda_1 z)\Phi(\lambda_0+\lambda_2 z), \quad z \in \mathbb{R}, \tag{8}$$

and we write $Z_{\lambda_0,\lambda_1,\lambda_2} \sim EGSN(\lambda_0,\lambda_1,\lambda_2)$.

Proposition 2. *For the special case $\lambda_0 = 0$,* (8) *becomes*

$$\frac{2}{F_{SN}\Big(0; \frac{-\lambda_1\lambda_2}{\sqrt{1+\lambda_1^2+\lambda_2^2}}\Big)}\phi(z)\Phi(\lambda_1 z)\Phi(\lambda_2 z), \quad z \in \mathbb{R},$$

which is equivalent to the two parameter generalized skew-normal distribution (GSN(λ_1,λ_2)) given in Jamalizadeh et. al. (2008).

To see this, we note that

$$\frac{2\pi}{\cos^{-1}\Big(\frac{-\lambda_1\lambda_2}{\sqrt{1+\lambda_1^2}\sqrt{1+\lambda_2^2}}\Big)} = \frac{1}{\Phi_2\Big(0,0; \frac{\lambda_1\lambda_2}{\sqrt{1+\lambda_1^2}\sqrt{1+\lambda_2^2}}\Big)} = \frac{2}{F_{SN}\Big(0; \frac{-\lambda_1\lambda_2}{\sqrt{1+\lambda_1^2+\lambda_2^2}}\Big)}.$$

Some properties of the model in (8) are stated below:

1. EGSN$(0, \lambda_1, \lambda_2)$ = GSN(λ_1, λ_2)

2. EGSN$(\lambda_0, 0, \lambda)$ = ESN(λ_0, λ)

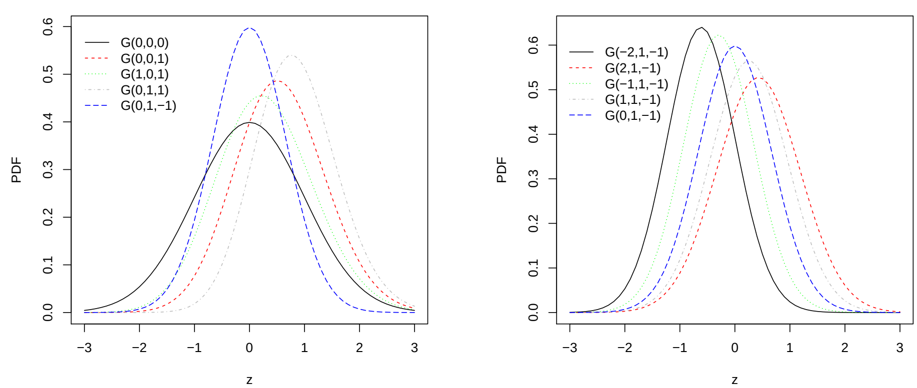
3. EGSN$(0, 0, \lambda)$ = EGSN$(0, \lambda, 0)$ = SN(λ)

4. EGSN$(0, 0, 0)$ = N(0,1)

5. EGSN$(\lambda_0, \lambda_1, \lambda_2)$ can be derived from the convolution of an independent skew-normal random variable and a truncated normal random variable

6. If $Z \sim$ EGSN$(\lambda_0, \lambda_1, \lambda_2)$, then $-Z \sim$ EGSN$(\lambda_0, -\lambda_1, -\lambda_2)$.

The parameters in the EGSN model can be easily interpreted: λ_1 is the population skewness inherent in the X variable, λ_2 is the skewness due to hidden truncation induced by the random variable $Y > 0$ and λ_0 is the shift parameter, which in some sense also regulates kurtosis (Arellano-Valle & Genton, 2010).



(a) (b)

Figure 1: (a) Plots of PDFs of EGSN; (b) Plots of PDFS to illustrate Kurtosis of EGSN.

Figure 1 shows densities of EGSN. This figure further illustrates some of the simple properties of the distribution. A comparison of the density EGSN(0,0,0) and EGSN(0,1,-1) shows that the latter is also symmetric but with different kurtosis.

Proposition 3. *The extended two-parameter generalized skew-normal density function is log concave.*

Proof. To prove that $\log f(z;\lambda_0,\lambda_1,\lambda_2)$ is a concave function of z, it suffices to show that the second derivative of $\log f(z;\lambda_0,\lambda_1,\lambda_2)$ is negative for all $z \in \mathbb{R}$. Now,

$$\frac{d^2 \log f(z;\lambda_0,\lambda_1,\lambda_2)}{dz^2} = -\Big[1+\lambda_1^2\Lambda(\lambda_1 z)\Big(\lambda_1 z+\Lambda(\lambda_1 z)\Big) \\ +\lambda_2^2\Lambda(\lambda_0+\lambda_2 z)\Big((\lambda_0+\lambda_2 z)+\Lambda(\lambda_0+\lambda_2 z)\Big)\Big],$$

where $\Lambda(.) = \phi(.)/\Phi(.)$. Since $\Lambda(.)$ is a positive function, we only need to show that $\lambda_1 z+\Lambda(\lambda_1 z)$ and $(\lambda_0+\lambda_2 z)+\Lambda(\lambda_0+\lambda_2 z)$ are positive for all $z \in \mathbb{R}$.

Case 1: If $\lambda_1 z \geq 0$, then $\lambda_1 z+\Lambda(\lambda_1 z)$ is clearly positive.

Case 2: If $\lambda_1 z < 0$, let $t = -\lambda_1 z$. Then, $\phi(\lambda_1 z) = \phi(-\lambda_1 z) = \phi(t)$ and $\Phi(\lambda_1 z) = 1-\Phi(-\lambda_1 z) = 1-\Phi(t)$. Thus, $\Lambda(\lambda_1 z)+\lambda_1 z = \phi(t)\big/(1-\Phi(t)) - t = r(t) - t$, where $r(t)$ is the failure rate of the standard normal distribution. It is known that $r(t) > t$, so the requirement is proved. The second part of the requirement, $(\lambda_0+\lambda_2 z)+\Lambda(\lambda_0+\lambda_2 z) > 0$, follows from the fact that $t+\Lambda(t) > 0\,\forall\, t \in \mathbb{R}$.

Proposition 4. *The extended two-parameter generalized skew-normal density function is unimodal.*

Proof. The proof follows from proposition 3 and the fact that a nondegenerate distribution F is strongly unimodal if and only if it has a log concave density f (Marshall & Olkin, 2007, p99, proposition B.2.)

3.2. Link between EGSN and CSN distribution

The EGSN distribution, like the GSN, can be linked with the CSN distribution. Briefly, the CSN distribution is defined as follows.

Definition 2. *Consider $p \geq 1$, $q \geq 1$, $\boldsymbol{\mu} \in \mathbb{R}^p$, $\boldsymbol{\nu} \in \mathbb{R}^q$, D an arbitrary $q \times p$ matrix, Σ and Δ positive definite matrices of dimensions $p \times p$ and $q \times q$, respectively. Then the PDF of the CSN distribution is given by:*

$$f_{p,q}(\mathbf{y}) = C\phi_p(\mathbf{y};\boldsymbol{\mu},\Sigma)\Phi_q(D(\mathbf{y}-\boldsymbol{\mu});\boldsymbol{\nu},\Delta), \quad \mathbf{y} \in \mathbb{R}^p, \tag{9}$$

with:

$$C^{-1} = \Phi_q(\mathbf{0};\boldsymbol{\nu},\Delta+D\Sigma D'),$$

where $\phi_p(.;\boldsymbol{\eta},\Psi)$, $\Phi_p(.;\boldsymbol{\eta},\Psi)$ are the PDF and CDF of a p-dimensional normal distribution with mean $\boldsymbol{\eta}\in\mathbb{R}^p$ and $p\times p$ covariance matrix Ψ. We write $\mathbf{y}\sim CSN_{p,q}(\boldsymbol{\mu},\Sigma,D,\boldsymbol{\nu},\Delta)$, if $\mathbf{y}\in\mathbb{R}^p$ is distributed as CSN distribution with parameters $q,\boldsymbol{\mu},D,\Sigma,\boldsymbol{\nu},\Delta$.

The EGSN model inherits properties of the CSN distribution via reparametrisation as equation (9). Thus, (8) can be re-written as

$$\frac{\phi(z)\Phi_2\Big(\lambda_1 z,\lambda_2 z;(0,-\lambda_0),1\Big)}{\Phi_2\Big(0,\lambda_0\Big/\sqrt{1+\lambda_2^2};\lambda_1\lambda_2\Big/\sqrt{1+\lambda_1^2}\sqrt{1+\lambda_2^2}\Big)},$$

which is a CSN density with parameters $\mu=0$, $\Sigma=1$, $D=(\lambda_1,\lambda_2)'$, $\boldsymbol{\nu}=(0,-\lambda_0)'$ and $\Delta=\boldsymbol{I}_2$. The corresponding CDF is given by

$$k(\lambda_0,\lambda_1,\lambda_2)\Phi_3\left(\begin{pmatrix} z\\ 0\\ 0\end{pmatrix};\begin{pmatrix}0\\ 0\\ -\lambda_0\end{pmatrix},\begin{pmatrix}1 & -\lambda_1 & -\lambda_2\\ -\lambda_1 & 1+\lambda_1^2 & \lambda_1\lambda_2\\ -\lambda_2 & \lambda_1\lambda_2 & 1+\lambda_2^2\end{pmatrix}\right).$$

Additional properties of the distribution can be derived using this link. For instance, the sum of $Z_{\lambda_0,\lambda_1,\lambda_2}\sim$ EGSN$(\lambda_0,\lambda_1,\lambda_2)$ and $X\sim N(0,1)$, is

$$\frac{1}{\sqrt{2}}\Big(Z_{\lambda_0,\lambda_1,\lambda_2}+X\Big)\sim CSN_{1,2}\left[0,1,(\lambda_1\Big/\sqrt{2},\lambda_2\Big/\sqrt{2})',(0,-\lambda_0)',\begin{pmatrix}1+\lambda_1^2/2 & \lambda_1\lambda_2/2\\ \lambda_1\lambda_2/2 & 1+\lambda_2^2/2\end{pmatrix}\right],$$

which is an EGSN$(\lambda_0,\lambda_1/\sqrt{2},\lambda_2/\sqrt{2})$ distribution.

The argument leading to density (8) has an implicit method for random number generation: the values of X such that $\lambda_0+\lambda_2 X>Y$ are sampled from its distribution.

3.3. Moments and Maximum Likelihood estimator of the EGSN model

Proposition 5. *If $M(t;\lambda_0,\lambda_1,\lambda_2)$ is the moment generating function of $Z_{\lambda_0,\lambda_1,\lambda_2}\sim EGSN(\lambda_0,\lambda_1,\lambda_2)$, then*

$$M(t;\lambda_0,\lambda_1,\lambda_2)=k(\lambda_0,\lambda_1,\lambda_2)e^{t^2/2}\Phi_2\left(\frac{\lambda_1 t}{\sqrt{1+\lambda_1^2}},\frac{\lambda_0+\lambda_2 t}{\sqrt{1+\lambda_2^2}};\frac{\lambda_1\lambda_2}{\sqrt{1+\lambda_1^2}\sqrt{1+\lambda_2^2}}\right), \qquad (10)$$

where $k(\lambda_0,\lambda_1,\lambda_2)$ is as given in (8).

Proof. The proof is immediate from the moment generating function of the CSN distribution.

The moments of $Z_{\lambda_0,\lambda_1,\lambda_2}$ can be obtained from (10). The mean is given as

$$\begin{aligned} E(Z_{\lambda_0,\lambda_1,\lambda_2}) =& k(\lambda_0,\lambda_1,\lambda_2)\Bigg\{\frac{1}{\sqrt{2\pi}}\frac{\lambda_1}{\sqrt{1+\lambda_1^2}}\Phi\Bigg(\frac{\lambda_0\sqrt{1+\lambda_1^2}}{\sqrt{1+\lambda_1^2+\lambda_2^2}}\Bigg) \\ &+\frac{\lambda_2}{\sqrt{1+\lambda_2^2}}\phi\Bigg(\frac{\lambda_0}{\sqrt{1+\lambda_2^2}}\Bigg)\Phi\Bigg(\frac{-\lambda_0\lambda_1\lambda_2}{\sqrt{1+\lambda_2^2}\sqrt{1+\lambda_1^2+\lambda_2^2}}\Bigg)\Bigg\}. \end{aligned} \tag{11}$$

To fit the model to data, one can introduce the affine transformation $Y = \mu + \sigma Z_{\lambda_0,\lambda_1,\lambda_2} \sim$ EGSN$(\mu, \sigma^2, \lambda_0, \lambda_1, \lambda_2)$. The density becomes

$$f(y;\mu,\sigma^2,\lambda_0,\lambda_1,\lambda_2) = \frac{\frac{2}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right)\Phi\left(\frac{\lambda_1(y-\mu)}{\sigma}\right)\Phi\left(\lambda_0+\lambda_2\left(\frac{y-\mu}{\sigma}\right)\right)}{F_{SN}\left(\frac{\lambda_0}{\sqrt{1+\lambda_2^2}};\frac{-\lambda_1\lambda_2}{\sqrt{1+\lambda_1^2+\lambda_2^2}}\right)}. \tag{12}$$

The log-likelihood function in this case is

$$\begin{aligned} l(\Xi) =& n\ln 2 - \frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \sum_{i=1}^{n}\frac{(y_i-\mu)^2}{\sigma^2} + \sum_{i=1}^{n}\ln\Phi\Bigg(\frac{\lambda_1(y_i-\mu)}{\sigma}\Bigg) \\ &+\sum_{i=1}^{n}\ln\Big[\Phi\Big(\lambda_0+\lambda_2\Big(\frac{y_i-\mu}{\sigma}\Big)\Big)\Big] - n\ln\Big[F_{SN}\Big(\frac{\lambda_0}{\sqrt{1+\lambda_2^2}};\frac{-\lambda_1\lambda_2}{\sqrt{1+\lambda_1^2+\lambda_2^2}}\Big)\Big], \end{aligned}$$

where $\Xi = (\mu, \sigma, \lambda_0, \lambda_1, \lambda_2)$.

The parameters in the EGSN model can be interpreted. If we write (11) as $E(Y) = \mu + \sigma E(Z_{\lambda_0,\lambda_1,\lambda_2})$, then μ is the theoretical mean in the original skew population that is not subjected to hidden truncation. This representation can be used for the evaluation of model fit. In addition, since the EGSN model is an extension of the ESN model, it suffers from parameter identifiability draw-backs as well. For instance, if $\lambda_1 = \lambda_2 = 0$, the distribution becomes the normal distribution regardless of the value of λ_0. The added advantage of the model is that the skewness parameters are distinct and in many practical applications, λ_1 and λ_2 will not be exactly zero simultaneously.

4. Illustrative Example

We consider a data set on married women's labour force participation (Wooldridge, 2002), with female wages as the outcome of interest. The outcome of interest is missing for 325 (43%) of the 753 women in the sample. As the data generating mechanism is from a sample selection setting, it is natural to model the observed part of the data using the EGSN and ESN models. Due to severe skewness, logarithm of wage is used as the response and we consider dependence on education status (1=graduate, 0= not) and city (1=city, 0=rural), i.e. $x = (1, educ, city)$. Our focus will be on complete case analysis with 428 women.

Table 1: MLEs for the Wage offer data using EGSN model and its sub-models

	Normal	S-Normal	GSN	ESN	EGSN
(Intercept)	-0.200	0.399	0.174	15.976	36.121
educ	0.107	0.108	0.109	0.111	0.111
city	0.066	0.096	0.095	0.092	0.086
σ	0.677	0.934	5.606	2.865	12.639
$\hat{\lambda}_0$	-	-	-	-34.671	-78.528
$\hat{\lambda}_1$	-	-	5.547	-	2.818
$\hat{\lambda}_2$	-	-1.853	-11.821	-6.138	-27.306
$\log(L(\theta))$	-440.808	-425.763	-421.844	-415.052	-413.069

Table 1 shows the results of fitting the EGSN and its sub-models to the wage offer data. Even with the logarithm transformation of the response, the effect of skewness is still pronounced (likelihood ratio statistic of Normal model vs. S-Normal model is 30.09, P<0.0001). To avoid the near identifiability problem highlighted in Arnold et al. (1993) for the ESN model, we estimated the parameters of the EGSN and ESN models using profile log-likelihood constructed as a function of λ_0. The ESN model fits better than the GSN for the same number of parameters. This indicates that the data was, perhaps, generated by selection above a threshold (expressed through λ_0) rather than by the double skewing of the GSN(λ_1, λ_2). This result is not surprising given that hidden truncation and sample selection models are related. Although both λ_1 and λ_2 control skewness in the EGSN and GSN models, their roles are not the same. The former is the inherent skewness in the original data while the latter is due to hidden truncation. Hence, the skewness due to hidden truncation in SN (Azzalini, 1985 model) and the ESN models is captured by λ_2.

5. EGSN model and Sample selection density

The data used in Table 1 arises from sample selection setting, hence the consistent results obtained from the ESN and EGSN models. Model (12) can be reparametrised as the continuous component of a sample selection density having skew-normal distribution features with the substitution $\mu = \beta' x$, $\lambda_0 = \gamma' x / \sqrt{1-\rho^2} \in \mathbb{R}$ and $\lambda = \rho / \sqrt{1-\rho^2} \in \mathbb{R}$ as

$$f(y) = \frac{\frac{2}{\sigma}\phi\left(\frac{y-\beta' x}{\sigma}\right)\Phi\left(\frac{\lambda_1(y-\beta' x)}{\sigma}\right)\Phi\left(\frac{\gamma' x+\rho\left(\frac{y-\beta' x}{\sigma}\right)}{\sqrt{1-\rho^2}}\right)}{F_{SN}\left(\gamma' x; 0, 1, \frac{-\lambda_1\rho}{\sqrt{1+\lambda_1^2-\lambda_1^2\rho^2}}\right)}, \tag{13}$$

where the outcome and the selection models are respectively $Y = \beta' x + \sigma\varepsilon_1$ and $S = \gamma' x + \varepsilon_2$, and

$$\begin{pmatrix}\varepsilon_1\\ \varepsilon_2\end{pmatrix} \sim SN_2\left\{\begin{pmatrix}0\\ 0\end{pmatrix}, \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}, \begin{pmatrix}\lambda_1\\ 0\end{pmatrix}\right\}.$$

SN_2 and x represents the bivariate skew-normal distribution and the covariates in the data, ρ is the correlation between Y and S, and λ_1 is the inherent skewness in the outcome from the population. The link that we have established can easily be used to study the properties of the sample selection model given in (13) as we have done in section 3. For example, equation (11) can be used to derive the conditional expectation of the observed data when the data is skew by using the reparametrisation above. This conditional expectation extends Heckman (1979) two-step method by an additional parameter, λ_1.

6. Concluding Remarks

We have introduced and studied an extended version of the two-parameter generalized skew-normal distribution (EGSN) of Jamalizadeh et. al. (2008). The distribution was derived using hidden truncation on a skew-normal random variable. A link between the model and sample selection models provides additional flexibility in modelling observed data arising from selection, which cannot be captured in the original GSN distribution.

The proposed model can also be considered as an extension of the extended skew-normal (ESN) distribution with an additional skewness parameter. Unlike the ESN model, the EGSN model can capture skewness inherent in the original population from which the observed data is derived. In addition, the parameters in the model have distinct interpretation; the location parameter estimates the theoretical mean for the covariates in the original population. Although the skewness parameters are distinct, we have not yet investigated their possible joint role in regulating skewness in the model.

Acknowledgements

We thank an Associate Editor and a Referee for helpful comments.

References

Arellano-Valle, R. B., Branco, M. D., Genton, M. G., (2006). A unified view of skewed distributions arising from selections. *Canad. J. Statist.* **34**, 581–601.

Arellano-Valle, R. B., del Pino, G., San Martin E., (2002). Definition and probabilistic properties of skew-distributions. *Statist. Probab. Lett.* **58(2)**, 111–121.

Arellano-Valle, R. B., Genton, M. G., (2010). Multivariate extended skew-t distributions and related families. *Metron* **68**, 201–234.

Arnold, B. C., (2009). Flexible univariate and multivariate models based on hidden truncation. *J. Statist. Plann. Inference* **139**, 3741–3749.

Arnold, B. C., Beaver, R. J., (2000a). Hidden truncation models. *Sankhyā Series A* **62**, 23–35.

Arnold, B. C., Beaver, R. J., (2000b). Some skewed multivariate distributions. *Amer. J. Math. Management Sci.* **20**, 27–38.

Arnold, B. C., Beaver, R. J., (2002). Skewed multivariate models related to hidden truncation and/or selective reporting. *Test* **11**, 7–54.

Arnold, B. C., Beaver, R. J., Groeneveld, R. A., Meeker, W. Q., (1993). The nontruncated marginal of a truncated bivariate normal distribution. *Psychometrika* **58**, 471–488.

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

Azzalini, A., (2014a). The R sn package : The skew-normal and skew-t distributions (version 1.0-0). Università di Padova, Italia. `http://azzalini.stat.unipd.it/SN`.

Azzalini, A., (2014b). *The Skew-Normal and Related Families*. IMS Monographs series, Cambridge University Press, Cambridge.

Birnbaum, Z. W., (1950). Effect of linear truncation on a multinormal population. *Ann. Math. Statist.* **21**, 272–279.

Copas, J. B., Li, H. G. (1997). Inference for Non-random Samples. *J. Roy. Statist. Soc. B* **59**, 55–95.

Heckman, J., (1979). Sample selection bias as a specification error. *Econometrica* **47**, 153–161.

Jamalizadeh, A., Behboodian, J., Balakrishnan, N., (2008). A two-parameter generalized skew-normal distribution. *Statist. Probab. Lett.* **78**, 1722–1728.

Marshall, A. W., Olkin, I., (2007). *Life Distributions: Structure of Nonparametric, Semiparametric, and Parametric Families*. Springer, New York.

Mudholkar, G. S., Hutson, A. D., (2000). The epsilon-skew normal distribution for analyzing near-normal data. *J. Statist. Plann. Inference* **83**, 291–309.

Ogundimu, E. O., Hutton, J. L. (2014). A unified approach to multilevel sample selection model. *Comm. Statist. Theory Methods*. (In Press).

O'Hagan, A., Leonard, T. (1976). Bayes estimation subject to uncertainty about parameter constraints. *Biometrika* **63**, 201–203.

Roberts, C., (1966). A correlation model useful in the study of twins. *J. Amer. Statist. Assoc.* **61**, 1184–1190.

Weinstein, M. A., (1964). The sum of values from a normal and a truncated normal distribution. *Technometrics* **6**, 104–105.

Wooldridge, J. M., (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Massachusetts.

LaTeX Souce Files

Click here to download LaTeX Souce Files: Latex source files.zip